1869 Development of Gene Targeted AFLP (GT-AFLP) Marker Systems and Their Applications in Cotton Genomics

Dr. Jinfa Zhang , New Mexico State University, Las Cruces, NM
Mr. Fei Wang , New Mexico State University, Las Cruces, NM
Mr. Chen Niu , New Mexico State University, Las Cruces, NM
Mr. Wu Wang , New Mexico State University, Las Cruces, NM
Mr. Mingxiong Pang , New Mexico State University, Las Cruces, NM
Mr. Yingzhi Lu , New Mexico State University, Las Cruces, NM
Ms. J. Curtiss , New Mexico State University, Las Cruces, NM
Ms. D. Miranda , New Mexico State University, Las Cruces, NM
Dr. Youlu Yuan , Cotton Research Institute, CAAS; Key Laboratory of Cotton Genetic Improvement, Ministry of Agriculture, P R China, Anyang, Henan, China
Dr. Shuxun Yu , Cotton Research Institute, CAAS; Key Laboratory of Cotton Genetic Improvement, Ministry of Agriculture, P R China, Anyang, Henan, China
Dr. Richard Percy , Arid-Land Agricultural Research Center, USDA-ARS, Maricopa, AZ
Dr. James McD Stewart , Department of Crop, Soil, and Environmental Sciences, University of Arkansas, Fayetteville, AR
Dr. Mauricio Ulloa , USDA-ARS, Western Intergrated Cropping Systems, Res. Unit, Shafter, CA
Dr. Roy Cantrell , Cotton Incorporated, Cary, NC

ACKNOWLEDGEMENTS:

This research was supported in part by grants from Cotton Incorporated, Cary, N.C., USA, USDA-ARS, and the New Mexico Agricultural Experiment Station, NM.

ABSTRACT

In higher plants, many genes exist in gene families with conserved sequence domains. To take advantage of genome sequence information, we have developed high throughput marker systems using a selective AFLP primer in combination with one gene targeted (GT) primer (collectively called GT-AFLP analysis) that target gene families in cotton. Our results demonstrate that this gene targeted AFLP-based strategy allows high-throughput, low cost detection of polymorphic functional markers in cotton. To demonstrate the robustness and applications of GT-AFLP, degenerate primers have been designed from gene families encoding for transcription factors (TF-AFLP), disease resistance analogs (RGA-AFLP), and pentatricopeptide proteins (PPR-AFLP). Examples of the effectiveness of GT-AFLP marker systems are provided in the areas of germplasm diversity and classification, genome-wide GT-AFLP marker mapping, candidate gene mapping, and gene expression studies. It is concluded that GT-AFLP is a very useful marker system to survey cotton genome for genetic diversity at the DNA level and for differential gene expression at the RNA level and is applicable to other plant species.

Since the advent of restricted fragment length polymorphism (RFLP), numerous molecular marker systems have been developed in the last 25 years. As genomic DNA sequences in most species were unknown, PCR-based marker technologies such as amplified fragment length polymorphism (AFLP) were developed (Vos et al., 1995). AFLP has been widely used to rapidly generate molecular markers among numerous organisms from bacteria to plants for various purposes including investigation of genetic diversity, germplasm fingerprinting, linkage and quantitative trait locus (QTL) mapping, gene isolation, and marker-assisted

selection in breeding. The most recent focus in DNA marker development is on the detection of single nucleotide polymorphism (SNP), and expressed sequence tags (EST) have provided an ample source for mining and developing SNPs.

Genomes of several plant species including *Arabidopsis*, rice and poplar have been sequenced. In *Arabidopsis*, approximately 1,000 gene families with more than 8,300 genes (>32% of 26,000 genes in the genome) have been identified (http://www.arabidopsis.org/browse/genefamily/index.jsp). A gene family is a group of genes coding for diverse proteins with related functions which, by virtue of their high degree of sequence similarity, are believed to have evolved from a single ancestral gene. For example, ~2,000 transcription factor (TF) genes (7.4% of *Arabidopsis* genes) belong to more than 60 gene families in Arabidopsis (http://datf.cbi.pku.edu.cn; http://arabtfdb.bio.uni-postsdam.de/ v1.1; Gong et al., 2004; Xiong et al., 2005). TFs are regulatory proteins in eukaryotes and often exhibit sequence-specific DNA-binding and are capable of activating or repressing transcription of multiple target genes. The DNA-binding domains of transcription factors are highly conserved and can be used to design degenerate primers for gene targeted (GT) PCR reactions to 5 survey for genetic diversity at the DNA level and for differential gene expression at the RNA level.

Plant disease resistance (R) genes with domains of nucleotide binding sites (NBS) and leucine rich repeats (LRR) represent one of the largest gene families (Martin et al., 2003). Recent genome analyses identified approximately 150 and 500 NBS-LRR genes in *Arabidopsis* (Meyers et al., 2003) and rice (Monosi et al., 2004), respectively. The conserved NBS domain comprising the P loop, the kinase-2 motif, and the GLPL motif has enabled the isolation of disease resistance analogues (RGAs) from numerous plant species (Martin et al., 2003). In *Arabidopsis* and rice genomes, each chromosome contains NBS-LRR genes, some of which are clustered. On most chromosomes it appears that on average every 5-20 cM contains a RGA or RGA cluster (Meyers et al., 2003; Monosi et al., 2004). Therefore, RGA could be used as chromosome anchored markers for disease resistance candidate gene mapping. Furthermore, many RGA may also be in the proximity with other genes or in gene-rich regions. Therefore, polymorphic RGA-AFLP markers can be also used for mapping other traits. Degenerate RGA primers in combination with selective AFLP primers provide numerous RGA-AFLP markers to conduct genome-wide mapping of R genes and RGA in any plant species in which no prior sequence information is required.

