Theor Appl Genet (1995) 91:1153–1161

G.-L. Wang · J.-M. Dong · A. H. Paterson

The distribution of *Gossypium hirsutum* chromatin in *G. barbadense* germ plasm: molecular analysis of introgressive plant breeding

Received: 12 July 1994 / Accepted: 12 May 1995

Abstract Cotton is unusual among major crop plants in that two cross-fertile species are widely cultivated for a common economic product, fiber. Both historical evidence and classical genetic studies suggest that many improved forms of *Gossypium barbadense* ("Sea Island", "Egyptian", and "Pima" cottons) may include chromatin derived from *G. hirsutum*. Using 106 restriction fragment length polymorphism (RFLP) loci well-distributed across the cotton genome, we revealed the amount and genomic distribution of *G. hirsutum* chromatin in 54 *G. barbadense* collections from around the world. The average *G. barbadense* collection was comprised of 8.9% alleles apparently derived from *G. hirsutum*. Pima cultivars (7.3%) had fewer *G. hirsutum* alleles than Sea Island (9.0%) or Egyptian (9.6%) cultivars. *G. hirsutum* alleles were not randomly distributed, as 57.5% of the total introgression observed was accounted for by five specific chromosomal regions that span less than 10% of the genome. The average length of an introgressed chromosome segment was $\geqslant 12.9$ cM. Overlap of introgressed chromatin in different breeding programs hints that retention of these *G. hirsutum* chromosomal segments may impart a selective advantage to *G. barbadense* genotypes. Although cluster analysis generally grouped germ plasm from common classes and/or breeding programs together, no 2 genotypes were identical – thus differences in the length and repertoire of introgressed chromosome segments also permit DNA fingerprinting of *G. barbadense* cultivars.

Key words Cotton · Restriction fragment length polymorphism (RFLP) · Genetic diversity · DNA fingerprinting · Fiber quality

Communicated by A. L. Kahler

G. -L. Wang[1] · J. -M. Dong · A. H. Paterson (✉)
Department of Soil and Crop Sciences, Texas AM University, College Station, Tx 77843, USA

Present address:
[1] Department of Plant Pathology, University of California, Davis, CA 95616, USA

Introduction

Cotton (*Gossypium* spp.) is the world's leading fiber crop, and the second most valuable oilseed. While four *Gossypium* species are cultivated, *G. barbadense* cottons are particularly prized for their superior fiber quality and command a premium price that often compensates growers for the lower yields obtained relative to *G. hirsutum* cultivars (at equivalent input levels).

Improved strains of *G. barbadense*, often referred to as "Sea Island", "Egyptian", or "Pima" cottons, are cultivated in the USA, Egypt, India, Former USSR, China, South America, and elsewhere and account for about 10% of the world's cotton. Sea Island cottons were a key progenitor of modern *G. barbadense* cultivars. Egyptian cotton is believed to have been derived from the hybridization of a Sea Island and a *G. barbadense* strain called Jumel's Tree Cotton (Niles and Feaster 1984). Pima, or American-Egyptian cotton traces back to both Sea Island and Egyptian cottons (as described in detail elsewhere; Niles and Feaster 1984). Stephens (1976) observed that occasional progeny of crosses between primitive *G. barbadense* cottons and *G. hirsutum* cottons resembled Sea Island cottons. This observation, together with additional evidence from the "corky" locus and nectary fringe hairs (Stephens 1974) led to the suggestion of an interspecific origin of Sea Island cultivars.

A detailed molecular map (Reinisch et al. 1994) provides the tools necessary to analyze the introgression of cotton chromatin in unprecedented detail. Percy and Wendel (1990) demonstrated *G. hirsutum* introgression into *G. barbadense* cultivars based on 24 isozyme loci. However, 24 loci can only detect linkage to a maximum of 1400 cM, or about 28% of the cotton genome (see Reinisch et al. 1994). Recombination during the course of *G. barbadense* breeding would further reduce the power of widely dispersed markers to identify introgressed chromatin.

The purpose of the study described here was to determine the levels, chromosomal distribution, and patterns across the gene pool of *G. hirsutum* chromatin in cultivated *G. barbadense* germplasm. By studying *G. barbadense* collections from breeding programs in different parts of the world, we sought to distinguish primary introgression events common to many *G. barbadense* genotypes from more recent events presumed to be a result of occasional crosses with *G. hirsutum* cottons. The persistence of common regions of introgression across many *G. barbadense* collections could reflect selective advantages accruing to genotypes introgressed in these regions, or it could simply indicate that *G. barbadense* germplasm has a narrow genetic base.

Materials and methods

Plant materials

Plant materials (Table 1) were generously provided by A. E. Percival, R. G. Percy, and J. F. Wendel. These 54 collections included Sea Island, Egyptian, and Pima cottons, and represented major breeding programs in most of the major cotton-growing countries of the world. To better represent the aggregate genotype of an accession, we extracted genomic DNA from bulk samples of at least four plants per collection, as described by Paterson et al. (1993).

RFLP analysis

A total of 106 RFLP loci distributed over the 22 largest linkage groups of the cotton molecular map (Reinisch et al. 1994) were selected for analysis. The initial set of markers were chosen at about 30 cM intervals and span approximately 3300 cM, or about 65% of the cotton genome. Additional markers were used to characterize particular regions in more detail, as discussed in the text. RFLP mapping procedures are described in Reinisch et al. (1994).

Data analysis

Genetic distances among all pairs of genotypes were estimated using Modified Rogers' Distance (MRD: Rogers 1972). Associations among genotypes were determined by PROC CLUSTER, subroutine WARD's minimum-variance method, and plotted with PROC TREE (SAS 1989).

