Euphytica (2009) 167:281–291
DOI 10.1007/s10681-008-9850-y

Promoter anchored amplified polymorphism based on random amplified polymorphic DNA (PAAP-RAPD) in cotton

Mingxiong Pang · R. G. Percy · Ed. Hughs · Jinfa Zhang

Received: 30 May 2008 / Accepted: 17 November 2008 / Published online: 3 December 2008
© Springer Science+Business Media B.V. 2008

Abstract Non-coding sequences account for a majority of the higher plant genome, some of which have important effects in gene regulation and plant development. In an effort to develop molecular marker systems to search for polymorphisms associated with high fiber yield and quality in cotton, we have developed a methodology that could specifically target the regulatory regions of the cotton genome. In this study we designed 10-nucleotide degenerate promoter primers based on conserved core promoter sequences and tested their applicability in PCR amplifications in combination with 10-mer random amplified polymorphic DNA (RAPD) primers. The amplified markers are called promoter anchored amplified polymorphism based on RAPD (PAAP-RAPD). Forty cotton genotypes with diverse genetic and geographical backgrounds were used to test the PAAP-RAPD system using polyacrylamide gel electrophoresis. Based on PAAP-RAPD markers amplified from 12 primer combinations, the 40 genotypes were classified into five distinctive groups: two Upland cotton (*Gossypium hirsutum*) groups from China, another two Upland cotton groups from the USA, and one group from American Pima cotton (*G. barbadense*). The groupings are in general consistent with their genetic and geographical origins. Thirty-six PAAP-RAPD and RAPD fragments were cloned and four of them were further subjected to sequence analysis. Signal scanning using software PLACE confirmed that they contained an array of cis-regulatory sequences in addition to the core promoter sequences. The results demonstrate the potential application of PAAP-RAPD as a new marker system specifically targeting regulatory regions of the plant genome.

Keywords Cotton · Molecular markers · Promoter anchored amplified polymorphism (PAAP)

M. Pang · J. Zhang (✉)
Department of Plant and Environmental Sciences, New Mexico State University, Las Cruces, NM 88003, USA
e-mail: jinzhang@nmsu.edu

R. G. Percy
USDA-ARS, Southern Plains Agricultural Research Center, College Station, TX 77845, USA

Ed. Hughs
USDA-ARS, Southwestern Cotton Ginning Research Laboratory, Mesilla Park, NM 88047, USA

Introduction

Numerous molecular marker systems have been developed in the last 28 years (Botstein et al. 1980; Hamada et al. 1982; Welsh and McClelland 1990; Williams et al. 1990; Dietrich et al. 1992; Vos et al. 1995). These DNA markers have been used in various areas including genetic diversity analysis, germplasm fingerprinting, linkage mapping, quantitative trait locus (QTL) dissection, gene tagging and isolation,

and marker-assisted selection in breeding. Of the many DNA marker systems developed so far, RFLP (restriction fragment length polymorphism; Botstein et al. 1980), RAPD (random amplified polymorphic DNA; Williams et al. 1990), and AFLP (amplified fragment length polymorphism; Vos et al. 1995) were extensively used as the landmarks before the 2000s. RFLP was used most extensively as the first generation of markers for genetic linkage mapping. But, the RFLP analysis usually requires radioactivity to label the probes and is time-consuming and tedious. Therefore, its broad use is restricted. Because of its reliability and high level of multiplex and polymorphism, AFLP based on polyacrylamide gel electrophoresis has been extensively applied as an alternative genetic marker system for RFLP. RAPD is based on PCR amplification of the genome using a single short (usually 10 nucleotides) primer of arbitrary sequence. The region to be amplified is usually <2 kb and contains two short sequences that are complementary to the primer in inverted orientation on opposite strands. The PCR products are usually separated using agarose gels, and visualized under ultraviolet light after stained by ethidium bromide. Due to their simplicity, rapidness, and no need for prior sequence information, the AFLP and RAPD marker systems have been highly attractive in the field of molecular marker development and molecular marker-assisted selection in plant breeding. However, the reproducibility of RAPD fragments has suffered a major criticism (Jones et al. 1997; Perez et al. 1998). Later on, when simple repetitive DNA sequences were found to be abundant and highly polymorphic in higher plants, flanking sequences have been used to develop numerous sequence-specific simple sequence repeat (SSR) or microsatellite markers (Weising et al. 2005). The most recent focus in DNA marker development is on the detection of single nucleotide polymorphism (SNP) based on expressed sequence tags (ESTs) (Weising et al. 2005).

Although most DNA marker systems can detect polymorphisms in individuals, it is unknown whether the polymorphism is derived from exons, introns or cis-regulatory regions of the genome except for these that are developed from specific gene sequences. It is well know that most of qualitative Mendelian traits are controlled by major genes that show sequence variations, the genetic basis for most QTLs as anchored by DNA markers is still unclear. However, several recent reports have demonstrated that the underlying quantitative trait nucleotides (QTN) for several cloned plant QTLs are associated with regulatory sequence variation (Salvi and Tuberosa 2005). Traditionally, functions of the regulatory sequences are characterized via mutation analysis, footprinting, and transformation studies. However, these molecular techniques are based on one gene at a time, so it is time-consuming, costly, and not high throughput. The recent development of functional genomics has generated considerable interests in identifying regulatory regions involved in gene expressions and networking of gene regulations. With the completion of genome sequencing in several model plants and several hundreds of conserved core promoter motifs available, a high throughput marker system specifically targeting the regulatory regions is now feasible because characterized promoters can be used as a platform for identifying and isolating promoters in a high throughput fashion followed by confirmation using bioinformatics and genetic experiments. Therefore, identification of potential promoter polymorphisms should enhance our understanding of wide and diverse gene expression regulations and facilitate the discovery of QTLs and expression QTLs (eQTLs) in plant breeding and genetics. However, up to date, there is no marker system that is specifically designed to target promoter and regulatory regions.

