Research Paper

Simple sequence repeat markers associated/linked with agronomic traits, as core primers, are eminently suitable for DNA fingerprinting in Upland cotton

Chengqi Li*, Bihua Chen, Xinjuan Xu, Dandan Li and Jinyuan Dong

Collaborative Innovation Center of Modern Biological Breeding, Henan Province/Cotton Research Institute, Henan Institute of Science and Technology, Xinxiang, Henan Province, 453003, China

Analyzing the genetic differences among crop germplasm resources scientifically and accurately is very important for the selection of core accessions, the identification of new cultivars, and the determination of seed purity. However, phenotypic selection *per se* is not sufficient to identify genetically distinct accessions. In this study, 26 out of 83 simple sequence repeat markers associated/linked with cotton important agronomic traits derived from our previous and other published research, corresponding to the 26 chromosomes of Upland cotton (*Gossypium hirsutum* L.), were selected as core primers for DNA fingerprinting construction. The 26 markers showed clear band patterns, good repeatability and high polymorphism. The average alleles, gene diversity index and polymorphism information content were 3.12, 0.4312 and 0.3830, respectively. Using TM-1, a genetic standard line for Upland cotton, as the control, DNA fingerprinting pattern and DNA barcodes were obtained based on the core primers. There was a significant positive correlation between genetic distance matrix determined using 26 core primers and that determined using more primers (335) derived from previous research, further suggesting that the core primers were eminently suitable for DNA fingerprinting in Upland cotton. This study provides a molecular basis for assessing identification, authenticity and seed purity of cotton cultivars.

Key Words: simple sequence repeat markers (SSRs), DNA fingerprinting, DNA barcode, Upland cotton, core primer.

Introduction

Cotton (*Gossypium* spp.) is an important economic fiber crop, and cotton production has a significant role in the global economy (Stephens and Mosley 1974). Among the four cultivated cotton species, namely Upland cotton (*Gossypium hirsutum* L.), Sea Island cotton (*Gossypium barbadense* L.), Asiatic cotton (*Gossypium arboreum* L.) and African cotton (*Gossypium herbaceum* L.), Upland cotton (2n = 52, AADD) is the most widely cultivated species worldwide due to its high yield and wide adaptability, representing 94% of the growing area and accounting for 95% of world cotton production (Chen *et al.* 2007). In the process of breeding and production of Upland cotton, identification of differences among cultivars or germplasm accessions mainly depends on the description of phenotypic traits, such as plant height, plant architecture, early maturity, yield,

fiber quality and disease resistance (Chen and Du 2006, Talib *et al.* 2015, Van Esbroeck *et al.* 1999). However, these traits vary according to environmental conditions, resulting in non-heritable phenotypic variation within a certain range occurring between different years or at different sites. Therefore, it is not reliable to identify individual accessions on the basis purely of phenotypic traits. Meanwhile, the genetic diversity of Upland cotton is much lower than that of the wild species because of the heavy reliance of commercial Upland cotton production on a limited number of genotypes as well as the long period of domestication of this crop under selection pressure for early maturity, insect resistance and disease resistance (Iqbal *et al.* 2001); as a consequence, the unambiguous identification of newly bred cultivars becomes more difficult. The application of transgenic technology to cotton breeding means that the number of cultivars with improvement in only a few or even a single gene has increased, so it is difficult to accurately distinguish the improved cultivars from their original parents by phenotypic traits. Therefore, establishing an objective, reliable and practicable identification technology to accurately identify different cotton cultivars is urgently needed.

With the rapid development of molecular marker technology, it has become possible to carry out fast and accurate identification of cultivars at the DNA level, which is not sensitive to environmental conditions (Gao *et al.* 2009). The International Union for the Protection of New Varieties of Plants (UPOV) has included identification of molecular markers into the Distinctness element of DUS (Distinctness, Uniformity and Stability) testing in crop varieties (See UPOV/INF/17 and UPOV/INF/18; http://www.upov.int/ information_documents/en/). In China, identification at the DNA level is also an important measure for cultivar quality monitoring, which also provides a theoretical and legal basis for cultivar protection (Wang *et al.* 2007). DNA barcodes refer to relatively short DNA fragments representing the species, and they are standardizable, easy to amplify, and show sufficient genetic variation (Zhao *et al.* 2012). A DNA barcode is a DNA fingerprint identity, a digital representation of DNA fingerprinting, which can be used to distinguish different cultivars accurately at the DNA level. Of the various molecular markers, simple sequence repeat markers (SSRs) have the advantages of high polymorphism, good reproducibility, co-dominant inheritance, short amplification products and widespread distribution (Kashi *et al.* 1997, Röder *et al.* 1998), making them among the preferred markers for constructing DNA barcodes. Presently, SSRs have been used for DNA fingerprinting or barcode construction in many crops, such as rice (Yan *et al.* 2011), maize (Li *et al.* 2005), wheat (Li *et al.* 2006), soybean (Gao *et al.* 2009), rapeseed (Chen *et al.* 2008), sugarcane (Liu *et al.* 2010), tomato (Dhaliwal *et al.* 2011), and sesame (Wei *et al.* 2017), providing a strong guarantee of cultivar identification and intellectual property protection.

In cotton, research into DNA fingerprinting using molecular marker technique has been reported. Punitha and Raveendran (2010) carried out a DNA fingerprinting study on colored- and white-linted genotypes, using randomly amplified polymorphic DNA (RAPD) markers, and cluster analysis showed clear-cut separation of the colored- and white-linted genotypes, forming three distinct clusters. Kuang *et al.* (2011) constructed a DNA fingerprinting database of 32 major Upland cotton cultivars from three main cotton regions of China, based on 36 SSRs. The result showed that ten cultivars could be distinguished using nine primers while thirty-two major cultivars could be identified using at least five primer combinations. In the study of Li *et al.* (2014), DNA fingerprinting of 30 major Upland cotton cultivars was analyzed using 20 SSRs. Four of the 30 cultivars had specific primers, i.e. each of the four cultivars could be distinguished from the others by using one specific primer; each of 26 cultivars could be distinguished from the others by using at least two primers. The molecular markers used in the above studies were all obtained from the cotton genome database and, therefore, were not directly related to cotton agronomic traits such as yield, fiber quality, plant architecture, etc., so the efficiency to identify and distinguish cultivars especially some new transgenic cultivars deferring from their parents only with respect to small genomic segments or a few traits, was not high.

In our previous studies, the SSRs distributed at approximately 10-cM intervals in each of 26 chromosomes the cotton tetraploid genetic map (Guo *et al.* 2007) and the SSRs linked to QTL for agriculturally and economically significant traits of cotton (Li *et al.* 2013, Mei *et al.* 2004, Nguyen *et al.* 2004, Qin *et al.* 2008, Shen *et al.* 2007, Song and Zhang 2009, Zhang *et al.* 2003) were selected to screen 172 Upland cotton cultivars. As a result, 331 polymorphic markers were obtained. Association mapping for important agronomic traits, including yield, fiber quality, early maturity and plant architecture, was conducted using these SSR markers, and some SSRs associated with target traits for cotton breeders were identified (Li *et al.* 2016a, 2016b, 2016c, 2017). On the basis of these studies, the SSR molecular markers associated with cotton agronomic traits were used in the current study as core primers to construct DNA barcodes of popular Upland cotton cultivars, providing a molecular basis for assessing identification, authenticity and seed purity of Upland cotton cultivars.

Materials and Methods

Plant materials

A total of 168 commercial Upland cotton cultivars, which have been or are being grown in different ecological cotton-growing areas of China, were selected as experimental materials. Of these, 61 were from the Yellow River basin, 26 were from the Yangtze River basin, 50 were from northwestern China, 20 were from northern China, and the remaining 11 were introduced from countries outside China. In 2015, all accessions were planted in the field of Henan Institute of Science and Technology, Xinxiang, Henan (113°52′ E, 35°18′ N, 95m asl) in single rows with 14–16 plants per row.

Sources and screening of core primers

"Core primers" refers to a set of primers with good characteristics of polymorphism, stability, and repeatability, which can be used as a preferred set of primers for related studies. In our previous study, association mapping for important agronomic traits of Upland cotton was carried out using 331 polymorphic SSR markers, and also a number of markers significantly associated with yield (Li *et al.* 2017), fiber quality (**Supplemental Table 1**, unpublished), early maturity (Li *et al.* 2016b, 2016c), and plant architecture (Li *et al.* 2016a) were identified. In this study, core primers will be screened from these traits-associated markers. In view of the fact that individual primers might not exhibit high polymorphism, stability and reproducibility, the SSR markers identified linked with at least two agronomic traits of cotton based on linkage mapping (Abdurakhmonov *et al.* 2009, Cai *et al.* 2014, Mei *et al.* 2013, Shao *et al.* 2014, Sun *et al.* 2012) were also used to screen the core primers. The three parameters, the number of alleles, the gene diversity index

(Di) and the polymorphism information content (PIC), were used to measure the polymorphism level of per primer for the 168 cotton accessions. The values of Di and PIC were estimated using the formulas for marker i:

$$Di = 1 - \sum_{j=1}^{n} P_{ij}^2, \text{ and}$$

$$PIC = 1 - \sum_{j=1}^{n} P_{ij}^2 - \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} 2P_{ij}^2 P_{ik}^2$$

where, n is the total number of alleles of marker i and P_{ij} is the frequency of the j_{th} allele of maker i in population.

