

Full Length Research Paper

Evaluation and association mapping of agronomic traits for drought tolerance in sorghum [*Sorghum bicolor* (L.) Moench]

Aleye Endre^{1*} and Kassahun Bante²

¹Amhara National Regional Agricultural Beureau, Dessie Soil Testing Laboratory, Dessie, Ethiopia.

²Jimma University, Department of Plant Sciences, Institute of Biotechnology, Jimma, Ethiopia.

Received 14 April, 2016; Accepted 25 August, 2016

Drought is the major sorghum production constraint in Ethiopia which necessitates the identification of sorghum genotypes that carry genes (quantitative trait locus, QTL) associated with drought tolerance thereby developing drought tolerant sorghum varieties. The objectives of this study were to identify drought tolerant sorghum genotypes, map chromosomal regions (QTLs) associated with agronomically important traits and identify simple sequence repeat (SSR) markers tightly linked with these QTLs. One hundred and sixty (160) sorghum genotypes (152 landraces and 8 released varieties) were genotyped with 39 SSRs markers and evaluated in the field at Kobo in the off-season using an alpha lattice design replicated three times. Phenotypic data including days to 50% flowering, plant height, panicle weight, grain weight, grain weight per panicle, panicle harvest index, one thousand grain weight and number of grains per panicle were collected. Analysis of variance showed highly significant ($P < 0.0001$) differences among the genotypes for all characters. Most of the characters showed moderate to high phenotypic and genotypic coefficient of variation. Linkage disequilibrium (LD) analysis indicated that in all accessions, 107 loci pairs (32.92%) had a significant ($p < 0.05$) mean LD of 0.19, with $R^2 > 0.2$ for 33 evaluated loci pairs. Population structure analysis showed that there were four distinct clusters in the studied materials. A total of 10 marker-trait associations were identified using seven different SSR markers. The percentage of the total variation explained by the markers ranged from 2.6% (Xtxp114 with THGT) to 17.76% (Xtxp145 with PHT). The seven SSR markers (xcup53, bSbCIR223, Xtxp114, mSbCIR248, Xtxp145, Xtxp278, and gbsp123) were located on chromosomes 1, 2, 3, 5, 6, 7 and 8, respectively, each chromosome harboring one marker. Most of the identified markers were localized in chromosomal positions that have been previously reported as positions for drought tolerance-related traits, supporting the present findings. The results of this study can serve as initial effort for the association mapping studies in sorghum particularly in Ethiopia as the associated SSR markers are potential candidates for marker-assisted selection to improve drought tolerance in sorghum. However, as this study is the first attempt in the identification of QTLs for drought tolerance using association mapping, the identified QTLs need to be validated in independent or related populations and in different environments before their use in marker-assisted selection.

Key words: Association mapping, drought, population structure, quantitative trait locus (QTLs), sorghum, simple sequence repeat (SSR).

INTRODUCTION

Sorghum [*Sorghum bicolor* (L.) Moench] is a largely self-pollinating (70-95%) monocot crop in the grass family of Poaceae with a diploid set of chromosomes ($2n=2x=20$) and an estimated genome size of 750 Mb (Doggett, 1976; Yonemaru et al., 2009). It serves as a staple food for the world's most food insecure people, particularly in the semi-arid tropics of Asia and Africa. More than 35% of world's sorghum production is dedicated to human consumption of which 95% is in developing countries; the rest being used mainly for animal feed, alcohol and industrial products (Dicko et al., 2006). Worldwide, sorghum ranked fifth in production after maize, wheat, rice and barley (FAOSTAT, 2010). In Ethiopia, it ranked third after maize and tef with a total production of 2.8 million metric tons and an average yield of 1,736 kg/ha (FAOSTAT, 2010).

Despite its importance, sorghum productivity is severely limited by drought accounting for more than 50% yield losses each year globally (Hao et al., 2011). The severity increases particularly in developing countries like Ethiopia, where the majority of the people depend on agriculture for their livelihood (Hao et al., 2011). In Ethiopia, sorghum is largely cultivated in moisture stress areas that cover nearly 66% of the total area of the country (Tadesse et al., 2008). This necessitates evaluation and identification of genomic regions that confer resistance to drought stress particularly at the reproductive stage thereby developing drought tolerant varieties. Sorghum is naturally moisture stress tolerant crop which can be more productive with moderate genetic improvement. In sorghum, two distinct drought responses are recognized, pre-flowering and post-flowering drought responses (Sanchez et al., 2002). The pre-flowering response occurs when the plants are under significant moisture stress before flowering (Tuinstra et al., 1996). Post-flowering drought response in sorghum is expressed when moisture stress occurs during the grain development stage (Tuinstra et al., 1996). Drought at any stage of crop development affects growth and production but drought during the flowering stage causes maximum crop damage (Ejeta and Knoll, 2007). Under water-limited environments, genetic improvement of crops for drought tolerance is a sustainable and economically feasible solution to reduce the impact (Tadesse et al., 2008). However, in quantitatively inherited traits such as drought tolerance that are controlled by many genes each with small effects (QTLs), selection by conventional methods based on phenotypic variations is inefficient and challenging due to the complex nature of the trait and the complicating effects of the environment. Through QTL

mapping, marker assisted selection (MAS) using DNA markers has become more efficient to tag such traits (Hausmann et al., 2002). The most common approach of QTL mapping is to identify QTLs in a bi-parental population (Shehzad et al., 2009). Another approach being applied is association mapping, which uses diverse populations to identify associations between allele frequencies and phenotypic variations (Sorkheh et al., 2008). Unlike QTL mapping where bi-parental crosses with contrasting genotypes are used to generate a mapping population, association mapping is an approach where a collection of cultivars, lines, and/or landraces, genotyped with densely spaced markers, can be used as mapping population (Sorkheh et al., 2008; Myles et al., 2009). Using a collection of cultivars has a number of advantages over the use of a bi-parental cross. Firstly, in the population, a broader genetic variation in a more representative genetic background will be available implying that one is not limited to the marker and trait loci that happen to differ between two parents. Secondly, LD mapping can attain a higher resolution because of the use of all meioses (recombinations) accumulated in the breeding (selection) history. Thirdly, historic phenotypic data on cultivars can be used to link markers to traits, without the need to develop bi-parental mapping populations (Sorkheh et al., 2008). Thus, originally developed for human genetics, association mapping strategy now is being exploited in several crop plants. In sorghum, Shehzad et al. (2009) reported a total of 14 significant SSR loci associated with traits including days to heading, days to flowering, number of panicles and panicle length using 107 representative sorghum accessions and 98 SSR markers. Wang et al. (2012) reported two SSR markers consistently associated with plant height at two environments. Upadhyaya et al. (2012) reported significant associations of five markers with maturity date and plant height on chromosomes 6, 9, and 10 using 242 sorghum accessions and 39 SSR markers. These works emphasize the possibility of genetic improvement in sorghum using the existing germplasm resources (including land races) by evaluating in field; and identifying and mapping QTLs associated with desired traits and selecting the genotypes (parents) that carry favorable alleles for gene introgression through marker assisted selection. Ethiopia is rich in sorghum landraces as sources of desirable genes to screen genotypes for better agronomic performance under moisture stress (Amsalu et al., 2000). However, no study to detect marker-trait association for moisture stress tolerance in sorghum have been previously reported in Ethiopia using association mapping strategy which initiated the present study.