In this study, we designed four degenerate promoter primers based on conserved core promoter motifs and integrated them with RAPD primers to identify polymorphisms in regulatory regions of the cotton genome. The amplified PCR products were revealed by non-denaturing polyacrylamide gel electrophoresis. This prompter anchored amplified polymorphism based on RAPD is called PAAP-RAPD. This marker system and its further modifications might have general applications in promoter polymorphism profiling in other species.

Materials and methods

Plant materials and DNA extraction

A total of 40 genotypes (Table 1) with geologically and genetically diverse backgrounds were selected in this study. Of which, 19 Upland cotton (*G. hirsutum*) were from China with a good representation of major

Table 1 Genotypes used in the experiment

Genotype	Species	Common name	Origin of country	Grouping
JF-1	*G. hirsutum*	Upland cotton	China	C-1
JF-2	*G. hirsutum*	Upland cotton	China	C-2
JF-5	*G. hirsutum*	Upland cotton	China	C-1
JF-6	*G. hirsutum*	Upland cotton	China	C-1
JF-7	*G. hirsutum*	Upland cotton	China	C-1
JF-9	*G. hirsutum*	Upland cotton	China	C-1
JF-10	*G. hirsutum*	Upland cotton	China	C-1
JF-11	*G. hirsutum*	Upland cotton	China	C-1
JF-12	*G. hirsutum*	Upland cotton	China	C-1
JF-13	*G. hirsutum*	Upland cotton	China	C-2
JF-14	*G. hirsutum*	Upland cotton	China	Ungrouped
CRI41	*G. hirsutum*	Upland cotton	China	C-2
CRI45	*G. hirsutum*	Upland cotton	China	C-1
LU1138	*G. hirsutum*	Upland cotton	China	C-1
73	*G. hirsutum*	Upland cotton	China	C-1
H963	*G. hirsutum*	Upland cotton	China	C-2
Phy-76	*G. barbadense*	Pima cotton	USA	Pima
PS-32	*G. barbadense*	Pima cotton	USA	Pima
K239	*G. barbadense*	Pima cotton	USA	Pima
K-240	*G. barbadense*	Pima cotton	USA	Pima
GB3-79	*G. barbadense*	Pima cotton	USA	Pima
57-4	*G. barbadense*	Pima cotton	USA	Pima
PS-1	*G. barbadense*	Pima cotton	USA	Pima
97H	*G. hirsutum*	Upland cotton	China	C-1
Ji1286	*G. hirsutum*	Upland cotton	China	US-1
SGKT	*G. hirsutum*	Upland cotton	China	C-1
TM-1	*G. hirsutum*	Upland cotton	USA	US-1
R8518	*G. hirsutum*	Upland cotton	USA	US-1
NM24016	*G. hirsutum*	Upland cotton	USA	US-2
A33(Acala)	*G. hirsutum*	Upland cotton	USA	US-1
Auburn 56	*G. hirsutum*	Upland cotton	USA	US-2
Auburn 634	*G. hirsutum*	Upland cotton	USA	C-1
DP 61	*G. hirsutum*	Upland cotton	USA	US-2
M-240	*G. hirsutum*	Upland cotton	USA	US-2
ST 213	*G. hirsutum*	Upland cotton	USA	US-2
M-249	*G. hirsutum*	Upland cotton	USA	Ungrouped
Coker 201	*G. hirsutum*	Upland cotton	USA	US-2
M-120	*G. hirsutum*	Upland cotton	USA	US-2
D8R8518	*G. hirsutum*	Upland cotton	USA	US-2
D2A	*G. hirsutum*	Upland cotton	USA	Ungrouped

cotton production areas; 14 Upland cotton (*G. hirsutum*) were from USA and seven were American Pima cotton (*G. barbadense*). Leaf tissues from 5 to 10 plants per genotype were collected in the greenhouse-grown plants and stored at −20°C until use. DNA from the 40 genotypes was extracted according to the rapid DNA isolation method described by Zhang and Stewart (2000).

Primer design

After data-mining, four degenerate promoter primers (G-1: GCCACSTGTC for G box; GC-1: NNNGGG CGGN for GC box; CA-1: YRRCCAATWSR for CAAT box; and TA-1: CTATAWAWASM for TATAAT box) were designed based on the conserved sequences of known promoter sequences (Butler and Kadonaga 2002) and synthesized from Sigma (Sigma–Aldrich Co, St. Louis, MO, USA). Three decamer oligonucleotide RAPD primers (UBC 001: CCTGGGCTTC; UBC 002: CCTGGGCTTG; and UBC693: GACGAGACGG) were synthesized from the University of British Columbia, Vancouver, Canada.

