Cereal Research Communications 41(3), pp. 359–365 (2013)
DOI: 10.1556/CRC.2013.0018
First published online 26 March 2013

Genetic Structure Analysis of Sorghum Parent Lines Based on SSR Markers

L.M. WANG[1,2]*, S.J. JIAO[2], Y.X. JIANG[2], H.D. YAN[2], D.F. SU[2], G.Q. SUN[2], X.F. YAN[1] and L.F. SUN[2]

[1]Northeast Forestry University Postdoctoral Programme, Heilongjiang Academy of Agricultural Sciences Postdoctoral Programme, Harbin 150086, P.R. China
[2]Crop Breeding Institute of Heilongjiang Academy of Agricultural Sciences, Harbin 150086, P.R. China

(Received 13 July 2012; accepted 26 October 2012;
Communicated by A. Anioł)

Genetic structure of 142 parent lines of sorghum [*Sorghum bicolor* (L.) Moench] was analyzed using model-based approach based on SSR markers. Forty-one selected from 103 SSR markers were used to analyze the parent lines, which generated 189 alleles revealed by each marker ranging from 2 to 11 with an average of 4.6 per marker. The polymorphic information content (PIC) value was 0.543 with a range of 0.089 to 0.850. All the parent lines were assigned to 7 subgroups, named Kafir, Kaoliang, Feterita, Shallu, Hegari, Milo and Durra. Parent lines without clear pedigree record were clustered into their corresponding groups, and genetic components of each line were estimated by Q-values. Information of this study would be useful for breeders to conclude their genetic background and select appropriate parents for germplasm improvement and hybrid breeding, and thus improve the efficiency of breeding programs.

Keywords: sorghum, parent lines, genetic structure, SSR markers, linkage disequilibrium

Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is one of the most important cereal crops, and grown for the production of grain, fodder, silage, syrup and other products. Knowledge of population structure and molecular relatedness among sorghum parent lines permits the classification of parent lines into heterotic groups, which is particularly important for hybrid breeding, germplasm improvement and association studies. The classification of parent lines into heterotic groups can be based on geographical origin, agronomical traits, pedigree data or molecular marker data (Ali et al. 2008). In the past, cultivated sorghums were classified into different groups of Milo, Kafir, Hegari, Feterita, Durra, Shallu and Kaoliang based on agronomical traits, and many new varieties and hybrids have originated by hybridization among these groups (Poehlman 1986). However, morphological variation does not reliably reflect the real genetic variation because of genotype-environ-

* Corresponding author; E-mail: dawnw@126.com

ment interactions and the largely unknown genetic control of polygenically inherited morphological and agronomic traits (Shehzad et al. 2009). With the development of molecular techniques, SSR markers have been successfully used to analyze population structure on various crops including sorghum (Wang et al. 2009; Mutegi et al. 2011), and give good discrimination among closely related individuals even when only a few loci are employed. On the other hand, molecular markers do not require previous pedigree information which is valuable for crops but frequently lacking (Kong et al. 2000). In recent years, population structure analysis based on linkage disequilibrium (LD) has been used in various crops (Choukan et al. 2006; Barro-Kondombo et al. 2010; Yu et al. 2011; Simko et al. 2012). Mutegi et al. (2011) used SSR markers to analyze genetic structure and relationships within and between cultivated and wild sorghum in Kenya.

In this study, parent lines of sorghum have been used frequently for creating segregating populations and hybrid breeding, but the pedigree information for many of them is not available or poorly documented. No information is available on the population structure analysis in Chinese sorghum parent lines. Therefore, it is necessary to analyze genetic structure of these parent lines in order to provide useful information to increase the efficiency in breeding program.

Materials and Methods

Plant materials

One hundred and forty-two parent lines of sorghum were used in this study; from China (55), America (24), India (15), Mexico (17), Russia (16) and Ukraine (15) (Table S1*).

DNA isolation and PCR assay

DNA was extracted from leaf tips of three to five seedlings of each line using a CTAB method according to Doyle and Doyle (1990).

A 103 SSR primer pairs which covered all the ten linkage groups of sorghum genome were selected based on the published information (http://sorgblast3.tamu.edu/Sorghum Genome/Mapping). From which, 41 pairs of primers with high levels of polymorphism and stable amplification were selected to genotype the 142 lines (Table 1).

