

Phenotypic characterization of elite quality protein maize (QPM) inbred lines adapted to tropical-highlands and the association studies using SSR markers

Demissew Ababulgu^{1,2*}, Hussein Shimelis¹, Mark Laing¹, Amelework Beyene¹

¹African Center for Crop Improvement, University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, Pietermaritzburg, South Africa

²Ethiopian Institute of Agricultural Research, Plant Protection Research Center, P.O. Box 37, West Shoa, Ambo, Ethiopia

*Corresponding author: d.abakemal@gmail.com

Abstract

A well-characterized germplasm is a requisite to develop productive and better adapted crop cultivars. Morphological and molecular markers and pedigree analyses are widely used in germplasm characterization, and to establish genetic diversity and relationship in maize. The objectives of this study were to characterise newly developed quality protein maize (QPM) inbred lines adapted to tropical-highlands using phenotypic traits and to determine the association with simple sequence repeat (SSR) markers. Thirty-six maize inbred lines (30 QPM and six non-QPM) adapted to tropical-highlands of Ethiopia were evaluated using 18 phenotypic traits and 26 selected SSR markers. Significant phenotypic variations were observed among inbred lines for all measured traits. Grain yield showed moderate and high genotypic and phenotypic coefficients of variation, respectively. Ear height and thousand kernel weight had high heritability (≥ 0.70) which could be a source of high genetic advance through selection. Principal component and cluster analysis using unweighted paired group method using arithmetic averages (UPGMA) revealed the presence of two to four distinct genetic groups, while population structure analysis based on phenotypic data showed the presence of four to six genetic groups. Seven QPM inbred lines (KIT32Q, 142-1eQ, SRSYN20Q, FS67(BC₂), FS170Q, FS60, and F7215) with complementary phenotypic traits and relatively better yield performance were selected using the phenotypic traits and SSR based genotyping for cultivar development and/or conservation.

Keywords: genotype; highland maize; phenotypic traits; QPM; SSR markers.

Abbreviations: CML_CIMMYT maize line; ECA_eastern and central Africa; masl_meters above sea level; NARS_National Agricultural Research System; SSR_simple sequence repeat; UPGMA_Unweighted Pair Group Method using Arithmetic Averages.

Introduction

Maize (*Zea mays* L.; $2n=2x=20$) is an important food security crop in the developing world, especially in sub-Saharan Africa (SSA) and Latin America. Farmers in the tropical highlands are yet growing unimproved local maize varieties, which are generally low yielders with long maturity period, tall plant height and susceptible to lodging. In addition, these varieties show narrow genetic bases and low variability for economic traits (Bjarnason, 1994). Limited researches have been conducted on tropical highland maize when compared to lowland tropical, subtropical, mid-altitude or temperate types. Recently, some public and private breeding programs are developing highland maize varieties due to the socio-economic importance of the crop in the highland agro-ecologies. Consequently, development and production of highland maize genetic resources are considerably expanding in the tropical highland agro-ecologies in sub-Saharan Africa.

The highland maize breeding program was established in 1998 in Ethiopia in partnership between the International Maize and Wheat Improvement Center (CIMMYT) and the National Agricultural Research Systems (NARS) of eastern

and central Africa (ECA) countries. This allowed the development and introductions of a large number of early generation maize lines with general adaptation to highland agro-ecologies (Krivaneck et al., 2007). Through this initiative, the highland non-QPM maize lines were also converted to QPM using backcrosses of the non-QPM local lines with QPM donor parents introduced from CIMMYT. The converted QPM inbred lines need to be systematically characterized using phenotypic and molecular markers for effective utilization in hybrid breeding programs to exploit heterosis. Parental selection based on individual performance, adaptability and yield stability have been used as the major selection criteria to predict hybrid performance (Bertan et al., 2007), though it demands field evaluations of large numbers of inbreds and hybrids in multi-environment trials. On the other hand, the use of molecular markers has been proposed as a more efficient method for selecting inbred lines and can also reduce the number of multi-location trials (Menkir et al., 2004; Barata and Carena, 2006). A well-characterized QPM germplasm would greatly contribute in cultivar development with enhanced

productivity and better adaptation to the target growing environments. Also, availability and maintenance of genetic variability is crucial to widen the genetic bases for effective breeding. Morphological and molecular markers and pedigree analyses are widely used in germplasm characterization and to establish genetic diversity and relationship in crop plants (Warburton et al., 2005). Morpho-agronomic traits are among the earliest markers widely used in germplasm characterization, management and utilization (Smith and Smith, 1992; Karanja et al., 2009; Sow et al., 2014). Abu-Alrub et al. (2006) used kernel characteristics as the best descriptors followed by ear traits for classifying Peruvian highland maize germplasm. Beyene et al. (2006) reported the presence of considerable genetic diversity among Ethiopian highland maize accessions based on morphological traits and SSR markers. Ranatunga et al. (2009) studied the extent of genetic diversity available in maize inbreds using qualitative and quantitative traits and simple sequence repeat markers.

The QPM inbred lines used in this study were converted versions of the non-QPM counterparts, and characterized using selected SSR markers (Demissew et al., 2015). Several studies have reported the effectiveness of phenotypic traits and molecular markers for estimating genetic diversity in maize germplasm; however, studies conducted on the association of phenotypic and molecular markers for elite QPM lines are limited. Therefore, the objectives of this study were to characterize newly developed elite QPM inbred lines adapted to tropical-highlands using phenotypic traits and to determine its association with SSR based genotyping for cultivar development and systematic conservation.

Results

Across locations analysis of variance for phenotypic traits

Combined analysis of variance showed the significant effects of inbred lines and their interaction with locations for most of the studied traits. The effect of environment (location) was significant on leaf length, number of leaves, ear length, and number of kernels per row (Supplemental Table 2). Combined analyses of variance were not carried out for anthesis date, anthesis-silking-interval, plant height, and ear diameter because of heterogeneity of error variance. Supplemental Table 3 shows the mean performance of lines across the two locations for 14 phenotypic traits.

