Population structure and linkage disequilibrium of ICRISAT foxtail millet (*Setaria italica* (L.) P. Beauv.) core collection

M. Vetriventhan[1], Hari D. Upadhyaya[1] , C. R. Anandakumar[2], S. Senthilvel[1, 4], R. K. Varshney[1] andHeiko K. Parzies[3]

(1) International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, 502 324, Andhra Pradesh, India

(2) Agricultural College and Research Institute, Tamil Nadu Agricultural University (TNAU), Madurai, 625 104, Tamil Nadu, India

(3) Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Fruwirthstr. 21, 70599 Stuttgart, Germany

(4) *Present address:* Directorate of Oilseeds Research (DOR), Rajendranagar, Hyderabad, 500 030, Andhra Pradesh, India

Euphytica
December 2013
DOI: http://dx.doi.org/10.1007/s10681-013-1044-6

Population structure and linkage disequilibrium of ICRISAT foxtail Millet (*Setaria italica* (L.) core collection

M. Vetriventhan, Hari D. Upadhyaya,* C.R. Anandakumar, S. Senthilvel, R.K. Varshney, Heiko K. Parzies

M. Vetriventhan, Hari D. Upadhyaya, S. Senthilvel, R.K. Varshney

International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru-502 324, Andhra Pradesh, India.

C.R. Anandakumar

Agricultural College and Research Institute, Tamil Nadu Agricultural University (TNAU), Madurai-625 104, Tamil Nadu, India.

Heiko K. Parzies

University of Hohenheim, Institute of Plant Breeding, Seed Science and Population Genetics, Fruwirthstr. 21, D-70599, Stuttart, Germany.

Present Address

S. Senthilvel

Directorate of Oilseeds Research (DOR), Rajendranagar-500 030, Hyderabad, Andhra Pradesh, India.

*Corresponding author: Hari D. Upadhyaya

E-mail address: h.upadhyaya@cgiar.org; Tel.: +91 40 3071 3333; Fax: +91 40 3071 3074

Abstract

Use of diverse germplasm is a key factor which allows high level of resolution due to extensive recombination in the history. Therefore, population used in association mapping should posses as many phenotypes as possible. One of the methods to obtain most of the phenotypes is to construct the core collection. The ICRISAT foxtail millet core collection consisting of 155 accessions was genotyped using 72 simple sequence repeat (SSR) markers to investigate the

genetic diversity, population structure and linkage disequilibrium. A high degree of molecular diversity among the accessions was found, with an average of 16.69 alleles per locus. STRUCTURE analyses classify the accessions into four subpopulations based on SSR allelic diversity. The Neighbor joining clustering and the principal coordinate analysis (PCoA) were in accordance with the racial classification. The distribution of molecular genetic variation among and within the four sub-populations and three races showed high degree of variability within each group, and low level of genetic distance among the groups. Linkage disequilibrium (LD) decay of less than 40 cM of genetic distance in foxtail millet core collection was observed, which suggests that it could be possible to achieve resolution down to the 40 cM level. From this investigation, it is evident that the foxtail millet core collection developed at ICRISAT is very diverse and could be a valuable resource for trait association mapping, crop breeding and germplasm management.

Key words: Analysis of molecular variance; Core collection; Foxtail millet; Genetic diversity; Linkage disequilibrium; Population structure; Simple sequence repeat

Introduction

Small millets are gaining importance because of their wider adaptability to drought and varied soil and environmental conditions, and nutritional benefits. Foxtail millet (*Setaria italica* (L.) P. Beauv.) (2n = 18), is one among the small millets, distributed widely around warm and temperate regions of Asia, Europe, North America, Australia and North Africa (Zhang et al. 2012). It is an ancient crop, its domestication in China dates back to 8,700 years ago (Lu et al. 2009), has greatly contributed to human civilizations both in Asia and Europe (Barton et al. 2009). Foxtail millet is as an important food and fodder crop in semi-arid tropics, particularly in changing climate. It produces substantial yield under varied environmental conditions, remains an essential food for home consumption in parts of India, China, Korea and Japan (Austin 2006). Taxonomically, foxtail millet is comprised of two subspecies, *S. italica* subsp. *italica* and subsp. *viridis*, and the wild ancestor of foxtail millet is *S. viridis* (Kihara and Kishimoto 1942; Li et al. 1945). Prasada Rao et al. (1987) have recognized three races of foxtail millet (*moharia*, *maxima* and *indica*) based on the morphological features.

Its relatively small genome size (~400Mb)(Bennetzen et al. 2012), inbreeding nature, short duration, C_4 nature, and wide geographic distribution and adaptability, make it an ideal model for grass functional genomics to investigate plant architecture, genome evolution, drought tolerance and physiology in the bioenergy grasses (Doust et al. 2009; Wang et al. 2010; Li and Brutnell 2011). A diversified germplasm collection plays a key role in both breeding and genomic research for any crop species. The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India has 1,534 foxtail millet accessions from 26 countries. To make use of this, Upadhyaya et al. (2008) developed a core collection consisting of 155 accessions, which represent 10.51% of 1,474 foxtail millet germplasm accessions conserved in ICRISAT genebank covering 23 countries. The entire germplasm collection of 1,534 accessions in foxtail millet is not large when compared with other crops like sorghum in ICRISAT Genebank (about 38,000 accessions), yet it is large in view of very low research priorities for neglected and under-utilized crops like foxtail millet. Moreover, this core collection can be evaluated extensively at a relatively low cost and information derived could be used a guide toward efficient utilization of the entire collection (Upadhyaya et al. 2008). Further, assessment of genetic diversity, population structure and linkage disequilibrium (LD) of foxtail

millet core collection could provide essential information for germplasm management, association mapping and crop improvement.

The phenotypic variations of many complex traits of agronomic importance are influenced by many genes, environment and interaction between genes and environment (Holland 2007). Linkage analysis and association mapping are the two most commonly used tools for dissecting complex traits (Zhu et al. 2008). Association mapping is an effective approach to detect QTL, if information on population structure and linkage disequilibrium is available. Association mapping has been proved to be an effective approach to mine the elite genes in germplasm resources (Zhang et al. 2011), has been successfully applied in several crops (Agrama et al. 2007; Malosetti et al. 2007; Zhao et al. 2007; Murray et al. 2009; Wen et al. 2009; Borba et al. 2010; Liu et al. 2010; Neumann et al. 2011; Dodig et al. 2012; Upadhyaya et al. 2012a; Upadhyaya et al. 2012b). One of the foremost factors is use of diverse germplasm in association mapping panel, which allows high level of resolution due to extensive recombination in the history (Wang et al. 2008). Therefore, choice of germplasm is the preliminary factor which determines the resolution of association mapping. Population used in association mapping should posses as many phenotypes as possible (Flint-Garcia et al. 2005). To achieve this, one of the methods to obtain most of the phenotypes is to construct the core collection (Zhang et al. 2011). A core collection (Frankel 1984) consists of a limited set of accessions (about 10%) derived from an existing germplasm collection, chosen to represent the genetic spectrum in the whole collection. In a species for which a core or mini core (Upadhyaya and Ortiz 2001) collections have been established, the core/mini core would be the ideal material for association mapping (Whitt and Buckler 2003).