PCR detection and genotyping

DNA was extracted from young leaves of cotton plants expressing the phenotype characteristic of the cultivar in each row, using the CTAB method (Paterson *et al.* 1993). The PCR amplification procedure was as follows: pre-denaturation was performed at 95°C for 4 min, denaturation was carried out at 94°C for 30 s, annealing was conducted at 57°C for 45 s, extension was performed at 72°C for 1 min, and the whole process was repeated for 30 cycles; at the end of the 30 cycles, extension was carried out at 72°C for 7 min, and then the temperature was maintained at 10°C for 10 min. The amplification products were separated by poly-acrylamide gel electrophoresis (PAGE) and detected by silver staining. SSR genotyping was carried out with reference to the methods of Zhao *et al.* (2012) and Mei *et al.* (2013), using TM-1, a genetic standard line for Upland cotton, as the control. Briefly, the band electrophoresis fingerprinting pattern in TM-1 (numbered 1 in the panel) was designated as 1, similar patterns were designated as 1, and different ones were in turn designated as 2, 3, 4, 5, and so on. In this way, the allelic variation matrix of all cultivars was consti-tuted with the fingerprinting codes at a marker locus.

Construction of DNA barcodes

Only one marker was selected from each of the 26 chro-mosomes of Upland cotton for construction of DNA bar-codes, based on the method developed by Yan *et al.* (2011) for the construction of the rice molecular identity database. As already noted, TM-1 was used as the control to obtain fingerprinting codes of each material. After the fingerprint-ing codes of all the accessions (including TM-1) were ob-tained using the core primers, they were arranged in order of chromosomes of Upland cotton, with the A subgenome (Chr.01–Chr.13) first, then D subgenome (Chr.14–Chr.26) next, to construct the DNA barcodes of the 168 Upland cot-ton cultivars.

Cluster analysis and Mantel test

When performing cluster analysis and Mantel test, the band electrophoresis fingerprinting pattern of each cultivar, such as 1, 2, 3, 4, 5, etc. was converted to the codes of 0 and 1, which represent the absence and presence of each of the corresponding alleles. In this way, the allelic variation matrix of cultivars was constituted with the codes of 0 and 1

at a marker locus. Nei's genetic distance of 168 cotton culti-vars based on both core primers and more primers was cal-culated using the provesti.dist() function of R package *poppr* (Kamvar *et al.* 2014). The Neighbor-Joining cluster dendro-gram for all cultivars based on the genetic distance matrix based on core primers was generated using the nei.dist() function of *poppr* (Kamvar *et al.* 2014). The correlation be-tween the genetic distance matrix determined using core primers and that determined using more primers was ana-lyzed by the Mantel test with 10000 permutations using the mantel function of R package *vegan* (Oksanen *et al.* 2017).

Screening and evaluation of core primers

A total of 78 SSRs, each associated with at least two ag-ronomic traits which had been identified using association mapping in our previous research based on (Li *et al.* 2016a, 2016b, 2016c, 2017), were located on 24 chromosomes with the exception of chr.04 and chr.05, and three unknown chro-mosomes (**Supplemental Table 2**). Here, these markers were repeatedly evaluated, and ultimately 21 SSRs located on 21 chromosomes with clear band patterns, good repeata-bility and high polymorphism information content were identified (**Table 1**). They were respectively CGR6078 on Chr.01 (A01), DPL0041 on Chr.02 (A02), MUSS162 on Chr.03 (A03), DPL0852 on Chr.07 (A07), CGR6103 on Chr.08 (A08), DPL0530 on Chr.09 (A09), NAU3467 on Chr.10 (A10), JESPR201 on Chr.11 (A11), CGR5193 on Chr.12 (A12), JESPR204 on Chr.13 (A13), NAU1070 on Chr.14 (D02), NAU3901 on Chr.15 (D01), TMB1268 on Chr.17 (D03), NAU2443 on Chr.18 (D13), NAU1102 on Chr.19 (D05), CGR6022 on Chr.20 (D10), NAU6966 on Chr.22 (D04), NAU5189 on Chr.23 (D09), DPL0068 on Chr.24 (D08), NAU3588 on Chr.25 (D06), and NAU3862 on Chr.26 (D12). In addition, five SSRs, MUSS193 on Chr.04 (A04) and DPL0131 on Chr.21 (D11) reported linked with at least two agronomic traits by linkage map-ping (Shao *et al.* 2014, Sun *et al.* 2012), NAU3269 on Chr.05 (A05), BNL3650 on Chr.06 (A06), and BNL2634 on Chr.16 (D07) reported associated with at least two agro-nomic traits by association mapping (Abdurakhmonov *et al.* 2009, Cai *et al.* 2014, Mei *et al.* 2013), were also repeatedly evaluated respectively. The amplification products of these five SSRs all had clear band patterns and good reproducibil-ity. As a result, a total of 26 SSRs, corresponding to the 26 chromosomes of Upland cotton, were identified as core primers for this study (**Table 1**).

The 26 SSRs revealed a total of 81 alleles ranging from 2 to 5 per primer with an average of 3.12. The Di and PIC ranged from 0.1346 to 0.6648 and from 0.1291 to 0.5907, respectively, and the average values were 0.4312 and 0.3830, respectively. The fingerprinting patterns of two rep-resentative primers, NAU1102 and DPL0041, on part of the cultivars, are shown in **Fig. 1**, where NAU1102 and DPL0041 amplified three and five alleles, respectively.

Table 1. Information of the 26 core SSR primers (markers) for DNA fingerprinting in Upland cotton

