Theor Appl Genet (2001) 103:547–554

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

A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting

Received: 23 May 2000 / Accepted: 23 January 2001

Abstract Reliable information about the evolutionary and genetic relationships of various germplasm resources is essential to the establishment of rational strategies for crop improvement. We used AFLPs to study the genetic relationships of 43 cultivars of *Gossypium hirsutum* representative of the genomic composition of modern 'Upland' cotton. The study also included representatives of the related tetraploid species *Gossypium barbadense*, as well as the diploid species *Gossypium raimondii*, *Gossypium incanum*, *Gossypium herbaceum* and *Gossypium arboreum*. We tested 20 primer combinations that resulted in a total of 3,178 fragments. At the species level, and above, genetic similarities based on AFLPs were in agreement with the known taxonomic relationships. Similarity indices ranged from 0.25 to 0.99. Representatives of the *G. hirsutum* germplasm resources utilized in North America, including secondary accessions collected by breeders in Central America ('Acala', 'Tuxtla', 'Kekchi') and the southwestern US ('Hopi Moencopi'), formed a single cluster with exceedingly limited genetic diversity (with many pairwise similarity indices >0.96) We concluded that these accessions were derived from the same genetic pool. The early maturing or 'latifolium' or 'Mexican Highlands' cultigens from which these cultivars were derived appear to have had extremely limited genetic diversity, perhaps as a result of a severe genetic bottleneck resulting from the selection pressures of domestication. Outside of the major *G. hirsutum* cluster, well-supported phylogenies were inferred. Inside this cluster, phylogenies were obscured by limited diversity, reticulation and lineage sorting. The implications of these findings for cotton improvement are discussed.

Keywords AFLP · Genetic diversity · Germplasm · Photoperiodism · Selection

Communicated by R. Hagemann

M.J. Iqbal and O.U.K. Reddy contributed equally to this paper

MJ Iqbal · KM El-Zik
Department of Soil and Crop Sciences, Texas A & M University, College Station, TX 77843–2474, USA

OUK Reddy · AE Pepper (✉)
Department of Biology, Texas A & M University, College Station, TX 77843–3258, USA
e-mail: apepper@bio.tamu.edu
Fax: 409–845–2518

Present address:
M.J. Iqbal, Department of Plant, Soil and General Agriculture, Southern Illinois University, Carbondale, IL 62901, USA

Introduction

The 'evolution under domestication' (Stephens 1967) of cultivated cotton (*Gossypium* spp.) presents a fascinating story with elements of biogeography, developmental biology, anthropology, and both natural and anthropogenic selection pressures. Only four species of *Gossypium* are presently cultivated, two New World tetraploid species *Gossypium hirsutum* L. and *Gossypium barbadense* L., and two Old World diploid species *Gossypium arboreum* L. and *Gossypium herbacium* L. The New World tetraploid species arose some 1–2 million years ago through the hybridization of an Old World taxon of the 'A genome' cytogenetic group, related to the present-day species *G. herbaceum* and *G. arboreum* L. (2n = 2*x* = 26), with a taxon of the 'D-genome' group, related to the new world species *Gossypium raimondii* Ulbrich and *Gossypium gossipioides* L. (2n = 2*x* = 26) (Beasley 1940, 1942; Wendel et al. 1992). After polyploidization, the nascent 'AD' disomic tetraploid (2n = 2*x* = 52) gave rise to five extant tetraploid species, including *G. barbadense* and *G. hirsutum* which dominate world cotton production for fiber and seed products.

G. hirsutum was probably first domesticated by pre-Columbian peoples of the Yucatan peninsula (Brubakar and Wendel 1994). The only truly wild *G. hirsutum* variety is 'yucatanense', a sprawling perennial shrub of littoral plant communities. Reproductive development of this species is under strict photoperiodic control, with flowering under short days. The primitive domesticated

variety ‘punctatum’ probably arose from within yucatanense. Though still a perennial shrub, many of the early punctatum cultivars had tufted ‘green’ seeds that were superior for primitive hand-ginning practices. These early domesticated forms of *G. hirustum* dispersed out of the Yucatan to the rest of Mesoamerica, as well as to northern South America and throughout the Caribbean basin. Ethnobotanical evidence suggests that the important landrace ‘latifolium’ arose from this germplasm as it was further domesticated in present day southern Mexico and Guatemala (Brubakar et al. 1999). Some accessions classified as latifoliums show photoperiodic flowering, while others are photoperiod independent (J. McCarty, personal communication). According to Cook (1905), the Kekchi indians of Guatemala traditionally intercropped their cotton with peppers (*Capsicum* spp.), and would remove cotton plants as soon as first bolls began to open in order to prevent competition with the developing peppers. This practice would have eliminated late-maturing genotypes. Such a rigorous selection regime for early maturity would have selected for reduced seed dormancy, compact ‘annualized’ growth habit, and possibly photoperiod-indendent flowering, thus creating genotypes suitable for cultivation at higher latitudes with longer summer daylengths. The early maturing latifolium genotypes diffused into the highlands of present-day southern central Mexico prior to the arrival of Europeans, and the resulting cultivars became collectively known as ‘Mexican Highlands’ (or ‘Mexican’) varieties.

