Development of SSR-based sorghum (Sorghum bicolor (L.) Moench) diversity research set of germplasm and its evaluation by morphological traits

著者	Shehzad Tariq, Okuizumi Hisato, Kawase Makoto, Okuno Kazutoshi
journal or publication title	Genetic resources and crop evolution
volume	56
number	6
page range	809-827
year	2009-09
権利	(C) Springer Science+Business Media B.V. 2009
URL	http://hdl.handle.net/2241/103626

doi: 10.1007/s10722-008-9403-1

Title Page

Authors: Tariq Shehzad[1], Hisato Okuizumi[2], Makoto Kawase[2] and Kazutoshi Okuno[1]*

Title: Development of SSR-based sorghum (*Sorghum bicolor* (L.) Moench) diversity research set of germplasm and its evaluation by morphological traits

1. Lab. of Plant Genetics and Breeding Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, 305-8572 Japan
2. Genebank, National Institute of Agrobiological Sciences, Tsukuba, 305-8602 Japan

* Corresponding author
Email; okusan@sakura.cc.tsukuba.ac.jp
Tel: +81-29853-4807
Fax: +81-29853-4807

Abstract

Assessment and utilization of diversity in plant genetic resources is vital for the improvement of plant species. A sorghum diversity research set (SDRS) was developed by using SSR markers. A total of 320 sorghum accessions were selected based on geographic distribution from more than 3500 germplasm accessions comprising Asia (East, Southeast, South and Southwest Asia) and Africa, conserved at NIAS Genebank, Japan. We selected 38 simple sequence repeats (SSR) markers which generated 146 alleles, covering ten chromosomes of sorghum from a three different published SSR linkage map of sorghum. The average percentage of polymorphic loci (P) and gene diversity (He) observed in this study were 82.8 and 0.217 respectively. Genetic analysis showed a positive correlation with geographic pattern of differentiation. Based on SSR assessment, 107 sorghum accessions were selected as diversity research set. There was no significant difference in pattern of genetic spectrum between SDRS and base population. Similarly no greater change was observed for variability parameters (Dice, % P, He) and almost all of the SSR alleles were retained in selected sorghum accessions except for the loss of a single allele at locus *Xtxp287*. SDRS was sown during sorghum sowing season in two replications. Data were recorded on 26 important morphological traits according to the standard sorghum descriptors at Genebank. Analysis of variance showed a highly significance difference among all accessions for all of the traits. Morpho-agronomic traits could not effectively classify the accessions according to geographic origin by using cluster analysis.

Key words SDRS. Microsatellites. Genetic diversity. Linkage map. Morphological traits

Introduction

Sorghum, *Sorghum bicolor* L. Moench is a major staple and fodder crop in tropical/semi tropical Africa and Asia (Doggett 1988). Among other crops, it is ranked as fifth most important crop in the world (FAO, 2004). Sorghum is an important target for plant genomics due to its adaptation to harsh environments, diverse germplasm collection, and relatively small genome size (Menzl et al. 2002). Identification of genetic variation underlying traits important in domestication and improvement of crops is an area of great interest to both evolutionary and applied biologists (Hamblin et al. 2004). The sorghum genome contains 750 Mb of DNA, which is 3 to 4-fold smaller than that of maize (2400 Mb) (Arumuganathan and Earle, 1991). In sorghum the most important agronomic form is S.bicolor ssp. bicolor, which is further classified into five races; bicolor, caudatum, durra, guinea and kafir, based on morphological characteristics of the inflorescence (Harlan and de Wet 1972).

Genetic diversity plays a vital role in the success of any breeding program (Ali et al. 2007). Molecular markers are an excellent tool for the assessment of genetic relationships and various types of markers have been used in many investigations of sorghum (Ritter et al. 2007). Many researchers have used different kinds of molecular markers e.g Restriction Fragment Length Polymorphisms (RFLPs) (Ahnert 1996; Deu et al. 1994; Tao et al. 1993), Randomly Amplified Polymorphic DNA (RAPDs) (Ayana et al. 2000; Tao et al. 1993; Uptmoor et al. 2003), microsatellites (SSRs) (Ali et al. 2007; Casa et al. 2005; Anas and Yoshida 2004; Folkertsma et al. 2005; Menz et al. 2004; Smith et al. 2000; Uptmoor et al. 2003) and Amplified Fragment Length Polymorphisms (AFLPs) (Menz et al.2004; Uptmoor et al. 2003) have all been successfully used to estimate the genetic diversity in sorghums germplasm. Morphological traits or pedigree are important estimates of genetic diversity in crops but such information is usually found unrealistic and mostly influenced by environmental factors (Fufa et al. 2005; Almanza-Pinzon et al. 2003). So molecular markers are supposed to be the most effective and efficient genetic tools for the actual estimate of genetic diversity.

Core collection is one of the most important practices of any Genebank management because of the difficulty to handle large number of germplasm for any collection. A core collection is a limited set of accessions representing, with a minimum of repetitiveness, the genetic diversity of a crop species and its wild relatives (Frankel 1984). The core collection concept has been around since 1984 and so far a large number of core collections have been created of different kinds of species. Multivaraite analysis is widely used approach to create groups of similar accessions. Data on genetic markers, agro-morphological characteristics or other characters can be used to construct a dendrogram and to group accessions, using a range of different cluster, discriminant or principal components analysis methods. Biochemical analysis and molecular markers assessment will certainly help to improve core collections. Similarly geographic information systems (GIS) will further analyze the patterns of distribution of genetic diversity in proper way (Guarino et al. 2001).

Mann et al. (1983) hypothesized that the origin and early domestication of sorghum took place in northeastern Africa however Wendorf et al. (1992) reported that carbonized seeds of sorghum, with consistent radiocarbon dates of 8000 years BP, were excavated at an early Holocene archaeological site E-75-6, at Nabta Playa, near the Egyptian-Sudanese border. It subsequently spread to eastern and southern Africa and by trade routes to Asia (Doggett 1988). Plant genetic resources for food and agriculture are crucial in feeding the world's population. They are the raw material that farmers and plant breeders use to improve the quality and productivity of our crops. Different organization and agencies tried to conserve the genetic resources of sorghum germplasm since its domestication. National Plant Germplasm System (USDA), International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) and Chinese Agricultural Science Academy National Seed Bank (CASANSB) has largest sorghum germplasm collections of 42221, 36719 and 14014 accessions, respectively.

Dahlberg et al. (2004) made a core collection from NPGS, USDA sorghum collection. The collection was broken out into its respective country of origin and random selection of 10% of the accessions from each country was used to develop the core. Based on these criteria, a core collection of 3011 accessions representing 77 different countries was developed. Similarly, ICRISAT has a core collection of 2246 sorghum accessions (10% of whole population) developed by Grenier et al. (2001). All of the accessions in whole population were stratified into different groups on the basis of geography and taxonomy. The effect morphological traits, including days to flowering and plant height on day-length

variation was evaluated. These two characters were highly correlated with latitudinal and racial distributions of landraces. Qingshan et al. (2001) has selected sorghum accessions from a large population conserved at CASANSB, China. The accessions were classified according to agronomic, nutrition, biotic and abiotic stress evaluation data. The accessions were comprised of mostly local, photoperiod insensitive, tall and big accessions.

The objectives of the study presented are i) detailed assessment of genetic diversity among 320 selected accessions, ii) selection of accessions for diversity research set iii) to identify the patterns of diversity according to geographic distribution and iv) to find out the relationship between molecular markers estimates and morphological traits analysis. The representative set of diverse sorghum germplasm is designated as Sorghum Diversity Research Set (SDRS), a pioneer work based on germplasm conserved at NIAS Genebank, Japan. It will improve the conservation and use of sorghum genetic resources maintained by Genebank and this information can be effectively utilized by researchers in sorghum breeding programs.

Materials and Methods

Plant materials and genomic DNA isolation

A total of 320 sorghum accessions (landraces) were selected from more than 3500 germplasm collection preserved at Genebank, National Institute of Agrobiological Science, Japan. Sorghum germplasm in Japan's national genebank was collected in a collaborative project with other national and international agencies. Our selection was based on the geographic distribution of sorghum over the regions around globe, particularly from Asia and Africa. Among Asian accessions, 103 were from East Asia (Japan, Korea, Taiwan and China), two from Southeast Asian (Cambodia and Myanmar), 99 from South Asia (India, Pakistan, Afghanistan, Bangladesh and Nepal) and two from Southwest Asia (Iran and Israel). While remaining 114 sorghum accessions were selected from Africa, covering almost all parts of African regions (Chad, Cong, Lesotho, Morocco, South Africa, Central Africa, Sudan, Nigeria, Algeria, Uganda, Ethiopia, Kenya, Zimbabwe and Tanzania). The passport and evaluation data of these accessions can be easily accessed through Genebank, NIAS website (http://www.gene.affrc.go.jp/about_en.php). The list of selected 320 sorghum accessions, its origin and other detail information is available on our laboratory home page: http://www.sakura.cc.tsukuba.ac.jp/~pbreed/index.html.

