Theor Appl Genet (1990) 79:529–542

© Springer-Verlag 1990

Allozyme evidence for the origin and diversification of *Gossypium barbadense* L.

R. G. Percy [1] **and J. F. Wendel** [2]

[1] USDA-ARS, in cooperation with the Arizona Agricultural Experiment Station, Maricopa Agricultural Center, 37860 W. Smith-Enke Rd., Maricopa, AZ 85339, USA
[2] Department of Botany, Bessey Hall, Iowa State University, Ames, IA 50011, USA

Received August 8, 1989; Accepted November 9, 1989
Communicated by A. R. Hallauer

Summary. *Gossypium barbadense* L. is a commercially important cotton species of tropical South American origin presently grown in many regions of the world. The species is morphologically diverse, consisting of a wide range of wild (or feral), commensal, landrace, and highly improved commercial forms. We performed allozyme analysis on 153 accessions representing the spectrum of *G. barbadense* diversity to ascertain the geographic origin of the species, its patterns of diffusion subsequent to domestication, and to reveal infraspecific relationships. Levels of genetic variation in *G. barbadense* are moderate. Of 59 loci scored, 24 were polymorphic, with a mean number of alleles per locus of 1.69 and an average panmictic heterozygosity of 0.062. Principal component analysis revealed geographic clustering of accessions into six relatively discrete regions. Gene frequencies at many loci are significantly heterogeneous among these regions, with an average G_{ST} of 0.272. Northwestern South America contains the greatest genetic variability; we suggest that this region is the ancestral home of the species. The data indicate separate diffusion pathways from this region into Argentina-Paraguay and into eastern and northern South America east of the Andes. Caribbean Island and Central American forms appear to be derived from the latter. These diffusion pathways are in accordance with morphological evidence and historical record. In contrast to expectations based on geographic proximity, Pacific Island forms have their closest affinity to accessions from eastern South America. Advanced cultivated stocks seem largely derived from western Andean material, but also contain introgressed *G. hirsutum* germ plasm. Introgression was relatively high (22%–50% of accessions) in commercial stocks and in forms from Argentina-Paraguay and various Pacific Islands, but was conspicuously low or absent in material from Central America and the Caribbean, where commensal and commercial forms of both species are sympatric.

Key words: *Gossypium barbadense* L. – Cotton – Isozymes – Genetic diversity – Introgression

Introduction

Gossypium barbadense L. is one of two tetraploid species grown commercially for the production of cotton lint. It is cultivated throughout temperate and tropical regions, with primary production areas in the USSR, Egypt, India, the United States, China, and Sudan. In 1987 global production of *G. barbadense* exceeded 1.04 billion kg, with an approximate market value of 2.6 billion dollars (Anonymous 1989). The unique fiber properties (greater length, strength, and fineness) of many *G. barbadense* cultivars have allowed it to command a 50%–60% price advantage over the other major commercial species, *G. hirsutum*.

Most details regarding the origins of improved domesticated forms of *G. barbadense* are obscure. The oldest archeological remains of *G. barbadense*, from coastal Peru, have been dated to 3,800–4,500 years ago (Stephens and Mosely 1974). The relatively primitive fruit and fiber properties and location of these remains are in accordance with the general belief that original domestication occurred in northwestern South America (Harland 1936; Hutchinson et al. 1947; Mauer 1930; Stephens and Mosley 1974).

Modern commercial cultivars, with few exceptions, are day-neutral annuals with extra-long staple fiber. Each of these characteristics is exclusively the product of domestication. In its native habitats, ranging from the coasts of western South America to intermountain valleys of the Andes to riverine environments of northern South America, noncultivated *G. barbadense* is an indeterminant perennial shrub or small tree that produces a

Table 1. Accessions included in isozyme study of *G. barbadense* L. Accessions are organized by geographic region and country of collection

Geographic region	Country	Collection no.	Collection site/information
Central America	Belize[a]	B231[c]	Stann Creek, Caribb. owner
		B233	Pomona, Caribb. owner
		B235	Punta Gorda, Caribb. owner
		B238	San Antonia, Mayan village
		B250	Orange Walk
		B253	Maskall
		K23[d]*	
	Guatemala[a]	K04*	
		K06*	
		K08*	
	Honduras[a]	B204	Puerto Cortes, dooryard
		B207	Villa Nueva, dooryard
		P.I.415098[e]*	
	Nicaragua[a]	P.I.415104	Camoapa
	Costa Rica[a]	B718	Puerto Limon, Prov. de Limon, tree cotton in dooryard
Caribbean	Cuba[a]	B131	Atkins Garden, Soledad
		B139	Santiago de Cuba
		B505	San Vicenti, Prov. del Rio
	Haiti[a]	B117	Croix de Bouquets, "Coton Violette"
		B125	Cap Haitien
		K37*	
	Dominican Rep.[a]	B106	Puerto Plata
		B107	Puerto Plata
	Guadeloupe[a]	B65	2 mi. south of Gozier
		B68	between Gozier-St. Anne
		B73	St. Francois
		K44*	
	Dominica[a]	B500	ex. Caribb. Reserve
	Martinique[a]	B55	St. Pierre
		B56	St. Pierre
	St. Lucia[a]	B51	Gros Islet, U.S. naval base (abandoned)
		B54	Marquis estate, northeast of island
		K50*	
	Trinidad[a]	B31	Piarco, near airport
South America	Colombia[a]	B315	"Lengupa", improved native cultivar, Inst. Fomento Algodonera
		B424	Villavicencio, Meta
		B444	Carmen de Apicala, Tolima
		B456	El Pinol
		B457	Mercaderes, Cauca, dooryard
		B460	Mercaderes, Cauca, dooryard
		B471	Florencia, Caqueta, dooryard
		B476	Florencia, Caqueta, dooryard
		B489	5 mi. north of Confluencia, Santander del Norte
		B490	Rio de Oro, Norte Santander
		B603	Rio Arquia, Darien, ChocoDept., Cuna Indians
		B666	"Lengupa" improved native cultivar, Dept. Santander
		B667	"La Mesericorda" CAT SP-3 ex Gutierrez
		B671	Puerto Asis, Putamayo
		K60*	
		K63*	
	Ecuador[a]	WB321[f]	Playas cliffs, Guayas
		WB334	El Morro, Guayas
		WB334	between Playas-El Morro, Guayas
		WB336	Guayaquil, Guayas, north of airport
		WB338	Guayaquil, Guayas, north of airport
		B339	Las Locas, north of Guayaquil, Guayas, dooryard
		B346	1 km north of Duran, Guayas, dooryard
		K65*	
		WB358	Aguas Verdes, "Tumbes wild"

Table 1. (continued)

