American Journal of Botany 79(11): 1291–1310. 1992.

GENETIC DIVERSITY IN GOSSYPIUM HIRSUTUM AND THE ORIGIN OF UPLAND COTTON[1]

JONATHAN F. WENDEL,[2,4] CURT L. BRUBAKER,[2] AND A. EDWARD PERCIVAL[3]

[2]Department of Botany, Iowa State University, Ames, Iowa 50011; and
[3]USDA, Agricultural Research Service, Route 5, Box 805, College Station, Texas 77845

Gossypium hirsutum has a large indigenous range encompassing most of Mesoamerica and the Caribbean, where it exhibits a diverse array of morphological forms spanning the wild-to-domesticated continuum. Modern, highly improved varieties ("Upland cotton"), which currently account for about 90% of world cotton commerce, are day-length neutral annuals derived from subtropical, perennial antecedents. To assess levels and patterns of genetic variation in the species and to elucidate the origin of Upland cotton, 538 accessions representing the full spectrum of morphological and geographical diversity were analyzed for allozyme variation at 50 loci. Levels of variation are modest overall but are low in Upland cotton. Relationships among accessions reflect pre-Columbian influences of aboriginal peoples and later European colonists superimposed on the preagricultural pattern. In contrast to expectations, two centers of diversity are evident, one in southern Mexico-Guatemala and the other in the Caribbean. Introgression of *G. barbadense* genes into *G. hirsutum* has been common in a broad area of sympatry in the Caribbean. The germplasm of present cultivars traces to Mexican highland stocks, which, in turn, were derived from material originally from southern Mexico and Guatemala. Despite the widespread belief that germplasm from several other species has been incorporated into modern Upland stocks through intentional breeding efforts, the 50 Upland cultivars examined contain no unique alleles, suggesting that retention of genes from transspecific sources has been minimal. The most recent infraspecific treatment, which recognizes seven races, does not adequately represent genetic relationships.

The genus *Gossypium* L. includes four species of cultivated cottons, providing the world's most important textile fiber and its second most valuable oil and meal seed. At present, American tetraploid species (*G. barbadense* L. and *G. hirsutum* L.) dominate worldwide cotton production, having displaced virtually all Old World diploid cultivars (*G. arboreum* L. and *G. herbaceum* L.; Hutchinson, 1959; Lee, 1984). *Gossypium barbadense* ("Extra-long staple," "Pima" or "Egyptian" cotton) is favored for some purposes due to its long, strong, and fine fibers, but its relatively low yield has limited its importance to about 8% of total world production (Lee, 1984). The bulk of the world's cotton is supplied by modern cultivars ("Upland cotton") of *G. hirsutum.* Upland cultivars currently are grown in more than 40 nations in both tropical and temperate latitudes, from 47° N in the Ukraine and 37° N in the United States to 32° S in South America and Australia (Niles and Feaster, 1984). In 1990, 12.4 million acres of cotton were grown in the United States alone, and its aggregate (fiber and seed) market value totaled approximately $5.5 billion (Anonymous, 1990).

Modern Upland cultivars are high-yielding, day-length neutral, early-cropping plants (hereafter "annuals") with easily ginned, abundant fiber. These "improved" characteristics resulted from human selection from perennial ancestors with shorter, sparser fiber. This domestication process is widely believed to have been accompanied by an extreme reduction in genetic diversity, relative to the less "improved" forms (Anonymous, 1972; Endrizzi, Turcotte, and Kohel, 1985). *Gossypium hirsutum* is a highly diverse species throughout most of its indigenous range, which includes most of Mesoamerica, northern and northeastern South America, the Caribbean, and numerous, distant islands in both the Pacific and Atlantic oceans (Fryxell, 1979). The wealth of morphological and ecological diversity in *G. hirsutum* is underscored by the taxonomic and nomenclatural history of the species—at least 30 specific epithets have been variously applied at one time or another to some portion of the species (see Fryxell, 1968, 1976, 1979). At present, most recognize only a single species for the assemblage, although Fryxell (1979) provisionally accepted *G. lanceolatum* Todaro (primarily from the western Mexican states of Guerrero and Oaxaca) as a distinct species, based, at least in part, on seed-protein electrophoretic results of Johnson (1975).

The most widely followed infraspecific classifications are those of Hutchinson, Silow, and Stephens (1947) and Hutchinson (1951). In the earlier treatment, three taxonomic varieties were recognized—var. *hirsutum,* which includes Upland cotton and other early-cropping forms, centered in Guatemala but encompassing much of Mesoamerica (and later worldwide); and two, mostly perennial varieties including a series of morphological forms and ecotypes in the wild-to-domesticated continuum: the primarily Central American var. *punctatum* (Schumacher) Hutch. and var. *marie-galante* (Watt) Hutch. from northern and northeastern South America and the Caribbean (geographical ranges illustrated in Hutchinson, Silow, and Stephens, 1947, p. 105). Hutchinson later replaced this classification with an informal system that recognizes seven geographical races (Hutchinson, 1951); this later treatment represents the last thorough consideration of infraspecific categories in *G. hirsutum.* It is also the system used in classifying accessions in the National Collection of *Gossypium* Germplasm, maintained in College Station, Texas (Percival, 1987).

Three of Hutchinson's geographical races correspond

[1] Received for publication 29 March 1992; revision accepted 10 July 1992.

The authors thank D. Bates, P. Bretting, P. Fryxell, R. Percy, J. Vreeland, and F. Wilson for many helpful comments on the manuscript; B. Loetel for Fig. 1; and T. Couch for technical assistance. This research was supported by NSF grants BSR-8619631 and BSR-8918041 to JFW.

[4] Author for correspondence.

to his earlier varieties, now recognized as races 'latifolium', 'punctatum', and 'marie-galante'. These three races were considered to have "spread further than the rest, and [to] have undergone further differentiation" (Hutchinson, 1951, p. 163). Race 'latifolium' occupies a relatively narrow indigenous range in Guatemala and Chiapas, Mexico, from where it was envisioned to have spread elsewhere. Race 'marie-galante' is distributed from El Salvador eastward through Costa Rica and Panama to the larger part of its range as described above. Race 'punctatum' has a more northern distribution in the Yucatán Peninsula, Gulf coastal states of Mexico, and in many areas of the Caribbean, from where it spread in post-Columbian times to many areas of the New World. The other four races, all perennial, are more narrowly distributed: race 'palmeri' (= *G. lanceolatum*) occurs primarily in the southern part of western Mexico; race 'morrilli' is from the Mexican plateau from Oaxaca, Puebla, and Morelos northward to Sonora and Sinaloa; race 'yucatanense' is narrowly distributed in the northern, coastal part of the Yucatán Peninsula; race 'richmondi' is confined to the Pacific side of the Isthmus of Tehuantepec (southern Mexico and Guatemala). In addition to geographical distribution, plant habit characters (e.g., size and general growth form, number of branches and their aspect) provide most of the morphological distinctions between the races, supplemented by several leaf and seed characters, although Hutchinson admitted that racial distinctions tend to be obscured in regions of racial sympatry and in the "old and often mutilated house yard and hedgerow plants" (Hutchinson, 1951, p. 165).