The inference that alleles common to *G. hirsutum* and *G. barbadense* were a result of introgression rather than variation among *G. barbadense* types was made based upon two independent pieces of data. First, we considered only those RFLP loci that showed alleles common to *G. barbadense* acc. K101 (*G. barbadense*) and *G. darwinii* acc. PW44 but a different allele common to several *G. hirsutum* cultivars and breeding lines ('TM-1', 'CAMD-E', 'Deltapine 50', 'Stoneville 825', 'Paymaster 145', 'Acala 1517-75', *G. hirsutum* race palmeri-race stock Tx01, and *G. hirsutum* race stock T25, as detected by A. H. Paterson and J. Dong, unpublished). This moderately reduced the number of loci that could be assayed and also precluded detection of *G. barbadense* introgression into *G. hirsutum* (which will be addressed under a separate cover). Second, evidence of introgression was considered unequivocal if and only if several linked DNA markers all showed the *G. hirsutum* allele in a (putative) *G. barbadense* background – such an event would be highly unlikely to occur by any means other than the introgression of a chromosome segment.

Lengths of introgressed chromosomal segments were estimated from prior mapping of the DNA markers (Reinisch et al. 1994) and using the following guidelines:

1) The minimum length of an introgressed chromosome segment is the distance spanned by all consecutive loci homozygous for the *G. hirsutum* allele. For example, the minimum length of the introgressed segment in collection K271-Sind S. I. on chromosome 1 was 20 cM (from marker *A1257* to *A1794*, Fig. 2),
2) The expected length of the introgressed chromosome segment is the minimum length, plus half of the distance between the terminal *G. hirsutum* allele and nearest *G. barbadense* allele. For example, the expected length of the introgressed segment in collection K271-Sind S. I. on chromosome 1 was 30.5 cM (from marker *pAR121* to *A1794*, Fig. 2),
3) If two introgressed regions were separated by a non-introgressed region, the length of the region most closely overlapping with the "concensus" introgressed site was used,
4) Incomplete genotypes were inferred in a manner that minimized introgression,
5) Heterozygous loci were considered *not* introgressed. Few were detected, and outcrosses were deemed at least as likely a cause of heterozygosity (or heterogeneity) as introgression.

Results

Cluster analysis of *G. barbadense* germplasm

Diversity among the *G. barbadense* collections for DNA marker alleles permitted the clustering of cultivars into groups that approximately reflect breeding lineages. The total variability explained by the clusters was 47.7%, and variability among groups was significantly ($P = 0.0001$) larger than within groups (Fig. 1). The first division, cluster H, separated 5 collections with a very high frequency of *G. hirsutum* alleles from all of the remaining collections. The second division included most Sea Island cottons (cluster G). Subsequent clusters delineated ancestral (group A) and modern (group C) Pima cottons, and Egyptian cottons from the Giza (C) and Bahtim (E) series. Group B included 5 obsolete collections from Peru ('Mollendo'), Colombia ('Lengupa'), Egypt ('Domains Sakel') , and USA ('Pima 3–79'), with an average of 7.3% *G. hirsutum* introgression. The only collection in group F, K237, was originally Yuma, which appears under its common name in group A. The high frequency of *G. hirsutum* alleles in 'K237' suggests that it has been contaminated by outcrossing to *G. hirsutum*.

Frequency of *G. hirsutum* alleles among the collections and across the genome

Among the 54 collections studied, the average frequency of *G. hirsutum* alleles was 8.9% (excluding 'Nevis Sea Island', 'Pima S-2', *G. barbadense* × *G. harknessii*, S. I. × 'Spears Green', and 'Spears Green', which each had more than 70% *G. hirsutum* alleles: if included, the average was 15.9%). A primitive Peruvian cultivar, 'Tanguis 45', had the lowest frequency of *G. hirsutum* alleles (3.4%). A total of 36 genotypes (66.7%) had less than 10% *G. hirsutum* alleles (Table 1). Thirteen genotypes had from 20–30% *G. hirsutum* alleles. Pima cultivars (7.3% *G. hirsutum* alleles) had fewer *G. hirsutum*

Table 1 Frequency of *Gossypium hirsutum* RFLP alleles among 54 obsolete and modern *G. barbadense* collections