Presence of population structure in the association panel, which is a division of the population into distinct subgroups related by kinship, leads to false positive association in association mapping (Yu and Buckler 2006; Zhu et al. 2008). Resolution of association studies depends on the structure of Linkage disequilibrium (LD) across the genome (Remington et al. 2001), low level of LD could lead to impractical whole-genome scanning because of the excessive number of markers required (Kruglyak 1999). Therefore, detailed information on population structure and extent of LD within the population are of fundamental importance for association mapping (Stich et al. 2005). Only a few previous researches on genetic diversity and population structure have been reported in foxtail millet. Wang et al. (2012) reported four subgroups in 250

landraces, which are in good accordance with eco-geographical distribution in China. Hirano et al. (2011) did population structure analysis by transposon display, which classified foxtail millet landraces into eight clusters that are closely related with geographic origins and suggest a monophyletic origin of foxtail millet domestication. Liu et al. (2011) investigated the population structure of foxtail millet and identified six groups, which matches with their pedigree information, in general, but not with their geographic origins.

In this study, a foxtail millet core collection (Upadhyaya et al. 2008) consisting of 155 accessions was used. The abundant variation of the foxtail millet core collection provides an important reservoir of genetic diversity and potential sources of beneficial alleles for its improvement (Upadhyaya et al. 2008). However, the knowledge on the level of genetic diversity and linkage disequilibrium in foxtail millet and its wild ancestor is very limited (Wang et al., 2010), which is necessary for dissecting DNA polymorphism underlying phenotypic variation using association mapping approach. To make use of foxtail millet core collection as association panel, the study was formulated to (1) examine the population structure of a foxtail millet core collection; (ii) investigate the genetic diversity within and among subpopulations, and (iii) identify the extent of LD within core collection. Results of this study would provide valuable information for trait association mapping using foxtail millet core collection, effective germplasm conservation, genomic studies and breeding applications.

Materials and Methods

Plant materials

A foxtail millet core collection consisting of 155 accessions, which is 10.51% of 1474 foxtail millet accessions from 23 countries conserved in ICRISAT genebank, was used (Upadhyaya et al. 2008). The collection consisted of 102 accessions (65.8%) from the race '*indica'*, 24 accessions (15.5%) from the race '*maxima'* and 29 accessions (18.7%) from the race '*moharia'*.

Genotyping by simple sequence repeat (SSR) markers

The details of SSR genotyping are available elsewhere (Vetriventhan et al. 2012), which have been followed in the present investigation. DNA was extracted from the 20 days old seedlings of 155 accessions using a high-throughput mini-DNA extraction method (Mace et al. 2003). A total of 72 SSR markers located across nine chromosomes of foxtail millet were used (Jia et al. 2009). The forward primers of all the SSRs were synthesized by adding M13-forward primer sequence (5′ CACGACGTTGTAAAACGAC3′) at the 5′ end of each primer. Genomic DNA of all the accessions was normalized to a uniform concentration of 5 ng μl^{-1} and the PCR were performed in 5 µl reaction. PCR products were then size-separated by capillary electrophoresis using an ABI Prism 3730*xl* DNA analyzer (Applied Biosystems Inc.). Raw data produced from ABI 3730*xl* DNA Analyser was analysed using Genemapper® software version 4.0 (Applied Biosystems, USA) and fragment size was scored in base pairs (bp) based on the relative migration of the internal size standard, LIZ 500 (Applied Biosystems, USA).

Population structure analysis

The model-based software program STRUCTURE 2.3.2 (Pritchard et al. 2000a; Pritchard et al. 2000b) was employed to subdivide accessions into genetic sub-population. No prior information was used to define sub-populations. To determine most appropriate *k* value, burn-in Markov Chain Monte Carlo (MCMC) replication was set to 10,000 and data were collected over 1,00,000 MCMC replications in each run. Five independent runs were performed setting the number of population (*k*) from 2 to 10 using a model allowing for no-admixture and correlated allele frequencies. The basis of this kind of clustering method is the allocation of individual genotypes to *k* clusters in such a way that Hardy-Weinberg equilibrium and linkage equilibrium are valid within clusters, whereas these kinds of equilibrium are absent between clusters. The *k* value was

determined based on the rate of change in LnP(D) between successive *k*, stability of grouping pattern across five run and germplasm information about the material under study.

Allele number, gene diversity, polymorphic information content (PIC), and heterozygosity (%) of each sub-population were calculated using PowerMarker version 3.25 (Liu and Muse 2005). An analysis of molecular variance (AMOVA) was performed based on 999 permutations using the software GENALEX 6.41 (Peakall and Smouse 2006) to estimate population differentiation among the sub-populations. Furthermore, pairwise genetic distance (Nei 1972), F_{ST} between sub-populations and Shannon's information index (*I*) were calculated using the software GENALEX 6.41. Pairwise Nei's genetic distance and F_{ST} indicates the genetic distance between the populations, measure of the extent of genetic differentiation among subpopulations. The low F_{ST} and Nei's genetic distance values between populations were indicative of minimal differentiation or high gene flow between the populations and *vice versa*. Principal coordinate analysis (PCoA) and neighbor-joining phylogenetic analysis was conducted to further assess the population subdivisions. PCoA was performed based on Nei (1973) distance matrix using GENALEX 6.41 and neighbor-joining tree was constructed based on the simple matching dissimilarity as implemented in DARwin 5.0.158 (2009-07-06) programme (Perrier and Jacquemoud-Collet 2006).

Linkage disequilibrium (LD)

The level of LD between pairs of locus was performed and the significance of pair-wise LD among all possible SSR loci was evaluated using TASSEL 2.1 (Bradbury et al. 2007) with the rapid permutation test in 10,000 shuffles. LD was estimated by squared allele-frequency correlations (r^2) between pairs of SSR loci. The pair of loci was considered to be significant in LD if *P* was ≤ 0.01.