Phenotypic traits variation among inbred lines

A considerable variation existed among inbred lines for all phenotypic traits considered in the study (Supplemental Table 3). The highest grain yield, recorded for the normal maize check inbred line (FS67-N), was 4.5 t ha^{-1} followed by three QPM lines such as 142-1-eQ (4.4 t ha^{-1}), KIT32Q (4.1 t ha^{-1}), and FS2-3SR (4.0 t ha^{-1}). All the four lines except 142-1-eQ had better plant and ear aspect scores. Line 142-1-eQ was superior in terms of ear length contributing to larger number of kernels per row as compared to other tested lines. Out of the six normal maize inbred lines and their QPM

counterparts, three (FS67-N, FS59-4N, and FS232N) had better grain yields than their QPM versions. However, the two non-QPM maize inbreds (KIT32N and SRSYN20N) were inferior in grain yield, while FS170N was on a par with its QPM counterparts on grain yield. All the six normal maize lines flowered relatively earlier than their QPM versions. Line FS60, with maximum leaf width of 13.1 cm and ear length of 14.4 cm, had a moderate grain yield of 3.4 t ha^{-1} . On the other hand, FS48 had the lowest leaf width but with bigger ear length (14.5 cm) and moderate grain yield (3.6 t ha^{-1}). CML144, CML491 and FS60 had bigger tassel size, relatively wider leaf area, and maximum foliage rating scores.

Variance components and heritability

Table 2 shows estimates of variance components of 18 phenotypic traits recorded from 36 QPM and non-QPM lines tested at two locations. Of the total phenotypic variances (σ_p^2), the proportion of genotypic variances (σ_g^2) was higher than environmental variance (σ_e^2) and the genotype-by-environment interaction variance (σ_{gxe}^2) for most traits except anthesis date, anthesis-silking-interval, and ear diameter. In most cases, variance components were greater than the respective standard errors of the evaluated traits (Table 2).

Estimates on genotypic coefficients of variations (GCV), phenotypic coefficients of variations (PCV), heritability in broad sense (H^2), and genetic advance as percent of the mean are presented in Table 3. Higher GCV values were recorded for plant aspect (22.8%) followed by thousand kernel weight (22.0%), and tassel size (21.9%). Traits such as ear aspect (17.9%), ear height (17.2%), grain yield (16.1%), leaf area (16.1%), and foliage rating (16.0%) had moderate GCV values. Lower GCV values were recorded for anthesis date (3.2%), anthesis-silking-interval (3.3%), ear diameter (5.9%) and rows per ear (9.9%). The highest PCV estimates were recorded for anthesis-silking-interval (56.3%), plant aspect (31.1%), tassel size (26.3%), grain yield (25.6%), ear aspect (25.5%), thousand kernel weight (24.7%), and foliage rating (20.6%). Conversely, low PCV values were observed for anthesis date (4.9%) and ear diameter (8.8), while moderate values were recorded for leaf area (19.7%), ear height (19.0%), and kernels per row (18.0%).

Broad sense heritability (H^2) estimates ranged from 0.4% for anthesis-silking-interval to 82% for ear height. Genetic advance as percent over the mean varied between 0.41% for anthesis-silking-interval and 40.29% for thousand kernel weight (Table 3). Most traits had heritability values greater than 50%. The traits with low heritability values (<0.50) include grain yield, anthesis date, anthesis-silking-interval and ear diameter. Anthesis date, anthesis-silking-interval, and ear diameter had low genetic advance estimates.

Ear height, tassel size, and thousand kernel weight had high heritability (≥ 0.70) expressing high genetic advance as percent over the mean at 5% selection intensity. High heritability and moderate genetic advance were exhibited by plant height, leaf width, leaf number, and ear length.

Table 1. Pedigree and protein profile of the 36 QPM and non-QPM inbred lines used in the study.

