GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Genetic dissection of upland cotton (*Gossypium hirsutum*) cultivars developed in Hubei Province by mapped SSRs

J.L. Tu[1]*, M.J. Zhang[2]*, X.Q. Wang[1], X.L. Zhang[1] and Z.X. Lin[1]

[1]National Key Laboratory of Crop Genetic Improvement,
National Centre of Plant Gene Research, Huazhong Agricultural University,
Wuhan, Hubei, China
[2]Department of Bioscience and Engineering, Huanggang Normal College,
Huanggang, China

*These authors contributed equally to this study.
Corresponding author: Z.X. Lin
E-mail: linzhongxu@mail.hzau.edu.cn

Genet. Mol. Res. 13 (1): 782-790 (2014)
Received January 9, 2013
Accepted May 24, 2013
Published January 31, 2014
DOI http://dx.doi.org/10.4238/2014.January.31.4

ABSTRACT. The genetic diversity of 51 upland cotton cultivars with different parental origins and breeding periods that were developed in Hubei Province was studied on the basis of 237 mapped simple sequence repeat markers covering the cotton genome. A total of 108 polymorphic primer pairs amplified 196 loci; the polymorphism information content range was 0.04 to 0.83, with an average of 0.46. A model-based clustering analysis (STRUCTURE) of the genomic data identified 3 clear subpopulations, and the result was confirmed by principal components analysis. The genetic similarity coefficient among 51 upland cotton cultivars was 0.598 on average, ranging from 0.378 to 0.817. The unweighted pair group method with arithmetic average cluster analysis revealed inconsistencies in other clustering patterns: "Tianmian1" was distinct from the rest of the materials and formed a separate cluster. This study will provide a guide for breeders to develop

new cultivars efficiently and to choose parents, and it supports the need to introduce new alleles into the gene pool of the upland cotton breeding program in Hubei Province.

Key words: Upland cotton; Simple sequence repeats; Genetic diversity

INTRODUCTION

Cotton belongs to Dicotyledoneae, produces fiber from seed, and occupies a very important status in China's textile industry. Hubei Province, in the middle reaches of the Yangtze River, is one of the important cotton bases of China and makes a significant contribution to the textile industry. The cotton cultivars of Hubei Province mainly experienced 3 periods: varieties introduction, systematic selection, and cross breeding. Before the 1960s, upland cotton varieties that were introduced from America were mainly cultivated, especially Deltapine 15. In the 1960s and 1970s, Eguangmian was mainly cultivated by Chinese cotton breeders. Subsequently, cotton cultivars with high-yield production and hybrid varieties were introduced. In the 21st century, high-yield, disease-resistant, and insect-resistant hybrid cultivars were mainly cultivated.

Molecular markers that detect the organ's genetic diversity effectively should have the following characteristics: rich polymorphism, stability, and high heredity. DNA molecular markers were the most effective method of genetic analysis at the DNA level. In cotton, various molecular markers have been applied in genetic diversity analysis such as simple sequence repeats (SSRs), rapid amplification of polymorphic DNA (RAPD), amplified fragment length polymorphism, inter-SSR, and sequence-related amplified polymorphism; the cotton species included *Gossypium hirsutum* (Xu et al., 2001; Rana et al., 2005; Zhang et al., 2005; Wu et al., 2010; Kalivas et al., 2011), *G. barbadense* (Westengen et al., 2005; El-Zanaty et al., 2011; Wang et al., 2011), and *G. arboretum* (Liu et al., 2006; Kantartzi et al., 2009).

An understanding of the magnitude and patterns of genetic diversity in upland cotton has important implications for breeding programs and for the conservation of genetic resources. Lu and Myers (2002) used RAPD data to evaluate genetic relationships and discriminate 10 influential upland cotton varieties (*G. hirsutum*). Chen and Du (2006) described the genetic diversity of 43 sources of upland cotton in China on the basis of SSR markers, and the results showed that the early maturity cotton accessions in the north were higher than those from the Huanghe and Yangtze growing areas; also, the molecular marker genetic similarity index of the domestic varieties was higher than that of the imported varieties. Kalivas et al. (2011) used SSR markers for the identification and the phylogenetic analysis of 29 cultivars of *G. hirsutum*. The polymorphism information content (PIC) ranged from 0 to 0.548 with a mean of 0.293; the unweighted pair group method with arithmetic average (UPGMA) analysis showed that the 30 genotypes formed 3 main groups. Zhang et al. (2011) estimated the genetic diversity of 59 core cotton cultivars from China's main cotton-growing areas based on genomic and newly developed expressed sequence tag (EST)-SSR markers.