Name	Collection no. (or source)[a]	Origin	% Introgression
Sea Island (AZK263)	GB245	SI	8.3
SI Seaberry	GB239	SI	9.8
Seaberry	GB246	SI	8.4
K275-OSI Ordinary	GB257	SI	9.3
K271-Sind S.I.	GB253	SI	9,8
K272-Russian S.I.	GB254	SI	4.0
K273-Russian	GB255	SI	8.3
K274-P. R. Regular	GB256	SI	8.8
Monserrat S.I.	GB228	SI	6.4
Nevis Sea Island	SA702	SI	91.2
K245 Nevis S.I.	GB227	SI	7.8
Feji Sea Island (sic)	SA704	SI	9.9
K276-Fiji S.I.	GB258	SI	8.3
SI Barbados	GB229	SI	11.3
Barbados Sea Island	SA705	SI	8.3
St Kitts Superfine	GB226	SI	10.8
St Vincent Superfine L^0L^0	GB249	SI	12.3
K270-St Vincent 135	GB252	SI	10.3
K277-BD St Vincent Rivers	GB259	SI	9.3
K269-Superfine V46	GB251	SI	8.8
K248 Ashmouni	GB230	Egyptian	7.3
K249 Bahtim 163	GB231	Egyptian	11.7
Bahtim 185	GB232	Egyptian	13.1
Brown Egyptian	GB234	Egyptian	8.8
Domains Sakel	K216	Egyptian	7.8
Giza7	GB236	Egyptian	10.7
Giza70	R. Percy	Egyptian	7.8
Giza75	R. Percy	Egyptian	8.2
Giza76	R. Percy	Egyptian	11.5
Yuma	R. Percy	Pima	8.2
K240-AMSAK	GB222	Pima	5.3
K241-Pima32	GB223	Pima	5.3
Earlipima	GB241	Pima	4.4
Pima 3-79	D. Stelly	Pima	6.3
PimaS1	GB224	Pima	8.7
PimaS2	GB225	Pima	85.9
PimaS3	SA1478	Pima	7.3
PimaS4	SA1479	Pima	11.5
PimaS5	SA1497	Pima	6.8
PimaS6	R. Percy	Pima	7.0
PimaS7	R. Percy	Pima	9.8
Mollendo	B377	Peruvian	4.7
Tanguis 45	R. Percy	Peruvian	3.4
Lengupa	B315	Columbia	9.3
Coastland RN-45	R. Percy	U. Georgia	5.8
K3130 Maared	GB243	Unknown	5.4
K3104	GB244	Unknown	13.4
K278-PEI152413	GB260	Unknown	11.3
K237 "Yuma"	GB219	Unknown	29.1
Barb. Tashkent St.	GB242	Russian	5.9
K268-VH Hybrid	GB250	Hybrids	10.7
G. barbadense × *G. harknessii*	SA121	Hybrids	73.5
S. I. × Spears Green	SA895	Hybrids	84.3
Spears Green	SA896	Hybrids	90.1

[a] SA no.: originally assigned by Mississippi Obsolete Variety Collection; K no.: originally assigned by Arizona K collection (*G. barbadense*); GB no.: originally assigned by Arizona *G. barbadense* collection, all as listed in the National collection of *Gossypium* germplasm (SCSB N. 321, June 1987)

alleles than Sea Island (9.0%) and Egyptian collections (9.6%).

Five genomic regions, cumulatively spanning 474 cM (9% of the genetic map), accounted for 57.5% of the putative *G. hirsutum* alleles detected in these 54 *G. barbadense* collections (Fig. 2). All of the collections showed introgression in at least one of these regions. The highly introgressed regions were equally distributed across subgenomes (two in A, two in D, one uncertain), and no cases of homoeology and no significant deviation from Hardy-Weinburg segregation ratios were apparent in a previously studied *G. hirsutum* × *G. barbadense* cross (Reinisch et al. 1994).

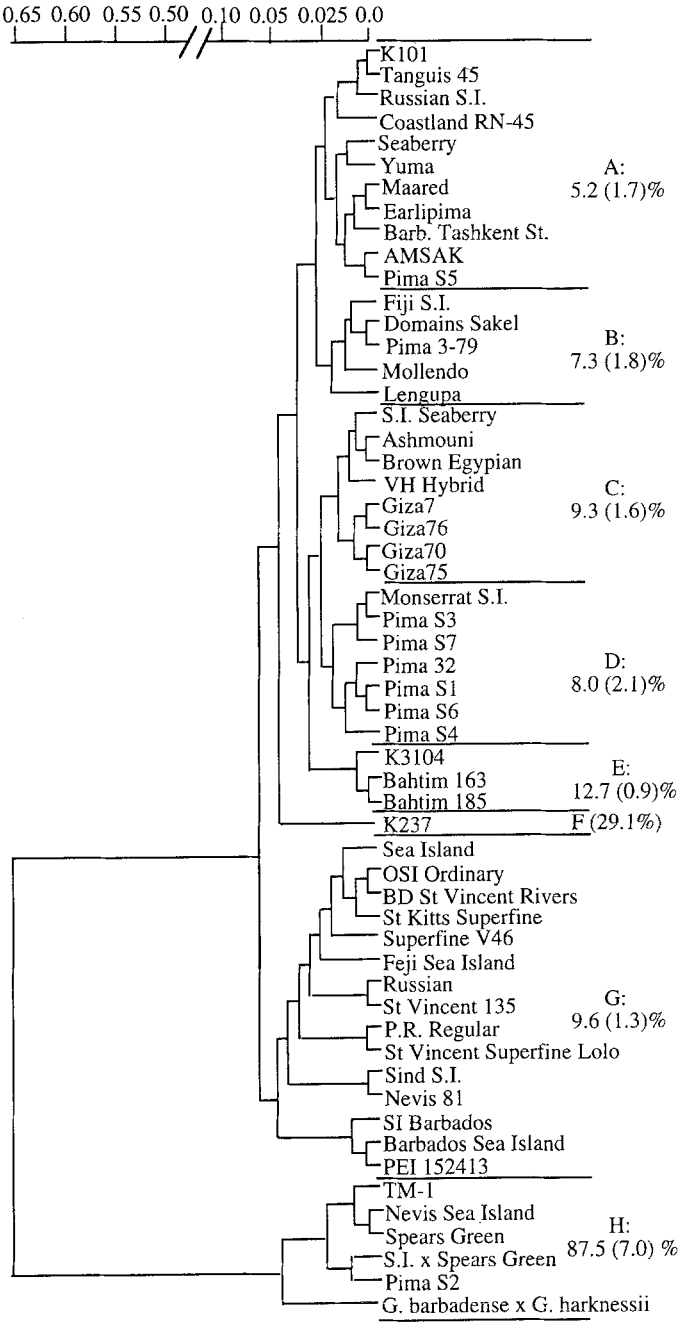

Fig. 1 Phenogram of 54 *G. barbadense* cultivars and *G. barbadense* × *G. hirsutum* hybrids produced by average linkage cluster analysis using Roger's genetic distance. The average (*a*) and standard deviation (*b*) of % *G. hirsutum* alleles are shown to exemplify differences among the eight clusters that were deemed meaningful. Note that 'K101' (*G. barbadense*) and 'TM-1' (*G. hirsutum*) are excluded from the calculations but are included in the phenogram for reference