Six of Hutchinson's geographical races (all but 'yucatanense') exist in various stages of domestication. Most common are "dooryard" cottons, which are commensals cultivated as solitary plants or in small plots for household or local needs, e.g., medicinal infusions, wound dressings, pillow stuffing, lamp wicks (Stephens, 1958). These plants typically are grown for many years, and may develop trunks of up to 30 cm in diameter, especially in race 'marie-galante'. They are widely believed to have undergone little deliberate selection and are thought to be derived largely from local progenitors. These cottons also are considered a frequent source of feral escapes, as they are often found in abandoned clearings, waste areas, and edges of villages. Commercial scale plantings primarily involve race 'latifolium', although early-cropping forms of race 'punctatum' have been developed in the Old World (Hutchinson, 1951).

The oldest archeological remains of *G. hirsutum* are from the Tehuacán Valley of Mexico, dating to from 4,000 to 5,000 years before present (BP) (Smith and Stephens, 1971), although this date should be considered tentative until additional stratigraphic and carbon-14 dating information become available (P. Fryxell and J. Vreeland, personal communication). These cottons appear to have been introduced, domesticated forms, suggesting that *G. hirsutum* has an ancient history of cultivation.

Many basic details of the natural history of *G. hirsutum* are uncertain. For example, its geographical range prior to human contact is unknown. Even the pre-Columbian range of *G. hirsutum* cannot be considered the original distribution; rather, it most likely reflects the native range expanded by several millenia of migrations and development in association with the movement of Amerindians. Later arrival of European colonists increased the complexity by dramatically increasing the rate of spread and exchange of germplasm throughout the New World tropics, together with subsequent transfer to the Old World tropics.

Because of the antiquity and complexity of human influences on the development and diffusion of *G. hirsutum*, we initiated a comprehensive electrophoretic study of accessions collected from throughout the species' range. Our purposes were to: 1) infer the time and place of domestication of *G. hirsutum*; 2) elucidate the origin of modern, Upland cotton and to characterize the genetic bottleneck that is thought to have accompanied its development; 3) describe levels and patterns of genetic variability and clarify genetic relationships among accessions collected from throughout the species' range; 4) evaluate whether Hutchinson's (1951) informal racial system provides a reasonable or sufficient perspective of infraspecific variation; and 5) quantify the magnitude of interspecific introgression of *G. barbadense* into *G. hirsutum* from different geographic regions.

MATERIALS AND METHODS

Plant materials—The National Collection of *Gossypium* Germplasm, College Station, Texas, is an assemblage of seed accessions of *Gossypium* species consolidated from smaller collections previously maintained in Texas, Mississippi, and Arizona. The germplasm originally accumulated from plant explorations, variety development programs, seed donations, and seed exchanges with other collections. Of the 1,990 accessions of feral, wild, or commensal *G. hirsutum* maintained at College Station, approximately 24% have been classified using Hutchinson's (1951) racial classification system. A sampling of 538 *G. hirsutum* accessions was selected for analysis (Table 1). Samples consisted of either original field-collected seed or seed derived from one to several cycles of self-pollination subsequent to collection. Accessions were selected to maximize geographical coverage and morphological diversity. Locality and collector information for most accessions are listed in Percival (1987). Most accessions show some level of human influence and represent feral escapes or commensal ('dooryard') cottons. If Fryxell's (1979) lint percentage (fiber weight/[seed weight + fiber weight]) figures are used as indicators of 'level of domestication' (25%–40% for commensal cottons, 18%–20% for feral forms, and 8%–10% for wild, littoral populations), of the 137 accessions included for which data are available, 61.3% can be considered commensal, 31.4% feral, and 7.3% putatively wild. One hundred seventy of the accessions studied (21.5%) have been classified to race: 56 'latifolium', 39 'marie-galante', four 'morrilli', 11 'palmeri', eight 'punctatum', two 'richmondi', and four 'yucatanense'. In addition to the feral, wild, and commensal accessions, 50 modern Upland cultivars representing the four categories of cotton commercially grown in the United States (Eastern, Plains, Delta, and Acala) were included. The 50 cultivars selected, along with their close relatives and derivatives, account for the majority of *G. hirsutum* cultivation in the United States.

TABLE 1. *Accessions of* Gossypium hirsutum *studied for allozyme variation, organized by geographic region and country of collection*

No.	Region	Colonial history[a]	Locale	Texas No.[b]
1	Bahamas	British	Eleuthera	807, 808, 809, 810, 811, 812, 813, 1340
			Great Exuma	1342
			New Providence	804, 805, 806
			Rum Cay	1343, 1344, 1345
			San Salvador	1346
2	Barbados	British	Maycock's Bay	1874
			Six Men's Bay	1875
			St. Philip	1876
3	Belize	British	Belize City	724, 725, 794
			Corozal	766, 789
			Pa Achaca	1442
			Punta Gorda	781
			San Antonio	784
			San Pedro	786
			Stann Creek	778
4	Cuba	Spanish	?	902(MG), 903(MG)
			Camaguey	801, 802, 803
			Sibaney	799
5	Dominica	British	Cailbishie	1556
			Carib Reserve	1075
			Colihaut	1555, 1807
			Morne Raquette	1554
			Roseau	1374
			Tarou	1553
6	Dominican Republic	Spanish	?	1260, 1261, 1262, 1263
			Azua	1572, 2265, 2266, 2267, 2268, 2270, 2271, 2272, 2274, 2276
			Bani	1816
			Barahona	911
			Bastidas	1822
			Boca Chica	1579, 1580
			Cortes	1820
			La Vega	1574
			Los Bancos	1573
			Peravia	2277
			San Pedro de Mecoris	1578, 1826
			Santiago	1577
			Santo Domingo	884(MG), 885(MG), 989, 2278
7	French West Indies I	French	Desirade	1841
			Guadeloupe	984, 1606(MG), 1608(MG), 1613(MG), 1619, 1620, 1624, 1627, 1630, 1644, 1655, 1750, 1757, 1759, 1760, 1859, 1862, 1863, 2058
			Ils des Saintes	1603, 1789, 1791, 1792
			Marie-Galante	1779(MG), 1842(MG), 1843(MG), 1848(MG), 1850(MG), 1851(MG), 1853(MG), 1854(MG), 2045(MG), 2049(MG)
8	French West Indies II	French	Martinique	866(MG), 1267, 1533, 1534, 1538, 1540, 1544, 1545, 1546, 1547, 1801, 1806
9	French West Indies III	French/Dutch	St. Barthelemy	1557, 1560, 1564, 1565, 1566, 1794, 1795, 1797, 1809, 2061
			St. Martin	1568, 1569, 1571
10	Cayman Islands	British	Grand Cayman	2249, 2250, 2252, 2253, 2254, 2255
11	Grenada/St. Lucia	British	Grenada	853(MG), 856(MG), 858(MG), 907
			St. Lucia	1870
12	Guatemala I	Spanish	Chiquimula	98(L), 106(L), 123(L), 142(L), 169(L), 209(L), 221(L), 240(L), 683, 685
			Jutiapa	93(L), 96(L), 97(L), 99(L), 100(L), 101(L), 111(MG), 119(L), 140(L), 141(MG), 150, 151, 153(L), 154(L), 167(L), 168(L), 184(MG), 196(L), 197(L), 198(L), 213(L), 214(L), 217(L), 234(L), 236(L), 237(L)
13	Guatemala II	Spanish	Escuintla	675
			Guatemala	107(L)
			Jalapa	122(L), 155(L), 156(L), 177(L), 200(L), 201(L), 220(L), 238(L)
			Retalhuleu	69(L)
			Santa Rosa	116(L), 149, 180(L), 367(MG), 372(MG), 379(MO)
			Zacapa	94(PU), 114(PU), 115(PU), 166(MO), 210(MO), 230(PU)