In colonial North America, cotton production initially consisted of a complex mixture of taxa, including *G. herbaceum* imported to the New World by the British, *G. barbadense* (‘Sea Island Cotton’) from the Caribbean, and several varieties of *G. hirsutum*, including tufted ‘green-seed’ types from Mexico. The invention of the cotton gin in 1793 increased the demand for high yield cultivars, eventually leading to the introduction of several non-tufted Mexican Highlands varieties including Burling’s Cotton in 1806 and Wyche cotton in 1857 (Smith et al. 1999). *G. hirsutum* of the ‘Mexican type’ may have been grown in the Stephen F. Austin colony in Texas as early as 1821, and numerous introductions were probably made by soldiers returning from the Mexican-American War (1846–1848). These cultivars were subjected to strong selection to create varieties adapted to local conditions in various cotton-growing regions of North America. Throughout this period, both intentional and unintentional outcrosses occurred between cultivars (Endrizzi et al. 1985). Collectively known as ‘American Upland Cotton’ the resulting high yielding and adaptable cultivars were dispersed to Europe, Africa and Asia.

In response to the dire boll weevil (*Ahthonomus grandis* Boh.) infestation that began at the end of the 19th century, breeders collected new germplasm from the southern Mexican state of Chiapas (‘Acala’ and ‘Tuxtla’, 1906), Guatamala (‘Kekchi’, 1904), and from the Hopi peoples of Arizona (‘Hopi Moencopi’). Although not resistant to the weevil, these cultigens were incorporated with germplasm from earlier Mexican Highlands introductions to provide the genetic foundation for modern Upland *G. hirsutum* cultivation worldwide.

An understanding of the evolutionary and genomic relationships of cotton species and cultivars is critical for further utilization of extant genetic diversity in the development of superior cultivars that will combine the favorable qualities (insect and pathogen resistance, fiber quality and yield traits, tolerance to environmental stresses) conditioned by diverse *G. hirsutum* germplasm resources (El-Zik and Thaxton 1989). Given the status of *G. hirsutum* as the world’s most important fiber crop, the genetic composition of various cultivars has not been adequately investigated. Studies using allozymes (Wendel et al. 1992) and RFLPs (Wendel and Brubaker 1993) have been limited by low levels of polymorphism at the intraspecific and interspecific levels. Recently, there have been several reports describing the genetic relationships of several cotton varieties based on PCR-based molecular techniques such as RAPDs (Multani and Lyon 1995; Tatineni et al. 1996; Iqbal et al. 1997) and AFLP (Abdallah et al. 2001). The amplified fragment length polymorphism (AFLP) technique (Zabeau and Vos 1993; Vos et al. 1995) generates a relatively large number of polymorphic amplified fragments compared to the RAPD method. Since each AFLP fragment is considered to be a genetically distinct locus, a large number of loci can be compared between different genotypes in relatively few experiments. We have examined 43 genotypes that are representative of most of the Upland *G. hirsutum* germplasm resources utilized in North America and Asia. Our set of taxa included the Acala, Tuxtla, Kekchi and Hopi Moencopi collections as well as several obsolete cultivars, such as ‘Western Stormproof’ and ‘Wannamaker’s Cleveland’ which are purported to be direct selections from 19th century germplasm introductions from Mexico (Smith et al. 1999).

Materials and methods

Plant materials and DNA isolation

Three diploid species of *Gossypium*, three *G. barbadense* cultivars, and 43 *G. hirsutum* accessions were used for the study (Table 1). Genomic DNAs were isolated from leaf tissue by the method of Iqbal et al. (1997). DNAs of cultivars Allepo, Nazelli, Cedix, LRA5166, CP 15/2, S-12, CIM-435, CIM-443 and CIM-1100 were provided by Dr. Yusuf Zafar, NIBGE, Faisalabad, Pakistan. After RNase treatment, the concentration of each sample was measured by a fluorimetric assay using Hoechst dye.