Core primer	Chr. (Subgenome)	Forward primer 5′-3′/Reward primer 5′-3′	Sequence source	Allele No.	Di	PIC	Agronomic traits associated/ linked with markers *	Literature cited
CGR6078	Chr.01 (A01)	CATGCAAGAAAGCTGCTCAA/ TAGGCATGTGTCTCCGTGTG	Genome	3	0.3515	0.3241	BW, FL, FS, FE, FU	Li *et al.* (2017); FL, FS, FE and FU are unpublished (**Supplemental Table 1**, similarly hereinafter)
DPL0041	Chr.02 (A02)	GCATCATATCATGTCCATTACAC/ GGGAGAGAGTGTAGTATGTTTGGG	Genome	5	0.4945	0.4666	LP, FL, FS, FE	Li *et al.* (2017); FL, FS and FE are unpublished
MUSS162	Chr.03 (A03)	TTGGTTGGTTAATTACGGGG/ GGCTTGTATCTCCCAGCAAG	EST	3	0.5675	0.5049	BW, LI	Li *et al.* (2017)
MUSS193	Chr.04 (A04)	GAAAATGAGCACTTCTCCGC/ AATGCGAATTGATCCAACAG	EST	2	0.4688	0.3589	**FL (12.9 cM), FM (2.22 cM), FU (6.20 cM), FE (19.71 cM)**	Sun *et al.* (2012)
NAU3269	Chr.05 (A05)	CGACTTAGCCGCCTATTAAA/ TTTATCCTGAACGACTTCC	EST	2	0.3809	0.3083	LY, SY, BN, LP, LI	Mei *et al.* (2013)
BNL3650	Chr.06 (A06)	TCGATTTCCTTATTTGATTTCTG/ AATTTGTCCAGATTCATTCTTCA	Genome	3	0.4808	0.4322	FL, FE	Abdurakhmonov *et al.* (2009)
DPL0852	Chr.07 (A07)	GTTCCAAATCAATCTCGTGT/ GGCTGTTACAGATCAAACTCCC	Genome	3	0.5184	0.4602	BN, FL, FS, FE, FU, HFFBN	Li *et al.* (2016c, 2017); FL, FS, FE and FU are unpublished
CGR6103	Chr.08 (A08)	CAAAGGATGGGACACAGGTAA/ TGCATTAGATACCGAAATGAGC	Genome	3	0.2725	0.2537	SY, FM	Li *et al.* (2017); FM is unpublished
DPL0530	Chr.09 (A09)	AGACTTACTTAAAGGCACCATTCG/ GCAGACTCTTCTGGTGTAACAGTG	Genome	3	0.4062	0.3705	FM, FL, FS, FU	FM, FL, FS and FU are unpublished
NAU3467	Chr.10 (A10)	AGCTAAGCGCTTCAAGTTGT/ ACGCATCCTAGAGGTCAGAA	EST	3	0.5159	0.4233	BP, BN, PH, EFB, FBA	Li *et al.* (2016a, 2016b, 2017)
JESPR201	Chr.11 (A11)	TCGATCAGTTAGGGTTTTGG/ CGAATCTCAACCAGATTTCC	Genome	3	0.5616	0.4807	FFBN, HFFBN	Li *et al.* (2016c)
CGR5193	Chr.12 (A12)	GGCACAGGTGCCCTCCTTA/ AGCAAGTCCGGCACAATC	Genome	3	0.1768	0.1692	LY, LP, LI	Li *et al.* (2017)
JESPR204	Chr.13 (A13)	CTCCAGGTTCAATGGTCTG/ CCCTCCATAACCAAAAGTTG	Genome	2	0.2934	0.2503	SP, BP, FM	Li *et al.* (2016b); FM is unpublished
NAU1070	Chr.14 (D02)	ACCAACAATGGTGACCTCTT/ AAGACAAAAGGCAAGGACAC	EST	3	0.6326	0.5549	LP, TFB, EFB, FBA	Li *et al.* (2016a, 2017)
NAU3901	Chr.15 (D01)	CTTGGAAAAAGGAAGGAGCAG/ AACAACATTGAAAGTCGGGG	EST	4	0.4734	0.4186	LP, LI, FM	Li *et al.* (2017); FM is unpublished
BNL2634	Chr.16 (D07)	CCCAGCTGCTTATTGGTTTC/ CAGGTACCATTGATGCCAAA	Genome	3	0.6111	0.5355	FL, FS, FM	Cai *et al.* (2014)
TMB1268	Chr.17 (D03)	CTCGAAACCTAGTGCCCTGT/ CGTTGGAGAAGGAAAGCCTAA	Genome	3	0.2917	0.2723	FBP, GP, FFBN	Li *et al.* (2016b, 2016c)
NAU2443	Chr.18 (D13)	AGCCTGCTTCATGTTCTTTT/ ATCTCTCTGTCTCCCCCTTC	EST	4	0.3289	0.3081	FS, FE, FU, HFFBN	Li *et al.* (2016c); FS, FE and FU are unpublished
NAU1102	Chr.19 (D05)	GCATATCTGGCGGGTATAAT/ TGTTTGGCATAAACCGAAAG	EST	3	0.6648	0.5907	GP, FFBN	Li *et al.* (2016b, 2016c)
CGR6022	Chr.20 (D10)	TTCTCTATAACCTCTACCCGCCTA/ ACATACGGGTTGAAATGTACTCCT	Genome	3	0.5709	0.5037	BW, SI	Li *et al.* (2017)
DPL0131	Chr.21 (D11)	ATGAATGCAGATCATTACGCCT/ GTCATCATTATCGTCAAGTC	Genome	3	0.3430	0.3164	**FE (0–14.4 cM), FS (0–14.4 cM)**	Shao *et al.* (2014)
NAU6966	Chr.22 (D04)	AAAGTGAGTTAAGAAAGGCT/ TGTCCCCAATCATATTTTC	Unknown	3	0.3909	0.3572	BW, LI, SI	Li *et al.* (2017)
NAU5189	Chr.23 (D09)	CAACTTCCCAAGCTCGTATT/ GTTCAACAGGTCTGTACCAGTTCC	EST	3	0.6602	0.5860	FM, PH, HFFBN	Li *et al.* (2016a, 2016c); FM is unpublished
DPL0068	Chr.24 (D08)	GCAAATGATCTCTGCCCTGTAA/ CCCATAGGGCATAACTTCTA	Genome	3	0.2997	0.2780	FBP, GP	Li *et al.* (2016b)
NAU3588	Chr.25 (D06)	GCCAACAAGACATAACAACAC/ TTGGAGAGGGAGATTGGTAG	EST	5	0.3216	0.3038	FM, FL, FS	FM, FL and FS are unpublished
NAU3862	Chr.26 (D12)	GGATGAACTTTGCTTTAGCC	EST	3	0.1346	0.1291	SY, LP	Li *et al.* (2017)

Di = gene diversity index; PIC = polymorphism information content; * BN = boll number per plant, BP = budding period, BW = boll weight, EFB = effective fruit branches, FBA = fruit branch angle, FBP = flowering and boll period, FE = fiber elongation, FFBN = first fruiting branch node, FL = fiber length, FM = fiber micronaire, FS = fiber strength, FU = fiber uniformity, GP = growth period, HFFBN = height of first fruiting branch node, LI = lint index, LP = lint percentage, LY = lint yield, PH = plant height, SI = seed index, SP = seedling period, SY = seed cotton yield, TFB = total fruit branches; The trait marked as plain means it is associated with marker by using association mapping, whereas the trait marked as bold means it is linked with marker with the genetic distance (cM) of marker-trait by using linkage mapping.
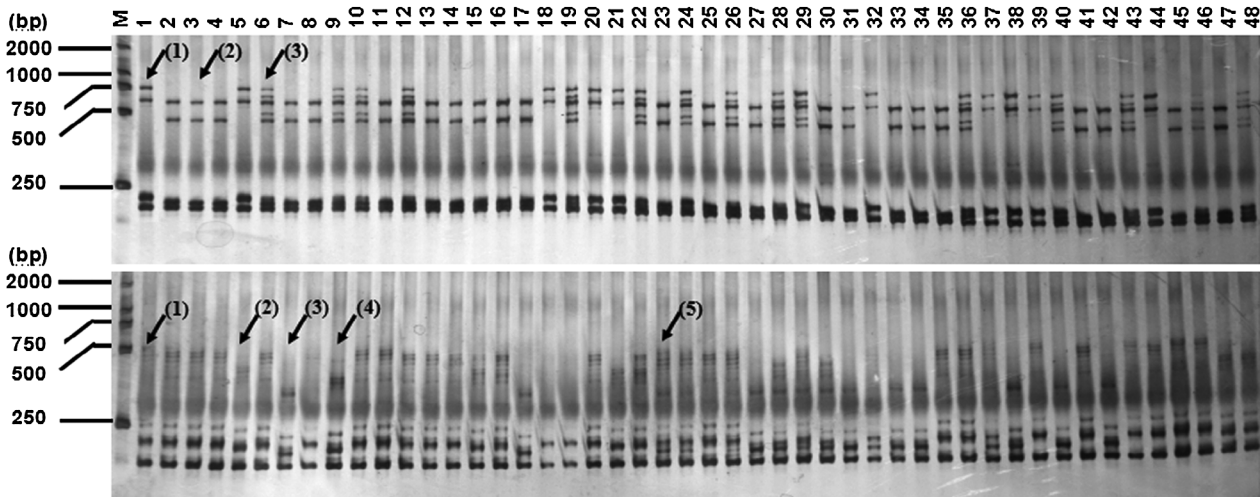
Fig. 1. Electrophoresis fingerprinting patterns of two representative primers, NAU1102 and DPL0041, for part of the Upland cotton cultivars. 1–48 corresponds to the materials in **Table 2**, respectively; Upper for primer NAU1102 and Lower for primer DPL0041; Dotted arrows mark different alleles.

Construction of DNA barcodes based on core primers

The genomic DNA of 168 cotton cultivars was amplified using the 26 core primers, and their corresponding fingerprints were obtained. Using TM-1 as the control, the electrophoresis fingerprinting pattern detected in TM-1 was denoted as 1, and fingerprinting codes of all cultivars (including TM-1) were obtained, which were recorded as 1, 2, 3, 4, 5 and so on. These codes were organized to form a data set according to the sequential arrangement of chromosome number and the order of A subgenome (Chr.01–Chr.13) first, followed by D subgenome (Chr.14–Chr.26), resulting in a code combination for the SSRs on 26 chromosomes for each cultivar. A total of 26 digits formed the SSR relative molecular identity number, i.e. DNA barcode, unique to each cotton cultivar (**Table 2**). The electrophoresis fingerprinting patterns corresponding to the barcodes are shown in **Fig. 2**. Barcodes and electrophoresis fingerprinting patterns can be used for verification of the identity of a specific cultivar, thus achieving the purpose of cultivar identification.

Cluster analysis of Upland cotton based on core primers

Revealing the genetic relationship among cultivars can provide important supporting information for cultivar identification. Based on genotype data from the 26 core primers, cluster analysis of the 168 Upland cotton cultivars was carried out and Neighbor-Joining dendrogram was obtained (**Fig. 3**). The 168 cultivars could be categorized into nine groups (A–I). Group A contained 17 cultivars, of which, three came from the Yellow River basin, one from the Yangtze River basin, 12 from northwestern China, and one from abroad. Group B contained 27 cultivars, of which, 10 came from the Yellow River basin, three from the Yangtze River basin, six from northwestern China, seven from northern China, and one from abroad. Group C contained 12 cultivars, of which, nine came from the Yellow River basin, the other three from the Yangtze River basin, northwestern China and northern China, respectively. Group D contained 33 cultivars, of which, 17 came from the Yellow River basin, nine from the Yangtze River basin, four from northwestern China, one from northern China, and two from abroad. Group E contained 13 cultivars, of which, six from northwestern China, four from northern China, and three from abroad. Group F contained 18 cultivars, of which, one came from the Yellow River basin, three from the Yangtze River basin, 11 from northwestern China, one from northern China, and two from abroad. Group G contained 21 cultivars, of which, 12 came from the Yellow River basin, two from the Yangtze River basin, three from northwestern China, and four from northern China. Group H contained 11 cultivars, of which, three came from the Yellow River basin, four from the Yangtze River basin, three from northwestern China, and one from abroad. Group I contained 16 cultivars, of which, six came from the Yellow River basin, three from the Yangtze River basin, four from northwestern China, two from northern China, and one from abroad.