TABLE 1. *Continued*

No.	Region	Colonial history[a]	Locale	Texas No.[b]
14	Guatemala III	Spanish	Alta Verapaz	1156, 1160, 1163, 1164, 1165
			Baja Verapaz	188(L), 241(L), 1443
			Huehuetenango	242(L), 1166
			Petén	479, 650, 654, 656, 660, 661
15	Haiti	French	Bedoue	1601
			Dufailly	1597
			Dufort	1582, 1583
			Jacmel	1585, 1586, 1587, 1588
			Miragoane	1829
			Mirebalais	1595, 1596
			Petionville	895(MG)
			Pierre-Payen	1600
			Port-au-Prince	893(MG)
			St. Louis du Sud	1594
			Thomonde	1598
			Vignier	1833
			Violet	1590, 1591
16	Honduras	Spanish	San Lorenzo	693
			Choluteca	691
			Cortes	695, 698, 699, 704, 706, 707, 708
17	Jamaica	British	?	996, 1216, 1229
			Cornwall	2231, 2232, 2233, 2234, 2235, 2236
			Middlesex	2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2228, 2229, 2237, 2238
			Palisadoes	967
			Surrey	2240, 2241, 2243, 2244, 2245, 2246, 2247
18	Mexico I	Spanish	Puebla	6(L), 746
			San Luis Potosí	762, 1329, 1332, 1336, 1366, 1379, 1382, 1449
			Tamaulipas	1041, 1220, 1458, 1459, 2309
			Veracruz	755, 756, 757(PA), 758, 948, 959, 960, 1112
19	Mexico II	Spanish	Campeche	1040, 2003, 2005, 2009, 2013, 2081, 2082, 2083(PA)
			Quintana Roo	1972, 1973, 1975, 1976, 1979, 1980, 1981
			Tabasco	2076
			Yucatán	481(PU), 488(PU), 491(PU), 493(L), 619, 670, 745, 1039(Y), 1046(Y), 1749, 1986, 2089(PA), 2094(Y), 2096(Y)
20	Mexico III	Spanish	Baja California	1182, 1183, 1247, 1257
			Sinaloa	776, 1114, 1115
21	Mexico IV	Spanish	Colima	1349, 1472, 1473, 2329, 2330, 2331
			Guerrero	225(L), 226(L), 315, 322(PA), 340(PA), 702, 1963, 1965, 1967, 1968
			Guerrero?	1(PA), 9(PA), 11[c]
			Michoacán	1113, 1117, 1530(PA), 1959, 1045(PA)
			Morelos	751
			Oaxaca	109(L), 192(MO), 232(L), 253(L), 303(PA), 316(PA), 460, 461(R), 750, 953
22	Mexico V	Spanish	Chiapas	21(L), 34(L), 44(PU), 58(L), 60(L), 770, 775, 1102(R), 1167, 2077
23	Curaçao	Dutch	Bullenbaai	2192
			Fuik Bay	2198, 2203
			Jan Kock	2193
			Newport	2204
			Weg Narr Fuik	2197
			West Pund Bay	2194
			Willemstad	2190, 2191, 2195
24	Bonaire	Dutch	Jato	2209
			Kraiendijk	2206
			Pis Pis	2208
			Playa Frans	2210
			Playa Pa Riba	2207
			Rincon	2212
25	Aruba	Dutch	Malmok	2218
			Oranjestad	2213, 2217
			Parkietenbos	2214
			Seroe Colorado	2216
26	Nicaragua	Spanish	?	1392
			Boaco	1376, 1383, 1385, 1391
			Carazo	1430

TABLE 1. *Continued*

No.	Region	Colonial history[a]	Locale	Texas No.[b]
			Chinandega	709
			Granada	713
			Isla de Maíz	720
			Managua	714, 956, 1091(MG), 1429
			Paso Caballos	718
			Puerto Corinto	710, 712
			San Benito	1316, 1317
			Somote	1010(MG)
27	Panama	Spanish	?	1314
			Aguadulce	1315
			Panama City	974
28	Puerto Rico	Spanish	?	2284
			Arecibo	2300
			Bahia de la Ballena	2288, 2290
			Cubo Rojo	883(MG)
			Fajardo	2280
			Guanica	878(MG), 882(MG)
			Guayama	2283
			La Pica	1352
			Mayaguez	1356, 1360, 1362
			Mona Island	816, 1240, 1278, 1289
			Parguera	2295
			Playa Cana Gorda	2287
			Playa Salinas	2298
			Ponce	2285
			Rincon	1089
			Salinas	2292
			Tamarind Beach	879(MG), 1243
29	El Salvador	Spanish	?	235(L), 380(MG), 390
			Matapan	1009(MG)
30	Trinidad & Tobago	British	Tobago	2185, 2186, 2188, 2189
			Trinidad	1069, 2122, 2125, 2126, 2131, 2134, 2137, 2141, 2144, 2147, 2152, 2155, 2160, 2164, 2171, 2173, 2175, 2178, 2182
31	UK Virgin Islands	British	Tortola Island	1367, 1368, 1369, 1370
			Virgin Gorda	1371
32	St. Kitts	British	St. Kitt's Island	871(MG), 1372, 1373, 2024, 2025
33	USA Virgin Islands	USA	St. Thomas	873(MG), 874(MG), 986, 1237
34	Modern Upland Cultivars	Acala	SJ-2	
		Western (Acala-like)	BR-110, BR-115	
		Delta		
			Arkot 518, Delcot 344, Deltapine 20, Deltapine 41, Deltapine 50, Deltapine 90, DES 119, Stoneville 112, Stoneville 302, Stoneville 453, Stoneville 825, Stoneville 1014, Stoneville 1324, Stoneville 6413, Stoneville 8911	
		Delta/Plains	Tamcot CAB-CS, Tamcot CAMD-E, Tamcot SP-215	
		Plains	All-Tex 857, All-Tex E-2, All-Tex Quickie, C4HUGBEH-1-2-86, CABUCD3H-1-86, Cascot 4, Cascot 41, Cascot 2910, Cascot L-7, Cascot L-13, Cencot, Dunn 325, Dunn 1047, Dunn HS-120, GSC 25, GSC 27, GSC 30, Lankart 311, Lankart 511, Lankart 571, Lankart LX571, Lankart PR75, Paymaster 145, Paymaster 505, Paymaster H86010, Paymaster H86048	
		Eastern	Coker 130-16905, Coker 315, Coker 139	

