Genetic Similarity Indices for Ancestral Cotton Cultivars and their Impact on Genetic Diversity Estimates of Modern Cultivars

G. A. Van Esbroeck, D. T. Bowman,* O. L. May, and D. S. Calhoun

ABSTRACT

The accuracy of pedigree-derived genetic-distance estimates depends on the availability of breeding records and on the validity of a number of assumptions. Pedigree analysis for cotton (*Gossypium hirsutum* L.) indicates a wide genetic base, which conflicts with other types of distance estimates indicating a narrow genetic base. The objectives of this study were to determine genetic similarity indices from agronomic and morphological traits for ancestral cotton cultivars and to determine their impact on the coefficient of parentage (CP) of recent cultivars. Twelve ancestral and two recent cultivars were grown at three locations (Clayton, NC; Florence, SC; and Stoneville, MS) during 1995 and evaluated for agronomic and fiber properties. Multivariate analysis of agronomic and fiber properties was used to create genetic similarity indices (*s*). There was little agreement between CP and *s*. Several ancestral cultivars developed from the earliest (pre-1800) introductions into the USA were very similar to the most recent (post-1900) introductions from Mexico. This suggested that all the original Mexican introductions were genetically very similar. Changing the assumptions in pedigree analysis to assume the original introductions were related by the average *s* among the ancestral cultivars ($s = 0.38$) decreased the level of diversity in frequently grown modern cultivars from CP = 0.16 to CP = 0.46 but had little effect on the relative ranking of cultivars. This occurred because effects of possible erroneous pedigree assumptions in distant generations became diluted. The high degree of similarity among ancestral cultivars suggests that pedigree-derived genetic distance estimates have overestimated the genetic diversity among today's cultivars but not relative relationships.

MAINTAINING GENETIC DIVERSITY among crop genotypes offers a measure of protection against potential widespread losses from crop pests and facilitates the creation of segregating populations from which plants with superior gene combinations can be selected. Accurate estimates of genetic distance may assist in crop improvement strategies. Genetic distances can be estimated from pedigree analysis or from multivariate analysis carried out on a large number of plant attributes (DNA polymorphisms, isozymes, morphological features, or agronomic performance). An advantage of pedigree-derived genetic distance estimates is their low cost, compared with the other estimates that require more time and resources to obtain. A disadvantage to the pedigree-derived genetic distance estimates is that their accuracy depends on the availability of accurate breeding records and on the validity of a number of assumptions.

G.A.Van Esbroeck and D.T. Bowman, Dep. of Crop Science, Box 8604, North Carolina State University, Raleigh, NC 27695-8604; O.L. May, USDA-ARS and Clemson University, Florence, SC 29506-9706; D.S. Calhoun, Mississippi Agriculture and Forestry Experiment Station, Delta Branch Exp. Stn., Stoneville, MS 38776. Received 10 Nov. 1997. *Corresponding author (daryl_bowman@ncsu.edu).

Published in Crop Sci. 39:323–328 (1999).

Despite widespread availability of pedigree-derived genetic distance estimates in cotton for the purposes of selecting parents (Calhoun et al., 1997; Bowman et al., 1997) and in monitoring genetic vulnerability (Van Esbroeck et al., 1998), the validity of the estimates have not been verified. Coefficient of parentage estimates indicate a greater level of genetic diversity in cotton than for most crops. Bowman et al. (1996) reported an average CP of 0.07 for cotton cultivars released between 1970 and 1990. In comparison, typical CP values among recently released cultivars were 0.21 for peanut (*Arachis hypogaea* L.) (Knauft and Gorbet, 1989), 0.16 to 0.24 for wheat (*Triticum aestivum* L.) (Cox et al., 1985b; Murphy et al., 1986), 0.13 for soybean [*Glycine max* (L.) Merr.] (Gizlice et al., 1993) and 0.08 for oat (*Avena sativa* L.) (Souza and Sorrells, 1989).