Four genomic regions, cumulatively spanning 125 cM (2.5% of the genetic map), were found to have no introgression (defined as at least 2 consecutive markers with 0% *G. hirsutum* alleles: Fig. 3). The rarely introgressed regions were similarly distributed across subgenomes (one in A, two in D, one uncertain), with no cases of homoeology. The region on chromosome 22 is associated with significant deviation from the Hardy-Weinburg expectation for segregation (Reinisch et al. 1994), but the deviation is a deficiency of *G. barbadense* homozygotes, contrary to the deficiency of *G. hirsutum* chromatin we found in our study.

Detailed mapping of introgressed chromosomal regions

The five chromosomal regions displaying the most prominent introgression were characterized in more detail with additional DNA markers (Fig. 2) and can be described as follows:

Chromosome 1 (A-subgenome)

Marker *A1257* best delineates a region in which the *G. hirsutum* allele is found in 48 (94%) of the 51 *G. barbadense* collections for which data were conclusive (Fig. 3). The mean minimal length of the introgressed segment is 14.4 cM, while the mean length of the introgressed segment after adjustment for the contribution of flanking intervals is 38.9 cM (Table 2). The latter calculation is undoubtedly biased upward due to the long distance (110 cM) between *pAR121* and *A1549*. The concensus region of introgression is to the right (as drawn) of *A1097* and to the left (as drawn) of *pAR121*. Most Sea Island collections carry *G. hirsutum* alleles at both *A1257* and *A1097* (9 cM apart). While the earliest available Egyptian cotton, 'Ashmouni', is not introgressed in this region, most subsequent Egyptian cottons are introgressed at *A1257*. Virtually all Pima cottons are likewise introgressed at *A1257* but not *A1097*, and most have an introgressed region that extends to *pAR121*, 21 cM from *A1257*.

Chromosome 5 (A-subgenome)

Marker *pAR112* best delineates a region in which the *G. hirsutum* allele is found in 38 (79%) of 48 *G. barbadense* collections for which data were conclusive (Fig. 3). The mean minimal length of the introgressed segment is 15.6 cM, while the mean length of the introgressed segment after adjustment for flanking intervals is 30.8 cM (Table 2). The concensus region of introgression appears to lie slightly to the right (as drawn) of *pAR112*, based on the observation that 'Nevis SI' (K245), the only introgressed collection that retains the *G. barbadense* allele at *pAR112*, shows the *G. hirsutum* allele at *PXP1-9a*, 32 cM to the left.

Several unexpected genotypes are found along chromosome 5. The Egyptian cultivar 'Ashmouni' and several subsequent Egyptian cultivars exhibit a marker genotype in this region that requires four recombination events to explain, with *G. hirsutum* alleles at *pAR112* and *P12-20a* but not at the intervening or flanking loci. A

