Theor Appl Genet (2001) 102:222–229

A.M. Abdalla · O.U.K. Reddy · K.M. El-Zik
A.E. Pepper

Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP

Received: 17 January 2000 / Accepted: 4 May 2000

Abstract *Gossypium* species (± 49) represent a vast resource of genetic diversity for the improvement of cultivated cotton. To determine intra- and inter-specific genetic relationships within a diverse collection of *Gossypium* taxa, we employed 16 AFLP primer combinations on three diploid species, *Gossypium herbaceum* L. (A1), *Gossypium arboreum* L. (A2) and *Gossypium raimondii* Ulbrich (D5), and 26 AD allotetraploid accessions (*Gossypium barbadense* L. and *Gossypium hirsutum* L.). A total of 1180 major AFLP bands were observed; 368 of these (31%) were polymorphic. Genetic similarities among all taxa ranged from 0.21 (between the diploid species *G. arboreum* and *G. raimondii*) up to 0.89 (within *G. barbadense*). Phenetic trees based on genetic similarities (UPGMA, N-J) were consistent with known taxonomic relationships. In some cases, well-supported phylogenetic relationships, as well as evidence of genetic reticulation, could also be inferred. UPGMA trees and principal coordinate analysis based on genetic similarity matrices were used to identify genetically distinct cultivars that are potentially important sources of germplasm for cotton improvement, particularly of fiber quality traits. We show that AFLP is useful for estimating genetic relationships across a wide range of taxonomic levels, and for analyzing the evolutionary and historical development of cotton cultivars at the genomic level.

Communicated by F. Salamini

The first two authors contributed equally

A.M. Abdalla
Agronomy Department, Faculty of Agriculture,
Cairo University, Giza, Egypt

K.M. El-Zik
Department of Soil and Crop Sciences,
Texas A & M University, College Station, TX 77843–2474, USA

O.U.K. Reddy · A.E. Pepper (✉)
Department of Biology, Texas A & M University,
College Station, TX 77843–3258, USA
e-mail: apepper@bio.tamu.edu
Fax: +1 979-862-4790

Keywords DNA fingerprinting · *Gossypium* species · Genomic evolution · Genetic similarity · Fiber quality

Introduction

As with many crops, the phylogenetic histories of cultivated cottons are reticulate in nature (Rieseberg and Noyes 1998), with individual taxa related to each other by a network, rather than as branches on a tree. The most important point of reticulation was the hybridization of an Old World 'A' genome taxon [related to extant species *Gossypium herbaceum* L. and *Gossypium arboreum* L. (2n = 2*x* = 26)] with a New World 'D' genome taxon [related to *Gossypium raimondii* Ulbrich and *G. gossipioides* L. (2n = 2*x* = 26)], followed by polyploidization to give rise to an 'AD' disomic tetraploid cotton (2n = 4*x* = 52) some 1–2 million years before thepresent (Beasley 1940, 1942; Wendel et al. 1992). The ancestral AD allotetraploid diverged in the New World to give rise to five extant AD tetraploid species including the cultivated cottons *Gossypium hirsutum* L. and *Gossypium barbadense* L., the world's most important textile crops.

The prehistoric distribution of *G. barbadense* was widespread, and included much of South America, southern Mesoamerica and the Caribbean basin (Fryxell 1979). *G. hirsutum* had a distribution throughout Mesoamerica, the shores of the Gulf of Mexico, and westward to the islands of the central Caribbean. Thus, there were significant regions of long-term overlap of *G. barbadense* and *G. hirsutum*. Given the present-day interfertility between the two species, it is likely that hybridization occurred where the two species were sympatric (Wendel et al. 1992). Subsequent reticulation of cotton genomes involved intentional interspecific and inter-cultivar crosses (Endrizzi et al. 1985), and natural outcrossing during field cultivation (which ranges from 1% to 32% depending on location and pollinator availability). As a result of this extensive reticulation, the genetic relationships and genomic composition of many cotton cultivars remain obscure.

The development of New World tetraploid cottons for intensive fiber production occurred during colonial times. 'Sea Island' cotton was developed in the Caribbean basin in the late 18th century through selection of *G. barbadense* for annual habit and day neutral flowering. The selection and cultivation of highly productive *G. hirsutum* 'Upland' cotton cultivars formed the foundation for the cotton-based economies of colonial North America. The involvement of specialized tetraploid cotton in Egypt began in the late 19th century, largely through the introduction of 'Sea Island'-type *G. barbadense* stocks. Egyptian breeding efforts led to the development of several elite *G. barbadense* cultivars, including the 'Giza' cultivars. Germplasm from Egyptian cultivars was utilized in the development of the 'Pima' cultivars (Calhoun et al. 1994), the only *G. barbadense* presently grown on a large scale in North America.

Modern cotton cultivars show significant variation for agriculturally important traits (El-Zik and Thaxton 1989). For example, the longest, strongest and finest cotton fibers are produced by the *G. barbadense* cultivars of the Egyptian, Sea Island, and Pima groups. However, Upland *G. hirsutum* cultivars have earlier maturity, higher yields, and adapt to a wider range of environments. An understanding of the genetic and genomic relationships of extant cotton species and cultivars is critical for the further utilization of cotton genetic diversity and genomic information in the development of superior cultivars that combine the favorable qualities conditioned by this diverse germplasm.