AFLP fingerprinting

AFLP fingerprinting (Zabeau and Vos 1993; Vos et al. 1995) was performed using Analysis System I (GIBCO-BRL Life Technologies) according to the manufacturer-supplied protocol with some modifications. Genomic DNA (225 ng) was digested with 0.625 units of *Eco*RI and *Mse*I for 2 h at 37°C. *Eco*RI/*Mse*I adapters were ligated to genomic DNA fragments using 0.25 Weiss units of T4 DNA ligase for 2 h at room temperature. Pre-amplification of adaptor-ligated DNA was carried out in a reaction containing 10 µl of pre-amplification primer mix, 1.25 µl of 10 × PCR buffer for

Table 1 List of cotton taxa examined in this study

Sr. No	Accession No	Taxa	Origin
1	SA 58	*G. hirsutum* L. var. Tuxtula	Mexico
2	SA 105	*G. hirsutum* L. var. Half & Half	USA
3	SA 219	*G. barbadense* L. var. Yuma	USA
4	SA 223	*G. barbadense* L. var. Pima 32	USA
5	SA 252	*G. hirsutum* L. var. Coker 100	USA
6	SA 259	*G. hirsutum* L. var. Coker's Foater 300	USA
7	SA 280	*G. hirsutum* L. var. Kekchi	Guatemala
8	SA 281	*G. hirsutum* L. var. Hopi Moencopi	USA
9	SA 296	*G. hirsutum* L. var. Wannamaker's Cleveland W.R.	USA
10	SA 311	*G. hirsutum* L. var. Ambassador (4B6)	USA
11	SA 395	*G. hirsutum* L. var. Deltatype Webber	USA
12	SA 464	*G. hirsutum* L. var. AHA 6–1-4	USA
13	SA 517	*G. hirsutum* L. var. Acala 111 Rogers	USA
14	SA 524	*G. hirsutum* L. var. Lankart 57	USA
15	SA 852	*G. hirsutum* L. var. Dixie Triumph	USA
16	SA 857	*G. hirsutum* L. var. Acala Original	Mexico
17	SA 874	*G. hirsutum* L. var. Deltapine 14	USA
18	SA 879	*G. hirsutum* L. var. Delfos Washington	USA
19	SA 883	*G. hirsutum* L. var. Lightening Express	USA
20	SA 892	*G. hirsutum* L. var. Mebane	USA
21	SA 966	*G. hirsutum* L. var. Paymaster 54B	USA
22	SA 970	*G. hirsutum* L. var. Stardel	USA
23	SA 1010	*G. hirsutum* L. var. Austin	USA
24	SA 1019	*G. hirsutum* L. var. Western Stormproof	USA
25	SA 1047	*G. hirsutum* L. var. Acala 4–42	USA
26	SA 1071	*G. hirsutum* L. var. Rex	USA
27	SA 1105	*G. hirsutum* L. var. Santan Acala 25	USA
28	SA 1113	*G. hirsutum* L. var. Empire	USA
29	SA 1122	*G. hirsutum* L. var. Lockett 4789	USA
30	SA 1157	*G. hirsutum* L. var. Dixie King	USA
31	SA 1159	*G. hirsutum* L. var. Fox 4	USA
32	SA 1163	*G. hirsutum* L. var. Stoneville 213	USA
33	SA 1165	*G. hirsutum* L. var. Carolina Queen	USA
34	SA 1185	*G. hirsutum* L. var. Coker 413–68	USA
35	SA 1549	*G. hirsutum* L. var. Paymaster 202	USA
36	GB-1023	*G. barbadense* L. var. Pima S-7	USA
37	N/A[a]	*G. hirsutum* L. var. Stoneville 887	USA
38	N/A	*G. hirsutum* L. var. Allepo	Syria
39	N/A	*G. hirsutum* L. var. Nazelli	Turkey
40	N/A	*G. hirsutum* L. var. Cedix	El Salvador
41	N/A	*G. hirsutum* L. var. LRA 5166	India
42	N/A	*G. hirsutum* L. var. CP 15/2	Pakistan
43	N/A	*G. hirsutum* L. var. S-12	Pakistan
44	N/A	*G. hirsutum* L. var. CIM 435	Pakistan
45	N/A	*G. hirsutum* L. var. CIM 443	Pakistan
46	N/A	*G. hirsutum* L. var. CIM 1100	Pakistan
47	D5-4	*G. raimondii* Ulbrich	N/A
48	E4-4	*G. incanum* Schwartz	N/A
49	A1-71	*G. herbaceum* L.	N/A
50	E3-1	*G. areysianum* Deflers	N/A
51	A2-78	*G. arboreum* L.	N/A

[a] N/A = not available

AFLP [200 mM Tris-HCl (pH 8.4), 15 mM $MgCl_2$, 500 mM KCl], 1.25 unit of *Taq* DNA polymerase (GIBCO BRL), and 1.25 µl of diluted, restricted and ligated template DNA.