Results

Overall genetic diversity

Seventy two SSRs detected a total of 1,202 alleles in the core collection. The average number of alleles per locus was 16.69, ranging from 4 to 35 (Table 1). The mean gene diversity was 0.73, and the mean PIC was 0.72. The average Shannon's information index was 1.99, ranging from 0.78 to 3.14. A total of 35 private or unique alleles that were present only in one accession and absent in the other accessions were detected. The average heterozygosity was 0.04, close to 100% homozygosity.

Population structure and genetic diversity of subpopulations

As per the STRUCTURE analysis, the number of population was assumed to be four (*k=4*) based on rate of change in *LnP(D)* between successive *k*, stability of grouping pattern across five runs and germplasm information about the material under study (Figure 1). Out of five runs for *k*=4, the run with highest likelihood value was selected to assign the posterior membership coefficient (Q) to each accession. A graphical bar plot was then generated with the posterior membership coefficient (Figure 2). The four sub-populations (SP) as inferred in the STRUCTURE analysis were named as SP_1, SP_2, SP_3 and SP_4, respectively. Overall proportions of membership of the accessions in each of the four subpopulations (SP_1 to SP_4) were 0.365, 0.045, 0.058 and 0.533, respectively. SP_1 contained 59 accessions and represented all three races, of which, *moharia* dominated with 27 accessions followed by *maxima* (18 accessions) and *indica* (14 accessions). SP_2 consisted of four accessions; three belongs to race, *maxima*, while all the seven accessions in SP_3 belong to *indica* race. Eighty-one of the 84 accessions in SP_4 belong to race *indica*.

The genetic diversity was assessed for each subpopulation (Table 1). SP_1 had the highest gene diversity (0.80), number of alleles per locus (14.40), Shannon's information index (2.14) and population specific alleles (417) compared with other subpopulations. Among 1,202 alleles detected in core collection, 1,037 alleles were found in SP_1, 180 alleles in SP_2, 185 alleles in SP_3 and 728 alleles in SP_4. SP_1 had the highest population-specific or private allele followed by SP_4. Among the 1,037 alleles detected in SP_1, 417 alleles were private alleles. A total of 13 private alleles in SP_2, 4 in SP_3 and 126 in SP_4 were found. Sample size of SP_1 is smaller than SP_4, exhibited relatively high gene diversity. Eighty-one of the 84 accessions in SP_4 belong to race

indica of which 74 accessions were from India, showed the maximum gene diversity, PIC and allele number after SP_1. Among 94 accessions from India, 74 accessions were grouped together under the SP_4 and seven accessions were grouped under SP_3. In total, among 155 accessions, 94 accessions are from India, 12 accessions from Sriya, seven accessions from Russia and CIS, six from China, five each from Korea and USA, and other countries are represented by less than three accessions (Table 2).

Assessment of population structure

The genetic relationship among the sub-populations was measured by Nei's genetic distance and pairwise estimate of F_{ST} (Table 3). The pairwise F_{ST} was the highest between SP_2 and SP_3 (0.385) followed by SP_2 and SP_4 (0.248). Pairwise estimates of F_{ST} values were found to be significant between populations (Table 3), indicate the existence of significant genetic differentiation among subpopulations. The genetic distance data agreed with the F_{ST} estimate. The SP_3 showed the lowest genetic distance with SP_4 (0.256), and SP_1 showed the lowest genetic distance with SP_4 (0.300) whereas, SP_2 showed the greatest genetic distance with SP_3 (1.141) followed by SP_2 with SP_4 (0.842).

PCoA and the neighbor-joining phylogenetic analysis were performed to further assess the population subdivisions. In the PCoA, the first three PCos explained 66.9% variation, of which, PCo1 and PCo2 contributed 37.3% and 16.4%, respectively of the SSR variation among the 155 accessions (Figure 3a and 3b). Plotting the first two PCos and coding of genotypes according to three biological races of foxtail millet (Figure 3a) shows clear separation of the race *indica*, most of which were present in SP_4 (Figure 2). The race *maxima* and *moharia* were not clearly separated as in SP_1. Further, plotting the first two PCos and coding of genotypes according to the four sub-populations identified using STRUCTURE shows the clear separation of four subpopulations (SP_1 to SP_4) (Figure 3b). The neighbor-joining tree of 155 core accessions and color coding of genotypes revealed that SP_1 (Red) and SP_4 (Blue) were the major subpopulations along with two small subpopulations, SP_2 (Green) and SP_3 (Black) (Figure 4), fairly corresponded to STRUCTURE analysis. Further, AMOVA on the basis of sub-populations and the three races of foxtail millet showed consistent relationship, representing high intra-population variation, which confirmed that the population has obvious structure (Table 4).

Linkage disequilibrium (LD)

The extent of LD was assessed among 2,556 marker pairs for all accessions. The r^2, the square of the correlation coefficient between two loci was used to measure LD. In a total of 2,556 pair wise comparisons (352 linked and 2,204 unlinked marker pairs) on the basis of 72 mapped SSR loci, 67%, 53% and 39% of SSR marker pairs showed significant LD at $P \leq 0.05$, $P \leq 0.01$ and $P \leq 0.001$, respectively. At the whole population level, the r^2 ranged from 0.0008 to 0.19 and 1,430 pairs of loci were significant at $P \leq 0.01$. Scatter plot of the LD values based on the r^2 values of 155 accessions are shown in Figure 5, where LD values for inter-chromosomal markers are compiled in a single file at 350 cM. Among the inter-chromosomal pairs, r^2 ranged from 0.01 to 0.17. At intra-chromosome level, LD was very common for distances 40 cM.

Discussion

Genetic diversity of foxtail millet core collection

In this investigation, a foxtail millet core collection was characterized using 72 SSR markers. High level of polymorphism was observed for all the SSR loci. The average number of alleles per locus was 16.69. The number of alleles ranged from 4 to 35, which was higher than the earlier reports in foxtail millet (6.16, Jia et al. 2009; 14.04, Liu et al. 2011; 2.4, Lin et al. 2012) and less than the Chinese foxtail millet landraces (6-47, Wang et al. 2012) and Chinese green foxtail millet (33.5, Jia et al. 2013a), illustrated that ICRISAT foxtail millet core collection contain high genetic diversity, which could provide valuable and important gene resources for foxtail millet breeding programs and for genomic study. Higher diversity reported in the 250 accession of Chinese foxtail millet (Wang et al. 2012) might be due to its larger size, represents 1% of foxtail millet kept in the Chinese National Gene Bank (CNGB), which is larger than foxtail millet conserved at ICRISAT, Patencheru, India. In comparison with that of foxtail millet landraces, Jia et al. (2013a) reported the higher alleles per locus in green foxtail millet, shows that, a large part of the genetic diversity in the wild gene pool was lost during domestication of foxtail millet and necessity for germplasm collection and protection of the wild relatives of crops (Jia et al. 2013a). Higher genetic diversity of germplasm is favorable for genetic marker development, construction of segregating population, functional gene cloning and association mapping and provides enriched gene resources for gene mining in the grass family (Wang et al. 2012). Compared with green foxtail millet, large part of the genetic diversity in the wild gene pool was lost during the domestication of foxtail millet. Average heterozygosity in the foxtail millet core collection was 0.04, which is less than earlier reports in foxtail millet and green millet (0.07-0.19) (Lin et al. 2012; Kumari et al. 2013; Jia et al. 2013a), which indicate that, the accessions used in the present study are very much close to inbred lines.