PCR amplification

To investigate the potential promoter anchored polymorphism (PAAP) among genotypes, polymerase chain reaction (PCR) using two primers (an RAPD primer and a degenerate promoter primer) was performed. The three RAPD primers and four promoter primers generated 12 primer combinations in the PAAP-RAPD analysis. The PCR reactions were carried out in a 20 µl volume with PCR conditions following Zhang and Stewart (2000). The amplification was performed in a PE Applied Biosystems GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA). The PCR products were electrophoresed in 5% non-denaturing polyacrylamide gels using Sequi-Gen GT Sequencing Cell (BioRad Laboratories, Hercules, CA, USA) and visualized by silver staining.

Gel scoring and data analysis

The PAAP-RAPD fragments are generally dominant and were recorded as present (1) or absent (0) for all the genotypes. During data collection and analysis, only the unique PAAP-RAPD bands (both monomorphic and polymorphic) produced by the primer combinations were scored using the following criteria. Common bands amplified by an RAPD primer alone and the RAPD primer/promoter primer combination were deemed as amplified by the RAPD primer. Similarly, common bands amplified by a promoter primer alone and the promoter primer/RAPD primer combination were deemed as amplified by the promoter primer. Therefore, these bands produced by RAPD primers or potential promoter primers alone were not recorded. To distinguish these three types of bands and for cloning purposes, four diverse genotypes (TM-1, NM24060, Pima 57–4, and 3–79, Table 1) were used. PCR products amplified by an RAPD primer, a promoter primer, and RAPD primer/promoter primer combination for the four genotypes were loaded on the gel side by side for a comparison (Fig. 1). Genetic similarity coefficients were calculated based on Jaccard's coefficient (JC) using the Numerical Taxonomy Multivariate Analysis System (NTSYSpc) Version 2.1 software package (Exeter Software, Setauket, NY, USA). The calculated JC coefficients were used to perform the cluster analysis using the un-weighted pair group method of arithmetic means (UPGMA).

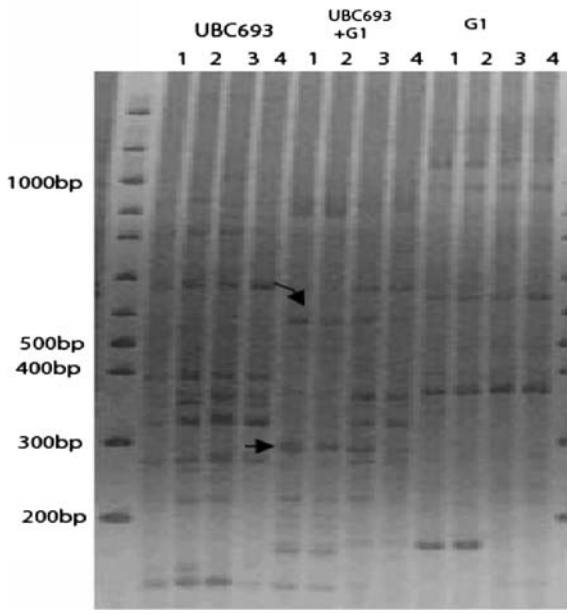


Fig. 1 PCR amplification and polyacrylamide gel electrophoresis using primers UBC 693 and G1 alone and their combination. *Arrows* indicate bands produced by the combination of UBC 693 and G1 in comparison with the bands produced by UBC 693 and G1 alone. For each primer or primer combination, the four lanes are, 1: TM-1; 2: NM 24016; 3: 57–4; and 4: 3–79

Cloning of polymorphic PAAP-RAPD fragments

Thirty-six polymorphic PAAP-RAPD or RAPD bands amplified from the 40 genotypes were selected, excised from polyacrylamide gels, and pooled. The excised fragments were eluted in 30 µl ddH_2O and

re-amplified using 5 µl of elute as DNA template, and the same primer combination and PCR conditions. After confirmed in a 1.4% agarose gel, the re-amplified fragment was cut and purified using the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA, USA). The PCR products were cloned using PCR cloning vector pGEM-T Easy system (Promega, Madison, WI, USA).

Sequencing and sequence analysis

For each transformation, 3 clones were randomly picked for DNA sequencing using Li-Cor 4200 Global IR2 system (Lincoln, NE, USA). The sequences were first searched for homology in GenBank using BLASTN (Altschul et al. 1997). Furthermore, in order to obtain more biological information about the cloned PAAP-RAPD fragments, 220 plant promoter sequences were downloaded from http://www.softberry.com and combined with the cloned PAAP-RAPD sequences amplified with both potential promoter primer and RAPD primer to conduct an analysis using Clustal W (Version 1.82 Chenna et al. 2003). To identify regulatory signals for transcription factors, the sequences were further analyzed using the promoter analysis software, PLACE (Higo et al. 1999).

Results

Comparison between PAAP-RAPD and RAPD markers

A comparison of PCR fragments amplified by PAAP-RAPD and RAPD was made using a promoter primer and an RAPD primer separately or in combination (Fig. 1). The RAPD primer UBC 693 and degenerate promoter primer G1 each amplified about 15 unique fragments of various sizes, some of which were polymorphic among the four genotypes tested. When the two primers were combined in a PCR reaction, some unique PAAP-RAPD fragments were produced, while some RAPD fragments disappeared due perhaps to primer competition for priming sites. The new bands were presumably produced by the primer combination of UBC 693 and G-1, and therefore were specific for the potential promoter. Using the comparative analysis as described in 'Materials and Methods', PAAP-RAPD bands were identified for each primer combination.