The PCR reactions were performed in 20 μL reaction volumes containing 2 μL genomic DNA, 1.5 μL Primer, 1.5 μL dNTP (10 mmol L^{-1}), 2 μL reaction Buffer (10×), 2 μL $MgCl_2$ (mmol L^{-1}), 0.5 μL *Taq* DNA polymerase (5 U μL^{-1}) and 10.5 μL dd H_2O. PCR reaction conditions consisted of 5 min at 94ºC for initial denaturation, followed by 35 cycles of amplification of 20 s at 94ºC, an annealing step of 30 s at the appropriate annealing temperatures, 40 s at 72ºC, and 10 min at 72ºC for final extension. PCR products were separated on 6% non-denaturing polyacrylamide gels in 1×TEB buffer, and then visualized with silver staining (Panaud et al. 1996).

* Further details about the Electronic Supplementary Material (ESM) can be found at the end of the article.

Table 1. Allele number, diversity index and PIC value of primers

SSR locus	Repeat motif	Linkage group	Number of alleles	Unbiased simpson index	PIC
Xtxp46	(GT)10	A	5	0.704	0.664
Xtxp329	(ATC)8+(CTT)22	Unknown	11	0.870	0.850
Xtxp324	(TA)18	Unknown	4	0.391	0.362
Xtxp58	(AG)13+(GA)16	A	4	0.706	0.648
Xtxp113	(GA)11	Unknown	7	0.825	0.794
Xtxp47	(GT)8(GC)5+(GT)6	H	6	0.726	0.678
Xtxp217	(GA)23	J	6	0.766	0.726
Xtxp67	(GA)28	I	6	0.582	0.554
Xtxp159	(CT)21	G	5	0.570	0.532
Xtxp83	(GT)9(AT)12	Unknown	4	0.425	0.376
Xtxp221	(TCC)6	Unknown	4	0.720	0.660
Xtxp208	(GGA)8	A	4	0.668	0.606
Xtxp350	(AAT)23	A	2	0.223	0.197
Xtxp258	(AAC)19	I	9	0.795	0.762
Xtxp61	(GA)13	A	7	0.595	0.565
Xtxp279	(CTT)10+(CTT)3+(CTT)6	A	4	0.507	0.444
Xtxp276	(CAA)9+(AAT)5	Unknown	2	0.168	0.153
Xtxp229	(GT)8	A	4	0.595	0.520
Xtxp328	(CT)6+(GGAC)6	Unknown	4	0.739	0.684
Xtxp296	(CA)18	B	3	0.569	0.498
Xtxp267	(TCT)20	Unknown	4	0.603	0.556
Xtxp212	(GT)10	D	3	0.439	0.396
Xtxp303	(GT)13	E	7	0.744	0.708
Xtxp57	(GT)21	F	4	0.481	0.442
Xtxp197	(AC)10	B	4	0.725	0.668
Xtxp82	(GA)12(GC)6	Unknown	3	0.651	0.573
Xtxp336	(CGG)4+(GAG)6	C	4	0.487	0.401
Xtxp338	(GAA)11	Unknown	4	0.534	0.494
Xtxp211	(CT)23	B	6	0.529	0.492
Xtxp250	(AAG)17AAT(AAG)4AAA(ACA)9	H	5	0.643	0.597
Xtxp93	(AC)9	Unknown	2	0.254	0.220
Xtxp98	(CTGT)5	Unknown	4	0.676	0.611
Xtxp201	(GA)36	B	6	0.809	0.774
Xtxp207	(CT)14	B	4	0.656	0.591
Xtxp228	(TC)12	C	4	0.719	0.660
Xtxp141	(GA)23	J	6	0.816	0.783
Xtxp91	(AG)19	Unknown	6	0.795	0.758
Xtxp205	(AG)12	C	3	0.540	0.468
Xtxp131	(CT)7	Unknown	3	0.503	0.382
Xtxp179	(GT)8AC(GT)5	B	2	0.094	0.089
Xtxp265	(GAA)19	F	4	0.197	0.329

Data analysis

Diversity index of primers was analyzed by using Genetics Statistics 3.0 (Chen 2005). Polymorphism information content (PIC) was calculated by using the formula developed by Anderson et al. (1993).

The model-based software STRUCTURE version 2.2 (Falush et al. 2003) was used to determine population structure and to assign parent lines to subgroups. The structure was run ten times by setting predefined K (the number of group in a population) from 1 to 10 using admixture models (Evanno et al. 2005). The highest likelihood value was selected to assign the posterior membership coefficients (Q) to each line.

Results

SSR marker polymorphism

Forty-one markers which generated 189 alleles were used to analyze the genetic structure of 142 parent lines. The number of alleles revealed by each marker ranged from 2 to 11 with an average of 4.6 per marker. The average diversity index of primers was 0.586 with a range of 0.094-0.870. The polymorphic information content (PIC) value for the SSR loci ranged from 0.089 for Xtxp179 to 0.850 for Xtxp329 with an average of 0.543 (Table 1).