A variety of molecular-marker technologies have been used to study the genetic diversity and relationships of crop species and their wild relatives. Of these methods, random amplified polymorphic DNAs (RAPDs) have been most widely used in cotton (Multani and Lyon 1995; Tatineni et al. 1996; Iqbal et al. 1997). Studies using allozymes (Wendal et al. 1992) and RFLPs (Wendal and Brubaker 1993) have been limited by low levels of polymorphism at the intraspecific and interspecific levels. The amplified fragment length polymorphism (AFLP) method (Zabeau and Vos 1993; Vos et al. 1995) has also been successfully used to analyze genetic diversity among a wide range of crop species and their wild relatives (Hill et al. 1995; Maughan et al. 1996; Powell et al. 1996). The major advantage of AFLP is its power to identify large numbers of potentially polymorphic loci; this property is referred to as a 'high multiplex ratio'. Vroh Bi et al. (1999) recently used AFLP markers to track the introgression of genetic material from a tri-species synthetic hybrid into *G. hirsutum*. The objective of the present study was to determine the utility of AFLP for estimating genetic diversity and phenetic relationships among a diverse collection of 29 accessions of diploid and tetraploid *Gossypium*, and for the identification of genetically unique cultivars as potentially important new sources of alleles for cotton improvement.

Materials and methods

Plant materials

Twentynine accessions from five species of the genus *Gossypium* were included in this study (Table 1). The cultivated tetraploid species were represented by eight *G. hirsutum* (AD1) and 17 *G. barbadense* (AD2) accessions. *G. barbadense* representatives included six Egyptian Giza, five Pima, and six Sea Island cultivars. Diploid cottons were represented by two *G. herbaceum* (A1), one *G. arboreum* (A2), and one *G. raimondii* (D5) accessions. Seed of *G. barbadense* cultivars and leaves of greenhouse-grown perennial diploid species were obtained from the cotton germplasm collection maintained by the USDA-ARS at the College Station, Tex. (Percival 1987). *G. hirsutum* seeds were obtained from the Texas A & M University Multi-Adversity Resistance (MAR) Cotton Program germplasm stocks.

AFLP fingerprinting

Genomic DNAs were isolated from leaf tissue by the method of Iqbal et al. (1997). AFLP fingerprinting (Zabeau and Vos 1993; Vos et al. 1995) was performed using Analysis System I (GIBCO-BRL Life Technologies) according to the manufacturer's supplied protocol, starting with 225 ng of each genomic DNA. Primary amplifications were performed with 'Mix I' on a Perkin-Elmer 9600 thermocycler. Both the *Eco*RI and *Mse*I primers used in secondary amplification had three extra 'selective' nucleotides at the 3' end

Table 1 Description of *Gossypium* taxa studied

Accession number	Taxon	Genome
Diploids		
A1-71	*G. herbaceum* L.	A1
A1-75	*G. herbaceum* L.	A1
A2-7004	*G. arboreum* L.	A2
D5-4	*G. raimondii* Ulbrich	D5
Tetraploids		
G. hirsutum L.		
SA-1583	Tamcot SP37	AD1
SA-1579	Tamcot CAMD-E	AD1
SA-2292	Tamcot Sphinx	AD1
SA-1668	Paymaster HS26	AD1
SA-1512	Deltapine 50	AD1
SA-1047	Acala 4-42	AD1
SA-2288	Acala Maxxa	AD1
SA-1577	Stoneville 112	AD1
G. barbadense L.		
GB-230	Giza 22	AD2
GB-605	Giza 47	AD2
GB-236	Giza 7	AD2
GB-603	Giza 64	AD2
GB-604	Giza 45	AD2
GB-738	Giza 36	AD2
GB-228	Montserrat Sea Island	AD2
GB-229	Barbados Sea Island	AD2
GB-679	Bleak Hall Sea Island	AD2
GB-1049	Seabrook Sea Island	AD2
GB-726	14-15 Bar	AD2
GB-727	14-25-5 Bar	AD2
GB-224	Pima S-1	AD2
GB-851	Pima S-3	AD2
GB-1497	Pima S-5	AD2
GB-1030	Pima S-6	AD2
GB-1023	Pima S-7	AD2

in order to reduce the number of amplified fragments. Primer nomenclature follows that of GIBCO-BRL Life Technologies; for example, the primer E-AGC denotes an *Eco*RI cohesive primer with the three selective nucleotides 5'-AGC-3'. The 16 selective primer combinations used in this study are listed in Table 2. Secondary amplifications containing ^{33}P labeled *Eco*RI primer were also carried out in a Perkin-Elmer 9600 thermocycler. After the addition of an equal volume (10 µl) of manual sequencing dye (98% formamide, 10 mM EDTA, 0.025% xylene cyanol, 0.025% bromophenol blue), the samples were heated at 94°C for 3 min and chilled on ice. A 2-µl aliquot of each sample was electrophoresed on a denaturing 5% polyacrylamide gel containing 7.5 M Urea and a 0.5 × TBE running buffer (45 mM Tris Borate, 1 mM EDTA, pH 8) at 60 W for 1.5 h, in a 30 cm × 40 cm manual sequencing apparatus (GIBCO-BRL Life Technologies). The gel was transferred to Whatman 3MM blotting paper, dried under vacuum, and exposed to X-ray film for up to 48 h.