In this study, 51 upland cotton cultivars that were developed from different periods and sources, according to the development of cotton breeding history of Hubei Province, were selected to analyze and evaluate their genetic relationship on the basis of mapped SSR markers. Combining 3 diversity analysis methods, 51 upland cotton cultivars could be deeply and objectively investigated.

MATERIAL AND METHODS

Plant materials and DNA extraction

Fifty-one upland cotton cultivars that were developed in Hubei Province were used in this study (Table 1). Cotton genomic DNA was extracted from young leaves with the modified cetyltrimethylammonium bromide (CTAB) method (Paterson et al., 1993).

Table 1. Name, origin, and breeding period of upland cotton cultivars developed in Hubei Province.

Cultivar	Code	Female parent	Male parent	Developed year
Deltapine15	HB01			1947
Yapengmian	HB02	Deltapine15 mutant		1956
Dongting 1	HB03	Deltapine15 Breeding		1957
Eguangmian	HB04	Eguangmian		1964
Emian 1	HB05	Deltapine15 Breeding		1950s
Emian 2	HB06	Deltapine15 x Yapengmian		1959
Emian 3	HB07	Deltapine15 Breeding		1950s
Emian 4	HB08	Deltapine15 Breeding		1962
Emian 5	HB09	Deltapine15 Breeding		1960s
Emian 6	HB10	Deltapine15 mutant		1966
Emian 9	HB11	Deltapine15 mutant		1964
Emian 10	HB12	Deltapine15 Breeding		1968
Emian11	HB13	Dai3599 x Gaomi933		1980
Emian 12	HB14	[(5006 x 1080)(3762 x Mujin)]F_1	Henan79	1985
Emian 13	HB15	CRI10		1990
Emian 14	HB16	2930	CRI7	1988
Emian 15	HB17	60Co radiation		1988
Emian 16	HB18	Ejing1	Esha28	1990
Emian 17	HB19			1988
Emian 18	HB20	Wan73-10	Shahu49	1993
Emian 19	HB21	Eing1 x Zhongmianxin10		1991
Emian 20	HB22	Ejing1 x Xiangmian10		1992
Emian 21	HB23	(5512 x PD4548)F_1 x CRI12		1994
Emian 23	HB24	98-38 x 4643 x 3247		2003
Emian 24	HB25	Simian3	87-28	2004
Esha28	HB26	Jingmian x Gangmian1	Jingmian4	1981
Ejing1	HB27	(3208 x Jingmian4)F_2	AntongSP21	1985
Ejing 92	HB28	3208	Jingmian4	1983
Ekangmian1	HB29	Jing3247 x Jiangsu9118		1990
Ekangmian 2	HB30	Ejing1 x Zhong2535		1998
Ekangmian 3	HB31	Eguangmian x Jingzhou3247	Zhong2535	1995
Ekangmian 4	HB32	Ejing1		1996
Ekangmian 5	HB33	Emian12 x PD0113		1992
Ekangmian 6	HB34	Ejing1	Xiangmian10	1997
Ekangmian 7	HB35	Jing7701	Jing1418	1997
Ekangmian 8	HB36	(3247 x Esha28)F_1 x CCRI12		1998
Ekangmian 9	HB37	Ejing1	Zhong7263+M03	1999
Ekangmian 10	HB38	(Jing3117 x Zhong12)F_1 x Jing3187		1999
Tianmian1	HB39	Humian204		1969
Shanong6	HB40	Eguangmian		1965
Deltapine16	HB41	Eguangmian	Deltapine45	1970
Gangmian1	HB42	Eguangmian		1972
Gangmian25	HB43	Gangmian1		1978
Ekangchong1	HB44	Simian3 import GFMcryIABt		2002
Yishuhong	HB45	Tuozimian		1953
Bayimian	HB46	Shuangtaomian x Eguangmian		1970
Xiaomian1	HB47	Eguangmian		
Jing1246	HB48	Ejing1	Xiangmian10	1997
Jingchu201	HB49	Ekangmian10	3259	2005
Huakangmian1	HB50	168	Sumian3	2002
Emian 22	HB51	Ejing1 x PD2165		1994

SSR marker analysis

SSR markers were the 237 pairs of SSRs that were selected from 26 chromosomes and were included 4 to 17 markers per chromosome (Wang et al., 2011). Polymerase chain reaction (PCR), electrophoresis and silver staining were performed as previously described by Wang et al. (2011).