Amplification was carried out for 20 cycles of 94°C for 30 s, 56°C for 1 min, 72°C for 1 min in Perkin-Elmer 9600 thermocycler. For selective amplification, 2.5 ng of *Eco*RI primer was labeled with ^{33}P γ-ATP (0.1 µl of 3,000 Ci/mol) using 0.1 units of T4 polynucleotide kinase for 1 h at 37°C. Selective amplification was carried out in a 10-µl volume containing labeled *Eco*RI primer, 2.25 µl of *Mse*I primer mix (6.7 ng/µl) with dNTPs, 3.95 µl of dH_2O, 1 µl of 10 × PCR buffer for AFLP, 0.25 units of *Taq* DNA polymerase (GIBCO BRL) and 2.5 µl of a 1:10 dilution of pre-amplified template DNA. Both *Eco*RI and *Mse*I primers used in selective amplification had three extra nucleotides at the 3′ end in order to reduce the number of amplified fragments. The primer combinations used were: (1) E-AAC + M-CAC, (2) E-AAC + M-CAG, (3) E-AAC + M-CTC, (4) E-AAC + M-CTG, (5) E-AAG + M-CTG, (6) E-AAG + M-CTT, (7) E-AAG + M-CAC, (8) E-ACA + M-CTA, (9) E-ACA + M-CAC, (10) E-ACC + M-CAC, (11) E-ACC + M-CTA, (12) E-ACG+ M-CTA, (13) E-ACT + M-CTC, (14) E-AGC + M-CAG, (15) E-AGC + M-CAA, (16) E-AGC + M-CTA, (17) E-AGG + M-CAC, (18) E-AGG + M-CAG, (19) E-AGG + M-CAT, (20) E-AGG + M-CTG. Amplification was carried out in a Perkin-Elmer 9600 thermocycler with a first cycle of 94°C for 30 s, 65°C for 30 s and 72°C for 1 min, followed by 12 cycles with a 0.7°C decrease in annealing temperature with each cycle. Further amplification was performed by an additional 23 cycles of PCR using 56°C as the annealing temperature.

After the addition of 10 µl of manual sequencing dye (98% formamide, 10 mM EDTA, 0.025% xylene cyanol, 0.025% bromophenol blue) samples were heated at 94°C for 3 min then

chilled on ice, and 2 µl of each was electrophoresed in a 5% denaturing acrylamide gel using 0.5 × TBE buffer (45 mM Tris-Borate, 1 mM EDTA, pH 8) at 60 W for 1.5 h. The gel was then transferred to 3MM blotting paper, dried in a vacuum dryer and exposed to X-ray film.

Data scoring and analysis

Data were scored on the basis of the presence or absence of amplified fragments using the criteria described by Abdalla et al. (2001). Each AFLP fragment was considered a different locus as it corresponds to a unique position in the genome (Vos et al. 1995). Genetic similarities based on Nei and Li's (1979) coefficient were calculated using the Numerical Taxonomy Multivariate Analysis System (NTSYS-pc) Version 2.0 software package (Rohlf 1993). Cluster analysis was performed using the unweighted pair group method of arithmetic means (UPGMA; Sokal and Michener 1958). Statistical support for various topological arrangements was determined by the bootstrapping method of Felsenstein
(1985) using 1,000 replicates. 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). Parsimony analyses were performed using the PAUP*4.0b2a (Swofford 1999) software package, and MacClade 3.05 (Maddison and Maddison 1993). For parsimony tests, heuristic searches were performed with a simple stepwise addition sequence, and tree-bisection-reconnection (TBR) branch swapping. Three *G. barbadense* cultivars were used as a monphyletic outgroup.

Results

AFLP analysis of 50 taxa identified a total of 3,178 polymorphic loci, presumably distributed throughout each of the *Gossypium* genomes. Genetic similarities among all taxa ranged from 0.25 (between *G. herbaceum* and *G. raimondii*) to 0.99 (see Table 1). Among tetraploid species, similarities ranged from 0.67 to 0.99. Similarities ranged from 0.83 to 0.99 among various *G. hirsutum* cultivars. Similarity matrices generated by our AFLP data set were analyzed by UPGMA clustering (Fig. 1). The taxa studied fell into well-supported clusters that were generally consistent with the established cytogenetic genome groups and taxonomic species. The D-genome diploid species *G. arboreum* and *G. herbacium* formed a well-supported cluster (bootststrap value = 100). Similarly, the AD tetraploids also formed a well-supported cluster (BV = 100), with distinct sub-clusters comprised of *G. barbadense* and *G. hirsutum* (BV = 100 for each). Interestingly, the poorly studied African-Arabian diploid species *G. incanum* (E4 genome group), which was originally included as an out-group for the analysis, formed a reasonably well-supported cluster with the tetraploid species (BV = 82) in UPGMA, as well as in parsimony analyses (data not shown).

Several cultivars from the Indian and Pakistani breeding programs, such as CIM443 and LRA5166, were found to be clearly within the *G. hirsutum* cluster (BV = 100), yet genetically diverse compared to the bulk of the *G. hirsutum* accessions. The American cultivar 'Deltapine 14' and the El Salvadoran accession 'Cedix' also fell into this category. The remaining 36 *G. hirsutum* ac-