Use of CP to compare the relative levels of genetic diversity among crops may not be valid, as the accuracy of the CP estimates may vary widely among crops. There are indications that the level of diversity in cotton, as determined by pedigree-derived genetic distances, may be overestimated. Over 30% of the cotton cultivars released between 1970 and 1990 were reselections (Bowman et al., 1996). Although cotton is normally an autogamous crop, cross-pollination rates as high as 50% have been reported when bees (*Bombus* spp.) were present (Loden and Richmond, 1951). Pedigree analysis normally treats a reselection as a compromise between an outcross to an unknown and a self pollination (Murphy et al., 1986; Bowman et al., 1997). The frequent reselections in cotton have the effect of continually incorporating unrelated germplasm into the CP estimates and, hence, probably exaggerate the diversity among cultivars. Moreover, in calculating CP for cotton, Bowman et al. (1997) assumed that all ancestors were equally unrelated. If, however, the original introductions were closely related to each other, genetic distances would be overestimated (MacCluer et al., 1983). Recent studies with isozymes and DNA markers showing that ancestors of modern upland cotton cultivars probably originated from a small area in Guatemala (Wendel et al., 1992; Brubaker and Wendel, 1994) suggest that ancestral cotton introductions and unknowns were closely related. Lack of DNA polymorphisms within cotton also suggests a lower level of diversity than obtained by pedigree analysis (Wendel et al., 1992; Brubaker and Wendel, 1994),

Upland cotton cultivars grown in the USA were derived from introductions that occurred during three periods: pre-1800s, early to mid-1800s, and early 1900s (Ware 1936, 1950). Prior to the 1800s, the cotton grown

Abbreviations: CP, coefficient of parentage; HVI, high volume instrument; *s*, genetic similarity index.

in the USA consisted of Asiatic cottons (Levant [*G. herbaceum* L.] and tree [*G. arboreum* L.]), green and black seeded *G. hirsutum* types from eastern Mexico, and sea island (*G. barbadense* L.) stocks from the Caribbean. These cottons were generally low yielding but some (*G. barbadense*) had excellent fiber properties. In the early to mid-1800s, a number of Mexican highland stocks were brought into the USA. These naturally hybridized or were artificially crossed to the then widely grown green seed, black seed, and sea island stocks (Moore, 1956). Breeders sought to combine the high yields of the Mexican types with superior fiber qualities of the older varieties. The introductions and hybridizations resulted in substantial variability among the types of cotton grown in the late 1800s (Moore, 1956).

When the boll weevil (*Anthonomus grandis* Boh.) became a threat to cotton production in the early 1900s, selection was undertaken for disease resistance and agronomic superiority within the often heterogeneous cultivars of the day (Ware, 1936, 1950). Extreme selection pressure for earliness led to an almost total loss of the late-maturing, long-staple, upland cottons. The cultivars Wannamaker Cleveland, Hartsville, Deltatype Webber, Express, Lone Star, and Dixie Triumph, which are included in the pedigree of most modern cultivars, were developed during this period. These ancestral cultivars are all presumed to contain genetic material from pre-1800 cultivars as well as from the Mexican introductions of the early to mid 1800s. The boll weevil epidemic also led to a number of germplasm collection trips and new introductions. The cultivars Kekchi, Young's Acala, and Hopi Moencopi, which occur frequently in pedigrees of modern cultivars, were introduced from Mexico in the early 1900s as sources of potential boll weevil resistance.

The genetic relationships among ancestors of soybean and maize (*Zea mays* L.) have been estimated through multivariate analysis of a large number of traits (Gizlice et al., 1993; Goodman and Bird, 1977). Traits previously used for multivariate analysis in cotton include morphological features (Tatineni et al., 1996), agronomic performance (Brown, 1991), isozymes (Wendel et al., 1992), and DNA polymorphisms (Brubaker and Wendel, 1994; Tatineni et al., 1996). DNA polymorphisms are considered the most suitable markers for genetic distance estimates because of their potentially large numbers and because they are not subject to environmental variation (Gepts, 1993). Use of DNA markers in cotton, however, is limited by the scarcity of polymorphic markers (Brubaker and Wendel, 1994; Tatineni et al., 1996). Morphological features that are least subject to environmental variation are most suitable for multivariate analysis. Goodman and Bird (1977) suggested using only those traits for which the ratio of cultivar variance to location and cultivar × location variance was >3.0. Where environmental influences are large, material may be grown in several environments and the response to environment used as a criteria to estimate genetic distance (Goodman and Paterniani, 1969).

