Australian Journal of Crop Science AJCS
AJCS 7(12):1912-1920 (2013) ISSN:1835-2707

Selection of core SSR markers for fingerprinting upland cotton cultivars and hybrids

Muhammad Mahmood Ahmed[1,2], Huanle Guo[1,2], Cong Huang[1], Xianlong Zhang[1] and Zhongxu Lin[1,*3]

[1]National Key Laboratory of Crop Genetic Improvement and National Centre of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, Hubei, PR China
[2]Co-authors equally contributed.

***Corresponding authors: linzhongxu@mail.hzau.edu.cn**

Abstract

Precise identification of cotton cultivars and hybrids is requisite to facilitate management of germplasm resources and successful hybridization programs. Fingerprinting based on minimal core set of highly informative primers will be more enlightening to unveil genetic constitution among cotton cultivars and hybrids from distinct growing regions of china. Thirty-eight upland cotton cultivars and 55 hybrids were selected from three cotton growing regions of china mainland i.e. yellow river cotton valley (YRCV), Yangtze river cotton valley (YzRCV) and north west dry region (NWDR); featured with perceptible climatology. Twelve randomly selected cultivar and hybrids (representative sample of the three regions) were employed to reveal polymorphism across mapped SSRs. Sixty-six genome-covered polymorphic SSRs were employed to assess genetic relatedness among all accessions. The results showed a polymorphism information content range from 0.34-0.86 and resolving power 0.04-2.45 for genome-covered SSRs. Higher PIC and Rp values rendered selection of 13 highly informative potential core SSRs, whose average PIC and Rp values were 0.80 and 1.64, respectively. Moreover, Jaccard's similarity coefficients of genome-covered and potential core SSRs were compared and found to be related. Potential core SSRs substantiated clustering results of genome-covered SSRs and successfully discriminated all accession. Overall clustering pattern and revealed genetic constitution of tested material suggested that the 13 potential core SSRs could assure comparable results with higher resolution as compared to that of genome-covered SSRs.

Keywords: Genetic diversity, molecular fingerprinting, SSR, upland cotton.
Abbreviations: CMD_cotton marker database, SSR_simple sequence repeats, PIC_polymorphism information content, Rp_ resolving power, YRCV_Yellow river cotton valley, YzRCV_Yangtze river cotton valley, NWDR_North west dry region, IPRs_ intellectual property rights.

Introduction

Cotton, the most important widely cultivated agricultural and industrial crop, is cultivated in varied climatic conditions on a large land area and occupies a pivotal rank in the world's economy. Moreover, it stands as the sixth largest source for vegetable oil production (Zhang et al., 2007). Among 50 species of genus *Gossypium* (Fryxell, 1992), two diploid species i.e. *G. arboreum* and *G. herbaceum* (2n = 2x = 26) and two tetraploid species i.e. *G. hirsutum* and *G. barbadense* (2n = 4x = 52) are widely cultivated and contribute to world's annual lint production (Wang et al., 2011), respectively.

China has been contributing a significant share in world's economy and unanimously has been the leading cotton producer for a decade. Land area suited for cotton cultivation in China could be categorized into distinct regions i.e. Yellow River Cotton Valley (YRCV), Yangtze River Cotton Valley (YzRCV) and North West Dry Region (NWDR), etc. Dissent geography and climatology has corroborated deviation in cotton yield, area under cultivation, cropping system, cropping intensity, selection of parents from germplasm for hybridization program, varieties under cultivation, etc. among the different regions. Ever increasing demand of high performance varieties and hybrids has confined cotton breeders to recurrently employ limited elite lines as parents, resulting in narrower genetic diversity in Chinese cultivars and hybrids of upland cotton. Employing broader genetic diversity is crucial for development of better-quality varieties and hybrids (Azmat and Khan, 2010) as it ensures species adaptability and survival. Worldwide and national economic importance of the cotton crop, availability of large number of cultivars and hybrids for cultivation, bottleneck effect and narrow genetic base have necessitated the development of an authentic and unquestionable method i.e. DNA fingerprinting to ameliorate cotton germplasm resources, distinguish among distant elite cultivars and hybrids and to protect plant varieties owners' intellectual property rights (IPRs) (Rojas et al., 2008).

In recent past, identification of cotton cultivars and hybrids has been executed based on morphological and phenological traits such as plant stature, number of sympodial branches, boll weight, flowering time, fiber length, fiber maturity, etc. Regarding morphological and phenological traits, it has always been befuddled to distinguish cultivars especially in the early stages of the cotton plant. Moreover, cultivar identification has been predicted to be complex in cotton because of polyploidy, simultaneous presence of lot of cultivars in a given time and area and narrow genetic base.
Successful breeding program and ameliorated management of genetic resources substantively depended on the ability to

distinguish cultivars, identify distant elite lines from germplasm resources and genetic diversity estimate (Owen et al., 2005). Moreover, precise identification of cotton cultivars and hybrids is requisite for effective selection management of genetic resources, varietal registration and to avoid adulteration of improved cotton cultivars and hybrids (Rahman and Rajora, 2002).

The exploitation of DNA polymorphism techniques rendered easy and quick approach for genetic diversity analysis and fingerprinting of cultivars and hybrids using molecular markers, especially at early stages. Several types of molecular markers such as RAPD (Asif et al., 2005; Hossain et al., 2002; Thomas et al., 2006), AFLP (Galovic et al., 2004; Sensi et al., 2003), ISSR (Pharmawati et al., 2005; Zhao et al., 2006) and SSR (Ercisli et al., 2011; Singh et al., 2007) have been employed by researchers in DNA fingerprinting and genetic diversity studies in plants. However, SSR markers have been preferred over others and applied to peach (Cheng and Huang, 2009), apricot (Zhang et al., 2012), rice (Tu et al., 2007), sugarcane (Hameed et al., 2012; Pandey et al., 2011), *Brassica napus* (Qu et al., 2012), jute, olive (Ercisli et al., 2011; Poljuha et al., 2008) and cotton (Chen and Du, 2006; Kalivas et al., 2011) for genetic diversity and fingerprinting studies. SSR markers are reproducible, co-dominant, abundant, robust, highly polymorphic, covering entire genome and analytically simple. At present, 17448 SSRs have been developed in cotton and are being added on daily basis to cotton marker database (CMD website, http://www.cottonmarker.org) to target specific goals by cotton researchers.