*Corresponding author. E-mail: aleye.endre@gmail.com.

MATERIALS AND METHODS

Phenotyping

The study area, plant materials and experimental design

The field experiment was conducted in Ethiopia at Kobo agricultural research site. Kobo is located 581 km north of Addis Ababa, at an altitude of 1468 masl. (12°9' N latitude and 39°38'E longitude). The entries for this study consisted of 160 sorghum (*S. bicolor*) accessions (152 land races and 8 released lines) which were sampled from more than 1800 sorghum accessions collected by the Institute of Biodiversity Conservation (IBC) of Ethiopia. The samples were selected on the basis of their geographic distribution from all regions of the country representing sorghum growing areas with altitude range of 500 to 1850; more accessions from areas predominantly affected by moisture stress (regions from which the accessions were collected are shown in Supplementary Table 1). The seeds of released varieties were provided by Melkasa and Sirinka Agricultural Research Centers. These experimental materials were grown at Kobo during the off season of 2011 using irrigation from January to June (Kobo is one of the best sites in the country for sorghum variety performance trial, particularly, for drought tolerance as the site is located in the area where sorghum is predominantly grown but frequently affected by moisture stress). Mean monthly temperature and rain fall at the experimental site during the cropping period is shown in Supplementary Table 2. The field experiment was laid down in alpha lattice design with three replications having 16 blocks per replication and 10 plots per block and a spacing of 75 cm and 20 cm between rows and plants, respectively. Forty (40) plants were planted per plot in two rows of 4 m long. The experimental plots were irrigated immediately after sowing to ensure uniform germination. Weekly interval irrigation was applied for the first three weeks. Starting from the fourth irrigation, water was supplied with 12 days interval till 50% flowering (Tuinstra et al., 1998; Xu et al., 2000). Irrigation was withheld when the majority of the entries reached 50% flowering creating a terminal water deficit which typifies the dry season of the semiarid tropics, where crops are usually grown on a depleted soil moisture profile. The recommended fertilizer rate of 100 kg Di ammonium phosphate (DAP) was applied by incorporating it into the soil during sowing the seeds followed by 25 kg urea ha⁻¹ by side dressing 55 days after the seeds were sown. Thinning was conducted after three weeks of sowing to maintain the plant distance and to balance the plant density. Karate was applied two times with a rate of 1 mm⁻¹ of water 30 and 45 days after emergence to protect against shoot fly. No herbicide was applied to control weeds. Bird damage was protected by covering the heads of 10 randomly chosen plants from each plot to obtain the average grain yield per panicle. Other agronomic practices such as weeding were applied uniformly to all plots according to the recommended practices. Phenotypic data were collected for seven traits (50%FL-days to 50% flowering, PHT-plant height, PWT-panicle weight, GWPP-grain weight per panicle, THGT-thousand seed weight, NGPP-number of seeds per panicle and PHI-Panicle harvest index) based on sorghum descriptor list (IBPGR and ICRISAT, 1993).

Genotyping

DNA extraction

The 160 accessions were grown in greenhouse and the fresh leaves of 10 plants from 14 days old seedlings were harvested in bulk and dried with silica gel. DNA was extracted following a modified cetyl trimethyl ammonium bromide (CTAB) extraction

protocol (Mace et al., 2003).

The quality and quantity of the isolated DNA was determined by comparing DNA samples with a known concentration of λ -DNA after running them on a 0.8% agarose gel (0.8 gm agarose dissolved in 100 ml 1X TBE buffer) that contained 0.3 μ g/ml ethidium bromide solution. At the end of electrophoresis, the gel was visualized using ultraviolet (UV) light and photographed using a video capture (Flowgen IS 1000). All samples were then normalized to the same concentration level (50 ng) and used for PCR.

SSR markers used

A total of 39 simple sequence repeat (SSRs) markers, including 22 di, 9 tri, and 4 tetra nucleotide or longer motifs, and 4 compound repeats were used. These SSR markers were selected based on their uniform distribution in the sorghum genome. Four of them from chromosome SBI-01, five of them from chromosome SBI-02, four of them from chromosome SBI-03, two of them from chromosome SBI-04, four of them from chromosome SBI-05 and chromosome SBI-06 each, five of them from chromosome SBI-07 and chromosome SBI-08 each and three of them from chromosome SBI-09 and chromosome SBI-10 each. These are the same set of markers that are selected and being used by the Generation Challenge Programme for genetic diversity assessment of global sorghum germplasm.

Polymerase chain reaction (PCR)

The PCR was performed in Nairobi (Kenya) using Gene-Amp PCR System 9600 (PE-Applied Biosystems) in 384-wells plates (ABGene, Rochester, New York.) in a total reaction volume of 10 μ l that consisted of 1 μ l DNA (50 ng), 1 μ l 10X PCR buffer, 1.5 μ l MgCl₂ (10 mM), 1 μ l reverse primer (2 pmoles), 1 μ l forward primer (2 pmoles), which were 5'-labelled with one of the 6-FAM, VIC, NED, PET fluorescent dyes (PE-Applied Biosystems), 0.5 μ l of each dNTP (2 mM), 0.04 μ l Taq DNA polymerase (5U) (PE-Applied Biosystems) and 3.46 μ l distilled water. The amplification profile consisted of initial denaturation of the template DNA at 95°C for 3 min, followed by 35 cycles, each for 30 s at 95°C (denaturation), 1 min at 56°C (annealing), and 1 min at 72°C (extension), and a final extension at 72°C for 30 min was included to minimise the +A overhang.

Capillary electrophoresis

After the PCR, a few samples from each primer pair product were randomly selected and checked for proper amplification by comparing DNA samples with a known molecular weight of λ -DNA after running them on 2% agarose gel. An ABI plate was prepared with a total volume of 10 μ l (9.0 μ l from a mix of an injection solution mixed by vortexing (1 ml) formamide (HIDI) (Perkin Elmer-Applied Biosystems) and 12.0 μ l GS500 LIZ (Perkin Elmer-Applied Biosystems) was aliquoted into 96-well plates and 1.0 μ l of pooled PCR products from each of the 6-FAM, VIC, NED and PET-labelled PCR products was added. DNA fragments were denatured at 95°C for 3 min, chilled quickly for five minutes and size-fractionated using ABI 3730 Capillary DNA sequencer (PE-Applied Biosystems). In this system, the labeled PCR products were detected using a laser and capillary electrophoresis based on their fluorescent dye and fragment size. The peaks were sized and the alleles called using Gene Mapper software version 3.7 (PE-Applied Biosystems) and presented as alleles scored as estimated fragment sizes in base pairs compared to the internal size standard GS500LIZ-3730.