Table 2 Statistics of monomorphic and polymorphic PAAP-RAPD bands amplified using a promoter anchored primer in combination with an RAPD primer

Primer combination	No. monomorphic bands	No. polymorphic bands	No. total
UBC001 + G-1	4	12	16
UBC001 + GC-1	0	7	7
UBC001 + CA-1	3	7	10
UBC001 + TA-1	8	9	17
UBC002 + G-1	0	7	7
UBC002 + GC-1	5	9	14
UBC002 + CA-1	6	11	17
UBC002 + TA-1	3	11	14
UBC693 + G-1	0	6	6
UBC693 + GC-1	0	13	13
UBC693 + CA-1	0	12	12
UBC693 + TA-1	2	8	10
Total	31	112	143

Polymorphism of PAAP-RAPD markers

Each primer combination produced six to 17 PAAP-RAPD markers for the 40 genotypes with an average of 12 fragments per primer combination. A total of 143 PAAP-RAPD fragments were produced from the 12 primer combinations (four potential promoter primers in combination with three RAPD primers). A monomorphic band was defined as being present in all genotypes tested. Since the 40 genotypes were highly diverse from two cultivated allotetraploid species, 112 (78.3%) PAAP-RAPD markers were found to be polymorphic, while only 31 (21.7%) were monomorphic (Table 2). Our results also showed that many polymorphic PAAP-RAPD markers are species specific while some are genotype specific (Fig. 1).

Clustering analysis

A dendrogram based on the 143 PAAP-RAPD alleles was obtained using UPGMA analysis on Jaccard's coefficients (JC). The JC ranged from 0.57 to 0.98, with the lowest JC between Upland and Pima cotton. The 40 genotypes were clustered into five main groups using JC of 0.78 as a dividing point (Fig. 2). As expected, all the seven Pima genotypes were

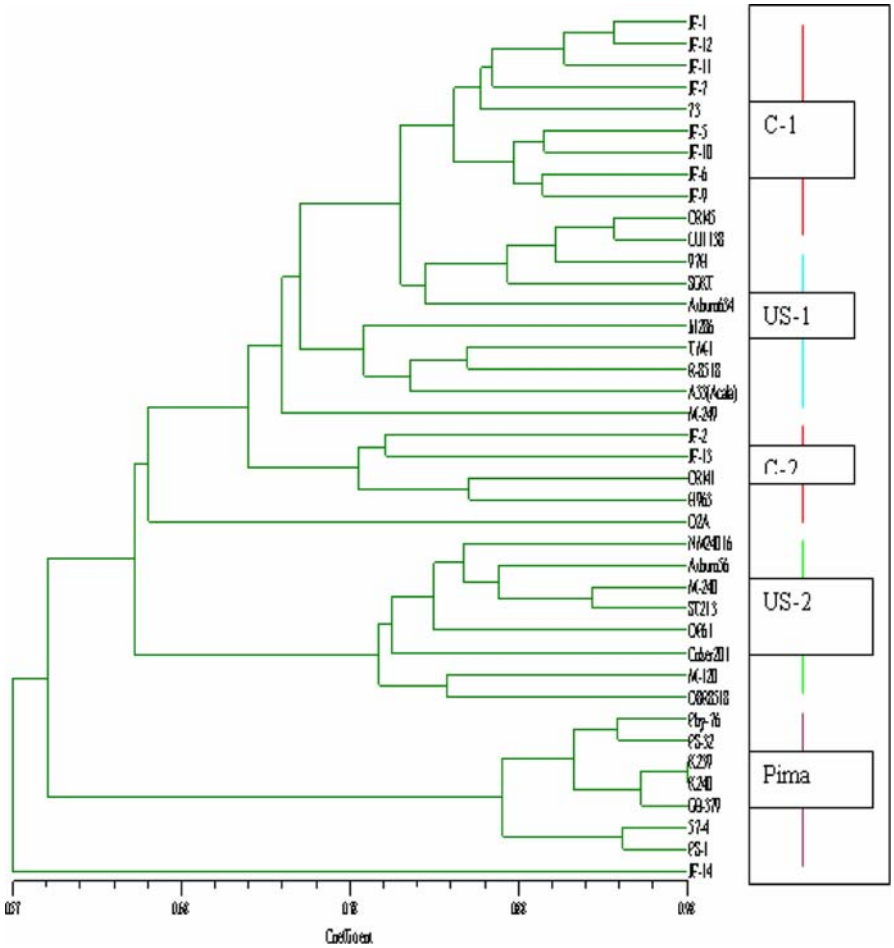


Fig. 2 A dendrogram of 40 cotton genotypes using the un-weighted pair-group method, arithmetic average (UPGMA) procedure based on Jaccard's coefficient from 143 PAAP-RAPD markers amplified by primer combinations of 10-mer and potential promoter primers

clustered into one group which can be further divided into three subgroups. Within Upland cotton, four groups were formed: C-1 group with 13 Chinese Upland cottons and one American Upland cotton genotype; US-1 group with three American Upland cotton genotypes and one Chinese Upland cotton genotype; C-2 group with four Chinese Upland cotton genotypes; and US-2 group with eight Upland cotton genotypes from the USA. Within each group, subgroups could be further divided. As expected, these four Upland cotton groups were grouped together before joining the Pima group. The major groupings are in general consistent with their genetic and geographic origins. Since original Chinese Upland cotton were introductions from the USA before the 1950s, it is not surprising that these modern Chinese cultivars tested are still closely related to the Upland cotton in the USA. However, it appeared that the modern Chinese Upland cotton cultivars have accumulated some unique genetic markers as revealed by PAAP-RAPD in the last 50 years that placed most of them into different groups from the USA Upland cotton.