In this study, the genetic diversity of cotton cultivars and hybrids from distinct cotton growing regions of china is revealed by employing SSR markers covering the whole genome. Then the core set of SSR primers are identified to alleviate discrimination, registration and avoid adulteration of improved cotton cultivars and hybrids.

Results

SSR polymorphism and resolving power

Sixty-six highly polymorphic primers generated 162 loci, ranging from 1 to 5 with an average of 2.45 fragments per primer and were designated as genome-covered SSRs. Polymorphism information content (PIC) varied from 0.86 (BNL3319) to 0.34 (BNL4047) with 0.68 as average. Resolving Power (Rp) of a primer is based on the ability of a primer to produce more variable banding patterns. Rp value ranged from 0.04 (NAU3943) to 2.45 (BNL3319) with an average of 0.98. However, higher average values were observed for number of amplified loci per primer N (3.61), PIC (0.8) and Rp (1.64) based on thirteen potential core SSRs (Table S1).

Estimation of genetic diversity

To estimate genetic divergence, 93 accessions were categorized in two ways such as type of accession, either cultivar/hybrid or origin which is one of the three major cotton growing regions in china mainland. All populations were subjected to genome-covered SSRs and potential core SSRs for estimation of genetic diversity parameters. Genetic divergence estimates of cultivars were found at par cultivars were found at par for both sets of SSRs, but hybrids displayed slightly higher diversity for potential core SSRs. However, cultivars exhibited higher percent of polymorphic loci for both sets of primers (Table 1). Similarly, 93 accessions were classified on the basis of origin. Accessions from Henan, Hebei and Shandong provinces were collectively designated as YRCV; Hubei, Hunan, Anhui and Jiangsu provinces as YzRCV and all accessions from Xinjiang were designated as NWDR. Genetic diversity estimates for genome-covered SSRs and potential core SSRs were found in following order; NWDR > YzRCV > YRCV and YzRCV > NWDR > YRCV, respectively (Table 2).

Utilization efficiency of SSR fingerprints

Encoding binary digit format for genotyping sequence format and implementing multiple sequence alignment tools exploited the utility of potential core SSRs to fingerprint cotton cultivar and hybrids. Determination of utilization efficiency of potential core SSRs was eased by pair wise comparison. The utilization efficiency of fingerprinting primers is presented in Table 3. Despite of displaying the highest Rp and PIC value, BNL3319 was able to produce two unique banding patterns which implied to massive genome of upland cotton, which is one of the limitations associated with its characterization. However, a greater number of variable banding patterns of BNL3319 were reckoned to have high PIC and Rp values, which contributed most when affixed to other loci and capable to successfully designate a unique sequence for all cultivar and hybrids. The addition of locus BNL1317 fingerprinted conjointly 15 accessions. Genotyping pattern of another locus BNL3255 conjoined and it enabled detection of 8 more unique banding patterns. Similarly, fingerprints of locus BNL3405 added which resulted in discrimination of 29 unique discriminating patterns and so on. SSR fingerprints of more than half accessions were distinguished using a combination of polymorphic alleles of conjoined eight loci only. At long last, 93 unique alleles were fingerprinted by appending 47 polymorphic alleles of 13 loci.

Genetic relatedness and cluster analysis of upland cotton cultivars and hybrids revealed by genome-covered SSRs

Band scores of genome-covered SSRs were compiled to build binary dataset and used to perform cluster analysis on cultivars and hybrids separately to better reveal genetic constitution, relatedness and comprehend relationships among accessions of the three regions. Cluster analysis was also employed to potential core SSRs to correspond to two sets of primers in characterizing and fingerprinting upland cotton cultivars and hybrids from three generative regions of china. The coefficient among 38 cultivars ranged from 0.30 to 0.93. Five pairs of cultivars exhibited similarity coefficients equal to or more than 0.90. However, the highly similar cultivars (0.93), JM38 and RH39, originated from Henan and Shandong provinces. The least similar cultivars were RH39 and ZMS49 (0.30) originated from Shandong and Henan as well.

All 38 cultivars could be divided into three main clusters (Fig. 1). The cluster A consisted of 7 cultivars, 3 cultivars (XLZH28, XLZ42 & XLZH47) from NWDR and 2 cultivars (LMY21 and XQ1) from YRCV. Two only cultivars of region YzRCV, GKM1 and SK1, appeared to be somehow similar with remaining 5 cultivars in this cluster. Presence of

Table 1. Genetic diversity estimates of two populations.

Population	Size	Genome-covered SSRs					Potential core SSRs				
		Polymorphic loci (%)	Na	Ne	I	He	Polymorphic loci (%)	Na	Ne	I	He
Cultivars	38	96.39%	1.95	1.46	0.43	0.29	96.88%	1.98	1.46	0.42	0.27
Hybrids	55	87.95%	1.81	1.47	0.41	0.27	93.75%	1.83	1.51	0.45	0.30
All	93	92.17%	1.88	1.47	0.42	0.28	95.31%	1.90	1.48	0.44	0.29

Na: number of different alleles, Ne: number of effective alleles, I: shannon's information index and He: expected heterozygosity.