Data analysis

Phenotypic data

All collected phenotypic data were subjected to one way ANOVA using SAS software version 9.2 (SAS Institute Inc., 2008) for variances and heritability.

Molecular data analysis

Linkage disequilibrium analysis

LD (R^2) between SSR markers /loci/ was evaluated using Tassel software version 2.0.1. The LD was calculated using the statistical coefficient of determination (R^2) (Shi et al., 2010). Alleles with frequencies less than 0.05 were not included for LD calculation.

Population structure (Q-matrix) and kinship (K-matrix) analysis

Population structure (Q-matrix) among 160 sorghum accessions was analyzed using 39 SSR markers by *STRUCTURE* software version 2.3.3 (Pritchard et al., 2000). By setting the number of k levels (number of sub groups) from 1 to 9 with five times repetition for each k, nine independent structure runs were performed with 100,000 burn-in time and 100,000 iterations for each run. All *STRUCTURE* runs were performed using the admixture model with the option of correlated allele frequency. Also, the measure of the degree of admixture, alpha, was allowed to be inferred from the data (Pritchard et al., 2000), and Lambda, the parameter of the distribution of allelic frequencies, was set to one. The matrix of kinship coefficient comparing all pairs of the 160 lines using 39 SSR markers was calculated by the software package SPAGeDi as described by Loiselle et al. (1995). Negative kinship values between two individuals, indicating that there was less relationship than that expected between two random individuals, were changed to 0 and the diagonal was set to 2 (Pritchard et al., 2000).

Analysis of marker trait association

By fitting the population structure and kinship matrix into the model to avoid spurious associations, the trait marker association was evaluated using a mixed linear model (MLM) in "TASSEL" software version 3.0.1. To achieve linear independence, the structure matrix (Q-matrix) with one column less than the number of sub populations was used (Pritchard et al., 2000). The statistical model used for identifying SSR markers associated with traits was as follows:

$$Y_{klmfn} = \mu + k + ML + Am(ML)Kn + \epsilon_{kmn}$$

Where, Y_{klmfn} is the phenotypic observation, μ is the general mean, k is the fixed effect of k^{th} subgroup of the population structure (Q-matrix), ML is the fixed effect of L^{th} marker, $Am(ML)Kn$ is the random effect of m^{th} accession nested in the L^{th} marker associated with n^{th} kinship coefficient, and ϵ_{kmn} is the error. Only markers with an allele frequency $\geq 5\%$ were included in the association analysis. The significance of associations between loci and traits was based on an F-test with P values calculated by TASSEL at 5% significant level (Wang et al., 2012).

RESULTS AND DISCUSSION

The analysis of variance (ANOVA) for the studied characters showed that there was a highly significant ($P <$

0.0001) difference among genotypes for all characters indicating wide variability in performance among the genotypes. Most of the characters showed moderate to high phenotypic and genotypic coefficient of variation. High heritability coupled with high genetic advance as percentage of mean is the most promising clue for possibility of improvement by selection and was observed for PHT, PWT, GWPP, THGT and NGPP.

Population structure

The population structure analysis showed that the 160 sorghum accessions contained four distinct sub groups (Figures 2 and 3). Actually, the plot of the average log likelihood values over five runs for each K (ranging the k-values from 1 to 9) showed that the log likelihood estimates increase progressively as K increases (Figure 1) and did not show a clear peak to determine the true K (number of sub groups). To reliably detect the most probable number of sub-populations, the ad hoc criterion described by Evanno et al. (2005) was used and the number of sub-populations were found to be 4 (Figure 2) which was selected and used for association analysis. Plots of ancestry estimates provided the estimated membership coefficients for each individual in each cluster. Each sorghum variety is represented by a single vertical line, partitioned into K colored segments that represent individual varieties estimated membership fraction in each of the K inferred clusters (Figure 3). The population structure analysis also indicated that sorghum accessions were not clustered according to their areas of collections; rather in each cluster were sorghum varieties from different areas of collections. For example, the first group, G-1 (Figure 3) consisted of 52 accessions of which 20 were from Amhara, 16 from Oromia, 9 from Tigray, 1 from Afar and other 5 accessions which their geographical origin was not available. Similarly in the 2nd (G-2), 3rd (G-3) and 4th (G-4) groups' clusters were composed of accessions from different areas of collections.

The distribution of accessions into the four groups without reflecting their region of origin might indicate the presence of wide variations among accessions within the regions as well as lack of strong regional differentiation which might be due to gene flow between the regions. Similar results that showed lack of clustering based on the collection sites of sorghum accessions were reported by Alemu (2009).

Level of linkage disequilibrium

In this study, all 39 SSR markers were used to estimate the presence of LD in all accessions. After filtration of the data to exclude markers with less than 5% allele frequencies from the analysis, there were 325 pair wise locus comparisons for all accessions and the majority of

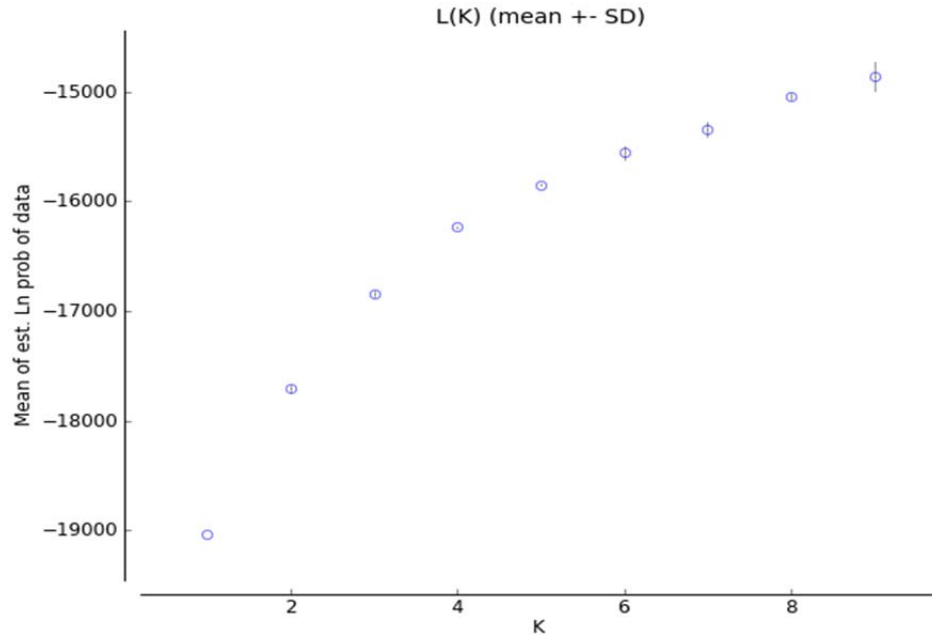


Figure 1. Posterior probability, $\ln P(D)$, of the data as a function of the number of subpopulations (k), where k was allowed to range from 1 to 9. Circles represent the average of the five independent runs for each value of k .