Weak correlations among various estimates of genetic distance highlight the potential weakness of reliance on any single estimate. For soybean, genetic distance based on pedigree analysis and multivariate analysis based on morphological features gave relatively similar results (Cox et al., 1985a; Gizlice et al., 1993); however, Cox et al. (1985a) reported higher correlations among estimates for recent cultivars ($r = 0.60$) than for older cultivars ($r = 0.24$). In wheat, correlations between pedigree-derived genetic distance and RFLP-derived estimates have varied from $r = 0.21$ to $r = 0.78$ (Cox et al., 1985b; Kim et al., 1997). The first objective of this study was to develop genetic similarity indices for ancestral cotton cultivars, on the basis of agronomic and morphological features. A second objective was to determine the impact of these similarity indices on CP estimates among modern cultivars.

MATERIALS AND METHODS

Twelve ancestral cotton cultivars and two more recent cultivars closely related to the ancestral lines (Lankart 57 , Tamcot SP21) were grown at Clayton, NC (Norfolk sandy loam; fine-loamy, siliceous, thermic Typic Paleudult), Florence, SC (Norfolk loamy sand; fine-loamy, siliceous, thermic Typic Kandiudult), and Stoneville, MS (Dundee silty clay; fine-silty, mixed thermic Aeric Ochraqualf) during 1995 (Table 2). Cultivars were grown in a randomized, complete-block design of four replicates with each plot consisting of two, 12-m-long rows with a 1-m row spacing. Normal agronomic practices were followed for each location. Cultivars were evaluated for a range of agronomic traits. Final plant height and nodes to first fruiting limb were determined on five plants per replicate. Fiber quality traits (elongation, strength, micronaire, length, uniformity) and lint percentage for North Carolina and Mississippi were determined for each replicate from the lint of 25 hand-harvested bolls. In South Carolina, fiber quality traits were obtained by combining the lint (25 bolls) from each of two replicates. Fiber samples were evaluated using high volume instrument (HVI) analysis in Mississippi and North Carolina and by individual instruments (fibrograph, stelometer, and micronaire) in South Carolina. Fibrograph 2.5% span length was considered equal to HVI upper half mean length. Stelometer strength was multiplied by 1.3 to estimate HVI strength for the South Carolina data. Plots were machine harvested and lint yield determined as seedcotton × percentage of lint, which was obtained from hand-harvested samples.

Highly correlated traits ($r > 0.80$) were excluded from further analysis. For each remaining trait, an analysis of variance was carried out over locations, with cultivar effects considered fixed and locations random. For each trait, the ratio of cultivar variance to location and cultivar × location variance was calculated (Goodman and Bird, 1977). After standardizing traits to unit variance, principal component analysis was performed using the PROC PRINCOMP procedure (SAS Institute, 1991) and eigenvalues and principal component coordinates determined for each cultivar. Using only those principal components with eigenvalues >1 (Goodman and Bird, 1977), pairwise genetic distances (D) between all cultivars were calculated as:

$$D_{ij} = [\sum(x_{ik} - x_{jk})^2/\lambda_k]^{0.5}$$

where D_{ij} is the genetic distance between the ith and jth cultivar, x_{ik} and x_{jk} are the kth principal component scores for the ith and jth cultivar, and λ_k is the kth eigenvalue and $\lambda_k > 1.0$.

Genetic distances were converted to relative genetic similarities as:

$$S_{ij} = 1 - (D_{ij}/D_{max})$$

Table 1. Mean squares for nine traits of 12 ancestral and two recent cotton cultivars.

Effect	df	Trait								
		Plant height	First fruiting node	Lint yield	Lint percentage	Boll† weight	Micronaire	Fiber strength	Elongation	Fiber length
Location (L)	2	7 057	12.60**	25 195	1 343.40**	13.63**	14.67**	263.2**	33.90**	95.3**
Replicates/L	6–9	2 339**	0.72*	72 139**	3.89*	0.56	0.28**	5.9**	0.73*	2.2*
Cultivar (C)	13	1 611**	2.17*	143 047	243.75**	4.92**	2.67**	55.4**	6.89**	38.8**
L × C	26	215**	0.85**	82 905**	11.88**	0.80**	0.15*	4.0**	1.94**	1.5
Error	90–116	100	0.35	11 679	1.65	0.31	0.08	1.6	0.36	1.0

*, ** Significant at the 0.05 and 0.01 probability levels, respectively.
† Data are for two locations; df are 1, 6, 13, 13, and 76 for L, Replicates/L, C, L × C, and error, respectively.

where S_{ij} is the similarity between two cultivars and D_{max} is the genetic distance for the two most dissimilar cultivars.