Among the gene families, the largest and perhaps the most mysterious is the PPR gene family, characterized by tandem arrays of the so-called pentatricopeptide repeats (PPRs), 35 degenerated amino acids (Small and Peeters, 2000). The gene family coding for PPR proteins expanded vastly during the evolution of the land plants: over 440 and 570 PPR genes are present respectively in the Arabidopsis (1-2% of all proteins) and rice genomes. The motif is found in a few animal and fungal proteins. The functions of these proteins are mostly unknown, but the vast majority of these proteins are predicted to be targeted to either mitochondria or chloroplasts, where the PPR proteins play constitutive, often essential roles. About half of the PPR are made of PPR motifs adjacent to each other and contain tandem repeated triple motifs (PLS), variants of the PPR motif and the PLS proteins can be further divided into four classes (A, E, E+ and DYW) based on the nature of their C-terminal motif. One important function of PPR proteins is involved in fertility restoration to cytoplasmic male sterility in plants (Chase, 2007).

Based on the notion that genetic diversity in gene-rich regions may have a higher probability of being associated with phenotypic expression, we have expended considerable effort to develop high throughput marker systems that target gene families by using a

selective AFLP primer in combination with one gene targeted (GT) primer, collectively called GT-AFLP. Our results have demonstrated that this gene targeted AFLP-based strategy allows high-throughput, low cost detection of polymorphic functional markers in cotton. Presentations of this work have been made in different professional meetings and published recently (Lu et al., 2006; Niu et al., 2006; Wang et al., 2007a, b; Zhang et al., 2006a, b; Zhang et al., 2007a, b). The objective of this paper is to briefly review the methodologies and applications of GT-AFLP analysis in cotton.

MATERIALS AND METHODS

Plant materials. For TF-AFLP analysis, 88 genotypes/accessions of the five tetraploid cotton species (*Gossypium hirsutum-Gh*, AD1; G. *barbadense*-Gb, AD2; G. *tomentosum*-Gt, AD3, G. *mustelinum*-Gm, AD4; and G. *darwinii*-Gd, AD5) were grown in either a greenhouse or a field in Las Cruces, NM. However, only 32 genotypes (see Fig. 1) were used for a comprehensive analysis and D5 (G. *raimondii*) was used as an out-group.

For RGA-AFLP analysis, eight genotypes were used, including four G. *hirsutum* cottons (TM-1, NM 24016, Acala 1517-99 and Acala Nem-X), and four G. barbadense cottons (SxP, Amsak, Pima 32, and Pima Phytogen 76). Among the four G. *hirsutum* cotton genotypes, TM-1 is the Upland cotton genetic standard, while the other three are Acala cottons with substantial germplasm introgression from G. barbadense cotton (Pima). For mapping RGA-AFLP markers, a recombinant inbred line (RIL) population derived from NM 24016 and Pima 3-79 were used (Lu et al., 2004).

For PPR-AFLP analysis, a backcrossing population was grown in the greenhouse. The population was produced from a cross between D_8 R, a CMS-D_8 fertility restorer line carrying the D8 restorer gene Rf_2 and SG 747, a sterility maintainer line with normal AD1 cytoplasm and recessive restorer allele rf_2 rf_2

.

DNA and RNA extraction. Leaf tissue was used for DNA extraction using a mini-prep method of Zhang and Stewart (2000). Leaves from Acala 1517-99, SG 747 and Pima Phytogen 76 and 10 days post-anthesis (DPA) bolls from 64 lines from a backcross inbred line population (BIL) were harvested for RNA extraction using a modified hot-borate method (Wan and Wilkins, 1994).

AFLP analysis. The AFLP analysis of DNA was performed as outlined by Vos et al. (1995) with minor modifications (Zhang et al., 2005). Briefly, genomic DNA was restricted with EcoRI and MseI, and ligated with EcoRI and MseI adaptors in the same reaction. The diluted, ligated solution was used in the first round of AFLP amplification using two pre-selective primers with a single selective nucleotide extension. Then, the second round of amplification was performed using the diluted pre-selective PCR reaction as a template with two selective AFLP primers. See Zhang et al. (2005) for details on PCR conditions for AFLP and the following GT-AFLP analysis. The PCR products were analyzed and sequenced using a CEQ 8000 Sequencer and Fragment Analysis Software (Beckman-Coulter Inc., Fullerton, CA), as described in Zhang et al. (2005).

TF-AFLP analysis. For DNA TF-AFLP analysis, one of the degenerate TF primers (Table 1) was used in combination with one of the selective AFLP primers in the second round of PCR amplification. The GT-AFLP PCR products were electrophoresed and analyzed using a CEQ 8000 Sequencer (Zhang et al., 2005). RGA-AFLP analysis. After the first round of AFLP

amplification using two pre-selective primers (pre-selective amplification), one of the degenerate RGA primers (Table 2) was used in combination with one of selective AFLP primers in the second round of PCR amplification and the PCR products were separated by electrophoresis through 5% (v/v) polyacrylamide gel followed by silver staining. PPR-AFLP analysis. PPR-AFLP analysis was the same as a regular AFLP, except that a PPR primer (Table 3) substituted an AFLP selective primer in the selective amplification (SA) reaction. All PCR reactions were performed in a PE Applied Biosystems Gene Amp PCR System 9700 (Applied Biosystem, Fostcil, CA, USA), and the PCR products were separated by electrophoresis through 5% (v/v) polyacrylamide gel followed by silver staining.