Cloning and sequence analysis of PAAP-RAPD markers

Thirty-six polymorphic PAAP-RAPD and RAPD fragments were cloned and sequenced, and no high homology to known sequences/genes was found in GeneBank using BLAST search (Altschul et al. 1997). Most of the cloned fragments contained either the RAPD or promoter primer sequences, while PAAP-RAPD primers were identified in four sequences (Fig. 3). They were clustered into different groups with the 220 plant promoter sequences and a portion of the cladogram with the four PAAP-RAPD sequences and some plant promoter sequences is shown in Fig. 4, implying that they are potential

Fig. 3 Cloned potential promoter sequences amplified by a potential promoter primer and an RAPD primer. The primer sequences are underlined

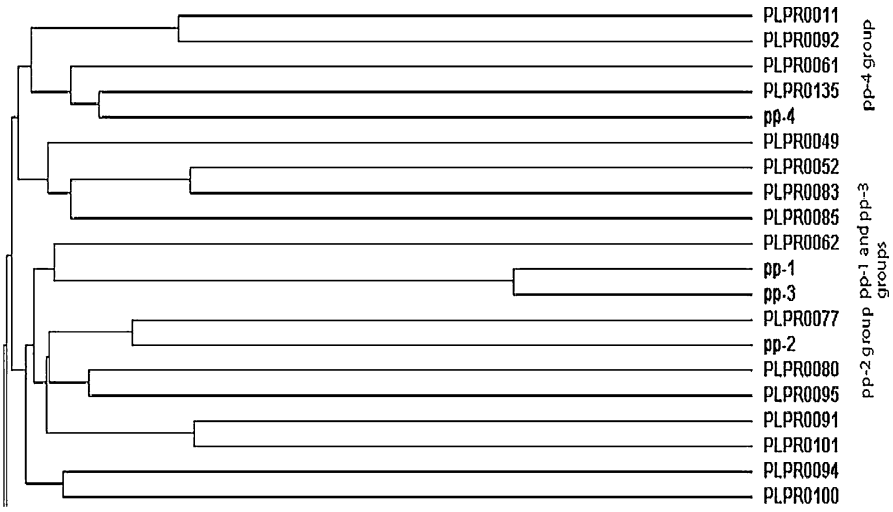
```
>1-2(UBC001+G1) both (pp-1)
GCCACGTGTCGAACATCTAATATGACAAGGAAAAATATCTAGCTGACTGGATAAGTGATGGGCATTGTTCACTTACTTA
CCCAGGAAGTATTCTCTACCTATAAAAGGATGATCTAAGGGGAAAAATCCTCTCATCCTTTAAGATGAGTCTAGAACGT
GGGTTAAGCTGGGAGATGCCCCTGTCAAGGTAAGAATCGTAACTTTACATGATATACTGCCACATAGTGATGAGTTTGT
TTGGTGAAGCAAAGCATGGTAAAGTTTTATTATTAGCTTTTAGTGTTTGAAGTGAAAGGAAATGAGTTAAATAAATAGA
TCTTTATGAGTTCTTATGGTTTATCGTTAAAGGTTATCGATCAACTCGATGGAAGCCCAGG

>2-1(UBC001+G1) both (pp-2)
GCCACGTGTCACCATGATAGTGATTATTAGTATCATATCAACACATTTTGTCGGCATTAATGAAGTACTTAAAAAGAGT
ACCAAATTAATTTATAATTTTTAAAATAGGTTTCAATTAAGTAAAAAAAAAAGTTAAATACTAAATAAGTACCTAACCCA
AATATTTAATTTTGGTAAGCTTGTATTACCAGAAAAGGTACACATAGAAATATCCGAAATTCAAAAAACTGGAATAAA
GGCAAACATGTCTCTTAAAAAAGAGTTGACATAAACAAGCGCGTACATTAATGATTGTTAAAAAAAAGGGAATTTTTA
TTTTTATATTATGCTAAAATTTGTAATATACTTTGGATGTAGGGTGATATCTTTCAATTTCTTTTATATAGAAATTTTATT
ATCTAGTGTACTTTTTTTTATTTCTTAAATTTAAAATGTTAATATTATATGGAATATTCATTTAAATTGAAACTTTAAAAA
TTAAATTTTATAGCTCAATCATTTTCAAAATAGAGATGTAAGGGTGTATTATATtATGAAGCCCAGG

>16-1(UBC001+G1) both (pp-3)
GCCACGTGTCGAGCATTTAATGGGACAAAGGAAAATATCTAGCTAACCGAGTCAATGATGGTACTATTCATTTACTTCT
TCGAGAAGTAACCTCAGTCTATATAATGATGAGCTAAGGGGGGAAAAATCCTCGTATCCTTTAAGATGAGTCAAAAAT
GAGAGGATTAGCTTTGAGATGTCCAATTTGTAAGGTAAGTTGTCAAACCTTACATGCTAAGTTGCCATAAGAGTAGATT
TGTTTAAGCTTAGTAATAAAATGGTAAAACCTTGTTGTACAATTCTAGTGTTCAAGTGAGAAAGGAAATATCTTGATGA
TTTTCTTGTGGTTTAACGTTTAAGAGCTGATGACCTTCTTGATGGAAGCCCAGG

>27(UBC693+G1) both (pp-4)
GACGAGACGGGAAGACCTAGGATGTAGAAGATCATAACATGAGTGGTTCGACAGATTTGGATTCCTATATATCCACTC
ATGTGGTACTTCATTGTACAATATATATAAGAATTATACGATAAGAATTATACGAATCCATCTGTATAGATATCATCAT
CTACATCCAGAAAGCCGTATGCTTTGGAAGAAGCTTGTACAGTTTGGGAAGGGGTTTTGATTGATCGATCAAAAAGAA
GAATCTACTTCAACCGATATGCCCTTAGGCACGGCCATACATAACATAGAAATCACACTTGGAAGGGGTGGACAATTA
GCTAGAGCAGCGGGTGCTGTTGCGAAACTGATTGCAAAGGAGGGGAAATCGGCCACATTAAAATTACCTTCTGGGGAG
GTCCGTTTGATATCCAAAAACTGCTCAGCAACAGTCGGACAGGTGGC
```