Cluster analysis

We tested possible genetic clusters in these parent lines using a model-based approach. After running K-value from 1 to 10, we observed the curve park at K = 7. This implied that the 7 subgroups were most reasonable to explain the 142 parent lines. The 7 sub-populations were designated with the names of typical lines in subgroups as Kafir, Kaoliang, Feterita, Shallu, Hegari, Milo and Durra (Table S1).

Table 2. Distribution of Q-value of parent lines in different groups

Group	Total number	Q < 0.6		Q ≥ 0.6		Q ≥ 0.8		Q ≥ 0.9	
		Number	Percentage (%)	Number	Percentage (%)	Number	Percentage (%)	Number	Percentage (%)
Kafir	20	9	45.0	11	55.0	9	45.0	6	30.0
Kaoliang	20	6	30.0	14	70.0	6	30.0	4	20.0
Feterita	18	7	38.9	11	61.1	8	44.4	8	44.4
Shallu	25	11	44.0	14	56.0	10	40.0	6	24.0
Hegari	19	6	31.6	13	68.4	8	42.1	5	26.3
Milo	18	6	33.3	12	66.7	5	27.8	0	0
Durra	22	6	27.3	16	72.7	9	40.9	2	9.1
Total	142	51	35.9	91	64.1	55	38.7	31	21.8

As shown in Table S1, Q-value denoted the genetic component of a parent line in the subgroups and molecular relatedness among different lines. The type with the largest Q-value indicated that the parent line was likely to be involved in this subgroup.

The 142 parent lines were categorized with the Q-value division of 0.6 (Table 2). The members and percentages of parent lines with high Q-values (0.8 or 0.9) were also shown in Table 2. Fifty-one lines had the maximum Q-value lower than 0.6, and only 38.7% or 21.8% lines showed Q-values higher than 0.8 or 0.9. This indicated that each subpopulation contained genetic components of other subpopulation, and revealed shared ancestry information among groups.

Discussion

Structure analysis of sorghum parent lines

Structure analysis based on LD can estimate the number of subpopulations and the genetic relatedness among assessed parent lines. In maize and sea buckthorn, inbred lines and cultivars with unknown pedigrees had also been clustered into known heterotic groups with considerable success, and the results based on molecular data were coincident with the estimation based on pedigrees (Choukan et al. 2006; Li et al. 2009). The results of this study demonstrated that genotyping sorghum parent lines with DNA markers could quickly reveal the genetic structure and genetic components of each line. Assessment of genetic background by molecular markers was an indispensable complementation to assessment by agronomical traits.

In this study, parent lines without clear pedigree record were clustered into their corresponding groups, and genetic components of each line were estimated by Q-values. Of all parent lines, 35.9% had maximum Q-values smaller than 0.6, which indicated that gene often exchange among lines during the procedure of introduction, crossing and backcrossing, thus resulted in complex genetic background. Similar results were found by Choukan et al. (2006) and Wu et al. (2010). Information of this study would be useful for breeders to conclude their genetic background and select appropriate parents for germplasm improvement and hybrid breeding, and thus improve the efficiency of breeding programs.

This research would also benefit for further studies to continue this molecular classification for additional lines to identify new sources of alleles for sorghum improvement.

Structure analysis used in sorghum breeding

Previous studies have used molecular markers to assign temperate line into known heterotic groups with considerable success. Maize is probably the most successful crop to assign inbred lines to heterotic groups by molecular markers, thus potentially increasing the efficiency of breeding programs (Reif et al. 2003; Choukan et al. 2006). Division of heterotic group is also fundamentally important in sorghum improvement. However, in China, no established heterotic groups of sorghum exist currently, and most of the parent lines have unknown pedigrees. Therefore, classification information presented in this study could be used as a basis for describing heterotic groups in sorghum.

In the past, breeders relied on qualitative pedigree information to estimate genetic contribution of parent lines. Our population structure analysis will allow breeders to quantify these relationships in the form of the Q matrix, and make more informed decisions about

potential crosses between these parent lines. To produce higher degree of heterosis, parent lines should be crossed rather with external groups than with the lines of the same group. Information generated from this study can be used to help breeders select parent lines in different groups to create higher degree of heterosis and promote further breeding progress.

Further research is needed to verify heterotic performance and combining abilities between the groups detected in this study, and facilitate the development of well-defined heterotic groups in sorghum.