Leaves from 40 days old seedlings were cut and then subjected to vacuum freeze drying method for dehydration. Genomic DNA was extracted from leaf tissues using the CTAB method described by Murray and Thompson (1980) with some modification. Extraction buffer was composed of 2% CTAB, 50 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.7 M NaCl, 0.1% SDS, 0.1 mg/ml Proteinase K, 2% insoluble PVP and 2% 2-mercaptoethanol. Chloroform extraction was performed to remove cellular debris and proteins by using chloroform-isoamyl alcohol (24:1 v/v), DNA was precipitated by addition 2-isopropanol and the precipitate was washed twice in 70% /90% ethanol. The final precipitate was dissolved in 50 µl of 1/10 TE solution containing RNase A, incubated at 42°C overnight, and stored at 4°C. The DNA concentration was measured on the NanoDrop ND-1000 (Thermo scientific) spectrophotometer and diluted to a working concentration of 5ng/µl.

Selection of SSR primers

Microsatellite primers were selected from published linkage maps of sorghum as revealed by Bhttaramakki et al. (2000), Kong et al. (2000) and Taramino et al. (1997). A total of 144 SSR primers were screened by using eight sorghum accessions with wide range of geographic distribution. 38 SSR primers showed a clear and polymorphic banding pattern which was then used on all 320 sorghum accessions while the other primer sets were discarded because of no band, complex banding pattern or non polymorphic nature. List of 38 SSR markers with primer sequences is given in Table 1.

PCR conditions and electrophoresis

PCR amplification of sorghum SSRs were performed in 10 uL reaction mixture containing 10 ng DNA template, 10 x PCR buffer (Mg^{2+} concentration: 20mM), 2mM dNTPs, 25 ng of each primer and 0.02 U of Blend Taq Plus polymerase (Toyobo Co., LTD. Japan) enzyme in either Eppendorf Master cycler or Applied Biosystems 9700 or 2700 PCR system Applied Biosystems 2720 thermal cycler. Annealing temperature was determined for all primers by using Eppendorf Master Cycler ep. gradient S. Thermal cycler protocol was set as denaturation at 98°C for 3 min, 30 cycles of 98°C (10s), 60°C (30s), and 72°C (30s), followed by for 7 min at 72°C and then cooling at 4°C.

PCR products were run on 10% polyacrylamide gel (10 cm in size) with constant supply of 200V power, 500mA current for 70 min to 120 min depending upon the size of PCR product. 10 x TBE buffer was used in making the gel while 1x Tris Glycine Buffer was subjected to the tank. Gel was stained in ethidium bromide solution and photograph was taken by using Kodak Digital Science EDAS 290 ver. 3.6 with Kodak ID Image analysis software ver. 3.5. Different bands of the same SSR primer was grouped according to their respective sizes by comparing with 50 bp DNA size marker ladder and all of the accessions were scored either "0" for absence or "1" for presence of specific allele.

Analysis of genetic diversity

Molecular data generated by SSR assays were analyzed by using NTSYS pc ver. 2.20q (Rohlf 2000). DICE similarity coefficient based on the proportion of shared alleles (DICE 1945, Nei and Li 1979) to measure genetic similarities between accessions based on the proportion of shared alleles with SIMQUAL module. The proportion of shared fragments dendogram was constructed by the clustering of accessions based on similarity matrix using unweighted pair group method with UPGMA arithmetic average algorithm in SAHN module (Fig. 1). Analysis of Molecular variance (AMOVA) based on PhiPT values, Allele frequencies (AF), Pairwise population Nei genetic distance values (Nei GD) and expected heterozygosity (He), also referred as unbiased gene diversity (D) were calculated by a software GenAlEx ver. 6 (Peakall and Smouse 2006).
Φ_{PT} is calculated as the proportion of the variance among populations, relative to the total variance.

$$\Phi_{PT} = \frac{V_{AP}}{(V_{AP} + V_{WP})}$$

V_{AP} the variance among populations and V_{WP} the variance within populations.
He (expected heterozygosity) can be calculated as

$$\overline{He} = \sum_{i=1}^{k} He\,(i)/k$$

Where He(i) is the expected heterozygosity in population i, and k is the number of populations.
Nei genetic distance is measured as Nei_D = -ln (I)
Where I is Nei's Genetic Identity and D is expected genetic heterozygosity.

Selection of accessions for SDRS

The number of accessions to be selected was decided on the basis of total number of alleles covered by the selected number of accessions (Fig. 2). The percentage of alleles covered by number of accessions was counted by repeated trails with each number of samples. As our target was to select a represented diversity set of accessions which has maximum genetic diversity and that can be utilized in the genetic studies of quantitative traits. There was only a loss of one allele at sample size of 107 to 139 and total alleles were recovered after sample size of 140 accessions. The sample size was eventually fixed at 107 at similarity index of 0.51 on the UPGMA dendogram. The dendogram constructed by the clustering of accessions based on similarity matrix using unweighted pair group method (UPGMA) divided into branches and one accession with highest diversity index was selected from each branch (Fig. 1). When the accessions in a branch were of similar diversity index then the origin was also given preference to maximize the representation of all geographic regions.

Phenotypic evaluation of SDRS

In year 2006, the selected 357 sorghum accessions from Genebank of National Institute of Agrobiological science (NIAS) were sown in field. Some of the accessions could not germinate and seeds were obtained only form 320 sorghum accessions. The panicles were covered with butter bags to avoid out crossing. Based on SSR evaluation, 107 sorghum accessions were selected as representative set of whole population and sown at NIAS field during growing season of year 2007. The statistical design used for field evaluation Randomized Complete Block Design (RCBD) with two replications. Ten plants per accession were planted with plant to plant distance of 20 cm and row to row distance of 1m. All standard agronomic practices were applied during the whole season. Data was recorded on 26 important morphological traits including both qualitative and quantitative characters, selected from standard NIAS, Genebank sorghum characters (descriptors). The traits can be divided into two groups according to the type of data collected i.e. continuous data and categorical data. Continuous type of data was recorded for 12 phenotypic traits including, days to heading (DTH), days to flowering (DTF), days to maturity (DTM), culm diameter (CD), grain weight per panicle (GWP), 100 grain weight (100GW), culm length (CL), number of tillers (NoT), number of panicles (NoP), panicle length (PL), leaf length (LL) and leaf width (LW). Similarly, 14 traits with categorical type of data were; panicle shape (PS), panicle type (PT), coleoptile's color (CC), quantity of lipid white powder on stem and leaves (LWP), color of midrib (MC), neck length of panicle (PNL), awn

presence (AP), glume color (GC), growth in early stage (GES), endosperm type (ET), aphid resistance (AR), number of regenerated tillers (NRT), regrowth (RG) and resistance to insecticides (RI). List of sorghum descriptors and related information can be accessed through NIAS, Genebank website i.e. http://www.gene.affrc.go.jp/about_en.php.

Mean values for all the traits obtained from two replications were standardized and analyzed by using NTSYS pc ver. 2.20q (Rohlf 2000) with Euqlidsq as coefficient of dissimilarity. The matrix was then subjected to UPGMA module for the construction of tree. A statistical software JMP ver.5 (SAS Institute 2002) was used to analyze both continuous and categorical data of 26 sorghum traits. Multivariate analysis, correlations, ANOVA and Tukey-Kramer HSD (honestly significant difference) test was performed for the 12 quantitative traits with continuous numeric data. While frequencies and contingency distribution with Pearson (χ^2) and likelihood-ratio Chi-square statistics were calculated for the 14 morphological traits having categorical data. χ^2 was calculated as:

$$\chi^2 = \sum_{cells} \frac{(\text{observed-expected})^2}{\text{expected}}$$

Results

SSR polymorphism

A total of 38 polymorphic SSR markers used in this study generated 146 alleles which were able to uniquely classify 320 sorghum accessions of diverse origin. The number of alleles revealed by each locus ranged from two to nine with average of 3.84. The selected SSR markers fairly cover the ten linkage groups of sorghum. Three of the SSRs have homology with genes responsible for a particular product such as *Xtxp212* (LG; D) is coding a putative protein in *A. thaliana*, *Xtxp92* (LG; E) is involved in the production of heat-shock-like protein in *Picea glauca* and *Xtxp273* (Pbbf) (LG; H) give a product of prolamine-box binding factor in *Z. mays* (Bhattramakki et al. 2000). Similarly three markers are derived from gene loci; *Cba* and *PepC* in LGs C and G, respectively, and *Kaf2* in LG J (Bhattramakki et al. 2000). Genetic similarity ranged from 0 (253 IN) to 0.8 (123 PK). Based on geographic distribution all of 320 sorghum accessions from 27 countries of origin can be classified into five major regions namely; East Asia, Southeast Asia, South Asia, Southwest Asia and Africa. Number of selected accessions from each region, average dice coefficient of similarity values and expected heterozygosity (gene diversity) for all 320 sorghum accessions, classified into five major regions are given in Table 2. Sorghum accessions from East Asia showed maximum value for dice coefficient (0.439) than the other groups. Gene diversity (He) and percent polymorphic loci (%P) were not so different for all accessions belonging to different geography. Analysis of molecular variance showed a total of 7% variation among regions while 93% variation within regions with significant value (Table 3).