The CP was calculated among the 14 cultivars using the pedigree information of Calhoun et al. (1997) and a modified FORTRAN program developed by D.M. Rodgers at Kansas State University. In calculating CP, it was assumed that ancestors and cultivars with unknown pedigrees were unrelated (CP = 0), cultivars obtained half their genes from each parent, all parents were homozygous and homogeneous, and the CP between a cultivar and a reselection was 0.75 [a compromise between an out cross (CP = 0.5) to an unknown and a self pollination (CP = 1)] (Murphy et al., 1986; Bowman et al., 1997). To determine the consequences of erroneously assuming that the original ancestors were unrelated, CP was recalculated for the 25 most frequently grown cultivars in 1995. In recalculating CP, the value CP = 0 (used for the relationship among ancestors) was replaced with the average s value obtained from multivariate analysis on the agronomic and fiber data.

RESULTS AND DISCUSSION

Within a location, seedcotton yield (lint + seed) and seed yield were highly correlated with lint yield ($r > 0.85$) and thus only lint yield was included in further analysis. Almost all traits in the ancestral cultivars were highly influenced by environment. The ratio of cultivar variance to location and cultivar × location variance was <1.5 for all traits. Furthermore, most traits were affected by a location × cultivar interaction ($P < 0.05$) (Table 1).

Lint yield, which ranged from 133 to 766 kg ha^{-1}, was the most variable of the traits measured. The significant ($P < 0.05$) cultivar × location interaction (Table 1) and the very low yields of some cultivars at some locations (Table 2) indicated that a great deal of selection for regional adaptability had occurred prior to the 1900s. Other studies (Brown, 1991) have also noted that cultivars tend to be superior in their region of development and differences among cultivars are most evident when they are grown outside their area of adaptation. Abou-El-Fittouh et al. (1969) attributed this regional variation in ranking among cultivars to a differential response to temperature. Deltapine 11A, which was developed from germplasm developed in Texas, Mississippi, and South Carolina, appeared to be the first widely adapted cultivar.

Variation among cultivars for lint percentage was more than 50% of the mean, whereas fiber length and strength varied by nearly 30% of the mean (Table 2). Although many fiber traits (lint percentage, micronaire, fiber strength, fiber strength, and elongation) were affected by environment, i.e., location, the small cultivar × location interaction relative to the cultivar effects indicated that, for the most part, the cultivars ranked consistently across locations. Relatively similar lint quality between cultivars developed from the pre- and post-1900 introductions reflects the fact that selection in the early 1900s emphasized the Mexican upland phenotype, characterized by earliness as a means to escape the boll weevil.

Multivariate analysis was carried out for each location-trait-cultivar value with the exception of micronaire, fiber strength, and fiber length. For these traits cultivar values were highly correlated across locations ($r > 0.8$) and thus the overall cultivar means were used in the multivariate analysis. Six principal components accounted for 87% of the variation among cultivars. The first two principal components accounted for 29 and 23% of the variation respectively. On average, fiber length and strength received highest weights on the first principal component, whereas plant height received highest weight on the second principal component. Similarity estimates (s) ranged from 0 to 0.76 (Table 3) and suggested that several cultivars that were distantly related based on pedigree information were in fact genetically very similar. Two of the post-1900 introductions from Mexico (Hopi Moencopi and Young's Acala), which in pedigree analysis were considered unrelated to all other cultivars, showed a high degree of genetic similarity to cultivars derived from the early and mid-1800 Mexican introductions. For example, Hopi Moencopi and Deltatype Webber were the two most similar cultivars ($s = 0.76$). Deltatype Webber was derived from pre-1800 introductions, while Hopi Moencopi was an introduction from Mexico in 1937. Both Kekchi and Hopi Moencopi (post-1900 introductions from Mexico) had yields and fiber properties very similar to cultivars that had undergone over 100 yr of breeding and selection in the USA. The general lack of diversity among the ancestors was illustrated by the fact that at Clayton, NC all cultivars initiated flowering over a 9-d period (data not shown).

There are several possible explanations for the preservation of the Mexican phenotype in the ancestors derived from the pre-1800 cultivars despite the occurrence of natural and controlled matings to other *Gossypium* species. Stephens (1949, 1950) reported that progeny from interspecific crosses were easily obtained but in subsequent generations progeny segregated to one of the parental types. In crosses between pre-1800 cultivars and Mexican introductions, breeders favored the Mexi-

Table 2. Agronomic and fiber traits of 14 cotton cultivars grown in North Carolina (NC), South Carolina (SC), and Mississippi (MS) during 1995.