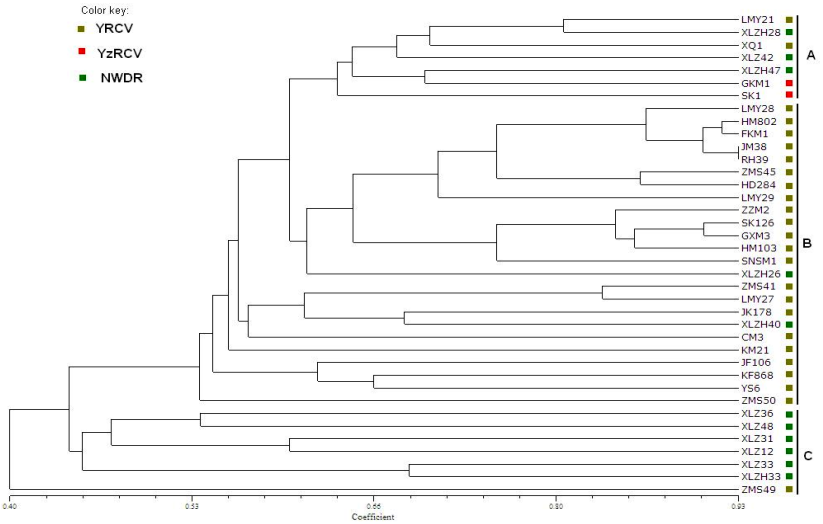

Fig 1. Cluster analysis of 38 upland cotton varieties originated from three growing regions (presented in color key) of China using 66 genome-covered SSRs analysis data from a Jaccard's similarity matrix and UPGMA method.

Table 2. Genetic diversity estimates of three populations.

Population	Size	Genome-covered SSRs					Potential core SSRs				
		Polymorphic loci (%)	Na	Ne	I	He	Polymorphic loci (%)	Na	Ne	I	He
YRCV	41	87.95%	1.810	1.420	0.390	0.260	96.88%	1.872	1.411	0.398	0.255
NWDR	11	84.34%	1.750	1.490	0.430	0.290	90.63%	1.723	1.475	0.414	0.276
YzRCV	41	84.34%	1.760	1.470	0.410	0.270	87.50%	1.766	1.525	0.453	0.305
All	93	85.54%	1.770	1.460	0.410	0.270	91.67%	1.787	1.470	0.422	0.279

YRCV: yellow river cotton valley, NWDR: North West dry region, YzRV: Yangtze river cotton valley, Na: number of different alleles, Ne: number of effective alleles, I: shannon's information index and He: expected heterozygosity.

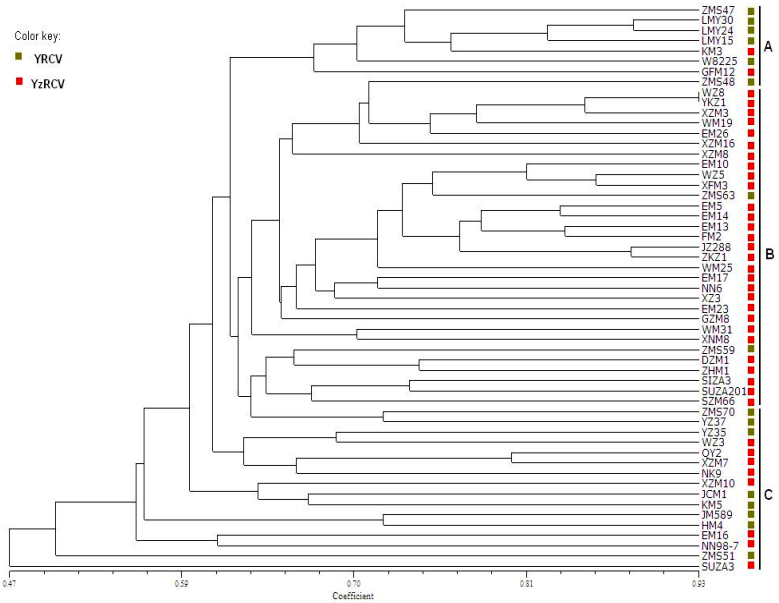

Fig 2. Cluster analysis of 55 upland cotton hybrids originated from two growing regions (presented in color key) of China using 66 genome-covered SSRs analysis data from a Jaccard's similarity matrix and UPGMA method.

cultivars from three regions imparted high variability to cluster A. The central and the biggest cluster, cluster B, consisted of 24 cultivars and completely occupied by cultivars from YRCV. Four most comparable cultivars i.e. HM802, JM38, RH39 and FKM1 with high similarity coefficient belonged to this cluster. It characterized enough variability to let in the two distinct cultivars from NWDR; XLZH26 and XLZH40. However, the rest of the cultivars belonged to different provinces of same region (YRCV). These findings substantiated higher reliability and dependability for investigating genetic relatedness among tested materials using potential core SSRs. The cluster C was occupied completely with cultivars from NWDR by placing more than half cultivars from NWDR together in a group. Bottom of the cluster was positioned by one of the most diverse cultivar ZMS49 from Henan (YRCV).

Similarity coefficients among 55 hybrids ranged from 0.38 to 0.93. The maximum value (0.93) between YKZ1 and WZ8 of region YzRCV suggested that both may share a common parentage (Fig. 2). Similarly, minimum value (0.38) between SUZA3 (Jiangsu) and WZ25 (Anhui) advocated existence of high level of divergence at DNA level. Overall there was slightly narrower genetic base found in hybrids than cultivars from different regions, and the results were in accordance with previous finding (Chen and Du, 2006).

All 55 hybrids could be categorized into three main clusters. Cluster A consisted of eight hybrids and predominately occupied by hybrids of Henan and Shandong provinces (YCRV). However, two hybrids KM3 (Jiangsu) and GFM12 (Anhui) originated from YzRCV also succeeded to get a place in Cluster A. The two most similar hybrids WZ8 and YKZ1 (0.93) from the same region (YzRCV) were placed on top of cluster B. The largest cluster B contained thirty-one hybrids and was almost completely occupied by hybrids of region YzRCV. However, two hybrids of Henan province (ZMS63 and ZMS59) were distantly placed in the mid of the cluster. The cluster C was characterized by a blend of hybrids from both regions. The most diverse hybrid SUZA3 (Jiangsu) occupied bottom position in the cluster. It was interestingly noted that hybrids of the two regions were piled up in small patches of 3-4 hybrids in the cluster (Fig. 2).