Data analysis. The AFLP and GT-AFLP markers were scored as present (1) or absent (0). To estimate the genetic similarities among genotypes a genetic distance matrix based on the Jaccard coefficient was used in the NTSYSpc, Numerical Taxonomy System, Version 2.1 (Exeter Software, Setauket, New York, USA). Phylogenetic trees were constructed using the unweighted pair group method of arithmetic means (UPGMA). This program was used to group genotypes that are genetically related to each other based on the genetic similarity matrix. 9 Linkage analysis. MAPMAKER/EXP, version 3.0 software (Lander et al., 1987) was used to calculate the genetic distance between Rf2 and associated markers and to construct the linkage map. The threshold LOD score was 3.0 and the maximum distance between markers was set at 50.0 cM. The Kosambi function was used to obtain genetic distance in centiMorgans (cM).

RESULTS AND DISCUSSION

Assessment of germplasm diversity and classification based on GT-AFLP. Thirty-two genotypes representing five species were used to compare the precision and efficacy of conventional AFLP and GT-AFLP in classifying the genotypes to species. It was recognized that although different restriction enzymes have been used in numerous AFLP analyses, conventional AFLP markers are anonymous in nature and many may reside in non-gene regions. To target gene sequences and create GT-AFLP, we substituted one selective AFLP primer with a degenerate gene family targeted (GT) primer in the second round of AFLP amplification (GT-AFLP). For TF-AFLP analysis, we selected seven transcription factor (TF) families (e.g., bZIP, bHLH, MYB, WRKY, NAC, MADS, and G2) and designed 10 degenerate primers from conserved regions. Twelve TF-AFLP primer combinations amplified a total of 645 fragments (29-103 fragments with an average of 54 fragments per primer combination) as visualized by capillary electrophoresis using the CEQ 8000 Sequencer. It is understandable that the average number of TF-AFLP markers is only 1/3 of that for AFLP (118-262 fragments with an average of 179 fragments per primer combination), since the TF-AFLP marker system specifically targets transcription factor genes, unlike AFLP which randomly amplifies the genome.

Using eight conventional AFLP primer combinations, a total of 1,434 fragments (179 fragments per primer combination) were resolved by the capillary electrophoresis system. The number of AFLP markers was sufficient to separate the 32 genotypes into five groups (Fig. 1), consistent with their origins and known species relationships. Accessions from AD1, AD3, and AD4 were grouped into separate groups without ambiguity. Even though the majority of AD2 and AD5 were grouped into two groups, a few accessions including two AD2 and one AD5 were placed into the same group and one AD2 accession was grouped with the AD5 group. These may represent hybrids or could be the result of mis-identification. The results also confirmed that AD1 is closer to AD3 than other species, while AD4 is the most distant species among the five tetraploid species.

With the TF-AFLP marker system, cluster analysis revealed five major species groups (Fig. 2) similar to AFLP markers (Fig. 1). All the AD1 accessions were grouped together. While the majority of AD3 and AD4 accessions were grouped into their own species groups, several genotypes were spread to several places in the dendrogram, perhaps attributable to the limited number of TF-AFLP markers. Since AD2 and AD5 accessions were closely related sister species, as expected, two accessions from AD2 and two accessions from AD5 were grouped into their sister species group (Fig. 2). A good separation of closely related accessions probably will require a greater number of TF-AFLP markers than reported here. However, the consistent amplification results obtained here suggest that the TF-AFLP can be used as a new marker system for genotyping, candidate gene mapping, and functional genomics. Further study will be needed to estimate the percentage of TF-AFLP markers that are amplified from the corresponding TF gene families and to map their genome locations in cotton.

Assessment of germplasm diversity and classification based on RGA-AFLP.

Eight genotypes from two species were compared using conventional and RGA-AFLP analyses. The RGA-AFLP analysis used one degenerate RGA primer designed from the regions of the R genes including NBS and LRR domains in combination with one selective AFLP primer in a PCR reaction. Of a total of 446 RGA-AFLP markers amplified by 22 RGA and AFLP primer combinations, 76 (17.0%) and 37 (8.3%) were polymorphic within four Upland cotton (G. *hirsutum*) genotypes and four Pima cotton (G. *barbadense*) genotypes, respectively. The polymorphism of RGA-AFLP was comparable with that of AFLP. The genetic similarity between the eight genotypes based on RGA-AFLP was highly correlated with that measured by AFLP ($r = 0.896$, $P = 0.01$), leading to similar results in genotype grouping at the species and sub-species levels (Figs. 3 and 4). This demonstrates that RGA-AFLP can representatively sample cotton genome and is a powerful and reliable marker system. It offers great flexibility for numerous primer combinations to a genome-wide search for RGAs. Due to the distribution of RGAs or RGA clusters on the plant genome, a genome-wide RGA-AFLP analysis will not only provide useful resources for candidate gene mapping of R genes or QTLs for disease resistances, RGA-AFLP markers could also serve as chromosomal anchors for mapping other traits.

Genome-wide GT-AFLP mapping. Degenerate primers designed based on conserved motifs of R genes were used in combination with AFLP techniques to analyze RGA in a recombinant inbred line (RIL) population developed from a cross between Pima (G. *barbadense*) 3–79 and the Upland cotton (G. *hirsutum*) line NM 24016. Eighty-eight polymorphic RGA markers were amplified by 8 pairs of RGA degenerate primers, while 131 polymorphic RGA-AFLP markers were produced using 6 RGA and AFLP primer pairs. Based on SSR markers with known chromosomal locations, 217 RGA and RGA-AFLP markers were assigned to 18 chromosomes using programs Mapmaker (Niu et al., 2006). Interestingly, these RGA and RGA-AFLP markers were not evenly distributed among chromosomes, in that 189 of the 217 markers (87%) were placed on three "giant" linkage groups (C6, C12, and C15). The results confirm that RGA in cotton tend to cluster together. Our work demonstrates that RGA-AFLP is a powerful DNA marker system for identifying resistance gene analogous with more sequence specificity. Using more RGA and AFLP primer combinations, most, if not all, RGA can be amplified and identified genome-wide in cotton. The identification and mapping of RGA and RGA-AFLP markers should provide a framework to facilitate marker-assisted selection (MAS) for disease resistance in cotton breeding and to understand the genome organization of R genes in cotton.