Data analysis

After observing the PCR electrophoresis results, the bands of DNA fragments were scored "1" for presence, "0" for absence, and "9" for missing data. Genetic diversity analysis and further study were made on the basis of these scores. The statistical methods and formulae that were used are shown below.

1) PIC = 1 - $\Sigma P i^2$ (i = 1 - K), where Pi represents the variation in frequency of the ith allele, and K represents the sum allele of one marker.

2) The STRUCTURE version 2.1 (Pritchard et al., 2000) software was used to analyze the population structure of these accessions. Markov Chain Monte Carlo was set to 50,000, burn-in was set to 100,000, and the K value was from 2 to 10 and each was repeated 3 times.

3) The genetic diversity statistics of upland cotton were analyzed using the frequency and analysis of molecular variance (AMOVA) functions of the GenAlEx6.2 software (Cruz et al., 2007), including polymorphism percentage, number of effective alleles, Shannon's information index, genetic differentiation (PhiPT), gene flow, and expected heterozygosity.

4) AMOVA was used to estimate the molecular diversity between and within populations, and principal coordinates analysis (PCA) based on Nei's genetic distance was used to confirm the Structure results.

5) Genetic similarity coefficients (Jaccard's coefficients) among varieties were calculated using the Qualitative Date program of the NTSYSpc 2.1 software (Biostatistics Inc., New York, USA). The dendrograms were constructed by UPGMA.

RESULTS

Amplified efficiency of SSRs

Among the 237 SSR primer pairs, 108 primer pairs (45.6%) were polymorphic and generated 196 polymorphic loci with an average of 1.82 loci per primer pair. Among them, 51 pairs of primers amplified only 1 polymorphic locus, 33 amplified 2 loci, 17 amplified 3 loci, and 7 pairs amplified 4 polymorphic loci. The PIC value for the SSR loci ranged from 0.04 to 0.83 with a mean of 0.46. The SSR with the highest average PIC value was BNL3257. When taking chromosomes into account, the highest average PIC values were for Chr17 (0.71), Chr04 (0.69), and Chr24 (0.66) (Table 2).

Population structure analysis

STRUCTURE V2.31 was used to infer the population structure using the entire SSR marker dataset. With K running, the distribution of L (K) did not show a clear mode for the

true K (Figure 1A). Thus, another method was used to overcome the difficulty of interpreting the real K values (Evanno et al., 2005), $\Delta K = m (|L(K + 1) - 2 L(K) + L(K - 1)|) / s[L(K)]$, where ΔK was developed and showed a clear peak where the true value of K was present. In this study, the highest value of ΔK for the 51 cultivars was K = 3 (Figure 1B), so value of K = 3 was chosen for the analysis. As shown in Figure 2, the cultivars were clearly classified into 1 of the 3 subpopulations (P1-P3), which included 7, 29, and 15 cultivars, respectively (Table 3).

Table 2. Amplified efficiency of markers on different chromosomes.

Chromosome	Number of markers	Amplified loci	PIC value	Mean PIC value
Chr1	6	4	0.04-0.66	0.35
Chr2	5	4	0.04-0.72	0.38
Chr3	7	5	0.40-0.72	0.61
Chr4	5	8	0.61-0.78	0.69
Chr5	11	4	0.04-0.44	0.14
Chr6	8	4	0.36-0.72	0.54
Chr7	4	1	0.29	0.29
Chr8	12	12	0.08-0.83	0.45
Chr9	11	12	0.26-0.82	0.63
Chr10	9	7	0.04-0.71	0.43
Chr11	17	10	0.04-0.82	0.27
Chr12	14	14	0.04-0.79	0.50
Chr13	11	13	0.08-0.8	0.51
Chr14	6	7	0.04-0.75	0.38
Chr15	11	6	0.11-0.72	0.46
Chr16	6	7	0.04-0.71	0.50
Chr17	5	3	0.70-0.72	0.71
Chr18	7	10	0.04-0.82	0.51
Chr19	13	8	0.11-0.75	0.45
Chr20	13	9	0.11-0.74	0.49
Chr21	6	10	0.08-0.80	0.61
Chr23	9	11	0.08-0.80	0.53
Chr24	11	5	0.53-0.78	0.66
Chr25	17	10	0.04-0.61	0.31
Chr26	13	12	0.14-0.78	0.50