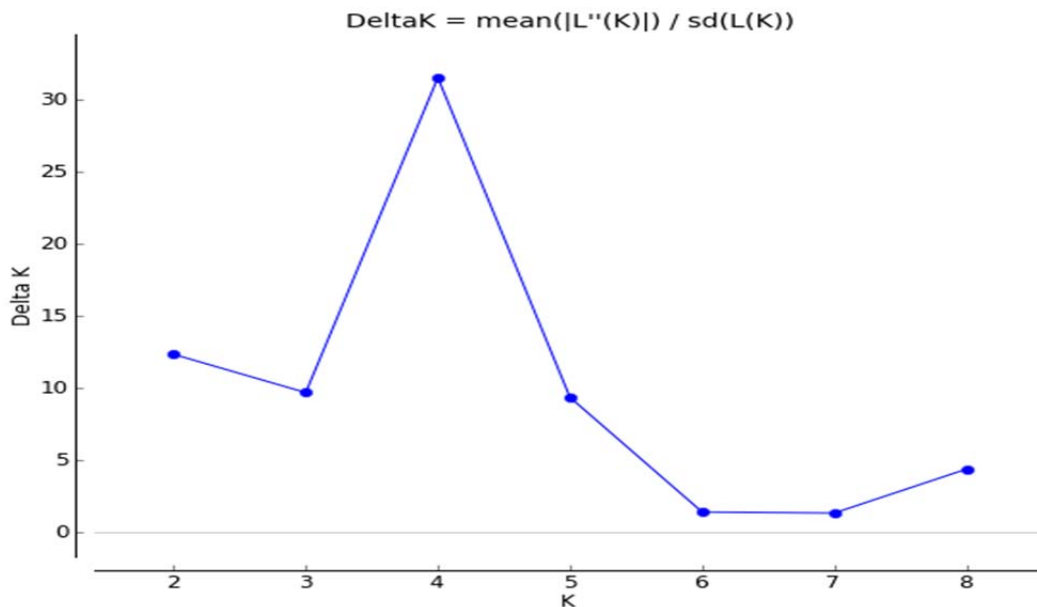


Figure 2. Values of K (x-axis), with its modal value used to detect the true K (y-axis) of four groups ($K = 4$).

loci pairs (67.077%) were independent loci (non-significant). In all accessions, 107 loci pairs (32.92%) had a significant ($p < 0.05$) mean LD of 0.19, with an $R^2 > 0.2$ for 33 evaluated loci pairs. However, the present study did not show a clear trend on linkage disequilibrium

decay (Figure 4) and no clear conclusions can be made regarding the decay of LD. This result might be explained by low number of markers used in this study. Similar results were reported by Shehzad et al. (2009) using 107 sorghum accessions and 98 SSR markers and Li et al.

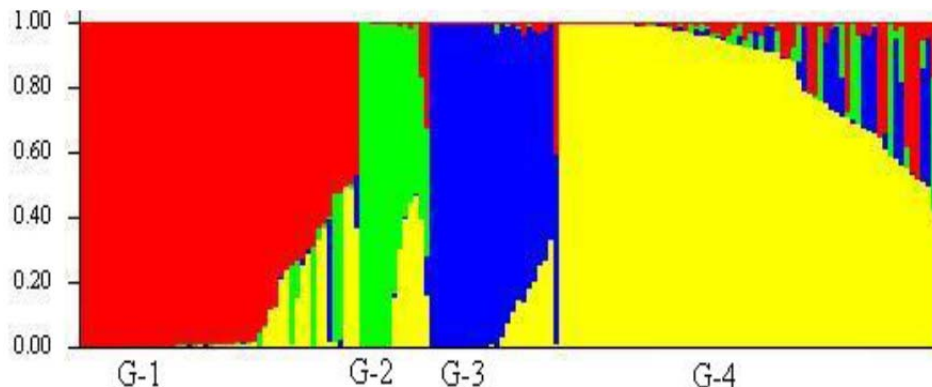


Figure 3. Population structure in the studied entries. The subpopulations obtained with K= 4 are represented by different colors as indicated at the bottom (G-1=red, G-2=green, G-3=Blue, and G-4 = yellow).

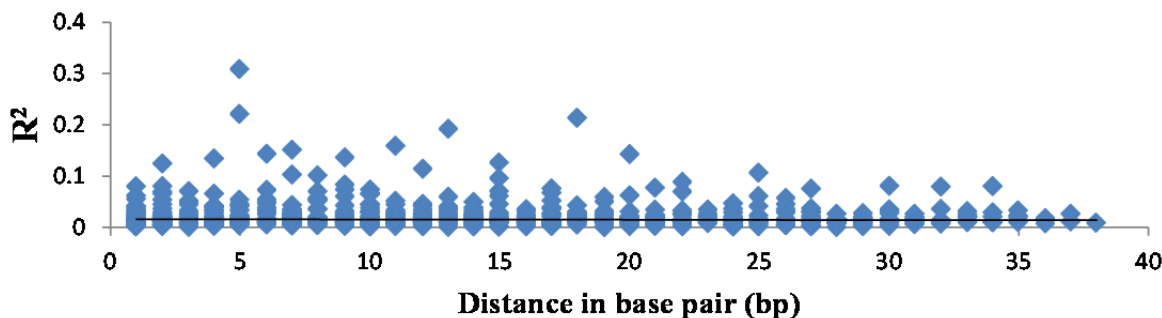


Figure 4. Linkage disequilibrium decay plot generated by 39 SSR markers.

Table 1. Associations between SSR markers and six agronomical traits.

Trait	Marker	Chr.	Position (Mb)	F-value	p-value	Marker R2
GWPP	gpsb123	C8	52.282	5.34565	0.0236	0.07626
PWT	gpsb123	C8	52.282	4.46473	0.03802	0.06574
PHI	gpsb123	C8	52.282	4.4633	0.03805	0.04638
PHT	mSbCIR223	C2	4.657	4.23342	0.04232	0.04077
PHT	mSbCIR248	C5	4.746	5.79222	0.01803	0.05389
THGT	Xcup53	C1	72.905	4.22234	0.04333	0.03798
THGT	Xtxp114	C3	60.794	4.72867	0.0313	0.02642
PHT	Xtxp145	C6	49.285	10.4089	0.0021	0.17757
50%FL	Xtxp145	C6	49.285	4.12649	0.04697	0.08347
50%FL	Xtxp278	C7	51.120	4.14561	0.04426	0.04183

GWPP=grain weight per panicle, PWT=panicle weight, PHI=panicle harvest index, PHT=plant height, THGT=thousands grain weight, 50%FL=Days to 50% flowering. Only SSR markers with a significant marker-trait associations are reported ($P < 0.05$). The P-value determines whether a QTL is associated with a marker, and the marker R2 evaluates the magnitude of the QTL effects (percentage of total variation explained by the marker).

