Theor Appl Genet (2006) 112: 1573–1581
DOI 10.1007/s00122-006-0261-y

ORIGINAL PAPER

Wangzhen Guo · Wei Wang · Baoliang Zhou
Tianzhen Zhang

Cross-species transferability of *G. arboreum*-derived EST-SSRs in the diploid species of *Gossypium*

Received: 11 June 2005 / Accepted: 1 March 2006 / Published online: 5 April 2006

Abstract Diploid species with a common *Gossypium* origin are highly diverse in morphology and have been classified into eight genomic groups designated A–G and K. In this study, the transferability of 207 *Gossypium arboreum*-derived expressed sequence tag-simple sequence repeat (EST-SSR) primer pairs was examined among 25 different diploid accessions representing 7 genomes and 23 *Gossypium* species. We found that 124 of the 207 (60%) primer pairs produced amplification products in all 25 accessions. The remaining 83 (40%) primer pairs produced amplification in only a subset of species, ranging from 13 to 22 species, which is consistent with some genome- and species-specific amplification. The cross-species amplification of these EST-SSRs in 22 diploid species was 96.5% in 4,554 combinations (207 SSRs×22 species), indicative of a high transferability among the *Gossypium* species. Furthermore, a high level of polymorphism with an average number of 6.53 alleles per SSR marker was detected. No correlation was found between the repeat motif type and cross-species amplification. DNA sequencing showed that the high-level polymorphism findings was mainly due to changes in the number of repeat motifs and that the high transferability can be attributed to a higher-level conservation in the flanking regions among these diploid *Gossypium* species. The transferability among these different diploid species presented here can increase the efficiency of transferring genetic information across species and further enhance their introgression into cultivated cotton species by the molecular tagging of important genes existing in these diploid species using the EST-SSR markers.

Abbreviations EST: Expressed sequence tag · PCR: Polymerase chain reaction · dpa: Day post-anthesis · bp: Base pair · SSR: Simple sequence repeats

Introduction

Morphological (Fryxell et al. 1992), meiotic (Menzel 1954), karyotypic (Beasley 1940), and genetic and molecular methods (Wendel 1989) have been employed in the classification of cotton (*Gossypium* spp.), which dates back to nearly a century. Currently, the most widely accepted classification of *Gossypium* follows Fryxell's model (1979, 1992), which recognizes 50 species including 45 diploids and 5 tetraploids. Although a common origin for *Gossypium* has been dated to just 5–15 million years ago, the diploid species in *Gossypium* are highly diverse in morphology and are represented genetically by eight different genome types including the A, B, C, D, E, F, G, and K genomes (Endrizzi et al. 1985; Stewart 1995). The F-genome species *G. longicalyx* diverged from the E genome, and the G-genome species diverged from the initial C genome. The diploid species distributed in the Australian Kimberley area were classified as K genome because of their distinctive geographical location, morphology, and ecology (Fryxell et al. 1992). Except for the A genome, all of the above diploid genomes are composed of wild species. These wild species are important genetic resources and have played and will continue to play significant roles in cotton breeding and cultivation (Stewart 1995). Nevertheless, only a small portion of the valuable genes has been utilized through introgression into cultivated

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00122-006-0261-y and is accessible for authorized users.

Communicated by J.S. Heslop-Harrison

W. Guo · W. Wang · B. Zhou · T. Zhang (✉)
National Key Laboratory of Crop Genetics & Germplasm Enhancement, Cotton Research Institute,
Nanjing Agricultural University, 210095 Nanjing,
Jiangsu Province, People's Republic of China
E-mail: cotton@njau.edu.cn
Tel.: +86-25-84395307
Fax: +86-25-84395307

cotton as of yet (Endrizzi et al. 1985; Meredith 1991; Stewart 1995; Brubaker et al. 1999).

Microsatellites or simple sequence repeats (SSRs) are tandem repeats of short (1–6 bp) DNA sequences; SSRs are present throughout the genome of an organism, in both non-coding and coding regions. Expressed sequence tags (EST)-derived-SSRs (EST-SSRs) from the transcribed regions of the DNA are generally more conserved across species than those from the untranscribed regions (Caudrado and Schwarzacher 1998). The distinguishing features of SSR loci include their high information content, co-dominant inheritance, even distribution along chromosomes, reproducibility, and locus specificity (Kashi et al. 1997; Röder et al. 1998a, b). In recent years, great efforts have also been made to develop genome SSRs (Reddy et al. 2001; Nguyen et al. 2004) (http://www.resgen.com) and EST-SSRs (Saha et al. 2003; Qureshi et al. 2004; Han et al. 2004) for cotton. These SSR markers have been widely applied in genetic and genomic studies such as quantitative trait locus (QTL) tagging (Zhang et al. 2003), MAS (Shen et al. 2001; Guo et al. 2003a), genetic diversity analysis (Guo et al. 2003b), and the construction of linkage maps (Zhang et al. 2002; Nguyen et al. 2004; Song et al. 2005). EST-derived polymerase chain reaction (PCR)-based markers for functional gene homologues in cotton have been reported previously (Chee et al. 2004). An especially promising attribute of EST-SSRs is their ability to differentiate among different genotypes within the cotton genus. Indeed genotype differentiation studies with the aim of reconstructing the species' genetic history have been conducted with the human genome as well as with the genomes of some plants (di Rienzo et al. 1994; Chen et al. 2002; Rossetto et al. 2002).