[a] Predominant colonial history is listed. Cultivars (group 34) are listed by their four major types under the column "Colonial history."

[b] Accessions that have been racially classified following Hutchinson (1951) are indicated parenthetically as follows: L = 'latifolium'; MG = 'marie galante'; MO = 'morrilli'; PA = 'palmeri'; R = 'richmondi'; Y = 'yucatanense'; PU = 'punctatum'.

[c] Morphological observations indicate that this accession is misclassified as race 'palmeri'.

Isozyme electrophoresis—Because most accessions have been maintained as inbred lines, little within-accession variation was expected (or evident); consequently, only four individuals were analyzed per accession. Tissue for electrophoresis consisted of imbibed seeds (24 hr) or young cotyledons. Sample preparation and electrophoresis buffer conditions are detailed in Percy and Wendel (1990), and staining protocols are described in Wendel and Weeden (1989). Fourteen enzyme systems were assayed: aspartate aminotransferase (AAT), endopeptidase (ENP), glutamate dehydrogenase (GDH), NADP-isocitrate dehydrogenase (IDH), malate dehydrogenase (MDH), NADH-dehydrogenase (= "menadione reductase," NAD), phosphoglucomutase (PGM), triose-phosphate isomerase (TPI), aconitate hydratase (ACO), alcohol dehydrogenase (ADH), 6-phosphogluconate dehydrogenase (PGD),

TABLE 2. *Allele frequencies*[a] *at 30 polymorphic allozyme loci in* Gossypium hirsutum

Region	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
N	16	3	10	6	7	30	34	12	13	6	5	36	23	16	19	9	29	23
Aat1-2	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-4	1.00	1.00	1.00	1.00	1.00	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aat2-1	0.53	0.33	0.20	0.17	0.00	0.08	0.00	0.08	0.12	0.00	0.00	0.03	0.11	0.16	0.00	0.00	0.03	0.26
-4	0.06	0.33	0.50	0.67	1.00	0.72	0.68	0.71	0.69	0.75	0.80	0.21	0.37	0.38	0.95	0.70	0.90	0.33
-6	0.41	0.33	0.30	0.17	0.00	0.20	0.32	0.21	0.19	0.25	0.20	0.76	0.52	0.47	0.05	0.30	0.07	0.41
-8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Aat4-2	0.37	0.33	0.40	0.00	0.00	0.03	0.00	0.08	0.15	0.00	0.00	0.03	0.04	0.19	0.00	0.05	0.00	0.04
-4	0.63	0.67	0.60	1.00	1.00	0.97	1.00	0.92	0.85	1.00	1.00	0.94	0.96	0.81	1.00	0.95	1.00	0.96
-6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
-n	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Aco1-4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00
-8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94	1.00	0.90	0.98	1.00
-n	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.10	0.00	0.00
Aco3-1	0.88	0.67	0.80	0.67	0.21	0.72	0.35	0.25	0.54	0.67	0.10	0.92	0.87	0.88	0.32	0.70	0.36	0.83
-3	0.06	0.33	0.20	0.33	0.79	0.28	0.65	0.75	0.46	0.17	0.90	0.08	0.13	0.12	0.68	0.30	0.60	0.17
-n	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
Aco5-2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-3	0.00	0.00	0.00	0.00	0.00	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-4	0.66	0.00	0.50	0.50	0.93	0.61	0.72	0.92	0.62	1.00	0.70	0.90	0.70	0.72	0.42	0.35	0.43	0.57
-6	0.34	1.00	0.50	0.50	0.07	0.23	0.28	0.08	0.39	0.00	0.30	0.10	0.30	0.28	0.58	0.65	0.57	0.44
Adh1-2	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.07	0.00
-4	0.97	1.00	1.00	1.00	1.00	0.95	1.00	0.92	1.00	1.00	1.00	1.00	1.00	1.00	0.90	1.00	0.93	1.00
-6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-n	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Adh2-4	0.97	1.00	1.00	0.67	1.00	0.75	0.97	1.00	0.96	1.00	1.00	1.00	0.96	1.00	0.76	0.80	0.91	0.96
-6	0.00	0.00	0.00	0.33	0.00	0.25	0.03	0.00	0.04	0.00	0.00	0.00	0.04	0.00	0.24	0.20	0.09	0.04
-n	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Arg1-3	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-4	0.56	0.33	0.80	0.33	0.43	0.85	0.24	0.25	0.12	0.83	0.20	1.00	0.96	0.81	0.50	0.90	0.59	0.87
-5	0.25	0.67	0.10	0.50	0.57	0.17	0.76	0.75	0.88	0.17	0.80	0.00	0.04	0.06	0.47	0.00	0.40	0.04
-n	0.13	0.00	0.10	0.17	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.03	0.10	0.02	0.09
Arg2-1	0.06	0.00	0.00	0.00	0.07	0.20	0.10	0.08	0.04	0.17	0.40	0.53	0.30	0.34	0.00	0.00	0.14	0.13
-2	0.94	1.00	0.95	1.00	0.93	0.78	0.90	0.92	0.96	0.83	0.60	0.47	0.70	0.66	1.00	1.00	0.83	0.87
-6	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
-n	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Enp1-3	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.08	0.09	0.00	0.00	0.25	0.00	0.44
-4	0.94	1.00	1.00	1.00	1.00	0.98	1.00	0.92	1.00	1.00	1.00	0.92	0.91	1.00	0.97	0.75	1.00	0.57
-5	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00
-n	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Enp2-2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00
-4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90	1.00
Gdh1-1	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-2	0.81	1.00	0.90	1.00	0.50	0.63	0.19	0.17	0.50	0.83	0.80	0.93	0.96	1.00	0.90	0.80	0.43	0.94
-6	0.19	0.00	0.10	0.00	0.50	0.37	0.78	0.83	0.50	0.17	0.20	0.07	0.04	0.00	0.11	0.20	0.57	0.07
-n	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Idh1-2	0.44	0.67	0.90	0.50	0.00	0.02	0.10	0.21	0.08	1.00	0.00	1.00	0.91	0.88	0.05	0.75	0.10	0.85
-4	0.00	0.00	0.00	0.00	0.00	0.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.26	0.00
-8	0.56	0.33	0.10	0.50	1.00	0.70	0.90	0.79	0.92	0.00	1.00	0.00	0.09	0.13	0.95	0.25	0.64	0.15
Idh2-4	1.00	0.67	1.00	0.83	0.86	0.88	0.96	1.00	0.96	1.00	0.60	1.00	1.00	1.00	1.00	1.00	0.85	1.00
-5	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-n	0.00	0.33	0.00	0.17	0.14	0.05	0.04	0.00	0.04	0.00	0.40	0.00	0.00	0.00	0.00	0.00	0.16	0.00
Leu1-2	0.28	0.00	0.20	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.09	0.00	0.00	0.00	0.00	0.00
-4	0.44	1.00	0.70	0.17	1.00	0.97	0.94	1.00	1.00	1.00	0.60	0.46	0.57	0.56	0.87	0.90	1.00	0.80
-5	0.28	0.00	0.10	0.17	0.00	0.03	0.06	0.00	0.00	0.00	0.40	0.51	0.35	0.44	0.13	0.10	0.00	0.20
-n	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mdh4-4	0.41	1.00	1.00	0.83	1.00	1.00	1.00	0.96	1.00	0.83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96
-6	0.59	0.00	0.00	0.17	0.00	0.00	0.00	0.04	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04