(2010) using 26 sorghum inbred lines.

Association mapping

In this study, a total of 10 significant marker-trait

associations ($P \leq 0.05$) were detected (Table 1) and the phenotypic effect of SSR marker alleles on the associated characters were identified (Table 2). The 10 significant marker-trait associations were identified using seven different SSR markers for six agronomical characters (50%FI, PHT, PWT, GWPP, THGT and PHI),

Table 2. The phenotypic effect of marker alleles at loci associated with traits and the number of accessions carrying each marker allele in the studied sorghum accessions.

Character	Marker	Chr.	Pos(Mb)	Genotype (bp)	Effect	Observation
50%FL	Xtxp145	C6	49.285	214:214	-7.1919	46
50%FL	Xtxp145	C6	49.285	212:212	0	15
50%FL	Xtxp278	C7	51.12	248:248	4.92614	62
50%FL	Xtxp278	C7	51.12	242:248	0	48
GWPP	gpsb123	C8	52.282	290:290	-15.866	44
GWPP	gpsb123	C8	52.282	292:292	0	34
PHI	gpsb123	C8	52.282	290:290	-0.063	44
PHI	gpsb123	C8	52.282	292:292	0	34
PHT	mSbCIR223	C2	4.657	105:111	0.17516	40
PHT	mSbCIR223	C2	4.657	105:105	0	62
PHT	mSbCIR248	C5	4.746	91:91	-0.2505	54
PHT	mSbCIR248	C5	4.746	101:101	0	46
PHT	Xtxp145	C6	49.285	214:214	-0.3974	46
PHT	Xtxp145	C6	49.285	212:212	0	15
PWT	gpsb123	C8	52.282	290:290	-17.307	44
PWT	gpsb123	C8	52.282	292:292	0	34
THGT	Xcup53	C1	72.905	182:182	3.5836	33
THGT	Xcup53	C1	72.905	182:186	0	48
THGT	Xtxp114	C3	60.794	231:231	-2.3798	37
THGT	Xtxp114	C3	60.794	233:233	0	112

50%FL = Days to 50% flowering, PHT = plant height, PWT = panicle weight, GWPP = grain weight per panicle, PHI = panicle harvest index, and THGT = thousands grain weight.

with R^2 ranging from 2.6% (Xtxp114 with THGT) to 17.76% (Xtxp145 with PHT) given subsequently.

Days to 50% flowering

Two SSR markers (Xtxp145 & Xtxp278) having a significant association ($P \leq 0.05$) with days to 50% flowering were detected on chromosome 6 and 7, respectively. Xtxp145 had an effect of explaining 8.35% of the total phenotypic variation, whereas Xtxp278 had an effect of 4.18% of the total phenotypic variation. SSR markers linked to QTLs that control flowering time in sorghum were previously reported on chromosome six by Mannai et al. (2011) using association mapping and on chromosome 7 by Sirinivas et al. (2009) and Shiringani et al. (2010) using conventional QTL mapping.

Plant height

Three loci (mSbCIR223, mSbCIR248, Xtxp145) having a significant association ($P \leq 0.05$) with PHT were detected on chromosome 2, 5 and 6, respectively. Marker mSbCIR223 had an effect of 4.1% of the total phenotypic variation; mSbCIR248 had an effect of 5.39% of the total phenotypic variation, whereas Xtxp145 had an effect of

17.76% of the total phenotypic variation. Wang et al. (2012), using pool based genome wide association mapping, reported four SSR markers that were closely associated with PHT on chromosomes 2 and 6. Similarly, Sirinivas et al. (2009), using conventional method confirmed the presence of QTLs for PHT on chromosome 6 and 7 in sorghum.

Panicle weight, grain weight per panicle, and panicle harvest index

Locus gpsb123 showed simultaneous significant associations ($P \leq 0.05$) with three characters, namely PWT, GWPP, and PHI on chromosome 8. This locus had an effect of explaining 6.6, 7.6, and 4.64% of the total phenotypic variation for PWT, GWPP, and PHI, respectively.

Thousand grain weight

Two loci (Xcup53 and Xtxp114) on chromosomes 1 and 3, respectively, showed significant association ($P \leq 0.05$) with THGT. Xcup53 had an effect of 3.8% of the total phenotypic variation, whereas Xtxp114 had an effect of explaining 2.64% of the total phenotypic variation. A QTL

controlling seed weight was previously reported on chromosome one by Sirinivas et al. (2009) using conventional QTL mapping.

The present study also shows that in each of the identified SSR marker loci there were two genotypes having variant alleles in the studied accessions (Table 2). Most of the two genotypes at each locus had different magnitudes on the expression of the phenotype. As shown in Table 2, for example, for Xcup53 on chromosome 1, there were two genotypes (182:182 and 182:186) which were significantly associated with THGT. The presence of allele 182 in its homozygous state (182:182) increased the weight of thousand seeds by 3.5836 g compared to its heterozygous state (182:186). Similarly, for Xtxp114 on chromosome 3 which also linked to THGT, there were two genotypes (231:231 and 233:233). For this trait (THGT), the difference between the two genotypes (231:231 and 233:233) was 2.37984. In the same way, Xtxp145 on chromosome 6 which associated simultaneously with two traits: 50%FL and PHT, had two homozygous genotypes (214:214 and 212:212). The presence of marker allele 214 in its homozygous form decreased the days to 50% FL in 46 lines by 7.19 and the PHT by 0.39742 compared to its variant allele (212:212) in 15 accessions. Similarly, Xtxp278 on chromosome 7 which was found to be significantly linked to days to 50%FL had two genotypes (248:248 and 242:248) for the studied accessions. The presence of marker allele (248:248) increased the days to 50% FL by 4.92614 days compared to its variant allele (242:248). In the same way, gbsp123 on chromosome-8, which was found to be significantly associated with PWT and GWPP, had two genotypes (290:290 and 292:292). The presence of allele 290 in its homozygous form on this locus decreased PWT and GWPP by 17.307 and 15.866, respectively in 44 accessions compared to its homozygous variant allele (292) in 34 accessions for both traits. The difference of the effect on the phenotype between the two genotypes of mSbCIR223 on chromosome 2, mSbCIR248 on chromosome 5, and Xtxp145 on chromosome 6 which were found to be linked to PHT and between the two genotypes of gbsp123 on chromosome 8 to PHI, was negligible. Actually, it is not the variant marker allele itself which causes the decrease or increase in the expression of the phenotype in any trait of interest. Rather, it is imagined that there is a causative gene that is tightly linked to the variant marker allele which is responsible to cause a decrease or an increase in the expression of the phenotype. Thus, by following the variant marker allele that is tightly linked to the causative gene, it is possible to follow the effect of the causative gene on the phenotype of the lines under study. Generally, most of the variant alleles on the identified SSR marker loci had differences in magnitude of their effects on the phenotype of the trait under study. Some of the variant alleles had an increasing effect on the expression of the phenotype while others had a reducing