Mantel test of genetic distance matrix

Correlation analysis between the genetic distance matrix determined by the 26 core primers and that determined by the 335 primers (330 from our previous work and the other five from published research) for 165 cultivars [the three cultivars, TM-1, Baimian5 and Yumian1(HN), were removed from 168 cultivars due to their absences in the detecting of 309 primers out of 26 core primers in our previous work] was carried out using the Mantel test with 10000 permutations. The R scripts can be found in **Supplemental Data 1**; the raw genotype data of 168 cultivars for 26 primers and 165 cultivars for 335 primers can be found in **Supplemental Data 2** and **Supplemental Data 3**, respectively.

Table 2. DNA barcodes of the 168 Upland cotton cultivars

No.	Cultivar	Ecological origin	DNA barcode *	No.	Cultivar	Ecological origin	DNA barcode *
1	TM-1	USA	1111111111111-1111111111111	85	Liaomian8	Northern China	1112123212311-2423131112111
2	KK1543	Former Soviet Union	1122112222221-2222221121112	86	Liaomian16	Northern China	1111112211121-1312133121112
3	Heishanmian	Northern China	1121121232111-2332221111221	87	Liaomian18	Northern China	2211112211211-1312121121111
4	Xinluzao1	Northwestern China	1121113212311-2332122211231	88	Liaomian19	Northern China	1111112211111-1312121111111
5	Xinluzao2	Northwestern China	1212112212211-2312112211311	89	Baimian985	Yellow River	1111111211111-1112112112111
6	Xinluzao3	Northwestern China	1121131212222-2312332112112	90	Yumian1(CQ)	Yangtze River	1131111211111-3312412112111
7	Xinluzao4	Northwestern China	1311113212212-2322122213111	91	Baimian5	Yellow River	3111211212111-1332112131111
8	Xinluzao5	Northwestern China	1111112211211-2412122312111	92	Yumian1(HN)	Yellow River	1112211211211-1311112113111
9	Xinluzao6	Northwestern China	1431112212211-3312333211131	93	Yumian5	Yellow River	1222113211311-2411132313111
10	Xinluzao7	Northwestern China	1112111232311-3312133313211	94	Yumian7	Yellow River	1131222221211-3412222111111
11	Xinluzao8	Northwestern China	1112111232311-3312133313211	95	Yumian12	Yellow River	1512131132211-3321133112121
12	Xinluzao9	Northwestern China	1111111212211-1412133213111	96	Yumian21	Yellow River	2131111231211-3322133132211
13	Xinluzao10	Northwestern China	1131111212111-3422321112111	97	Lumian1	Yellow River	1131211312211-3412112113111
14	18-3	Northern China	1122121212111-2332421113111	98	Lumian4	Yellow River	1131231311211-2312112113111
15	Xinluzao11	Northwestern China	1111111211211-2312123112141	99	Lumian6	Yellow River	1311111112211-1311112111111
16	Xinluzao12	Northwestern China	3131132211211-2412121133111	100	Lumian10	Yellow River	1111111211211-1312112113111
17	Xinluzao13	Northwestern China	1311111211212-1422122113111	101	Lumianyan21	Yellow River	1111131212111-1412112113111
18	Xinluzao15	Northwestern China	1111122211131-2312112113111	102	Lumianyan28	Yellow River	1111211212211-2412112123111
19	Xinluzao16	Northwestern China	2112213223112-1322132111111	103	Lumianyan29	Yellow River	1121112211111-2312332111111
20	Xinluzao17	Northwestern China	1111112211111-1312112312111	104	Shiyuan321	Yellow River	1131111211131-3423121111111
21	Xinluzao18	Northwestern China	1221111211111-2312112113131	105	Fenwu195	Yellow River	1121111212231-2433122311111
22	Xinluzao19	Northwestern China	1221111211211-2312133113121	106	Qianjiang9	Yangtze River	1111111111111-1432132113111
23	Xinluzao20	Northwestern China	3521113232112-2222422311131	107	Xiangmian3	Yangtze River	1112111232111-1312112133111
24	Xinluzao21	Northwestern China	1521113212112-2322132312111	108	Xiangmian10	Yangtze River	1121213211111-2332132213111
25	Xinluzao22	Northwestern China	1111211221212-1322121121321	109	Yanmian48	Yangtze River	1112111212211-1312332312211
26	Xinluzao23	Northwestern China	1511111331312-3322131132221	110	Jiangsumian1	Yangtze River	1112211211111-1322322313111
27	Xinluzao24	Northwestern China	2311113111212-1332121121311	111	Sumian1	Yangtze River	1312111212211-1332322112211
28	Xinluzao25	Northwestern China	1412111212211-2131133111111	112	Sumian6	Yangtze River	3221211211231-2322121111142
29	Xinluzao26	Northwestern China	1512112211111-3312112111111	113	Sumian9	Yangtze River	3112211211111-1312312113111
30	Xinluzao27	Northwestern China	3412113211211-2312122222211	114	Sumian10	Yangtze River	1132131212111-3322121132111
31	Xinluzao28	Northwestern China	2311113321212-1332121121311	115	Sumian12	Yangtze River	3311111212211-1312122312111
32	Xinluzao29	Northwestern China	2521211231111-2332311113111	116	Sumian16	Yangtze River	3331212311222-3333121332111
33	Xinluzao30	Northwestern China	2321213223212-2312422211131	117	Xuzhou142	Yangtze River	3112221212111-2412313312111
34	Xinluzao31	Northwestern China	2311113321212-1321121121321	118	Qiannong465	Yangtze River	1532132332311-3312133112111
35	Xinluzao32	Northwestern China	2111113211111-1422421213111	119	Dongting1	Yangtze River	1111212111111-1311111112111
36	Xinluzao33	Northwestern China	1112111211312-1322431113111	120	Xinqiu1	Yellow River	1111131331311-3113112112111
37	Xinluzao34	Northwestern China	1512111212111-2322111311111	121	Simian2	Yangtze River	1112131211111-3112332113111
38	Xinluzao35	Northwestern China	1321211211111-2322132111111	122	Simian3	Yangtze River	1133211232211-3312132112112
39	Xinluzao36	Northwestern China	1131121222111-3412111113311	123	Simian4	Yangtze River	1111211231111-1422122113111
40	Xinluzao37	Northwestern China	1312111321112-1422132131111	124	Chuanmian56	Yangtze River	2531132231221-2322333113112
41	Xinluzao38	Northwestern China	1122131213211-2323221223111	125	Jinzhong169	Northern China	1121111111111-2312122113151
42	Xinluzao39	Northwestern China	2311113121212-1322121121311	126	Jinzhong200	Northern China	1111131121132-2312222111112
43	Xinluzao40	Northwestern China	2111112211212-1322133112212	127	Jinmian5(SX)	Northern China	1222121221111-2322121112231
44	Xinluzao41	Northwestern China	1112211212211-2422112113111	128	Jinmian6	Northern China	1211211211212-2311121111331
45	Xinluzao42	Northwestern China	2122121222211-2412122331111	129	Jinmian8	Northern China	1531131331111-2313131113211
46	Baimian1	Yellow River	1111131212211-1412131113111	130	Jinmian9	Northern China	1232221211131-2411131113213
47	Xinluzao46	Northwestern China	2211113211211-1412121121111	131	Jinmian13	Yellow River	1131121211111-3112332113111
48	Xinluzao47	Northwestern China	1232112321212-1322131131111	132	Jinmian14	Northern China	1111112212232-1332121111111
49	Xinluzao48	Northwestern China	3121113213111-2332312113111	133	Jinmian24	Northern China	1131131212211-3232113111211
50	Xinluzao49	Northwestern China	1321112113211-2322312113111	134	Jinmian29	Yellow River	1111112212211-1132122111151
51	Xinluzao51	Northwestern China	3112131211211-2322312313111	135	Jinmian36	Yellow River	1321112112212-1332122111111
52	Xi9	Yellow River	3111113311211-2322312132111	136	Jinmian45	Yellow River	1112211212111-1312122133111
53	Xinluzhong36	Northwestern China	1212113211311-3322312112111	137	Jinmian1	Northern China	1122221212111-2312122133111
54	CRI8	Yellow River	2111111211213-3322112131111	138	Jinmian2	Northern China	3132221211311-3312132132221
55	CRI10	Yellow River	1211111212211-2422121111111	139	Jinmian4	Northern China	1122111221212-2322132312131
56	CRI12	Yellow River	1132211211122-3322112113111	140	Jinmian5	Northern China	1111121211211-2412122111111
57	CRI13	Yellow River	1112111221111-2323122113121	141	Guoxinmian3	Yellow River	3132132212212-1322132133121
58	CRI14	Yellow River	3112113213311-2321232212131	142	Jimian958	Yellow River	1111211212211-2132111123111
59	CRI15	Yellow River	1111111211211-1412111111111	143	Jimian1	Yellow River	1121221211211-2131311113111
60	Zhong1707	Yellow River	1211111211111-1323112111111	144	Jimian7	Yellow River	1122122213312-1332111113121
61	CRI17	Yellow River	1112211231211-2323132112111	145	Jimian12	Yellow River	1112111321311-1312123113131
62	CRI18	Yellow River	1112121211111-2322133133111	146	Ejing1	Yangtze River	1111111331111-1322123113121
63	CRI19	Yellow River	3132131212211-2332123113111	147	Emian3	Yangtze River	1511211212211-1412111111111
64	CRI20	Yellow River	1122132221212-2412132112111	148	Emian14	Yangtze River	2521131211211-2222132122111
65	CRI22	Yellow River	1122232211211-3422121311111	149	Esha28	Yangtze River	1531132212311-3332132112111
66	CRI23	Yellow River	1111211211221-3312112133112	150	Jing8891	Yangtze River	1111111211111-1312121113131
67	CRI24	Yellow River	1111111231111-1112113111111	151	Daihongdai	Yangtze River	1211111211211-1312122211311
68	CRI25	Yellow River	1112111211211-2412112122111	152	STV2B	USA	1122112312211-2312133113111
69	CRI26	Yellow River	1132111232211-3412111112211	153	DPL15	USA	1112211211211-2322133113111
70	CRI27	Yellow River	1111113212211-1412111113311	154	DPL16	USA	3512131211211-1331112111121
71	CRI30	Yellow River	1222111222211-2312133112111	155	Shanmian4	Yellow River	1112232212211-1312113132111
72	CRI33	Yellow River	1111111211211-1412122111121	156	Shan1155	Yellow River	1112233211211-2322121112111
73	CRI34	Yellow River	1111111211211-1322111111111	157	Shan2365	Yellow River	1121211211211-2412122123111
74	CRI35	Yellow River	1111111212211-1432111111111	158	Handan802	Yellow River	1132131211211-3112133132111
75	CRI36	Yellow River	1111111211211-1432122111121	159	Handan885	Yellow River	1131121211211-3412122313111
76	CRI37	Yellow River	1112131212311-1312132332111	160	Ganmian8	Yangtze River	1111211211211-1312112113111
77	CRI40	Yellow River	1122113211211-2332112111111	161	Uganda3	Uganda	1121121211131-2312122113113
78	CRI50	Yellow River	1131112211112-1322132112321	162	Lvzao254	Northwestern China	1111132231211-1312132112111
79	CRI58	Yellow River	1511111211111-1313112113231	163	Shixuan87	Uzbekistan	1111122211211-2311121111111
80	CRI64	Yellow River	1111112311211-1422122311111	164	Beishinuo	USA	1122112212211-2122122313111
81	Zhongzhimian2	Yellow River	1112231211311-3412122122111	165	99M4	USA	1121112231211-2322132113111
82	Liaomian4	Northern China	1111212211211-1312132111121	166	99M7	USA	1331112311232-3333132122113
83	Liaomian5	Northern China	1131212211211-2312132111111	167	99M8	USA	1522112212311-2312122113111
84	Liaomian7	Northern China	1112111212311-1312122131111	168	Lamagan77	Northwestern China	3522113211112-2333332313211