SSR cluster analysis

Based on dendogram generated from 146 shared alleles by UPGMA cluster analysis, 320 sorghum accessions were clearly classified into three distinct clusters i.e. Cluster I and Cluster II and Cluster III (Fig.1). Cluster I, the major cluster is sub divided into six groups (a, b, c, d, e, f) comprised of 162 sorghum accessions derived from Asian origin with few exceptions. Among cluster I, group “a” and “b” is mainly composed of East Asian accessions from Japan, China and Korea while group “c” includes South Asian accessions from India and Pakistan. Small groups including group “d” contains only seven accessions (Sudan:2, Nigeria:1, Zimbabwe:1, Japan:1, Korea:1 and India:1) while group “e” is composed of four accessions (Japan:2 and South Africa: 2). Similarly group “f” is composed of accessions from India and Pakistan. One accession (289PK) from Pakistan grouped exclusively separated from others in this cluster. Cluster II and III can be again sub divided into five groups namely; a, b, c, d and e. Cluster II is relatively mix type of cluster with 82 sorghum accessions from Africa and Asia. Eight accessions from East Asia also falls in this cluster (Japan:4, Korea:1, China:2, Taiwan:1) while the only one accession from Iran, Israel and Myanmar also present in this cluster. In this cluster two accessions from African origin (239ET and 306KE) were uniquely placed and remain separated from other accessions of the cluster. A total of 71 accessions were included in cluster III mainly derived from Africa along with 16 accessions from other origins (Japan:2, China:4, Cambodia:1, India:2, Pakistan:5 and Nepal:1). Four sorghum accessions having

ID as 233 (IN), 66 (ZA), 29 (IN) and 72 (IN), were separated form all other accessions made unique positions in the dendogram.

The same results were reconfirmed by principal coordinate analysis using correlation genetic similarity matrix. The first two principal coordinates with eigen value effect of 17.5% and 5.8% respectively, resulted into a scatter plot which discriminately classified all 320 accessions into three major groups (Fig. 3). Basically East Asian accessions made a distinct group from except few accessions witch showed similarity with accessions from African origin. Accessions from South Asian countries spread over the plot and linked between the two clusters mainly created by East Asia and Africa. The only two sorghum accessions from Southeast Asia and two from Southwest Asia also lies in the plate of African accessions.

Characteristics of SDRS

The core set of accessions include 107 sorghum accessions from 27 countries with diverse origin from Asia and Africa pole of the globe. Each country contributed different number of accessions to the total, such as Japan; 11, Korea; 7, Taiwan; 1, China; 6, Cambodia; 1, Myanmar; 1, India; 8, Pakistan; 13, Afghanistan; 2, Bangladesh; 1, Nepal; 2, Iran; 1, Israel; 1, Chad; 2, Congo; 1, Lesotho; 3, Morocco; 5, South Africa; 7, Central Africa; 1, Sudan; 11, Nigeria; 4, Algeria; 1, Uganda; 4, Ethiopia; 5, Kenya; 3, Zimbabwe; 3 and Tanzania; 2. The list of SDRS with relevant information can be accessed through our laboratory website (http://www.sakura.cc.tsukuba.ac.jp/~pbreed/) or by request to corresponding author. There was no drastic change in dice average value, percent polymorphic loci (%P) and gene diversity (He) in SDRS in comparison with base populations except for East Asian accessions where the percent polymorphic loci, % P is reduced from 82.86 to 78.32 while gene diversity, He changed from 0.22 to 0.20 (Table 2). Analysis of molecular variance for SDRS showed a total of 6% variation among geographic regions while 94% variation within geographic regions with significant value (Table 3).

Phenotypic characterization of SRDS

To confirm the spectrum of diversity in morphological traits, data was recorded for the diversity representative set of 107 sorghum accessions) which were selected from the SSR assessments. Analysis of variance (ANOVA) tested for 12 morphological traits of continuous type numeric data including days to heading (DH), days to flowering (DF), days to maturity (DM), culm diameter (CD), grain weight per panicle (GWP), 100 grain weight (100GW), culm length (CL), number of tillers (NoT), number of panicles (NoP), panicle length (PL), leaf length (LL) and leaf width (LW) showed highly significant difference for all the accessions according to *F* statistics. ANOVA table with distribution of 12 morpholigical traits are given in Table 4. According to correlation analysis, GWP was non-significantly correlated with CD (Table 5). Similarly CL had no significant correlation with CD and GWP. NoT was only highly significantly correlated with CD while the remaining combinations were non-significant. Similarly PL was showed only highly significance with DH while non-significantly correlated with other traits. LL had no significant correlation with GWP, NoT, NoP and PL. LW was non significant with GWP, 100GW, CL and PL.

The accessions in SDRS were classified into five major regions according to geographic distribution (East Asia: 25, Southeast Asia: 2, South Asia: 26, Southwest Asia: 2 and Africa: 52). Mean values of 12 continuous traits were compared for regions by using Tukey-Kramer HSD test. According to the test, the five regions comprised of 107 accessions showed no significant difference for mean comparison of 12 traits except for culm diameter where accessions in Africa group was separated from others. Similarly for 14 morphological traits, the accessions in SRDS distributed with different frequencies according to the ranks of sorghum descriptors list. Majority of accessions were non glutinous endosperm type in nature (100 accessions) and accessions with glutinous endosperm were mostly form East Asian origin (Japan, China and Korea). Most of the accessions showed susceptibility towards the application of insecticides i.e. 91 out of 107 accessions. The character of growth in early stage was recoded for maximum number of accessions (44 accessions) while one accession each showed very poor and extremely poor growth in early stage. Coleoptile color was green for most of the accessions (55 accessions) while white midrib color was recorded for majority of accessions (71 accessions). Aphid resistance, quantity of lipid powder on stem and leaves, number of regenerated tillers, regrowth, panicle shape, panicle type, neck length of panicle, awn presence and glume color varied with different categories among all sorghum accessions. The association among accessions of five geographic regions for 14 categorical traits was also compared by using Pearson Chi-square test at confidence interval of 0.95 (Table 6). Geographic regions

were found to be independent for six traits including ET, RI, CC, PS, PT and AP while for other traits (GES, CM, MC, LWP, NRT, RG, PNL and GC) they were positively associated with each other.

Clustering on the basis of morphological traits

By using all 26 morphological traits dendogram was generated based on Euqlidsq coefficient of genetic dissimilarity with UPGMA model of arithmetic algorithm (Fig. 4). The dendogram can be divided into two main clusters each with four groups (a, b, c and d) respectively. All 107 sorghum accessions distributed between the clusters independent of their origin. Cluster I is composed of 52 sorghum accessions, majority is from Asian origin (36 accessions) while 16 accessions are from Africa. Similarly among accessions in Cluster II, 17 out of 55 total accessions are from Asian origin while remaining 38 accessions are African based origin.

Multivariate analysis of 12 quantitative traits

Multivariate analysis of the 12 quantitative characters gives a representation of the morphological diversity of 107 sorghum accessions. The first axis accounts for 40.8% of the variance and the principal component scores are influenced by DH, DF, DM, CD, CL, LL and LW as shown in Table 7. The second axis accounts for 20.9% of the variance which is mainly influenced by NoT, NoP and CL. Similarly third axis accounts for 12% of the total variance and mainly influenced by GWP, PL, LL and LW. Scores from principal component analysis were used to classify all 107 in scatter plot design (Fig.4). Based on first three principal components, all of the accessions made clusters with different combinations. Accessions can be divided into five major groups i.e. East Asia, Southeast Asia, South Asia, Southwest Asia and Africa. All sorghum accessions separated unequally on both axis and made mix clusters, which is independent of its origin. Fig. 5 (a) shows all sorghum accessions made mix cluster based first two principal components. Similarly 5 (b) represents the same type of mix clustering behavior by using first and third principal component.

Discussions

Genetic diversity assessments

This study gives a wide range of genetic diversity among 320 sorghum accessions selected from diverse origins based on geographic distribution in Asia and Africa. A large number of sorghum SSR markers i.e. 144, selected from three different published SSR linkage maps by Bhttaramakki et al. (2000), Kong et al. (2000) and Taramino et al. (1997). Only 38 unique markers were selected on the basis of their clear distinguishing polymorphic pattern. The total number of alleles generated from these markers (146) and size varies from the source data which might be because of the difference in genotypes as our material was composed of diverse collection from different origins. SSR markers used in this study showed a high level of polymorphism which is because of its unique replication slippage mechanism (Pejic et al, 1998). The mean number of alleles per locus (3.84) was similar as recorded by Ali et al. (2007) on 72 entries (3.22) and Schloss et al. (2002) on 25 sorghum lines (3.4) but lower than that reported by Agrama and Tuinstra (2003) and Smith et al. (2000) with mean allele per locus of 4.3 and 5.9, respectively. The average percentage of polymorphic loci (P) and gene diversity (He) observed in this study are 82.8 and 0.217 respectively recorded in our study represents a diverse spectrum in all sorghum accessions which is in correspondence with the findings of Djè et al. (2000). But the gene diversity observed in this population is less than reported by Ali *et al*. (2007) i.e. 0.40 and value of 0.46, 0.62, 0.58 reported by Schloss et al. (2002), Agrama and Tuinstra (2003) and Smith et al. (2000), respectively.

One SSR Locus *Xtxp 331*was identified to possess rare alleles and exhibited gene diversity index of less than 0.05 in whole population and also in SDRS. Rare alleles are defined as a frequency of <0.05 (Somers et al. 2007; Casa et al. 2005). These rare alleles play an important role if they are uniquely linked to some particular genotypes. Such alleles are of major interest because they may be diagnostic for particular genotypes or for particular regions of the genome specific to a particular type of sorghum (Arama and Tuinstra 2003). Another locus *Cba* also behaved as comprised of rare alleles in East Asia group among 320 sorghum accessions while the remaining geographic groups showed only *Xtxp331* as rare. Similarly in SDRS *Xtxp331* was the only locus with rare alleles in all five geographic groups. One allele of the locus

named as *Xtxp287*, was not selected and consequently caused the loss one allele in total of 146 alleles due to the selection of representative set of accessions.