effect, this phenomenon might have useful application in molecular breeding. For example, if an interest arises to develop a variety of a grain cereal having a relatively short stature and earliness in flowering, a statistically linked marker allele with a reducing effect on plant height and days to 50% flowering shall be the target of the breeder. On the other hand, if the interest is to develop a variety for green forage having a relatively taller plant height and late flowering with high biomass accumulation, the marker allele with an increasing effect on plant height and days to 50% flowering shall be the target allele as plants with taller plant height and late flowering tend to accumulate high biomass for the purpose of green forage than with short stature and earliness in flowering (Habyarimana et al., 2004).

In sorghum, several reports have been published using conventional QTL mapping and some of them are in agreement with the present study. For example, in the study of Srinivas et al. (2009), one major QTL and two other QTLs were detected on chromosome 1 controlling seed weight which is similar to the present finding where locus Xcup53 on chromosome 1 is found to be associated with THGT. Among the nine significant QTLs associated with PHT in the study of Shiringani et al. (2010), one was found to be located on chromosome 2 which corresponds to the present result where locus mSbCIR223 on chromosome 2 found to be associated with PHT. Among the five QTLs detected with 50%FL in the study of Shiringani et al. (2010), two of them were located on chromosome 6 and 7. Similar results were obtained from the present study where markers Xtxp145 on chromosome 6 and Xtxp278 on chromosome 7, both significantly associated with 50% FL. Thus, some of the QTLs detected in previous reports were also detected in the present study. However, there are also some discrepancies between the findings of the present study and the previous QTL mapping studies. For example, in Srinivas et al. (2009), QTLs controlling for PWT and GWPP were detected on chromosome 6 where as in the present study PWT, GWPP and PHI were associated with locus gbsp123 on chromosome 8. Moreover, in both Srinivas et al. (2009) and Shiringani et al. (2010), several QTLs simultaneously localized on more than three chromosomes were detected for each trait whereas in the present study only PHT, 50%FL, and THGT were associated with more than one chromosome simultaneously. This discordance suggests two possible reasons; one reason is that this study may not have detected all the existing major QTLs because of the small number of markers used. Another cause is that a major QTL detected by a bi parental cross QTL mapping may not have large effect in the phenotypic variation of a germplasm collection and may be difficult to be detected with association mapping approach (Shahzad et al., 2009). Beyond the causes of discrepancy issues, the above presented cases noticed useful insight in the application of plant breeding. The significant association

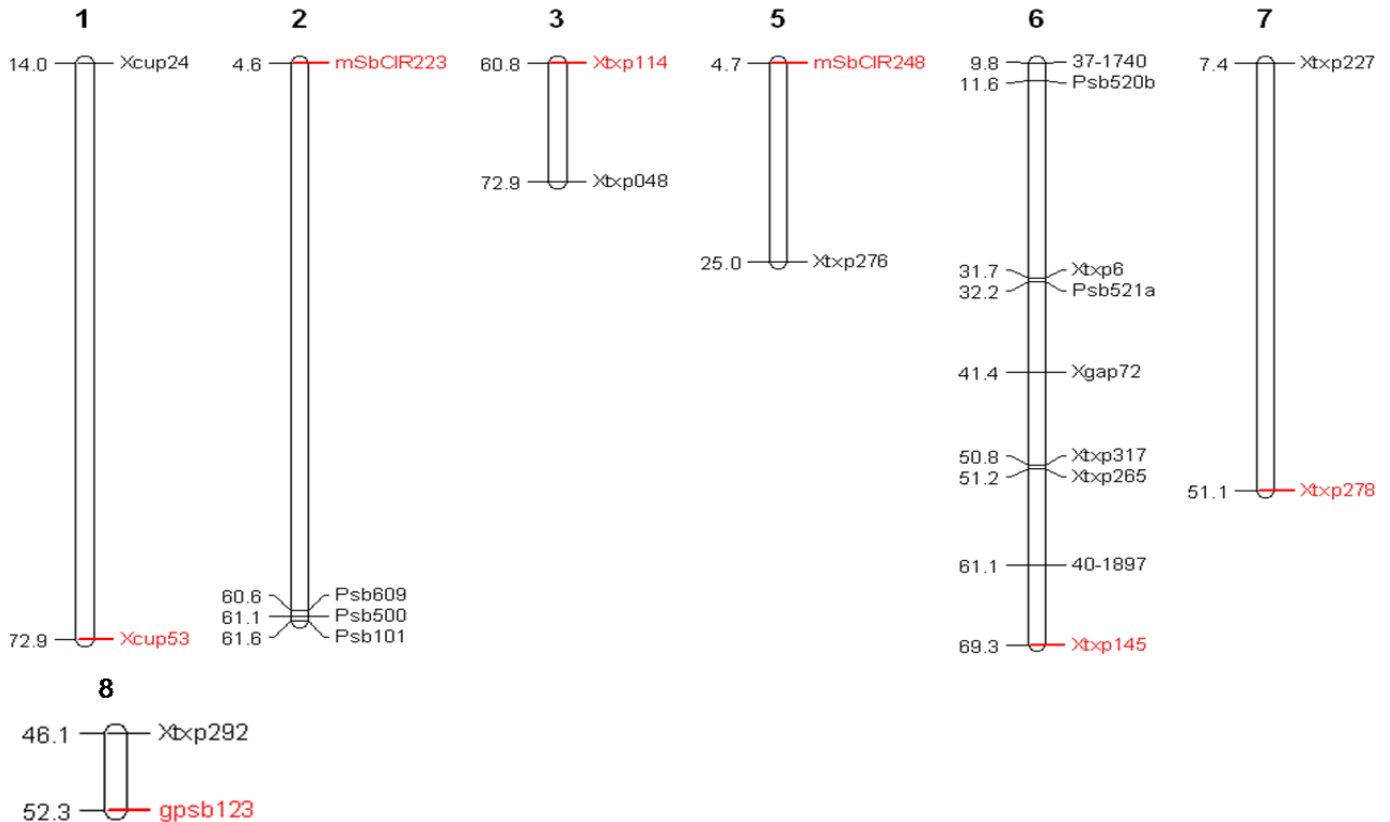


Figure 5. Chromosomal regions of marker trait associations. The linked markers highlighted in red are the present findings where as those in black are previously identified ones. Physical positions in base pairs (Mbp) are indicated on the left of the map and the corresponding marker names are indicated on the right. The sources for marker positions are Mace et al. (2009) and Wang et al. (2012).