PIC = polymorphism information content.

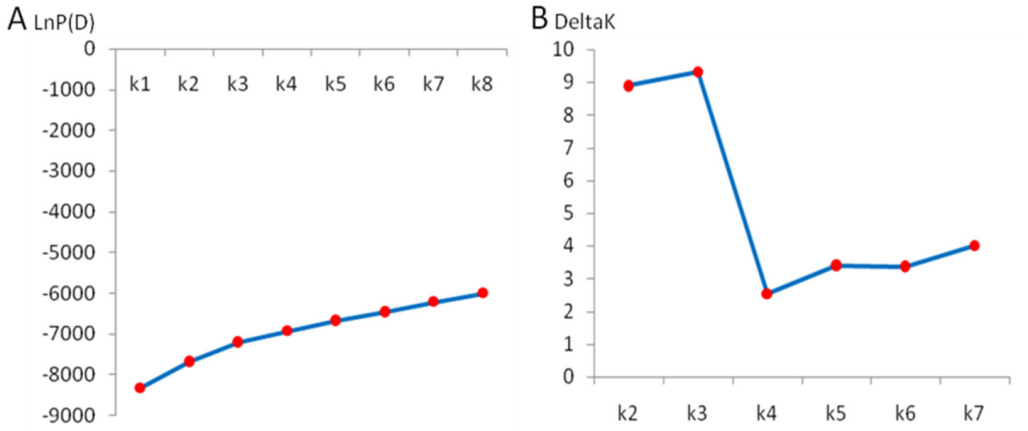

Figure 1. Two different methods of determining the value of K. **A.** The *ad hoc* procedure described by Pritchard et al. (2000). **B.** The second order statistics (ΔK) developed by Evanno et al. (2005).

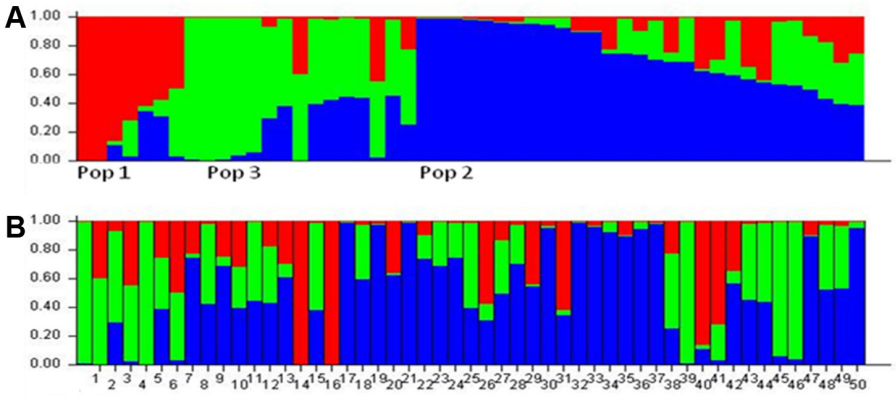

Figure 2. A. Model-based ancestry for each of the 51 upland cotton cultivars based on SSR markers used to build the Q matrix. Red represents Pop 1 and blue represents Pop 2 while green is representative of Pop 3. **B.** Population structure of 51 upland cotton cultivars; the height of vertical axis means that the material is divided into a special cluster where the similarity of each line is more than 50%.

Table 3. Cultivars' distribution and names of STRUCTURE-based populations.