Geographic region	Country	Collection no.	Collection site/information
	Peru[a]	B350	12 km south of Chilete, Cajamarca, dooryard
		B351	25 km north of Chilete, Cajamarca, dooryard
		B353	Chivato, near Tumbes, Tumbes
		B354	Chivato, near Tumbes, Tumbes
		B363	17 km south Nazareth, Maranon Valley
		B366	between bagua Grande-Chachapoyas, Utcubamba Valley
		B373	Saña Valley, dooryard
		B374	Saña Valley, dooryard
		B375	Ilo Valley, near Chamas
		B377	"Mollendo" cultivated type
		B453	Hacienda Uchumayo, near Quillabamba, Cuzco
		B454	Hacienda Potrero, 35 km south of Quillabamba, Cuzco
		WB620	Zarumilla, Sector Canario
		WB621	Zarumilla, Sector Canario
		B646	Pampa Grande, Tumbes River, dooryard
		B699	Padre Island, Iquitos
		CB4073[g]*	Tumbes
		K83*	
		K93*	
		K97*	
	Boliva*	B402	Santa Cruz del Valle Ameno, La Paz
		B404	Carapari del Rio Grande, Chuquisaca
		B405	Tomina, Chuquisaca
		B406	Santa Catalina, La Paz
		B407	Aten, La Paz
		K101*	
		K103*	
	Venezuela[a]	B516	Orinoco
		B559	8 km from La Grita, Tachira Prov.
		B572	Irapa, Sucre Prov.
		B573	Mendoza, Merida Prov., Lake Maracaibo
	Surinam[a]	B380	Nieuw Amsterdam, Suriname
		B384	Near Jodensavanna, Indian village, dooryard
		B385	Calcutta? (Cottica?) dooryard
	Brazil[a]	B409	Bandeira, border of Amapa & Para Prov., Jari River
		B410	Bandeira, border of Amapa & Para Prov., Jari River
		B447	Near Belem, Para Prov.
		B475	Belem, Para Prov., dooryard
		B607	Sete Lagos, Minas Gerais Prov.
		B652	Southwest of Belem, Para Prov.
		B655	Near Recife, Pernambuco Prov., Guararape village
		B691	Near Porto Grander, Amapa
		K195*	
		K198*	
	Paraguay[a]	K187*	
		K189*	
	Argentina[b]	K133*	
		K136*	
		K144*	
		K162*	
		K172*	
		K173*	
Atlantic Ocean	Bermuda[b]	B170	Paget East
		B171	St. George's West
Pacific Basin	Hawaii[b]	B903	Aiea, "Caravonida" ex. J. B. Smith
		B912	ex. Phillips, 1961
	Solomon Is.[b]	B906	Honiadra
		B907	Honiara
	Marquesas[b]	B918	Anaho, northeast of Nuku-Hiva
		B919	Haamene (Tahaa: Iles-sous-le vent)
		B937	Hatiheu, Nuku-Hiva, dooryard

Table 1. (continued)

Geographic region	Country	Collection no.	Collection site/information
	Tuamotus[b]	B922	Vahitahi Is.
	Tonga[b]	B939	Dooryard
	Galapagos	B1213	San Christobal Is., in garden, reputedly from Guayaquil, Ecuador
		B1214	San Christobal Is., stored seed, reputedly same plant as 1213
Africa	Angoila[b]	B1002	Angola "Wild" Portuguese
India	India[b]	B1009	Margao, Goa
		B1011	Ponda, Goa
Improved Cultivars	United States[b]	K238*	"Old Pima"
		K241*	"Pima 32"
		K242*	"Pima S-1"
		K243*	"Pima S-2"
		—	"Pima S-6"
		—	"Pima S-3"
	Egypt[b]	K248*	"Ashmouni"
		K250*	"Bahtim 185"
		K254*	"Giza 7"
		—	"Menoufi"
		—	"Karnak"
		CB3032*	"Domains Sakel"
	Caribbean	K245*	"Nevis Sea Island"
		K246*	"Monserratt Sea Island"
		K256*	"Sea Island 12B2"
		K257*	"S.I. Seaberry"
		K267*	"St. Vincent Superfine"
		—	Sea Island
		—	Sea Island
	Russia[b]	K260*	"Barbadense Tashkent"
		P.I.441014*	"Termez"
		P.I.441018	"Syrkhandar'in"
	—	—*	"Aires"
	—	—*	"Hallmark"
	Tonga[b]	—*	"Tonga"
	Peru[a]	CB287*	"Tanguis 5-2/A"

[a] Country partially or entirely within *Gossypium barbadense* native range
[b] Country outside of native range of *G. barbadense*
[c] Accessions with numbers preceded by the letter B are from a collection assembled and maintained by S.G. Stephens at North Carolina State University, Raleigh/NC, and presently maintained by the USDA-ARS at Maricopa/AZ
[d] Accessions with numbers preceded by the letter K are from a working collection assembled and maintained by the USDA-ARS at Maricopa/AZ
[e] Accessions with numbers preceded by the letters P.I. have been deposited in the National Germ Plasm Collection at Fort Collins/CO. Accessions bearing P.I. numbers were obtained from the collection
[f] Accessions with numbers preceded by the letters WB are part of the "B" collection of S. G. Stephens. The "W" denotes accessions thought by Stephens to be wild on the basis of morphology or locality of collection
[g] Accessions with numbers preceded by the letters CB were part of a collection maintained at the "Cotton Branch" of the USDA-ARS National Headquarters in Beltsville/MD. This collection is now maintained in the National Germ Plasm Collection at Fort Collins/CO, and in a working collection at Maricopa/AZ
* Unknown or unspecified collection site

short, coarse fiber. These wild or primitively domesticated forms display a more extensive range of variation in floral and vegetative traits than do modern commercial cultivars. Particularly distinctive variants have been observed in the Galapagos Islands and northern Brazil (Watt 1907; Hutchinson et al. 1947); these have been given formal varietal status as *G. barbadense* var. *darwinii* (Watt) J. B. Hutch. and *G. barbadense* var. *braziliense* (Rafin.) Fryx. (Hutchinson et al. 1947; Fryxell 1979).

The practice of categorizing accessions within germ plasm collections on the basis of relative degree of "improvement" is a tacit recognition of human influence on variation patterns in *G. barbadense*. Within germ plasm collections, four general categories of accessions are found. Few truly wild specimens of *G. barbadense* are thought to exist, and wild accessions constitute the smallest category of collections. A second category, "dooryard cottons," is probably the best represented cat-

egory in collections; these cottons differ from wild cottons in having a history of casual use by local peoples throughout South and Central America and the Caribbean. Dooryard cottons perhaps should best be thought of as commensals. These cottons are not grown in large populations as is the case with landraces, but occur as single plants found near habitations. Dooryard cottons have received little population selection and are thought to derive in many cases directly from local, wild progenitors. In many regions, improved *G. barbadense* has become established as part of the natural vegetation; these feral cottons are often difficult to distinguish from wild forms and dooryards, but their past cultivation may be evidenced in the degree of manipulation for agronomic properties and displacement they display. Landraces, a third category, are derived from locally available cottons, but are cultivars grown on a commercial scale. *G. barbadense* landraces are often short-day flowering perennials, as are the wild and dooryard forms. Finally, there are improved cultivars, which display the greatest evidence of manipulation for agronomic properties and displacement from their geographic regions of origin. Although cultivation has imposed a degree of morphological and physiological uniformity on commercial *G. barbadense*, a fair amount of cultivar variation remains, e.g., in environmental adaptations and several growth and fiber characteristics.

The purpose of the present investigation was to assess, by means of allozyme analysis, infraspecific relationships and the level and patterns of genetic variability within *G. barbadense*. We were particularly interested in clarifying the origins of cultivated forms, defining their diffusion pathways subsequent to domestication, and quantifying the proportion of the *G. barbadense* gene pool that has been "captured" in widely utilized cultivars. Because both anecdotal information and historical record (Kerr 1960; Feaster and Turcotte 1962; Stephens 1975) document introgression of *G. hirsutum* germ plasm into cultivated *G. barbadense*, an additional objective was to use molecular markers to describe the extent and magnitude of this phenomenon.

Materials and methods

Plant materials

A geographic sampling of 153 *Gossypium barbadense* L. accessions from the USDA-ARS working germ plasm collection at Maricopa/AZ was selected for analysis (Table 1). Accessions included in the sample varied in their origins and histories. Those carrying the designation "B" originated in a collection assembled and formerly maintained at North Carolina State University by S. G. Stephens (this collection was acquired by the USDA-ARS in 1973). Accessions given "CB", "K", or "P.I." designations were maintained at Beltsville/MD by the USDA-ARS Cotton Branch prior to incorporation into the present collection. Among the contributors to the above collections were the Stoneville/MS Obsolete Cultivar Collection maintained by the state of Mississippi and an Argentine collection maintained by A. Gutierrez. Accessions of the USDA-ARS germ plasm working collection have been maintained by periodic seed renewal in greenhouses in Phoenix or Maricopa/AZ, or by renewal in winter nurseries at Iguala or Tecoman, Mexico. To insure genetic purity of accessions, seed renewal under field conditions has been accompanied by culling of off-type plants and self-pollination.