Candidate gene mapping. Petunia *Rf* (Bentolila et al., 2002), radish *Rfo* (Brown et al., 2003) and rice *Rf-1* (Komori et al., 2004; Wang et al., 2006) all encode a PPR protein with 2 to 26 PPR motifs organized in a tandem array. The PPR proteins target to organelle mitochondria or chloroplasts, and contain a RNA-binding structure (Small and Peeters, 2000). There is evidence that an *Rf* gene can change the RNA transcript profile of CMS associated genes. In rice, two restorer genes $Rf1_a$ and $Rf1_b$ encode PPR proteins (Wang et al., 2006). Both RF1A and RF1B proteins can silence the mRNA of CMS associated novel mitochondrial gene *B-atp6/orf79* by different mechanisms. The RF1A cleaves the mRNA while RF1B degrades the mRNA (Wang et al., 2006). In our study, seven primers were designed from conserved motifs of pentatricopeptide repeat (PPR) proteins (Lurin et al., 2004) of *Arabidopsis thaliana* (Table 3). Those PPR primers were combined with AFLP primers for the identification of candidate PPR protein genes that are associated with Rf_2 . Primer combination of PPR-S/AFLP-M8 amplified a polymorphic 155-bp fragment only present in the fertile plants (Fig. 5). At least two other primer combinations (PPR-E+/AFLP-E2 and PPR-S/AFLP-E8) were found to produce faint polymorphic bands between eight fertile and eight sterile plants (Fig. 5). But the reproducibility was not consistent, indicating a need to redesign primers or change PCR conditions. The PPR-AFLP marker PPR-S-M8 was found to be linked to Rf_2 gene (Fig. 6) and its genetic distance to UBC188-500 marker was similar to that between Rf_2 and UBC188 as reported by Zhang and Stewart (2004) and Zhang et al. (2005). This indicates that the PPR-AFLP marker PPR-S-M8 might be a candidate for Rf2 and the cotton Rf_2 gene may encode a PPR protein. However, only high resolution mapping and cloning of Rf_2 will confirm these conjectures.

Gene expression analysis. Recent reports of RGA studies in cotton were focused on a limited number of PCR amplification products for isolation and sequencing of putative R genes from genomic DNA. Using degenerate primers to amplify cotton genomic DNA, Tan et al. (2003) cloned 33 putative cotton RGAs containing the highly conserved NBS R-protein motif. He et al. (2004) utilized degenerate primers designed from NBS and membrane-spanning motifs and identified 61 RGA sequences. Hinchliffe et al. (2005) isolated 57 RGA sequences and mapped nine RGA-STS markers to two homeologous chromosomes in cultivated tetraploid cotton. However, the strategy of using degenerate primers to directly amplify cotton R genes has not been reported. The cDNA RGA-AFLP strategy will be a more straightforward way to identifying R genes and study their expression profiles.

In an attempt to use degenerate primers to amplify putative R genes, cDNA from three cotton genotypes (Acala 1517-99, Pima Phy 76 and SG 747) were amplified using primers designed from disease resistance gene analogues (RGA) or combined with AFLP primers (RGA-AFLP). 810 fragments were cloned and sequenced. Cluster analysis was used to group the expressed RGA and RGA-AFLP into different categories. About 90% of the fragments were homologous to cotton ESTs deposited in Genebank and many isolated fragments were also homologous, rendering an opportunity for SNP identification (Niu et al., 2007). About 10% of the fragments were novel, representing new genes that have not been isolated before. Many cDNA RGA and RGA-AFLP fragments were found to be putative disease resistant, abiotic stress related or similar to nucleotide binding sequences, indicating the usefulness of the two analysis systems (RGA and RGA-AFLP).

The reliability of cDNA GT-AFLP analysis can be indirectly judged from our previous analysis using an arbitrary TRAP primer (Hu and Vick, 2003) in combination with a degenerate TF primer (Zhang et al., 2007b). PCR and electrophoresis were performed on a BIL population of 64 cDNA samples. Different primer combinations with the same random primer but different TF primers produced PCR products with different banding patterns. The high reliability of the TF-TRAP system can be gauged from the high proportion of monomorphic

bands amplified by the same primer combination in all the 64 cDNA samples. It is apparent that there existed many fragments that were polymorphic in the BIL population, while many other fragments showed intensity differences that might reflect the differential expression of corresponding genes among the BIL lines. The results demonstrated that degenerate TF primers can be used in combination with a fixed primer for genome-wide profiling of TF gene expression.

REFERENCES

Bentolia, S., A.A Alfonso, and M.R. Hanson. 2002. A pentatricopeptide repeat-containing gene restores fertility to male sterile plants. Proc. Natl. Acad. Sci. USA. 99: 10887-10892.

Brown, G.G., N. Formanova, H. Jin, R. Wargachuk, C. Dendy, P. Patil, M. Laforest, J. Zhang, W.Y. Cheung, and B.S. Landry. 2003. The radish Rf_0 restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. Plant J. 35: 262-272.

Chase, C.D. 2007. Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. Trends Genet. 23: 81-90.

Gong, W., Y.P. Shen, L.G. Ma, Y. Pan, Y.L. Du, et al. 2004. Genome-wide ORFeome cloning and analysis of *Arabidopsis* transcription factor genes. Plant Physiol. 135: 773-782.