S/N	Name	Pedigree	Source	*Protein %	Tryptophan %	Quality Index (%)	Heterotic group [†]	Remark
1	142-1-eQ	Unknown (derived from Ecuador-573)	ETHIOPIA	9.44	0.092	0.98	Ecuador	
2	CML144	Pob62c5HC182-2-1-2-B-B-3-1-#-#	CIMMYT	10.82	0.081	0.75	B	
3	CML176	(P63-12-2-1/P67-5-1-1)-1-2-B-B	CIMMYT	10.48	0.065	0.62	Unknown	
4	CML491	(6207QB/6207QA)-1-4-#-2-2-B-B	CIMMYT	10.09	0.083	0.82	A	
5	F7215Q	Unknown (derived from Kitale Syn. II)	ETHIOPIA	10.83	0.055	0.51	Kitale	
6	FS111	[POOL9Ac7-SR(BC2)]FS111-6-1-1-2-1-#-#	CIMMYT-EIAR	10.33	0.078	0.76	Ecuador	
7	FS112	[POOL9Ac7-SR(BC2)]FS112-4-2-1-1-2-#-#	CIMMYT-EIAR	10.71	0.060	0.56	Unknown	
8	FS151-3SR	[POOL9Ac7-SR(BC2)]FS151-3SR-1-2-1-1-#-#	CIMMYT-EIAR	11.44	0.071	0.62	Pool 9A	
9	FS170N	[POOL9Ac7-SR(BC2)]FS170-2-1-3-2-2-1-#-#	CIMMYT-EIAR	—	—	—	Unknown	Non-QPM
10	FS170Q	[POOL9Ac7-SR(BC2)]FS170-2-1-3-1-#-#	CIMMYT-EIAR	12.81	0.075	0.59	Unknown	
11	FS211-1SR	[POOL9Ac7-SR(BC2)]FS211-1SR-1-1-1-#-#	CIMMYT-EIAR	8.72	0.068	0.78	Kitale	
12	FS232N	[POOL9Ac7-SR(BC2)]FS232-4-1-3-1-2-1-3-#-#-#-#	CIMMYT-EIAR	—	—	—	Pool 9A	Non-QPM
13	FS232Q	[POOL9Ac7-SR(BC2)]FS232-4-1-3-1-#-#	CIMMYT-EIAR	10.17	0.068	0.67	Pool 9A	
14	FS2-3SR	[POOL9Ac7-SR(BC2)]FS2-3SR-2-1-2-#-#	CIMMYT-EIAR	9.32	0.064	0.69	Unknown	
15	FS4-3SR	[POOL9Ac7-SR(BC2)]FS4-3SR-1-1-1-#-#	CIMMYT-EIAR	11.30	0.085	0.75	Unknown	
16	FS45	[POOL9Ac7-SR(BC2)]FS45-3-2-2-1-#-#	CIMMYT-EIAR	11.51	0.091	0.79	Ecuador	
17	FS48	[POOL9Ac7-SR(BC2)]FS48-1-1-1-1-1-#-#	CIMMYT-EIAR	9.24	0.080	0.87	Kitale	
18	FS48-1SR	[POOL9Ac7-SR(BC2)]FS48-1SR-2-1-2-1-#-#	CIMMYT-EIAR	10.85	0.085	0.78	Kitale	
19	FS59-2	[POOL9Ac7-SR(BC2)]FS59-2-2-1-1-#-#	CIMMYT-EIAR	11.26	0.067	0.59	Kitale	
20	FS59-4N	[POOL9Ac7-SR(BC2)]FS59-4-1-2-1-1-#-B-B-B-B	CIMMYT-EIAR	10.03	0.047	0.47	Ecuador	Non-QPM
21	FS59-4Q	[POOL9Ac7-SR(BC2)]FS59-4-1-2-1-1-#-#	CIMMYT-EIAR	10.78	0.066	0.61	Ecuador	
22	FS60	[POOL9Ac7-SR(BC2)]FS60-2-1-1-1-#-#	CIMMYT-EIAR	11.70	0.065	0.56	Pool 9A	
23	FS67(BC1)	[POOL9Ac7-SR(BC2)]FS67-1-2-1-1-1-#-#	CIMMYT-EIAR	9.72	0.078	0.81	Kitale	
24	FS67(BC2)	[POOL9Ac7-SR(BC2)]FS67-1-2-3-1-#-#	CIMMYT-EIAR	10.73	0.065	0.61	Kitale	
25	FS67-N	[POOL9Ac7-SR(BC2)]FS67-1-2-3-1-#-B-B-B-B	CIMMYT-EIAR	10.14	0.048	0.47	Kitale	Non-QPM
26	FS68(BC1)	[POOL9Ac7-SR(BC2)]FS68-1-1-2-1-1-#-#	CIMMYT-EIAR	11.63	0.067	0.57	Kitale	
27	FS68(BC2)	[POOL9Ac7-SR(BC2)]FS68-1-1-2-1-1-#-#	CIMMYT-EIAR	11.65	0.056	0.48	Kitale	
28	KIT12	[KIT/SNSYN[N3/TUX]]c1F1-##(GLS=1)-12-2-1-#-#	CIMMYT-EIAR	10.17	0.072	0.70	Ecuador	
29	KIT29	[KIT/SNSYN[N3/TUX]]c1F1-##(GLS=2)-29-35-2-3-#-#	CIMMYT-EIAR	7.28	0.070	0.96	Unknown	
30	KIT31	[KIT/SNSYN[N3/TUX]]c1F1-##(GLS=1.5)-31-17-1-1-#-#	CIMMYT-EIAR	8.95	0.059	0.66	Unknown	
31	KIT32N	[KIT/SNSYN[N3/TUX]]c1F1-##(GLS=2.5)-32-1-1-1-#-#-#-#	CIMMYT-EIAR	—	—	—	Ecuador	Non-QPM
32	KIT32Q	[KIT/SNSYN[N3/TUX]]c1F1-##(GLS=2.5)-32-1-1-1-#-#-#-#	CIMMYT-EIAR	9.71	0.075	0.77	Ecuador	
33	KIT34	[KIT/SNSYN[N3/TUX]]c1F1-##(GLS=2.5)-34-2-1-1-#-#-#-#	CIMMYT-EIAR	10.39	0.102	0.99	Ecuador	
34	SRSYN20N	SRSYN95[ECU//SC/ETO]F1-##(GLS=3.5)-20-2-1-1-#-#-#-#	CIMMYT-EIAR	—	—	—	Pool 9A	Non-QPM
35	SRSYN20Q	SRSYN95[ECU//SC/ETO]F1-##(GLS=3.5)-20-2-1-1-#-#-#-#	CIMMYT-EIAR	12.06	0.069	0.57	Pool 9A	
36	SRSYN48	SRSYN95[ECU//SC/ETO]F1-##(GLS=3.5)-48-1-1-1-#-#-#-#	CIMMYT-EIAR	10.12	0.077	0.76	Ecuador	

[†] Putative heterotic grouping based on phenotypic data of the non-QPM counterparts before conversion to QPM

*Protein quantity and quality of the inbred lines were determined in 2011 at the CIMMYT Maize Nutrition Quality and Plant Tissue Analysis Laboratory following procedures described by Nuritt et al. (2009). Whole grain samples were used to determine concentrations of nitrogen and tryptophan. Percent tryptophan (% Trp) was determined by the colorimetric method based on glyoxalic acid while percent nitrogen (% N) was determined by Micro-Kjeldahl methodology. Protein content was determined based on the formula for maize: Protein = %N*6.25. Quality Index (QI) was calculated as QI = (%Trp / %Protein)*100.

Table 2. Estimates of components of variances and their standard errors from pooled data for 18 phenotypic traits of the 36 inbred lines evaluated at Ambo and Kulumsa, Ethiopia, in 2013.

Trait	$\sigma^2_g \pm SE$	$\sigma^2_e \pm SE$	$\sigma^2_{gxe} \pm SE$	$\sigma^2_p \pm SE$
Grain yield	0.25 ± 0.10	0.20 ± 0.03	0.18 ± 0.07	0.63 ± 0.20
Anthesis date	10.90 ± 4.46	3.55 ± 0.61	11.26 ± 3.14	25.71 ± 8.21
AnthesisSilking Interval	0.01 ± 0.52	0.60 ± 0.10	2.78 ± 0.73	3.39 ± 1.35
Ear aspect	0.25 ± 0.09	0.08 ± 0.01	0.18 ± 0.05	0.51 ± 0.15
Plant aspect	0.38 ± 0.12	0.18 ± 0.03	0.15 ± 0.06	0.71 ± 0.21
Ear height	143.75 ± 37.01	18.42 ± 3.11	12.14 ± 5.33	174.31 ± 45.48
Plant height	284.86 ± 77.62	32.91 ± 5.55	59.93 ± 18.3	377.70 ± 101.47
Leaf length	43.30 ± 11.15	15.22 ± 2.54	0.00 ± 2.00	58.52 ± 15.69
Leaf width	1.48 ± 0.39	0.28 ± 0.05	0.24 ± 0.09	2.00 ± 0.53
Leaf area	6154.0 ± 1654.0	3055.0 ± 521.0	0.00 ± 433	9209.0 ± 2608.0
Tassel size	1.29 ± 0.35	0.51 ± 0.09	0.05 ± 0.09	1.85 ± 0.53
Number of leaves	0.38 ± 0.10	0.08 ± 0.01	0.03 ± 0.02	0.49 ± 0.13
Foliage rating	0.67 ± 0.20	0.25 ± 0.04	0.19 ± 0.08	1.11 ± 0.32
Ear length	2.12 ± 0.58	0.39 ± 0.07	0.40 ± 0.15	2.91 ± 0.80
Ear diameter	0.05 ± 0.02	0.06 ± 0.01	0.00 ± 0.01	0.11 ± 0.04
Number of rows/ear	1.46 ± 0.40	0.50 ± 0.09	0.16 ± 0.11	2.12 ± 0.60
Number of kernels per row	10.92 ± 3.32	4.70 ± 0.78	3.05 ± 1.35	18.67 ± 5.45
Thousand kernel weight	3165.75 ± 857.8	86.1 ± 14.85	756.62 ± 191.6	4008.47 ± 1064.25