In selecting accessions, care was taken to obtain a representative, non-biased sampling. We omitted accessions displaying "improved" characteristics in favor of wild or commensal accessions in order to maximize the possibility of discerning geographic patterns of variability and regional relationships. A separate representative subset of improved cultivars was included in the analysis. Preference was given to Stephens' "B" collection due to its detailed locality information. In addition, most of these consisted of either original field-collected seed or seed from a low number of renewal cycles, thus minimizing bias due to possible drift from original genotypes. Relatively little information (e.g., only country of origin) was available for many of the remaining accessions (Table 1).

The Galapagos Islands' endemic *G. barbadense* var *darwinii* is not included in the present study, as we consider it a distinct species, *G. darwinii* Watt (following Fryxell 1979). Inter-island and intra-island variation patterns and evolution of *G. darwinii* will be presented elsewhere.

Isozyme analysis

Starch gel electrophoresis was performed on crude protein extracts of leaf tissue and from seeds that had been imbibed for 24 h. Preliminary surveys indicated little or no variation within accessions; consequently, few individuals (an average of four) were analyzed per accession. Sample preparation consisted of placing approximately 40 mg of tissue in a 0.5-ml microfuge tube and homogenizing it with a power-driven acetal pestle (on ice) in 75 µl of 75 m*M* Na-phosphate, pH 7.5, containing 0.5% BSA, 5% (w/v) sucrose, 10% (w/v) polyvinylpolypyrrolidone, 14 m*M* mercaptoethanol, 100 m*M* ascorbic acid, 10 m*M* dithioerythritol, and 10 m*M* diethyldithiocarbamate. Extracts were frozen at −70 °C until electrophoresis.

Enzymes were separated in 12% (w/v) starch gels. Seventeen enzymes could be satisfactorily and consistently resolved using five different electrophoretic buffer systems. (1) Electrode buffer – 76 m*M* TRIS–5 m*M* citrate (pH 8.6); gel buffer – 0.3 *M* boric acid titrated to pH 8.0 with NaOH; used for aspartate aminotransferase (AAT) and phosphoglucose isomerase (PGI). (2) Electrode buffer – 0.19 *M* boric acid adjusted to pH 8.3 with lithium hydroxide (final LiOH molarity is approximately 0.038); gel buffer – 1 part electrode buffer to 9 parts 52 m*M* TRIS-8 m*M* citrate, pH 8.3, used for endopeptidase (ENP), catalase (CAT), and triose-phosphate isomerase (TPI). (3) Electrode buffer – 40 m*M* citric acid titrated to pH 6.1 with N-(3-aminopropyl)-morpholine; gel buffer – 1 : 19 dilution of the electrode buffer; used for 6-phosphogluconate dehydrogenase (PGD) and arginyl-specific aminopeptidase (ARG). (4) Electrode buffer – 65 m*M* L-histidine–19 m*M* citrate, pH 6.5; gel buffer – 1 : 6 dilution of the electrode buffer; used for aconitate hydratase (ACO), alcohol dehydrogenase (ADH), NADP-isocitrate dehydrogenase (IDH), NADH-dehydrogenase (="menadione reductase", NAD), malate dehydrogenase (MDH), and phosphoglucomutase (PGM). (5) Electrode buffer – 180 m*M* TRIS–100 m*M* boric acid–4 m*M* EDTA, pH 8.6; gel buffer – 1 : 3 dilution of the electrode buffer; used for glutamate synthetase (GS), glutamate dehydrogenase (GDH), and formate dehydrogenase (FDH). Leucyl-aminopeptidase (LEU) was as-

Table 2. Allelic frequencies, by region, of 24 polymorphic loci in *Gossypium barbadense*

Locus	Alleles	West of Andes	Caribbean	East of Andes	Argentina Paraguay	Central America	Pacific Islands	Improved cultivars	Across regions[a]	Pure *G. barbadense*[b]
(*N*)		36	21	23	6	15	10	23	134	111
Mdh4	4	0.64	0.05	0.24	0.08	0.20	0.30	0.37	0.270	0.320
	6	0.36	0.95	0.76	0.92	0.80	0.70	0.63	0.730	0.680
Gdh1	2	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.010	0.000
	4	0.06	0.48	0.61	0.00	0.93	0.70	0.02	0.400	0.420
	6	0.94	0.52	0.39	1.00	0.07	0.30	0.91	0.590	0.580
Idh1	1	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.010	0.030
	2	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.010	0.000
	4	0.90	1.00	1.00	1.00	1.00	1.00	0.96	0.980	0.970
Idh2	4	0.40	0.91	0.35	0.00	0.87	0.00	0.30	0.400	0.490
	5	0.60	0.09	0.65	1.00	0.13	1.00	0.70	0.600	0.510
Enp1	4	0.17	0.00	0.17	0.00	0.07	0.20	0.24	0.120	0.120
	5	0.83	1.00	0.83	1.00	0.93	0.80	0.76	0.880	0.880
Enp2	4	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.998	0.995
	9	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.002	0.005
Tpi3	4	1.00	1.00	1.00	0.83	1.00	1.00	1.00	0.980	0.990
	6	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.020	0.010
Tpi6	4	1.00	1.00	0.96	1.00	1.00	1.00	0.96	0.990	0.990
	9	0.00	0.00	0.04	0.00	0.00	0.00	0.04	0.010	0.010
Tpi7	4	0.00	0.00	0.04	0.00	0.00	0.00	0.06	0.020	0.009
	5	1.00	1.00	0.96	1.00	1.00	1.00	0.94	0.980	0.991
Arg2	1	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.006	0.000
	2	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.003	0.000
	3	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.004	0.009
	4	0.78	1.00	0.83	1.00	1.00	1.00	0.94	0.934	0.892
	6	0.17	0.00	0.17	0.00	0.00	0.00	0.00	0.048	0.090
	9	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.004	0.009
Aat2	2	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.016	0.036
	4	0.89	1.00	1.00	1.00	1.00	1.00	0.96	0.978	0.964
	6	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.006	0.000
Pgd1	1	0.00	0.00	0.00	0.00	0.00	0.30	0.00	0.043	0.027
	3	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.012	0.000
	4	1.00	1.00	1.00	1.00	1.00	0.70	0.91	0.945	0.973
Nad1	4	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.009	0.000
	9	1.00	1.00	1.00	1.00	1.00	1.00	0.93	0.991	1.000
Aco1	4	0.79	1.00	1.00	1.00	1.00	0.90	0.96	0.950	0.923
	6	0.10	0.00	0.00	0.00	0.00	0.10	0.00	0.028	0.041
	8	0.11	0.00	0.00	0.00	0.00	0.00	0.04	0.022	0.036
Aco3	1	0.00	0.05	0.00	0.00	0.00	0.00	0.04	0.013	0.009
	4	1.00	0.90	0.87	0.50	1.00	0.90	0.96	0.876	0.919
	7	0.00	0.00	0.00	0.42	0.00	0.00	0.00	0.059	0.023
	8	0.00	0.05	0.13	0.08	0.00	0.10	0.00	0.052	0.050
Aco5	1	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.012	0.027
	4	0.84	1.00	1.00	1.00	1.00	1.00	1.00	0.976	0.946
	9	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.012	0.027
Aco6	2	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.004	0.009
	4	0.86	0.76	1.00	1.00	1.00	1.00	0.96	0.940	0.910
	9	0.11	0.24	0.00	0.00	0.00	0.00	0.04	0.056	0.081
Leu1	2	0.08	0.00	0.04	0.00	0.00	0.00	0.00	0.018	0.036
	4	0.89	1.00	0.96	1.00	1.00	1.00	0.96	0.972	0.955
	5	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.006	0.000
	9	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.004	0.009
Pgm1	4	0.60	0.91	0.91	0.50	1.00	1.00	0.96	0.839	0.806
	9	0.40	0.09	0.09	0.50	0.00	0.00	0.04	0.161	0.194
Pgm3	2	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.006	0.000
	4	1.00	1.00	1.00	1.00	1.00	1.00	0.94	0.994	1.000