* mean the DNA codes were organized according to the sequential arrangement of chromosome number and the order of A genome first, followed by D genome: CGR6078/[Chr.01 (A01)], DPL0041/[Chr.02 (A02)], MUSS162/[Chr.03 (A03)], MUSS193/[Chr.04 (A04)], NAU3269/[Chr.05 (A05)], BNL3650/[Chr.06 (A06)], DPL0852/[Chr.07 (A07)], CGR6103/[Chr.08 (A08)], DPL0530/[Chr.09 (A09)], NAU3467/[Chr.10 (A10)], JESPR201/[Chr.11 (A11)], CGR5193/[Chr.12 (A12)], JESPR204/[Chr.13 (A13)], NAU1070/[Chr.14 (D02)], NAU3901/[Chr.15 (D01)], BNL2634/[Chr.16 (D07)], TMB1268/[Chr.17 (D03)], NAU2443/[Chr.18 (D13)], NAU1102/[Chr.19 (D05)], CGR6022/[Chr.20 (D10)], DPL0131/[Chr.21 (D11)], NAU6966/[Chr.22 (D04)], NAU5189/[Chr.23 (D09)], DPL0068/[Chr.24 (D08)], NAU3588/[Chr.25 (D06)], NAU3862/[Chr.26 (D12)].
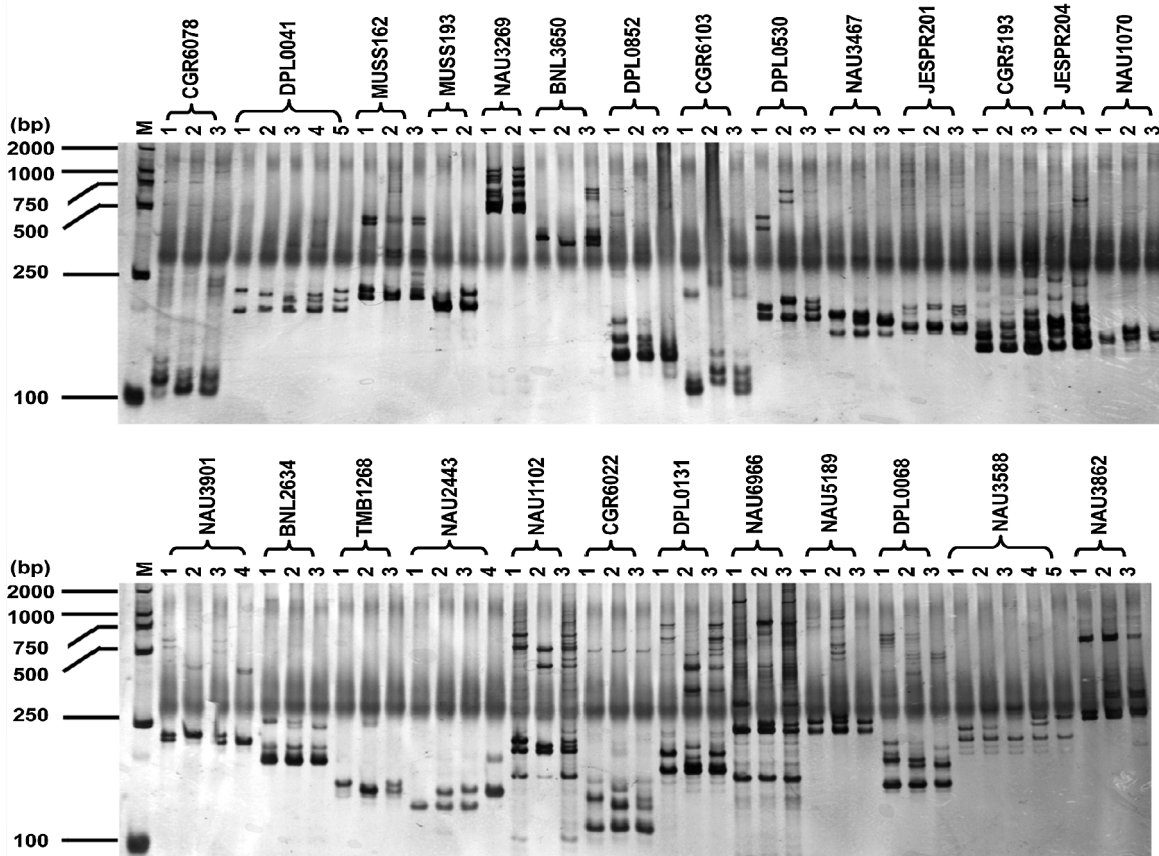
Fig. 2. Electrophoresis fingerprinting patterns of 26 core primers corresponding to DNA barcodes.
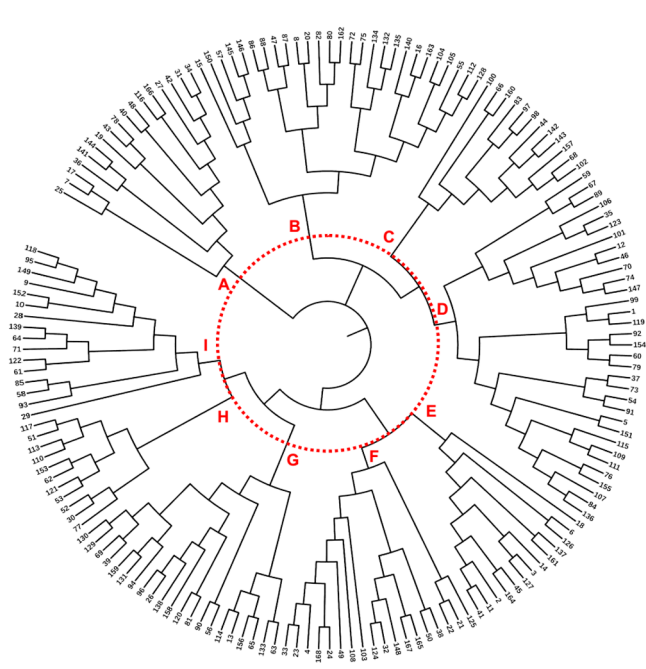

Fig. 3. Neighbor-Joining (NJ) dendrogram for 168 Upland cotton cultivars based on the genetic distance matrix determined using 26 core primers (All cultivars were categorized into nine groups, Group A–I).