Data analysis

Since the validity of every polymorphic AFLP band could not be evaluated by an independent method (such as segregation analysis in an F_2 or recombinant-inbred population), only distinct, major, reproducible bands were scored. Minor polymorphic AFLP bands were excluded from the analysis because these can arise artifactually from differences in genomic DNA quality and other factors (Lin and Kuo 1995; Schondelmaier et al. 1996). Presence or absence of each AFLP fragment was scored as a binary unit character (1 = present, 0 = absent). Genetic similarities based on Jaccard's coefficient (Jaccard 1908) were calculated using the SIMQUAL program of the Numerical Taxonomy Multivariate Analysis System (NTSYS-pc) Version 2.0 software package (Rohlf 1993). The resulting genetic similarity matrices were used to generate an unweighted pair group method of arithmetic means (UPGMA) trees (Sokal and Michener 1958) using the NTSYS-pc or the UPGMA tree searching algorithm of the Phylogenetic Analysis Using Parsimony (*and Other Methods) (PAUP*4.0b2a) software package (Swofford 1998). Secondary analyses were performed using the Neighbor Joining (N-J) method (Saitou and Nei 1987) in the PAUP*4.0b2a Software package. The robustness of phenetic and phylogenetic trees was evaluated by comparing topologies generated by different methods, and by bootstrapping (Felsenstein 1985) using PAUP*4.0b2a. Principal coordinate analysis (PCA) based on genetic similarity matrices was performed using the DCENTER and EIGEN algorithms of the NTSYS-pc software package (Rohlf 1993).

Results

Sixteen primer combinations were tested for selective amplification of DNA fragments from a diverse collection of 29 genotypes representing five *Gossypium* species. A total of 1180 major AFLP bands were observed; 368 of these (31%) were polymorphic within the collection (Table 2). The number of polymorphic bands generated by individual primer pairs ranged from 9 (E-AAC/M-CAC) to 33 (E-AAG/M-CTG and E-AGG/M-CTG). Within the tetraploid cotton cultivars, the primer combination E-ACC/M-CTG produced the greatest number of polymorphic products (24 in total). Out of the 368 bands that were polymorphic in our collection, 143 were shared between a subset of tetraploid cotton cultivars and a subset of A-genome accessions (*G. herbaceum* or *G. arboreum*). We have designated these as 'A-related markers.' Eighty four markers were shared between the tetraploids and *G. raimondii*; these were designated as 'D-related markers'.

Genetic similarities among all taxa ranged from 0.21 (between *G. arboreum* and *G. raimondii*) to 0.72 (between *G. barbadense* and *G. hirsutum*), with a mean similarity of 0.54. Mean genetic similarities within and between taxonomic species are shown in Table 3. The mean genetic similarity within our geographically and historically diverse set of *G. barbadense* accessions (0.89) was similar to our *G. hirsutum* collection (0.86). However, the number of *G. barbadense* accessions sampled was larger, and the total genetic diversity was greater than that of *G.*

Table 2 Polymorphism detected by AFLP in five cotton species

Primers pairs	Total amplified bands	Polymorphic bands	Number of bands that are polymorphic within the indicated genome group				
			Tetraploids	AD1	AD2	Diploids	A
E-AGG/M-CTG[a]	102	33	16	6	11	24	10
E-AGG/M-CAT	63	17	4	2	1	6	3
E-AGG/M-CTG	51	12	8	7	2	8	0
E-AGC/M-CAA	69	24	10	3	7	21	8
E-AGC/M-CTG	48	27	15	6	11	16	6
E-AGC/M-CAC	55	17	15	11	14	11	3
E-AGC/M-CTA	110	29	14	2	12	21	5
E-AAG/M-CAC	66	19	13	0	13	11	3
E-AAG/M-CTG	80	33	18	11	9	15	5
E-AAG/M-CAT	64	21	3	2	1	14	6
E-AAG/M-CTA	104	31	18	5	16	20	4
E-ACC/M-CAT	39	15	14	9	12	11	2
E-ACC/M-CTG	97	28	24	15	22	15	1
E-AAC/M-CAT	109	28	10	7	2	19	7
E-AAC/M-CTG	70	25	6	3	1	18	9
E-AAC/M-CAC	53	9	2	0	0	7	2
Mean	73.8	23	11.9	5.6	8.4	14.8	4.6
Total	1180	368	190	89	134	237	74

[a] E = *Eco*R1, M = *Mse*1

Table 3 Estimated genetic similarities within and among *Gossypium* species based on AFLP

Gossypium spp.	*G. barb.*	*G. hirs.*	*G. herb.*	*G. arbo.*	*G. raimo.*
G. barbadense	0.89	0.72	0.42	0.43	0.34
G. hirsutum		0.86	0.41	0.41	0.34
G. herbaceum			0.94	0.66	0.22
G. arboreum				NA	0.21
G. raimondii					NA