Table 2. (continued)

Locus	Alleles	West of Andes	Caribbean	East of Andes	Argentina Paraguay	Central America	Pacific Islands	Improved cultivars	Across regions[a]	Pure *G. barbadense*[b]
Pgm4	4	0.89	0.69	0.87	1.00	1.00	0.70	0.96	0.872	0.851
	9	0.11	0.31	0.13	0.00	0.00	0.30	0.04	0.128	0.149
Pgm5	2	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.004	0.009
	4	0.90	1.00	1.00	1.00	1.00	1.00	0.41	0.902	0.968
	6	0.07	0.00	0.00	0.00	0.00	0.00	0.59	0.094	0.023
Pgm6	2	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.006	0.000
	4	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.994	1.000
Pgm7	2	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.012	0.000
	4	0.99	0.55	0.63	1.00	0.20	0.70	0.91	0.711	0.698
	9	0.01	0.45	0.37	0.00	0.80	0.30	0.00	0.277	0.302

[a] Accessions of unknown locality removed from data set
[b] Accessions that had introgressant alleles and all improved cultivar accessions removed from data set

sayed both on systems (1) and (3), above. Enzymes were visualized using staining methods detailed in Wendel and Weeden (1989).

Genetic interpretations of isozyme and allozyme phenotypes were based on observed patterns of variation, typical patterns of subcellular localization and gene expression in other plants, and knowledge of the quarternary structure of the protein products (reviewed in Weeden and Wendel 1989). Support for these interpretations comes from formal genetic analysis involving numerous interspecific and intraspecific F_2 and BC progenies among *G. barbadense*, *G. hirsutum*, and *G. tomentosum* (J. F. Wendel, in preparation). Loci encoding the most anodally migrating isozyme for each enzyme system were assigned the numerical designation 1, with additional loci numbered sequentially in order of decreasing electrophoretic mobility. Similarly, allozymes at each locus were given numerical designations in order of decreasing electrophoretic mobility.

Data analysis

A variety of statistical techniques was employed to summarize the gene frequency data obtained for each accession. Multivariate relationships among accessions were revealed with principal component analysis, using a covariance matrix derived from allele frequencies (Sneath and Sokal 1973). Recognition of accession groups based on these results allowed the computation of "regional" gene frequencies. These were used in cluster analysis (Sneath and Sokal 1973) and in apportioning genetic variation among regions (Nei 1987). This latter technique partitions total variation (H_T) into within- and among-region components (H_S and D_{ST}, respectively); G_{ST} ($=D_{ST}/H_T$) is a measure of the proportion of total variation accounted for by regional differentiation. Genetic distance and identity statistics (D and I) were calculated following Nei (1987). Many of the above computations were expedited by the computer programs BIOSYS (D. Swofford, Illinois Natural History Survey) and NTSYS (Exeter Publishing Ltd, Setauket/NY).

Results

Genetic variability

Of the 17 enzyme systems examined, 4 were invariant and were postulated to be under the genetic control of eight loci: CAT(2), FDH(1), GS(1), and PGI(4). Polymorphisms were observed in at least one zone of activity for the remaining 13 enzyme systems. These 13 enzymes are encoded by a minimum of 51 genetic loci: AAT(4), ACO(7), ADH(2), ARG(2), ENP(2), GDH(1), IDH(2), LEU(2), MDH(6), NAD(2), PGD(6), PGM(7), and TPI(8). These gene number estimates are minima, in that poorly resolved isozymes are not included, and because additional unobserved isozymes were likely present as comigrating duplicated products. Including the monomorphic enzymes, we estimate that 59 genetic loci were examined in this study.

Of the 59 loci scored, 35 were fixed for the same allele in all accessions examined. At least one locus was variable for 13 enzyme systems, resulting in a total of 24 polymorphic loci ($P=40.7\%$ for the species). A total of 65 alleles was detected at the polymorphic loci. Twelve of the polymorphic loci were minimally variable, in that only two alleles were detected per locus. An additional 9 loci were tri-allelic, while the remaining 3 loci (*Arg2, Aco3, Leu1*) were multi-allelic (6, 4, and 4 alleles, respectively). Including the 35 monomorphic loci, the mean number of alleles per locus is calculated to be 1.69 (2.71 per polymorphic locus).

In accordance with expectations, a single multi-locus genotype was observed within most accessions, although low levels of within-accession variation were sometimes observed. In the majority of cases where variation was observed within an accession, it was limited to only one or two loci. Similarly, observed heterozygosity was nearly nonexistent; when an accession was polymorphic, it usually consisted of a mixture of alternate homozygotes.

Gene frequencies for the entire sample (Table 2) demonstrate that the majority of polymorphic loci are only weakly polymorphic; at 15 loci the frequency of the most common allele is greater than 0.90. Consequently, most loci have relatively low estimates of panmictic heterozygosity ($=1-\Sigma(p_i)^2$, where p_i are allele frequencies).

Table 3. Genetic differentiation[a] among seven regions of *Gossypium barbadense* L. for 24 polymorphic loci

Locus	H_S	H_T	G_{ST}	χ^2 (*df*)	*P*
Mdh4	0.33	0.39	0.17	53.8 (6)	0.00
Gdh1	0.26	0.49	0.48	139.4 (12)	0.00
Idh1	0.04	0.04	0.07	29.2 (12)	0.01
Idh2	0.25	0.48	0.48	88.5 (6)	0.00
Enp1	0.20	0.21	0.08	15.5 (6)	0.02
Enp2	0.01	0.01	0.00	2.7 (6)	0.84
Tpi3	0.04	0.05	0.15	42.9 (6)	0.00
Tpi6	0.02	0.02	0.03	7.8 (6)	0.26
Tpi7	0.03	0.03	0.04	10.3 (6)	0.11
Arg2	0.11	0.13	0.11	53.5 (30)	0.01
Aat2	0.04	0.04	0.08	32.1 (12)	0.01
Pgd1	0.09	0.11	0.22	95.6 (12)	0.00
Nad1	0.02	0.02	0.06	14.6 (6)	0.02
Aco1	0.09	0.10	0.09	35.6 (12)	0.00
Aco3	0.18	0.23	0.22	135.5 (18)	0.00
Aco5	0.04	0.05	0.11	34.2 (12)	0.00
Aco6	0.10	0.11	0.13	32.7 (12)	0.01
Leu1	0.05	0.06	0.05	27.3 (18)	0.07
Pgm1	0.20	0.27	0.26	56.8 (6)	0.00
Pgm3	0.01	0.01	0.04	9.7 (6)	0.14
Pgm4	0.19	0.22	0.13	26.5 (6)	0.00
Pgm5	0.10	0.18	0.46	123.1 (12)	0.00
Pgm6	0.01	0.01	0.04	9.8 (6)	0.14
Pgm7	0.27	0.42	0.36	119.6 (12)	0.00
Mean	0.11	0.15	0.28	–	–
Mean with 35 invariant loci	0.05	0.15	0.28	–	–
Mean of non-introgressant loci[b]	0.13	0.18	0.28	–	–

[a] H_S, H_T, and G_{ST} are gene diversity statistics of Nei (1987). The final two columns present Chi-square values from tests of gene frequency homogeneity across regions (degrees of freedom), and their probability values, under the null hypothesis of no genetic differentiation

[b] Non-introgressant loci were polymorphic loci that had no alleles in common with *G. hirsutum*. These ten loci were *Mdh4, Idh2, Enp2, Tpi6, Aco1, Aco3, Aco5, Aco6, Pgm4,* and *Pgm5*

The most equitable gene frequencies, and consequently the highest expected heterozygosities, were obtained for *Mdh4, Gdh1, Idh2,* and *Pgm7* (H_T in Table 3). Averaged across polymorphic loci, the mean panmictic heterozygosity in *G. barbadense* is 0.15; including the 35 monomorphic loci, it is 0.06.