The results showed that there was a significant positive correlation between the two genetic distance matrices ($r = 0.3078$, $P = 9.999e-05$), indicating that the genetic relationship revealed by the 26 core primers was highly similar to that revealed by the 335 primers, further illustrating that using these core primers for diversity evaluation of Upland cotton cultivars was reliable and they were eminently suitable for DNA fingerprinting in Upland cotton.

Discussion

DNA barcode is the digital representation of DNA fingerprinting. Different researchers use different DNA barcoding methods for experimental materials. Most of the studies code marker genotypes as 1 or 0, according to the presence or absence of the band, respectively, and construct DNA fingerprints using 0-1 numeric string (Dhaliwal *et al.* 2011, Li *et al.* 2006, Liu *et al.* 2010, Pan *et al.* 2008) and denary numeric string (Li *et al.* 2014, Pan *et al.* 2010). As Upland cotton (2n = 52, AADD) is an allotetraploid, the number of SSR bands is large and the bands appear in groups, so that the SSR band pattern of Upland cotton is more complex than for diploid crops. Thus marking the existence or nonexistence of each band with 1 or 0 might cause the phenomenon of misjudgment. Therefore, assignment of different

electrophoresis fingerprinting patterns to different cultivars was carried out in this study based on the method of Zhao *et al.* (2012). In view of the fact that TM-1, a genetic standard line for Upland cotton, is highly homozygous and has wide application in genetic and breeding research, it was used as the control and its electrophoresis fingerprinting pattern was assigned the code 1. The genotype codes of other cultivars for individual primers were sorted by the order of chromosomes where the markers were located and the order of chromosomes from genome A first and genome D next, resulting in the construction of unique DNA barcode of each of the cotton cultivars. Generally, these codes are less than 10, and also, the electrophoresis fingerprinting patterns corresponding to the codes are shown in **Fig. 2**. For the above reasons, our method for DNA fingerprints in Upland cotton is more reasonable and convenient than previous traditional methods. We propose that TM-1 could be used as the standard control material in the future for construction of all DNA barcodes of Upland cotton, so that DNA barcodes from different research groups can be used for comparison and identification on the DNA level among different cultivars.

In recent years, cotton DNA fingerprinting has been widely reported. Guo *et al.* (1996) constructed DNA fingerprinting of 9 main Upland cotton cultivars in China using RAPD markers. Song *et al.* (1999) distinguished 8 cotton cultivars from each other using 26 AFLP markers. Wang and Li (2002) obtained the DNA fingerprinting of brown cotton "Three Lines" (sterile, maintainer, restorer) and their hybrid F_1 using AFLP markers. However, these studies are in the early stages of molecular marker development, and the experimental materials for constructing fingerprinting are mainly confined in the target materials without considering more background materials. In the process of DNA fingerprinting construction, it is necessary to select more background germplasm for screening primer polymorphism and for enhancing the specificity of cultivar identification (Li *et al.* 2011). Because of the low level of polymorphism among Upland cotton cultivars, in order to identify SSR loci that were involved in every chromosome, which exhibited high polymorphism and effectively distinguished genomic characteristics of different cotton accessions, 168 cultivars of Upland cotton were selected as experimental materials for repeated screening of primers. These cultivars acted as not only target materials but also background materials. They were either from one of the four major cotton-growing areas in China, each with different environmental characteristics, or introduced from foreign countries, so the range of pedigrees was very rich, including members of the Stoneville pedigree, such as Lumian6, Sumian1 and Sumian12, the Foster pedigree, such as Xinluzao11, Liaomian10 and Liaomian18, the King pedigree, such as Heishanmian, Xinluzao10 and Xinluzao23, the Deltapine pedigree such as Xinluzao5, CRI19 and Yumian1, Trice cotton, such as Qiannong465, and the Uganda pedigree such as CRI12, Yumian21 and Uganda3. Neighbor-Joining cluster dendro-gram based on the 26 core primers further reflected the genetic relationship among cultivars, which were categorized into nine groups. Therefore, the 168 cultivars were used as background materials to identify genome-wide polymorphic markers, and as target materials to identify stable polymorphic markers on 26 chromosomes of Upland cotton for the follow-up DNA fingerprinting.

SSR markers are among the most widely-used molecular markers at present. SSR markers are co-dominant, simple and easy to automate and abundant in quantity. Since their amplification products are stable, they have great advantages for use in the analysis of cotton genome evolution, genetic diversity and DNA fingerprinting (Guo *et al.* 2003, Kantartzi *et al.* 2009, Khan *et al.* 2000). SSRs can be classified into three types (Li *et al.* 2014). The first type are markers within genes, which can identify genes associated with important agronomic traits in crops. The second type are markers closely linked to traits, which can identify disease resistance and stress tolerance genes and some important quality trait genes in crops. The third type are markers loosely linked with trait genes. When using DNA fingerprinting to identify crop cultivars and to determine their purity, priority should be given to the first two types of markers considering the correlation between markers and traits. In the past, not only was the number of basic primers used for analysis of cotton fingerprinting limited, but these primers were mostly of the third type, so that the polymorphism level of primers was low, or there was no corresponding relationship between polymorphism and phenotypic differences among cultivars, resulting in low detecting efficiency. In this study, the basic primers used for identifying core primers were from two sources. One source was the SSR markers selected from 26 chromosomes of Upland cotton based on the genetic map constructed by previous research (Guo *et al.* 2007). Although these markers were mainly of the third type, they had high polymorphism themselves among accessions because they are derived from the published genetic map. The other source was published SSR markers closely linked to important trait genes of cotton (Li *et al.* 2013, Mei *et al.* 2004, Nguyen *et al.* 2004, Qin *et al.* 2008, Shen *et al.* 2007, Song and Zhang 2009, Zhang *et al.* 2003). These markers were mainly of the first and the second types. Therefore, compared with previous studies, the basic primers selected in this study had a wider range of sources and could be used for effective screening of polymorphisms among accessions.

Determination of core primers is an important part of a DNA fingerprinting analysis system, and it is also a key step in the commercialization of DNA fingerprint identification. It not only greatly reduces the cost of primer synthesis, but also greatly reduces the intensive work of screening primers. It also allows fingerprinting results from different research groups to be compared and integrated. Pan *et al.* (2008) used 12 accessions with significantly different phenotypes and genetic backgrounds as the panel germplasm for screening 5,914 pairs of SSR primers, and the results

showed that 319 pairs of primers exhibiting suitable amplification and clear bands could be regarded as core primers for determining fingerprints. Moreover, they recommended 13 SSR primer pairs expressing polymorphism in Upland cotton, Sea Island cotton and Asiatic cotton, which could be regarded as the first-choice markers for DNA fingerprinting and germplasm identification. Zhao *et al.* (2012) selected 12 cotton cultivars derived from different pedigrees and different cotton-growing areas to identify SSR markers exhibiting high levels of polymorphism. Twenty-six SSR primer pairs were tagged onto the corresponding 26 chromosomes of cultivated tetraploid cotton species, and were recommended as the first-selected primer pairs to establish DNA barcodes of cotton cultivars, while the other 25 SSR primers could be used as candidate primers. These core primers have laid the foundation for the establishment of the standard DNA fingerprint database of Upland cotton. However, the markers used in the above studies were all obtained from cotton genome database, and are not directly related to cotton agronomic traits, so the efficiency to identify and distinguish cultivars is not high. In particular, owing to the application of transgenic technology on cotton breeding in recent years, some new transgenic cultivars differ from their parents only with respect to small genomic segments or a few traits. Therefore it is very difficult to accurately distinguish the accessions which have close relationship using limited molecular markers unrelated to phenotypic traits. In this study, the selection of core primers depends on both the level of marker polymorphism and the relationships between markers and agronomic traits. Although the average alleles (3.12), Di (0.4312) and PIC (0.3830) of the 26 core primers were lower than those reported by Lacape *et al.* (2007) (an average alleles of 5.6 and PIC of 0.55) and Moiana *et al.* (2012) (an average alleles of 6.9 and PIC of 0.646), they were similar to those reported by Bertini *et al.* (2006) (an average alleles of 2.13 and PIC of 0.40), Fang *et al.* (2013) (an average alleles of 2.64 and PIC of 0.2869) and Zhao *et al.* (2015) (an average alleles of 2.26, Di of 0.3502, and PIC of 0.2857), and above all, these markers were all associated/linked with at least agronomic traits of cotton. Therefore, the 26 core primers identified in this study are eminently suitable for DNA fingerprinting in Upland cotton. This is exactly the innovation of this research.