Population structure and relationship between population structure with racial classification and geographical distribution

Association mapping with diverse germplasm or wild populations can identify new superior alleles that were not captured by breeding practices and supports introgression of these alleles into elite breeding germplasm (Kumar et al. 2007). However, understanding population structure is essential to avoid spurious association between phenotype and genotype in association

mapping (Pritchard et al. 2000b). A model based approach implemented in the software STRUCTURE is the most frequently used approach. In the present study STRUCTURE analysis revealed the existence of population subdivision and identified four sub-populations in foxtail millet core collection. This population structure might be due to the presence of three races of foxtail millet, *indica*, *maxima* and *moharia*. The total number of alleles, number of alleles per locus, PIC and gene diversity was more in SP_1 and SP_4, which can be explained by the difference in the sample size and diversity within the sub-population, which can lead to differential allelic richness (Yang et al. 2010). Sample size of SP_1 is smaller than SP_4, exhibited relatively high gene diversity, and is due to presence of all three races from different geographic regions.

The 155 accessions of the foxtail millet core collection were from 23 countries, with the majority being from India, and a few accessions were from other countries, we did not obtain a clear-cut grouping of the accessions based on either countries or regions of origin. However, the population subdivisions detected through STRUCTURE, PCoA and neighbor-joining phylogenetic analysis, F_{ST} and genetic distance were in accordance with racial classification of foxtail millet. Eighty-one of the 84 accessions in SP_4 belong to race *indica* of which 74 accessions were from India, showed the maximum gene diversity, PIC and allele number after SP_1 suggesting that, there is lot of genetic variation in foxtail millet germplasm collected from India. The distribution of molecular genetic variation among and within the four sub-populations and three races revealed high degree of variability within each group, and low level of genetic distance among the groups. This results in agreement with earlier study in foxtail millet (Wang et al. 2012) and other crops (Abdurakhmonov et al. 2008; Peleg et al. 2008a; Peleg et al. 2008b; Jun et al. 2008).

Linkage disequilibrium (LD) in ICRISAT foxtail millet core collection

Foxtail millet, as a self-pollinating species, is expected to have a high level of LD (Wang et al. 2012). Wang et al. (2010) reported the increased level of LD in the domesticated foxtail millet (extends to 1 kb), while it decayed rapidly to a negligible level within 150 bp in wild green foxtail, suggested that, the increased level of LD in the cultivated foxtail millet was mainly due to the change of population size during the domestication process. Wang et al. (2012) reported the LD decay of less than 20 cM of genetic distance using SSR markers with 250 foxtail millet landraces. In self-pollinating crops such as barley, LD commonly extends for distances of up to

10 cM (Kraakman et al. 2004), and in some *Arabidopsis* populations, LD exceeds 50 cM (Nordborg et al. 2002). LD is high in rice, and LD decay values are 20 to 30 cM (Agrama et al. 2007). In our study, LD decay of approximately 40 cM of genetic distance was observed, which suggests that it could be possible to achieve resolution down to the 40 cM level.

The choice of germplasm is a key factor which determines the resolution of AM. The core collection, which represents the diversity of the entire collection of that species, would be the ideal material for association mapping (Whitt and Buckler 2003). The core collection was effectively used as association mapping panel in several crops *viz.*, rice (Borba et al. 2010), wheat (Dodig et al. 2012), sorghum (Shehzad et al. 2009; Upadhyaya et al 2012a; Upadhyaya et al. 2012b), common bean (Blair et al. 2009) etc. Availability of huge number of molecular markers in foxtail millet (Jia et al. 2009; Kumari et al. 2013; Pandey et al. 2013; Muthamilarasan et al. 2013) provides an immense applicability in germplasm characterization, phylogenetics, gene/quantitative trait loci discovery and comparative mapping. It is now possible to do genome-wide association mapping. Pandey et al. (2013) identified 28,342 microsatellite repeat-motifs spanning 405.3Mb of foxtail millet genome, of the 28,342 microsatellites, 21,294 primer pairs were successfully designed, and a total of 15,573 markers were physically mapped on 9 chromosomes of foxtail millet. Muthamilarasan et al. (2013) developed 5123 intron-length polymorphic (IPL) markers of which 4049 were physically mapped onto 9 chromosomes of foxtail millet. Jia et al. (2013b) sequenced 916 diverse foxtail millet varieties, identified 2.58 million SNPs and used 0.8 million common SNPs to construct a haplotype map of the foxtail millet genome. Also phenotyped the 916 diverse foxtail millet under five different environments and identified 512 loci associated with 47 agronomic traits by genome-wide association studies.

The phenotypic variations of many complex traits of agronomic importance are influenced by genotype and genotype × environment interaction; hence replicated multi-environment testing is prerequisite and useful for correctly identifying QTLs associated with the trait of interest (Tao et al. 2000). This core collection is manageable in size (155 accessions), represents diversity of 1,474 foxtail millet accessions, can be extensively evaluated under replicated multi-environments for various economically important traits of interest, using relatively less resources. Availability of large number molecular markers in foxtail millet (Jia et al. 2009; Kumari et al. 2013; Pandey et al. 2013; Muthamilarasan et al. 2013), can be used to identify marker trait association using this core collection as an association panel.

Acknowledgements

The authors gratefully acknowledge the financial support of the BMZ/GTZ project on "Sustainable conservation and utilization of genetic resources of two underutilized crops-finger millet and foxtail millet- to enhance productivity, nutrition and income in Africa and Asia" funded by the Federal Ministry for Economic Cooperation and Development (BMZ), Germany to carry out this activity.