Clustering of sorghum accessions

Based on 38 sorghum SSR loci with 146 alleles, all of the sorghum accessions (320) distinctly classified according to geographic distribution. Our results indicated that genetic variation was more closely related to geographic origin which is supported by Morden et al. (1989) while it is contrary to the findings of Djè et al. (2000) that 11% of total genetic variation at microsatellite loci was due to geographic differentiation. Our collection was composed of race bicolor that showed a relative high level of genetic diversity than reported for other races of *Sorghum bicolor* (Djè et al. 2000). Bicolor is thought to be the race most closely related to wild sorghums (Harlan and de Wet 1972).

The 26 sorghum traits studied, both quantitative and qualitative could not classify 107 selected sorghum accessions according to geographic distribution. Our results could not confirm any clear relationship between molecular and morphological estimates. This in contrast to the findings of Geleta et al. (2006) which were reported a significant correlation between SSR and morphological traits among 45 accessions by using ten SSR primers. It could be because of our large population size and greater number of SSR markers. While Ayana et al. (2001) showed no genetic relationship (allozyme plus RAPD) with quantitative agro-morphological traits, which shows no correlation between molecular markers and morphological traits. Similarly Dhalberg et al. (2002) reported insufficient relationship between RAPD markers and agronomic descriptors. Morphological traits are more influenced by environment thus there is more chance of fluctuation with change in environmental factors while genetic structure is more consistent with less influence of environment. Our results are supported by the findings of Fufa et al. (2005) and Almanza-Pinzon et al. (2003).

High levels of correlation between DNA markers and agronomic descriptors may be difficult but still morphological traits are important for developing core collection (Ayana et al. 2001). Another possible reason for such mix type of clusters of sorghum accessions for morphological traits could be intense human selection pressure for such traits since beginning in the process of sorghum domestication. While SSRs are usually not the targets of selection, and its reduced variation at a linked site during a selective sweep depends not only on the strength of selection, but also on the rate and mode of mutation and the rate of recombination between the selected site and the marker assayed (Casa et al. 2005). SSR loci may quickly recover lost diversity and, therefore, give non-significant deviation from neutrality (Vigouroux et al. 2005). Sorghum is considered to be originated from Africa (Harlan and Stemler 1976; Wendorf et al. 1992). This study clearly identifies the pattern of sorghum domestication from Africa to Asia (Fig.2). Our results classified sorghum accessions in two distinct parts of East Asia and Africa while the accessions belonging to South Asia origin joins the two parts by spreading from Africa to East Asia. It suggests that sorghum was introduced from Africa to East Asia through South Asia.

The representativeness of the SDRS can be judged by comparing with original population. The diversity research set developed in this study retained 99.93 % of total alleles present in whole population. Similarly, no drastic change was observed for genetic diversity attributes like % polymorphic loci, similarity coefficient and gene diversity regarding base collection. Based on our results, we were able to construct a diversity research set of 107 accessions from 320 diverse geographic origins of Africa and Asia i.e. 33 % of the total number. Our collection size is more than recommended by Brown et al. (1989b) i.e. 10 % of whole population size which reflects 70 % of total genetic diversity. The size of our collection is in accordance with Noirot et al. (1996) who suggested that higher percentages (20–30%) are needed, particularly where the objective is to capture the genetic diversity of quantitatively inherited characters.

USDA and ICRISAT collections lack accessions from East Asia and mostly composed of African and South Asian accessions. Chinese sorghum collection is composed of mostly native germplasm with introduction of some African accessions which shows less diversity in traits. While, the SDRS developed in our study has a diverse collection of landraces without any improved variety, selected from all parts of Africa and Asia. Our results reflect the process of sorghum domestication from its place of origin to East Asia.

Conclusions

Molecular markers especially SSR are most efficient source of diversity studies in crop species. In this report 38 SSR markers clearly classified 320 sorghum accessions into geographic pattern and showed the process of domestication from Africa to East Asia through South Asia. Our study will serve as a pioneer work in construction of sorghum core collection in Genebank, Japan. The use of population diversity based approaches and SSR assessments can be used for identifying genomic regions of interest. Based on the data obtained, there is still a need to further investigate the proper method of diversity studies and making a core collection. Whole genome sequence information can serve as a landmark for developing more authentic approach for making core collection and the area of functional genomics can be explored to establish a solution for such constrains.

Our further targets include the utilization of SDRS in linkage-disequilibrium (LD) analysis and association between molecular markers and morphological traits and association mapping. The cryptic variations associated with SSR loci and long distance chromosomal regions can be further utilized for the selection of parents in Qtl mapping techniques. Our study will serve as a pioneer work in sorghum diversity and genomic studies.

Acknowledgment

This research was supported by the project of Genebank, NIAS under the Ministry of Agriculture, Forestry and Fisheries (MAFF), Japan.

Literature

Agarma HA, Tuinstra MR (2003) Phylogenetic diversity and relationships among sorghum accessions using SSRs and RAPDs. African Journal of Biotechnology 2:334-340

Ahnert D, Lee M, Austin D, Livini C, Woodman W, Openshaw S, Smith J, Porter K, Dalon G (1996) Genetic diversity among elite sorghum inbred lines assessed with DNA markers and pedigree information. Crop Sci 36:1385–1392

Ali ML, Rajewski JF, Baenziger PS, Gill KS, Eskridge KM, Dweikat L (2007) Assessment of genetic diversity and relationship among a collection of US sweet sorghum germplasm by SSR markers. Mol Breeding DOI 10.1007/s11032-007-9149-z

Alamnza-Pinzon MI, Khairullah M, Fox PN, Warburton ML (2003) Comparison of molecular markers and coefficients of parentage for the analysis of genetic diversity among spring bread wheat accessions. Euphytica 130:77–86

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep 9:208–218.

Ayana A, Bryngelsson T, Bekele E (2000) Geographic and altitudinal allozyme variation in sorghum (*Sorghum bicolor* (L.) Moench) landraces from Ethiopia and Eritrea. Hereditas 135:1-12.

Brown AHD (1989) Core collections: A practical approach to genetic resources management. Genome 31:818–824

Casa A, Mitchell S, Hamblin M, Sun H, Bowers J, Paterson A, Aquadro C, Kresovich S (2005) Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. Theor Appl Genet 111:23–30

Anas, Yoshida T (2004) Genetic diversity among Japanese cultivated sorghum assessed with simple sequence repeats markers. Plant Prod Sci 7:217–223

Dahlberg JA, Zhang X, Hart, GE, Mullet JE (2002) Comparative assessment of variation among sorghum germplasm accessions using seed morphology and RAPD measurements. Crop Sci. 42:291–296

Dhalberg JA, Burke JJ, Rosenow DT (2004) Development of a sorghum core collection: Refinement and evaluation of a subset from Sudan. Economic Botany 58:556-567

De Wet JMJ (1978) Systematics and evolution of Sorghum Sect. Sorghum (Gramineae). Am J Bot 65:477–484.

Deu M, Gonzalez-de-Leon D, Glaszmann J, Degremont I, Chantereau J, Lanaud C, Hamon P (1994) RFLP diversity in cultivated sorghum in relation to racial differentiation. Theor Appl Genet 88:838–844

Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297–302.

Djè Y, Heuertz M, Lefèbvre C, Vekemans X (2000) Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers. Theor Appl Genet 100:918–925.

Escribano P, Viruel MA, Hormaza JI (2008) Comparison of different methods to construct a core germplasm collection in woody perennial species with simple sequence repeat markers. A case study in cherimoya (*Annona cherimola*,Annonaceae), an underutilised subtropical fruit tree species. Ann Appl Biol doi:10.1111/j.1744-7348.2008.00232.x

Folkertsma R, Frederick H, Rattunde W, Chandra S, Raju G, Hash C (2005) The pattern of genetic diversity of Guinea-race *Sorghum bicolor* (L.) Moench landraces as revealed with SSR markers. Theor Appl Genet 111:399–409

Frankel OH (1984) Genetic perspectives of germplasm conservation. In: Arber W, Llimensee K, Peacock WJ, Starlinger P (eds) Genetic Manipulation: Impact on Man and Society, Cambridge University Press, Cambridge, pp 161-170

Fufa H, Baenziger PS, Beecher BS, Dweikat I, Graybosch, Eskridge KM (2005) Comparison of phenotypic and molecular marker-based classifications of hard red winter wheat cultivars. Euphytica 145:133–146

Geleta N, Labuschangne, MT, Chris D, Viljoen, CD (2006) Genetic diversity analysis in sorghum germplasm as estimated by AFLP, SSR and morpho-agronomical markers. Biodiversity and Conservation 15:3251–3265

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

Grenier C, Deu M, Kresovich S, Bramel-Cox PJ, Hamon P (2000) Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs non-random sampling procedures B. Using molecular markers. Theor Appl Genet 101:197–202.

Guarino L A, Jarvis RJ, Hijmans , Maxted N (2001) Geographic information systems (GIS) and the conservation and use of plant genetic resources. In: Engels JMM, Ramanatha Rao V, Brown AHD, Jackson MT (eds) Managing Plant Genetic Diversity, CABI, Oxon, UK, pp. 171–178

Harlan JR, De Wet JMJ (1972) A simplified classification of cultivated sorghum. Crop Sci 12:172–176.