Molecular genetic studies suggest that genetic correlations among different quantitative traits may be a result of gene interaction or pleiotropy (Lehner 2011). Based on the above reasons, molecular markers associated with cotton agronomic traits including yield, fiber quality, early maturity, and plant architecture (Abdurakhmonov *et al.* 2009, Cai *et al.* 2014, Li *et al.* 2016a, 2016b, 2016c, 2017, Mei *et al.* 2013, Shao *et al.* 2014, Sun *et al.* 2012) were selected as core primers to conduct DNA fingerprinting in this study. Because the DNA differences corresponded to phenotypic differences, and also because 11 out of the 26 core primers were ESTs (expressed sequence tags, **Table 1**) corresponding to parts of a specific expressed gene sequence, the abili-

ty to distinguish different cultivars, especially those with only small phenotypic differences, was more effective. There was a significant positive correlation between the genetic distance matrix calculated by the 26 core primers and that by all 335 pairs of primers, further suggesting that these core primers for were eminently suitable for DNA fingerprinting in Upland cotton.

In recent years, the number of new cotton cultivars has been increasing year by year. As the number of cotton cultivars is not fixed and new cultivars bred by organizations and companies constantly enrich the Germplasm Bank, DNA fingerprint data have also been growing with the input of fingerprint data from the new cultivars. At the same time, with the continuing growth of the cultivar pool, the identification efficiency of existing core primers will be affected; therefore, the DNA fingerprint identity data of cultivars should also be extensible. When the existing core primers cannot meet the needs of cultivar identification, new effective core primer data can be added to the DNA fingerprinting database. Presently, the genome of the Upland cotton has been sequenced (Li *et al.* 2015, Zhang *et al.* 2015), and a total of 77,996 SSR markers were identified based on the sequence information (Wang *et al.* 2015). Further, a genome-wide single-nucleotide polymorphism (SNP) chip of Upland cotton, NAUSNP80K array, was successfully developed based on sequencing of 'TM-1' and re-sequencing of 100 different cultivars in Upland cotton with 5× coverage on average (Cai *et al.* 2017), which provided an important guarantee for exploring SNP markers related to cotton important agronomic traits. In the research of Wei *et al.* (2017), a total of 140 polymorphic markers including 46 SSRs, 47 SNPs and 47 InDels were recommended as a core set of molecular markers to establish sesame cultivars' fingerprinting. Especially, 9 SSRs 15 SNPs and 14 InDels were sufficient to distinguish all sesame cultivars. Therefore, further research needs to be focused on developing more SSR markers, or even SNP and InDel markers, associated with specific cotton agronomic traits, which would greatly ensure the accuracy, timeliness and practicality of a cotton DNA fingerprinting database, and its applicability to cotton improvement.

Literature Cited

Abdurakhmonov, I.Y., S. Saha, J.N. Jenkins, Z.T. Buriev, S.E. Shermatov, B.E. Scheffler, A.E. Pepper, J.Z. Yu, R.J. Kohel and A. Abdukarimov (2009) Linkage disequilibrium based association

mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. Genetica 136: 401–417.

Bertini, C.H.C.M., I. Schuster, T. Sediyama, E.G. Barros and M.A. Moreira (2006) Characterization and genetic diversity analysis of cotton cultivars using microsatellites. Genet. Mol. Biol. 2: 321–329.

Cai, C., G. Zhu, T. Zhang and W. Guo (2017) High-density 80 K SNP array is a powerful tool for genotyping *G. hirsutum* accessions and genome analysis. BMC Genomics 18: 654.

Cai, C.P., W.X. Ye, T.Z. Zhang and W.Z. Guo (2014) Association analysis of fiber quality traits and exploration of elite alleles in Upland cotton cultivars/accessions (*Gossypium hirsutum* L.). J. Integr. Plant Biol. 56: 51–62.

Chen, B.Y., X.M. Wu, D.X. Zhang, F.L. Liu, G.Y. Lu, K. Xu and G.Z. Gao (2008) SSR marker fingerprinting of winter rapeseed varieties in national field trails. Mol. Plant Breed. 6: 709–716.

Chen, G. and X.M. Du (2006) Genetic diversity of basal germplasm phenotypes in Upland cotton in China. Acta Bot. Boreal.-Occident. Sin. 26: 1649–1656.

Chen, Z.J., B.E. Scheffler, E. Dennis, B.A. Triplett, T.Z. Zhang, W.Z. Guo, X.Y. Chen, D.M. Stelly, P.D. Rabinowicz, C.D. Town *et al.* (2007) Toward sequencing cotton (*Gossypium*) genomes. Plant Physiol.145: 1303–1310.

Dhaliwal, M.S., M. Singh, K. Singh and D.S. Cheema (2011) Genetic diversity analysis and DNA fingerprinting of elite genetic stocks of tomato using SSR markers. Indian J. Genet. Plant Breed. 71: 341–348.

Fang, D.D., L.L. Hinze, R.G. Percy, P. Li, D. Deng and G. Thyssen (2013) A microsatellite-based genome-wide analysis of genetic diversity and linkage disequilibrium in Upland cotton (*Gossypium hirsutum* L.) cultivars from major cotton-growing countries. Euphytica 191: 391–401.

Gao, Y.L., R.S. Zhu, C.Y. Liu, W.F. Li, H.W. Jiang, C.D. Li, B.C. Yao, G.H. Hu and Q.S. Chen (2009) Establishment of molecular ID in soybean varieties in Heilongjiang, China. Acta Agron. Sin. 35: 211–218.

Guo, W.Z., T.Z. Zhang, J.J. Pan and J. He (1996) Analysis of RAPD fingerprinting on main cotton cultivars in China. Chin. J. Agric. Biotechnol. 4: 129–134.

Guo, W.Z., K. Wang and T.Z. Zhang (2003) A and D genome evolution in *Gossypium* revealed using SSR molecular markers. Acta Genet. Sin. 30: 183–188.

Guo, W.Z., C.P. Cai, C.B. Wang, Z.G. Han, X.L. Song, K. Wang, X.W. Niu, C. Wang, K.Y. Lu, B. Shi *et al.* (2007) A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in *Gossypium*. Genetics 176: 527–541.

Iqbal, M.J., O.U.K. Reddy, K.M. El-Zik and A.E. Pepper (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.

Kamvar, Z.N., J.F. Tabima and N.J. Grünwald (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2: e281.

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

Kashi, Y., D. King and M. Soller (1997) Simple sequence repeats as a source of quantitative genetic variation. Trends Genet. 13: 74–78.

Khan, S.A., D. Hussain, E. Askari, J.M. Stewart, K.A. Malik and Y. Zafar (2000) Molecular phylogeny of *Gossypium* species by DNA fingerprinting. Theor. Appl. Genet. 101: 931–938.

Kuang, M., W.H. Yang, H.X. Xu, Y.Q. Wang, D.Y. Zhou and X.A. Feng (2011) Construction of DNA fingerprinting and analysis of genetic diversity with SSR markers for cotton major cultivars in China. Sci. Agric. Sin. 44: 20–27.

Lacape, J.M., D. Dessauw, M. Rajab, J.L. Noyer and B. Hau (2007) Microsatellite diversity in tetraploid *Gossypium* germplasm: assembling a highly informative genotyping set of cotton SSRs. Mol. Breed. 19: 45–58.

Lehner, B. (2011) Molecular mechanisms of epistasis within and between genes. Trends Genet. 27: 323–331.

Li, C.Q., X.Y. Wang, X.F. Zhang and Q.L. Wang (2011) SSR fingerprinting establishment of Baimian series cotton inbred varieties (lines). Cotton Sci. 23: 228–234.

Li, C.Q., X.Y. Wang, N. Dong, H.H. Zhao, Z. Xia, R. Wang, R.L. Converse and Q.L. Wang (2013) QTL analysis for early-maturing traits in cotton using two upland cotton (*Gossypium hirsutum* L.) crosses. Breed. Sci. 63: 154–163.

Li, C.Q., H.P. Li, H.H. Zhao, N. Dong, Q.L. Wang and N.J. Ai (2014) SSR core primers reveal DNA fingerprinting of 30 Upland cotton (*Gossypium hirsutum* L.) varieties in China. Acta Bot. Bor-Occid. Sin. 34: 2210–2219.

Li, C.Q., N.J. Ai, Y.J. Zhu, Y.Q. Wang, X.D. Chen, F. Li, Q.Y. Hu and Q.L. Wang (2016a) Association mapping and favorable allele exploration for plant architecture traits in Upland cotton (*Gossypium hirsutum* L.) accessions. J. Agric. Sci. 154: 567–583.