References

Abdurakhmonov IY and Abdukarimov A (2008) Application of association mapping to understating the genetic diversity of plant genetic resources. Int J Plant Genomics. doi: 10.1155/2008/574927

Agrama HA, Eizenga GC and Yan W (2007) Association mapping of yield and its components in rice cultivars. Mol Breed 19: 341-356

Austin DF (2006) Foxtail millets (*Setaria*: Poaceae) - Abandoned food in two hemispheres. Econ Bot 60: 143-158

Barton L, Newsome SD, Chen Fa-Hu, Wang H, Guilderson TP and Bettinger RL (2009) Agricultural origins and the isotopic identity of domestication in northern China. PNAS 106: 5523-5528

Bennetzen JL, Schmutz J, Wang H et al (2012) Reference genome sequence of the model plant *Setaria***.** Nat Biotechnol 30: 555-561

Blair MW, Díaz LM, Buendía HF and Duque MC (2009) Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris L.*). Theor Appl Genet 119: 955-972

Borba TCO, Brondani RPV, Breseghello F, Coelho ASG, Mendonça JA, Rangel PHN and Brondani C (2010) Association mapping for yield and grain quality traits in rice (*Oryza sativa* L.). Genet Mol Biol 33: 515-524

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y and Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635

Dodig D, Zoric M, Kobiljski B, Savic J, Kandic V, Quarrie S and Barnes J (2012) Genetic and association mapping study of wheat agronomic traits under contrasting water regimes. Int J Mol Sci 13: 6167-6188

Doust AN, Kellogg EA, Devos KM and Bennetzen JL (2009) Foxtail Millet: A sequence- driven grass model system. Plant Physiol 149: 137-141

Flint-Garcia SA, Thuillet AC, Yu JM, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM and Buckler ES (2005). Maize association population: A high-resolution platform for quantitative trait locus dissection. Plant J 44: 1054-1064

Frankel OH (1984) Genetic perspective of germplasm conservation. In: Arber W, Illmensee K, Peacock WJ and Starlinger P (eds) Genetic manipulations: Impact of man and society. Cambridge: Cambridge University Press: 161-170

Hirano R, Naito K, Fukunaga K, Watanabe KN, Ohsawa R and Kawase M (2011). Genetic structure of landraces in foxtail millet (*Setaria italica* (L.) P. Beauv.) revealed with transposon display and interpretation to crop evolution of foxtail millet. Genome 54:498-506

Holland JB (2007) Genetic architecture of complex traits in plants. Curr Opin Plant Biol 10:156-161.

Jia X, Zhang Z, Liu Y, Zhang C, Shi Y, Song Y, Wang T and Li Y (2009) Development and genetic mapping of SSR markers in foxtail millet (*Setaria italica* (L.) P. Beauv.). Theor Appl Genet 118: 821-829

Jia G, Shi S, Wang C, Niu Z, Chai Y, Zhi H and Diao X (2013a) Molecular diversity and population structure of Chinese green foxtail [*Setaria viridis* (L.) Beauv.] revealed by microsatellite analysis. J Exp Bot 64: 3645-3655

Jia G, Huang X, Zhi H, Zhao Y et al (2013b) A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). Nat Genet 45: 957–961

Jun TH, Van K, Kim MY, Lee SH and Walker DR (2008) Association analysis using SSR markers to find QTL for seed protein content in soybean. Euphytica 162: 179-191

Kraakman ATW, Niks RE, Van Den Berg PMMM, Stam P and Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168: 435-446

Kihara H and Kishimoto E (1942) Bastrade zwischen *Setaria italica* und *S. viridis.* (in Japanese with German summary). Bot Mag Tokyo 56:62-67

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22: 139-144

Kumar R, Qiu J, Joshi T, Valliyodan B, Xu D and Nguyen HT (2007) Single feature polymorphism discovery in rice. PLoS ONE 2:e284

Kumari K, Muthamilarasan M, Misra G, Gupta S, Subramanian A, Parida SK, Chattopadhyay D, Prasad M (2013) Development of eSSR-Markers in *Setaria italica* and their applicability in studying genetic diversity, cross-transferability and comparative mapping in millet and non-millet species. PLoS ONE 8(6): e67742. doi: 10.1371/journal.pone.0067742

Li HW, Li CH and Pao WK (1945) Cytological and genetical studies of the interspecific cross of cultivated foxtail millet, *Setaria italica* (L.) Beauv., and the green foxtail millet, *S. viridis* (L.). J Am Soc Agron 37: 32-54

Li P and Brutnell TP (2011) *Setaria viridis* and *Setaria italica*, model genetic systems for the panicoid grasses. J Exp Bot 62: 3031-3037

Lin HS, Liao GI, Chiang CY, Kuoh CS and Chang SB (2012) Genetic diversity in the foxtail millet (*Setaria italica*) germplasm as determined by agronomic traits and microsatellite markers. Aust J Crop Sci 6(2): 342-349

Liu K and Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker data. Bioinformatics 21:2128-2129

Liu L, Wang L, Yao J, Zhang Y and Zhao C (2010) Association mapping of six agronomic traits on chromosome 4A of wheat (*Triticum aestivum* L.). Mol Pl Breed 1(5): pp10

Liu Z, Bai G, Zhang D, Znu C, Xia X, Cheng Z and Shi Z (2011) Genetic diversity and population structure of elite foxtail millet (*Setaria italica* (L.) P. Beauv.) germplasm in China. Crop Sci 51:1655-1663

Lu H, Zhang J, Liu KB, Wu N, Li Y, Zhou K, Ye M, Zhang T, Zhang H, Yang X, Shen L, Xu D and Li Q (2009) Earliest domestication of common millet (*Panicum miliaceum*) in East Asia extended to 10,000 years ago. PNAS 106: 7367-7372

Mace ES, Buhariwalla HK and Crouch JH (2003) A high-throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol Biol Repor 21: 459a-459h

Malosetti M, van der Linden CG, Vosman B and van Eeuwijk FA (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. Genetics 175:879-889

Muthamilarasan M, Venkata Suresh B, Pandey G, Kumari K, Kumar Parida S and Prasad M (2013) Development of 5123 intron-length polymorphic markers for large-scale genotyping applications in foxtail millet. DNA Res. doi:10.1093/dnares/dst039

Murray SC, Rooney WL, Hamblin MT, Mitchell SE and Kresovich S (2009) Sweet sorghum genetic diversity and association mapping for brix and height. Plant Genome 2:48-62

Nei M (1972) Genetic distance between populations. Amer Nat 106: 283-392

Nei M (1973) Analysis of gene diversity in subdivided populations. PNAS 70: 3321-3323

Neumann K, Kobiljski B, Denčić S, Vershney RK and Börner A (2011) Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). Mol Breed 27: 37-58

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in Arabidopsis thaliana. Nat Genet 30: 190-193