Infraspecific relationships and regional variability

Inspection of the geographic distribution of alleles suggested that many had potential systematic significance in that their distribution was nonrandom. In order to reduce the dimensionality of the data set, principal component analysis was performed on the covariance matrix of allele frequencies. Accessions were plotted according to their coordinates along the first two principal components (which accounted for 36% of the total variance). The resulting plot led to the recognition of more or less discrete clusters of accessions that correspond to particular geographic regions (Fig. 1). The first principal component (PCA1) separated mainland South American accessions along an east-west axis, apparently into regions east and west of the Andes mountains. Clustering with accessions from west of the Andes mountains were accessions from trans-Andean sites (those from inter-montane valleys), accessions from Argentina and Paraguay, and modern improved cultivars. Mainland accessions collected east of the Andes mountains were grouped with accessions of Caribbean, Central American, and Pacific Island origin. PCA2 further separated the Pacific Island and Caribbean accessions into two discrete clusters within the larger east-of-Andes cluster. A less distinct regional cluster comprising the Central American accessions could be discerned nested within the Caribbean Island cluster. Thus, principal component analysis led to six geographically interpretable clusters and a seventh comprising modern improved cultivars.

Mean allele frequencies for the 24 polymorphic loci are presented in Table 2 for each geographic region. Casual inspection of this table reveals that many loci differ among regions. For example, *Gdh1-6* is nearly fixed in material from west of the Andes, Argentina, Paraguay, and in improved cultivars, while in other regions its frequency does not exceed 0.52. Not surprisingly, this locus, as well as *Mdh4* and *Pgm7*, had the highest factor loadings into PCA1. The highest factor loadings for PCA2 were for *Idh2* and *Pgm5*. A significance test ($p < 0.05$) for genetic differentiation among regions (Chi-square tests of homogeneity of gene frequencies) shows that gene frequencies for 19 of the 24 polymorphic loci are heterogeneous (Table 3).

Effects of regional divergence in gene frequencies on apportionment of genetic diversity were quantified by the gene diversity statistics of Nei (1987). The proportion of the genetic variation resulting from differences among regions (G_{ST}) ranged from 0.0 for *Enp2* (equivalent gene frequencies in all regions) to 0.48 for *Gdh1*. Averaged over the 24 polymorphic loci, the proportion of total variation arising as a consequence of regional differentiation was 0.28. This value did not change when calculated using only the 10 non-introgressant loci.

Overall similarity between regions was summarized by the genetic identity coefficient (*I*) of Nei (1987). These ranged from 0.994 for the east-of-Andes–Pacific Island pair to 0.952 for the Central America–Argentina-Paraguay pair (Table 4). As expected from the PCA results, high genetic identities were found among Pacific Island, Caribbean, Central American, and east-of-Andes regional pairings. Principal component analysis indicated that the first three of the preceding regions clustered

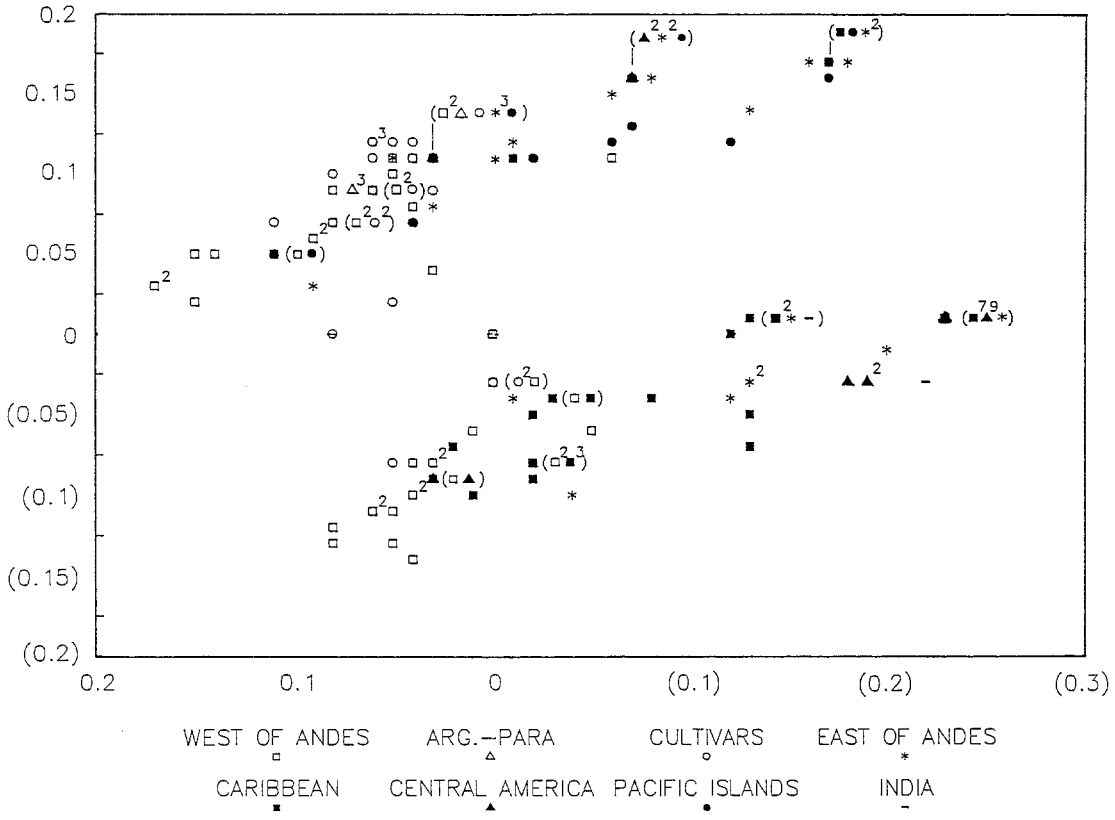

Fig. 1. Principal component analysis of 153 *Gossypium barbadense* accessions based on the covariance matrix of gene frequencies. The first two axes explain 36% of the total variance. Superscripts occurring above regional symbols denote the number of accessions that plotted to a common point. In cases where numerous accessions of differing regional affiliation plotted to a common point, their symbols appear in parentheses above or to the right of the plot point

Table 4. Nei's genetic identities among geographic regions of *Gossypium barbadense*

Regions being paired		$\bar{I}$
East of Andes	Pacific Islands	0.994
Caribbean	Central America	0.991
East of Andes	Caribbean	0.991
East of Andes	Central America	0.989
West of Andes	Cultivars	0.989
West of Andes	East of Andes	0.985
West of Andes	Argentina-Paraguay	0.984
East of Andes	Cultivars	0.984
Argentina-Paraguay	Cultivars	0.981
East of Andes	Argentina-Paraguay	0.981
Pacific Islands	Cultivars	0.980
Pacific Islands	Caribbean	0.979
Pacific Islands	West of Andes	0.978
Pacific Islands	Central America	0.978
Pacific Islands	Argentina-Paraguay	0.977
Caribbean	West of Andes	0.977
Caribbean	Cultivars	0.976
Caribbean	Argentina-Paraguay	0.969
Central America	Cultivars	0.963
Central America	West of Andes	0.963
Central America	Argentina-Paraguay	0.952

primarily within the east-of-Andes cluster. In contrast, the lowest estimates of I were obtained between improved cultivars, west-of-Andes region, or Argentina-Paraguay and Central America. Pairings of improved cultivars, west-of-Andes region, and Argentina-Paraguay with the Caribbean region also produced low identity coefficients. A west-of-Andes east-of-Andes regional pairing produced an intermediate genetic identity value.