Harlan JR, Stemler A (1976) The races of Sorghum in Africa. In: Harlan JR, de Wet JMJ, Stemler A (eds) Origin of African plant domestication, Mouton Press, The Hague, Netherlands, pp 465–478

Kong L, Dong J, Hart GE (2000) Characteristics, linkage-map positions, and allelic differentiation of *Sorghum bicolor* (L.) Moench DNA simple-sequence repeats (SSRs). Theor Appl Genet 101: 438–448.

Mann JA, Kimber CT, Miller FR (1983) The Origin and Early Cultivation of Sorghums in Africa. Bulletin 1454, Texas Agriculture Experiment Station, College Station, TX, USA.

Menz M, Klein R, Unruh N, Rooney W, Klein P, Mullet J (2004) Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. CropSci 44:1236–1244

Menz1 MA, Klein RR, Mullet JE, Obert JA, Unruh NC, Klein PE (2002) A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926AFLP, RFLP and SSR markers. Plant Molecular Biology 48: 483–499

Morden CW, Doebley K, Schertz KF (1990) Allozyme variation among the spontaneous species of *Sorghum* section Sorghum (Poaceae). Theor Appl Genet 80:296–304

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl acad Sci USA 76:5269–5273

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

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in excel, population genetic software for teaching and research. Molecular Ecology Notes 6: 288-295

Pejic I, Ajmone-Marsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G, Motto M (1998) Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. Theor Appl Genet 97: 1248-1255.

Qingshan L, Dahlberg JA (2001) Chinese Sorghum Genetic Resources. Economic Botany 55: 401-425

Ritter KB, Lynne McIntyre C, Godwin ID, Jordan DR, Chapman, SC (2007) An assessment of the genetic relationship between sweet and grain sorghums, within *Sorghum bicolor* ssp. bicolor (L.) Moench, using AFLP markers. Euphytica 157:161–176

SAS Institute Inc, Cary NC, USA (2002) JMP statistical and graphics guide, version5

Schloss SJ, Mitchell SE, White GM, Kukatla R, Bowers JE, Paterson AH, Kresovich S (2002) Characterization of RFLP probe sequences for gene discovery and SSR development in *Sorghum bicolor* L. Moench. Theor Appl Genet 105:912–920.

Smith JSC, Kresovich S, Hopkins MS, Mitchell SE, Dean RE, Woodman WL, Lee M, Porter K (2000) Genetic diversity among elite sorghum inbred lines assessed with Simple sequence repeats. Crop Sci 40:226–232.

Somers DJ, Banks T, Depauw R, Clarke J, Fox S, Pozniak C (2007) Genome-wide linkage disequilibrium in wheat. Plant and Animal Genome XV conference, 13–17 Jan., 2007, San Diego, CA, Poster abstract P268

Tao Y, Manners J, Ludlow M, Henzell R (1993) DNA polymorphisms in grain sorghum (*Sorghum bicolor* (L.) Moench). Theor Appl Genet 86:679–688

Taramino G, Tarchini R, Ferrario S, Lee M, Pe ME (1997) Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*. Theor Appl Genet 95: 66–72.

Uptmoor R, Wenzel W, Friedt W, Donaldson G, Ayisi K, Ordon F (2003) Comparative analysis on the genetic relatedness of *Sorghum bicolor* accessions from Southern Africa by RAPDs, AFLPs and SSRs. Theor Appl Genet 106:1316–1325

Vigouroux Y, Mitchell SE, Matsuoka Y, Hamblin MT, Kresovich S, Smith JSC, Jaqueth J, Smith OS, Doebley J (2005) Ananalysis of genetic diversity across the maize genome using microsatellites.Genetics169:1617–1630

Wendorf F, Close AE, Schild R, Wasylikowa K, Housley RA, Harlan JR, Krolik H (1992) Saharan exploitation of plants 8000 years BP. Nature (London) 359: 721–724

Legends of Tables (7 Tables)

Table1 List of Sorghum SSR primers; Bhattramakki et al. (2000) and Kong et al. (2000)

Table 2 Total number of selected accessions, Dice coefficient average value, mean % polymorphic loci (P) and expected Heterozygosity, He (Gene diversity) of base collection 320 sorghum accessions in comparison with SDRS (107 accessions) assayed by 38 SSR markers (146 alleles)

Table 3 Analysis of molecular variance (AMOVA) of base population (320 sorghum accessions) and SDRS (107 accessions) with 38 SSR markers

Table4 Analysis of variance (ANOVA) and distributions of 12 quantitative traits with continuous type of data recorded on 107 selected sorghum accessions in 2 replications

Table 5 Correlations between 12 quantitative traits with continuous type recorded for 107 sorghum accessions

Table 6 Pearson chi-square tests of 14 morphological traits with categorical data recorded for 107 sorghum accessions

Table7 Principal Components based on Correlations with eigen estimates and their effect resulted from 12 morphological traits with continuous type of data

Tables and Figures

Table1 List of Sorghum SSR primers; Bhattramakki et al. (2000) and Kong et al. (2000)

Locus	LG	No.alleles	Type(s) of SSR(s) amplified and number of repeats in BTx623c	Sequence of forward primer	Sequence of reverse primer	Size range (bp)	Ann. temp(°C)
Xtxp284	A	4	$(AAG)_{19}$	CCAGATTGGCTGATGCATACACACT	AAGGGTAATTTATGCACTCCAAGGTAGGAC	200-245	60
Xtxp335	A	5	$(GT)_{12}$	TATTTCCTCTTGAAAGAATCAGGG	TATTCATCGAGCAAAAGGCA	160-245	60
Xtxp149	A	3	$(CT)_{10}$	AGCCTTGCATGATGTTCC	GCTATGCTTGGTGTGGG	160-165	60
Xtxp56	B	4	$(GA)_{39}$	TGTCTTCGTAGTTGCGTGTTG	CCGAAGGAGTGCTTTGGAC	310-450	60
Xtxp348	B	4	$(TAA)_{37}$	CGACATCAGCGTTGTCTTTCTA	GCTTACGAATAGGGCAAAAGAACT	280-325	60
Xtxp197	B	3	$(AC)_{10}$	GCGTCAATTAATCCAAACAGCCTC	GAGTTCCTATTCCCGTTCATGGTGAT	150-160	60
Xtxp201	B	3	$(GA)_{36}$	GCGTTTATGGAAGCAAAAT	CTCATAAGGCAGGACCAAC	225-265	60
Xtxp285	C	5	$(CTT)_{11}CTC(CTT)_{16}$	ATTTGATTCTTCTTGCTTTGCCTTGT	TTGTCATTTCCCCCTTCTTTCTTTT	205-260	60
Cba	C	3	$(TA)_{18}$	AAAGCTCGGCGTTAGAAATA	CGCTTAACAACTCCTACCATC	195-230	60
Xtxp205	C	3	$(AG)_{12}$	CCTGCCGTGTCTTCC	TATATGCATGCCGTAGATTT	190-200	60
Xtxp228	C	3	$(TC)_{12}$	ACAGGTTGGCGATGTTTCTCT	TTCTTTTTCGAATTCATTCCTTTT	230-250	60
Xtxp59	C	2	$(GGA)_5$	GAAATCCACGATAGGGTAAGG	GACCCAGAATAGAAGAGAGG	195-200	60
Xtxp60	D	2	$(GT)_4GC(GT)_5$	GCTAGCTGACGCACGTCTCTG	TGCAACCGAGCGGTGACTA	220-225	60
Xtxp51	D	2	$(TG)_{11}$	TCTCGGACTCAAGAGCAGAGG	GGACAGCAGCGGCTTCAG	225-230	60
Xtxp212	D	3	$(GT)_{10}$	TTTCCCCTCTTTCTTGTGTC	CTCGGCGTCGTCGTA	135-145	60
Xtxp312	E	6	$(CAA)_{26}$	CAGGAAAATACGATCCGTGCCAAGT	GTGAACTATTCGGAAGAAGTTTGGAGGAAA	90-185	60
Xtxp92	E	2	$(GAA)_5$	ACTTGCAGGTTAATTTCGTCC	GGCGAGCTTGCGGTAG	155-165	60
Xtxp295	E	4	$(TC)_{19}$	AAATCATGCATCCATGTTCGTCTTC	CTCCCGCTACAAGAGTACATTCATAGCTTA	145-160	60
Xtxp278	E	2	$(TTG)_{12}$	GGGTTTCAACTCTAGCCTACCGAACTTCCT	ATGCCTCATCATGGTTCGTTTTGCTT	240-250	60
Xtxp67	F	4	$(GA)_{28}$	CCTGACGCTCGTGGCTACC	TCCACACAAGATTCAGGCTCC	145-170	60
Xtxp258	F	5	$(AAC)_{19}$	CACCAAGTGTCGCGAACTGAA	GCTTAGTGTGAGCGCTGACCAG	180-225	60
Xtxp287	F	5	$(AAC)_{21}$	GCAAGCGAGCTGACTTATGTAACGAGA	CAAAGTGCTACTAAACCTATGCAGGGTGAA	330-360	60
Xtxp10	F	3	$(CT)_{14}$	ATA CTA TCA AGA GGG GAG C	AGT ACT AGC CAC ACG TCA C	125-140	50
Xtxp217	G	4	$(GA)_{23}$	GGCCTCGACTACGGAGTT	TCGGCATATTGATTTGGTTT	150-170	60
Xtxp270	G	6	$(GAA)_{12}(GAAA)_6 + (GAA)_{21} + (GTA)_5 + (GTA)_3 + (GTA)_3$	AGCAAGAAGAAGGCAAGAAGAAGG	GCGAAATTATTTTGAAATGGAGTTGA	280-320	60
Xtxp331	G	9	$(GAT)_{32}$	AACGGTTATTAGAGAGGGAGA	AGTATAATAACATTTTGACACCCA	170-280	60
PepC	G	3	$(AT)_{10}$	TGGGAAGCAGCTCAGG	AGGGTGGTGATGTAGGGA	200-220	60
Xtxp20	G	5	$(AG)_{21}$	TCT CAA GGT TTG ATG GTT GG	ACC CAT TAT TGA CCG TTG AG	180-240	60
Xtxp273(Pbbf)	H	4	$(TTG)_{20}$	GTACCCATTTAAATTGTTTGCAGTAG	CAGAGGAGGAGGAAGAGAAGG	180-220	60
Xtxp47	H	2	$(GT)_8(GC)_5 + (GT)^6$	CAATGGCTTGCACATGTCCTA	GGTGCGAGCTAGTTAAGTGGG	280-290	60
Xtxp354	H	5	$(GA)_{21}$ + (AAG)3	TGGGCAGGGTATCTAACTGA	GCCTTTTTCTGAGCCTTGA	130-170	60
Xtxp105	H	4	$(TG)_5 + (CT)_6GTCT(GT)_7$	TGGTATGGGACTGGACGG	TGTTGACGAAGCAACTCCAAT	290-325	60
Xtxp145	I	5	$(AG)_{22}$	GTTCCTCCTGCCATTACT	CTTCCGCACATCCAC	200-230	60
Xtxp274	I	6	$(TTC)_{19}$	GAAATTACAATGCTACCCCTAAAAGT	ACTCTACTCCTTCCGTCCACAT	280-320	60
Xtxp97	I	3	$(CA)_8+(GCC)_6$	CAAATAAACGGTGCACACTCA	GTATGATTGGAGACGAGACGG	120-130	60
Xtxp303	J	4	$(GT)_{13}$	AATGAGGAAAATATGAAACAAGTACCAA	AATAACAAGCGCAACTATATGAACAATAAA	140-160	60
Kaf2	J	3	$(CAA)_9$	TCGGCGAGCATCTTACA	TACGTAGGCGGTTGGATT	260-275	60
Xtxp65	J	3	$(ACC)_4+(CCA)_3CG(CT)_8$	CACGTCGTCACCAACCAA	GTTAAACGAAAGGGAAATGGC	120-130	60