Pandey G, Misra G, Kumari K, Gupta S, Parida SK, Chattopadhyay D and Prasad M (2013) Genome-wide development and use of microsatellite markers for large-scale genotyping applications in foxtail millet [*Setaria italica* (L.)]. DNA Res 20: 197-207

Peakall R and Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecology Notes 6(1): 288-295

Perrier X and Jacquemoud-Collet JP (2006) DARwin software (2006) http://darwin.cirad.fr/darwin

Peleg Z, Fahima T, Abbo S, Krugman T and Saranga Y (2008a) Genetic structure of wild emmer wheat populations as reflected by transcribed versus anonymous SSR markers. Genome 51: 187-195

Peleg Z, Saranga Y, Krugman T, Abbo S, Nevo E and Fahima T (2008b) Allelic diversity associated with aridity gradient in wild emmer wheat populations. Plant Cell Environ 31: 39-49

Prasada Rao KE, de Wet JMJ, Brink DK and Mengesha MH (1987) Intraspecific variation and systematics of cultivated *Setaria italica*, foxtail millet (Poaceae). Econ Bot 41:108-116

Pritchard JK, Stephens M, Rosenberg NA and Donnelly P (2000a) Association mapping in structured populations. Am J Hum Genet 67:170-181

Pritchard JK, Stephens M and Donnelly P (2000b) Inference of population structure using multilocus genotype data. Genetics 155: 945-959

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM and Buckler ES IV (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. PNAS 98: 11479-11484

Shehzad T, Iwata H and Okuno K (2009) Genome-wide association mapping of quantitative traits in sorghum (*Sorghum bicolor* (L.) Moench) by using multiple models. Breed Sci 59: 217-227

Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M and Reifl JC (2005). Linkage disequilibrium in European elite maize germplasm investigated with SSRs. Theor Appl Genet 111: 723-730

Tao YZ, Henzell RG, Jordan DR, Butler DG, Kelly AM, McIntyre CL (2000) Identification of genomic regions associated with stay green in sorghum by testing RILs in multiple environments. Theor Appl Genet 100: 1225-1232

Upadhyaya HD and Ortiz R (2001) A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. Theor Appl Genet 102: 1292-1298.

Upadhyaya HD, Pundir RPS, Gowda CLL, Reddy VG and Singh S (2008) Establishing a core collection of foxtail millet to enhance utilization of germplasm of an underutilized crop. Plant Genet Resour: Characterization Utiliz 7:177-184

Upadhyaya HD, Wang Yi-Hong, Sharma Shivali and Singh S (2012a) Association mapping of height and maturity across five environments using sorghum mini core collection. Genome 55:471-479

Upadhyaya HD, Wang Yi-Hong, Sharma Shivali, Singh S and Hasenstein KH (2012b) SSR markers linked to kernel weight and tiller number in sorghum identified by association mapping. Euphytica 187: 401-410

Vetriventhan M, Upadhyaya HD, Anandakumar CR, Senthilvel S, Parzies HK, Bharathi A, Varshney RK and Gowda CLL (2012). Assessing genetic diversity, allelic richness and genetic relationship among races in ICRISAT foxtail millet core collection. Plant Genet Resour: Characterization Utiliz 10: 214-223

Wang C, Chen J, Zhi H, Yang L, Li W, Wang Y, Li H, Zhao B, Chen M and Diao X (2010) Population genetics of foxtail millet and its wild ancestor. BMC Genetics 11: 90

Wang C, Jia G, Zhi H, Niu Z, Chai Y, Li W, Wang Y, Li H, Lu P, Zhao B and Diao X (2012) Genetic diversity and population structure of Chinese foxtail millet [*Setaria italica* (L.) Beauv.] landraces. G3: Gene Genome Genet 2: 769-777

Wang R, Yu Y, Zhao J, Shi Y, Song Y, Wang T and Li Y (2008) Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China. Theor Appl Genet 117:1141-1153

Wen W, Mei H, Feng F, Yu S, Huang Z, Wu J, Chen L, Xu X and Luo L (2009) Population structure and association mapping on chromosome 7 using a diverse panel of Chinese germplasm of rice (*Oryza sative* L.). Theor Appl Genet 119: 459-470

Whitt SR and Buckler ES (2003) Using natural allelic diversity to evaluate gene function. Methods Mol Biol 236:123-139

Yang X, Yan J, Shah T, Warburton ML, Li Q, Li L, Chai YG, Fu Z, Zhou Y, Xu S, Bai G, Meng Y, Zheng Y and Li J (2010) Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. Theor Appl Genet 121: 417-431

Yu J and Bucklerv (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotech 17:155-160

Zhang P, Li J, Li X, Liu X, Zhao X and Lu Y (2011) Population structure and genetic diversity in a rice core collection (*Oryza sativa* L.) investigated with SSR Markers. PLoS ONE 6: e27565

Zhang G, Liu X, Quan Z et al (2012) Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. Nat Biotech 30:549-554

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P and Nordborg M (2007) An *Arabidopsis* example of association mapping in structured samples. PLoS Genet 3(1):e4

Zhu CS, Gore M, Buckler ES and Yu JM (2008) Status and prospects of association mapping in plants. Plant Genome 1:5-20

Table 1 Molecular diversity of sub-populations detected by STRUCTURE using 72 SSR markers

Statistics	Overall	Sub-population (SP) detected by STRUCTURE software			
		SP_1	SP_2	SP_3	SP_4
Sample size	155	59	4	7	85
Total number of alleles	1202	1037	180	185	728
Average number of alleles per locus	16.69 (4-35)[a]	14.40 (3-30)	2.50 (1-4)	2.57(1-5)	10.04 (1-23)
Standard Error (SE)	0.920	0.790	0.080	0.120	0.700
Shannon's information index (*I*)	1.99 (0.78-3.14)	2.14 (0.38-3.25)	0.80 (0-1.21)	0.70 (0-1.55)	1.14 (0-2.28)
Standard Error (SE)	0.092	0.005	0.037	0.048	0.095
Gene Diversity	0.73 (0.06-0.95)	0.80 (0.16-0.96)	0.50 (0-0.75)	0.41 (0-0.78)	0.63 (0-0.93)
Standard Error (SE)	0.026	0.022	0.021	0.027	0.032
Heterozygosity	0.04 (0-0.22)	0.04 (0-0.25)	0.01 (0-0.25)	0.01 (0-0.33)	0.04 (0-0.26)
Standard Error (SE)	0.004	0.039	0.007	0.007	0.005
PIC[b]	0.72 (0.06-0.94)	0.78 (0.15-0.95)	0.43 (0-0.70)	0.36 (0-0.73)	0.61 (0-0.92)
Population specific alleles number	35	417	13	4	126