σ^2_e = error variance; σ^2_g = genotypic variance; σ^2_p = phenotypic variance; σ^2_{gxe} = genotype x environment variance; SE = standard error

Table 3. Estimates of genotypic (GCV) and phenotypic (PCV) coefficients of variation, broad sense heritability (H^2) and genetic advance (GA) based on pooled data for 18 phenotypic traits of the 36 maize inbred lines evaluated at Ambo and Kulumsa, Ethiopia, in 2013.

Trait	GCV (%)	PCV (%)	H^2	GA	GA (% of Mean)
Grain yield	16.1	25.6	0.40	0.65	20.96
Anthesis date	3.2	4.9	0.42	4.43	4.26
Anthesis/Silking Interval	3.3	56.3	0.004	0.01	0.41
Ear aspect	17.9	25.5	0.49	0.72	25.79
Plant aspect	22.8	31.1	0.54	0.93	34.35
Ear height	17.2	19.0	0.82	22.46	32.28
Plant height	11.3	13.0	0.75	30.24	20.20
Leaf length	10.0	11.4	0.74	11.68	17.37
Leaf width	12.6	14.6	0.74	2.16	22.35
Leaf area	16.1	19.7	0.67	132.30	27.10
Tassel size	21.9	26.3	0.70	1.96	37.76
Number of leaves	11.5	13.1	0.78	1.12	20.89
Foliage rating	16.0	20.6	0.60	1.31	25.67
Ear length	11.8	13.8	0.73	2.56	20.78
Ear diameter	5.9	8.8	0.45	0.31	8.21
Number of rows/ear	9.9	11.9	0.69	2.07	16.90
Number of kernels per row	13.8	18.0	0.58	5.21	21.71
Thousand kernel weight	22.0	24.7	0.79	103.15	40.29

Table 4. Principal components (PCs) for 12 phenotypic traits of the 36 maize inbreds evaluated in 2013 at Ambo and Kulumsa, Ethiopia.

Variables	Eigenvectors/principal components			
	PC1	PC2	PC3	PC4
Grain yield	0.383	0.255	0.188	-0.046
Ear aspect	-0.368	-0.214	0.010	0.145
Plant aspect	-0.399	0.043	0.046	0.201
Ear height	0.348	0.138	-0.353	0.222
Plant height	0.277	0.112	-0.392	0.316
Leaf length	0.161	-0.345	-0.068	0.640
Leaf width	0.251	-0.381	0.141	-0.421
Leaf area	0.288	-0.486	0.051	0.066
Foliage rating	0.159	-0.497	0.076	-0.106
Ear length	0.256	0.220	0.457	0.116
Number of kernels per row	0.213	0.195	0.514	0.192
Thousand kernel weight	0.229	0.149	-0.422	-0.368
Eigen value	3.859	2.548	2.160	1.119
Individual variation explained (%)	32.2	21.2	18.0	9.3
Cumulative variation explained (%)	32.2	53.4	71.4	80.7

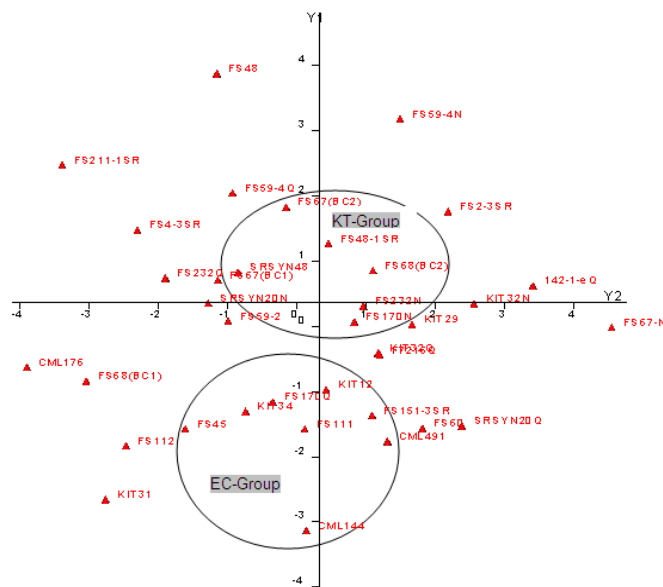


Fig 1. Scatter plot of PC1 (32.2%) and PC2 (21.2%) of the 36 QPM and non-QPM inbred lines based on 12 correlated phenotypic traits combined over the two locations.

positive weights, and leaf width and thousand kernel weights, with negative weights, were associated with the fourth PC (PC4) that explained 9.3% of the total variation. Based on principal component scores using the two most important principal components, PC1 and PC2, a bi-plot was constructed providing a recognizable cluster of the genotypes. From the bi-plot, two major heterotic groups (the Ecuador group denoted by the symbol 'EC' and Kitale by 'KT') represented half of the tested inbred lines, whereas the remaining 50% did not show defined groups (Fig.1).