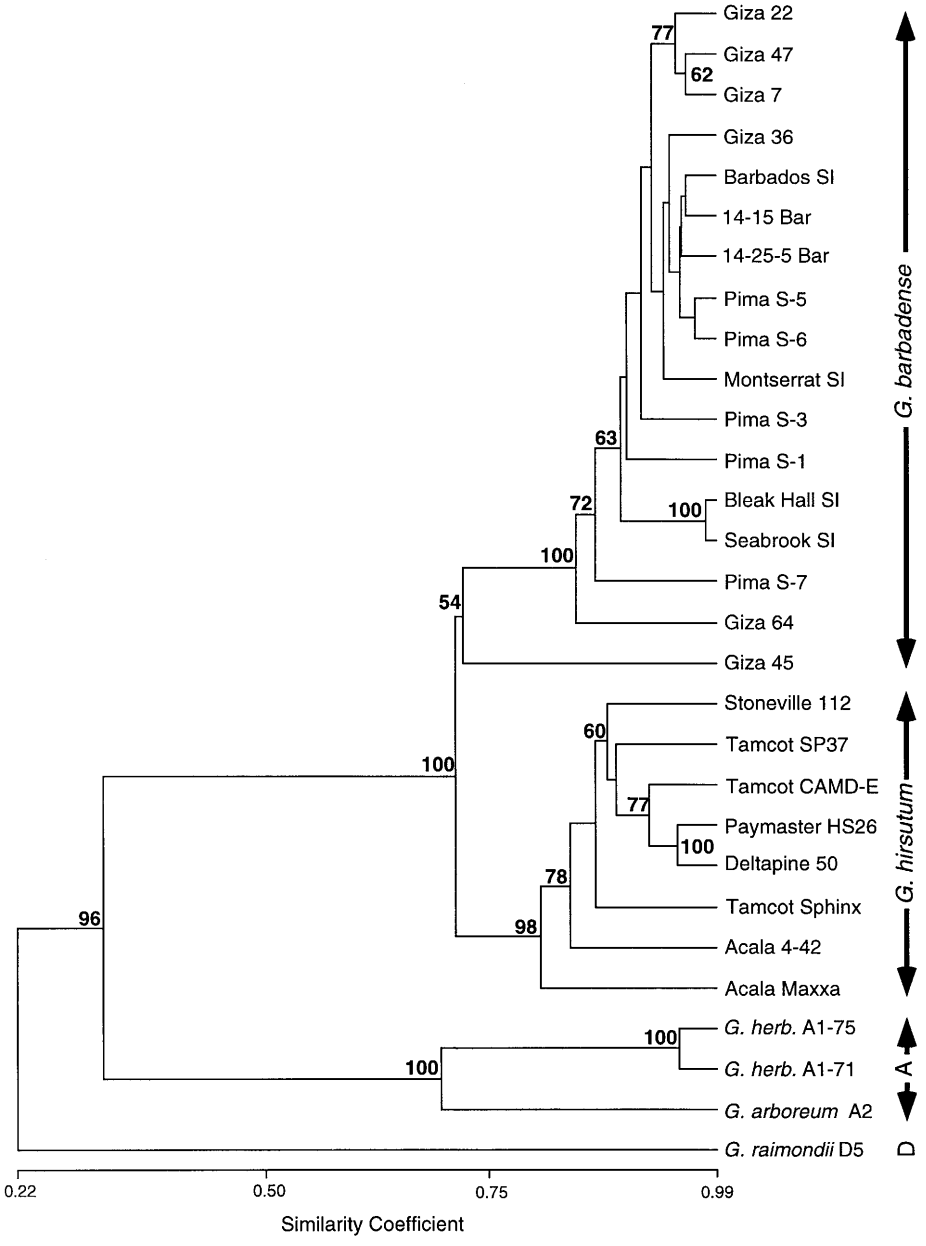

Fig. 1 An UPGMA phenogram based on AFLP data from 16 primer combinations. The numbers adjacent to some nodes indicate bootstrap confidence values (1000 bootstrap replicates). Nodes without numbers had bootstrap values of less than 50. Cytogenetic groups and taxonomically defined species are indicated at the right

hirsutum. The levels of similarity between the A and D diploid species were extremely low (0.21 to 0.22) indicating that AFLP is a reliable source of informative genome-wide markers for phylogenetic studies of cultivated and wild diploid *Gossypium* species.

A data set consisting of all 368 polymorphic AFLP loci was analyzed by UPGMA clustering (Fig. 1). The taxa fell into four well-supported clusters that are consistent with the cytogenetic genome groups and traditional taxonomic species: 17 cultivars fell into a *G. barbadense* cluster, eight cultivars into a *G. hirsutum* cluster, three taxa into an 'A-genome' cluster, and one taxon, *G. raimondii*, comprised the 'D-genome' cluster. The Egyptian cultivar Giza 45 occupied a basal position within the

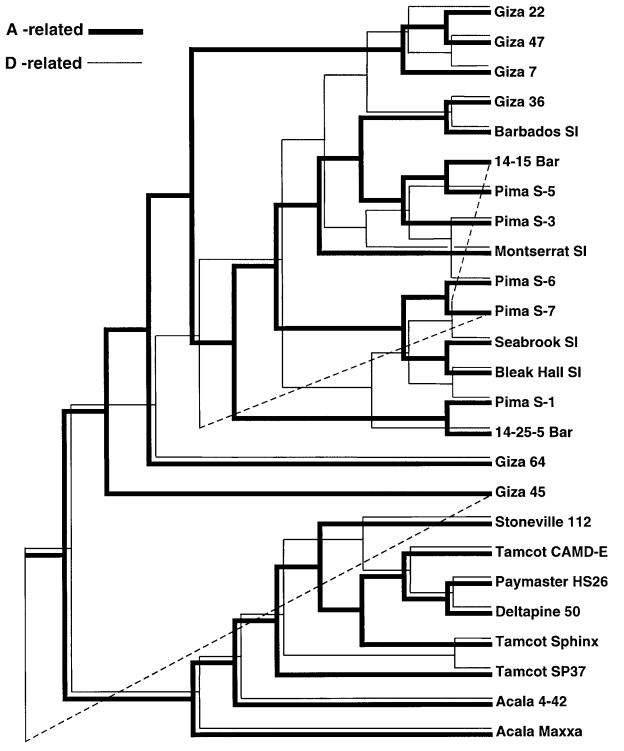

Fig. 2 A comparison of dendrograms based on UPGMA analysis of 143 A-related AFLP loci (*heavy lines*) versus 84 D-related AFLP loci (*light lines*). *Dashed lines* indicate major incongruities in the topologies generated by the two data sets

G. barbadense cluster. Likewise, the Upland cultivar Acala Maxxa occupied a basal position within the *G. hirsutum* cluster. Further, the *G. hirsutum* and *G. barbadense* groups clustered together to form a large, well-supported clade representing the combined descendants of the original AD allotetraploid.