[a] Values in parenthesis represents range

[b] PIC, Polymorphic information content

Table 2 Details of the accessions present in each subpopulation detected by STRUCTURE analysis

Country/Subpopulation			Race	Total
Sub Population I : Total 59 accessions				
	moharia (27)[#]	*maxima* (18)	*indica* (14)	
Afghanistan	2	-	-	2
China	-	5	-	5
Hungary	1	-	-	1
India	1	3	7	11
Iran	1	-	-	1
Korea,	-	4	-	4
Lebanon	3	-	-	3
Malawi	-	-	1	1
Nepal	-	1	-	1
Pakistan	2	-	-	2
Russia and CIS	4	2	-	6
Spain	1	-	-	1
Sri Lanka	-	-	1	1
Syria	7	1	3	11
Taiwan	1	1	1	3
Turkey	1	1	-	2
United Kingdom	-	-	1	1
USA	2	-	-	2
Unknown	1	-	-	1
Sub Population II : Total 4 accessions				
	maxima (3)	*indica* (1)	*moharia* (0)	
India	1	1	-	2
Korea	1	-	-	1
Russia and CIS	1	-	-	1
Sub Population III: Total 7 Genotypes				
	indica (7)	*maxima*	*moharia*	
India	7	-	-	7
Sub Population IV: Total 85 accessions				

	indica (81)	*moharia* (2)	*maxima* (2)	
China	1	-	-	1
Ethiopia	1	-	-	1
India	71	2	1	74
Kenya	1	-	-	1
Malwi	1	-	-	1
Nepal	0	-	1	1
Pakistan	1	-	-	1
South Africa	1	-	-	1
Syria	1	-	-	1
USA	3	-	-	3

[#]Numbers in parenthesis indicate number of accessions in each group

Table 3 Pairwise estimates of Nei's genetic distance (GD) and F_{ST} among the four subpopulations (SP) detected by STRUCTURE

Subpopulation	SP1	SP2	SP3	SP4
SP1		0.557	0.550	0.300
SP2	0.126**		1.141	0.842
SP3	0.172**	0.385**		0.256
SP4	0.100**	0.248**	0.139**	

F_{ST} estimate appear below the diagonal (** significant at $P \leq 0.01$) and pairwise Nei's genetic distance appears above the diagonal

Table 4 Analysis of molecular variance (AMOVA) based on four subpopulations (SP_1 to SP_4) detected by STRUCTURE and three races in the foxtail millet core collection

Source	df	Sum of square	Mean sum of square	Estimated Variance	Percentage of variation (%)
Based on subpopulations inferred by STRUCTURE					
Among populations	3	1250.3	416.8	11.1	10
Within population	151	15060.7	99.7	99.7	90
Total	154	16311.0		110.9	
Based on three races viz., *indica*, *maxima* and *moharia*					
Among populations	2	818.48	409.24	7.81	7
Within population	152	15495.80	101.95	101.95	93
Total	154	16314.28		109.75	

Fig. 1 Rate of change in *LnP(D)* between successive *k* (*k* averaged over the five run). *k* ranged from 2 to 10

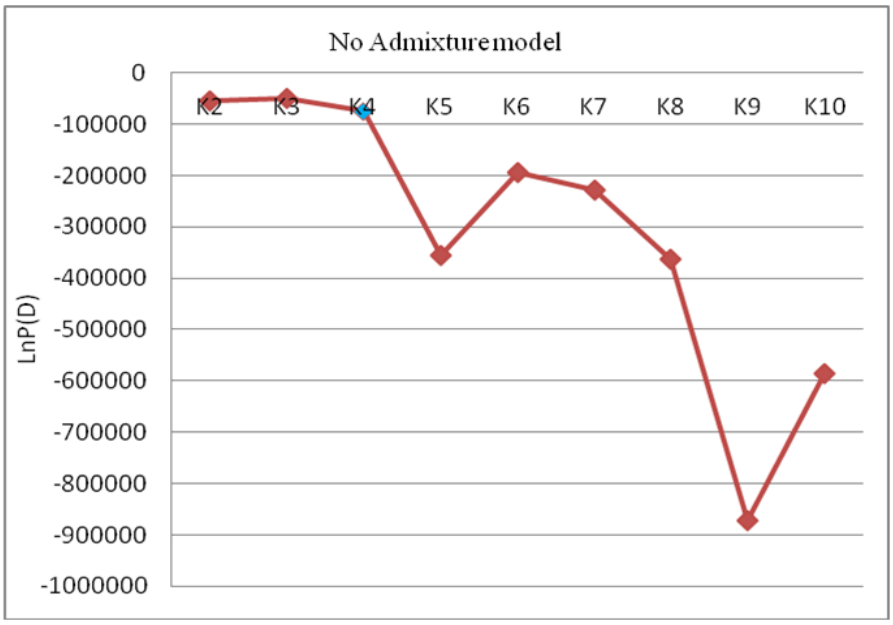


Fig. 2 Population structure of foxtail millet core collection based on 72 SSR markers. Numbers in the 'y' axis show the subgroup membership and each accession represented by thin vertical line, which is partitioned into four colored segments that represent the individual membership to the subpopulation. SP denotes subpopulation