Acknowledgements

This research was financially supported by the earmarked fund of China Agriculture Research System (CARS-06-01-04) and International Science & Technology Cooperation Program of China (2011DFR30840).

References

Ali, M.L., Rajewski, J.F., Baenziger, P.S., Gill, K.S., Eskridge, K.M., Dweikat, I. 2008. Assessment of genetic diversity and relationship among a collection of US sweet sorghum germplasm by SSR markers. Mol. Breed. **21:**497–509.

Anderson, J.A., Churchill, G.A., Autrique, J.E., Sorrellls, M.E., Tanksley, S.D. 1993. Optimizing parental selection for genetic linkage maps. Genome **36:**181–186.

Barro-Kondombo, C., Sagnard, F., Chantereau, J., Deu, M., vom Brocke, K., Durand, P., Gozé, E., Zongo, J.D. 2010. Genetic structure among sorghum landraces as revealed by morphological variation and microsatellite markers in three agroclimatic regions of Burkina Faso. Theor. Appl. Genet. **120:**1511–1523.

Chen, Q.S. 2005. Comparative analysis of specialty index and genetic similarity on soybean germplasm with resistance to Cercospora sojina Hara. China Biotech. **25**(suppl.):155–158. (in Chinese)

Choukan, R., Hossainzadeh, A., Ghannadha, M.R., Warburton, M.L., Talei, A.R., Mohammadi, S.A. 2006. Use of SSR data to determine relationships and potential heterotic groupings within medium to late maturing Iranian maize inbred lines. Field Crops Res. **95:**212–222.

Doyle, J.F., Doyle, J.L. 1990. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Focus **12:**13–15.

Evanno, G., Reganut, S., Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecology **14:**2611–2620.

Falush, D., Stephens, M., Pritchard, J.K. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics **164:**1567–1587.

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

Li, H., Ruan, C.J., da Silva, J.A.T. 2009. Identification and genetic relationship based on ISSR analysis in a germplasm collection of sea buckthorn (*Hippophae* L.) from China and other countries. Scientia Hortic. **123**:263–271.

Mutegi, E., Sagnard, F., Semagn, K., Deu, M., Muraya, M., Kanyenji, B., de Villiers, S., Kiambi, D., Herselman, L., Labuschagne, M. 2011. Genetic structure and relationships within and between cultivated and wild sorghum (*Sorghum bicolor* (L.) Moench) in Kenya as revealed by microsatellite markers. Theor. Appl. Genet. **122:**989–1004.

Panaud, O., Chen, X., McCouch, S. R. 1996. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol. General Genet. **252:**597–607.

Poehlman, J.M. 1986. Breeding Field Crops. Van Nostrand Reinhold Company, New York, USA, pp. 522–527.

Reif, J.C., Melchinger, A.E., Xia, X.C., Warburton, M.L., Hoisington, D.A., Vasal, S.K., Beck, D., Bohn, M., Frisch, M. 2003. Use of SSRs for establishing heterotic groups in subtropical maize. Theor. Appl. Genet. **107**:947–057.

Shehzad, T., Okuizumi, H., Kawase, M., Okuno, K. 2009. Development of SSR-based sorghum (*Sorghum bicolor* (L.) Moench) diversity research set of germplasm and its evaluation by morphological traits. Genet. Res. Crop Evol. **56:**809–827.

Simko, I., Eujayl, I., van Hintum, T.J.L. 2012. Empirical evaluation of DArT, SNP, and SSR marker-systems for genotyping, clustering, and assigning sugar beet hybrid varieties into populations. Plant Sci. **184:**54–62.

Wang, M.L., Zhu, C., Barkey, N.A., Chen, Z., Erpelding, J.E., Murray, S.C., Tuinstra, M.R., Tesso, T., Pederson, G.A., Yu, J. 2009. Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. Theor. Appl. Genet. **120:**13–23.

Wu, C.L., Zhang, Q.Q., Dong, B.X., Li, S.F., Zhang, C.Q. 2010. Analysis of genetic structure and relationships of maize inbred lines in China. Acta Agron. Sin. **36:**1820–1831.

Yu, X., Bai, G., Luo, N., Chen, Z., Liu, S., Liu, J., Warnke, S. E., Jiang, Y. 2011. Association of simple sequence repeat (SSR) markers with submergence tolerance in diverse populations of perennial ryegrass. Plant Sci. **180:**391–398.

Electronic Supplementary Material (ESM)

Electronic Supplementary Material (ESM) associated with this article can be found at the website of CRC at http://www.akademiai.com/content/120427/

Electronic Supplementary *Table S1.* Parent lines of sorghum