Genetic relatedness and cluster analysis of upland cotton cultivars and hybrids revealed by potential core SSRs

Similarity coefficient of 38 cultivars using 13 highly informative SSRs varied from 0.22 to 0.95. Potential core SSRs successfully discerned and placed together the four most similar cultivars (JM38, RH39, HM802 and FKM1) originated from YRCV, as recognized by genome-covered SSRs. The similarity coefficient value between RH39 and ZMS49 (0.32) was at par with that of genome-covered SSRs (0.30), but it was not the lowest value. Cultivars JM38 and XLZ33 were designated to constitute the most diverse genetic base and belonged to different regions (YRCV and NWDR). Thus, potential core SSRs not only substantiated similarity coefficients estimates of genome-covered SSRs, but also exploited benevolent potential of 13 informative SSRs to explicate true figure of genetic constitution among 38 cultivars. Overall, slightly greater variability of tested materials revealed because of higher resolution capacity of potential core SSRs.

Based on UPGMA dendrograms (Fig. 3), cluster A comprised of six cultivars belonged to different regions. It was interestingly noted that potential core SSRs validated the placing of three cultivars i.e. XLZH47, XLZH28 and LMY21 in the cluster. However, their higher resolution capacity did not let the cultivars GKM1 and SK1 (YzRCV) to be resided in cluster A and were placed in cluster B. Moreover, potential core SSRs successfully stacked 18 cultivars, originated from YRCV, in the cluster B. Three mostly similar cultivars i.e. HM802, RH39 and JM38 resided in cluster B, as same as clustered by genome-covered SSRs. Similarly, potential core SSRs successfully inhabited most cultivars of NWDR region to pile on each other and grouped with eight cultivars of the region in cluster C, whereas only six were heaped up by genome-covered SSRs. Potential core SSRs replaced ZMS49 (Henan) with CM3 (Hebei), originated from same region, at bottom position of the cluster. However, these validated the presence of ZMS49 based on genome-covered SSRs in the cluster.

Similarity coefficients of 55 commercial hybrids (0.27 to 0.94) were more varying as compared to those of genome-covered SSRs. For genome-covered SSRs, two highly similar hybrids (YKZ1 and WZ8) belonged to different provinces. However, the two most similar hybrids (LMY30 and LMY 24) were originated from Shandong province (YRCV), based on potential core SSRs. Thus, the fact logically asserted high resolving capability of potential core SSRs. Similarly, five pair of hybrids exhibited lowest similarity coefficient value (0.27), each belonged to different provinces. Whereas, two most diverse hybrids (ZMS63 and ZMS59) revealed by genome-covered SSRs shared common origin (Henan).

Four hybrids (LMY15, LMY24, LMY30 and W8225) were piled very close to each other by two sets of SSRs but in different clusters, all originated from Shandong (YRCV). However, hybrids belong to other provinces of YRCV region piled in small groups and invaded different clusters for both sets of SSRs. The cluster B exhibited relatively higher variability among the hybrids of both regions. Upper half of the cluster predominately occupied by hybrids of YzRCV, while bottom of the cluster by that of YRCV. LMY 30 and LMY 24 resided together at the bottom of the cluster with other hybrids of the region. Hybrids originated from YRCV and YzRCV coalesced to constitute cluster C which was characterized to be more variable cluster (Fig 4).

Comparing genome-covered and potential core SSRs

Jaccard's similarity coefficients among upland cotton varieties and hybrids employing genome-covered SSRs and potential core SSRs were presented in supplementary data (Table S2-S5)**.** In general, relatively lower similarity coefficients were based on potential core SSRs among varieties and hybrids than those of genome-covered SSRs (Table 4). Similarly, scrutiny of average number of polymorphic loci amplified, PIC and Rp based on two sets of primers revealed the proficient higher power of potential core SSRs to resolve and explicate relatively clearer picture of genetic constitution of tested materials (Fig. 5). Deviated similarity coefficients between two sets of primers ascribed specifically to higher PIC and Rp values and comprehensively to amplified number of loci (N). Moreover, Mantel test was performed among similarity coefficients obtained for 38 varieties and depicted variable trends for genome-covered and potential core SSRs i.e. R^2= 0.554 (Fig. 6a). Likewise, regression diagram for 55 hybrids varied to extent R^2=0.5755 for both sets of primers (Fig. 6b). Overall, slightly lower regression was observed for 55 hybrids.

Discussion

Screening for polymorphism and distribution of mapped SSR primers rendered the selection of 66 primers with minimum two pair of primers from each homologous chromosome, consequently, representing all possible linkage maps in the present study. In present study, the average number of polymorphic bands for two sets of SSRs (2.45 and 3.6) per primer was at par with that of previous researchers (Azmat and Khan, 2010; Bertini et al., 2006; Kalivas et al., 2011; Liu et al., 2006).

Average PIC and Rp values were found to be 0.68 and 0.98, respectively. Minimum, maximum and average PIC values were found higher than those of previous reports (Bertini et al., 2006; Liu et al., 2000). This fact might be due to the reason that varying climatic conditions of three distinct regions hoarded more genetic variations at DNA levels than that of exhibited by 97 BC_4F_4 accessions (Liu et al., 2000). Polymorphic primers exhibiting PIC>0.5 have been regarded as highly polymorphic primers (Zhang et al., 2012). Thus, the fact manifested that most primers utilized in our study were highly polymorphic. Out of 66 SSR primers, 41 pair of primers displayed PIC above average and 32 pair of primers showed Rp value above average. Neither PIC nor Rp alone did provide enough informative reliability to exclusively fingerprint and estimate the genetic diversity. Instead, the combination of PIC and Rp and number of alleles amplified has been exploited to examine markers for their ability to discriminate among accessions by many investigators (Hameed et al., 2012; Korkovelos et al., 2008). Therefore, a core set 13 primers, potential core SSRs, which performed better than the average execution of all 66 primers were employed and their enormous potential was exploited to identify minimum core set of highly informative primers.