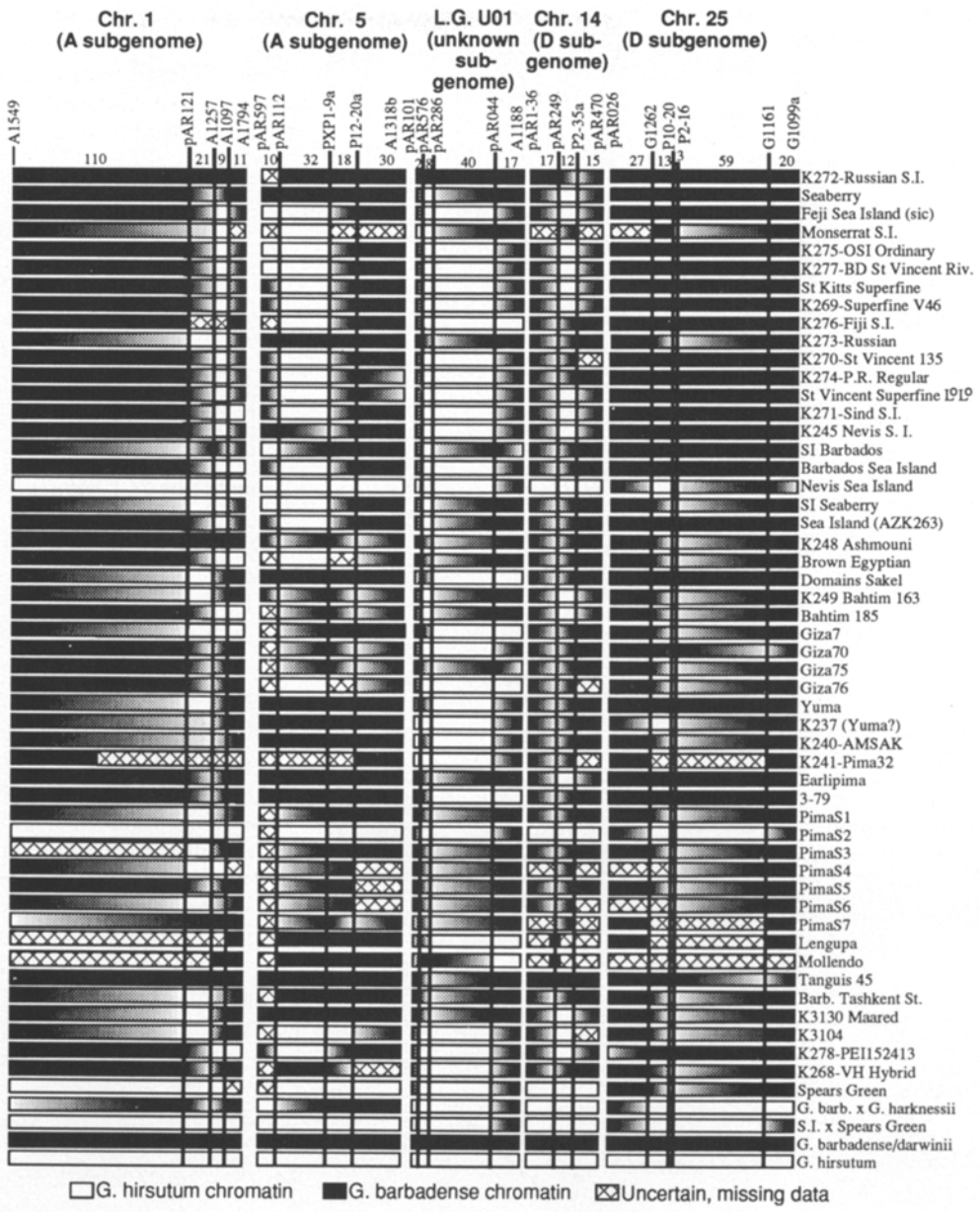

Fig. 2 Chromatin from *G. hirsutum* most commonly introgressed into *G. barbadense* cultivars. Chromosomes or linkage groups, DNA markers, and distances between markers are shown as previously determined (Reinisch et al. 1994). *Open regions* are flanked by loci homozygous for the *G. hirsutum* allele, while solid-filled regions are flanked by loci homozygous for the *G. barbadense* allele. *Gray-scale regions* harbor a recombination site, with the *intensity of shading* used to emphasize the genotypes of the respective flanking loci (as shown in legend). *Cross-hatched regions* represent missing data

similar recombinant chromosome segment has either recurred in, or been transmitted to, 'Pima S-7'. Older Pima collections, specifically 'Yuma', 'Earlipima', and '3–79' appear to be largely free of introgression along chromosome 5, while 'Pima S-1' and more recent Pima collections carry the *G. hirsutum* allele at both *P2-16* and *P10–20* (3 cM away). This unusual chromosome segment appears to have first occurred early in the Egyptian breeding program, perhaps being derived from 'Jumel's Tree Cotton' as such a genotype is not found among any of the Sea Island collections studied ('Ashmouni' is derived from a cross between a Sea Island type, and 'Jumel's'; Niles and Feaster 1984). However, two Sea Island collections, 'Puerto Rican Regular' and 'St Vincent Superfine L^0L^0', exhibit a "mirror-image" event, with *G. barbadense* chromatin at *P12-20a* flanked by *G. hirsutum* chromatin on each side. These unusual events are not explained by ambiguity in the order of markers along the genetic map – the distances between these markers (Reinisch et al. 1994) permit accurate ordering.

L G. U01(uncertain subgenome)

Marker *pAR286* best delineates a region in which the *G. hirsutum* allele is found in 53 (98%) of the 54 *G. barbadense* collections studied (Fig. 3). The mean minimal length of the introgressed segment is 26.9 cM, while the mean length of the introgressed segment after adjustment for the contribution of flanking intervals is 42.2 cM (Table 2). This is the most frequently introgressed region

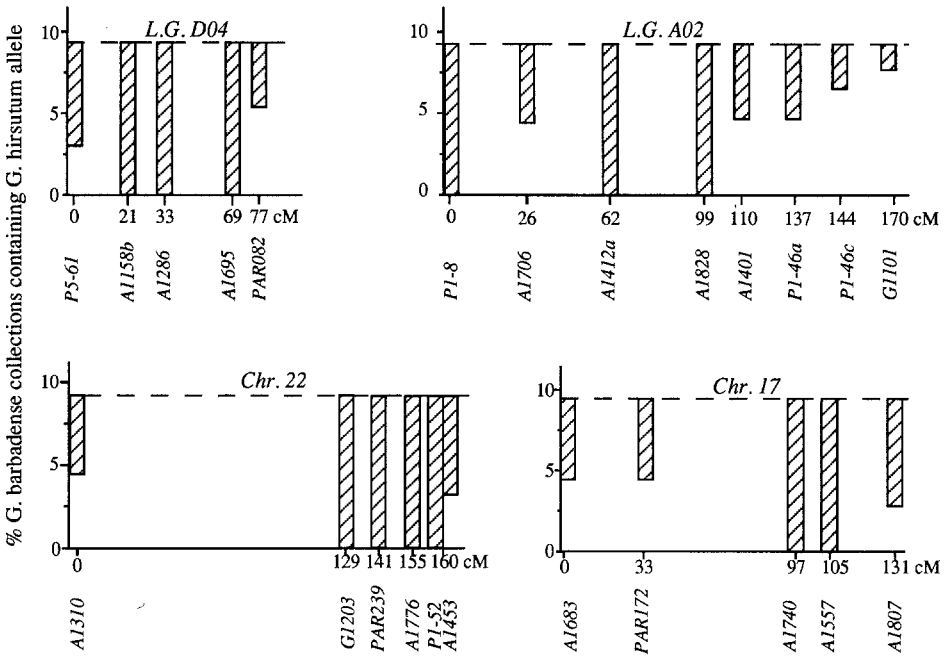

Fig. 3 Chromatin from *G. hirsutum* least-commonly introgressed into *G. barbadense* cultivars. Chromosomes or linkage groups, DNA markers, and distances between markers are shown as previously determined (Reinisch et al. 1994). The *dotted line* drawn across each graph at 9% represents the average frequency of *G. hirsutum* alleles in *G. barbadense* cultivars across all loci and cultivars studied. *Bars* start at 9%, and indicate the frequency of *G. hirsutum* alleles in *G. barbadense* cultivars at the corresponding RFLP loci studied