Table 2 AFLP-based similarities of selected representitive taxa

Taxon	Tuxtula	Kekchi	Hopi	AHA	Acala	West	Cedix	LRA	S-12	C4	C1	Yuma	Pima 32	Pima S-7	*G. rai.*	*G. herb.*	*G. arbo.*
Tuxtula	1.00																
Kekchi	0.96	1.00															
Hopi Moencopi	0.96	0.97	1.00														
AHA 6-1-4	0.90	0.90	0.89	1.00													
Acala Original	0.96	0.97	0.97	0.90	1.00												
Western Stromprf	0.96	0.98	0.97	0.90	0.97	1.00											
Cedix	0.91	0.92	0.91	0.87	0.91	0.91	1.00										
LRA 5166	0.92	0.92	0.91	0.88	0.91	0.92	0.92	1.00									
S-12	0.91	0.91	0.90	0.88	0.91	0.91	0.91	0.95	1.00								
CIM 443 (C4)	0.85	0.85	0.85	0.83	0.86	0.86	0.85	0.89	0.90	1.00							
CIM 1100 (C1)	0.91	0.91	0.91	0.88	0.92	0.91	0.91	0.95	0.95	0.90	1.00						
Yuma	0.74	0.75	0.75	0.71	0.75	0.75	0.73	0.72	0.72	0.68	0.73	1.00					
Pima 32	0.73	0.74	0.74	0.70	0.74	0.74	0.72	0.71	0.71	0.67	0.71	0.71	0.90				
Pima S-7	0.74	0.74	0.74	0.70	0.74	0.74	0.72	0.71	0.71	0.67	0.72	0.91	0.99	1.00			
G. raimondii	0.29	0.29	0.29	0.28	0.29	0.29	0.30	0.29	0.29	0.28	0.29	0.29	0.29	0.29	1.00		
G. herbacium	0.32	0.32	0.32	0.31	0.32	0.32	0.32	0.32	0.31	0.31	0.31	0.30	0.31	0.31	0.25	1.00	
G. arboreum	0.31	0.31	0.31	0.31	0.31	0.31	0.32	0.31	0.31	0.31	0.31	0.30	0.31	0.31	0.27	0.57	1.00