Li, C.Q., X.J. Xu, N. Dong, N.J. Ai and Q.L. Wang (2016b) Association mapping identifies markers related to major early-maturing traits in upland cotton (*Gossypium hirsutum* L.). Plant Breed. 135: 483–491.

Li, C.Q., J.B. Zhang, G.H. Hu, Y.Z. Fu and Q.L. Wang (2016c) Association mapping and favorable allele mining for node of first fruiting/sympodial branch and its height in Upland cotton (*Gossypium hirsutum* L.). Euphytica 210: 57–68.

Li, C.Q., N. Dong, Y.Z. Fu, R.R. Sun and Q.L. Wang (2017) Marker detection and elite allele mining for yield traits in Upland cotton (*Gossypium hirsutum* L.) by association mapping. J. Agric. Sci. 155: 613–628.

Li, F.G., G.Y. Fan, C.R. Lu, G.H. Xiao, C.S. Zou, R.J. Kohel, Z.Y. Ma, H.H. Shang, X.F. Ma, J.Y. Wu *et al.* (2015) Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat. Biotechnol. 33: 524–530.

Li, G.Y., S. Dreisigacker, M.L. Warburton, X.C. Xia, Z.H. He and Q.X. Sun (2006) Development of a fingerprinting database and assembling an SSR reference kit for genetic diversity analysis of wheat. Acta Agron. Sin. 32: 1771–1778.

Li, X.H., X.H. Li, W.W. Gao, Q.Z. Tian, M.S. Li, F.M. Ma and S.H. Zhang (2005) Establishment of DNA fingerprinting database of maize hybrids and its application in parentage identification. Acta Agron. Sin. 31: 386–391.

Liu, X.L., L. Ma, X.K. Chen, X.M. Ying, Q. Cai, J.Y. Liu and C.W. Wu (2010) Establishment of DNA fingerprint ID in sugarcane cultivars in Yunnan, China. Acta Agron. Sin. 36: 202–210.

Mei, H.X., X.F. Zhu and T.Z. Zhang (2013) Favorable QTL alleles for yield and its components identified by association mapping in Chinese Upland cotton cultivars. PLoS ONE 8: e82193.

Mei, M., N.H. Syed, W. Gao, P.M. Thaxton, C.W. Smith, D.M. Stelly and Z.J. Chen (2004) Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). Theor. Appl. Genet. 108: 280–291.

Moiana, L.D., P.S.V. Filho, M.C. Goncalves-Vidigal, G.F. Lacanallo, M.Z. Galván, L.P. Carvalho, M.P. Maleia, C.M. Pacheco, T. Ribeiro,

H.Z. Neto *et al.* (2012) Genetic diversity and population structure of cotton (*Gossypium hirsutum* L. race latifolium H.) using microsatellite markers. Afr. J. Biotechnol. 11: 11640–11647.

Nguyen, T.B., M. Giband, P. Brottier, A.M. Risterucci and J.M. Lacape (2004) Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. Theor. Appl. Genet. 109: 167–175.

Oksanen, J., F.G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P.R. Minchin, R.B. O'Hara, G.L. Simpson, P. Solymos *et al.* (2017) Vegan: Community Ecology Package. R package version 2.4-4. https://CRAN.R-project.org/package=vegan

Pan, Z.E., J.L. Sun, X.W. Wang, Y.H. Jia, Z.L. Zhou, B.Y. Pang and X.M. Du (2008) Screening of SSR core primers with polymorphism on a cotton panel. Biodivers. Sci. 16: 555–561.

Pan, Z.E., X.W. Wang, J.L. Sun, Z.L. Zhou, Y.H. Jia, S.P. He, J. Wang, L.R. Wang, B.Y. Pang and X.M. Du (2010) Construction of digital fingerprint of Zhongmiansuo 48. Chin. Agric. Sci. Bull. 26: 31–35.

Paterson, A.H., C.L. Brubaker and J.F. Wendel (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.

Punitha, D. and T.S. Raveendran (2010) DNA fingerprinting studies in coloured cotton genotypes. Plant Breed. 123: 101–103.

Qin, H.D., W.Z. Guo, Y.M. Zhang and T.Z. Zhang (2008) QTL mapping of yield and fiber traits based on a four-way cross population in *Gossypium hirsutum* L. Theor. Appl. Genet. 117: 883–894.

Röder, M.S., V. Korzun, B.S. Gill and M.W. Ganal (1998) The physical mapping of microsatellite markers in wheat. Genome 41: 278–283.

Shao, Q.S., F.J. Zhang, S.Y. Tang, Y. Liu, X.M. Fang, D.X. Liu, D.J. Liu, J. Zhang, Z.H. Teng, A.H. Paterson *et al.* (2014) Identifying QTL for fiber quality traits with three Upland cotton (*Gossypium hirsutum* L.) populations. Euphytica 198: 43–58.

Shen, X.L., W.Z. Guo, Q.X. Lu, X.F. Zhu, Y.L. Yuan and T.Z. Zhang (2007) Genetic mapping of quantitative trait loci for fiber quality and yield trait by RIL approach in Upland cotton. Euphytica 155: 371–380.

Song, G.L., C.Q. Zhang, J.Z. Jia, K. Wang and R. Cui (1999) Cotton AFLP analysis with silver-staining and primary report of variety fingerprinting based on it. Cotton Sci. 11: 281–283.

Song, X.L. and T.Z. Zhang (2009) Quantitative trait loci controlling plant architectural traits in cotton. Plant Sci. 177: 317–323.

Stephens, S.G. and M.E. Mosley (1974) Early domesticated cottons from archaeological sites in central coastal Peru. Am. Antiquity 39: 109–122.

Sun, F.D., J.H. Zhang, S.F. Wang, W.K. Gong, Y.Z. Shi, A.Y. Liu, J.W. Li, J.W. Gong, H.H. Shang and Y.L. Yuan (2012) QTL mapping for fiber quality traits across multiple generations and environments in Upland cotton. Mol. Breed. 30: 569–582.

Talib, I., S. Ali, M.W. Pervez, M.U. Hassan, R. Hussain, K. Rehman, R. Firdous and S. Hussain (2015) Study of genetic diversity in germplasm of Upland cotton (*Gossypium hirsutum* L.) in Pakistan. Am. J. Plant Sci. 6: 2161–2167.

Van Esbroeck, G.A., D.T. Bowman, O.L. May and D.S. Calhoun (1999) Genetic similarity indices for ancestral cotton cultivars and their impact on genetic diversity estimates of modern cultivars. Crop Sci. 39: 323–328.

Wang, L.X., Y.F. Li, L.F. Chang, L. Huang, H.B. Li, L.L. Ge, L.H. Liu, J. Yao and C.P. Zhao (2007) Method of ID constitution for wheat cultivars. Acta Agron. Sin. 33: 1738–1740.

Wang, Q., L. Fang, J.D. Chen, Y. Hu, Z.F. Si, S. Wang, L.J. Chang, W.Z. Guo and T.Z. Zhang (2015) Genome-wide mining, characterization, and development of microsatellite markers in *Gossypium* species. Sci. Rep. 5: 10638.

Wang, X.D. and Y.Y. Li (2002) Development of cytoplasmic malesterile, maintainer and restorer lines in colored cotton and construction of their DNA fingerprints. J. Zhejiang Univ. 28: 1–6.

Wei, L.B., H.M. Miao, F.F. Xu, J.J. Kong and H.Y. Zhang (2017) Chinese sesame cultivars, DNA Fingerprinting, and two-dimensional barcodes using single nucleotide polymorphisms, insertions or deletions, and simple sequence repeat markers. Crop Sci. 57: 1941–1947.

Yan, J.W., D.G. Tian, Y. Xu and F. Wang (2011) Constructing SSR molecular database for identity system of hybrid rice parents. Fujian J. Agric. Sci. 26: 148–152.

Zhang, T.Z., Y.L. Yuan, J. Yu, W.Z. Guo and R.J. Kohel (2003) Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection. Theor. Appl. Genet. 106: 262–268.

Zhang, T.Z., Y. Hu, W.K. Jiang, L. Fang, X.Y. Guan, J.D. Chen, C.A. Saski, B.E. Scheffler, D.M. Stelly, A.M. Hulse-Kemp *et al.* (2015) Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. Nat. Biotechnol. 33: 531–537.

Zhao, L., C.P. Cai, H.X. Mei and W.Z. Guo (2012) Screening of microsatellite loci for identifying genome barcoding of cotton cultivars. Acta Agron. Sin. 38: 1810–1817.

Zhao, Y.L., H.M. Wang, W. Chen, Y.H. Li, H.Y. Gong, X.H. Sang, F.C. Huo and F.C. Zeng (2015) Genetic diversity and population structure of elite cotton (*Gossypium hirsutum* L.) germplasm revealed by SSR markers. Plant Syst. Evol. 301: 327–336.