Table 2 Estimated lengths of individual chromosome segments introgressed from *G. hirsutum* into *G. barbadense* cultivars

Chromosome or linkage group[a]	Concensus marker	Minimum length[b]	Adjusted length[c]
1	*A1257*	14.4	38.9
5	*pAR112*	15.6	30.8
U01	*pAR286*	26.9	42.2
14	*pAR249*	5.3	19.8
25	*P2-16*	2.3	38.0
Grand mean		12.9	33.9

[a] As described in Reinisch et al. (1994)
[b] The distance spanned by all consecutive loci homozygous for the *G. hirsutum* allele
[c] The minimum length (above), plus half the distance between the terminal *G. hirsutum* allele and nearest *G. barbadense* allele. See Materials and methods for more details

found, with only 'K272'-'Russian Sea Island' exhibiting the same genotype as *G. barbadense* 'K101' and *G. darwinii* 'PW44'. This introgressed segment also shows the greatest length of any introgressed segments we found.

Chromosome 14 (D-subgenome)

The marker *pAR249* best delineates a region in which the *G. hirsutum* allele is found in 50 (96+) of the 52 *G. barbadense* collections for which data were conclusive (Fig. 3). The mean minimal length of the introgressed segment is 5.3 cM, while the mean length of the introgressed segment after adjusting for the contribution of flanking intervals is 19.8 cM (Table 2). A total of 25 (46%) *G. barbadense* cultivars show introgression only at the *pAR249* locus, while in an additional 15 (28%) the introgressed segment extends at least 12 cM to the *P2-35a* locus. The observation that 1 collection, 'K272'-Russian Sea Island', was introgressed at the *P2-35a* locus but not the *pAR249* locus suggests that the concensus region of introgression is between *pAR249* and *P2-35a* (e.g., slightly to the right of *pAR249*, as drawn). 'Tanguis 45' and 'Coastland RN-45', the only collections that exhibit the same genotype as *G. barbadense* 'K101' and *G. darwinii* 'PW44' in this region, are closely related (Jenkins 1953).

Chromosome 25 (D-subgenome)

Marker *P2-16* best delineates a region in which the *G. hirsutum* allele is found in 28 (56%) of the 50 *G. barbadense* collections for which results were certain (Fig. 3). The mean minimal length of the introgressed segment is 2.3 cM, while the mean length of the introgressed segment after adjustment for the contribution of flanking intervals is 38 cM (Table 2). The former is clearly an underestimate, due to the use of 2 very closely-spaced markers, while the latter is probably biased upward by the 59 cM distance between *P2-16* and *G1161*. The concensus region of introgression is slightly to the right (as drawn) of *P2-16*, based on the observation that 'Tanguis 45' and 'Giza 70' are introgressed only at *G1161*, 59 cM to the right.

Curiously, most Sea Island collections are free of *G. hirsutum* alleles in this region. In contrast, all but 2 Egyptian collections carry the *G. hirsutum* allele at *P2-16* and *P10-20* (3 cM away). Several of the first Pima collections, specifically 'Yuma', 'Earlipima', and '3–79' are apparently free of introgression, while 'Pima S-1' and

subsequent Pima collections carry the *G. hirsutum* allele at both *P2-16* and *P10-20* (3 cM away).

Discussion

Our results support the hypothesis of Stephens (1976) that improved *G. barbadense* cottons are of interspecific origin, and we document the specific *G. hirsutum* chromosome segments found in Sea Island cottons and their antecedents. Moreover, we complement the results of Percy and Wendel (1990), with a more comprehensive picture of *G. hirsutum* introgression into *G. barbadense*. Percy and Wendel (1990) reported that 21.7% of improved *G. barbadense* cultivars are introgressed at 1 or more of 24 isozyme loci, with individual cultivars being introgressed at up to 50% of loci. The more extensive genome coverage afforded by the genetic map (Reinisch et al. 1994) revealed that 100% of improved *G. barbadense* cultivars are introgressed, with individual cultivars being introgressed at up to 85.9% of loci (excluding 'Nevis Sea Island', a probable anomaly).

Sources of *G. hirsutum* chromatin in cultivated *G. barbadense*

It is likely that introgression of *G. hirsutum* chromatin occurred fairly recently, e.g, during the course of breeding improved *G. barbadense* strains. Naturally occurring genetic exchange most commonly involves the introgression of *G. barbadense* chromatin into *G. hirsutum* (Brubaker et al. 1993), thus is unlikely to account for the introgression observed herein. A minimum estimate of the length of *G. hirsutum* chromosome segments in *G. barbadense* genotypes was 12.9 cM (Table 2), which was determined by calculating the average distance (from Reinisch et al. 1994) spanned by linked loci introgressed into each cultivar for each of the five commonly introgressed regions. A breeding program equivalent to eight backcrosses would account for a 12.9-cM average length of introgressed chromatin segments (Hanson 1959b). This suggests that a relatively small number of opportunities for recombination may have passed since introgression of the *G. hirsutum* chromatin. The region of chromosome 25 near *G1161* has clearly become prominent in the gene pool of cultivated *G. barbadense* only recently, as it shows no introgression in the vast majority of Sea Island cotton examined but is introgressed in most Egyptian and Pima cottons.