SSR primers have been extensively employed for fingerprinting the large number of accessions in peach (Cheng and Huang, 2009), jute and rice (Chakravarthi and Naravaneni, 2006). Previously, researchers were able to assemble a core set of very few primers and had also testified their informative power such as in Greek olive (Roubos et al., 2010) and rice (Tu et al., 2007). However, our study found a minimum of 13 primers to differentiate among 93 cotton cultivars and hybrids from three distinct regions of China. A question may arise that why a large number of primers were needed in cotton? One possible answer implies that the larger genome size of cotton, the possible evolutionary influence of more than 50 species of genus *Gossypium* and narrowed genetic diversity due to bottleneck effect have collectively contributed to the loss of genetic variation at DNA level, which could be considered requisite for developing a fingerprinting core set of primers. Moreover, a core set of minimum primers (less than six or eight) in cotton is not under notice in our knowledge.

Regarding 38 varieties, average genetic similarity coefficients employing genome-covered SSRs and potential core SSRs were 0.55 and 0.53, respectively. This implied that analyzed set of varieties owned moderate diversity and genetic variation and might be considered as a good representative of upland cotton diversity of the region. Moreover, varieties originated in Xinjiang (NWDR) depicted vast diversity, greater extent of genetic variation at the DNA level as compared to cultivars and hybrids of YRCV and YzRCV. The argument laid because mean genetic similarity coefficients of all cultivars from Xingjian observed to vary from 0.48 to 0.50 for two sets of SSRs. These all findings were in correspondence of previous genetic diversity study of source germplasm of upland cotton originated in China (Chen and Du, 2006).

Genetic diversity estimates were measured employing both sets of SSRs separately on 93 accessions by categorizing in different populations to unveil the genetic constitution of present elite cultivars and hybrids in china. In conclusion, genetic constitution and variability of accessions from regions NWDR and YzRCV were noticed to be at par, but narrower than accessions from YRCV. These findings were also reported by other researchers (Chen and Du, 2006), thus revealed the fact that bottleneck effect, selection of more similar plants for hybridization programs and extensively repeated selection for specific desirable traits led to minute genomic differences among present cultivars and hybrids of YRCV regions.

Potential core SSRs were able to not only discriminate among accessions, but also successfully piled varieties and hybrids intimately based on respective regions of origin. This imparted significance to scope of employing potential core SSRs for being used in preliminary screening studies and amending upland cotton germplasm resources of the three generative regions of China by identifying distant parental lines for a purpose oriented breeding program.

Materials and Methods

Plant materials

Thirty eight traditional upland cotton cultivars and 55 commercial cotton hybrids, collectively 93 accessions, were employed in the present study. All accessions were broadly categorized based on type of material i.e. traditional variety or commercial hybrids and their origin i.e. Yellow River Cotton Valley (YRCV), Yangtze River Cotton Valley (YzRCV) or North West Dry Region (NWDR). Among the 38 cultivars, 25 were taken from Henan, Hebei and Shandong provinces (YRCV), one from Anhui and Jiangsu province each (YzRCV) and 11 cultivars from Xinjiang municipality and regiment (NWDR). Meanwhile, 16 hybrids were included from Henan, Hubei and Shandong provinces (YRCV) and 39 hybrids were taken from Anhui, Jiangsu, Hubei and Hunan province (YzRCV) made total of 55 hybrids (Table S6).

DNA extraction

Seeds of all varieties and hybrids were sown in polythene bags, 2-3 seeds per bag. Polyethylene bags were placed under controlled environmental conditions of glass house. Two weeks later, fresh leaves were taken and immediately placed in ice. DNA was extracted from leaf samples according to procedure proposed earlier for cotton (Paterson et al., 1993).

SSR analysis

Primarily, 324 SSR primers were screened for polymorphism, selected from linkage map constructed by (Yu et al., 2011), presenting primers from each chromosome across 12 accessions (cultivars and hybrids presenting three distinct regions). Ultimately, genome-covered SSRs were chosen, minimum two for each chromosome covering the whole genome of cotton, and employed to unveil genetic constitution of cotton cultivars and hybrids of three generative regions of China. These markers were selected initially as a consequence of screening which accomplished the criteria: (1) high polymorphism, (2) large number of

Table 3. Utilization efficiency of highly informative SSR fingerprints.

No.	Locus combination	Number of accessions distinguished
1	BNL3319	2
2	BNL3319 +BNL1317	15
3	BNL3319 +BNL1317 +BNL3255	23
4	BNL3319 +BNL1317 +BNL3255 +BNL3405	29
5	BNL3319 +BNL1317 +BNL3255 +BNL3405 +BNL1414	38
6	BNL3319 +BNL1317 +BNL3255 +BNL3405 +BNL1414 +NAU5499	48
7	BNL3319 +BNL1317 +BNL3255 +BNL3405 +BNL1414 +NAU5499+ HAU3236	55
8	BNL3319 +BNL1317 +BNL3255 +BNL3405 +BNL1414 +NAU5499+ HAU3236 +GH277	62
	All 13 loci	93

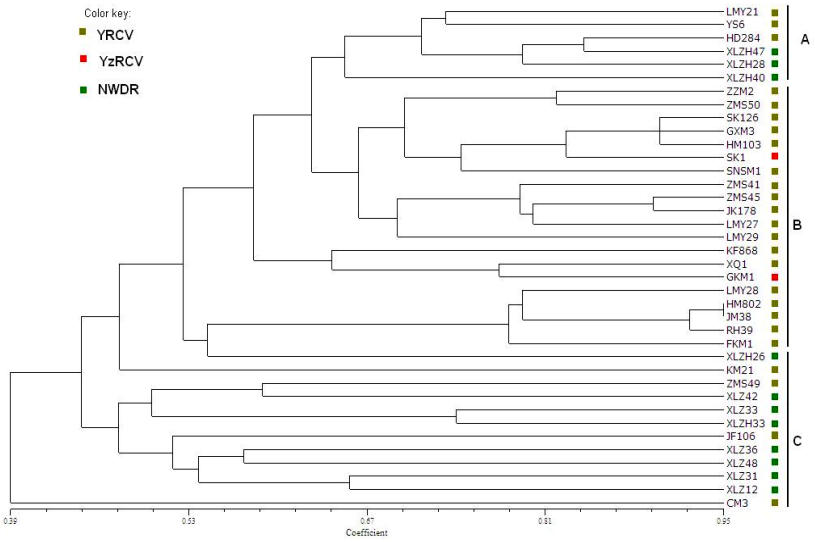

Fig 3. Cluster analysis of 38 upland cotton varieties originated from three growing regions (presented in color key) of China using 13 potential core SSRs analysis data from a Jaccard's similarity matrix and UPGMA method.