Locus with bold face has homology with genes. LG; stands for linkage group (Bhattramakki et al. 2000)

Table 2 Total number of selected accessions, Dice coefficient average value, mean % polymorphic loci (P) and expected Heterozygosity, He (Gene diversity) of base collection 320 sorghum accessions in comparison with SDRS (107 accessions) assayed by 38 SSR markers (146 alleles)

Origin		**No. of Selected Accessions**	**Dice Average**	**%P**	**Gene Diversity (He)**
East Asia	Base Population	103	0.44	82.86	0.22
	SDRS	25	0.43	78.32	0.2
Southeast Asia	Base Population	2	0.3	82.89	0.22
	SDRS	2	0.3	82.89	0.22
South Asia	Base Population	99	0.35	82.27	0.22
	SDRS	26	0.35	81.68	0.21
Southwest Asia	Base Population	2	0.29	84.21	0.22
	SDRS	2	0.28	84.21	0.22
Africa	Base Population	114	0.32	81.81	0.21
	SDRS	52	0.31	81.88	0.21
Total	Base Population	320	0.34	82.81	0.22
	SDRS	107	0.33	82.27	0.22

Table 3 Analysis of molecular variance (AMOVA) of base population (320 sorghum accessions) and SDRS (107 accessions) with 38 SSR markers

Source		df	SS	MS	Est. Var.	%	Stat	Value	Prob
Among regions	Base Population	4	428.613	107.153	1.615	7%			
	SDRS	4	177.749	44.437	1.294	6%			
Within regions	Base Population	315	6337.778	20.120	20.120	93%	**PhiPT**	0.074	0.020
	SDRS	102	2239.008	21.951	21.951	94%		0.056	0.010
Total	Base Population	319	6766.391	127.273	21.735				
	SDRS	106	2416.757	66.388	23.245				

Table4 Analysis of variance (ANOVA) and distributions of 12 quantitative traits with continuous type of data recorded on 107 selected sorghum accessions in 2 replications

Source	df	Sum of Squares	Mean Squares	F ratio	Prob>F	Mean	Std Dev	Std Err Mean	upper 95% Mean	lower 95% Mean
DH	106	85323.4	804.9	6.8	<0.0001	88.5	19.5	1.9	92.2	84.8
Error	107	12742.5	119.1							
DF	106	93261.8	879.8	7.2	<0.0001	92.0	20.4	2.0	95.9	88.1
Error	107	13128.5	122.7							
DM	106	85323.4	804.9	6.8	<.0001	114.8	22.9	2.2	119.1	110.4
Error	107	12742.5	119.1							
CD	106	10583.3	99.8	5.0	<.0001	23.2	7.1	0.7	24.5	21.8
Error	107	2134.0	19.9							
GWP	106	55808.3	526.5	2.8	<.0001	35.4	15.9	1.5	38.4	32.3
Error	107	19793.0	185.0							
100GW	106	158.1	1.5	6.9	<.0001	2.9	0.8	0.1	3.0	2.7
Error	107	23.3	0.2							
CL	106	990522.1	9344.6	18.7	<.0001	178.3	67.7	6.5	191.3	165.4
Error	107	53550.0	500.5							
NoT	106	1654.3	15.6	17.5	<.0001	2.9	2.8	0.3	3.4	2.4
Error	107	95.5	0.9							
NoP	106	3462.8	32.7	7.6	<.0001	4.3	4.0	0.4	5.1	3.6
Error	107	461.0	4.3							
PL	106	15840.5	149.4	24.1	<.0001	27.1	8.5	0.8	28.7	25.5
Error	107	662.5	6.2							
LL	106	20503.0	193.4	3.0	<.0001	63.8	9.3	0.9	65.6	62.0
Error	107	6980.5	65.2							
LW	106	509.8	4.8	4.1	<.0001	7.9	1.5	0.1	8.2	7.6
Error	107	126.8	1.2							

DH; days to heading DF; days to flowering DM; days to maturity CD; culm diameter GWP; grain weight per panicle 100GW; 100 grain weight CL; culm length NoT; number of tillers NoP; number of panicles PL; panicle length LL; leaf length and LW; leaf width

Table 5 Correlations between 12 quantitative traits with continuous type recorded for 107 sorghum accessions

	DH	DF	DM	CD	GWP	100GW	CL	NoT	NoP	PL	LL
DF	0.10***										
DM	0.83***	0.84***									
CD	0.58***	0.58***	0.58***								
GWP	-0.37***	-0.38***	-0.45***	-0.13 ^{NS}							
100GW	-0.73***	-0.74***	-0.72***	-0.31**	0.52***						
CL	0.46***	0.47***	0.50***	0.05 ^{NS}	-0.17 ^{NS}	-0.51***					
NoT	-0.06^{NS}	-0.04^{NS}	-0.05 ^{NS}	-0.33***	-0.00 ^{NS}	-0.13 ^{NS}	0.15 ^{NS}				
NoP	-0.15***	-0.14 ^{NS}	-0.14 ^{NS}	-0.41***	0.04 ^{NS}	-0.04 ^{NS}	0.12 ^{NS}	0.94***			
PL	-0.05***	-0.06 ^{NS}	-0.11 ^{NS}	0.07 ^{NS}	0.07 ^{NS}	0.06 ^{NS}	-0.07 ^{NS}	0.18 ^{NS}	0.10 ^{NS}		
LL	0.53***	0.53***	0.51***	0.39***	0.05 ^{NS}	-0.47***	0.61***	0.12 ^{NS}	-0.04 ^{NS}	0.18 ^{NS}	
LW	0.36***	0.35***	0.30**	0.60***	0.10 ^{NS}	-0.10 ^{NS}	0.01 ^{NS}	-0.42***	-0.49***	0.19 ^{NS}	0.42***

DH; days to heading DF; days to flowering DM; days to maturity CD; culm diameter GWP; grain weight per panicle 100GW; 100 grain weight CL; culm length NoT; number of tillers NoP; number of panicles PL; panicle length LL; leaf length and LW; leaf width

Significant differences at *** P < 0.001, ** 0.01, * 0.05, respectively and NS No correlation (not significant)

Table 6 Pearson chi-square tests of 14 morphological traits with categorical data recorded for 107 sorghum accessions