Based on the publicly available EST sequences generated from a 7–10 day post-anthesis (dpa) fiber cDNA library of diploid A-genome *G. arboreum* (http://www.genome.clemson.edu/projects/cotton), 544 EST-SSRs from a total of 931 ESTs or cDNAs containing SSRs were developed in our laboratory. Among them, 103 ESTs were mapped on the corresponding chromosomes based on our backbone genetic map constructed using an inter-specific BC_1 mapping population [(TM-1×Hai7124)×TM-1] (Han et al. 2004). In the present research, 207 out of the 544 *G. arboreum*-derived EST-SSRs were randomly chosen to study the cross-species transferability for the first time in *Gossypium*. The cross-species transferability of these EST-SSRs from the A-genome *G. arboreum* to 25 different diploid accessions representing 7 genomes (A-G) and 23 species, the levels of polymorphisms across species, and the genome specificity of amplification are presented.

Table 1 The diploid species in *Gossypium* used in the present research

Species name	Distribution	Genomic group
G. herbaceum L. cv. Hongxin	Old world cultigen	A_1
G. herbaceum var. *africanum*	Africa	A_1
G. arboreum L. cv. Jinhua	Old world cultigen	A_2
G. arboreum L. cv. Shixiya	Old world cultigen	A_2
G. capitis-viridis Mauer	Africa, Cape Verde Islands	B_3
G. sturtianum J.H.Willis	Australia	C_1
G. robinsonii F. Muell.	Australia	C_2
G. thurberi Tod.	Mexico, Arizona	D_1
G. armourianum Kearn.	Galapagos Islands, Mexico	$D_{2\text{-}1}$
G. harknessii Brandg.	Mexico	$D_{2\text{-}2}$
G. klotzschianum Anderss.	Galapagos Islands	$D_{3\text{-}k}$
G. davidsonii Kell.	Mexico	$D_{3\text{-}d}$
G. aridum (Rose & Standl.) Skov.	Mexico	D_4
G. raimondii Ulbr.	Peru	D_5
G. gossypioides (Ulbr.) Standl.	Mexico	D_6
G. lobatum Gentry	Mexico	D_7
G. trilobum (Moc. & Sess. ex DC.) Skov. emend. Kearn.	Mexico	D_8
G. laxum Phillips	Mexico	D_9
G. stocksii Mast. ex Hook.	Arabian Peninsula	E_1
G. somalense (Gurke) Hutch.	Northeast Africa	E_2
G. incanum (Schwartz) Hillc.	Arabian Peninsula	E_4
G. longicalyx Hutch. & Lee	East Africa	F_1
G. bickii Prokh.	Australia	G_1
G. nelsonii Fryx.	Australia	G
G. australe F. Muell.	Australia	G

Genome designations follow Stewart (1995); taxonomy and distribution follow Endrizzi et al. (1985)

Materials and methods

Plant materials and DNA extraction

A total of 25 accessions including 7 genomes and 23 species—2 A-, 11 D-, 3 E-, 2 C-, 3 G-, 1 B-, and 1 F-genome—in *Gossypium* were collected. Their geographic distribution and genomic groups are given in Table 1. One *G. herbaceum* and two *G. arboreum* species were collected in the Jiangpu Breeding Station, Nanjing Agricultural University (NAU). All other wild species were collected from the National Wild Cotton Plantation in Hainan Island, China. DNA was isolated from the plants' young leaves using the cetyltrimethylammonium bromide (CTAB) method (Paterson et al. 1993).

EST-SSR primer pairs derived from *G. arboreum*

The cross-species transferability to 22 species representing genome types A–G was assessed for 207 randomly selected *G. arboreum*-derived EST-SSR primer pairs which had previously been developed in our laboratory. These 207 EST-SSRs contained 2–6 nucleotide repeat motifs. Among them, 75 had been mapped onto 24 corresponding chromosomes/linkage groups based on our previously described backbone genetic map for tetraploid cotton (Han et al. 2004). All data for these SSRs including their primer sequences, putative

functions, GenBank accession numbers, and mapping loci in tetraploid cotton have been publicly released at http://www.njau.edu.cn, http://cottondb.tamu.edu and http://www.mainlab.clemson.edu/cmd/projects/nau.

PCR amplification

Polymerase chain reaction amplification was carried out in four 96-well plates placed in a PTC 225 thermocycler (MJ Research, Watertown, MA, USA), according to previously described reaction parameters (Zhang et al. 2003). The high fidelity ExTaq DNA polymerase [Takara Biotech (Dalian)] was used in the PCR amplification. PCR products were size-separated on 6% polyacrylamide gels, visualized via silver staining according to our protocols (Zhang et al. 2000), and photographed using an electrophoresis image analysis system. The consistency of the amplification was verified by running at least two independent PCR amplifications.

Data scoring and statistical analysis

Gossypium arboreum cv. Jinhua and Shixiya were chosen to test amplification, allele average size, and banding morphology (cluster of stutter). The size of the most intensely amplified band for each SSR locus, or the average of the stutter if the intensity was the same, was scored using a 50-bp DNA marker [Takara Biotech (Dalian)] as the standard. Null alleles were assigned to genotypes in which no amplification product was produced under the standard conditions. The number of amplified fragments in each species was counted regardless of polymorphism. Amplification was quantified according to the following equation:

$$\text{Amplification}\,(\%) = \frac{\text{numbers of amplified} \times 100}{\text{total combinations of}\,(\text{SSRs} \times \text{tested cotton species})} \quad (1)$$

The percentage of cross-species transferability was calculated as the percentage of amplification of the SSR markers amplified in 22 species excluding the *G. arboreum* species (Kuleung et al. 2004). The correlation between repeat motif and cross-species transferability was completed with the SAS software (SAS 1989).