Cluster analyses using phenotypic and SSR data

The dendrogram constructed using UPGMA clustering based on phenotypic data classified the test genotypes into three major groups (Groups I to III) and seven outlier inbreds allocated in Group IV (Fig.2A). The estimated cophenetic correlation value was poor fit ($r_{cop}= 0.63$). Outlier inbreds were identified for having either highest or lowest data values for one or more traits. The dendrogram constructed using the UPGMA clustering algorithm based on the SSR data allocated test inbred lines into three major groups (Fig 2B). These groupings, in some cases, revealed associations related to their pedigree records. The cophenetic correlation value ($r_{cop}= 0.80$) observed for the molecular data was in the range of good fit. The association analysis between phenotypic and molecular markers exhibited the presence of positive but non- significant correlation between the two marker systems ($r= 0.032$, $P= 0.700$).

The phenotypic dendrogram, which demonstrated the clustering patterns of the 36 inbred lines (Fig 2A), showed that Group III was dominated by three inbred lines belonging to Kitale heterotic group and three lines from other heterotic groups. The lines in cluster III were characterized by short ear heights with average grain yield, number of leaves above the ear, ear diameter, and ear aspect. Group II consisted of 50% of the inbred lines which mainly belong to Ecuador heterotic group expressing high grain yield, good ear and plant aspects, and good ear length and diameter. This cluster also consisted of lines from Pool9A and Kitale heterotic groups, and previously uncategorized lines. Group I, on the other hand, contained five lines (each from Kitale and Ecuador heterotic group, and three unrecognized lines) that were characterized by low grain yield, increased anthesis-silking-interval, and short ear height and ear length.

The dendrogram based on 26 SSR markers revealed four genetic groups (designated as Group I, II, III and IV) consisting of 5, 14, 14 and 3 inbred lines, respectively (Fig 2B). Three inbred lines, which were not previously categorized under the known heterotic groups, are now classified under Group IV. Group III consisted of 14 inbreds of which six lines were from Ecuador heterotic group, four from Kitale heterotic group, one from Pool9A heterotic group, one from previously uncategorized lines, and two of the CMLs series (CML144 and CML491). Three lines which were clustered under Group III are non-QPM counterparts. Similarly, 14 inbred lines were clustered in Group II, which comprised of five lines extracted from Kitale heterotic group, four from Ecuador heterotic group, another four from Pool9A heterotic group, and one previously uncategorized inbred line. Three lines allocated in Group II are non-QPM counterparts of the QPM converted inbred lines. Two previously uncategorized inbreds, and one from Kitale and

one from Pool9A heterotic group, and CML176 are clustered under Group I.

The population structure of the 36 highland inbred lines was analyzed by employing STRUCTURE software utilizing the Bayesian clustering approach that divided the population into sub-groups (Fig 3 and Supplemental Table 4). The assumed values of probable sub-groups (K) were determined by choosing higher ΔK value, an ad hoc quantity related to the second order change of the log probability of data with respect to the number of clusters inferred by Structure (Pritchard et al., 2000). From the population structure analysis of the phenotypic traits, it was possible to determine the optimal number of genetic groups using the method proposed by Evanno et al. (2005). According to Evanno table output, the $K = 4$ and $K = 6$ were observed to be the possible population numbers due to the high ΔK peak value of 2.85 and 3.10, respectively, among the assumed K (Fig 3D). At $K= 4$ the clustering pattern followed the heterotic grouping, whereas at $K=7$ the clustering seemed to relate with pedigree relationships.

Discussion

Knowledge on genetic diversity and relationships present in maize germplasm is a pre-requisite for effective breeding or conservation. Genetic diversity and interrelationship analysis is a foundation for selection of parents for crossing, and to enhance the genetic bases of breeding programs. According to FAO's report, an additional 60 million tonnes of maize grain will be needed in the tropics by 2030 from the annual global harvest (Paliwal et al., 2000). To meet this demand, conservation and efficient utilization of suitable genetic diversity is crucial.

This study revealed the presence of considerable phenotypic and genotypic variability among elite tropical highland adapted QPM inbred lines evaluated using agromorphological traits and SSR markers (Tables 2 to 4; Fig 1 to 3). Beyene et al. (2006) observed high genetic diversity among the traditional Ethiopian highland maize accessions using both morpho-agronomic traits and molecular markers. Legesse et al. (2009) reported significant genetic variability in the combining ability and heterotic response of highland maize inbred lines using agronomic traits. Kumar et al. (2012) detected high genetic diversity among 91 maize genotypes using morpho-physiological traits and SSR markers.

The broad range of phenotypic variations detected in the QPM lines present a great opportunity for the development of improved open-pollinating or hybrid varieties of QPM adapted to the highland environments. The combined analysis of variance showed highly significant location x line interaction for most of the tested traits. This suggested the differential performance of the lines across the two locations. It was interesting to note that out of the six normal maize inbred lines included in the study with the QPM lines, three had better grain yield and one with equivalent yield performance and two had lower grain yield values than their QPM converted counterparts. The present result conforms to previous report by (Atlin et al., 2011) who reported the possibility of developing agronomically better QPM genotypes than the normal maize counterparts. Therefore, QPM cultivars can be promoted and adopted in sub-Saharan Africa given their superior agronomic performance which can be similar or better than the non-

QPM counterparts. This will enhance farmers' acceptance and creates market potential of QPM cultivars providing improved protein quality due to their enhanced lysine and tryptophan contents (Krivanek et al., 2007).

Phenotypic variance and heritability are direct indicators of the response to selection (Falconer, 1960). In the studied inbred lines, the proportions of genotypic variances (σ^2_g) were higher than both environmental variances (σ^2_e) and genotype x environment interactions (σ^2_{gxe}) for most of the studied traits. Therefore, the newly developed elite QPM lines are useful genetic resources for hybridization and selection programs.

Phenotypic traits evaluated in the present study such as thousand kernel weight, ear height, and tassel size exhibited relatively high genotypic coefficient of variation and high heritability and genetic advance estimates. Selection of the top 5% of the inbred lines with higher thousand kernel weight may lead to increased thousand kernel weight by 40.3%. Dagne (2008) reported that QPM line development should focus on thousand kernel weight as selection criterion to increase grain yield. However, heritability and selection response were low for days to anthesis, anthesis-silking-interval, and ear diameter. Shanthi et al. (2011) reported low broad sense heritability values and genetic advance as percent of the mean for days to 50% tasseling and silking, protein and oil contents.

Principal component analysis adequately represented the studied traits into four principal components. The cumulative variation explained by the four PCs (80.7%) in the current study is greater than reported by (Alika et al., 1993) who indicated 71.8% of total variation in 62 traditional Ethiopian highland maize accessions that was represented by the first four PCs. Similarly, Dagne (2008) found a comparable result of 78% total variation explained by the first six PCs when evaluating 35 CIMMYT developed QPM inbred lines. Alika et al. (1993) reported the major role of morphological traits in ascribing phenotypic variation.