of *gsb123* with PWT, GWPP and PHI indicated that this locus on chromosome8 simultaneously influenced the expression of the three traits notifying the presence of pleiotropic effect (*Xtxp145* on chromosome6 showed the same pleiotropic effect on 50%FL and PHT). This implies that in variety development improving for one trait helps for the improvement of the other which simplifies fixation in breeding materials. On the other hand, the expression of 50%FL, PHT, and TGHT were controlled by QTLs at different chromosomes indicating the presence of epistatic effects. A direct implication of epistasis is that the effects of the single-locus QTLs are mostly dependent on the genotypes of other loci, and the effect of a QTL can sometimes be negated by the genotypes of a second locus; thus an attempt for utilization of these QTLs in the breeding programs has to take in to account for such epistatic effects (Xing et al., 2002). To sum up, despite some discrepancies, many of the loci that were identified in this study were associated with traits common with previous studies indicating the creditability of the results, though relatively small number of markers used. Also, though the percentages of the variance explained by the associated markers are seemingly low ranging from 2.6 to 17.76% (Table 1), these estimates are lower than the

real QTL effects because in association mapping approach incomplete LD between marker and QTL leads to an underestimation of the variance explained by the QTL (Wurschum, 2012). Comparable results between biparental mapping population QTL analysis and association mapping could be observed when LD is perfect ($r^2=1$) and the same alleles segregate in both populations (Myles et al., 2009). Even if LD was perfect, underestimation of the phenotypic variance could arise from allelic frequency differential in the association mapping population (Stich et al., 2008). The maximum proportion of the variance explained by a marker is observed for allele frequencies of 0.5, as expected in biparental populations such as recombinant inbred lines or F_1 -derived doubled haploids. For a germplasm collection, the allele frequencies are expected to be considerably different from 0.5, especially when multi-allelic markers such as SSRs are used (Stich et al., 2008). Thus, the proportion of the variance explained by a marker is notably lower in association mapping approach despite the same underlying allelic effect (Stich et al., 2008). Taking this into consideration, the associated markers in this study can facilitate marker assisted selection and gene introgression to develop desirable cultivars in

sorghum and the study can serve as an initial effort in Ethiopia to select and map desirable genotypes or alleles using association mapping approach. However, validation of the associated markers by increasing the marker density and evaluating the phenotypes in representative environments will improve the variance explained by the associated markers and provide a more accurate estimation of the impact that the favorable alleles will have in a breeding program. In the present study, the identified seven SSR markers (Xcup53, mSbCIR223, Xtxp114, mSbCIR248, Xtxp145, Xtxp278, and gbsp123) were localized (Voorrips, 2002) on chromosomes 1, 2, 3, 5, 6, 7 and 8, respectively, harboring one marker each (Figure 5).

Conclusion

This study was conducted to identify drought tolerant sorghum genotypes, map chromosomal regions (QTLs) associated with agronomically important traits under moisture stress and identify SSR markers tightly linked with these QTLs. One hundred sixty (160) sorghum genotypes were evaluated in the field at Kobo in the off-season using irrigation in an alpha lattice design replicated three times. The phenotypic data were collected including days to 50% flowering, plant height, panicle weight, grain weight, grain weight per panicle, panicle harvest index, one thousand grain weight and number of grains per panicle. Analysis of variance showed highly significant ($P < 0.0001$) differences among the genotypes for all characters. Most of the characters showed moderate to high phenotypic and genotypic coefficient of variation. Heritability was high for all of the studied characters. Linkage disequilibrium (LD) analysis showed that in all accessions, 107 locus pairs (32.92%) had a significant ($p < 0.05$) mean LD of 0.19, with an $R^2 > 0.2$ for 33 evaluated locus pairs. Population structure analysis showed four distinct clusters in the studied materials. A total of 10 marker-trait associations were identified using 7 different SSR markers with R^2 ranging from 2.64 to 17.76. The seven SSR markers were localized on chromosomes 1, 2, 3, 5, 6, 7 and 8 harboring one marker each (xcup53, bSbCIR223, Xtxp114, mSbCIR248, Xtxp145, Xtxp278, and gbsp123, respectively).

Most of the identified markers were localized in chromosomal positions that have been previously reported as positions for drought tolerance-related traits, supporting the present findings. Hence, the associated SSR markers are potential candidates for marker assisted selection to improve drought tolerance in sorghum. Therefore, based on further validation in independent or related populations and in different environments, the markers that showed association with traits can be used to select genotypes with desirable features for a trait and land races which were found to be superior in their performance can be used for developing new varieties.

Conflict of interests

The authors have not declared conflict of interests.