Table 4. Comparing similarity coefficients for upland cotton varieties and hybrids originated in China.

		Cultivars			Hybrids		
Primer set type	Size	Minimum	Maximum	Mean	Minimum	Maximum	Mean
Genome-covered SSRs	66	0.31	0.93	0.55	0.38	0.93	0.61
Potential core SSRs	13	0.22	0.95	0.53	0.27	0.95	0.55

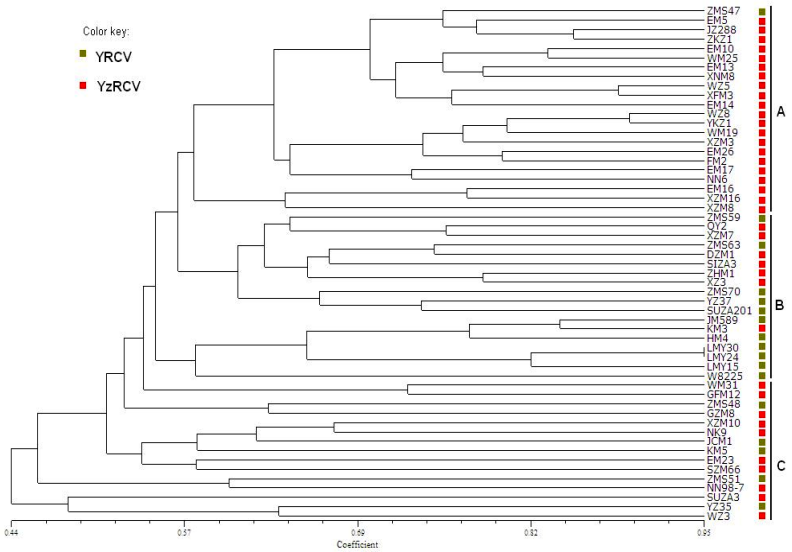

Fig 4. Cluster analysis of 55 upland cotton hybrid originated from two growing regions (presented in color key) of China using 13 potential core SSRs analysis data from a Jaccard's similarity matrix and UPGMA method

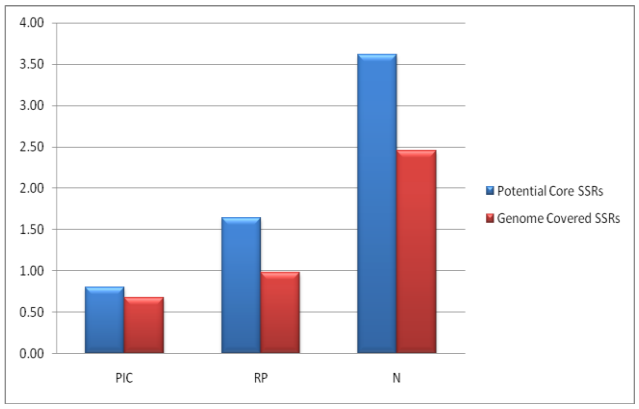

Fig 5. Comparing polymorphism information content (PIC), resolving power (Rp) and number of amplified alleles (N) between genome-covered and potential core SSRs.

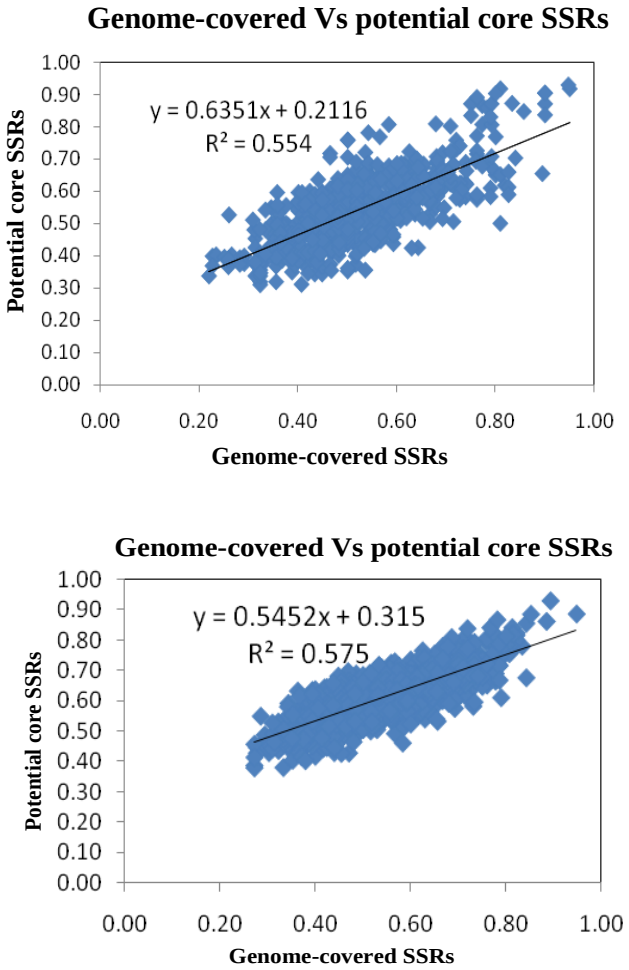

Fig 6. Regression diagrams of Mantel test between similarity coefficients of genome-covered SSRs vs. potential core SSRs. (a) regression among 38 upland cotton cultivars of different origin R^2=0.554; (b) regression among 55 upland cotton hybrids of different origin R^2= 0.575

identifying loci (3) ability to produce more polymorphic alleles and (4) ability to clearly differentiate among different varieties and hybrids. Verge of higher values of PIC (0.75) and Rp (1.05) provided with selection of 13 primers which were highly informative and engaged to constitute a core set of primer which displayed the benevolent potential of discrimination among all accessions. Thus, a minimal core set of 13 potential core SSRs was utilized to develop DNA fingerprints of 93 upland cotton cultivars and hybrids.