Fig. 4 A portion of cladogram representation of 4 cloned potential promoter sequences with other known promoter sequences. *pp* denotes PAAP-RAPD sequences (see Fig. 3). *PLPR* denotes plant promoter sequences in the database



promoter-like sequences in the cotton genome. To identify other cis-acting elements for transcription factors in the cloned PAAP-RAPD sequences, we further used the promoter analysis software, PLACE (Higo et al. 1999), to scan regulatory signals of the PAAP-RAPD sequences. Table 3 presents a detailed example for one of the PAAP-RAPD sequences. The results indicated that G box, TATA box, and CAAT box were all detected, as expected, since they are in physically close proximity with one another and are located on the upstream of numerous genes. There were also other potential signals for transcription factors including WRKY and MYB. Thus, it is likely that there is a functional gene down-stream of this PAAP-RAPD sequence.

Discussion

During plant development, different genes are required for the completion of different stages (Jakoby and Schnittger 2004). Furthermore, because the binding of RNA polymerase II to its promoters is the key step in the initiation of gene expression, it is

Table 3 Edited signal scanning results of one cloned PAAP-RAPD sequence, 16- (UBC 001 + G1) see Fig. 4 for the PAAP-RAPD sequence

Motif	Position	Strand	Motif sequence	Site #
ABREAT CONSENSUS	1	(−)	YACGTGGC	S000406
ABRELATERD1	4	(+)	ACGTG	S000414
ABRELATERD1	3	(−)	ACGTG	S000414
CAAT BOX1	53	(+)	CAAT	S000028
CAAT BOX1	181	(+)	CAAT	S000028
CAAT BOX1	276	(+)	CAAT	S000028
CACGTG MOTIF	3	(+)	CACGTG	S000042
CACGTG MOTIF	3	(−)	CACGTG	S000042
CACTFTPPCA1	62	(+)	YACT	S000449
DOFCOREZM	27	(+)	AAAG	S000265
DOFCOREZM	297	(+)	AAAG	S000265
DOFCOREZM	138	(−)	AAAG	S000265
DOFCOREZM	169	(−)	AAAG	S000265
MYB1AT	324	(−)	WAACCA	S000408
MYB CORE	44	(−)	CNGTTR	S000176
MYB ST1	134	(−)	GGATA	S000180
NOD CON2GM	337	(−)	CTCTT	S000462
ROOT MOTIF TAPOX1	34	(−)	ATATT	S000098
ROOT MOTIF TAPOX1	303	(−)	ATATT	S000098
S1 F BOX SORPS1L21	58	(+)	ATGGTA	S000223
S1 F BOX SORPS1L21	257	(+)	ATGGTA	S000223
SEF4 MOTIF GM7S	151	(−)	RTTTTTR	S000103
SP8 BFIBSP8BIB	62	(+)	TACTATT	S000184
TAAA GSTKST1	138	(−)	TAAAG	S000387
TATA BOX4	99	(+)	TATATAA	S000111
W BBOX PCWRKY1	148	(−)	TTTGACT	S000310
W BOX ATNPR1	51	(−)	TTGAC	S000390
W BOX ATNPR1	149	(−)	TTGAC	S000390
W BOX ATNPR1				
WRKY 71OS	346	(+)	TGAC	S000447
WRKY 71OS	51	(−)	TGAC	S000447
WRKY 71OS	149	(−)	TGAC	S000447
WRKY 71OS	199	(−)	TGAC	S000447
WUSATAg	17	(+)	TTAATGG	S000433

suggested that specific sequences may exist in the promoters which control the temporal and spatial gene expressions (Kreiman 2004). It is increasingly evident that the interactions of a multitude of different overlapping cis-elements and transcription factors are responsible for diverse gene expression events. The advantage of fully sequenced model plants such as Arabidopsis facilitates an easier promoter identification and comparative analysis. Seki et al. (2002) identified promoter regions by comparing data of the 5′-ends of isolated full-length cDNAs from Arabidopsis to its genomic sequences and constructed a promoter database using data from a plant transcription factor database, PLACE (Higo et al. 1999). There are other promoter databases such as PlantCARE- a database of plant cis-acting regulatory elements, enhancers, and repressors (Lescot et al. 2002), Eukaryotic Promoter Database (EPD) (Schmid et al. 2004), the Hematopoiesis Promoter Database (HemoPDB) (Pohar et al. 2004), and the CORG database (Dieterich et al. 2003). Identification of promoters and their regulatory elements is one of the major challenges for both

molecular biology and bioinformatics. The efforts require integration of comparative, structural, and functional genomics. Many bioinformatics methodologies have been designed to detect conserved motifs in a set of genes and the findings of the conserved motifs would be useful to identify and molecularly profile the regulatory regions of the genome.