He, L., C. Du, L. Covaleda, Z. Xu, A.F. Robinson, J.Z. Yu, R.J. Kohel, and H.B. Zhang. 2004. Cloning, characterization, and evolution of the NBS-LRR-encoding resistance gene analogue family in polyploid cotton (*Gossypium hirsutum* L.). Mol. Plant Microbe Interact. 17: 1234–1241.

Hinchliffe, D.J., Y. Lu, C. Potenza, C. Segupta-Gopalan, R.G. Cantrell, and J.F. Zhang. 2005. Resistance gene analogue markers are mapped to homeologous chromosomes in cultivated tetraploid cotton. Theor. Appl. Genet. 110: 1074-1085.

Hu, J., and B.A. Vick. 2003. TRAP (target region amplification polymorphism), a novelmarker technique for plant genotyping. Plant Mol. Biol. Rep. 21: 289–294.

Komori, T., S. Ohta, N. Murai, Y. Takakura, Y. Kuraya, S. Suzuki, Y. Hiei, H. Imaseki, and N. Nitta. 2004. Map-based cloning of a fertility restorer gene, *Rf-$_1$*, in rice (*Oryza sativa* L.). Plant J. 37: 315-325.

Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln, and L. Newburg. 1987. MAPMAKER: an interactive computer package for construction primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181.

Lu, Y., M. Pang, S. Higbie, D. Hinchliffe, C. Niu, Y. Yuan, S. Yu, and J.F. Zhang. 2006. Development of high throughput DNA marker systems in cotton based on gene and regulatory sequences. p. 828. *In* Proc. Beltwide Cotton Conf., San Antonio, TX. 3-6 Jan. 2006. Natl. Cotton Counc. Amer., Memphis, TN.

Lu, Y., J.F. Zhang, R.G. Percy, and R.G. Cantrell. 2004. An integrated SSR-STS-AFLP-SRAP genetic map using recombinant inbred line populations in tetraploid cottons. p. 1156-1161.

In Proc. Beltwide Cotton Conf., San Antonio, TX. 5-9 Jan. 2004. Natl. Cotton Counc. Amer., Memphis, TN.

Lurin, C, C. Andres, S. Aubourg, M. Bellaoui, F. Bitton, C. Bruyere, M. Caboche, C. Debast, J. Gualberto, B. Hoffmann, A. Lecharny, M. Le Ret, M.L. Martin-Magniette, H. Mireau, N. Peeters, J.P. Renou, B. Szurek, L. Taconnat, and I. Small. 2004. Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell 16: 2089-2103.

Martin, G.B., A.J. Bogdanove, and G. Sessa. 2003. Understanding the functions of plant disease resistance proteins. Annu. Rev. Plant Physiol. Plant Mol. Biol. 54: 23–61.

Meyers, B.C., A. Kozik, A. Griego, H. Kuang, and R.W. Michelmore. 2003. Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell 15: 809–834.

Monosi, B., R.J. Wisser, L. Pennill, and S.H. Hulbert. 2004. Full-genome analysis of resistance gene homologues in rice. Theor. Appl. Genet. 109: 1434–1447.

Niu, C., M. Pang, and J.F. Zhang. 2007. Disease resistance gene analogues in cotton: expression and sequence analysis. p. 244-251. *In* Proc. Beltwide Cotton Conf., New Orleans, LA. 10-12 Jan. 2007. Natl. Cotton Counc. Amer., Memphis, TN.

Niu, C., and J.F. Zhang. 2006. Plant disease resistance gene analogues in cotton: mapping and expression. p. 924. *In* Proc. Beltwide Cotton Conf., San Antonio, TX. 3-6 Jan. 2006. Natl. Cotton Counc. Amer., Memphis, TN.

Small, I.D., and N. Peeters. 2000. The PPR motif—a TPR-related motif prevalent in plant organellar proteins. Trends Biochem. Sci. 25: 46–47.

Tan, H., F.E. Callahan, X.D. Zhang, M. Karaca, S. Saha, J.N. Jenkins, R.G. Creech, and D.P. Ma. 2003. Identification of resistance gene analogs in cotton (*Gossypium hirsutum* L.). Euphytica 134: 1–7.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23: 4407-4414.

Wan, C.H., and T.A. Wilkins. 1994. A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). Anal. Biochem. 223: 7-12.

Wang, F., J. McD. Stewart, and J.F. Zhang. 2007a. Molecular markers linked to the Rf_2 fertility restorer gene in cotton. Genome 50: 818-824.

Wang, F., M. O'Connell, J. McD. Stewart, and J.F. Zhang. 2007b. Mapping of restorer gene, Rf2 and RFLP analysis of MtDNA in CMS-D8 cotton. p. 2106. *In* Proc. Beltwide Cotton Conf., New Orleans, LA. 10-12 Jan. 2007. Natl. Cotton Counc. Amer., Memphis, TN.

Wang, Z., Y. Zou, X. Li, Q. Zhang, L. Chen, H. Wu, D. Su, Y. Chen, J. Guo, D. Luo, Y. Long, Y. Zhong, and Y.G. Liu. 2006. Cytoplasmic male sterility of rice with Boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. Plant Cell 18: 676-687.

Xiong, Y., T. Liu, C. Tian, S. Sun, J. Li, and M. Chen. 2005. Transcription factors in rice: a genome-wide comparative analysis between monocots and eudicots. Plant Mol. Biol. 59: 191-203.

Zhang, J.F., and J. McD. Stewart. 2000. Economic and rapid method for extracting cotton genomic DNA. J. Cotton Sci. 4: 193-201.