Group	Cultivar No.
Pop 1 (7)	HB07 HB15 HB17 HB27 HB32 HB41 HB42
Pop 2 (29)	HB06 HB08 HB10 HB11 HB13 HB14 HB18 HB19 HB20 HB21 HB22 HB23 HB24 HB25 HB28 HB29 HB30 HB31 HB33 HB34 HB35 HB36 HB37 HB38 HB43 HB48 HB49 HB50 HB51
Pop 3 (15)	HB01 HB02 HB03 HB04 HB05 HB09 HB12 HB16 HB26 HB39 HB40 HB44 HB45 HB46 HB47

PCA and AMOVA results

PCA was conducted to further assess the population subdivisions that were identified by STRUCTURE. In the PCA graph (Figure 3), principal coordinates 1, 2, and 3 explained 27.52, 19.12, and 14.97% of variation, respectively. The PCA results absolutely confirmed the results that were obtained from STRUCTURE, and genetic diversity statistics showed that the order of all statistical values was P2 > P3 > P1. Moreover, the 3 populations had different statistical values in pairwise comparisons (Table 4).

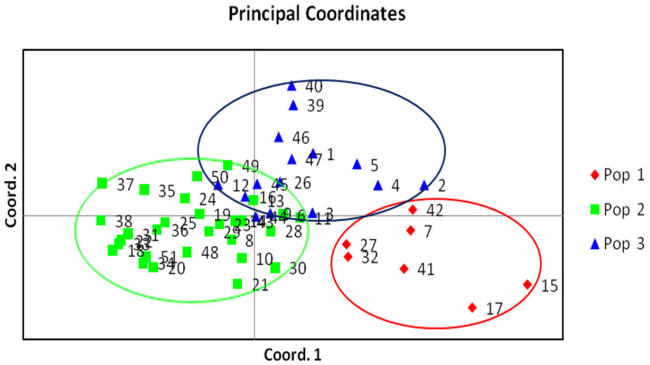

Figure 3. PCA graph of 51 upland cotton cultivars developed in Hubei Province.

Table 4. Comparison of genetic diversity among subgroups.

Population	N	N_A	N_E	I	H_E	UH_E	Polymorphism (%)
Pop 1	7	1.347	1.317	0.285	0.188	0.203	57.28
Pop 2	29	1.751	1.383	0.351	0.229	0.233	84.51
Pop 3	15	1.512	1.362	0.325	0.214	0.221	71.36
Total	51	1.537	1.354	0.321	0.210	0.219	71.05

N_A = number of alleles; N_E = effective number of alleles; I = Shannon's information index; H_E = expected heterozygosity; UH_E = unbiased H_E.

According to the Structure result, the 51 cultivars were divided into 3 subgroups. AMOVA showed that the between-population component of genetic variance was 15%, in contrast to 85% for the within-population component (Table 5). At $P < 0.01$, the general PhiPT was 0.146.

Table 5. Analysis of molecular variance.

Source	d.f.	SS	MS	Estimated variation	%
Among population	2	172.36	86.18	4.22	15%
Within population	48	1184.11	24.67	24.67	85%
Total	50	1356.47		28.89	100%

d.f. = degrees of freedom; SS = sum of squares; MS = mean squares.

UPGMA analysis

Genetic similarity coefficient matrices were calculated by the NTSYS-pc 2.1 software package (Rohlf, 2000) based on Jaccard's coefficient. The resulting similarity coefficient was used to construct a genetic distance matrix using the SAHN procedure based on UPGMA (Figure 4). The genetic distance between these selected cultivars ranged from 0.378 to 0.817 with an average of 0.598, and the highest value that were observed was between HB01 and HB47, which were both introduced from America in the 1960s. According to the UPGMA dendrogram, HB39 was distinct from the rest of materials and formed a separate cluster. The rest of the cultivars could be separated into 2 clusters. HB02, HB04, HB15, HB17, HB27, HB41, and HB42 were grouped together to form the second cluster, and the others formed the third cluster.