A phenogram produced from an average linkage cluster analysis using Rogers' genetic distance is presented in Fig. 2 (cophenetic correlation = 0.849). The resulting topology is fully consistent with regional relationships suggested by the PCA and genetic identity computations.

Each region differed not only in its allelic array and gene frequencies, but also in amount of genetic diversity. Total number of alleles varied from 82 in the west-of-Andes region and improved cultivars to 64 in the Argentina-Paraguay and Central American regions (Table 5). Accordingly, the mean number of alleles per locus (including monomorphic loci) ranged from 1.39 west of the Andes and in improved cultivars to 1.08 in Argentina-Paraguay and Central America. The west-of-

Andes region and improved cultivars also had the greatest number of unique alleles (nine and ten, respectively), whereas the Caribbean, east-of-Andes, and Central American regions possessed none (Table 5). West of the Andes was the only region with six loci having more than two allelic variants. Variation within regions, as measured by the percentage of polymorphic loci, ranged from 35.6% in improved cultivars to 6.8% in the Argentina-Paraguay region. The percentage of unique multi-locus genotypes per region ranged from 72.2% in the west-of-Andes region to 6.6% in Central America. Mean panmictic heterozygosity followed a similar trend, ranging from 0.067 in the west-of-Andes region to 0.019 in Central America (Table 5). The west-of-Andes region, by all the preceding means of measurement, appeared to be the most genetically diverse, whereas Central America and Argentina-Paraguay contained the lowest levels of genetic variation.

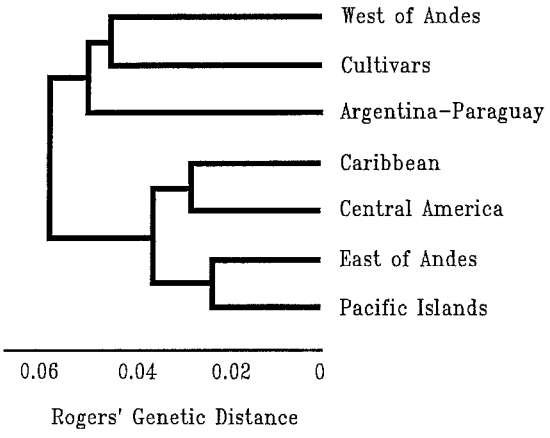

Fig. 2. Phenogram of *Gossypium barbadense* regional relationships produced by average linkage cluster analysis using Roger's genetic distance. Cophenetic correlation = 0.849

Gossypium hirsutum introgression

Ongoing isozyme studies of over 500 accessions of *Gossypium hirsutum* have identified 14 allozyme loci that distinguish *G. hirsutum* from *G. barbadense* (J. F. Wendel, unpublished data). At each of these loci the two species are fixed or nearly fixed for alternate alleles; thus, these loci serve as species-specific markers for the detection of interspecific introgression. In fact, the large number of such loci provides a sensitive tool for the detection of introgressed germ plasm. In the present study, *G. hirsutum* alleles were detected at all 14 loci. The number of introgressed alleles varied in individual accessions from 0 (most accessions) to 7 in accession K189, a Paraguayan accession. Moreover, the occurrence of introgressant alleles was not random among geographic regions, but was concentrated in accessions from Argentina and Paraguay (50% of accessions), the Pacific Island region (33.3% of accessions), and in improved cultivars (21.7% of accessions). A smaller frequency of accessions with introgressant alleles was observed in accessions from South America west of the Andes mountains (11.1%), the Caribbean (4.3%), and east of the Andes (4.3%). No accessions from Central America displayed introgressant *G. hirsutum* alleles at any locus.

Discussion

Genetic variation in Gossypium barbadense

Levels of isozyme variation have been published for most major domesticated plants and their wild relatives (reviewed in Doebley 1989), but little comparable information has been forthcoming for cultivated cottons (Bourdon 1986; Wendel et al. 1989). Although isozyme studies of other crops vary widely in many important respects, including number of loci assayed and thoroughness of germ plasm sampling, these data provide a com-

Table 5. Genetic diversity by region for 59 loci (including 35 invariant loci) of *Gossypium barbadense* L.

Region	N	A[a]	$\%P$[b]	H_S[c]	Total no. of alleles	No unique alleles[d]	% unique genotypes[e]
West of Andes	36	1.39	27.1	0.067	82	9	72.2
Cultivars	23	1.39	35.6	0.058	82	10	52.2
Argentina-Paraguay	6	1.08	6.8	0.026	64	2	16.7
East of Andes	23	1.20	20.3	0.053	71	0	65.2
Caribbean	21	1.15	13.6	0.040	68	0	42.9
Central America	15	1.08	8.5	0.019	64	0	6.6
Pacific Islands	10	1.13	13.6	0.047	67	1	60.0

[a] A – Mean number of alleles per locus
[b] $\%P$ – Percent polymorphic loci
[c] H_S – Mean subpopulation diversity = panmictic heterozygosity
[d] Number of alleles, found at any of the 59 loci, which were unique to one region
[e] Percentage of genotypes found in a region, which occurred only once in the *G. barbadense* sample

parative context for the amount and distribution of genetic diversity among taxa. Forty-one percent of surveyed loci in *Gossypium barbadense* are polymorphic, a value that is comparable to the mean for domesticated crop plants (49%; Doebley 1989). Comparisons of other variability measures, such as mean number of alleles per locus (1.69 versus a mean of 2.15 for other crops) and expected heterozygosity (0.06 versus a mean of 0.19), support the conclusion that *G. barbadense* is neither genetically depauperate nor exceptionally diverse; rather, it contains moderate levels of allozyme variation.

Origin and diffusion pathways of Gossypium barbadense

Measures of variability within wild and commensal *G. barbadense* indicate a center of variation in northwestern South America west of the Andes. This region exceeded other regions in all measures of genetic variability, including the percentage of unique genotypes found among its accessions, allelic richness, percent polymorphic loci, and panmictic heterozygosity (Table 5). Moreover, it is the only part of the native range of the species containing unique alleles. We regard these data as evidence for a probable center of origin in northwestern South America. Isozyme evidence for a west Andean center of variation and origin is consistent with the archeological record (Stephens and Mosely 1974) and previous observations by Mauer (1930) and Harland (1936), who placed the center of variability of the species west of the central cordillera of Colombia. Hutchinson et al. (1947) were more inclusive in defining the center of variability as "the Andean valleys from Bolivia to Colombia" and the center of origin as the "mountain valleys of northwestern South America." Accessions contributing to isozyme variation in the west-of-Andes region include materials of trans-Andean and Bolivian locality, and thus conform to Hutchinson et al.'s broader interpretation of the center of variation and origin.

Reconstruction of historical events leading to the diffusion of *Gossypium barbadense* from the western Andean center of origin to other geographic regions is an uncertain undertaking, but several insights are suggested from the allozyme data. These are largely based on genetic identity coefficients, principal component analysis, and various measures of regional variability, supplemented in most cases by morphological evidence and/or historical record.