	Plant height (cm)			Lint yield (kg ha^{-1})			Lint percent			First fruiting node			Micronaire		
Cultivar	NC	SC	MS	NC	SC	MS	NC	SC	MS	NC	SC	MS	NC	SC	MS
Deltatype Webber	111	72	90	446	336	390	36.8	31.8	22.1	5.8	5.1	5.2	4.5	3.5	3.2
Dixie Triumph	116	85	90	476	318	523	37.1	35.7	29.5	5.8	5.2	7.3	4.9	4.5	3.9
Deltapine 11A	113	105	108	644	601	620	45.5	42.8	34.8	5.7	5.0	7.1	5.3	4.6	4.3
Express 432	126	107	117	548	344	479	38.9	36.5	28.6	6.0	5.4	5.9	5.1	4.5	3.9
Half & Half	112	91	85	361	290	444	45.0	43.9	37.7	5.2	4.7	5.6	5.8	4.7	4.3
Hartsville	128	95	109	385	443	589	39.0	39.9	30.4	6.2	5.6	5.6	5.0	4.7	4.6
Hopi Moencopi	108	78	119	368	397	252	34.9	35.0	23.8	5.6	5.2	6.0	4.5	4.0	3.7
Kekchi	117	95	103	158	676	769	41.3	47.6	36.2	5.7	4.8	6.2	5.5	4.7	4.2
Lankart 57	107	93	90	438	566	461	40.2	42.3	30.3	5.7	5.4	6.2	4.7	4.7	3.4
Lone Star	136	113	134	450	589	125	39.8	41.2	29.3	7.3	5.8	7.0	4.6	3.9	3.4
Rowden	121	105	118	133	213	163	23.2	23.9	20.1	6.3	5.1	5.2	6.4	5.9	5.7
Tamcot SP 21	98	78	85	539	438	766	40.4	44.4	34.2	5.6	4.8	5.3	4.5	4.1	4.0
Wannamaker Cleveland	122	96	119	418	444	348	36.5	35.6	26.7	6.3	5.5	7.2	5.6	4.7	4.4
Youngs Acala	124	112	124	339	577	275	37.3	39.7	27.7	6.8	6.0	5.9	5.0	4.4	3.8
Range	38	41	49	511	463	644	22.3	23.7	17.6	2.1	1.3	2.1	1.9	2.4	2.5
Mean	117	95	107	407	445	443	38.3	38.6	29.5	6.0	5.2	6.1	5.1	4.5	4.0
LSD (0.05)†	13	18	11	65	244	88	2.4	2.2	1.2	1.1	0.6	0.8	0.5	0.5	0.4

	Uniformity‡			Strength (kN m kg^{-1})			Elongation (%)			Boil weight (g)			Fiber length (mm)		
Cultivar	NC	SC	MS	NC	SC	MS	NC	SC	MS	NC	SC	MS	NC	SC	MS
Deltatype Webber	81.7	46.6	83.4	25.5	26.4	30.0	6.4	9.8	6.4	5.2		4.3	27.4	29.7	31.1
Dixie Triumph	79.4	52.0	82.6	20.7	21.1	23.9	6.8	9.0	7.4	5.1		4.5	23.4	22.6	25.7
Deltapine 11A	81.7	47.4	83.6	23.6	22.2	24.9	7.5	9.8	7.3	5.1		4.7	25.7	26.3	28.3
Express 432	81.2	45.4	81.7	23.7	20.9	26.3	6.9	7.3	6.2	4.4		4.0	25.0	26.4	27.6
Half & Half	77.6	54.3	80.2	17.5	18.7	22.0	7.5	8.3	7.1	5.8		4.5	20.1	20.8	22.5
Hartsville	82.0	52.0	82.9	23.8	24.5	25.9	6.9	7.3	7.3	5.2		5.9	24.7	24.9	26.7
Hopi Moencopi	82.4	46.1	83.1	25.1	25.1	27.5	6.5	7.1	6.9	5.8		4.7	27.8	29.2	29.0
Kekchi	81.0	47.9	82.3	22.1	21.8	25.7	7.2	9.1	6.8	4.7		4.1	24.5	25.7	28.3
Lankart 57	81.2	48.5	83.4	22.3	21.2	25.7	7.6	9.5	7.0	7.9		6.9	26.0	25.9	29.1
Lone Star	82.1	49.3	83.5	23.3	20.3	25.3	7.9	9.9	7.6	6.5		4.7	26.2	27.7	29.0
Rowden	82.8	51.7	83.0	20.5	17.4	23.5	8.3	8.1	9.5	3.8		4.1	24.3	24.7	24.6
Tamcot SP 21	80.6	46.9	83.3	23.7	20.9	26.5	5.6	8.1	6.2	5.2		4.0	26.4	26.9	29.0
Wannamaker Cleveland	81.7	50.1	84.2	23.1	22.0	28.0	7.3	7.5	7.1	4.8		4.3	24.8	27.2	29.0
Youngs Acala	82.2	50.9	83.8	25.9	23.6	30.1	5.5	6.4	5.8	5.9		5.0	25.0	27.6	28.0
Range	5.2	8.9	4.0	8.4	9.0	8.1	2.8	3.5	3.7	4.1		2.9	7.7	8.9	8.6
Mean	81.2	49.2	82.9	22.5	21.8	26.1	6.9	8.4	7.0	5.4		4.7	25.1	26.1	27.7
LSD (0.05)†	1.5	1.7	1.3	1.8	1.5	2.1	0.9	1.1	0.5	0.5		1.0	1.3	1.2	1.7