Constructing SSR fingerprints using potential core SSRs

Binary digital format of 13 highly informative primers were encoded to genotyping sequence format, contained (A) for presence and (C) for the absence of alleles, then used to construct SSR fingerprints of 93 accessions. Forty seven loci amplified by 13 primers were queued up forming a character string presenting SSR fingerprints of all individuals. To determine utilization efficiency of these SSR fingerprints, a separate genotyping sequence file was built for each cultivar and hybrid. Thus, a genotyping sequence file constituted overall scoring of an individual for 47 amplified loci by 13 SSR primers in an appended sequence order, as described by (Pan et al., 2007). All these files were processed as DNA sequence and aligned with Multiple Sequence Alignment program in DNAMAN® software (Lynnon Biosoft, Vaudreuil, Quebec, Canada). Multiple sequence alignment tools sorted all individuals into different homology groups; each group contained those individuals which featured less divergence in genotyping sequence among each other. This categorization eased further statistical analysis in determining the utilization efficiency of the constructed fingerprints. Subsequently, pair wise comparison conducted to all individuals from same homology group and genotyping sequence was compared. If a unique sequence was found, it was designated as fingerprint key of one particular cultivar or hybrid.

Polymerase chain reaction (PCR) and electrophoresis

PCR mixture (10 µL) consisted of 25 ng template DNA, 0.3 µM of each primer, 1.5 mM $MgCl_2$, 0.2mM dNTPs, 1X reaction buffer and 0.5 unit DNA polymerase (MBI). The mixture was subjected to amplification in a BIO-RAD thermal cycler for 5 min at 94°C and followed by 30-34 cycles of 1 min at 94°C, 1 min at specific annealing temperature 'Tm' (X°C) of each prime and then a final extension at 72°C for 5 minutes. After PCR, amplified products were separated on 6% (w/v) denaturing polyacrylamide (1:19 bis: acrylamide) gel and bands were visualized on silver staining.

Data Analysis

After staining, bands of DNA fragments were scored and converted into character format as (1) indicated the presence and (0) indicated absence, thus developing a binary digital format containing (0) and (1) for all polymorphic bands produced by 66 primers. Polymorphism information content (PIC) of each SSR locus calculated, based on calculation of allele frequency, using the following equation: $PIC=\sum pi^2$
Where, '*pi'* is the frequency of '*ith'* allele of a primer. Resolving power (Rp) calculated by following the procedure opted by (Dalamu et al., 2012). Both PIC and Rp were calculated using an excel worksheet.

The genetic diversity statistics were assessed using the frequency procedure of the GenAlEx6.2 software (Von Mark et al., 2007) for percentage of polymorphic loci (P), number of different alleles (Na), number of effective alleles (Ne), Nei's expected heterozygosity (He) and Shannon's information index (I). Pair wise Jaccard's similarity coefficients (Jaccard, 1980) were estimated employing the Numerical Taxonomy Multivariate Analysis System (NTSYS-pc), Version 2.1, software package followed by cluster analysis using the Un-weighted Pair Group Method with Arithmetic mean (UPGMA) (Rohlf, 2000).

Conclusion

The principal objective of the present study was to identify a set of primers which have the proficient potential to differentiate among cotton cultivars and hybrids of the region. Implementation of potential core of SSRs would turn to be a helpful preliminary tool for cotton breeders to successfully pick up genetically more distant parents for hybridization programs. Moreover, it would also assist in amelioration of cotton germplasm resources of three most productive regions of China, and eventually, to protect intellectual property (IP) rights of cotton breeders and assist in varietal registration.

Acknowledgements

This work was financially supported by the project "Molecule detection technology research on cotton seed authenticity and purity".

References

Asif M, Rahman M, Zafar Y (2005) DNA fingerprinting studies of some wheat (*Triticum aestivum* L.) genotypes using random amplified polymorphic DNA (RAPD) analysis. Pak J Bot 37:271-277.

Azmat MA, Khan AA (2010) Assessment of genetic diversity among the varieties of *Gossypium arboreum* and *Gossypium hirsutum* through random amplification of polymorphic DNA (RAPD) and simple sequence repeats (SSR) markers. Pak J Bot 42:3173-3183.

Bertini CHC, Schuster I, Sediyama T, Barros EG, Moreira MA (2006) Characterization and genetic diversity analysis of cotton cultivars using microsatellites. Genet Mol Biol 29:321-329.

Chakravarthi BK, Naravaneni R (2006) SSR marker based DNA fingerprinting and diversity study in rice (*Oryza sativa*. L). Afr J Biotechnol 5:684-688.

Chen G, Du XM (2006) Genetic diversity of source germplasm of upland cotton in China as determined by SSR marker analysis. Acta Genet Sin 33:733-745.

Cheng Z, Huang H (2009) SSR fingerprinting chinese peach cultivars and landraces (Prunus persica) and analysis of their genetic relationships. Scientia Hortic 120:188-193.

Dalamu TKB, Gaikwad AB, Saxena S, Bharadwaj C, Munshi AD (2012) Morphological and molecular analyses define the genetic diversity of Asian bitter gourd (*Momordica charantia*, L.). Aust J Crop Sci 6:261-267.

Ercisli S, Ipek A, Barut E (2011) SSR marker-based DNA fingerprinting and cultivar identification of olives (*Olea europaea*). Biochem Genet 49:555-561.

Fryxell PA (1992) A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). Rheedea 2: 108-165.

Galovic VO, Mladenovic-drinic S, Jelovac D, Navalusic J (2004) Application possibilities of AFLP fingerprinting technique in maize DNA profiling and plant variety protection. Genetika 36:133-142.