TABLE 2. *Continued*

19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34		
28	7	32	10	10	6	5	19	3	25	4	23	5	5	4	50	Mean1	Mean2
0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.009	0.008
1.00	1.00	1.00	1.00	1.00	1.00	0.80	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.990	0.992
0.63	0.00	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.085	0.098
0.17	0.43	0.48	0.20	0.90	1.00	0.60	0.74	1.00	0.82	0.25	0.96	1.00	0.90	1.00	0.07	0.625	0.540
0.20	0.57	0.47	0.75	0.10	0.00	0.20	0.26	0.00	0.10	0.75	0.04	0.00	0.10	0.00	0.93	0.283	0.360
0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.005	0.002
0.68	0.00	0.02	0.10	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.02	0.00	0.00	0.00	0.00	0.078	0.083
0.32	1.00	0.98	0.90	1.00	1.00	1.00	0.95	1.00	0.88	1.00	0.98	1.00	1.00	1.00	0.98	0.918	0.915
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.001	0.001
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.001	0.001
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.001	0.001
1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.993	0.994
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.006	0.006
0.78	1.00	0.84	0.60	0.70	1.00	0.00	0.37	0.33	0.58	0.50	0.13	0.40	0.50	0.00	0.90	0.570	0.636
0.18	0.00	0.16	0.40	0.30	0.00	1.00	0.63	0.67	0.42	0.50	0.87	0.60	0.50	1.00	0.10	0.420	0.357
0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.008	0.007
0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.004	0.002
0.00	0.00	0.00	0.00	0.30	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.018	0.017
0.14	0.86	0.42	0.30	0.65	0.67	1.00	0.42	0.67	0.66	0.25	1.00	1.00	0.90	0.50	0.40	0.618	0.590
0.86	0.14	0.58	0.70	0.05	0.00	0.00	0.58	0.33	0.34	0.75	0.00	0.00	0.10	0.50	0.60	0.358	0.392
0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.013	0.014
1.00	0.86	1.00	1.00	1.00	0.83	1.00	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00	0.980	0.982
0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.004	0.002
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.001	0.002
1.00	0.93	1.00	1.00	1.00	1.00	0.90	0.42	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00	0.939	0.932
0.00	0.07	0.00	0.00	0.00	0.00	0.10	0.58	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.059	0.066
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.001	0.002
0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.002	0.004
0.93	1.00	0.55	1.00	0.50	0.50	0.40	0.95	1.00	0.34	0.75	0.87	0.00	0.60	0.00	1.00	0.614	0.692
0.07	0.00	0.00	0.00	0.45	0.33	0.60	0.00	0.00	0.52	0.25	0.13	1.00	0.40	1.00	0.00	0.335	0.245
0.00	0.00	0.42	0.00	0.05	0.17	0.00	0.05	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.047	0.059
0.07	0.00	0.07	0.25	0.15	0.33	0.20	0.00	0.00	0.04	0.50	0.00	0.00	0.20	0.00	0.43	0.141	0.167
0.93	1.00	0.94	0.75	0.70	0.67	0.80	1.00	1.00	0.96	0.50	1.00	1.00	0.80	1.00	0.57	0.850	0.826
0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.005	0.005
0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.001	0.002
0.00	0.57	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.059	0.065
1.00	0.43	0.76	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	1.00	0.80	1.00	1.00	0.938	0.932
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.001	0.002
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.001	0.002
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.001	0.003
0.00	0.00	0.00	0.10	0.00	0.17	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.012	0.011
1.00	1.00	1.00	0.90	1.00	0.83	1.00	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.986	0.986
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.006	0.004
0.95	1.00	1.00	1.00	0.40	0.17	0.20	0.66	1.00	0.38	1.00	0.96	0.00	0.40	0.00	1.00	0.687	0.735
0.05	0.00	0.00	0.00	0.55	0.83	0.80	0.29	0.00	0.60	0.00	0.04	1.00	0.40	1.00	0.00	0.301	0.258
0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.003	0.004
0.88	1.00	1.00	1.00	0.00	0.00	0.00	0.34	0.00	0.24	0.75	0.00	0.00	0.40	0.00	1.00	0.442	0.520
0.05	0.00	0.00	0.00	0.30	0.33	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.20	0.00	0.00	0.043	0.047
0.07	0.00	0.00	0.00	0.70	0.67	1.00	0.66	1.00	0.68	0.25	1.00	1.00	0.40	1.00	0.00	0.513	0.433
0.96	1.00	1.00	1.00	0.70	1.00	0.80	1.00	1.00	0.88	1.00	1.00	1.00	0.80	1.00	1.00	0.937	0.956
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.007	0.006
0.04	0.00	0.00	0.00	0.30	0.00	0.20	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.055	0.039
0.00	0.00	0.02	0.10	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.042	0.030
0.89	0.14	0.69	0.40	1.00	1.00	0.80	0.76	1.00	0.94	1.00	0.96	1.00	0.80	1.00	0.03	0.776	0.719
0.11	0.86	0.26	0.50	0.00	0.00	0.20	0.16	0.00	0.06	0.00	0.04	0.00	0.20	0.00	0.97	0.179	0.249
0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.000	0.002
0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	0.85	1.00	1.00	1.00	0.86	0.958	0.951
0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.15	0.00	0.00	0.00	0.14	0.041	0.049