Developing markers (such as EST-SSR or SNP) targeting gene sequences based on ESTs is easy as long as one has the EST sequences. However, regulatory sequences including the core promoters are not included in the EST sequences in most cases if not all. Unless one knows the upstream regulatory sequences for a gene, designing specific primers to regulatory sequences of the gene would be impossible. In this study, we have described a strategy named promoter anchored amplified polymorphism (PAAP) based on RAPD (PAAR-RAPD). For this marker system, we use degenerate promoter primers designed based on the conserved promoter sequences in combination with RAPD 10-mer primers to specifically profile potential promoter regions of the cotton genome. The major advantage of this marker system is its potential in identifying and mapping promoter regions genome-wide. However, individual promoter sequences are not needed for the PAAP marker development. In this regard, PAAP-RAPD has similar advantages and disadvantages to RAPD, but targeted to potential promoter regions. Even though RAPD-based markers may not consistently yield similar results, they are still being widely used. Because of their biological importance in gene expression and physical proximity to genes, this PAAP marker system may also amplify the 5′-UTRs (untranslated regions) and the first exon of their down-stream genes.

Several DNA sequence variations including SNP, insertion, and deletion produce genetic polymorphisms. As with AFLP, RFLP, SSR or SNP, RAPD or PAAP-RAPD can detect these variations. For example, a nucleotide replacement (point mutation) in a core promoter sequence can affect primer binding, resulting in no amplification (i.e., absence of a PAAP-RAPD band). An insertion or deletion between a core promoter and gene sequence down-stream or up-stream that has a binding site for an RAPD primer could result in PAAP-RAPD length polymorphism (in this case, it could be co-dominance). Sequence variation in regulatory elements and length between them and the down-stream genes may affect transcription factor binding and transcription efficiency, which in turn affects the levels of gene expression. The differential gene expression levels may eventually affect plant growth and development such as yield and quality. Therefore, PAAP may provide an important and focused link between DNA sequence polymorphism and QTLs.

However, functional and clear descriptions of genomic segments that contain elements required to activate transcription would be helpful but are still unavailable for most genes, even though some regulatory motifs of specific genes have been investigated experimentally in detail. Through using potential promoter primers designed based on the core conserved promoter sequences to specifically target promoters in the genome, we tested our hypothesis using 40 cotton genotypes with diverse genetic and geographical backgrounds based on the PCR amplification using 12 primer combinations of RAPD and degenerate promoter primers. The genotype classification based the PCR amplification bands is overall consistent with their genetic and geographical origins. Furthermore, in order to evaluate the cloned potential promoters we downloaded the characterized dicot plant promoters from database at http://www.softberry.com/with a total of 220 promoter sequences from dicot plants. We cloned and then sequenced 36 polymorphic PAP-RAPD and RAPD bands from the polyacrylamide gels. As an example, four cloned PAAP-RAPD sequences from potential promoter regions amplified with both degenerate promoter and 10-mer random primer were analyzed for homology with the 220 promoters from the databaser using ClustalW (Version 1.82, Chenna 2003). We found that all of the four fragments were clustered into various promoter groups (Fig. 4). The results suggest that the sequences cloned may be potential promoters. In this study, individual polymorphic PAAP-RAPD or RAPD bands were isolated and cloned. The efficiency of cloning was low. Even though the efficiency of this PAAP-RAPD marker system targeting promoter regions cannot be validated without tremendous molecular characterization of individual sequences, this strategy can be used to amplify potential promoter regions using a number of primer combinations to construct genome-wide PAAP-RAPD libraries for high throughput cloning and sequencing. The sequences obtained could be used for designing more sequence-specific primers to develop promoter-specific markers.

As indicated in Fig. 1, the PCR fragments are generated from three different sources. The first is the bands that are produced by the RAPD primers only. The second source comes from the PCR amplification with only potential promoter primers. The third is produced by the primer combination of RAPD and degenerate promoter primers. Strictly speaking, the second and third ones are most interesting because of the possibility to find polymorphisms from potential promoter regions in the genome. However, for fingerprinting purpose, the distinction may be unnecessary. For mapping, it becomes much easier since only two parental lines are usually used for producing a mapping population. One can always add the parental DNA to run PCR using RAPD primer only, promoter primer only, and the primer combination when genotyping the mapping population with the primer combination. In this way, unique PAAP-RAPD markers can be identified in the mapping population in comparison with the banding patterns from the parents. In doing so, polymorphic markers amplified by RAPD primer only or promoter primer only can also be distinguished and scored for mapping.