One dispersal hypothesis is suggested by the clear affinity between accessions from east of the Andes, the Caribbean, and Central America (pairwise range of $I=0.989-0.991$). This close relationship has long been suspected from the presence of a set of distinctive morphological traits (connate or "kidney seed", absence of fringe trichomes above floral nectaries, distinctive single-limbed trichomes, glabrous stems, short internodes, and long, slender, tapered bolls) that are common within these regions (Watt 1907; Hutchinson et al. 1947; T. Lee unpublished data; E. L. Turcotte and R. G. Percy, in preparation). Historical records (Watt 1907) report the indigenous nature of these distinctive cottons in northeastern South America. Principal component analysis (Fig. 1) shows that accessions from these three regions form a nested series, with Central American accessions embedded within a Caribbean cluster which, in turn, is embedded in a larger east-of-Andes cluster.

The highest genetic identity between any of these three regions and the west Andean center of variation is for accessions from east of the Andes. This suggests direct dispersal from west- to east-of-the-Andes range. In contrast, some of the lowest genetic identity values obtained are between accessions from the Caribbean and Central American regions with accessions from west of the Andes. This argues against a major role for direct dispersal to Central America or the Carribean from west of the Andes, and suggests instead indirect diffusion into these regions via northeastern South America.

If this hypothesis is correct, one might expect at least two opportunities for loss of genetic variation, the first during trans-Andean dissemination and the second during diffusion into Central America and the Caribbean. Consistent with this hypothesis, all measures of diversity calculated from the allozyme data (Table 5) demonstrate the expected pattern of sequential reduction of genetic variation from west of the Andes to east of the Andes to the Caribbean and Central America. In fact, this latter region is genetically depauperate: only four multi-locus genotypes were observed in accessions from Central America, one of which constituted 67% of the sample.

It is uncertain whether dispersal of *G. barbadense* from northeastern South America to the Caribbean and Central America occurred in pre- or post-Columbian times. It is known that by the closing decades of the eighteenth century, kidney-seeded cottons of northeastern South America constituted the chief cotton crop of the West Indies (Watt 1907). The advent of Sea Island cottons in the early nineteenth century, emerging ginning technology, and fiber improvement in *G. hirsutum* displaced kidney-seeded cottons as a commercial crop; they have since persisted in a commensal or feral state.

A subset of the morphological traits common to accessions from east of the Andes has been considered distinctive enough collectively to warrant varietal status as *Gossypium barbadense* var. *braziliense* (Rafin.) Fryx. (Watt 1907; Hutchinson et al. 1947; E. L. Turcotte and R. G. Percy, in preparation). Based on allozyme evidence alone there is little reason to support this distinction; accessions from east of the Andes contain neither unique alleles nor multi-locus distinctiveness from other non-"kidney-seed" accessions from this or other regions.

On the basis of geography alone, one might expect a close affinity between Pacific Island accessions and materials from west of the Andes. This appears not to be the case, however. Rather, Pacific Island accessions appear to be derived from material east of the Andes. The genetic identity ($I=0.994$) between these two regions was the highest observed. This close relationship is graphically depicted by the results of both principal component (Fig. 1) and cluster (Fig. 2) analyses. More distant relationships between Pacific Island accessions and those from the Caribbean and Central America are suggested by the data sets (Table 4, Figs. 1 and 2).

There is both historical and morphological support for a derivation of Pacific island accessions from east-of-Andes material. Stephens (1963) has described, using original sources, the importation of cotton seed and the establishment of a cotton industry in the Pacific Islands during the latter half of the nineteenth century. In his article, Stephens cites a specimen located in the U.S. National Herbarium, which was collected on Fiji by the 1840 U.S. Exploring Expedition. The specimen (and numerous others collected in the Pacific) is kidney-seeded, a common morphological feature of cottons of northeastern South America. An earlier arrival of east-of-Andes cotton to the Pacific is a possibility. Lewton (1920) proposed an export of kidney-seeded cottons to the Indian subcontinent and Malaysia by both the Portuguese and British trading companies. Watt (1907) has cited specimens and several early nineteenth century sources describing kidney-seeded cottons from Malaysia, Polynesia, and India, and their cultivation by agents of the East India Company. Another possible agent of dispersal were the Portuguese who, like the British Empire, had outposts in northeastern South America, India, and Southeast Asia. In this regard, it is noteworthy that two accessions collected from the old Portuguese colony of Goa in India (B1009, B1011) clustered with east-of-Andes accessions in the principal component analysis. However, the Goa accessions were associated on the PCA2 axis with Central American and Caribbean accessions and not with Pacific Island accessions. The loci which differentiate Central American and Pacific accessions on the PCA2 axis also divide east-of-Andes accessions into two nearly equal groups. One might speculate that several, perhaps multiple, independent introductions of *G. barbadense* from northeastern South America to the Indian subcontinent, Southeast Asia, and the Pacific Islands have occurred.

A final dispersal pathway concerns Argentine and Paraguayan accessions, which were interspersed among accessions from west of the Andes in principal component analysis (Fig. 1). The derivation of Argentina-Paraguay accessions from west-of-Andes materials was further supported by genetic identity statistics (Table 4) and cluster analysis (Fig. 2). It has been conjectured (Hutchinson et al. 1947) that the timing of diffusion into Argentina and Paraguay from a northwestern South American source was post-Columbian. This is likely, given the observation that 50% of the Argentina-Paraguay accessions display allozyme evidence of introgression with *G. hirsutum*. It is certain that introgression was post-Columbian, as Argentina and Paraguay are far outside of *G. hirsutum's* native range. Also, many of these accessions are day-neutral, productive annuals (Percival 1987), implying improvement efforts accompanying dispersal. The lack of an extra-long staple (ELS) fiber type among the Argentine and Paraguayan accessions excludes an improved cultivar origin for these accessions and supports a West Andean origin.

Origin of improved cultivars of Gossypium barbadense

In the majority of crop species levels of genetic variation are lower than in their wild progenitors (reviewed in Doebley 1989). This loss presumably reflects genetic bottlenecks experienced by the crop during the domestication process. In *Gossypium barbadense*, however, improved cultivars possess greater genetic variation than all regional groupings of wild and commensal forms, with the exception of accessions from west of the Andes mountains (Table 5). Cultivars exceed the west-of-Andes region in percent polymorphic loci and number of unique alleles, and have equivalent allelic richness.

The high variability observed in improved cultivars relative to regional variation was unexpected. Modern *G. barbadense* cultivars trace to a single Sea Island germ plasm source for their extra-long staple fiber, and thus the level of cultivar variation was expected to be low due to the effects of an extreme genetic bottleneck. Allozyme evidence suggests that any constriction that occurred during cultivar development has been compensated for by interspecific hybridization. Accessions with introgressant *G. hirsutum* alleles constituted 22% of the cultivar sample. Of the 20 polymorphic loci in cultivars, 9 were polymorphic due to introgression alone, and an additional 3 displayed both introgressant and non-introgressant polymorphisms. Moreover, 9 of the 10 unique alleles in cultivars were observed to be common alleles in *G. hirsutum*. This contrasts with material from west of the Andes, where all 9 unique alleles appear to have had a native *G. barbadense* origin.

Evidence from breeding lineages of commercial *G. barbadense* (Kerr 1960; Feaster and Turcotte 1962) supports the interpretation that introgression has contributed considerable genetic variation to modern cultivars. The frequency with which introgression has occurred during *G. barbadense* crop improvement efforts suggests that, in many cases, it has been intentional and has proven desirable. Stephens (1975) has suggested that

the distinctive ELS fiber of Sea Island cottons may have had its origin in an introgressive event.