TABLE 2. *Continued*

Region	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
N	16	3	10	6	7	30	34	12	13	6	5	36	23	16	19	9	29	23
Mdh6-4	0.09	0.00	0.25	0.67	0.14	0.40	0.53	0.21	0.50	0.17	0.00	0.57	0.46	0.28	0.40	0.60	0.59	0.30
-6	0.78	1.00	0.70	0.25	0.07	0.28	0.10	0.29	0.04	0.83	1.00	0.43	0.46	0.53	0.26	0.30	0.28	0.70
-n	0.13	0.00	0.05	0.08	0.78	0.32	0.37	0.50	0.46	0.00	0.00	0.00	0.09	0.19	0.34	0.10	0.14	0.00
Nad1-4	1.00	1.00	1.00	0.83	1.00	1.00	0.97	1.00	0.73	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
-n	0.00	0.00	0.00	0.17	0.00	0.00	0.03	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pgd1-1	0.06	0.00	0.40	0.17	0.29	0.58	0.37	0.17	0.27	0.83	0.00	0.15	0.30	0.56	0.66	0.85	0.66	0.50
-3	0.94	1.00	0.60	0.83	0.71	0.42	0.63	0.83	0.73	0.17	1.00	0.85	0.70	0.44	0.34	0.15	0.34	0.50
Pgd2-4	1.00	1.00	1.00	1.00	1.00	0.73	1.00	1.00	1.00	0.83	1.00	1.00	1.00	1.00	1.00	1.00	0.86	1.00
-8	0.00	0.00	0.00	0.00	0.00	0.27	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.00
Pgd3-4	1.00	1.00	1.00	0.83	1.00	0.98	1.00	1.00	0.96	1.00	1.00	0.96	1.00	0.94	0.90	0.65	1.00	1.00
-6	0.00	0.00	0.00	0.17	0.00	0.02	0.00	0.00	0.04	0.00	0.00	0.04	0.00	0.06	0.11	0.35	0.00	0.00
Pgi3-4	1.00	1.00	1.00	1.00	1.00	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
-6	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pgm2-1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-2	0.00	0.00	0.20	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.00
-4	0.66	0.67	0.20	0.33	0.00	0.11	0.15	0.29	0.19	0.17	0.40	0.94	0.74	0.69	0.11	0.60	0.03	0.67
-8	0.34	0.33	0.60	0.67	1.00	0.86	0.85	0.71	0.81	0.83	0.60	0.06	0.26	0.31	0.90	0.40	0.83	0.33
Pgm6-1	0.06	0.67	0.00	0.50	0.29	0.30	0.71	0.71	0.89	0.25	0.60	0.03	0.09	0.06	0.68	0.20	0.40	0.17
-2	0.13	0.00	0.30	0.17	0.07	0.30	0.07	0.21	0.00	0.33	0.00	0.88	0.59	0.53	0.00	0.30	0.03	0.26
-3	0.81	0.33	0.70	0.00	0.14	0.18	0.10	0.00	0.08	0.42	0.00	0.03	0.28	0.41	0.11	0.25	0.05	0.57
-4	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00
-8	0.00	0.00	0.00	0.33	0.50	0.22	0.10	0.08	0.04	0.00	0.40	0.00	0.04	0.00	0.21	0.25	0.52	0.00
Pgm7-1	0.13	0.00	0.10	0.00	0.64	0.03	0.56	0.42	0.39	0.00	0.80	0.00	0.04	0.00	0.08	0.05	0.12	0.00
-2	0.88	1.00	0.90	1.00	0.21	0.97	0.35	0.50	0.58	1.00	0.20	0.83	0.87	0.94	0.92	0.85	0.88	1.00
-n	0.00	0.00	0.00	0.00	0.14	0.00	0.09	0.08	0.04	0.00	0.00	0.17	0.09	0.06	0.00	0.10	0.00	0.00
Skd1-3	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-4	0.94	1.00	1.00	0.75	0.86	1.00	0.96	1.00	1.00	0.67	1.00	0.99	1.00	1.00	0.95	1.00	0.83	0.94
-5	0.06	0.00	0.00	0.08	0.00	0.00	0.03	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.05	0.00	0.12	0.07
-6	0.00	0.00	0.00	0.00	0.14	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00
-n	0.00	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	0.00	0.00	0.03	0.00
Tpi3-4	0.88	0.67	0.80	0.50	0.14	0.56	0.21	0.13	0.58	1.00	0.40	0.97	0.96	0.94	0.18	0.50	0.33	0.98
-6	0.13	0.33	0.20	0.50	0.86	0.44	0.79	0.88	0.42	0.00	0.60	0.03	0.04	0.06	0.82	0.50	0.64	0.02
-n	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
Tpi4-4	0.00	0.00	0.10	0.00	0.00	0.00	0.03	0.08	0.04	0.00	0.00	0.70	0.52	0.38	0.05	0.20	0.00	0.15
-6	1.00	1.00	0.90	1.00	1.00	1.00	0.97	0.92	0.96	1.00	1.00	0.31	0.48	0.63	0.95	0.80	1.00	0.85
Tpi7-4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00
-6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00

[a] Geographic regions number 1 to 34 are illustrated in Fig. 1; accessions studied from each region (total = "*N*" in this table) are listed in Table 1. The allele symbol "n" designates "null" variants. Two estimates of mean allele frequencies are shown: Mean1 = unweighted arithmetic mean across regions; Mean2 = weighted arithmetic mean across regions, where the weights are equal to the number of accessions analyzed.

phosphoglucose isomerase (PGI), shikimate dehydrogenase (SKD), and both arginyl-specific (ARG) and leucyl-specific (LEU) forms of aminopeptidase. Isozyme and allozyme nomenclature follow Percy and Wendel (1990).

Data analysis—Standard statistics for characterizing genetic variability were computed for all accessions and various groups of accessions, including the proportion of polymorphic loci (*P*), the mean number of alleles per locus (*A*), and mean panmictic (= expected) heterozygosity ($H = 1 - \Sigma (p_i)^2$, where the p_i's represent allele frequencies). Multivariate relationships among accessions were revealed with principal component analysis using a variance-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), principal component analysis, and in apportioning genetic variation among regions (Nei, 1987). This latter technique partitions total variation (H_T) into within-group and among-group 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. Homogeneity of gene frequencies among regions was tested by contingency chi-square analysis (Workman and Nieswander, 1970). Genetic distance and identity statistics (*D* and *I*) were computed following Nei (1978) and Rogers (1972). Many of the above computations were expedited by the microcomputer programs BIOSYS (D. Swofford, Illinois Natural History Survey, Champaign, IL) and NTSYS (Exeter Publishing Ltd., Setauket, NY).