† When there were missing plots LSD was calculated using the average number of plots per cultivar.
‡ Recorded as uniformity index for NC and MS and uniformity ratio in SC.

Table 3. Coefficient of parentage (upper values) and relative genetic similarity estimated from agronomic traits (lower, italic values) among 14 cotton cultivars.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Deltatype Webber	–													
Dixie Triumph	0.04	–												
	0.54													
Deltapine 11A	0.04	0.28	–											
	0.33	*0.38*												
Express 432	0.00	0.14	0.27	–										
	0.39	*0.57*	*0.40*											
Half and Half	0.07	0.10	0.07	0.00	–									
	0.33	*0.60*	*0.24*	*0.43*										
Hartsville	0.14	0.07	0.05	0.00	0.17	–								
	0.30	*0.29*	*0.30*	*0.32*	*0.46*									
Hopi Moencopi	0.00	0.00	0.00	0.00	0.00	0.00	–							
	0.76	*0.50*	*0.36*	*0.43*	*0.41*	*0.53*								
Kekchi	0.00	0.00	0.00	0.00	0.00	0.00	0.00	–						
	0.04	*0.02*	*0.46*	*0.16*	*0.09*	*0.39*	*0.20*							
Lankart 57	0.00	0.10	0.10	0.18	0.00	0.00	0.00	0.00	–					
	0.29	*0.21*	*0.23*	*0.00*	*0.27*	*0.45*	*0.40*	*0.14*						
Lone Star	0.00	0.14	0.17	0.32	0.00	0.00	0.00	0.00	0.18	–				
	0.41	*0.62*	*0.48*	*0.45*	*0.36*	*0.27*	*0.41*	*0.06*	*0.25*					
Rowden	0.00	0.19	0.23	0.42	0.00	0.00	0.00	0.00	0.24	0.42	–			
	0.38	*0.25*	*0.31*	*0.21*	*0.24*	*0.41*	*0.50*	*0.27*	*0.27*	*0.23*				
Tamcot SP21	0.02	0.03	0.02	0.02	0.13	0.04	0.00	0.19	0.01	0.05	0.02	–		
	0.54	*0.44*	*0.51*	*0.53*	*0.43*	*0.53*	*0.65*	*0.42*	*0.29*	*0.32*	*0.38*			
W. Cleveland	0.07	0.15	0.08	0.00	0.20	0.20	0.00	0.00	0.00	0.00	0.04	0.00	–	
	0.55	*0.61*	*0.59*	*0.67*	*0.47*	*0.51*	*0.64*	*0.32*	*0.28*	*0.62*	*0.64*	*0.47*		
Youngs Acala	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	–
	0.20	*0.24*	*0.16*	*0.37*	*0.38*	*0.69*	*0.42*	*0.22*	*0.21*	*0.22*	*0.43*	*0.43*	*0.19*	

can phenotypes, which were noted for their disease resistance, ease of harvest, and early maturity (Niles and Feaster, 1984). The almost complete elimination of the non-*hirsutum* phenotype in early 1900 cultivars is supported by DNA data. DNA analysis could not detect *G. barbadense* alleles in numerous modern cultivars despite frequent reports of natural outcrosses to *G. barbadense* in the early cultivars (Wendel et al., 1992). The Mexican introductions evaluated in this study had been subjected to some degree of selection when first introduced in order to conform to the USA growing season and the cotton industry standards. This may have contributed to the generally similar phenotypes in the two groups. However, the fact that fiber traits are relatively stable and only slowly changed by breeding supports the view that there was limited variability among the various introductions from Mexico.