Endosperm type				**Number of regenerated tillers**			
Test	*ChiSquare*	*Prob>ChiSq*	*RSquare(U)*	*Test*	*ChiSquare*	*Prob>ChiSq*	*RSquare(U)*
Likelihood Ratio	12.587	0.0135	0.2434	Likelihood Ratio	35.003	0.3274	0.0775
Pearson	11.382	0.0226		Pearson	35.584	0.3033	
Resistance to Insecticides				**Regrowth**			
Test	*ChiSquare*	*Prob>ChiSq*	*RSquare(U)*	*Test*	*ChiSquare*	*Prob>ChiSq*	*RSquare(U)*
Likelihood Ratio	14.124	0.0069	0.1564	Likelihood Ratio	42.369	0.104	0.0924
Pearson	16.735	0.0022		Pearson	45.147	0.0616	
Growth in early stage				**Panicle shape**			
Test	*ChiSquare*	*Prob>ChiSq*	*RSquare(U)*	*Test*	*ChiSquare*	*Prob>ChiSq*	*RSquare(U)*
Likelihood Ratio	27.252	0.7058	0.0739	Likelihood Ratio	74.578	0.0002	0.1962
Pearson	23.657	0.8566		Pearson	77.96	<.0001	
Coleoptile color				**Panicle type**			
Test	*ChiSquare*	*Prob>ChiSq*	*RSquare(U)*	*Test*	*ChiSquare*	*Prob>ChiSq*	*RSquare(U)*
Likelihood Ratio	29.482	0.0209	0.1191	Likelihood Ratio	31.137	0.0129	0.1219
Pearson	28.808	0.0253		Pearson	30.771	0.0144	
Color of midrib				**Neck length of panicle**			
Test	*ChiSquare*	*Prob>ChiSq*	*RSquare(U)*	*Test*	*ChiSquare*	*Prob>ChiSq*	*RSquare(U)*
Likelihood Ratio	28.246	0.1037	0.1168	Likelihood Ratio	37.168	0.243	0.0856
Pearson	31.84	0.045		Pearson	36.183	0.2795	
Aphid rsistance				**Awn presence**			
Test	*ChiSquare*	*Prob>ChiSq*	*RSquare(U)*	*Test*	*ChiSquare*	*Prob>ChiSq*	*RSquare(U)*
Likelihood Ratio	38.921	0.1863	0.0858	Likelihood Ratio	51.713	<.0001	0.3268
Pearson	34.926	0.3307		Pearson	92.652	<.0001	
Quantity of lipid white powder				**Glume color**			
Test	*ChiSquare*	*Prob>ChiSq*	*RSquare(U)*	*Test*	*ChiSquare*	*Prob>ChiSq*	*RSquare(U)*
Likelihood Ratio	28.782	0.4236	0.0737	Likelihood Ratio	39.732	0.1636	0.0974
Pearson	28.296	0.4489		Pearson	65.445	0.0004	

where RSquare(U) = –log likelihood for Regions/–log likelihood for C. Total (corrected total)

Table7 Principal Components based on Correlations with eigen estimates and their effect resulted from 12 morphological traits with continuous type of data

Eigenvalue	4.9004	2.5067	1.4743	0.9995	0.7257	0.3796	0.335	0.2803	0.2037	0.16	0.0319	0.0028
Percent	40.837	20.889	12.286	8.3296	6.0477	3.1631	2.7917	2.3361	1.6976	1.3332	0.2656	0.0237
Cum Percent	40.837	61.726	74.012	82.341	88.389	91.552	94.344	96.68	98.378	99.711	99.976	100
	Eigenvectors											
DH	0.4217	0.0462	-0.044	-0.049	0.1725	0.3841	-0.151	0.2241	0.2524	-0.162	-0.02	-0.689
DF	0.4238	0.0548	-0.055	-0.05	0.1746	0.3614	-0.145	0.2213	0.2221	-0.096	0.0186	0.7225
DM	0.4101	0.0632	-0.106	-0.023	0.0884	-0.039	0.2248	0.107	-0.29	0.81	-0.047	-0.046
CD	0.3051	-0.274	0.1278	-0.221	0.3714	-0.183	0.638	-0.166	-0.14	-0.374	0.0383	0.0089
GWP	-0.189	-0.132	0.492	0.4624	0.3666	0.4793	-0.038	-0.128	-0.317	0.083	0.0432	0.0005
100GW	-0.356	-0.221	0.1496	0.0809	0.1174	-0.053	0.3635	0.6346	0.4585	0.1838	0.0006	0.0028
CL	0.2542	0.2644	0.0841	0.5016	-0.458	-0.078	0.2495	0.3773	-0.286	-0.291	0.1395	0.0016
NoT	-0.062	0.564	0.2196	-0.154	0.2852	-0.177	-0.031	-0.007	0.0966	0.0558	0.6924	-0.024
NoP	-0.116	0.5576	0.1504	-0.106	0.3112	-0.15	-0.012	0.1888	-0.15	-0.123	-0.67	0.0166
PL	-0.014	0.0194	0.5493	-0.587	-0.475	0.3165	0.0975	0.0579	-0.099	0.0569	-0.03	0.0057
LL	0.3023	0.0824	0.4478	0.3	-0.145	-0.24	0.0339	-0.4	0.5549	0.1425	-0.208	0.0092
LW	0.2131	-0.379	0.3503	-0.066	0.1036	-0.487	-0.539	0.3181	-0.203	-0.031	0.0448	-0.004

DH; days to heading DF; days to flowering DM; days to maturity CD; culm diameter GWP; grain weight per panicle 100GW; 100 grain weight CL; culm length NoT; number of tillers NoP; number of panicles PL; panicle length LL; leaf length and LW; leaf width

Legends of Figures (5 Figures)

Fig.1 Dendogram generated by UPGMA cluster analysis showing relationship among 320 sorghum accessions based on Nei and Li similarity estimate (Nei and Li, 1997)

Fig. 2 Percentage of alleles retained in specific sample size of accessions

Fig. 3 Scores from principal component analysis of the coefficient matrix determined on 38 SSR markers for 320 diverse sorghum genotypes: Component1 [contribution rate: 56.13 (17.5%)] vs. Component 2 [contribution rate: 18.56 (5.8%)]

Fig. 4 Dendogram of 107 sorghum accessions generated by UPGMA model with EuqSql matrix based on 26 morphological traits

Fig. 5 Scatter plot designs developed by first tree principal components derived from 12 quantitative traits of 107 sorghum accessions: a) PC1 [contribution rate: 4.90 (40.8%)] vs. PC2 [contribution rate: 2.50 (20.9%)] and b) PC1 [contribution rate: 4.90 (40.8%)] vs. PC3 [contribution rate: 1.47 (12.3%)]

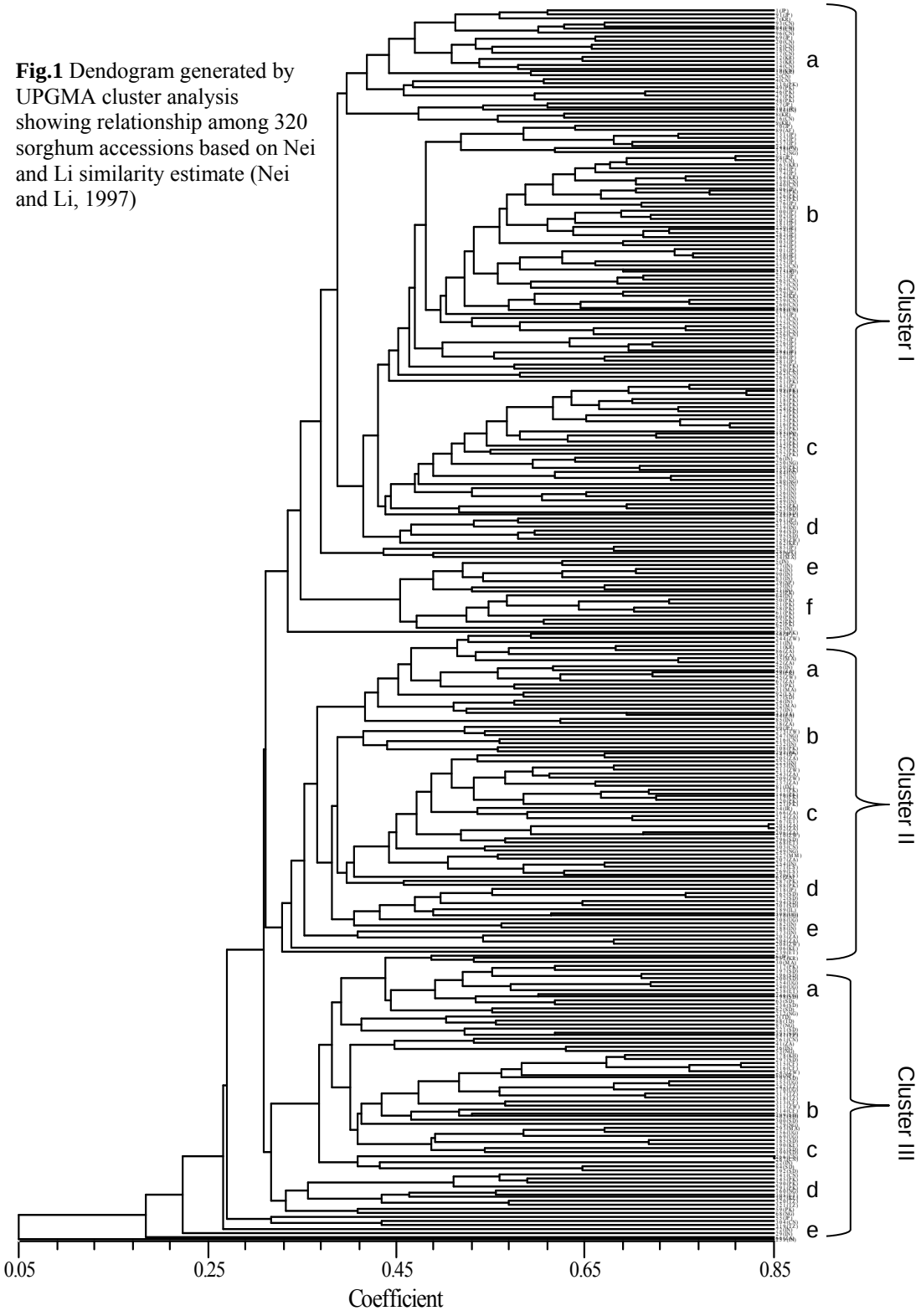


Fig.1 Dendogram generated by UPGMA cluster analysis showing relationship among 320 sorghum accessions based on Nei and Li similarity estimate (Nei and Li, 1997)