Dendrograms constructed using the phenotypic traits and the SSR markers classified the test inbred lines into four genetic groups for systematic selection (Fig. 2A and Fig. 2B). Similarly, four and seven genetic groups were also recognized from population structure analysis using the phenotypic data (Fig 3). The lack of clear cut in the number of populations estimated may partly be due to few number of phenotypic markers used in this study and the peculiarities of many phenotypic traits and lack of common phenotypic assaying methods across environments (Amelework et al., 2016). In general, a distinct population structure was observed among the studied highland maize inbred lines. A model-based population structure analysis using similar inbred lines and SSR markers with the current study revealed the presence of three genetic groups, which is generally consistent with pedigree information and partly with heterotic grouping (Demissew et al., 2015).

Although the diversity analyses based on phenotypic or molecular markers resulted in a similar number of distinct groups and a similar concentration of genotypes in each group, the correlation between the two markers systems was low. In agreement to this finding, Dagne (2008) reported a low correlation coefficient ($r= 0.09$) between morphological and SSR-based genetic distance measures among 35 QPM inbred lines using 40 SSR markers. Karanja et al. (2009) found a correlation coefficient of 0.2323 with p

value of 0.012 between SSR and morphological data in assorted maize inbred lines. Amelework et al. (2016) also reported that there was no significant relationship between molecular and morphological distances in a study conducted using 30 SSRs and 36 sorghum genotypes. Lack of significant association between the phenotypic and SSR data could in part be attributed to the relatively small number SSRs used in this study and the molecular markers did not adequately sample the genomic regions that are responsible for the phenotypic variation among the inbreds (Alves et al., 2013). In addition, several factors such as the distribution of markers in the genome, the number of markers used and the nature of the evolutionary mechanism underlying the variation measured can affect the genetic distance estimates (Powell et al., 1996). A relatively small number of SSR markers were used because they were selected from the list of markers that had previously been effective and used for genetic characterization of CIMMYT maize germplasm (Warburton et al., 2002). The classification of some of the inbred lines using the two methods followed similar patterns of clustering of the same lines using combining ability analysis and pedigree information by Twumasi-Afriyie et al. (2003). Lucchin et al. (2003) clustered 20 Italian flint maize landraces using 34 morphological and agronomic traits. Wietholter et al. (2008) pointed out the significance of some traits such as plant height, ear insertion, female flowering, male flowering and kernel row number per ear in the classification of Brazilian corn landraces. Abu-Alrub et al. (2006) reported that tassel related traits were found to be less reliable descriptors unlike kernel and ear traits for classifying Peruvian highland maize germplasm. From the foregoing researches, it appears that the contribution of phenotypic traits in characterizing maize germplasm is not well exploited due to the need for intense evaluation of a given genetic pool or population at the target production environments. Therefore, the information generated in this study can possibly serve as a platform for detailed characterization and heterotic grouping with large number of QPM genotypes and molecular markers by making crosses and field evaluations for confirmation of the heterotic groups.

Materials and methods

Germplasm

Thirty-six maize inbred lines were used in this study. The lines comprised of 31 highland adapted maize inbred lines from the highland maize breeding program, two from mid-altitude breeding programs in Ethiopia, and three CIMMYT derived maize lines (CMLs). In Ethiopia, the highland and transitional highland sub-humid maize growing agro-ecologies lie within the altitudinal ranges of 1800-2400 masl, while the mid-altitude sub-humid lies in the ranges of 1000-1800 masl (Kelemu and Mamo, 2002). The details of the tested lines including protein (tryptophan) levels are presented in Table 1. Twenty seven of the test inbred lines are converted elite QPM genotypes developed through a backcross breeding procedure (Demissew et al., 2015).

Experimental sites

Two experiments were conducted under rainfed conditions during the main rainy season (June–September) of 2013 at two locations, Ambo and Kulumsa Agricultural Research Centers, in Ethiopia. Ambo is situated at 8°57' N latitude, 38°07' E longitude with an altitude of 2225 meters above sea level (masl) and Kulumsa at 8°13' N latitude, 39°13' E longitude and 2180 masl. Both locations represent the highland sub-humid maize growing agro-ecology in Ethiopia. The soil type at Ambo is clay (heavy vertisol) with a pH of 7.8 for the most top soil (0–30 cm). The long-term total annual rainfall of this site is 1115 mm, and average minimum and maximum temperatures are 11.7°C and 25.4°C, respectively. Kulumsa has clay soil (eutric vertisol) with a pH of 6.8 at 0–30 cm of soil depth. The long term total annual rainfall at Kulumsa is 824 mm, with average minimum and maximum temperatures of 10°C and 23°C, respectively.

Experimental design and field procedures

The test inbred lines were evaluated using a 12 x 3 alpha lattice design with two replications at each location. In both locations, planting was done during the first week of June 2013. All entries were sown and thinned to one plant per hill after 35 days of emergence giving a population density of 53,333 plants ha⁻¹. Each entry was planted in two rows, each 5.25m long, with intra-row spacing of 0.25 m and inter-row spacing of 0.75 m. The experimental units at both locations received 69 kg ha⁻¹ of phosphate (P₂O₅) in the form of diammonium phosphate fertilizer (DAP) and one-third of 119 kg nitrogen (N) ha⁻¹ in the form of urea as basal dressing during planting. Second and third doses of N (each one-third of 119 kg) were side-dressed 35 days after the crop emergence and before tasseling, respectively. Weeds were controlled manually.

DNA extraction and genotyping

Seedlings of all genotypes were grown in plastic seed trays for 3 weeks in a screen house at the Biosciences for east and central Africa (BecA) hub in Nairobi, Kenya. Leaf tissue from each line was cut into pieces with scissors, and transferred into 1.2 ml strip tubes that contained two 4 mm stainless steel grinding balls. The tissue was freeze-dried (lyophilized) for 3 days using a Labconco freeze dryer (<http://www.labconco.com>). The lyophilized leaf samples were ground into fine powder using GenoGrinder-2000 at 500 strokes per minute for 6 minutes. Genomic DNA was extracted using a modified version of the mini-prep Cetyltrimethyl Ammonium Bromide (CTAB) method of CIMMYT protocol (http://www.generationcp.org/capcorner/chile_wksp_2005/manuals/manual_01.pdf). The quality of the isolated DNA was checked after running aliquots of DNA samples on a 0.8% agarose gel that contained 0.3 µg/ml GelRed (Biotium). DNA concentration was measured using NanoDrop ND-800 Spectrophotometer (Thermo Scientific).