DNA sequence analysis

To confirm the polymorphic amplification of the expected *G. arboreum* microsatellites among the different tested diploid species, three SSR loci having tri-, tetra-, and hexa-repeat motifs, respectively, were recovered and sequenced from different accessions. Both NAU980 and NAU905 were found to be located within the 3′UTR and NAU986 was found to be located within the 5′UTR and partial coding region.

The amplified fragments were excised from polyacrylamide gels. DNA recovery, vector ligation, and transformation of DH5α *Escherichia coli* competent cells were conducted as described previously (Guo et al. 2003a). At least three positive clones from each amplified fragment were simultaneously sequenced by Invitrogen Biotech (Shanghai), China.

Sequences were edited using Editseq in the DNA Star package (http://www.dnastar.com) after the vector sequences were removed. The DNA sequences were then aligned using the MegAlign program within DNA Star and adjusted manually to obtain multiple gaps within the repeat region. The consensus sequence was used as the reference in sequence comparisons and base mutation analyses. The rate of base substitution in the flanking regions was calculated following the method described by Chen et al. (2002).

Results

Cross-species amplification of EST-SSR primer pairs derived from *G. arboreum* across 23 diploid species in *Gossypium*

All 207 *G. arboreum*-derived EST-SSR primer pairs yielded microsatellite products in type A-genome cultivars *G. arboreum* cv. Jinhua and Shixiya, as well as in *G. herbaceum* cv. Hongxin and var. *africanum*. Of these 207 SSRs, 124 (60%) amplified products from all 25 accessions, composed of 23 cotton species including *G. arboreum* itself (Fig. 1a, b). Meanwhile 83 (40%) amplified products from only a subset of the accessions; these subsets ranged from 13 to 22 species. The 124 SSRs with the ability to amplify products of all of the tested species had different repeat motifs with differing cross-species transferability such that the tetra-nucleotide repeat motif SSR had the highest transferring rate (78.6%) and was followed by the tri- (61.5%), hexa- (60.5%), di- (53.3%), and penta-nucleotide repeats (42.8%). However, there was no correlation ($r = -0.05188$) between the repeat motif type and cross-species transferability. When all combinations (207 SSRs×22 species = 4,554) were examined, as high as 96.5% cross-species transferability was observed. Thus it is clear that these *G. arboreum*-derived EST-SSRs have high transferability across the diploid species in *Gossypium*.

Different genomes had different transferability. The B-, C-, E-, F-, and G-type genome species had higher cross-species transferabilities than that of the D-genome

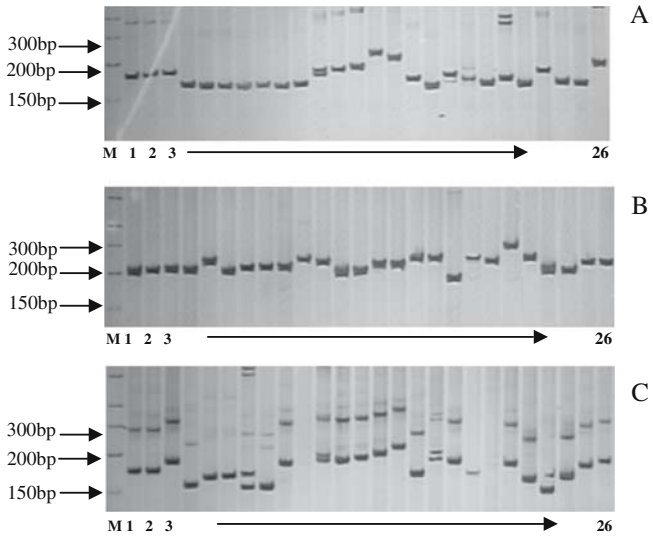

Fig. 1 Polyacrylamide gel electrophoresis patterns of microsatellite alleles amplified using three EST-SSR markers. **a** and **b** show that PCR products could be amplified from all 25 accessions by NAU980 (**a**) and NAU986 (**b**), and **c** shows that null alleles were detected in *Gossypium thurberi* and *G. trilobum* by NAU905 (**c**). *M* shows the molecular weight marker, and 1–26 are represented for *G. herbaceum* cv. Hongxin (*1*); *G. herbaceum* var. *africanum* (*2*); *G. arboreum* cv. Jinhua (*3*); *G. capitis-viridis* (*4*); *G. nelsonii* (*5*); *G. australe* (*6*); *G. robinsonii* (*7*); *G. sturtianum* (*8*); *G. bickii* (*9*); *G. thurberi* (*10*); *G. armourianum* (*11*); *G. harknessii* (*12*); *G. harknessii* (*13*); *G. klotzschianum* (*14*); *G. davidsonii* (*15*); *G. aridum* (*16*); *G. raimondii* (*17*); *G. gossypioides* (*18*); *G. lobatum* (*19*); *G. trilobum* (*20*); *G. laxum* (*21*); *G. stocksii* (*22*); *G. somalense* (*23*); *G. incanum* (*24*); *G. longicalyx* (*25*), and *G. arboreum* cv. Shixiya (*26*), respectively

Table 2 Cross-species transferability of *G. arboreum*-derived EST-SSRs among different genomes in *Gossypium*

Genome	No. of SSRs amplified wholly	Percentage amplified wholly (%)	No. of SSRs amplified partially	No. null amplified
A	207	100	0	0
B	195	94.2	0	12
C	199	96.1	6	2
D	156	75.4	49	2
E	186	89.9	16	5
F	196	94.7	0	11
G	196	94.7	8	3

Whole amplification means that all SSR primer pairs could produce PCR products in a given genome; partial amplification means that there were no amplicons in some species in a given genome; and null amplification means that no amplicon of expected size was produced in any species of a given genome

species (Table 2). This analysis revealed that the A-genome species were closely related with the B-, C-, F-, G-, and E-genome species.