Four of the five commonly introgressed regions in *G. barbadense* can be accounted for by introgression already present in (putatively ancestral) Sea Island cotton. Hutchinson and Manning (1945) proposed a Peruvian origin of Sea Island cotton, a proposal supported both by Percy and Wendel (1990) and our finding that Peruvian cotton showed the lowest levels of *G. hirsutum* introgression (Table 3), suggesting the greatest similarity to wild *G. barbadense*. While the specific pathway of domestication of Sea Island cotton remains somewhat clouded (Percy and Wendel 1990), introduction of *G. hirsutum* chromatin at specific sites in the genome appears to have at least accompanied, and perhaps been partly responsible for, Sea Island improvement.

Table 3 Variation in frequency of *G. hirsutum* alleles among different classes of *G. barbadense* cultivars

Type of cultivar	Number	Range (%)	Average (SD) (%)
Sea Island[a]	20	4.0–12.3	9.0 (1.8)
Pima[b]	12	5.3–11.5	7.3 (2.1)
Egyptian	9	7.8–13.1	9.6 (2.1)
Peruvian	2	3.4– 4.7	4.1 (0.9)
Miscellaneous others	11	5.8–90.1	30.8 (34.1)

[a] Excluding 'Nevis Sea Island' (91.2% *G. hirsutum*). If included, average is 13.1%
[b] Excluding 'Pima S-2' (85.9% *G. hirsutum*). If included, average is 13.8%

Persistence of *G. hirsutum* chromatin in the cultivated gene pool of *G. barbadense*

Some introgressed chromatin may persist in particular breeding programs as a result of repeated crossing between closely related genotypes. For example, the recurrence in several Egyptian cottons of a segment of chromosome 5 which requires four distinct chromatid exchanges to be accounted for (Fig. 2) seems more likely to be due to inbreeding than to multiple independent occurrences.

However, the finding of recombination events in several different intervals surrounding common introgressed regions shows that there have been opportunities for chromatin of *G. hirsutum* and *G. barbadense* to recombine and/or for introgressed chromatin to be eliminated. The finding of at least 1 *G. barbadense* cotton which was not introgressed in each of the five regions (based on the markers studied) corroborates that the *G. barbadense* gene pool is not monomorphic for *G. hirsutum* alleles in these regions.

At least some *G. hirsutum* chromosome segments may have been maintained in the *G. barbadense* gene pool due to favorable direct effects and/or interactions with *G. barbadense* alleles. This postulate is based on the persistence of several regions of *G. hirsutum* chromatin in *G. barbadense* cotton through multiple cycles of recombination (as evidenced by the overlapping, but different, *G. hirsutum* chromosome segments) and on the observation that new *G. hirsutum* chromosome segments have become prominent in the *G. barbadense* gene pool during the course of modern plant breeding. It has previously been suggested that the introgression of *G. hirsutum* chromatin into Pima cotton has broadened the genetic base, increased productivity, and improved the plant type and adaptability to environmental stress (Meredith 1991).

Several regions of the genome appeared free of introgression (Fig. 3). It is difficult to place any importance on such regions in this study, since the average level of introgression was only 9%. However, if the genes in these regions did impart desirable *G. barbadense* fiber characteristics, selection might have insulated them from introgression. The mapping of quantitative trait loci accounting for differences in fiber quality between *G. hirsutum* and *G. barbadense* will elucidate this possibility.

G. barbadense genotypes with anomalous genome composition

Two cultivars, 'Pima S-2' and 'Nevis Sea Island', had unexpectedly high frequencies of *G. hirsutum* alleles. 'Pima S-2' was putatively derived from a cross between 'Pima S-1' and '3-79' (Feaster and Turcotte 1962), which showed only 8.7% and 6.3% *G. hirsutum* alleles, respectively. On the basis of published data (USDA-ARS 1987), 'Pima S-2' is similar to *G. hirsutum* for several traits, most notably high yield, short stature, and early maturity. Anecdotal information suggests that 'Pima S-2' may have been the result of an outcross (E. Turcotte, personal communication). Similarly, two different accessions both designated 'Nevis Sea Island' show 7.8%, and 91% *G. hirsutum* alleles, respectively. The latter is likely to be due to outcrossing or other errors during maintenance of genetic stocks (R. Percy, personal communication).

Cultivar identification (DNA fingerprinting) in *G. barbadense*

No two *G. barbadense* collections we studied carried the same repertoire of *G. hirsutum* chromosome segments. Consequently, *G. hirsutum* introgression provides a simple means for DNA fingerprinting of *G. barbadense* collections using the DNA markers described herein.

Different *G. barbadense* collections with similar names sometimes showed markedly different genotypes. For example, 'SI Seaberry' (GB239) and 'Seaberry' (GB246) showed marked differences in the repertoire (chromosomes 5, 25) and length (chromosomes 1, 14) of introgressed chromosome segments. In contrast, 'Feji' (sic) 'Sea Island' (SA704) and 'Fiji S. I.' (GB258) share a common repertoire of introgressed segments, although with minor differences in the length of two segments (l.g. U03, chromosome 14). Relatively informal documentation of names of Sea Island varieties prior to 1900 (Brown and Ware 1958) may suggest "convergence" in the names of some Sea Island collections. Consequently, we have sought to place minimal weight on evidence from any single collection and instead studied the patterns of introgression across groups of cultivars.