The genomic complement of modern *G. hirsutum* and *G. barbadense* is composed of two distinct sub-genomes, one derived from the A-genome ancestor (chromosomes 1–13) and another derived from the D-genome ancestor (chromosomes 14–26) (Beasley 1940, 1942; Endrizzi et al. 1985). To the extent that A-related and D-related AFLP products may be derived from the A- and D-genome progentors, respectively (see discussion), the two groups of markers can be used as independent data sets. The data sets for 143 A-related markers and 84 D-related markers were analyzed independently by UPGMA, then compared (Fig. 2) to evaluate the robustness of inferred phylogenetic topologies, and to detect evidence of possible reticulation. Based on the total data set, Giza 22, Giza 47 and Giza 7 formed a distinct cluster that was well-supported by bootstrap analysis (Fig. 1) and UPGMA trees from A-related and D-related marker data sets (Fig. 2). Where topologies were well-supported, and there was no evidence for reticulation (from molecular markers or from traditional pedigrees), phylogenetic relationships could be inferred. Thus, Giza 22, Giza 47 and Giza 7 formed a distinct monophyletic clade. Similarly, the *G. hirsutum* cultivars Deltapine 50, Paymaster HS26 and Tamcot CAMD-E also formed a distinct clade. In contrast, Giza 45 had a well-supported position at the base of the *G. barbadense* cluster when only A-related markers were considered (bootstrap value = 100), but was basal to all tetraploid cottons when D-related markers were analyzed (bootstrap value = 100). This inconsistency suggests that the history of Giza 45 included the introduction of D-related germplasm from outside the combined tetraploid cluster as defined by this study.

The overall robustness of phenetic and putative phylogenetic topologies was evaluated by: (1) bootstrap analysis of UPGMA trees generated using the total data set, (2) comparison of UPGMA trees with N-J trees generated using the total data set (data not shown), and (3) comparison of trees generated using A-related and D-related markers. By all three criteria, the topologies of the *G. hirsutum* group were, in general, more robust than those of the *G. barbadense* group.

Principal coordinate analysis (PCA) based on genetic similarity matrices was used to visualize the genetic relationships within the tetraploid taxa. The first three eigenvectors accounted for 61% of the variation observed. PCA (Fig. 3) placed most of the cultivars into two clusters: a tight *G. hirsutum* cluster, and a moderately loose *G. barbadense* cluster containing a dense sub-cluster of historically and geographically diverse cultivars including traditional Sea Island cottons (Montserrat, Barba-

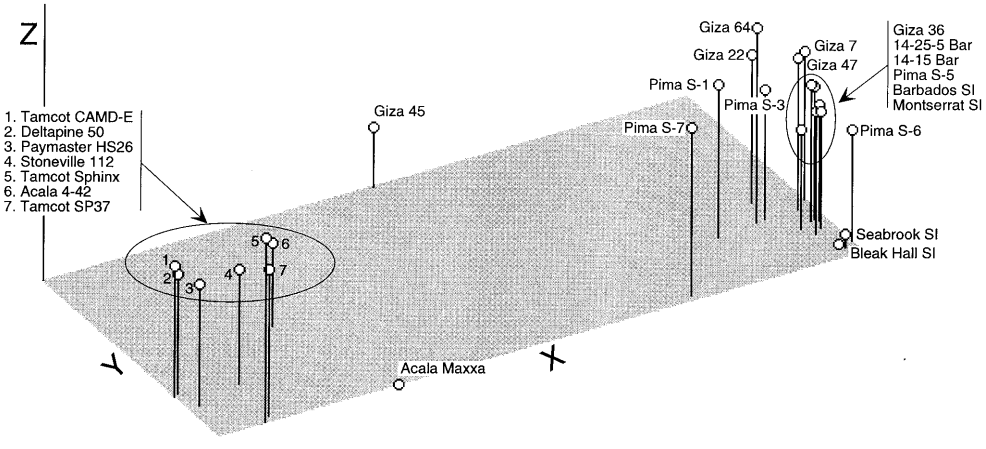

Fig. 3 Relationships of 25 tetraploid cotton taxa visualized by principal coordinate analysis (PCA) of AFLP-based genetic similarities. X, Y and Z axes are indicated

dos), Giza 36, and Pima S-5. The Sea Island cultivars Seabrook and Bleak Hall also formed a distinct subcluster within the greater *G. barbadense* group. Remarkably, Giza 45 and Acala Maxxa were found to be genetically quite distant from either the *G. hirsutum* or *G. barbadense* clusters.