REFERENCES

- Alemu T (2009). Analysis of genetic diversity among Ethiopian sorghum [*Sorghum bicolor* (L.) Moench] germ plasm collections. An MSc. Thesis presented to the school of graduate studies of Hawassa University, 98 p.
- Amsalu A, Endeshaw B, Bryngelsson T (2000). Genetic variation of Ethiopian and Eritrean sorghum (*Sorghum bicolor* (L.) Moench) germplasm assessed by random amplified polymorphic DNA (RAPD). *Genet. Resour. Crop Evol.* 47:471-482.
- Dicko MH, Gruppen H, Traore AS, Alphons GJ, Voragen AGJ, Van Berkel WJH (2006). Sorghum grain as human food in Africa: Relevance of content of starch and amylase activities. *Afr. J. Biotechnol.* 5:384-395.
- Doggett H (1976). *Sorghum bicolor* (Gramineae-Andropogoneae). In: N.W. Simmonds (Ed.). *Evolution of crop plants*. Longman and New York.
- Ejeta G, Knoll JE (2007). Marker-assisted selection in sorghum pp. In: R. Vashney and R. Tuberosa (eds) *Genomic Assisted Crop Improvement: Vol 2, Genomic Applications in Crops*. pp.187-205.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software Structure: A simulation study. *Mol. Ecol.* 14:2611-2620.
- FAOSTAT-Food and Agriculture Organization (United Nations) Statistics Division (2010). accessed 07 October 2010.
- Habyarimana E, Laureti D, Deninno M, Lorenzoni C (2004). Performance of biomass sorghum [*sorghum bicolor* (L.) Moench] under different water regimes in mediterranean region. *Ind. Crops Prod.* 20:23-28.
- Hao ZX, Li C, Xie J, Weng M, Li D, Zhang X, Liang L, Liu S, Zhang S (2011). Identification of Functional Genetic Variations Underlying Drought Tolerance in Maize Using SNP Markers. *J. Integr. Plant Biol.* 53(8):641-652.
- Hausmann BIG., Hess DE, Seetharama N, Welz HG, Geiger HH (2002). Construction Of A Combined Sorghum Linkage Map From Two Recombinant Inbred Populations Using Aflp, Ssr, Rflp, And Rapd Markers, And Comparison With Other Sorghum Maps. *Theor. Appl. Genet.* 105:629-637.
- IBPGR/ICRISAT (International Plant Genetic Resources Institute/International Board for Plant Genetic Resources (1993). *Descriptors for Sorghum* [*Sorghum bicolor* (L.) Moench]. IBPGR/ICRISAT. Rome, Italy.
- Li Y, Bhosale S, Hausmann BI, Stich B, Melchinger AE, Parzies HK (2010). Genetic diversity and linkage disequilibrium of two homologous genes to maize D8: *Sorghum* SbD8 and pearl millet PgD8. *J. Plant Breed. Crop Sci.* 2(5):117-128.
- Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am. J. Bot.* 82:1420-1425.
- Mace E, Rami J, Bouchet S, Klein P, Klein R, Kilian A, Wenzl P, Xia L, Halloran K, Jordan R (2009). A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DART) markers. *BMC Plant Biol.* 9(13):1-14.
- Mace ES, Buhariwalla HK, Crouch JH (2003). A high-throughput DNA extraction protocol for tropical molecular breeding programs. *Plant Mol. Biol. Rep.* 21:459a-459h.
- Mannai YE, Shehzad T, Okuno K (2011). Variation in flowering time in sorghum core collection and mapping of QTLs controlling flowering time by association analysis. *Genet Resour Crop Evol.* 58:983-989.
- Myles, Brown JPJ, Ersoz ES, Zhiwu Zhang, Costich E, Buckler ES (2009). Association Mapping: Critical Considerations Shift from Genotyping to Experimental Design. *Plant Cell* 21:2194-2202.
- Prichard JK, Stephens M, Rosenberg NA, Donnelly P (2000). Inference of population Structure using multilocus genotype data. *Genetics* 155:945-959.

- Sanchez AC, Subudhi PK, Rosenow DT, Nguyen HT (2002). Mapping Qtl's Associated With Drought Resistance In Sorghum (*Sorghum Bicolor* (L.) Moench). *Plant Mol. Biol.* 48:713-726.
- SAS Institute Inc. (2008). Statistical analysis software version 9.2, Cary NC: SAS institute Inc. USA.
- Shehzad T, Iwata H, 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.
- Shi A, Chen P, Zhang B, Hou A (2010). Genetic diversity and association analysis of protein and oil content in food-grade soybeans from Asia and the United States. *Plant Breed.* 129(3):250-256.
- Shiringani AL, Frisch M, Friedt W (2010). Genetic mapping of QTLs for sugar-related traits in a RIL population of *Sorghum bicolor* L. Moench. *Theor. Appl. Genet.* 121:323-336.
- Sorkheh K, Malysheva-Otto LV, Wirthensohn MG, Tarkesh-Esfahani S, Martínez-Gómez P (2008). Linkage disequilibrium, genetic association mapping and gene localization in crop plants. *Genet. Mol. Biol.* 131:805-814.
- Srinivas G, Satish K, Madhusudhana K, Seetharama N (2009). Exploration and mapping of microsatellite markers from subtracted drought stress ESTs in *Sorghum bicolor* (L.) Moench. *Theor. Appl. Genet.* 118:703-717.
- Stich B, Piepho HP, Schulz B, Melchinger AE (2008) Multi-traits association mapping in sugar beet (*Beta vulgaris* L.). *Theor. Appl. Genet.* 117:947-954.
- Tadesse T, Tesso T, Ejeta G (2008). Combining ability of introduced sorghum parental lines for major morpho-agronomic traits. *J. SAT Agric. Res.* 6:1-7.
- Tuinstra MR, Ejeta G, Goldsbrough P (1998). Cell biology & molecular genetics: Evaluation of near-isogenic sorghum lines contrasting for qtl markers Associated with drought tolerance. *Crop Sci.* 38:835-842.
- Tuinstra MR, Grote EM, Goldsbrough PB, Ejeta G (1996). Identification of quantitative trait loci associated with pre-flowering drought tolerance in sorghum. *Crop Sci.* 36:1337-1344.
- Upadhyaya HD, Wang Y, Sharma S, Singh S (2012). Association mapping of height and maturity across five environments using the sorghum mini core collection. *Genome* 55(6):471-481.
- Voorrips RE (2002). Map Chart: Software for the graphical presentation of linkage maps and QTLs. *J. Heredity* 93(1):77-78.
- Wang Y, Loganantharaj R, Bible P, Upadhyaya HD (2012). Identification of SSR markers associated with height using pool-based genome-wide association mapping in sorghum. *Mol. Breed.* 30(1):281-292.
- Wurschum T (2012) Mapping QTL for agronomic traits in breeding populations. *Theor. Appl. Genet.* 125:201-210.
- Xing YZ, Tan YF, Hua JP, Sun XL, Xu CG, Zhang Q (2002). Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. *Theor. Appl. Genet.* 105:248-257.
- Xu W, Subudhi PK, Crasta OR, Rosenow DT, Mullet JE, Nguyen HT (2000). Molecular Mapping of QTLs Conferring Stay-Green in Grain Sorghum (*Sorghum Bicolor* L. Moench). *Genome* 43:461-469.
- Yonemaru J, Ando T, Mizuybayashi T, Kasuga S, Matsumoto T, Yano M (2009). Development of genome-wide simple sequence repeat markers using whole-genome shotgun sequences of sorghum (*Sorghum bicolor* (L.) Moench). *DNA Res.* 16:187-193.

Supplementary Table 1. Regions from which the accessions were collected.

Region	Afar (1)	Amhara (2)	Eritrea(3)	Gambella(4)	NA(5)	Oromia(6)	RV (7)	SNNS(8)	Tigray (9)	Total
No. of acc.	2	47	4	13	6	41	8	18	21	160

Numbers in brackets are code numbers used to represent the names of regions. RV = Released varieties, NA = Information not available, and No. of acc. = number of accessions.

Supplementary Table 2. Mean monthly temperature and rain fall data at the experimental site during the cropping period (January-June 2010/2011).

Month	Temperature (°C)		Rain fall (mm)	
	Minimum	Maximum	Minimum	Maximum
December	5.5	20.55	-	-
January	6.75	21.35	-	-
February	6.0	22.75	0.85	2.3
March	8.75	23.85	1.15	3.15
April	9.5	24.5	1.5	31.05
May	10.5	25.15	2.5	5.5
June	12.5	27.05	1.15	2.5
Mean	8.5	23.6	1.43	8.9