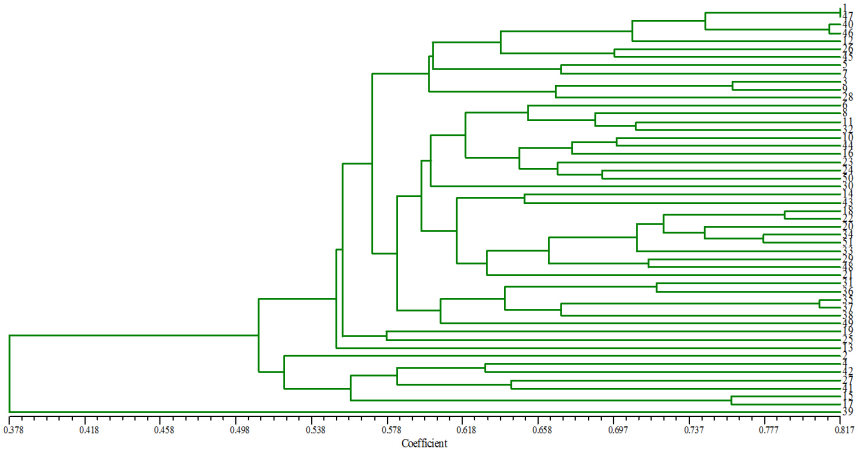

Figure 4. UPGMA cluster analysis of 51 upland cotton cultivars based on SSR markers.

DISCUSSION

To evaluate the overall genetic variation among different cotton populations, the genetic diversity of 51 upland cotton cultivars that were developed in different periods in Hubei Province were analyzed. The range of PIC calculated in this study was 0.04-0.83 with an average of 0.46. The average genetic similarity coefficient of 51 cultivars was 0.598 and ranged from 0.378 to 0.817. This result indicates that the genetic variation among those cultivars is very low, which is consistent with previous studies (Wendel et al., 1992; Iqbal et al., 1997, 2001; Xu et al., 2001). These indicate that the breeding of upland cotton faces a bottleneck effect. Therefore, introduction of new cultivars and new breeding methods is urgent and necessary. The comparison of the genetic diversity of 3 subpopulations derived from the analysis of population structure implicated that the genetic variation among these 3 subpopulations was very small.

AMOVA showed that the between-population component of genetic variance was 15%, and the within-population component was 85%. That is, the variation among different cotton sources mainly contributed to the overall variation. At $P < 0.01$, the general PhiPT was 0.145, which indicated that the molecular variation was mostly restricted to the subpopulations.

Identifying a proper K value to divide the population structure is a prerequisite to assess the genetic variation and structure at the molecular level using the computational program STRUCTURE (Zhao et al. 2010). It was very difficult to define the K value based on the LnP (D) value with slow increasing tendency. Therefore, the peak value could be easily observed when the K value was set to 3 based on the ΔK variation that was calculated using the reported method (Evanno et al., 2005). Besides Structure, UPGMA and PCA were used to divide the population structure (Shim and Jørgensen, 2000; Li et al., 2010; Sakiroglu et al., 2010). In our study, subpopulations were categorized by 3 different methods. According to our results, Structure and PCA analysis were highly consistent. Comparatively, the results consistently showed that HB15, HB17, HB27, HB41, and HB42 were grouped together by Structure, PCA, and UPGMA; on the other hand, the results of UPGMA showed some difference from the other 2 methods because the different methods analyzed genetic diversity according to different theories, resulting in different results among them.

In summary, systematic evaluation of the genetic diversity of upland cotton cultivars in Hubei Province has been carried out in this study. This study will provide very useful information for future cotton breeding in Hubei Province. This research also provides a very strong theoretical basis to obtain an excellent and scientifically established repository of the Hubei cotton varieties.

ACKNOWLEDGMENTS

Research supported by the National Basic Research Program (#2011CB109303) and the Fundamental Research Funds for the Central Universities (Program #2012MBDX007).

REFERENCES

Chen G and Du XM (2006). Genetic diversity of source germplasm of Upland cotton in China as determined by SSR marker analysis. *Yi. Chuan Xue. Bao.* 33: 733-745.

Cruz VMV, Luhman R, Marek LF, Rife CL, et al. (2007). Characterization of flowering time and SSR marker analysis of spring and winter type *Brassica napus* L. germplasm. *Euphytica* 153: 43-57.

El-Zanaty AM, Salem KFM and Esmail RM (2011). Detection of genetic diversity in Egyptian cotton (*Gossypium barbadense* L.) varieties using RAPD markers and morphological traits. *J. Am. Sci.* 7: 1107-1115.