Fig. 2 Percentage of alleles retained in specific sample size of accessions

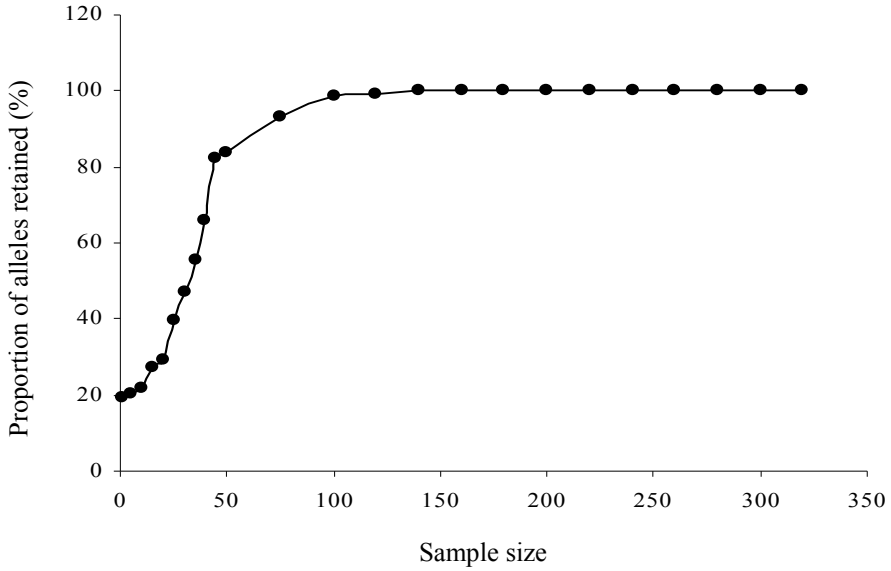
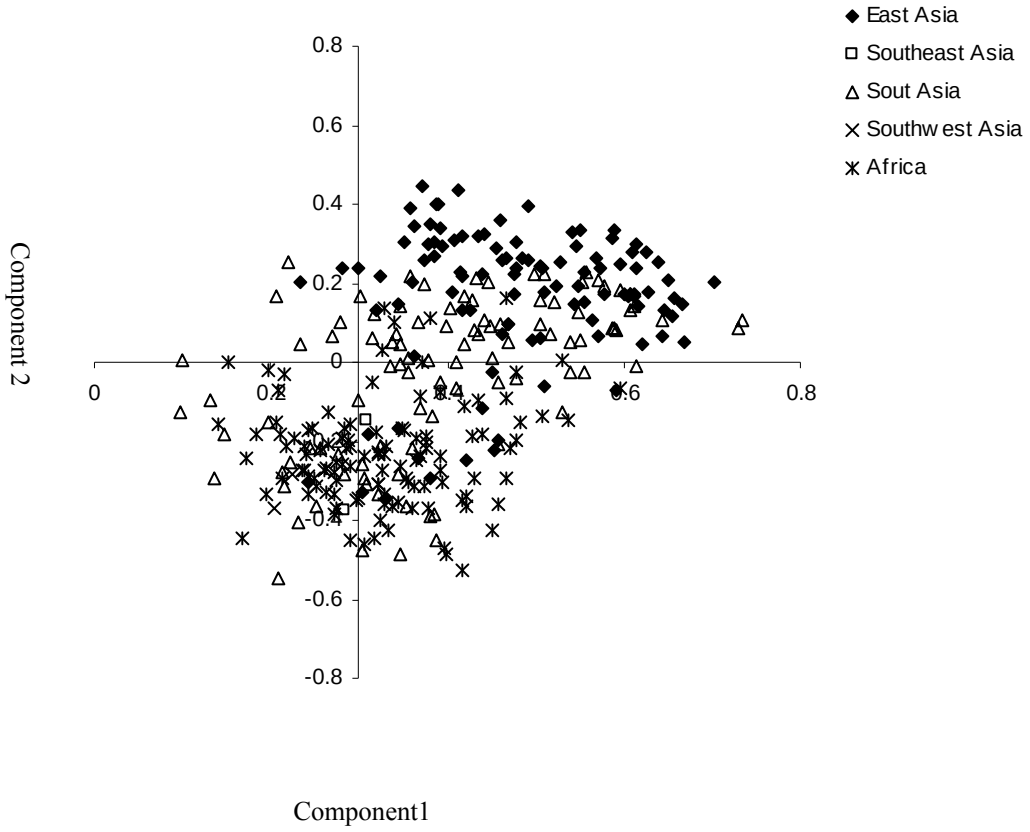


Fig. 3 Scores from principal component analysis of the coefficient matrix determined on 38 SSR markers for 320 diverse sorghum genotypes: Component1 [contribution rate: 56.13 (17.5%)] vs. Component 2 [contribution rate: 18.56 (5.8%)]



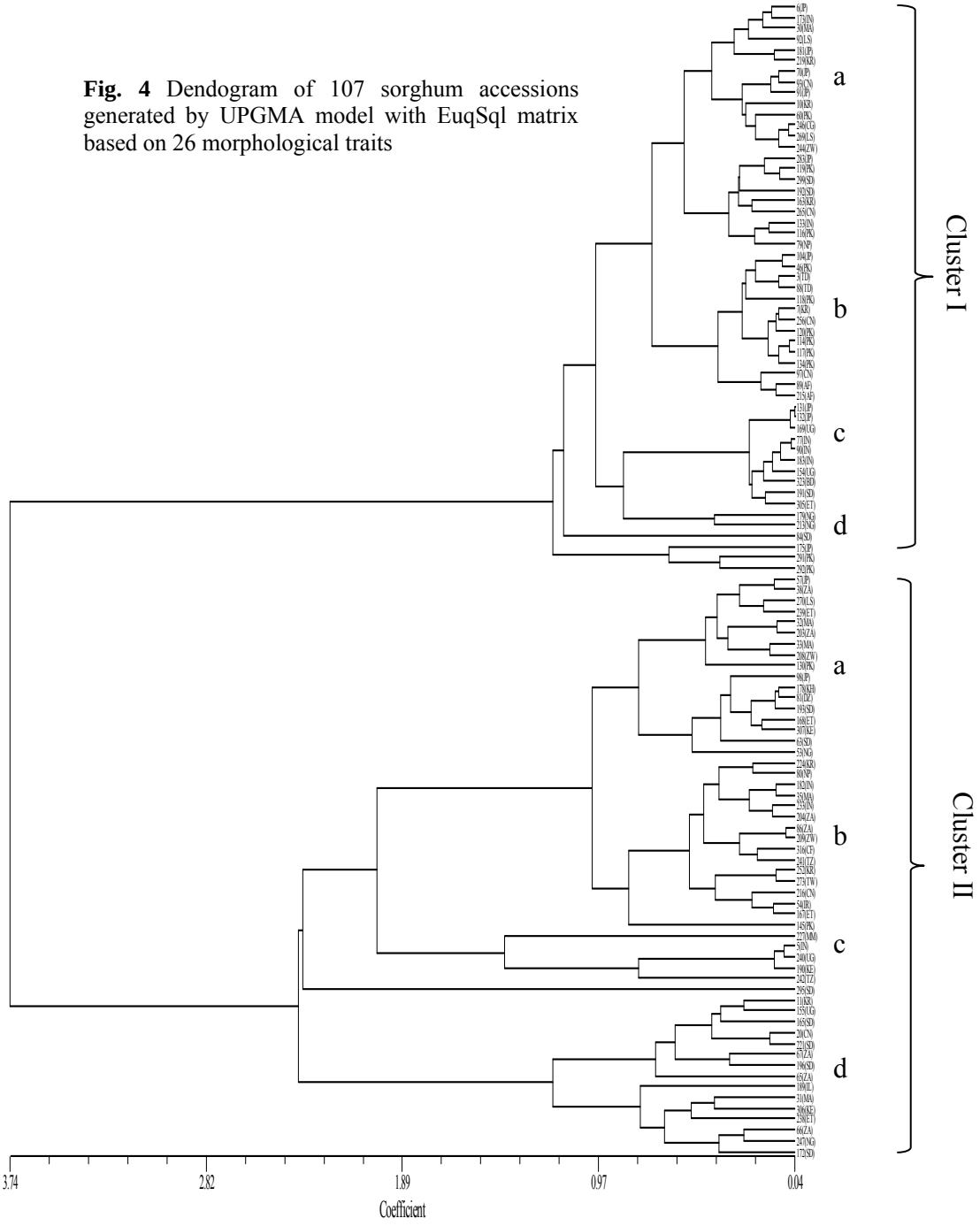


Fig. 4 Dendogram of 107 sorghum accessions generated by UPGMA model with EuqSql matrix based on 26 morphological traits

Fig. 5 Scatter plot designs developed by first tree principal components derived from 12 quantitative traits of 107 sorghum accessions: a) PC1 [contribution rate: 4.90 (40.8%)] vs. PC2 [contribution rate: 2.50 (20.9%)] and b) PC1 [contribution rate: 4.90 (40.8%)] vs. PC3 [contribution rate: 1.47 (12.3%)]