Genome- and species-specific amplification of *G. arboreum*-derived EST-SSR

Of the 83 EST-SSRs that did not produce cross-species amplification in all 25 of the accessions surveyed, most (69.9%) produced amplicons from the diploid species. Interestingly, some of the SSRs were found to be genome-specific (Table 3). For example, the SSR NAU955 produced the PCR product in only A- and D-type genome species, indicating that it may be considered as an A- and D-genome-specific SSR marker. Additionally, five SSRs (NAU908, NAU945, NAU955, NAU1169, and NAU1362) did not produce any product in all 3 E-genome species; and two of these (NAU1223 and NAU1360) did not produce any product in any of the 11 D-genome species. Thus, these SSRs might be considered as negative E- and D-genome SSR markers. Furthermore, the numerous negative B- and F-genome-specific SSRs observed (Table 3) might be used as dominant markers in identifying DNA fragments introgressed into other species in *Gossypium*.

NAU905 was active at amplicons in 9 of the 11 D-genome species examined; *G. thurberi* and *G. trilobum* were the two species that did not produce amplified products with NAU905 (Fig. 1c, Table 3). So, this SSR marker may have species specificity. Amplification failures with species specificity were also detected in other D-genome species by 49 EST-SSR primer pairs. Higher frequencies of amplification failures were found with *G. lobatum* (59.2%) and *G. raimondii* (38.8%) than with the other D-genome species. These findings suggest that the *G. lobatum* and *G. raimondii* species may have diverged considerably from the other D-genome species during their evolution.

As to the five Australian diploid species, both NAU1336 and NAU1362 could produce amplification in the two C-genome species, *G. sturtianum* and *G. robinsonii*, but not in the 3 G-genome species, *G. nelsonii*, *G. australe*, and *G. bickii*. Thus NAU1336 and NAU1362 may serve as markers to distinguish C- and G-genome species in Australian diploid species. It should be noted that there existed great differences in amplification between *G. bickii* and the other two G-genome species. Among the 3 G-genome species, NAU887, NAU908 and NAU1005 could produce an SSR band from *G. bickii*, but not from *G. nelsonii* and *G. australe*. Meanwhile NAU877 and NAU1169 could produce an SSR band from *G. nelsonii* and *G. australe*, but not from *G. bickii* (Table 3). These findings provide strong evidence indicating that *G. nelsonii* and *G. australe* have greater DNA sequence similarity with each other than either has with *G. bickii*.

Polymorphism and DNA sequencing of *G. arboreum*-derived EST-SSRs in different diploid species in *Gossypium*

A total of 1,352 alleles was detected in the amplification experiments with 25 accessions using 207 EST-SSRs. Only NAU1044 and NAU1055 produced a monomorphic band and allele polymorphism was as high as 99%. In the polymorphic locus which gave amplified fragments of expected size in *G. arboreum*, most of the loci were

Table 3 Amplification features of genome-specific and partial species-specific EST-SSR primer pairs

Primers	Genome						
	A	B	C	D	E	F	G
NAU869	√	√	√	★ (null in *G. lobatum*)	√	×	√
NAU877	√	√	√	√	√	√	★ (null in *G. bickii*)
NAU887	√	√	√	√	√	√	★ (null in *G. australe* and *G. nelsonii*)
NAU901	√	×	√	√	√	√	√
NAU905	√	√	√	★ (null in *G. thurberi* and *G. trilobum*)			
NAU907	√	×	√	√	√	√	√
NAU908	√	√	√	√	×	√	★ (null in *G. australe* and *G. nelsonii*)
NAU915	√	√	√	√	√	×	√
NAU921	√	√	√	√	√	×	√
NAU945	√	√	√	★ (null in *G. aridum* and *G. lobatum*)	×	√	√
NAU955	√	×	×	√	×	×	×
NAU962	√	√	√	√	√	×	√
NAU963	√	√	√	√	√	×	√
NAU967	√	√	√	★ (null in *G. lobatum*)	★ (null in *G. incanum*)	×	√
NAU971	√	√	√	√	★ (null in *G. incanum*)	×	√
NAU973	√	√	√	√	√	×	√
NAU977	√	×	√	★ (null in *G. lobatum*)	√	√	√
NAU982	√	×	√	√	√	√	√
NAU987	√	√	√	★ (null in *G. lobatum*)	√	×	★ (null in *G. australe*)
NAU1005	√	√	√	√	√	√	★ (null in *G. australe* and *G. nelsonii*)
NAU1008	√	×	√	√	√	√	√
NAU1025	√	×	√	√	√	√	√
NAU1052	√	×	★ (null in *G. sturtianum*)	★ (null in *G. lobatum* and *G. laxum*)	★ (null in *G. incanum*)	√	★ (null in *G. nelsonii*)
NAU1127	√	×	√	★ (null in *G. lobatum*)	√	√	√
NAU1169	√	×	×	★ (null in *G. lobatum*)	×	×	★ (null in *G. bickii*)
NAU1223	√	√	√	×	√	√	√
NAU1336	√	×	√	√	√	√	×
NAU1360	√	√	√	×	√	√	√
NAU1362	√	×	√	★ (null in *G. raimondii*)	×	√	×

"×" means no amplified products in a given genome, "√" indicates that there were amplification products in a given genome, and "★" means that no amplicon was produced from one and/or two species included in brackets in this genome

characterized by a single SSR band, whereas some had two SSR bands such as the NAU980 primers in *G. armourianum*, and the NAU905 primers in *G. robinsonii* and *G. raimondii* (Fig. 1). The average number of alleles per SSR marker was 6.53 with a range of 1–10 alleles based on the dominant scoring of the SSR bands characterized by the presence or absence of a particular band. The greatest variation of SSR alleles was found for NAU883, which interacted with 10 alleles in 25 accessions.