Zhang, J.F., and J. McD. Stewart. 2004. Identification of molecular markers linked to the fertility restorer genes for CMS-D8 in cotton (*Gossypium hirsutum* L.). Crop Sci. 44: 1209-1217.

Zhang, J.F., and J. McD. Stewart. 2005. Linkage analysis between gametophytic restorer Rf_2 gene and genetic markers in cotton. Crop Sci. 45: 147-156.

Zhang J.F., Y. Lu, M. Pang, S. Higbie, D. Hinchliffe, C. Niu, R.G. Percy, and R.G. Cantrell. 2006b. Development of high throughput DNA marker systems in cotton based on gene and regulatory sequences. Plant & Animal Genomes XIV Conf., San Diego, CA. 14-18 Jan. 2006.

Zhang, J.F., Y.Z. Lu, and S.X. Yu. 2005. Cleaved AFLP (cAFLP), a modified amplified fragment length polymorphism analysis for cotton. Theor. Appl. Genet. 111: 1385-1395.

Zhang, J.F., M. Pang, C. Niu, W. Wang, F. Wang, D. Miranda, J. Curtiss, Y. Yuan, and S. Yu. 2006a. Development of high throughput functional DNA markers based on gene families. ASA Annual Meeting, Indianapolis, IN. 12-16 Nov. 2006.

Zhang, J.F., W. Wang, R. Esmail, M. Pang, R. G. Percy, J. McD. Stewart, B. Yue, and J. Hu. 2007b. Genetic diversity of tetraploid cotton species based on AFLP and GT-AFLP analysis. p. 235-243. *In* Proc. Beltwide Cotton Conf., New Orleans, LA. 10-12 Jan. 2007. Natl. Cotton Counc. Am., Memphis, TN.

Zhang, J.F., Y. Yuan, C. Niu, D.J. Hinchliffe, Y. Lu, S. Yu, R.G. Percy, and R.G. Cantrell. 2007a. AFLP-RGA markers in comparison with RGA and AFLP in cultivated tetraploid cotton. Crop Sci. 47: 180-187.

Table 1. Degenerate transcription factor (TF) gene primers.

Primer	Sequence
bHLH-1	NNNGARMGINNNMGIMGIGAR
bZIP-1	AAYMGIGARTCNGCNNNNMGIAGY
bZIP-2	NNNMGIAGYMGINNNMGIAAR
G2-1	AGYCAYCTNCARAARTAYAGR
MADS-1	NNNAARMGIMGINNNGGNTTR
MADS-2	GARTTRATNNNNTTRTGYMGI
Myb-1	GGNAAGTCNTCYMGIYTNMGIT
Myb-2	CCNGGNMGICANGAYAAYGAA
NAC-1	NNNTGGNNNATGCAYGARTAY
WRKY-1	TGGMGIAARTAYGGNCARAAG

Table 2. Degenerate primers used to amplify putative resistance gene analogue (RGA) and AFLP-RGA markers.

Type	Primer	Sequence
Kinase	Pto kin-1	GCATTGGAACAAGGTGAA
	Pto kin-2	AGGGGGACCACCACGTAG
	RLK-for	GAYGTNAARCCIGARAA
	RLK-rev	TCYGGYGCRATRTANCCNGGITGICC
NBS	NBS-F1	GGAATGGGNGGNGTNGGNAARAC
	NBS-R1	YCTAGTTGTRAYDATDAYYYTRC
GGBGKTT GLPLAL	S2	GGIGGIGTIGGIAAIACIAC
	AS3	IAGIGCIAGIGGIAGICC
LRR	CLRR-for	TTTTCGTGTTCAACGACG
	CLRR-rev	TAACGTCTATCGACTTCT
	RLRR-for	CGCAACCACTAGAGTAAC
	RLRR-rev	ACACTGGTCCATGAGGTT
	NLRR-for	TAGGGCCTCTTGCATCGT
	NLRR-rev	TATAAAAAGTGCCGGACT
	XLRR-for	CCGTTGGACAGGAAGGAG
	XLRR-rev	CCCATAGACCGGACTGTT

Table 3. Sequences and relative parameters of PPR motif primers.

Primer	Sequence
PPR-E	TAYGNCTNCTNCTNAAYNNNTAYGCN
PPR-E+	CAYGARTTYGTNGCNGGNGAY
PPR-L2	GTNTCYTCNCAYNNNGGNYTNGTN
PPR-P	ATGAARGARAARGGNATNNNNCC
PPR-S	ATGTAYGCNAARTGTGGN
DYW-1	CAYTCNGARAARCTNGCNNNNGCNTTYGGNCT
DYW-2	TGYTCNTGYNNNGAYTAYTGG

FIGURE CAPTIONS

Figure 1. An UPGMA dendrogram based on AFLP markers. D-D5, PI 530947; Gh(A2)-Unknown race stock; Gh01- Pima NM020032-7; Gh02- Acala 1517-99; Gh05- Acala Phy 72; Gh06- Acala NemX; Gt01- AD3, unknown origin; Gt03- AD3 A29-1236; Gt04- AD3 Ma+Pa; Gt04- AD3 Ma+Pa 1; Gb02- Ashmouni; Gb21- Giza 80; Gb23- Giza 88; Gb25- Pima 32; Gb26- Pima S-1; Gb27- Pima S-7; Gb29- Xinhai, China; Gb31- Unknown origin; Gd01- AD5, PI 499696; Gd02- AD5, unknown origin; Gd03- AD5-03; Gd04- AD5-06; Gd05- AD5-07; Gm01- AD4-01; Gm02- AD4-06; Gm03- AD4-08; Gm04- AD4-09; Gm05- AD4-10; Gm07-AD4-11; Gm09- AD4-17; Gm10- AD4-22; Gm17- AD4-X.