RESULTS

Genetic diversity—Genetic interpretations of isozyme phenotypes were based on the quaternary structure of each

TABLE 2. *Continued*

19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34		
28	7	32	10	10	6	5	19	3	25	4	23	5	5	4	50	Mean1	Mean2
0.18	0.50	0.26	0.20	0.10	0.00	0.20	0.74	0.67	0.48	0.75	0.33	0.20	0.10	0.00	0.01	0.320	0.344
0.73	0.21	0.71	0.80	0.30	0.50	0.00	0.16	0.00	0.12	0.13	0.00	0.20	0.50	1.00	0.99	0.372	0.432
0.09	0.29	0.03	0.00	0.60	0.50	0.80	0.11	0.33	0.40	0.13	0.67	0.60	0.40	0.00	0.00	0.307	0.225
1.00	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.80	1.00	1.00	0.98	0.978	0.984
0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.02	0.021	0.016
0.09	0.43	0.68	0.50	0.50	0.83	0.20	0.84	0.67	0.64	0.75	0.15	0.40	0.40	0.00	0.12	0.414	0.404
0.91	0.57	0.32	0.50	0.50	0.17	0.80	0.16	0.33	0.36	0.25	0.85	0.60	0.60	1.00	0.88	0.585	0.596
0.95	1.00	1.00	1.00	0.70	0.67	1.00	1.00	1.00	0.92	1.00	1.00	1.00	0.80	1.00	1.00	0.954	0.957
0.05	0.00	0.00	0.00	0.30	0.33	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.20	0.00	0.00	0.045	0.043
1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.50	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.947	0.959
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.052	0.041
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.998	0.996
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.001	0.004
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.004	0.001
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.014	0.014
0.18	1.00	0.40	0.70	0.30	0.17	0.40	0.21	0.00	0.04	0.25	0.00	0.80	0.50	0.00	0.91	0.376	0.411
0.82	0.00	0.60	0.30	0.70	0.83	0.60	0.79	0.83	0.96	0.63	1.00	0.20	0.50	1.00	0.09	0.603	0.574
0.07	0.21	0.08	0.00	0.65	0.67	1.00	0.08	0.00	0.58	0.00	0.00	0.60	0.50	1.00	0.00	0.356	0.275
0.16	0.14	0.68	0.70	0.30	0.33	0.00	0.68	0.00	0.20	1.00	0.00	0.00	0.00	0.00	0.94	0.273	0.366
0.75	0.64	0.24	0.30	0.05	0.00	0.00	0.11	0.00	0.14	0.00	0.07	0.40	0.20	0.00	0.06	0.214	0.219
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.006	0.007
0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.11	1.00	0.08	0.00	0.94	0.00	0.20	0.00	0.00	0.148	0.132
0.05	0.00	0.00	0.00	0.35	0.58	1.00	0.00	0.00	0.36	0.25	0.83	0.60	0.20	1.00	0.00	0.254	0.182
0.95	1.00	0.97	1.00	0.65	0.42	0.00	1.00	1.00	0.62	0.75	0.13	0.40	0.60	0.00	1.00	0.717	0.783
0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.04	0.00	0.20	0.00	0.00	0.028	0.035
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.004	0.002
0.96	1.00	0.90	1.00	1.00	0.83	1.00	0.95	1.00	1.00	1.00	0.91	1.00	0.80	1.00	0.93	0.949	0.947
0.04	0.00	0.07	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.20	0.00	0.07	0.039	0.039
0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.004	0.007
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.002	0.005
0.79	1.00	1.00	0.90	0.50	0.50	0.20	0.53	0.00	0.50	0.75	0.04	0.40	0.60	0.00	1.00	0.566	0.641
0.07	0.00	0.00	0.10	0.50	0.50	0.80	0.47	1.00	0.50	0.25	0.96	0.60	0.40	1.00	0.00	0.427	0.350
0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.005	0.009
0.07	0.29	0.57	0.40	0.00	0.00	0.00	0.34	0.00	0.12	0.75	0.00	0.00	0.00	0.00	0.36	0.151	0.205
0.93	0.71	0.44	0.60	1.00	1.00	1.00	0.66	1.00	0.88	0.25	1.00	1.00	1.00	1.00	0.64	0.848	0.795
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.999	0.999
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.001	0.001

enzyme, typical patterns of subcellular localization and expression in other plants (reviewed in Weeden and Wendel, 1989), and from formal genetic analyses of numerous interspecific and intraspecific F_2 and BC progenies (Wendel, unpublished data). This evidence suggests that the 14 enzymes scored are encoded by at least 50 genes. This estimate is a minimum, because poorly resolved isozymes are not counted and because the products of some loci may have been obscured by co-migration with isozymes encoded by other loci. Twenty loci were fixed for the same allele in all accessions examined: *Aat3, Aco2, Aco4, Mdh1, Mdh2, Mdh3, Mdh5, Nad2, Pgd4, Pgd5, Pgd6, Pgi2, Pgm1, Pgm3, Pgm4, Pgm5, Tpi1, Tpi2, Tpi5, Tpi6.* At least one locus was variable for all 14 enzymes assayed, yielding a total of 30 polymorphic loci (P = 60% for the species). Nine polymorphic loci were biallelic, nine were triallelic, and the remaining 12 were multiallelic, resulting in a total of 95 allelic variants. This number includes the unusual observation of putative "null" alleles at 16 different loci. Although most of these have not been confirmed through formal allelism tests, all were scored as bona fide allelic variants because: 1) null alleles are often detected in tetraploid cotton (Percy and Wendel, 1990; Wendel and Percy, 1990), consistent with its high chromosome number and consequent "excess" of coding loci for most enzyme systems (see Weeden and Wendel, 1989); 2) all null alleles detected were phenotypically verified in additional electrophoretic runs; 3) all nulls tested in genetic crosses behave in accordance with Mendelian expectations (unpublished data). Mean allelic frequencies for most nulls were low (<0.05) to very low (<0.01), the exceptions being *Mdh6-n* (0.225) and *Arg1-n* (0.059). Including the 20 monomorphic loci, 115 alleles were detected in *G. hirsutum,* yielding 2.30 as the estimate of the mean number of alleles per locus.