There are several possible explanations for the lack of agreement between the two genetic distance estimates. Breeders may have been simultaneously selecting for a similar phenotype in unrelated cultivars, such that cultivars became agronomically similar despite maintaining their genetic distance. Selection for similarly acting genes may have occurred but from different sources (pedigree analysis measures gene identity by descent not identity in state). Outcrossing to other cultivars of the day could have resulted in a higher level of heterogeneity and mixing of the ancestral germplasm than is accounted for in the pedigree calculations.

Pedigree analysis assumes that a cultivar obtains an equal number of genes from each of its parents. The morphological data, however, suggests that breeders selectively favored specific traits and hence, genes. This is supported by RFLP data in corn, which showed that following a cross between two inbreds, F_2 derived lines may inherit a disproportionate share of alleles from one parent (Bernardo et al., 1996).

The high degree of similarity among the ancestral cultivars in this study suggests that pedigree analysis may be overestimating the level of diversity among cotton cultivars. For example, Van Esbroeck et al. (1998), assuming that ancestors were unrelated, determined that CP for 24 cultivars, each occupying >1% of the hectarage in the USA in 1995, was 0.16. It is probably impossible to determine the precise genetic distance among ancestors; however, CP was undoubtedly greater than zero. Our estimate for the average *s* value among the ancestors (all cultivars in the study except Lankart 57 and Tamcot SP 21) was 0.38. This value may be an overestimation, i.e., an artifact of the statistical methodology employed. We, however, recalculated CP for modern cultivars assuming ancestors were related by a CP of 0.38 to determine the impact of underestimating the genetic relationship among ancestors on estimates of genetic diversity in modern cultivars. The result was an average CP of 0.46 for the 24 cultivars in the Van Esbroeck et al., (1998) study. Despite the discrepancy between the original and revised CP estimates, the two were highly correlated ($r = 0.99$). This occurred because ancestors usually entered the pedigree only once and there were few subsequent crosses to the ancestral lines. Moreover, since the number of parents included in a pedigree increases with each generation, errors in any single parent dating back more than five generations had little impact on the relative CP among pairs of cultivars. Thus, erroneously assuming all ancestors were unrelated may have overestimated the overall diversity among cultivars but had little impact on the relative genetic distances among cultivars.

The agronomic and fiber data from this study supports isozyme and DNA data showing a narrow genetic base in cotton. Despite originating from several sources over a number of years, the ancestors of today's cultivars were remarkably similar for yield and fiber qualities. Pedigree analysis, based on the assumption of unrelated ancestors, probably overestimated the genetic diversity among cultivars but appeared accurate in estimating the relative genetic relationships among cultivars. If relationships among ancestors are unknown, CP may not accurately estimate the true level of genetic diversity within a crop species.

REFERENCES

Abou-El-Fittouh, H.A., J.O. Rawlings, and P.A. Miller. 1969. Classification of environments to control genotype by environment interactions with an application to cotton. Crop Sci. 9:135–140.

Bernardo, R., A. Murigneux, and Z. Karaman. 1996. Marker-based estimates of identity by descent and alikeness in state among maize inbreds. Theor. Appl. Genet. 93:262–267.

Bowman, D.T., O.L. May, and D.S. Calhoun. 1996. Genetic base of upland cotton cultivars released between 1970 and 1990. Crop Sci. 36:577–581.

Bowman, D.T., O.L. May, and D.S. Calhoun. 1997. Coefficients of parentage for 260 cotton cultivars released between 1970 and 1990. USDA Tech. Bull. 1852.

Brown, J.S. 1991. Principal component and cluster analysis of cotton variability across the U S. cotton belt. Crop Sci. 31:915–922.

Brubaker, C.L., and J.F. Wendel. 1994. Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*: Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). Am. J. Bot. 81:1309–1326.