Figure 2. An UPGMA dendrogram based on TF-AFLP markers. D-D5, PI 530947; Gh(A2)-Unknown race stock; Gh01- Pima NM020032-7; Gh02- Acala 1517-99; Gh05- Acala Phy 72; Gh06- Acala NemX; Gt01- AD3, unknown origin; Gt03- AD3 A29-1236; Gt04- AD3 Ma+Pa; Gt04- AD3 Ma+Pa 1; Gb02- Ashmouni; Gb21- Giza 80; Gb23- Giza 88; Gb25- Pima 32; Gb26- Pima S-1; Gb27- Pima S-7; Gb29- Xinhai, China; Gb31- Unknown origin; Gd01- AD5, PI 499696; Gd02- AD5, unknown origin; Gd03- AD5-03; Gd04- AD5-06; Gd05- AD5-07; Gm01- AD4-01; Gm02- AD4-06; Gm03- AD4-08; Gm04- AD4-09; Gm05- AD4-10; Gm07-AD4-11; Gm09- AD4-17; Gm10- AD4-22; Gm17- AD4-X.

Figure 3. An UPGMA dendrogram based on AFLP-RGA markers.

Figure 4. An UPGMA dendrogram based on AFLP markers.

Figure 5. PPR-AFLP marker, PPR-S-M8-155 (A) linked to Rf_2 and three other Rf_2 associated PPR-AFLP markers amplified by primers PPR-E+/AFLP-E2 (B) and PPR-S/AFLP-E8 (C) as indicated by arrows.

Figure 6. Locations of markers flanking the Rf2 locus on a linkage group. Genetic distance in centimorgans(cM) between adjacent markers were estimated based on the analysis of 112 backcross plants.