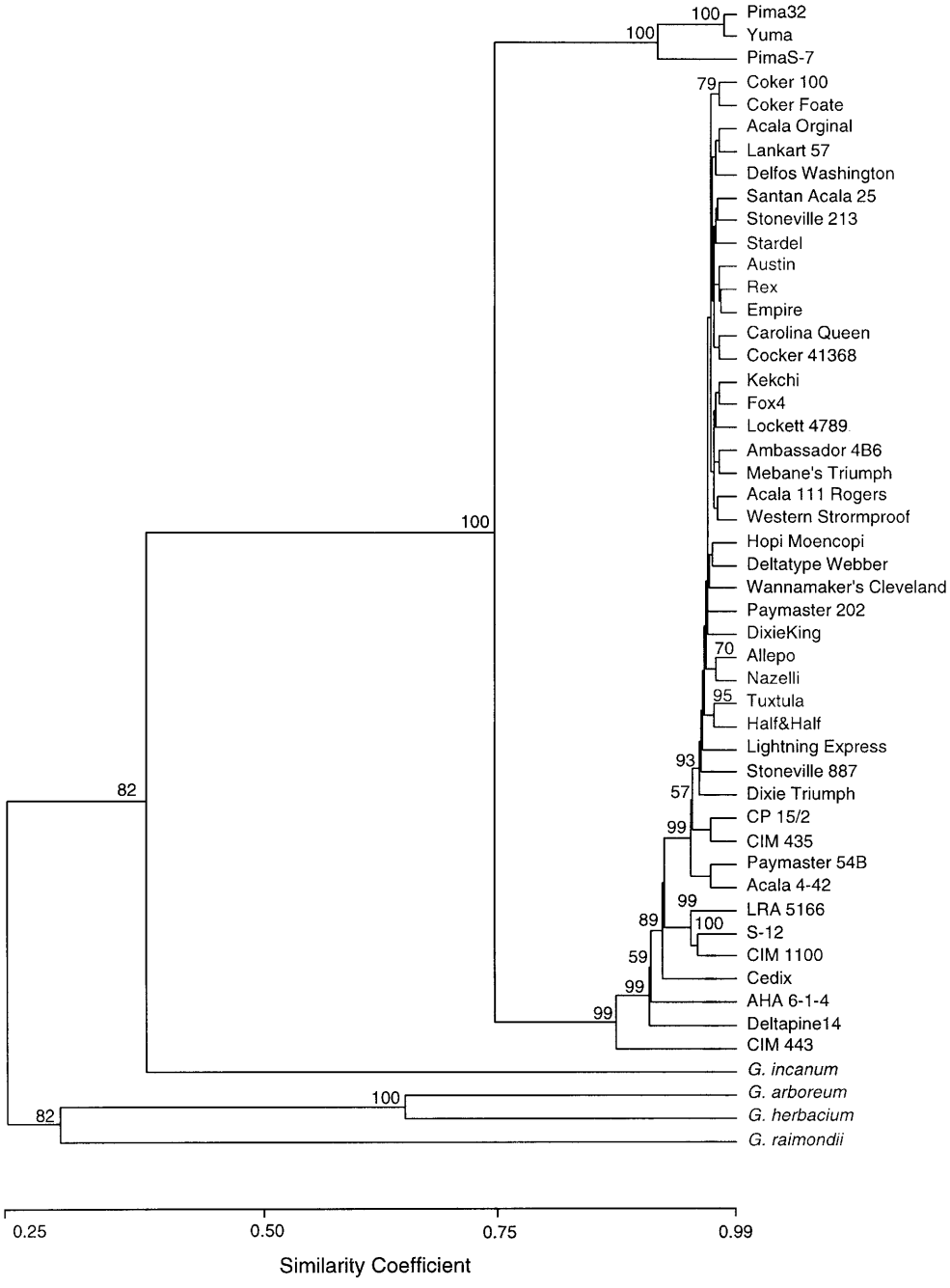

Fig. 1 An UPGMA phenogram based on AFLP data from 20 primer combinations. Bootstrap values from 1,000 replicates are indicated

cessions, including the Acala, Kekchi, and Hopi Moencopi primary germplasm introductions, and the Allepo and Nazelli cultivars from the Near East, formed a single well-supported cluster (BV = 99) characterized by exceedingly limited internal genetic differentiation. Many of the pairwise similarity indices in this cluster were in the range of 0.96 to 0.99 (Fig. 1, Table 2). Few topological arrangements within this cluster were supported by bootstrap analysis.

Principal coordinate analysis (PCA) based on genetic similarity matrices was used to visualize the genetic relationships of *G. hirsutum* types along with representatives of *G. barbadense* (Fig. 2). The first three eigen vectors used for the analysis accounted for 80% of the observed variation. Representatives of the *G. hirsutum* germplasm resources of North America, along with Allepo and Nazelli, formed a very dense cluster, indicative of very limited genetic differentiation. In contrast, most of the Indian/Pakistani cultivars formed a relatively diverse 'scatter' that included the Central American cultivar 'Cedix'. The direction of the scatter was not skewed towards *G. barbadense*, as has been observed previously in *G. hirsutum* lines known to have introgressions of *G. barbadense* germplasm (Abdalla et al. 2001).

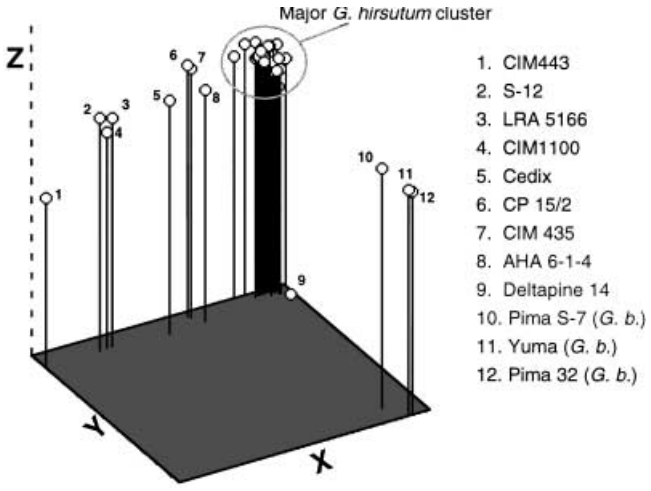

Fig. 2 Genetic relationships of 43 *G. hirsutum* cultivars and three representative *G. barbadense* (*G.b.*) visualized by principle coordinate analysis (PCA) of AFLP-based genetic similarities

Discussion

AFLP analysis was used to derive well-supported hypotheses for genetic relationships across a wide range of taxonomic levels, from among the relatively distant diploid *Gossypium* species to within the group of *G. hirsutum* cultivars. *G. hirsutum* grown around the world is usually attributed to importations of 'American Upland Cotton' in the 19th and early 20th centuries. Our results show this is indeed the case for the Near Eastern cultivars Allepo and Nazelli, and some of the Indian and Pakistani cultivars, such as CP15/2 and CIM435. However, the cotton breeding programs of India and Pakistan (which were intertwined until the partitioning of 1947) apparently utilized a much more genetically diverse pool of germplasm resources. The Indian sub-continent has a long tradition of cultivation of the diploid species *G. herbacium* and *G. arboreum*, which extends back at least 4,300 years (Gulatti and Turner 1928). This cultivation persists to this day, particularly in regions of low rainfall. Throughout much of the 19th century there were numerous importations of New World tetraploid cultivars to India, including *G. barbadense* 'Sea Island Cotton', and *G. hirsutum* from North America ('American Upland Cotton'), and South America (I.M. Katageri, personal communication). More recently, synthetic tetraploids derived from colchicine-doubled diploid species have been incorporated into Indian and Pakistani breeding programs. In our PCA (Fig. 2), we did not observe that the scattered group of Indian/Pakistani cultivars was skewed toward *G. barbadense*. Furthermore, analysis of individual loci using MacClade 3.01 failed to reveal a consistent pattern of allele sharing with *G. barbadense*, or with any of the individual diploid taxa. There was, however, a consistent pattern of allele sharing with Cedix, a *G. hirsutum* originally collected by J.R. Schwendiman in El Salvador (Wilson and Brown 1991).