Hameed U, Pan YB, Muhammad K, Afghan S, Iqbal J (2012) Use of simple sequence repeat markers for DNA fingerprinting and diversity analysis of sugarcane (*Saccharum* spp) cultivars resistant and susceptible to red rot. Genet Mol Res 11:1195-1204.

Hossain MB, Haque S, Khan H (2002) DNA Fingerprinting of Jute Germplasm by RAPD. J Biochem Mol Biol 35:414-419.

Jaccard P (1980) Nouvelles researchers sur la distribution florale Société. Vaudoise des Sciences Naturelles 44:22-27.

Kalivas A, Xanthopoulos F, Kehagia O, 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.

Korkovelos AE, Mavromatis AG, Huang WG, Hagidimitriou M, Giakoundis A, Goulas CK (2008) Effectiveness of SSR molecular markers in evaluating the phylogenetic relationships among eight *actinidia* species. Scientia Hortic 116:305–310.

Liu D, Guo X, Lin ZX, Nie Y, Zhang XL (2006) Genetic diversity of Asian cotton (*Gossypium arboreum* L.) in China evaluated by microsatellite analysis. Genet Resour Crop Ev 53:1145-1152.

Liu S, Cantrell RG, McCarty JC, Stewart JM (2000) Simple sequence repeat-based assessment of genetic diversity in cotton race stock accessions. Crop Sci 40:1459-1469.

Owen CA, Bita EC, Banilas G, Hajjar SE, Sellianakis V, Aksoy U, Hepaksoy S, Chamoun R, Talhook SN, Metzidakis I, Hatzopoulos P, Kalaitzis P (2005) AFLP reveals structural details of genetic diversity within cultivated olive germplasm from the Eastern Mediterranean. Theor Appl Genet 110:1169-1176.

Pan YB, Scheffler BS, Richard J (2007) High-throughput genotyping of commercial sugarcane clones with microsatellite (SSR) DNA markers. Sugar Tech 9:176-181.

Pandey A, Mishra RK, Mishra S, Singh YP, Pathak S (2011) Assessment of genetic diversity among sugarcane cultivars (*Saccharum officinarum* L.) using simple sequence repeats markers. OnLine J Biol Sci 11:105-111.

Paterson AH, Brubaker CL, 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

Pharmawati M, Yan G, Finnegan PM (2005) Molecular variation and fingerprinting of *Leucadendron* cultivars (Proteaceae) by ISSR markers. Ann Bot 95:1163-1170.

Poljuha D, Sladonja B, Setic E, Milotic A, Bandelj D, Jakse J, Javornik B (2008) DNA fingerprinting of olive varieties in Istria (croatia) by microsatellite markers. Scientia Hortic 115:223-230.

Qu C, Hasan M, Lu K, Liu L, Liu X, Xie J, Wang M, Lu J, Odat N, Wang R, Chen L, Tang Z, Li J (2012) Genetic diversity and relationship analysis of the Brassica napus germplasm using simple sequence repeat (SSR) markers. Afr J Biotechnol 11:6923-6933.

Rahman M.H., Rajora O.P. (2002) Microsatellite DNA fingerprinting, differentiation, and genetic relationships of clones, cultivars, and varieties of six poplar species from three sections of the genus Populus. Genome 45:1083-1094.

Rohlf FJ (2000) NTSYS-pc numerical taxonomy and multivariate analysis system. version 2.1. Exeter Software, Setauket, NewYork.

Rojas G, Méndez MA, Muñoz C, Lemus G, Hinrichsen P (2008) Identification of a minimal microsatellite marker panel for the fingerprinting of peach and nectarine cultivars. Electron J Biotechno 11:4-5

Roubos K, Moustakas M, Aravanopoulos FA (2010) Molecular identification of Greek olive (*Olea europaea*) cultivars based on microsatellite loci. Genet Mol Res 9:1865-1876.

Sensi E, Vignani R, Scali M, Masi E, Cresti M (2003) DNA fingerprinting and genetic relatedness among cultivated varieties of *Olea europaea* L. estimated by AFLP analysis. Scientia Hortic 97:379-388.

Singh R, Nagappan J, Tan SG, Panandam JM, Cheah SC (2007) Development of simple sequence repeat (SSR) markers for oil palm and their application in genetic mapping and fingerprinting of tissue culture clones. Asia-Pac J Mol Biol 15:121-131.

Thomas G, Mohapatra T, Rao AR, Sharma RP (2006) Distinguishing Indian commercial wheat varieties using RAPD based DNA fingerprints. Indian J Biotechnol 5:200-206.

Tu M, Lu BR, Zhu Y, Wang Y (2007) Abundant Within-varietal genetic diversity in rice germplasm from yunnan province of China revealed by SSR fingerprints. Biochem Genet 45:789–801.

Von Mark VC, Luhman R, Marek LF, Rife CL, Shoemaker RC, Brummer EC, Gardner CA (2007) Characterization of flowering time and SSR marker analysis of spring and winter type Brassica napus L. germplasm. Euphytica 153: 43-57.

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

Yu Y, Yuan D, Liang S, Li X, Wang X, Lin Z, Zhang X (2011) Genome structure of cotton revealed by a genome-wide SSR genetic map constructed from a BC_1 population between *gossypium hirsutum* and *G. barbadense*. BMC Genomics 12:1-14.

Zhang J, Zhu WY, Xuan C, Sun H, Li Y (2012) Molecular fingerprinting and relative relationship of apricot cultivars in China by simple sequence repeat (SSR) markers. Afr J Biotechnol 11:2631-2641.

Zhang YX, Lin ZX, Li W, Tu LL, Nie YC, Zhang XL (2007) Studies of new EST-SSRs derived from *Gossypium barbadense*. Chinese Sci Bull 52:2522-2531.

Zhao WG, Miao XX, Zang B, Zhang L, Pan YL, Huang YP (2006) Construction of fingerprinting and genetic diversity of mulberry cultivars in China by ISSR markers. Acta Genet Sin 33:851-860.