In order to study how DNA polymorphisms are produced, the amplified products from NAU905, NAU980, and NAU986 were sequenced. Both NAU980 and NAU986 amplified SSR products from all 25 accessions, but NAU905 could not amplify SSR products from the *G. thurberi* and *G. trilobum* species (Fig. 1). The different allele lengths, repeat types, and repeat numbers are shown as Supplementary material. Amplicon sequence alignment showed that although a great deal of variation exists in the repeat types and lengths, all the primer binding regions are highly conserved. Allelic diversity could be attributed mainly to differences in repeat type and length in the microsatellite regions. Variations in the flanking regions due mainly to base substitution and/or indel mutations were also observed. Both base indels and substitutions were found within NAU980 and NAU986 loci, but only base substitutions were identified within NAU905. There were 0.99, 0.95, and 2.12% base substitutions detected in the flanking regions of NAU980, NAU986, and NAU905 loci, respectively. There was one base-deletion in *G. capitis-viridis* and *G. raimondii*, two base-deletions in *G. bickii*, and one base-insertion in *G. lobatum* and three E-genome cotton species in NAU980 loci. Additionally, one base-deletion in *G. capitis-viridis, G. raimondii*, and *G. stockii*, five base-deletions in *G. nelsonii* and *G. australe* species, and one 23 base-deletion in *G. gossypioides* were found in the NAU986 loci. And in the NAU905 loci, variations in the repeat regions involved changes in the repeat number or a combination of repeat number changes

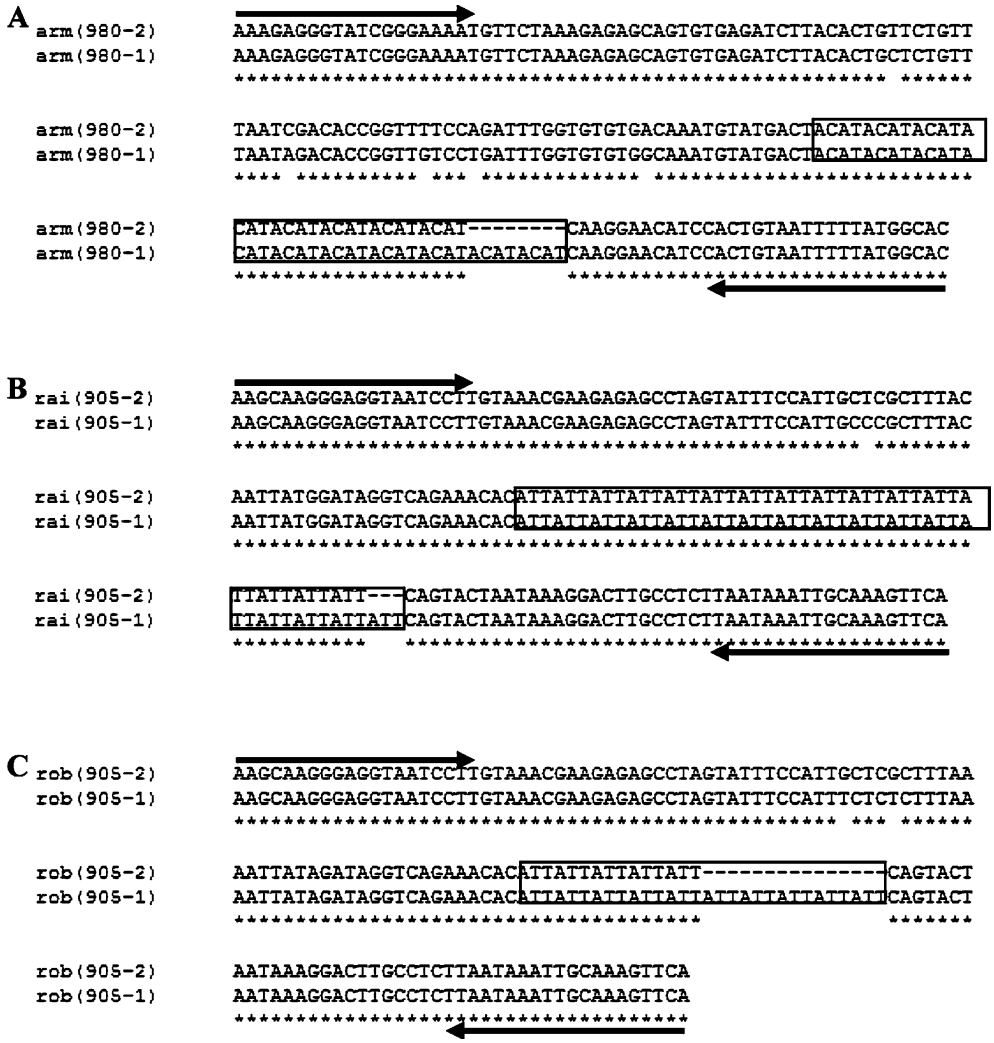

Fig. 2 CLUSTAL alignment of sequences obtained from two SSR bands amplified by NAU980 in *G. armourianum* (arm, **a**); NAU905 in *G. raimondii* (rai, **b**), and *G. robinsonii* (rob, **c**). Alignments include primer sequences (*arrows*). Repetitive sequences are indicated in *box*. Conserved nucleotide positions are marked by *asterisks*. The suffix 1 has bigger molecular weight than the suffix 2

and insertions of additional types of repeat motif sequences.