Historical data concerning the origin of commercial cultivars is limited. All modern ELS cottons are believed to have originated with the Sea Island cottons of the West Indies (Stephens 1976). However, the origin of Sea Island cotton is obscure. Hutchinson and Manning (1945) speculated a Peruvian origin of Sea Island, but there appears to be no historical evidence to support this conjecture. Stephens (1975), in an attempt to recreate the events leading to the creation of a day-neutral, flowering, ELS-fibered plant, made crosses between postulated parents of Sea Island cotton. These putative parents were a West Indian commensal of *G. barbadense* and a wild accession of *G. hirsutum*. Stephens reported partial success in generating an ELS fiber type after several generations of backcrossing and selection. Subsequent to the development of Sea Island cotton, "Egyptian" ELS cotton varieties were developed from hybridization of a perennial Peruvian *G. barbadense* (Jumel's tree cotton) with Sea Island (Balls 1912). It was from this hybridization that modern American, Egyptian, and Russian ELS cottons were derived.

Allozyme evidence suggests that improved cultivars originated, at least in part, from accessions west of the Andes (Figs. 1 and 2; Tables 3 and 5). This supposition is not surprising for many cultivars, given the presumed Peruvian parentage of "Egyptian" cotton. More significant is the affinity of Sea Island accessions to those from west of the Andes. Sea Island accessions clearly cluster with West Andean accessions on the east-west gradient of the PCA1 axis of Fig. 1. PCA2 serves to further separate Sea Island cottons from Central American and Caribbean accessions. This observation lends support to Hutchinson and Manning's (1945) proposal of a west Andean parentage for Sea Island cottons. It also brings into question Stephens' (1975) choice of a West Indies commensal as a possible Sea Island parent. Contradictory morphological, genetic (Stephens 1974), and allozyme evidence does not invalidate Stephens' theory of an interspecific ELS fiber origin, but suggests a more complex ancestry for Sea Island cottons than has been previously proposed.

Patterns of interspecific introgression

The pattern and frequency of interspecific introgression observed among regions and cultivars was contrary to prior expectations. Introgression was expected to be highest within Central American and Caribbean accessions, where *G. barbadense* and *G. hirsutum* are sympatric in commensal and cultivated states. In fact, no introgression was observed within commensal accessions from Central America, and uncultivated accessions from the Caribbean Islands produced the next lowest observed frequency of introgression. While it is possible that some level of introgression exists that went undetected, introgressant alleles were observed at 14 of the 24 polymorphic loci in accessions from other regions; thus, the probability of undetected introgression would seem to be small.

Given the sympatry and consequent frequent opportunity for hybridization between *G. hirsutum* and *G. barbadense* in the above regions, the most plausible explanation for the absence of introgression would appear to be the operation of some form of isolating mechanism. Several potential isolating mechanisms between the species have been described, including differences in the timing of pollen shed (Stephens and Phillips 1972), selective fertilization (Kearney and Harrison 1932), and partial ecological isolation (Mauer 1930). An additional possibility involves the expression of species-specific allelic variants at the "corky" locus (Stephens 1946; Stephens and Phillips 1972). When heterozygous for species-specific alleles, plants are less competitive and partially female sterile. Significantly, Stephens and Phillips (1972) observed that the highest frequency of one of the two species-specific alleles is in regions of sympatry, and that a third "neutral" allele reaches its highest frequency in regions of allopatry. A final isolating mechanism that has been described in cotton is the often observed but variously interpreted phenomenon of "F_2" breakdown" (a high frequency of aberrant recombinant morphologies) and selective elimination of donor parent genotypes in interspecific backcrosses (Harland 1939; Knight 1945; Stephens 1949). Net results of these phenomena include the rapid achievement of recurrent genotypes in backcrossing and the majority recovery of parental types in later-generation segregates.

Each of the above isolating mechanisms has the common characteristic of being incomplete or inefficient. Assuming that some form of reproductive isolation is occurring in Central American and Caribbean commensal accessions, a means of reconciling its presence with observed (high) levels of introgression in cultivar, Argentine, and Paraguayan accessions is necessary. In particular, selective forces for the retention of introgressant traits would appear to be required. Cultivar, Argentine, and Paraguayan accessions share a set of physiological and morphological traits indicative of agronomic manipulation. One might postulate, therefore, that agronomically favorable introgressant traits observed during cultivar improvement efforts would receive positive selection pressure and that, where introgression has persisted in *G. barbadense*, it has been largely due to human manipulation.

Acknowledgements. This work was supported by a grant from the National Science Foundation (BSR-8619631 to JFW). We thank T. Couch for technical assistance and E. Turcotte for his comments during manuscript preparation.

References

Anonymous (1989) Cotton: Review of world situation. Bull Int Cotton Advis Comm 42:1–15

Balls WL (1912) The cotton plant in Egypt. Macmillan, London

Bourdon C (1986) Enzymatic polymorphism and genetic organization of two cotton tetraploid cultivated species, *G. hirsutum* and *G. barbadense*. Cotton Fibres Trop 41:191–210

Doebley JF (1989) Isozyme evidence and the evolution of crop plants. In: Soltis DE, Soltis PS (eds) Isozymes in plant biology. Dioscorides Press, Portland/OR (in press)

Feaster CV, Turcotte EL (1962) Genetic basis for the varietal improvement of Pima cottons. Crops Re, USDA-ARS 34-31, pp 1–15

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

Harland SC (1936) Some notes on cotton in Columbia. Trop Agric/Trinidad 13:1–13

Harland SC (1939) Genetical studies in the genus *Gossypium* and their relationship to evolutionary and taxonomic problems. Proc 7th Int Genet Congr Edinburgh, pp 138–143

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

Hutchinson JB, Silow RA, Stephens SG (1947) The Evolution of *Gossypium*. Oxford University Press, London

Kearney TH, Harrison GJ (1932) Pollen antagonism in cotton. J Agric Res 44:191–226

Kerr T (1960) The potentials of barbadense cottons. Proc 12th Ann Cotton Imp Conf, Memphis/TN, pp 57–60

Knight RL (1945) The theory and application of the backcross technique in cotton breeding. J Genet 47:76–86

Lewton FL (1920) The history of kidney cotton. J Wash Acad Sci 10:591–597

Mauer FM (1930) The cottons of Mexico, Guatemala, and Colombia. Bull Appl Bot Genet Pl Breed Suppl 47:543–553

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Percival AE (1987) The National Collection of *Gossypium* germ plasm. Southern Coop Ser Bull 321:309–362

Sneath PHA, Sokal RR (1973) Numerical taxonomy. Freeman, San Francisco

Stephens SG (1946) The genetics of "corky". I. The New World alleles and their possible role as an interspecific isolating mechanism. J Genet 47:150–161

Stephens G (1949) The cytogenetics of speciation in *Gossypium*. I. Selective elimination of the donor parent genotype in interspecific backcrosses. Genetics 34:627–637

Stephens SG (1963) Polynesian cottons. Ann Mo Bot Garden 50:1–22

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

Stephens SG (1975) Some observations on photoperiodism and the development of annual forms of domesticated cottons. Econ Bot 30:409–418

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

Stephens SG, Mosely ME (1974) Early domesticated cottons from archeological sites in central coastal Peru. Am Antiq 39:109–122

Stephens SG, Phillips LL (1972) The history and geographical distribution of a polymorphic system in New World cottons. Biotropica 4:49–60

Watt G (1907) The wild and cultivated cotton plants of the world. Longmans, Green, London

Weeden NF, Wendel JF (1989) Genetics of plant isozymes. In: Soltis DE, Soltis PS (eds) Isozymes in plant biology. Dioscorides Press, Portland/OR, (in press)

Wendel JF, Olson PD, Stewart JM (1989) Genetic diversity, introgression, and independent domestication of Old World cultivated cottons. Am J Bot (in press)

Wendel JF, Weeden NF (1989) Visualization and interpretation of plant isozymes. In: Soltis DE, Soltis PS (eds) Isozymes in plant biology. Dioscorides Press, Portland/OR (in press)