As with RAPD, the reproducibility of RAPD-based PAAP markers (PAAP-RAPD) may be questionable using ordinary low-resolution agarose gels, because many faint bands can not be clearly stained for a reliable scoring. However, our results indicated that separation of PAAP-RAPD using polyacrypamide gels tremendously enhanced the resolution of the system and therefore reliability of fragment recording. In our study, we did not specifically clone and sequence sufficient PAAP-RAPD fragments to exam the efficiency of this new marker strategy. However, of the 36 randomly chosen RAPD and PAAP-RAPD fragments, 4 (>10%) were in fact amplified by PAAP-RAPD primers. This relatively low efficiency was also encountered in the TRAP marker system. To ensure a high efficiency of PCR amplification from both short primers, primer design should be improved in the future. Furthermore, longer PAAP primers integrated in AFLP or other marker systems such as TRAP (Hu and Vick 2003) and SRAP (Li and Quiros 2001) should further increase the reproducibility and throughput of the current PAAP marker system.

Through cloning and sequencing of polymorphic PAAP-RAPD, sequences from the cloned fragments with RAPD and degenerate promoter primers can be further analyzed. The current work represents the first attempt in designing a DNA marker system to profile polymorphism in promoter or regulatory regions. With a better understanding of plant promoter/regulatory sequences in the future, more promoter primers can be further designed and refined. Using numerous primer combinations, polymorphic PAAP markers can be mapped in the cotton genome for a genome-wide mapping of promoter regions and their relationships with quantitative traits. Once co-localization of a PAAP marker and a QTL is confirmed, the relationship between the PAAP marker and the QTL can be further studied to identify candidate DNA sequences for the quantitative trait using various molecular techniques. The PAAP will provide a good start point to characterize differences in regulatory sequences among genomes and progenies.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. doi:10.1093/nar/25.17.3389

Botstein DRLW, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Butler JEF, Kadonaga JT (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev 16:2583–2592. doi:10.1101/gad.1026202

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the clustal series of programs. Nucleic Acids Res 31:3497–3500. doi:10.1093/nar/gkg500

Dieterich C, Wang H, Rateitschak K, Luz H, Vingron M (2003) CORG: a database for comparative regulatory genomics. Nucleic Acids Res 31:55–57. doi:10.1093/nar/gkg007

Dietrich W, Katz H, Lincoln SE, Shin HS, Friedman J, Dracopoli NL, Lander ES (1992) A genetic map of the mouse suitable for typing intraspecific crosses. Genetics 131:423–447

Hamada H, Petrino MG, Kakunaga T (1982) A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc Natl Acad Sci USA 79:6465–6469. doi:10.1073/pnas.79.21.6465

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database. Nucleic Acids Res 27:297–300. doi:10.1093/nar/27.1.297

Hu J, Vick BA (2003) Target region amplification polymorphism: a novel marker technique for plant genotyping. Plant Mol Biol Rep 21:289–294. doi:10.1007/BF02772804

Jakoby M, Schnittger A (2004) Cell cycle and differentiation. Curr Opin Plant Biol 7:661–669. doi:10.1016/j.pbi.2004.09.015

Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, van de Wiel C, Bredemeijer G, Vosman B, Matthes M, Daly A,

Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevschi A, Marmiroli N, Aert R, Volckaert G, Rueda J, Linacero R, Vazquez A, Karp A (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol Breed 3:381–390. doi:10.1023/A:1009612517139

Kreiman G (2004) Identification of sparsely distributed clusters of cis-regulatory elements in sets of co-expressed genes. Nucleic Acids Res 32:2889–2900. doi:10.1093/nar/gkh614

Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S (2002) Plant CARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30:325–327. doi:10.1093/nar/30.1.325

Li G, Quiros CF (2001) Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in Brassica. Theor Appl Genet 103:455–461. doi:10.1007/s001220100570

Perez T, Albornoz J, Dominguez A (1998) An evaluation of RAPD fragment reproducibility and nature. Mol Ecol 7:1347–1357. doi:10.1046/j.1365-294x.1998.00484.x

Pohar TT, Sun H, Davuluri RV (2004) HemoPDB: Hematopoiesis Promoter Database, an information resource of transcriptional regulation in blood cell development. Nucleic Acids Res 32:D86–D90. doi:10.1093/nar/gkh056

Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci 10:297–304. doi:10.1016/j.tplants.2005.04.008

Schmid CD, Praz V, Delorenzi M, Perier R, Bucher P (2004) The eukaryotic promoter database EPD: the impact of in silico primer extension. Nucleic Acids Res 32:D82–D85. doi:10.1093/nar/gkh122

Seki M, Narusaka M, Kamiya A, Ishida J, Satou M, Sakurai T, Nakajima M, Enju A, Akiyama K, Oono Y, Muramatsu M, Hayashizaki Y, Kawai J, Carninci P, Itoh M, Ishii Y, Arakawa T, Shibata K, Shinagawa A, Shinozaki K (2002) Functional annotation of a full-length *Arabidopsis* cDNA collection. Science 296:141–145. doi:10.1126/science.1071006

Vos P, Hogers R, Bleeker M, Reijans M, Tvd Lee, Hornes M, Friters A, Pot J, Paleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414. doi:10.1093/nar/23.21.4407

Weising K, Nybom H, Wolff K, Kahl G (2005) DNA fingerprinting in plants: principles, methods, and applications. 2nd edn. CRC Press

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18:7213–7218. doi:10.1093/nar/18.24.7213

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535. doi:10.1093/nar/18.22.6531

Zhang J, Stewart JM (2000) Economical and rapid method for extracting cotton genomic DNA. J Cotton Sci 4:193–201