Two SSR bands of expected size were amplified in *G. armourianum* by NAU980, in *G. robinsonii* and *G. raimondii* by NAU905. In addition to repeat motif number variation, some base substitutions—such as five in *G. armourianum* (Fig. 2), two in *G. robinsonii*, and one in *G. raimondii*—were revealed in their flanking regions by sequence alignment of the two SSR bands. This finding indicates that there was duplication in the SSR flanking sequences in some species and that some cross-species introgression has occurred in these species. Our sequence analysis indicated that the high-level transferability of *G. arboreum*-derived EST-SSRs in the 25 tested accessions was largely attributable to a higher-level conservation of the transcriber DNA SSR flanking sequences among these diploid species in *Gossypium*.

Discussion

EST-SSR transferability in diploid cotton species

The transferability with genomic SSR markers is generally low when cross-species analyses are conducted, while polymorphisms are numerous. Meanwhile EST-SSR markers derived from transcribed regions of the DNA produce a higher rate of transferability, but fewer polymorphisms. Based on wheat genomic SSR markers, the transferability from wheat to rye was found to be only 17% (Kuleung et al. 2004); however, based on EST-SSR markers, the transferability from wheat to 18 *Triticum-Aegilops* species was found to be as high as 84% (Bandopadhyay et al. 2004), and from Tall fescue (*Festuca arundinacea* Schreb.) to seven grass species was found to be nearly 92% (Saha et al. 2004).

Using 20 EST-SSRs and 22 genome-derived SSRs for genotyping the A- and B-type genomes of wheat, the EST-SSRs produced a 25% polymorphism rate, whereas the genome-derived SSR markers produced a 53% polymorphism rate (Eujayl et al. 2002). Furthermore, polymorphism frequency may be related to the mode of plant reproduction. It is low in self-pollinated species such as rice (43%) and wheat (38%), but is high in cross-pollinated species such as tall fescue and ryegrass (66%) using the same markers (Saha et al. 2004). Similar results were also reported in cross-species studies examining *Medicago* spp. and the subspecies of legume crops (70%) (Eujayl et al. 2004), as well as a study examining the *Triticum-Aegilops* species (88%) (Bandopadhyay et al. 2004).

In the present study, high cross-species amplification (60%), transferability (96.5%), and polymorphism (99%) were detected in the diploid species in *Gossypium* using 207 *G. arboreum*-derived EST-SSR markers, which indicates that these EST-SSRs derived from ESTs or cDNAs cloned from a 7–10 dpa fiber cDNA library of *G. arboreum* had higher conservation with other genome species. SSR repeat number changes in protein-coding, 5′UTR, and 3′UTR regions may lead to a gain or loss of gene function or may disrupt other cellular functions (for review see Li et al. 2004). The high cross-species transferability of EST-SSRs from the fiber development stage in this paper might imply that there are some genes that are related to fiber quality in other genomes. Indeed there are a wide range of *Gossypium* seed covering types, which include some that are nearly glabrous to the naked eye (e.g. *G. klotzschianum* and *G. davidsonii*), other that are short stiff, dense, brown hairs (*G. australe* and *G. nelsonii*), and still others that have long white fibers that are considered highly improved forms of the two Asiatic cultivated species in these 25 diploid accessions. This variability may also be the reason that these *G. arboreum*-derived EST-SSRs that mapped not only in At-, but also in Dt-subgenomes of tetraploid cotton (Han et al. 2004), have high cross-species transferability, and many QTLs for fiber related traits (Jiang et al. 1998; Kohel et al. 2001) that have been detected in the Dt-subgenome of tetraploid cotton. Sequence alignment analysis of amplified fragments from NAU980 and NAU986 loci revealed that the primer binding regions were highly conserved and that microsatellite repeats were present at all loci within all accessions, but that the repeat number varied. Different fiber phenotypes might be the manifestation of complex mutational events involving insertion or deletion in SSR regions with high conservation in primer binding regions of the SSR flanking sequence. It is reasonable to suppose that these variations of EST-SSR regions may have regulatory functions during fiber development or may even be critical genetic factors in the fiber development process. In fact, some SSRs having various putative functions have been mapped. For example, NAU1272 was associated with sucrose synthase (Nolte et al 1995; Ruan et al. 2003) and NAU1221 was associated with fiber protein E6 (John 1996), which is preferentially or specifically expressed during primary and secondary fiber cell wall formation. In addition, NAU1215 was associated with an MYB family transcription factor, which might be one of a large group of transcription factors involved in plant development and tissue differentiation (Karpinska et al 2004). Functional genomics of cell elongation in developing cotton fibers including all the EST-SSRs used in this study were analyzed systematically by microarray technique (Arpat et al. 2004). Those data will be helpful for establishing the cross-genome homologies across the entire *Gossypium* genus not only in molecular tagging of target genes and identification of candidate genes, but also in the studies of genome evolution among these diploid species.

Genetic relationship in diploid cotton species

Eight different genome types genetically represent all of the diploid species with a common origin in the *Gossypium* genus. These eight genome type groups are comprised of four major lineages which encompass three continents: Australia (C-, G-, and K-genome species), America (D-genome species), and Africa/Arabia (1st lineage of the A-, B-, and F-genome species, and 2nd lineage of the E-genome species) (Fryxell 1979, 1992). Extensive molecular biological evidence from nuclear and/or chloroplast genome research is consistent with this classification (Wendel and Albert 1992; Cronn et al. 1996; Seelanan et al. 1997, 1999; Liu et al. 2001; Cronn et al. 2002).