Twenty-six SSR markers selected from the list of markers that had previously been used for genetic characterization of CIMMYT maize germplasm (Warburton et al., 2002) were used for this study. The SSRs were chosen based on prior

information, including chromosomal distribution, minor allele frequency (MAF), polymorphic information content (PIC), and repeat length. Polymerase chain reaction (PCR) was performed in 96-wells plates. Both DNA extraction and SSR genotyping were done at the BecA hub, Kenya.

Data analyses

Observations were made on 18 phenotypic traits according to maize descriptors of International Board for Plant Genetic Resource (IBPGR) and CIMMYT (IBPGR and CIMMYT, 1991). The list of traits used in the study with their abbreviations and description are summarised in Supplemental Table 1.

Analysis of variance (ANOVA) for each experimental site was done separately using the restricted maximum likelihood (REML) (Patterson and Thompson, 1971) analysis in GenStat® Release 14 statistical software (Payne et al., 2007). Inbred lines were assigned as fixed factor, and replicates and incomplete blocks within replicates were random factors. Data for both locations were then pooled after homogeneity of variance tests using Hartley's F_{\max} -test (Hartley, 1950) and combined ANOVA conducted.

Variance components and corresponding standard errors were further estimated to identify genetic variability among lines and to determine the genetic and environmental effects on phenotypic traits. The REML analysis was used to estimate variance components as genotypic, environmental (locations) and their interactions. Additionally, phenotypic variance (σ^2_p), genotypic variance (σ^2_g), environmental variance (σ^2_e) and genotypic variance by environmental variance ($\sigma^2_{g \times e}$) and the error variances (σ^2_e) were calculated (Hallauer et al., 2010). Phenotypic and genotypic coefficients of variations, heritability in the broad sense and genetic advance expressed as percent of the mean (Johnson et al., 1955) were computed using variance components.

Principal components analysis (PCA) was conducted to assess the importance and contribution of the phenotypic traits in explaining variations. Since principal components of a set of traits depend on the scales used to measure the traits, standardization of the measurements was made to ensure that all have equal weight in the analysis (Manly, 2005). Standardization of the data was thus carried out by subtracting from each observation the mean value of the trait and subsequently dividing by its respective standard deviation.

Cluster analyses were carried out using Euclidean dissimilarity matrices on both phenotypic and molecular data to determine phenotypic and genotypic interrelationship among the inbred lines. The sequential agglomerative hierarchical nested cluster analysis (SAHN) method (Sneath and Sokal, 1973) of the NTSYS-pc 2.10 software (Rohlf, 2000) was used to compute Euclidean distances. The distance matrices were further clustered in the form of dendrograms using the Unweighted Pair Group Method using Arithmetic Averages (UPGMA) with average linkage (Sneath and Sokal, 1973). Cophenetic correlation coefficients were estimated using COPH and MXCOMP options of the NTSYS-pc 2.10 to test the goodness of fit of the cluster analyses to the dissimilarity matrices (Mantel, 1967). Association analysis between Phenotypic and SSR markers' data was carried out using GenAlex version 6.5 (Peakall and Smouse, 2007).

The Bayesian genotypic clustering approach of STRUCTURE 2.3.4 (Pritchard et al., 2000) was used to validate the population structure among the inbred lines. An admixture model with independent allele frequencies, without prior population information, was used to simulate the population. This model assumes that the genome of each individual is a mixture of genes originating from K unknown ancestral populations. For joint inference of the population substructure, the model was run for 20 replicate analyses for each K value ranging from 1 to 10, with a burn-in period of 100000 and Markov Chain Monte Carlo (MCMC) steps of 100000 iterations. Evanno et al. (2005) test was performed after 20 simulations per K value to estimate the best K value.

Conclusion

The present study found considerable genetic variability among elite QPM inbred lines adapted to tropical-highland agro-ecologies. Seven inbred lines [KIT32Q, 142-1eQ, SRSYN20Q, FS67(BC₂), FS170Q, FS60, and F7215] with complementary phenotypic traits and relatively better yield performance were selected using phenotypic traits and SSR-based markers. The selected elite QPM lines are useful genetic resources for breeding and conservation.

Acknowledgments

The authors would like to thank the Alliance for a Green Revolution in Africa (AGRA) for providing PhD scholarship for the first author and Generation Challenge Program (GCP) for funding the genotyping work.