Analysis of *G. barbadense* as a model for molecular characterization of introgressive plant breeding

The study of genome-wide introgression from *G. hirsutum* into *G. barbadense* extends the concept of "tagging" introgressed chromatin harboring specific genes (Osborn et al. 1987; Tanksley and Hewitt 1988; Young et al. 1988; Sarfatti et al. 1989; Jena et al. 1992) to global characterization of many introgression events that may have been associated with long-term selection for complex measures of productivity and/or quality.

Interspecific introgression has been implicated as a primary agent of genetic change in many plant groups (Stebbins 1969; Grant 1981; Rieseberg et al. 1988). Several lines of evidence hint that retention of *G. hirsutum* chromatin in common regions of *G. barbadense* collections could be due to selection for favorable phenotypes. While gross phenotypic differences between cluster H (*G. hirsutum*-like) and the remaining *G. barbadense* cottons are evident from published data (USDA-ARS 1987), more subtle differences likely to be associated with specific *G. hirsutum* chromosome segments in *G. barbadense* will require simultaneous multienvironment testing (which has not yet been done). Subsequent investigations will evaluate the association between levels and patterns of introgression and specific complex traits in cotton.

Acknowledgments The authors thank A. E. Percival, R. Percy, and J. Wendel for seed, K. El-Zik, J. Gannaway, R. Percy, C. W. Smith, J. Wendel, and the anonymous reviewers for helpful comments, and F. M. Xie for help with cluster analysis. Aspects of this work were supported by the Texas Higher Education Coordinating Board (ATP 999902-148 to AHP), Texas Agricultural Experiment Station, and USDA-National Research Initiative (91-37300-6570 to AHP).

References

Bryan WE (1955) Prospects for Pima S-1 cotton. Progr Agric Ariz 7:6

Brown HB, Ware JO (1958) Cotton, 3rd edn. McGraw-Hill, New York

Brubaker CL, Koontz JA, Wendel JF (1993) Bidirectional cytoplasmic and nuclear introgression in the New World cottons, *Gossypium barbadense* and *G. hirsutum* (Malvaceae). Am J Bot 80:1203–1208

Feaster CV, Turcotte E (1962) Genetic basis for varietal improvement of Pima cottons. USDA-ARS 34–31

Grant V (1981) Plant speciation. Columbia University Press, New York, pp 234

Hanson WD (1959b) The breakup of initial linkage blocks under selected mating systems. Genetics 44:857–868

Hutchinson JB (1959) The application of genetics to cotton improvement. Cambridge University Press, London

Hutchison JB, Manning HL (1945) The Sea Island cottons. Emp J Exp Agric 13:80–92

Jena KK, Khush GS, Kochert G (1992) RFLP analysis of rice (*Oryza sativa* L.) introgression lines. Theor Appl Genet 83: 1027–1034

Jenkins JG (1953) Coastland – a new long-staple cotton for the Southeast. Ga Coastal Plain Exp Stn Bull 53

Kearney TH (1943) Egyptian-type cottons: their origin and characteristics. Report Division Cotton and other Fiber Crops and Diseases, USDA Mimeo

Meredith WR (1991) Contribution of plant introductions to cotton improvement. In: Shands HL, Wiesner (eds) Use of plant introductions in cultivar development, part I. Crop Sci Soc Am, Madison, Wis., pp 127–146

Niles GA, Feaster CV (1984) Breeding. In: Kohel RJ, Lewis CF (eds) Cotton Agron Monogr 24, ASA, Madison, Wis., pp 201–231

Osborn TC, Alexander DC, Fobes JF (1987) Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. Theor Appl Genet 73:350–356

Paterson AH, Brubaker CL, Wendel JF (1993) A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP and PCR analysis. Plant Mol Biol Rep 11:122–127

Percy RG, Wendel JF (1990) Allozyme evidence for the origin and diversification of *Gossypium barbadense* L. Theor Appl Genet 79:529–542

Reinisch AJ, Dong D, Brubaker CL, Stelly DM, Wendel JF, Paterson AH (1994) A detailed RFLP map of cotton, *G. hirsutum* × *G. barbadense*: chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829–847

Rieseberg LH, Soltis DE, Palmer JD (1988) A molecular reexamination of introgression between *Hellanthus annuus* and *H. bolanderi* (compositae). Evolution 42:227–238

Rogers JS (1972) Measurements of genetic similarity and genetic distance. (Studies in Genetics VII). University of Texas Publication 7213, Austin, Tex.,

Sarfatti M, Katan J, Fluhr R, Zamir D (1989) An RFLP marker in tomato linked to the *Fusarium oxysporum* resistance gene 12. Theor Appl Genet 78:755–759

SAS Institute (1989) SAS/STAT users guide, version 6, 4th edn., vol 1. SAS Institute, Cary, N.C.

Stebbins GL (1969) The significance of hybridization for plant taxonomy and evolution. Taxon 18:26–35

Stephens SG (1974) The use of two polymorphic systems, nectary fringe hairs and corky alleles, as indicators of phylogenetic relationships in New World cottons. Biogeotropica 6:194–201

Stephens SG (1976) The origin of Sea Island cotton. Agric History 50:391–399

Tanksley SD, Hewitt J (1988) Use of molecular markers in breeding for soluble solids content in tomato-a re-examination. Theor Appl Genet 75:811–823

USDA-ARS (1987) The national collection of *Gossypium* germplasm. South Coop Ser Bull 321

Young ND, Zamir D, Ganal M, Tanksley SD (1988) Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the *TM2a* gene in tomato. Genetics 120:579–585