Evanno G, Regnaut S and Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620.

Iqbal MJ, Aziz N, Saeed NA, Zafar Y, et al. (1997). Genetic diversity evaluation of some elite cotton varieties by RAPD analysis. *Theor. Appl. Genet.* 94: 139-144.

Iqbal MJ, Reddy OUK, El-Zik KM and Pepper AE (2001). A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. *Theor. Appl. Genet.* 103: 547-554.

Kalivas A, Xanthopoulos F, Kehagia O and Tsaftaris AS (2011). Agronomic characterization, genetic diversity and association analysis of cotton cultivars using simple sequence repeat molecular markers. *Genet. Mol. Res.* 10: 208-217.

Kantartzi SK, Ulloa M, Sacks E and Stewart JM (2009). Assessing genetic diversity in *Gossypium arboreum* L. cultivars using genomic and EST-derived microsatellites. *Genetica* 136: 141-147.

Li JQ, Schulz B and Stich B (2010). Population structure and genetic diversity in elite sugar beet germplasm investigated with SSR markers. *Euphytica* 175: 35-42.

Liu DQ, Guo XP, Lin ZX, Nie YC, et al. (2006). Genetic diversity of Asian cotton (*Gossypium arboreum* L.) in China evaluated by microsatellite analysis. *Genet. Resour. Crop Evol.* 53: 1145-1152.

Lu J and Myers O (2002). Genetic relationships and discrimination of ten influential Upland cotton varieties using RAPD markers. *Theor. Appl. Genet.* 105: 325-331.

Paterson AH, Brubaker CL and Wendel JF (1993). A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. *Plant Mol. Biol. Rep.* 11: 122-127.

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

Rana MK, Singh VP and Bhat KV (2005). Assessment of genetic diversity in upland cotton (*Gossypium hirsutum* L.) breeding lines by using amplified fragment length polymorphism (AFLP) markers and morphological characteristics. *Genet. Resour. Crop Evol.* 52: 989-997.

Rohlf FJ (2000). NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System, Version 2.1. Exeter Software, New York.

Sakiroglu M, Doyle JJ and Charles BE (2010). Inferring population structure and genetic diversity of broad range of wild diploid alfalfa (*Medicago sativa* L.) accessions using SSR markers. *Theor. Appl. Genet.* 121: 403-415.

Shim SI and Jørgensen RB (2000). Genetic structure in cultivated and wild carrots (*Daucus carota* L.) revealed by AFLP analysis. *Theor. Appl. Genet.* 101: 227-233.

Wang XQ, Feng CH, Lin ZX and Zhang XL (2011). Genetic diversity of sea-island cotton (*Gossypium barbadense*) revealed by mapped SSRs. *Genet. Mol. Res.* 10: 3620-3631.

Wendel JF, Brubaker CL and Percival AE (1992). Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. *Am. J. Bot.* 79: 1291-1310.

Westengen OT, Huaman Z and Heun M (2005). Genetic diversity and geographic pattern in early South American cotton domestication. *Theor. Appl. Genet.* 110: 392-402.

Wu DP, Fang XX, Ma MN, Chen JH, et al. (2010). Genetic relationship and diversity of the germplasms in *Gossypium barbadense* L. from four different countries using SSR markers. *Cotton Sci.* 22: 104-109.

Xu QH, Zhang XL and Feng CD (2001). Genetic diversity analysis on cultivars (*G. hirsutum* L.) developed by Hebei Province and CCRI by RAPD markers. *Acta Gossypii Sin.* 13: 238-242.

Zhang JF, Lu Y, Cantrell RG and Hughs E (2005). Molecular marker diversity and field performance in commercial cotton cultivars evaluated in the Southwestern USA. *Crop Sci.* 45: 1483-1490.

Zhang Y, Wang XF, Li ZK, Zhang GY, et al. (2011). Assessing genetic diversity of cotton cultivars using genomic and newly developed expressed sequence tag-derived microsatellite markers. *Genet. Mol. Res.* 10: 1462-1470.

Zhao WG, Chung JW, Cho YI, Rha WH, et al. (2010). Molecular genetic diversity and population structure in *Lycium* accessions using SSR markers. *C. R. Biol.* 333: 793-800.