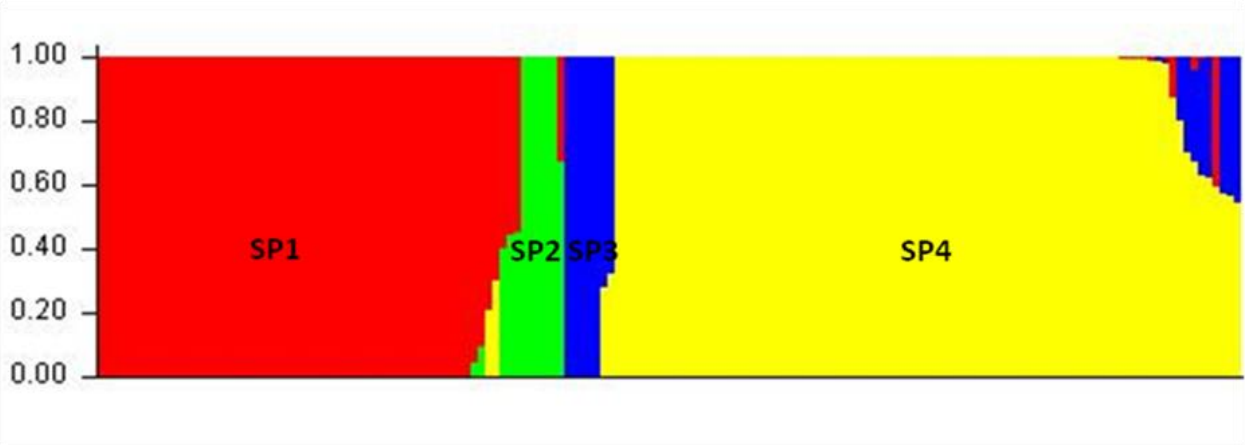
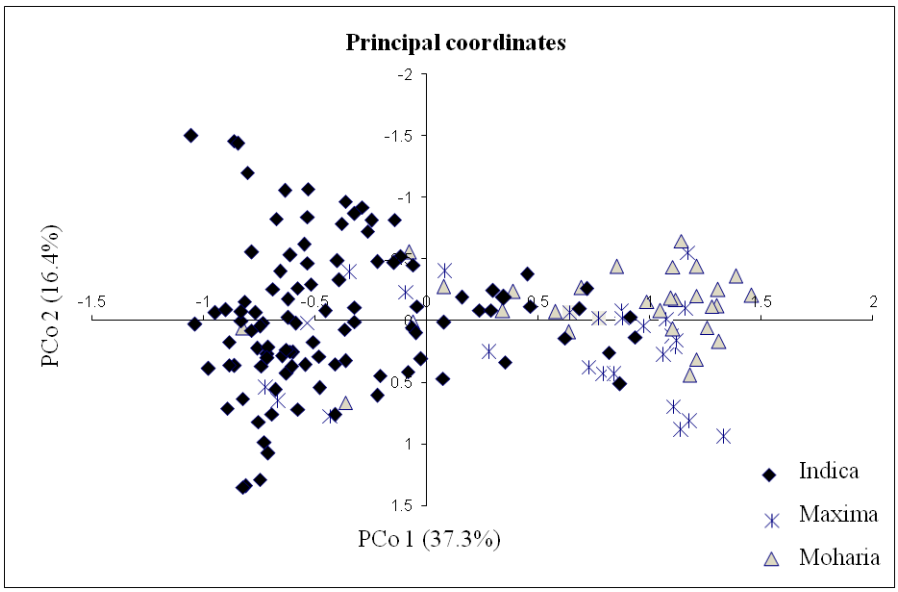


Fig. 3 Principal coordinate analysis (PCoA) of foxtail millet core collection accessions using 72 SSR markers based on Nei (1973) distance estimates. PCo1 and PCo2 are the first and the second principal coordinates, respectively. Numbers in parentheses refers to the proportion of variance explained by the principal coordinate. SP denotes subpopulation.

a. PCoA: coding based on three biological races



b. PCoA: coding based on four subpopulations identified in STRUCTURE analysis

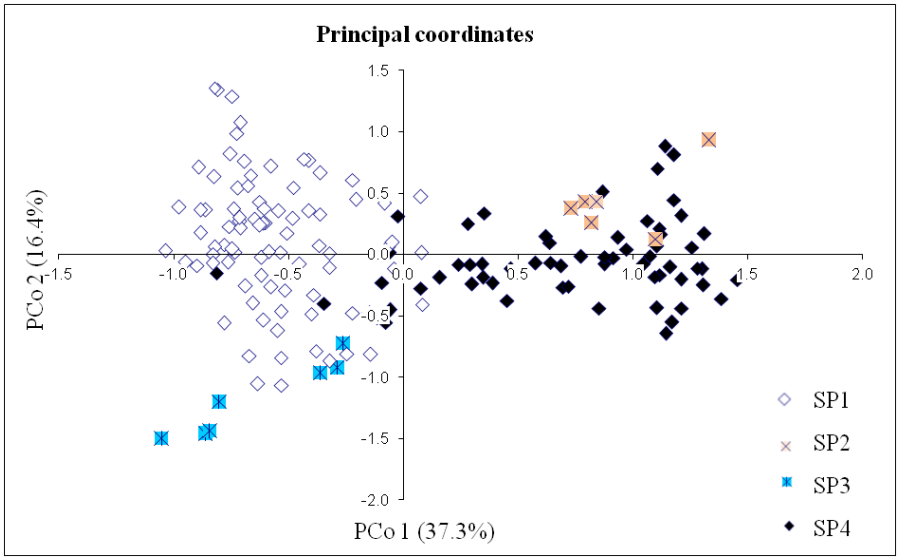


Fig. 4 Neighbor-joining tree based on the simple matching dissimilarity matrix of 72 markers genotyped across the foxtail millet core collection. Each color represents the different subpopulation (SP) identified in STRUCTURE analysis. Red: SP_1, Green: SP_2; Black: SP_3; Blue: SP_4.

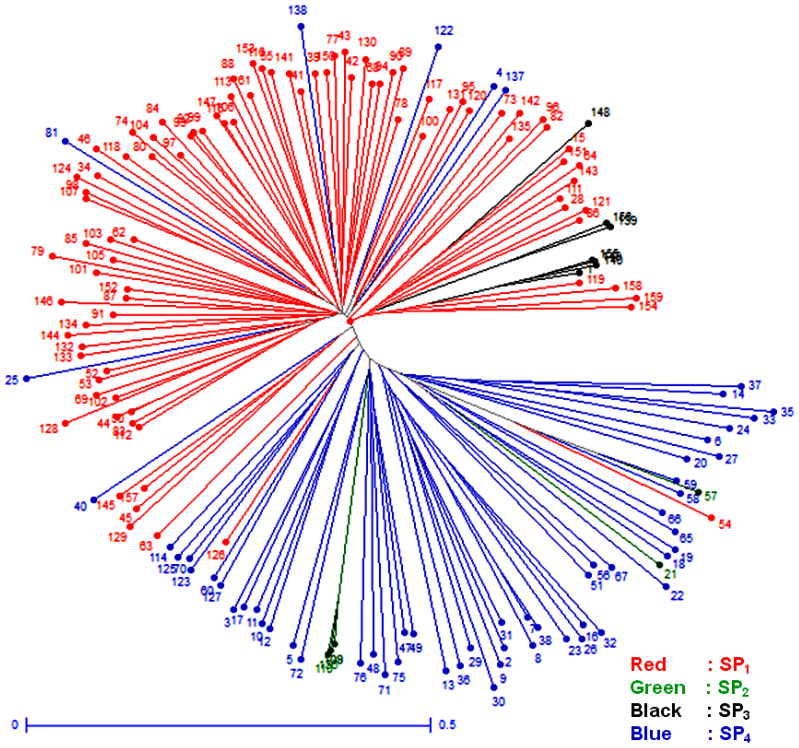


Fig. 5 The pattern of LD for 72 SSR loci indicating correlations of allele frequency (r^2) value against genetic distance (cM) between all loci pairs