Calhoun, D.S., D.T. Bowman, and O.L. May. 1997. Pedigrees of upland and pima cotton cultivars released between 1970 and 1997. Missi. Agric. and For. Exp. Sta. Bull. 1069.

Cox, T.S., Y.T. Kiang, M.B. Gorman, and D.M. Rodgers. 1985a. Relationship between coefficient of parentage and genetic similarity indices in the soybean. Crop Sci. 25:529–532.

Cox, T.S., G.L. Lookhart, D.E. Warker, L.G. Harrell, L.D. Albers, and D.M. Rodgers. 1985b. Genetic relationships among hard red winter wheat cultivars as evaluated by pedigree analysis and gliadin polyacrylamide gel electrophoretic patterns. Crop Sci. 25: 1058–1063.

Gepts, P. 1993. The use of molecular and biochemical markers in crop evolution studies. p. 51–94. *In* M.K. Hecht (ed.) Evolutionary biology. Vol. 27. Plenum Press, New York.

Gizlice, Z., T.E. Carter, Jr., and J.W. Burton. 1993. Genetic diversity in North America soybean: I. Multivariate analysis of founding stock and relation to coefficient of parentage. Crop Sci. 33:614–620.

Goodman, M.M., and R.M. Bird. 1977. The races of maize IV. Tentative grouping of 219 Latin American races. Econ. Bot. 31:204–221.

Goodman, M.M., and E. Paterniani. 1969. The races of maize: III Choices of appropriate characters for racial classification. Econ. Bot. 23:265–273.

Kim, H.S., and R.W. Ward. 1997. Genetic diversity in eastern U.S. soft winter wheat (*Triticum aestivum* L. em. Thell.) based on RFLPs and coefficients of parentage. Theor. Appl. Genet. 94:472–479.

Knauft, D.A., and D.W. Gorbet. 1989. Genetic diversity among peanut cultivars. Crop Sci. 29:1417–1422.

Loden, H.O., and T.R. Richmond. 1951. Hybrid vigor in cotton-cytogenetic aspects and practical applications. Econ. Bot. 31:387–408.

MacCluer, J.W., A.J. Boyce, B. Dybe, L.R. Weitkamp, D.W. Pfenning, and C.J. Parsons. 1983. Inbreeding and pedigree structure in standard bred horses. J. Hered. 74:394–399.

Moore, J.H. 1956. Cotton breeding in the old south. Agric. History 30:95–104.

Murphy, J.P., T.S. Cox, and D.M. Rodgers. 1986. Cluster analysis of red winter wheat cultivars based on coefficients-of-parentage. Crop Sci. 26:672–676.

Niles, G.A., and C.V. Feaster. 1984. Breeding. p. 201–231. *In* R.J. Kohel and C.F. Lewis (ed.) Cotton. Agron. Monogr. 24. ASA, CSSA, and SSSA, Madison, WI.

SAS Institute. 1991. SAS/STAT user's guide. version 6.03. SAS Inst., Cary, NC.

Souza, E., and M.E. Sorrells. 1989. Pedigree analysis of North American oat cultivars released from 1951 to 1985. Crop Sci. 29:595–601.

Stephens, S.G. 1949. The cytogenetics of speciation in Gossypium. 1. Selective elimination of the donor parent genotype in interspecific backcrosses. Genetics. 34:627–637.

Stephens, S.G. 1950. The internal mechanism of speciation in Gossypium. Bot. Rev. 16:115–149.

Tatineni, V., R.G. Cantrell, and D.D. Davis. 1996. Genetic diversity in elite cotton germplasm determined by morphological characteristics and RAPDs. Crop Sci. 36:186–192.

Van Esbroeck, G.A., D.T. Bowman, D.S. Calhoun, and O.L. May. 1998. Changes in the genetic diversity of cotton in the U.S. from 1970 to 1995. Crop Sci. 38:33–37.

Ware, J.O. 1936. Plant breeding and the cotton industry. p. 657–744. *In* USDA Yearbook of Agric. U.S. Gov. Print. Office, Washington, DC.

Ware, J.O. 1950. Origin, rise and development of American upland cotton varieties and their status at present. Mimeo Publ. (unnumbered) Univ. of Arkansas, College of Agric., Agric. Exp. Stn., Fayetteville, AR.

Wendel, J.F., C.L. Brubaker, and A.E. Percival. 1992. Genetic diversity in Gossypium hirsutum and the origin of upland cotton. Am. J. Bot. 79:1291–1310.