Cedix was successfully used as a source of germplasm for resistance to the cotton leaf crumple geminivirus (Wilson and Brown 1991). Much like the Indian/Pakistani cultivars, Cedix showed no consistent pattern of allele sharing with *G. barbadense* or any of the individual diploid species. We hypothesize that Cedix and the Indian/Pakistani cultivars may be representative of a more diverse *G. hirsutum* germplasm pool than that utilized in the typical North American cultivars. In the Indian/Pakistani cultivars, the source of this diversity may have been importations of *G. hirsutum* from South America.

Despite the size of our AFLP data set we failed to obtain a clear and robust definition of several of the phylogenies and genetic relationships within a very large clade containing predominantly North American Upland cottons as well as early 20th century germplasm collections. This lack of phenetic and phylogenetic definition may be attributable to several causes. There was a remarkable paucity of phylogenetically informative AFLP loci within the large *G. hirsutum* clade, particularly considering the fact that 3,178 polymorphic loci were documented in the study. In addition, the poorly supported phylogenies within this clade could also be explained by the phenomenon of 'lineage sorting' (Wendel and Doyle 1998). If there was polymorphism present at a particular locus in the Mexican germplasm-source populations, then various 'samplings' of this gene pool by pre-Mendelian 'seedsmen' and by breeders might give rise to lineages with different genotypes at this locus. Any attempts to derive genetic relationships or phylogenies based on this locus would be unfounded. Finally, reticulation of Upland cotton lineages is known to have occurred through intentional inter-cultivar crosses documented in pedigree records of varying reliability (Calhoun et al. 1994; Smith et al. 1999), and through unintentional outcrossing mediated by natural pollinators. Such gene flow would have further obfuscated much of the 'molecular signature' of the various phylogenetic lineages.

To test the role of reticulation in the major *G. hirsutum* clade we employed parsimony analysis on the 36 taxa of the major *G. hirsutum* clade (utilizing the *G. barbadense* cultivars Yuma, Pima 32 and Pima S-7 as a monophyletic outgroup). Heuristic parsimony searches yielded poorly supported topologies much like those obtained from UPGMA analysis. To determine the effect of reticulation, we performed parsimony analysis on two additional data sets, each excluding ten of the 36 taxa. One data set excluded several direct germplasm collections, such as Tuxtla and Kekchi, as well as cultivars that were purportedly direct selections from 19th century germplasm introductions (Calhoun et al. 1997; Smith et al. 1999). A single most-parsimonious tree from this data set had a consistency index (CI) of 0.5770 (0.5118 excluding uninformative characters) and a homoplasy index (HI) of 0.4230 (0.4882 excluding uninformative characters). Our comparison data set excluded ten taxa known from pedigree records to be derived from wide inter-cultivar crosses,

such as Deltatype Webber, Empire and Coker 413. The CI from the latter data set was only slightly higher at 0.5947 (0.5296 excluding un-informative characters), and HI was only slightly lower at 0.4053 (0.4704 excluding un-informative characters). These results indicate that although reticulation plays a limited role in obfuscating the phylogenetic signal, other factors almost certainly exert an effect. The lack of robustness in trees within the major *G. hirsutum* clade is probably at least partially attributable to severely limited genetic differentiation and to lineage sorting or 'sampling' effects. In the near future, information on the linkage and linkage-disequilibrium relationships of the AFLP loci used in the present study will allow us to determine the phylogenetic histories of individual chromosomal segments, greatly enhancing our understanding of the relative effects of reticulation and lineage sorting (Rieseberg and Noyes 1998).

The limited genetic diversity of cultivated Upland *G. hirsutum* has been observed previously (Wendel et al. 1992; Wendal and Brubaker 1993; Tatineni et al. 1996; Iqbal et al. 1997). One of several alternative hypotheses to explain this apparent lack of diversity is that genetic bottlenecks occurred upon importation of small quantities of seed from Mexico in the 19th century introductions. For example, Burling's cotton (1806) was reportedly smuggled out of Mexico in the stuffing of dolls (Wailes 1854; Lyman 1868). It has been further hypothesized that the limited number of importations ('samplings') captured only a fraction of the genetic diversity present in Mexican cultigens. However, we found that secondary germplasm introductions from Mexico (Acala, Tuxtla) and Guatamala (Kekchi) show no greater genetic diversity than samples in the collection of mostly North American Upland cultivars. The Hopi Moencopi cultivar, traditionally cultivated by Hopi peoples of northeast Arizona, also fell within the large clade of genetically similar cultivars. Since *G. hirsutum* seeds were widely traded among the indigenous peoples of the Americas, it is not surprising that Hopi Moencopi also appears to be derived from the same pool of latifolium cultigens as Kekchi, Acala, Tuxtla and the 19th Century Mexican Highlands introductions. Indeed, the only cultivars with any appreciable diversity were Cedix and several of the Indian/Pakistani cultivars, which almost certainly represent germplasm sources from outside the 'Mexican Highlands' latifolium genetic pool.