Discussion

Despite being costly and technically cumbersome (discussed by Schondelmaier et al. 1996), we found that AFLP is an effective and appropriate tool for detecting DNA-level polymorphism in cultivated tetraploid cottons and related diploid species. Significant and informative polymorphism was detected across a wide range of taxonomic levels, from between distantly related diploid *Gossypium* species to within a group of modern Upland cultivars. For example, within the tetraploid taxa, each primer pair detected an average of approximately 12 polymorphic loci (Table 1). UPGMA and N-J analysis of genetic similarities based on 368 polymorphic markers places 25 AD tetraploid cotton cultivars into two distinct monophyletic clades that are in complete agreement with the traditional taxonomic arrangement of these cultivars into *G. barbadense* and *G. hirsutum.* Among the limited taxa included in our study, *G. barbadense* had a greater diversity, or broader 'genetic base' than *G. hirsutum*. However, much of this diversity may, in fact, be derived from *G. hirsutum* and other *Gossypium* species (discussed below).

Our AFLP data-set records the phylogenetic histories of a large number of individual loci through the evolution and development of cottons, providing useful information for discovering patterns of genome evolution in cultivated cotton. Out of the 368 bands that were polymorphic in our collection, 143 were shared between a subset of tetraploid cotton cultivars and a subset of A-genome accessions (*G. herbaceum* or *G. arboreum*) and 84 markers were shared between a subset of the tetraploids and *G. raimondii* (D-genome). The loss of an AFLP product by mutation should be more probable than the gain of a product, as is the case for restriction endonuclease fragments (Nei and Li 1987). Shared AFLP products are therefore more likely to be sympleisiomorphic (ancestral) alleles than homoplastic alleles arising by molecular convergence. Furthermore, the level of AFLP similarity within the A-genome taxonomic group (0.800) is much greater than similarities between the A- and D-genome taxa (0.215). Logic dictates that alleles exclusively shared between the A-genome and the tetraploids are more likely to have been contributed unilaterally by the A-genome progenitor than present in both A and D progenitors and lost in the extant D-genome representative analyzed in this study. Although we can not make definitive assignments of the A-related alleles to the A sub-genome, the A-related markers have a greater probability of being derived from the A-genome progenitor than markers obtained at random from the total data set. Similar arguments can be made for the probability of origin of the D-related markers in the D-genome progenitor.

The finding that there were fewer D-related markers than A-related markers could reflect significant divergence of modern *G. raimondii* from the ancestral D-genome parent (as has been speculated by Jiang et al. 1998), or could reflect the estimated 50–60% larger size of the ancestral A genome relative to that of the D genome (Edwards et al. 1974; Geever et al. 1989; Reinisch et al 1994). Based on the probable linkage association of markers within the A-related and D-related marker sets, respectively, we partitioned our data for UPGMA cluster analysis (Fig. 2). One striking result of this partitioned analysis is that most taxa show overall topological congruence of the two data sets while a small minority of taxa (14–15 Bar, Pima S-7, Giza 45) show major incongruities, indicating a vastly different pattern of genomic evolution. One explanation for the origin of this incongruity is wide intra- or inter-specific hybridization.

The patterns of AFLP diversity were useful for assessing the genetic contributions of various lineages to modern cultivars. The progenitor Egyptian cotton cultivar, Ashmouni, was derived from an intraspecific cross in the early 1900 s between a ''Sea Island'' cotton and an obscure *G. barbadense* tree cotton growing as a garden ornamental. Traditional pedigrees (Anonymous 1992) record that Ashmouni was utilized in the subsequent development of a number of Egyptian cultivars. Further, the Pima cultivars also trace the major part of their genetic ancestry to Ashmouni derivatives. These records are supported at the molecular level by the clustering of several Pima cultivars with Giza and Sea Island lines in UPGMA analysis (Fig. 2) and PCA (Fig. 3). In addition, the effect of the introduction of alleles from *G. hirsutum* during the development of the Pima cottons (Calhoun et al. 1994) is possibly detected by PCA as a minor shift in the position of Pima cultivars S-1 and S-7 towards the G. *hirsutum* group. Similarly, Tamcot Sphinx, Tamcot SP37 and Acala 4–42 are shifted slightly toward the *G. barbadense* cluster, perhaps reflecting the introduction of *G. barbadense* germplasm into these lineages (El-Zik and Thaxton 1989; Calhoun et al. 1994).

We found high levels of similarity within sub-clusters of the *G. barbadense* group (Figs. 1, 3). Several of the topological arrangements within the *G. barbadense* clade were not well-supported by bootstrap analysis, and there was greater inconsistency between the A-related and D-related marker sets than in *G. hirsutum*. Theoretically, a hybridization event followed by selfing (causing fixation of alleles) would be predicted to decrease the number of loci that are polymorphic in subsequent generations by 50%. In addition to creating a set of closely related descendent cultivars, this event would obfuscate phylogenetic studies based on unlinked markers situated throughout the genome, as the various markers would have independently become fixed to one or the other parental allele. Thus, one explanation for both the poorly

supported topologies and high levels of similarity within *G. barbadense* sub-clusters is that these lineages have been subjected to a greater degree of inter-cultivar gene flow than *G. hirsutum* (Kellogg et al. 1996; Wendel and Doyle 1998). Poorly supported and incongruous phylogenies can also be explained by the phenomenon of 'lineage sorting' (Wendel and Doyle 1998). If there were significant heterozygosity present at the time that a new cultivar was selected (a divergence event) then this heterozygosity will be 'sorted out' into its component alleles in each of the separate descendent lineages. However, since the cultivars in question arose in different parts of the world, and at different times, this explanation is implausible.