In this report, a high level of polymorphism (99%) was observed in all 25 diploid accessions. Furthermore, using Dice (1945) similarity index and the NTSYS-pc package (Rohlf 1998), our cluster analysis based on the DNA data also indicated that these 25 accessions could be classified as six groups. Note that this conclusion is highly congruent with previous genome designations and geographical distribution (data not shown). Meanwhile, the high transferability of the EST-SSRs among the 25 tested accessions provides evidence that the different diploid species evolved from one common ancestor despite their subsequent distinctly different and monophyletic evolutionary paths produced by geographical isolation and variations in ecological conditions (Fryxell 1992).

Our transferability, polymorphism, and sequence analysis data indicated that the G-genome species *G. bickii* had more genetic similarity with the C-genome species *G. robinsonii* and *G. sturtianum* than with the G-genome species *G. nelsonii* and *G. australe*. This finding was somewhat contradictory with previous reports suggesting that *G. bickii* shares a more recent common ancestor with its close morphological allies, *G. nelsonii* and *G. australe* (Fryxell 1992; Liu et al. 2001). This discrepancy may be due to natural inter-specific hybridization which has been reported for *G. aridum* (Wendel and Albert 1992), *G. gossypioides* (Zhao et al.

1998), *G. bickii* (Liu et al. 2001), and the B-genome species (Cronn et al. 2002). To be prudent, the relationships among the five Australian diploid species remain to be further studied using different methods in future research.

There have been many reports of studies examining cross-species transferability in other plants using EST-SSR markers (Bandopadhyay et al. 2004; Eujayl et al. 2004; Saha et al. 2004). This is the first report to assess the inter-species relationships in diploid cotton species in *Gossypium*. It is clear that EST-SSR markers are very useful in characterizing species relationships and in the introgression and search for desirable alleles from wild germplasm pools in *Gossypium* that may be employed due to their high cross-species transferability. In the future, our line of research will focus on (1) large-scale development of new EST-SSR markers, (2) detection of the gene order across these diploid species, (3) molecular tagging of target genes from these diploid species, and (4) introgression of the identified genes into cultivated cotton species by MAS using these transferable markers.

Acknowledgments The study was financially supported in part by grants from the National Natural Science Foundation of China (30270806, 30471104), the State Key Basic Research and Development Plan of China (2002CB111301), the Program for Changjiang Scholars and Innovative Research Team in University and the Natural Science Foundation of Jiangsu Province (BK2003414). Thanks to M. Xia and S. Z. Ke for their help in formatting the original data.

References

Arpat AB, Waugh M, Sullivan JP, Gonzales M, Frisch D,Main D, Wood T, Leslie A, Wing RA, Wilkins TA (2004) Functional genomics of cell elongation in developing cotton fibers. Plant Mol Biol 54:911–929

Bandopadhyay R, Sharma S, Rustgi S, Singh R, Kumar A, Balyan HS, Gupta PK (2004) DNA polymorphism among 18 species of *Triticum-Aegilops* complex using wheat EST-SSRs. Plant Sci 166:349–356

Beasley JO (1940) The origin of American tetraploid *Gossypium* species. Am Nat 74:285–286

Brubaker CL, Brown AHD, Stewart JMcD, Kilby MJ, Grace JP (1999) Production of fertile hybrid germplasm with diploid Australian *Gossypium* species for cotton improvement. Euphytica 108:199–213

Caudrado A, Schwarzacher T (1998). The chromosomal organization of simple sequence repeats in wheat and rye genomes. Chromosome 107:587–594

Chee PW, Rong J, Williams-Coplin D, Schulze SR, Paterson AP (2004) EST derived PCR-based markers for functional gene homologues in cotton. Genome 47:449–462

Chen X, Cho YG, McCouch SR (2002) Sequence divergence of rice microsatellites in *Oryza* and other plant species. Mol Genet Genomics 268:331–343

Cronn RC, Zhao X, Paterson AH, Wendel JF (1996) Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. J Mol Evol 42:685–705

Cronn RC, Small RL, Haselkorn T, Wendel JF (2002) Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. Am J Bot 89:707–725

Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297–302

Di Rienzo A, Peterson AA, Garza JC, Valdes AM (1994) Mutational processes of simple sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

Endrizzi JE, Turcotte EL, Kohel RJ (1985) Genetics, cytology, and evolution of *Gossypium*. Adv Agron 23:271–375

Eujayl I, Sorrells ME, Baum M, Wolters P (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407

Eujayl I, Sedge MK, Wang L, May GD, Chekhovskij K, Zwonitzer JC, Mian MAR (2004) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago*spp. Theor Appl Genet 108:414–422

Fryxell PA (1979) The natural history of the cotton tribe. Texas A&M University Press, College Station

Fryxell PA, Craven LA, Stewart JM (1992) A revision of *Gossypium sect. Grandicalyx* (Malvaceae), including the description of six new species. Syst Bot 17:91–114

Guo WZ, Zhang TZ, Sheng XL, John Y, Kohel RJ (2003a) Development of SCAR marker linked to a major QTL for high fiber strength and its molecular marker assisted selection in Upland cotton. Crop Sci 6:2252–2256

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

Han ZG, Guo WZ, Song XL, Zhang TZ (2004) Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. Mol Genet Genomics 272:308–327

Jiang CX, Wright RJ, El-zik KM, Paterson AH (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). Proc Natl Acad Sci USA 95:4419–4424