References

- Abu-Alrub I, Christiansen JL, Madsen S, Sevilla R, Ortiz R (2006) Assessing tassel, kernel and ear variation in Peruvian highland maize. *Plant Genet Resour Newsl.* 137:34-41.
- Alika JE, Aken'Ova ME, Fatokun CA (1993) Variation among maize (*Zea mays* L.) accessions of Bendel State, Nigeria. Multivariate analysis of agronomic data. *Euphytica* 66:65-71.
- Alves AA, Bhering LL, Rosado TB, Laviola BG, Formighieri EF, Cruz CD (2013) Joint analysis of phenotypic and molecular diversity provides new insights on the genetic variability of the Brazilian physic nut germplasm bank. *Genetics and Molecular Biology* 36: 371-381.
- Amelework B, Shimelis H, Laing M (2016) Genetic variation in sorghum as revealed by phenotypic and SSR markers: implications for combining ability and heterosis for grain yield. *Plant Genetic Resources*; 1-13.
- Atlin GN, Palacios N, Babu R, Twumasi-Afryie S, Friesen DK, Groote HD, Vivek B, Pixley K (2011) Quality protein maize: Progress and prospects. *Plant Breed Rev.* 34:83-130.
- Barata C, Carena MJ (2006) Classification of North Dakota maize inbred lines into heterotic groups based on molecular and testcross data. *Euphytica* 151: 339-349.
- Bertan I, Carvalho FIF, Oliveira AC (2007) Parental selection strategy in plant breeding programs. *J Crop Sci Biotechnol.* 10: 211-222.
- Beyene Y, Botha AM, Myburg AA (2006) Genetic diversity in traditional Ethiopian highland maize accessions assessed by simple sequence repeat (SSR) markers. *Genet Resour Crop Evol.* 53:1579-1588.
- Bjarnason M (ed) (1994) The subtropical, midaltitude and highland maize subprograms. Maize Program Special Report. CIMMYT, Mexico, D.F.
- Dagne W (2008) Genotypic variability and combining ability of quality protein maize inbred lines under stress and optimal conditions. PhD Thesis. University of the Free State, South Africa.
- Demissew A, Hussein S, Derera J, Kassa S (2015) Genetic purity and patterns of relationships among tropical highland adapted quality protein and normal maize inbred lines using microsatellite markers. *Euphytica* 204:49-61.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol.* 14:2611-2620.
- Falconer DS (1960) Introduction to quantitative genetics. Oliver Boyd, Edinburg, London.
- Hallauer AR, Carena MJ, Miranda JB (2010) Quantitative genetics in maize breeding. Springer Science + Business Media, New York.
- Hartley HO (1950) The maximum F-ratio as a short-cut test for heterogeneity of variance. *Biometrika* 37:308-312.
- IBPGR, CIMMYT (1991) Descriptors for maize. International Board for Plant Genetic Resources, Rome and International Maize and Wheat Improvement Center, Mexico City, Rome.
- Johnson HW, Robinson HF, Comstock RE (1955) Estimates of genetic and environmental variability in soybeans. *Agron J.* 47:314-318.
- Karanja J, Amugune NO, Ininda J, Kimatu JN, Danson JW (2009) Microsatellite analysis of the correlation between molecular and morphological traits in assorted maize inbred lines. *Afr Crop Sci J.* 17:133 - 144.
- Kelemu F, Mamo G (2002) Suitable zones for growing maize in Ethiopia. pp. 195-202. In: Nigussie M, Tanner D (eds) Proceedings of the Second National Maize Workshop of Ethiopia, Addis Ababa, Ethiopia, 12-16 November 2001.
- Krivanek AF, Groote HD, Gunaratna NS, Diallo AO, Friesen DK (2007) Breeding and disseminating quality protein maize (QPM) for Africa. *Afr J Biotechnol.* 6:312-324.
- Kumar A, Rakshit A, Mangilipelli NK, Varalaxmi Y, Vijayalakshmi T, Vanaja JM, Yadav SK, Venkateswarlu B, Maheswari M (2012) Genetic diversity of maize genotypes on the basis of morpho-physiological and simple sequence repeat (SSR) markers. *Afr J Biotechnol.* 11:16468-16477.
- Legesse BW, Pixley KV, Botha AM (2009) Combining ability and heterotic grouping of highland transition maize inbred lines. *Maydica* 54:1-9.
- Lucchin M, Barcaccia G, Parrini P (2003) Characterization of a flint maize (*Zea mays* L. convar. mays) Italian landrace. I. morpho-phenological and agronomic traits. *Genet Resour Crop Ev.* 50:315-327.
- Manly BFJ (2005) Multivariate statistical methods: A primer, 3rd edn. CRC Press LLC, Florida.
- Mantel NA (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209-220.
- Menkir A, Melake-Berhan A, The C, Ingelbrecht I, Adepoju A (2004) Grouping of tropical mid-altitude maize inbred lines on the basis of yield data and molecular markers. *Theor Appl Genet.* 108:1582-1590.

- Nurit E, Tiessen A, Pixley KV, Palacios-Rojas N (2009) Reliable and inexpensive colorimetric method for determining protein bound tryptophan in maize kernels. *J Agric Food Chem.* 57:7233-7238.
- Paliwal RL, Granados G, Lafitte HR, Violic AD, Marathe JP (2000) Tropical maize: improvement and production. The Food and Agriculture Organization of the United Nations, Rome, Italy.
- Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545-554.
- Payne RW, Harding SA, Murray DA, Soutar DM, Baird DB, Glaser AI, Welham SJ, Gilmour AR, Thompson R, Webster R (2007) GenStat® Release 14 statistical software for windows. VSN International Ltd., UK.
- Peakall R, Smouse PE (2007) GenAlex 6: genetic analysis in excel. Population genetic software for teaching and research. *Mol Ecol Notes.* 6:288-295.
- Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Dingey C and Rafalski A (1996) The comparison of RLFP, RAPD, and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* 2: 225-238.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Ranatunga MAB, Meenakshisundaram P, Arumugachamy S, Maheswaran M (2009) Genetic diversity analysis of maize (*Zea mays* L.) inbreds determined with morphometric traits and simple sequence repeat markers. *Maydica* 54:113-123.
- Rohlf FJ (2000) NTSYS-pc: numerical taxonomy and multivariate analysis system. ver. 2.1, Exeter Software. Satauket, New York, USA.
- Shanthi P, Satyanarayana E, Suresh-Babu G, Sai-Kumar R (2011) Studies on genetic variability for phenological, yield and quality parameters in quality protein maize (QPM) (*Zea mays* L.). *Crop Research Hisar.* 41:188-191.
- Smith JSC, Smith OS (1992) Fingerprinting crop varieties. *Adv Agron.* 47:85-140.
- Sneath PHA, Sokal RR (1973) Numerical taxonomy. Freeman, San Francisco.
- Sow M, Sido A, Laing M, Ndjiondjop MN (2014) Agromorphological variability of rice species collected from Niger. *Plant Genet Resour C.* 12:22-34.
- Twumasi-Afriyie S, Yihun K, Napir G (2003) Exploitation of combining ability and heterotic response in maize germplasm to develop cultivars for the eastern African highlands. Book of Abstracts. The Arnel R. Hallauer International Symposium on Plant Breeding, 17-22 August 2003. CIMMYT, Mexico D.F.
- Warburton ML, Ribaut JM, Franco J, Crossa J, Dubreuil P, Betran FJ (2005) Genetic characterization of 218 elite CIMMYT maize inbred lines using RFLP markers. *Euphytica* 142:97-106.
- Warburton ML, Xia X, Crossa J, Franco J, Melchinger AE, Frisch M, Bohn M, Hoisington D (2002) Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. *Crop Sci.* 42:1832-1840.
- Wietholter P, Melo-Sereno MJC, Freitas-Terra TD, Anjos-E-Silva SD, Barbosa-Neto JF (2008) Genetic variability in corn landraces from southern Brazil. *Maydica* 53:151-159.