Analysis of AFLP diversity identified genetically unique cultivars that are potentially important sources of diverse cotton germplasm. For example, PCA suggested that the distinctive germplasm of the Seabrook and Bleak Hall Sea Island cottons had not been extensively utilized in the development of modern Pima and Giza cultivars. Perhaps the most intriguing result from the present study is the apparent genetic distinctness of Giza 45 and Acala Maxxa. Grown mainly in Egypt for specialty markets, Giza 45 is renowned for having the highest fiber quality among the world's cottons [Giza 45 fiber is extra long-staple with an average length of 40 mm, fiber strength of 392 mN/tex and fineness of 3.0 micronaire units (Anonymous 1994)]. However, it exhibits below-average yields and poor adaptability to a range of growing environments. According to pedigree records, Giza 45 is a progeny from a cross between Giza 7 and Giza 28, the latter of which is itself derived from a cross between obscure Egyptian cultivars ''Shaka 3'' and ''Shaka 4'' (Anonymous 1992). The pronounced genetic differentiation of Giza 45 from Giza 7 suggests, by deduction, that Shaka 3 or Shaka 4 (or both) is the likely source of the highly dissimilar germplasm observed – particularly in the D-related markers – in Giza 45.

In a *G. hirsutum* cv Tamcot CAMD-E × *G. barbadense* cv Sea Island Seaberry interspecific cross, Jiang et al. (1998) discovered that most of the QTLs influencing fiber quality were found in the D sub-genome of cotton. Jiang et al. (1998) hypothesized that in the wild A-genome ancestor to the AD tetraploid, loci relevant to fiber development may have already contained many ''favorable'' alleles for fiber characteristics. In this scenario, modern cotton fiber quality is ultimately limited by the amount of genetic variation incorporated into the D sub-genome at the origin of the AD allotetraploid and mutations that have arisen since. Our discovery of highly divergent D-related genetic material in Giza 45 hints at the possibility of an alternate avenue for the incorporation of distinct D sub-genome variation that positively effects fiber quality.

Acala Maxxa, grown mainly in the western US, produces some of the highest quality fibers of the *G. hirsutum* group. Its pedigree is complex (Calhoun et al. 1994), and includes germplasm introductions from Pima cultivars and from the synthetic 'triple hybrid' lineage. This triple hybrid was created beginning in 1938 through crossing the diploids *G. arboreum* (A2) and *Gossypium thurberi* (D1), followed by chromosome doubling to create an (A2D1) allotetraploid (Beasely 1942), and introgression into *G. hirsutum*. The introduction of alleles from Pima lines, which are largely *G. barbadense* at the genomic level, and from *G. arboreum* (A2) and *G. thurberi* (D1), are likely sources of the genetic distinctness of Acala Maxxa. Since both Acala Maxxa and Giza 45 produce superior quality fiber and – based on our analysis – are genetically distant from typical *G. hirsutum* and *G. barbadense*, they could be important sources of allelic diversity for the improvement of fiber quality. The existence of substantial AFLP polymorphism between Giza 45 and Acala Maxxa and the more commonly utilized elite *G. barbadense* and *G. hirsutum* lineages indicates that large numbers of polymorphic markers will be available to facilitate the identification and mapping of loci that control fiber quality traits.

Since AFLP is a proven tool for genetic linkage mapping (Joerg et al. 1995; Thomas et al. 1995; Mackill et al. 1996; Schondelmaier et al. 1996) as well as for phenetic and phylogenetic studies (Heun et al. 1997), the AFLP loci we identified can readily be placed onto linkage maps derived from intraspecific (e.g. *G. hirsutum* × *G. hirsutum*) and interspecific (e.g. *G. barbadense* × *G. hirsutum*) crosses. Rouppe Van Der Voort et al. (1997) used shared AFLP markers to align genetic maps generated in different potato genotypes. Similarly, Waugh et al. (1997) discovered that AFLP markers that were polymorphic in different barley mapping populations mapped to the same genetic locus. These data indicated that the AFLP products studied in each case were ancestral or sympleisiomorphic in nature, and demonstrated the utility of AFLP for tagging a particular genomic region and identifying homoeologous regions in other taxa – at least over short evolutionary distances. An integrated approach combining linkage mapping with genetic fingerprinting of extant species will provide detailed phylogenetic histories of individual chromosomal segments throughout the development of modern cotton cultivars, and allow for the development of a comprehensive picture of the genome dynamics leading to the emergence of modern cotton cultivars. Combined with phenotypic information (concerning fiber quality and other traits) from diverse germplasm sources, this enhanced understanding of cotton genome evolution will be an important new tool for identifying the loci that control important traits and for uncovering interactions between these loci.

Acknowledgments This work was supported by USDA-NRICGP (Plant Genome) grant 97–35300–4585. Financial assistance provided by the Egyptian government to A.M. Abdalla for his Ph.D. research at Texas A & M University is gratefully acknowledged. We are also indebted to Ed Percival and Peggy Thaxton for valuable information and advice, to Thomas Brooks for assistance with data analysis, and to Ashok Singh and members of the Pepper laboratory for critical reading of the manuscript. All experiments described herein comply with the current laws of the United States of America.