We hypothesize that one or more genetic bottlenecks may have occurred during the later stages of the development of *G. hirsutum* latifolium, possibly as a result of rigorous selection for early maturity. This type of 'catastrophic selection' is well recognized as a potential mechanism for genetic bottlenecks (Lewis 1962). Much of the original genetic diversity of *G. hirsutum*, including valuable alleles that confer resistance to insects, pathogens and environmental adversities, would have been lost during this phase of its domestication. The overwhelming majority of Upland cotton improvement efforts have thus far utilized germplasm drawn from among the set of highly similar taxa represented by the large *G. hirsutum* clade (Figs. 1 and 2). Efforts to reduce the 'genetic vulnerability' of *G. hirsutum* must therefore focus on taxa outside of this clade, i.e., on the more diverse cultivated varieties such as Cedix and the Indian/Pakistani cultivars, as well as primitive domesticated landraces (such as 'marie-galante', 'morrilli', 'palmeri', 'punctatum' and 'richmonii') and the wild *G. hirsutum* yucatanense. Additional important germplasm resources include the other allotetraploid species *G. barbadense*, *Gossypium darwinii* Watt, *Gossypium tomentosum* Nuttall, and *Gossypium mustelinum* Miers ex Watt. However, the majority of these potential germplasm resources show photoperiodic control of flowering, a major obstacle to the efficient introgression of valuable alleles into modern cultivars (Percival et al. 1999). Therefore, programs to utilize these important germplasm sources should be preceded by vigorous efforts directed towards understanding the genetic and molecular basis of photoperiodic flowering in *Gossypium*.

Acknowledgements This work was supported by USDA-NRICGP (Plant Genome) grant 97-35300-4585. We thank Y. Zafar, NIBGE, Faisalabad, Pakistan, for providing DNA of some cotton varieties. R.J. Kohel provided plant materials and A.S. Reddy provided valuable advice. We also thank I.S. Katagiri for numerous helpful discussions. All experiments described herein comply with the current laws of the United States of America.

References

Abdalla AM, Reddy OUK, El-Zik KM, Pepper AE (2001) Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. Theor Appl Genet 102:222–229

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

Brubaker CL, Wendel JF (1994) Re-evaluating the origin of domesticated cotton (*Gossypium hirsutum*: Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). Am J Bot 81: 1309–1326

Brubaker CL, Bourland FM, Wendel JF (1999a) The origin and domestication of cotton. In: Smith WC, Cothren T (eds) Cotton: origin, history, technology and production. John Wiley and Sons, New York, pp 3–31

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

Cook OF (1905) Cotton culture in Guatemala. In: Yearbook of the United States department of Agriculture: 1904. U.S. Government Printing Office, Washington, D.C., pp 475–488

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

Gulatti AM, Turner AJ (1928) A note on the early history of cotton. Indian Central Cotton Committee, Tech Lab Bull No. 17

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

Lewis H (1962) Catastrophic selection as a factor in speciation. Evolution 16: 257–271

Lyman JB (1868) A treatise on cotton culture. Orange, Judd and Co., New York

Maddison WP, Maddison DR (1992) MacClade: analysis of phylogeny and character evolution. Sinaur Associates, Sunderland, Massachusetts

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

Nei N, Li W (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76: 5269–5273

Percival AE, Wendel JF, Stewart JM (1999) Taxonomy and germplasm resources. In: Smith WC, Cothren T (eds) Cotton: origin, history, technology and production. John Wiley and Sons, New York, pp 33–64

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

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

Smith CW, Cantrell RG, Moser HS, Oakley SR (1999) The origin and domestication of cotton. In: Smith WC, Cothren T (eds) Cotton: origin, history, technology and production. John Wiley and Sons, New York, pp 99–171

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

Stephens SG (1967) Evolution under domestication of the New World cottons (*Gossypium* spp.). Cienc Cult 19: 118–134

Swofford DL (1999) 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

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

Wailes BLC (1854) Report on the agriculture and geology of Mississippi. State of Mississippi Printing Office

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

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

Wendel JF, Doyle JJ (1998) Phylogenetic Incongruence: Window into Genome History and Molecular Evolution. In: Soltis DE, Soltis PS and Doyle JJ (eds) Molecular Systematics of Plants II, DNA Sequencing. Kluwer Academic, Boston/Dordrecht/London, pp 265–296

Wilson FD, Brown JK (1991) Inheritance of resistance to cotton leaf crumple virus in cotton. J Hered 82: 508–509

Zabeau M, Vos P (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Application number: 924026297, Publication Number EP 0534858