John ME (1996) Structural characterization of genes corresponding to cotton fiber mRNA, E6: reduced E6 protein in transgenic plants by antisense gene. Plant Mol Biol 30:297–306

Karpinska B, Karlsson M, Srivastava M, Stenberg A, Stenberg A, Schrader J, Sterky F, Bhalerao R, Wingsle G (2004) MYB transcription factors are differentially expressed and regulated during secondary vascular tissue development in hybrid aspen. Plant Mol Biol 56:255–270

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

Kohel RJ, Yu J, Park YH, Lazo GR (2001) Molecular mapping and characterization of traits controlling fiber quality in cotton. Euphytica 121:163–172

Kuleung C, Baenziger PS, Dweikat I (2004) Transferability of SSR markers among wheat, rye, and triticale. Theor Appl Genet 108:1147–1150

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function and evolution. Mol Biol Evol 21:991–1007

Liu Q, Brubaker CL, Green AG, Marshall DR, Sharp PJ, Singh SP (2001) Evolution of the *FAD2–1* fatty acid desaturase 5′UTR intron and the molecular systematics of *Gossypium* (Malvaceae). Am J Bot 88:92–102

Menzel MY (1954) A cytological method for genome analysis in *Gossypium*. Genetics 40:214–223

Meredith WRJ (1991) Contributions of introductions to cotton improvement. In: Shands HL, Wiesner LE (eds) Use of plant introductions in cultivar development. Crop science of America, Madison, pp 127–146

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

Nolte KD, Hendrix DL, Radin JW, Koch KE (1995) Sucrose synthase location during initiation of seed development and trichome differentiation in cotton ovules. Plant Physiol 109:1285–1293

Paterson AH, Brubaker C, 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
Qureshi SN, Saha S, Kantety RV, Jenkins JN (2004) EST-SSR: a new class of genetic markers in cotton. J Cotton Sci 8:112–123
Reddy OUK, Pepper AE, Abdurakhmonov I, Saha S, Jenkins JN, Brooks T, Bolek Y, El-Zik KM (2001) New dinucleotide and trinucleotide microsatellite marker resources for cotton genome research. J Cotton Sci 5:103–113
Röder MS, Korzun V, Gill BS, Ganal MW (1998a) The physical mapping of microsatellite in wheat. Genome 41:278–283
Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganal MW (1998b) A microsatellite map of wheat. Genetics 149:2007–2023
Rohlf FJ (1998) NTSYS-pc: numerical taxonomy and multivariate analysis system, vers. 2.0. Applied Biostatistics Inc, New York
Rossetto M, Mcnally J, Henry RJ (2002) Evaluating the potential of SSR flanking regions for examining taxonomic relationships in the Vitaceae. Theor Appl Genet 104:61–66
Ruan YL, Llewellyn DJ, Furbank RT (2003) Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. Plant Cell 15:952–964
Saha S, Karaca M, Jenkins JN, Zipf AE, Reddy UK, Kantety RV (2003) Simple sequence repeats as useful resources to study transcribed genes of cotton. Euphytica 130:355–364
Saha MC, Mian MAR, Eujayl I, Zwonitzer JC, Wang LJ, May GD (2004) Tall fescue EST-SSR markers with transferability across several grass species. Theor Appl Genet 109:783–791
SAS institute (1989) SAS/STAT user's guide, version 6, 4th edn. SAS Institute, Cary
Seelanan T, Brubaker CL, Stewart JM, Craven LA, Wendel JF (1999) Molecular systematics of Australian *Gossypium* section *Grandicalyx* (Malvaceae). Syst Bot 24:183–208
Seelanan T, Schnabel A, Wendel JF (1997) Congruence and consensus in the cotton tribe. Syst Bot 22:259–290
Shen XL, Yuan YL, Guo WZ, Zhu XF, Zhang TZ (2001) Genetic stability of a major QTL for fiber strength and its marker-assisted selection in Upland cotton. High Tech Lett 11:13–16
Song XL, Wang K, Guo WZ, Zhang J, Zhang TZ (2005) A comparison of genetic maps constructed from haploid and BC_1 mapping populations from the same crossing between *Gossypium hirsutum* L. × *G. barbadense* L. Genome 48:378–390
Stewart JM (1995) Potential for crop improvement with exotic germplasm and genetic engineering. In: Constable GA, Forrester NW (eds) Challenging the future: proceedings of the world cotton research conference-1, Brisbane Australia. CSIRO, Melbourne, pp 313–327
Wendel JF (1989) New world tetraploid cottons contain old world cytoplasm. Proc Natl Acad Sci USA 86:4132–4136
Wendel JF, Albert VA (1992) Phylogenetics of the cotton genus (*Gossypium* L.): Character-state weighted parsimony analysis of chloroplast DNA restriction site data and its systematic and biogeographic implications. Syst Bot 17:115–143
Zhang J, Guo WZ, Zhang TZ (2002) Molecular linkage map of allotetraploid cotton (*Gossypium hirsutum* L. × *Gossypium barbadense* L.) with a haploid population. Theor Appl Genet 105:1166–1174
Zhang J, Wu YT, Guo WZ, Zhang TZ (2000) Rapid detection of SSR with PAGE/silver staining. Cotton Sci 12:267–269
Zhang TZ, Yuan YL, Yu J, Guo WZ, Kohel RJ (2003) Molecular tagging of a major QTL for fiber strength in Upland cotton. Theor Appl Genet 106:262–268
Zhao XP, Si Y, Hanson RE, Crane CF, Price HJ, Stelly DM, Wendel JF, Paterson AH (1998) Dispersed repetitive DNA has spread to new genomes since polyploidy formation in cotton. Genome Res 8:479–492