References

Anonymous (1992) Egyptian cotton varieties report. Cotton Research Institute, Agriculture Research Center, Ministry of Agriculture, Giza, Egypt

Anonymous (1994) Fiber and spinning characteristics of Egyptian cotton varieties. Cotton Research Institute, Agriculture Research Center, Ministry of Agriculture, Giza, Egypt

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

Beasley JO (1942) Meiotic chromosome behavior in species, species hybrids, haploids and induced polyploids of *Gossypium*. Genetics 27: 25–54

Calhoun DS, Bowman DT, May OL (1994) Pedigrees of Upland and Pima cotton cultivars released between 1970 and 1990. Mississippi State University, Agricultural and Forestry Experiment Station, Bulletin 1017

Edwards GA, Endrizzi AJE, Stein R (1974) Genome DNA content and chromosome organization in *Gossypium*. Chromosoma 47: 309–326

El-Zik KM, Thaxton PM (1989) Genetic improvement for resistance to pests and stresses in cotton. In: Frisbee RE, El-Zik KM, Wilson LT (eds) Integrated pest managment systems and cotton production. John Wiley and Sons, New York, pp 191–224

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

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

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

Geever RF, Katterman FRH, Endrizzi JE (1989) DNA hybridization analyses of a *Gossypium* allotetraploid and two closely related diploid species. Theor Appl Genet 77:553–559

Heun M, Schafer-pregl R, Klawan D, Castagna R, Accerbi M, Borghi B, Salamini F (1997) Site of einkorn wheat domestication identified by DNA fingerprinting. Science 278: 1312–1314

Hill M, Witsenboer H, Zabeau M, Vos P, Kesseli R, Michelmore R (1995) PCR-based fingerprinting using AFLPs as a tool for studying genetic relationships in *Lactuca* spp. Theor Appl Genet 93: 1202–1210

Iqbal MJ, Aziz N, Saeed NA, Zafar Y (1997) Genetic diversity evaluation of some elite cotton varieties by RAPD analysis. Theor Appl Genet 94:139–144

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44:223–270

Jiang C-X, 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

Joerg B, Vos P, Martin K, Salamini F, Heun M (1995) Combined mapping of AFLP and RFLP markers in barley. Mol Gen Genet 249:65–73

Kellogg EA, Appels R, Mason-Gamer RJ (1996) When genes tell different stories: the diploid genera of Titicaceae (Gramineae). Systematic Bot 21:321–347

Lin J-J, Kuo J (1995) AFLP: a novel PCR-based assay for plant and bacterial DNA fingerprinting. Focus 17: 66–70

Mackill DJ, Zhang Z, Redona ED, Colowit PM (1996) Level of polymorphism and genetic mapping of AFLP markers in rice. Genome 39:969–977

Maughan PJ, Saghai Maroof MA, Buss GR, Huestis GM (1996) Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near-isogenic line analysis. Theor Appl Genet 93: 392–401

Multani DS, Lyon BR (1995) Genetic fingerprinting of Australian cotton cultivars with RAPD markers. Genome 38:1005–1008

Nei N, Li W (1987) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76: 5269–5273

Percival AE (1987) The national collection of *Gossypium* germplasm. South Coop Series Bull 321

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238

Reinish AJ, Dong J-M, Brubaker CL, Stelly DM, Wendel JF, Paterson AH (1994) A detailed map of *Gossypium hirsutum* × *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829–847

Rieseberg LH, Noyes RD (1998) Genetic map-based studies of reticulate evolution in plants. Trends Plant Sci 7:254–259

Rohlf FJ (1993) NTSYS-pc. Numerical taxonomy and multivariate analysis system. Exeter Software, Setauket, New York, USA

Rouppe Van Der Voort JNAM, Van Zandvoort P, Van Eck HJ, Folkertsma RT, Hutten RCB, Draaistra J, Gommers FJ Jacobsen E, Helder J, Bakker J (1997) Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. Mol Gen Genet 255:438–447

Saitou M, Nei N (1987) The neighbor joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Schondelmaier J, Steinruecken G, Jung C (1996) Integration of AFLP markers into a linkage map of sugar beet (*Beta vulgaris* L.). Plant Breed 115:231–237

Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. Univ Kansas Sci Bull 28:1409–1438

Swofford, D. L. (1998) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts

Tatineni V, Cantrell RG, Davis DD (1996) Genetic diversity in elite cotton germplasm determined by morphological characteristics and RAPD. Crop Sci 36: 186–192

Thomas CM, Vos P, Zabeau M, Jones DA Norcott KA, Chadwick BP, Jones JDG (1995) Identification of amplified restriction fragment polymorphism (AFLP) markers tightly linked to the tomato Cf-9 gene for resistance to *Cladosporium fulvum*. Plant J 8:785–794

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

Vroh Bi I, Maqet A, Baudoin J-P, du Jardin P, Jacuemin JM, Mergeai G (1999) Breeding for ''low gossypol seed and high gossypol plants'' in upland cotton. Analysis of tri-species hybrids and backcross progenies using AFLPs and RFLPs. Theor Appl Genet 99:1233–1244

Waugh R, Bonar N, Baird E, Thomas B, Graner A, Hayes P, Powell W (1997) Homology of AFLP products in three mapping populations of barley. Mol Gen Genet 255:311–321

Wendel JF, Brubaker CL (1993) RFLP diversity in *Gossypium hirsutum* L. and new insights into the domestication of cotton. Am J Bot 80(SUPPL 6):71

Wendel JF, Doyle JJ (1998) Phylogenetic incongruence: window into genome history and molecular evolution. In: Soltis DE, Soltis PS, Doyle JJ (eds) Molecular systematics of plants. II. DNA Sequencing. Kluwer Academic, Boston Dordrecht London, pp 265–296

Wendel JF, Brubaker CL, Percival AE (1992) Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. Am J Bot 79: 1291–1310

Zabeau M, Vos P (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Application number: 92402629.7, Publication Number EP 0534858