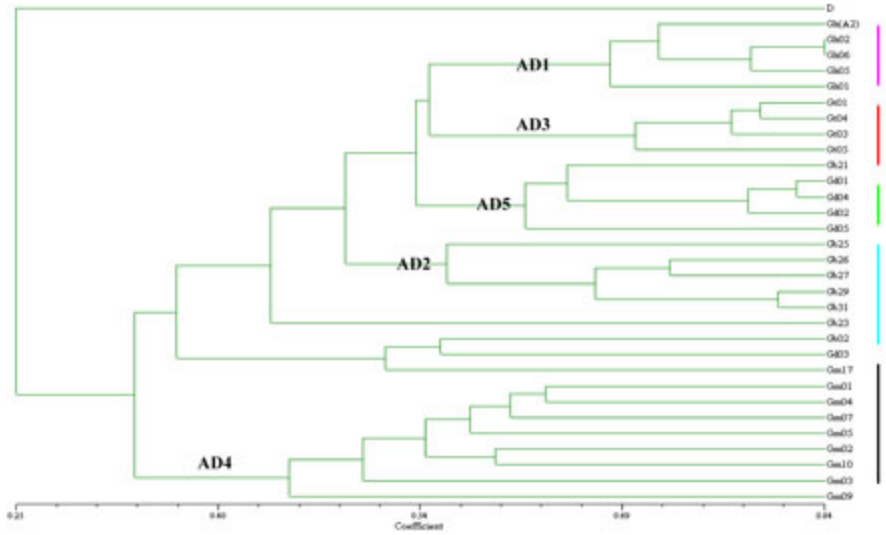

Figure 1

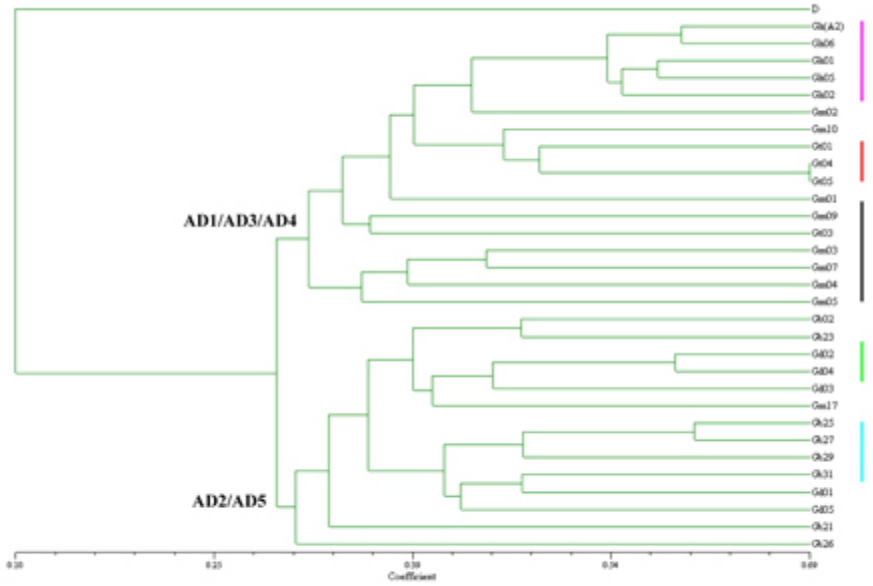

Figure 2

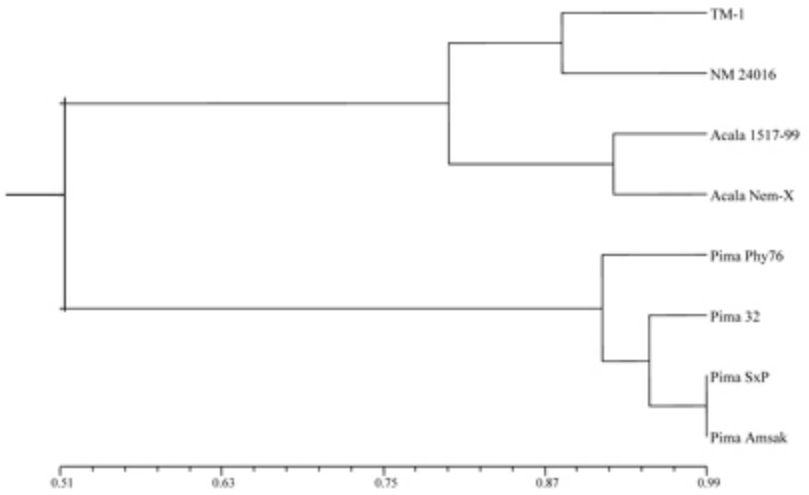

Figure 3

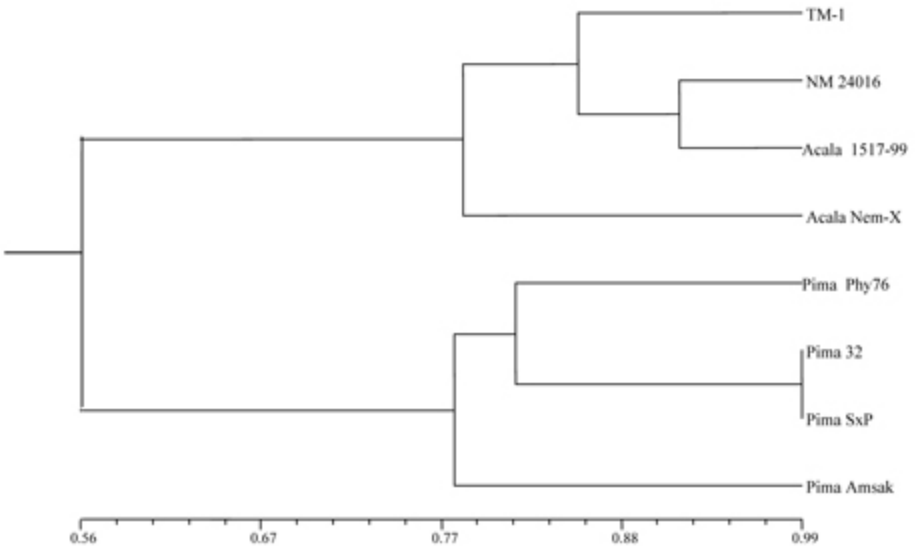

Figure 4

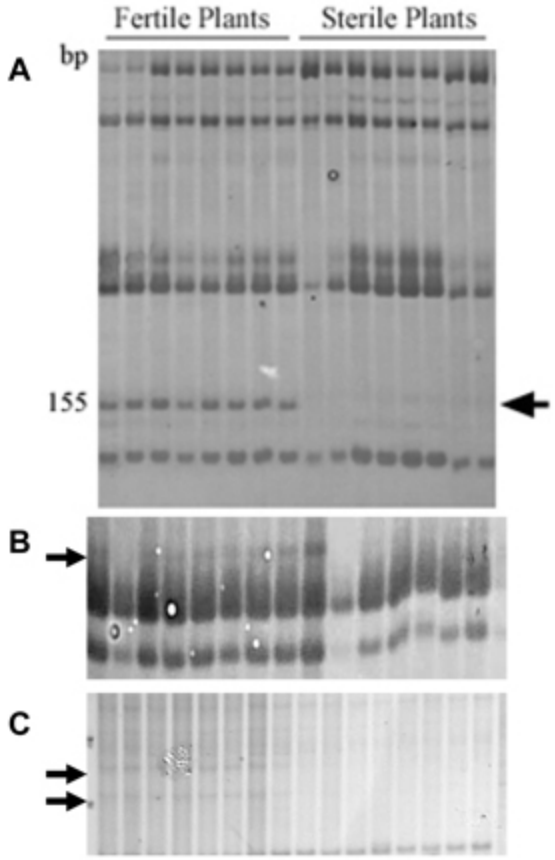

Figure 5

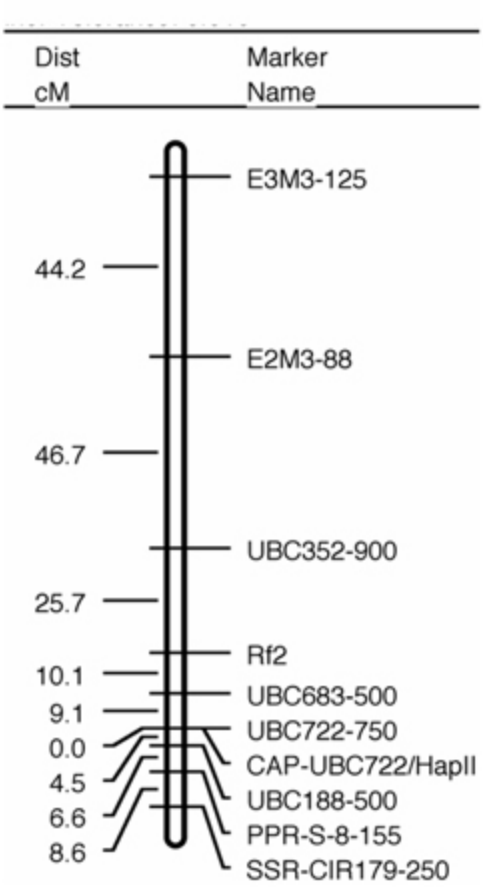
Dist
cM
Marker
Name
E3M3-125
44.2
E2M3-88
46.7
UBC352-900
25.7
Rf2
10.1
UBC683-500
9.1
UBC722-750
0.0
CAP-UBC722/HapII
4.5
UBC188-500
6.6
PPR-S-8-155
8.6
SSR-CIR179-250

Figure 6