As expected, individuals within accessions were fixed for the same allele at most loci, so that a single multilocus genotype was often sufficient to describe an accession's allelic profile. Generally, wherever within-accession variability was encountered, it consisted of alternate homo-

TABLE 3. *Genetic differentiation*[a] *among accessions from 34 geographic regions of* Gossypium hirsutum *for 30 polymorphic allozyme loci*

Locus	H_T	D_{ST}	G_{ST}	P_1	P_2	P_3
Aat1	0.019	0.002	0.090	0.00	0.49	0.00
Aat2	0.521	0.152	0.292	0.00	0.00	0.00
Aat4	0.150	0.039	0.263	0.00	0.00	0.00
Aco1	0.014	0.000	0.000	0.03	0.00	0.97
Aco3	0.498	0.144	0.290	0.00	0.00	0.00
Aco5	0.489	0.122	0.250	0.00	0.00	0.00
Adh1	0.039	0.002	0.054	0.00	0.00	0.25
Adh2	0.114	0.027	0.233	0.00	0.00	0.00
Arg1	0.508	0.189	0.373	0.00	0.00	0.00
Arg2	0.256	0.038	0.149	0.00	0.00	0.00
Enp1	0.116	0.031	0.263	0.00	0.00	0.00
Enp2	0.028	0.002	0.065	0.00	0.00	0.00
Gdh1	0.436	0.206	0.472	0.00	0.00	0.00
Idh1	0.538	0.311	0.577	0.00	0.00	0.00
Idh2	0.119	0.017	0.141	0.00	0.00	0.00
Leu1	0.363	0.139	0.382	0.00	0.00	0.00
Mdh4	0.080	0.021	0.268	0.00	0.00	0.00
Mdh6	0.664	0.204	0.307	0.00	0.00	0.00
Nad1	0.041	0.005	0.130	0.00	0.72	0.00
Pgd1	0.485	0.121	0.250	0.00	0.00	0.00
Pgd2	0.086	0.014	0.160	0.00	0.00	0.00
Pgd3	0.099	0.029	0.295	0.00	0.00	0.00
Pgi3	0.004	0.000	0.000	0.00	—	0.06
Pgm2	0.493	0.157	0.318	0.00	0.00	0.00
Pgm6	0.730	0.286	0.392	0.00	0.00	0.00
Pgm7	0.420	0.187	0.445	0.00	0.00	0.00
Skd1	0.098	0.008	0.087	0.00	0.00	0.15
Tpi3	0.496	0.204	0.412	0.00	0.00	0.00
Tpi4	0.257	0.088	0.341	0.00	0.00	0.04
Tpi7	0.001	0.000	0.000	0.98	—	0.97
Mean[b]	0.163	0.055	0.336			

[a] Total variation (H_T) is partitioned into within-group and among-group 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 (Nei, 1987). H_T, D_{ST}, and G_{ST} were estimated using the Wright78 implementation of BIOSYS. The final three columns present probability values from chi-square tests of gene frequency homogeneity among regions under the null hypothesis of no genetic differentiation: P_1 is among all 34 geographic regions (Table 1; Fig. 1); P_2 is among mainland Meso-American regions (numbers 3, 12–14, 16, 18–22, 26, 27, and 29); P_3 is among Caribbean island regions (numbers 1, 2, 4–11, 15, 17, 23–25, 28, and 30–33).

[b] Includes 20 monomorphic loci. Mean G_{ST} = mean D_{ST}/mean H_T.

zygotes at one to several loci; consequently, observed heterozygosity was very low.

Allele frequencies averaged across all accessions are given in Table 2 (as "Mean 1"). Exactly half of the polymorphic loci were only slightly variable, in that the frequency of the most common allele exceeded 0.90; these loci have correspondingly low estimates of panmictic heterozygosity (<0.150; H_T of Table 3). Allele frequencies were more equitable for the other 15 polymorphic loci (Table 2), with 12 exhibiting heterozygosity estimates in excess of 0.400 (Table 3). Averaged across polymorphic loci, mean panmictic heterozygosity in *G. hirsutum* is 0.272; with the 20 monomorphic loci, it drops to 0.163.

Geographical patterns of diversity—Inspection of the allele frequency data for the 538 accessions suggested that many of the 95 alleles detected at the 30 polymorphic loci were distributed in a nonrandom fashion over the range of the species. Because of the size and complexity of the data set, an initial principal component analysis was performed on the variance-covariance matrix of the entire sample of allele frequencies (dimensions = 538 × 95). Accessions were projected onto a plane defined by the first two principal components, which accounted for 27.2% and 9.2% of the total variance, respectively. This analysis revealed that accessions tended to cluster in multivariate space with other accessions from the same geographical area (data not presented). This observation provided the rationale for grouping accessions into 34 geographical regions (Table 1; Fig. 1). These regions varied widely in the number of included accessions, which ranged from three, for both Barbados and Panama, to 50, for Upland cultivars. Mainland Mesoamerican accessions were grouped into 13 regions, whereas 20 clusters of accessions were recognized from various combinations of Caribbean islands. Five of the 538 accessions (TX-1, TX-9, TX-11, TX-2083, and TX-2089) were omitted from regional groupings because of questionable locality information.

Allele frequencies for each of the 34 geographical regions were estimated as unweighted arithmetic means of accession allele frequencies (Table 2). These data indicate that alleles differ widely in their distribution and frequency among regions (Tables 2, 4). Twelve alleles are restricted to single geographical regions: for example, *Aat2-8* was detected only in accessions from Aruba, and *Enp2-2* was detected only in accessions from Nicaragua. Twenty alleles, all relatively high frequency variants, were detected in all 34 regions (Table 4). The remaining 63 alleles occur in more than one but not all regions. Inspection of Table 4 reveals many distribution patterns, with the following noteworthy examples: 1) *Aat4-6* was detected only in accessions from one of three Guatemalan regions (#12 of Fig. 1 = Departments of Jutiapa and Chiquimula) and from Upland cultivars; 2) Many alleles were restricted in their distribution to several or more islands in the Caribbean (e.g., *Aco5-3, Arg2-6, Enp1-5, Idh2-5*), whereas far fewer alleles were confined to some portion of the species' mainland range (e.g., *Aco1-n*); 3) Most common were alleles that were variously distributed in portions of both mainland Mesoamerica and the Caribbean (Table 4); 4) Every allele detected in the 50 Upland cultivars was also detected from some portion of the indigenous range of *G. hirsutum.*

Each geographical region differed not only in its allelic array (Table 4) and allele frequencies (Table 2), but also in amount of genetic diversity (Table 5). Although accessions from individual regions varied widely in genetic diversity (a portion of which may result from limited sampling in some regions), few clear geographical patterns emerged. Mean estimates for all variability measures (*A, P, H* and total number of alleles) were nearly identical for accessions from mainland Mesoamerica and the Caribbean; consequently, estimates of the mean amount of genetic diversity in regions from both of these areas approximates the mean among all 34 regions (Table 5). In contrast to suggestions from the literature (Hutchinson, 1951, 1959; Stephens, 1967; Lee, 1984), these data do not suggest a center of exceptional genetic diversity for *G. hirsutum* in southern Mexico and Guatemala. The data do, however, highlight the relatively low level of genetic diversity contained in Upland cultivars: more accessions were included in this "region" than in any other, yet

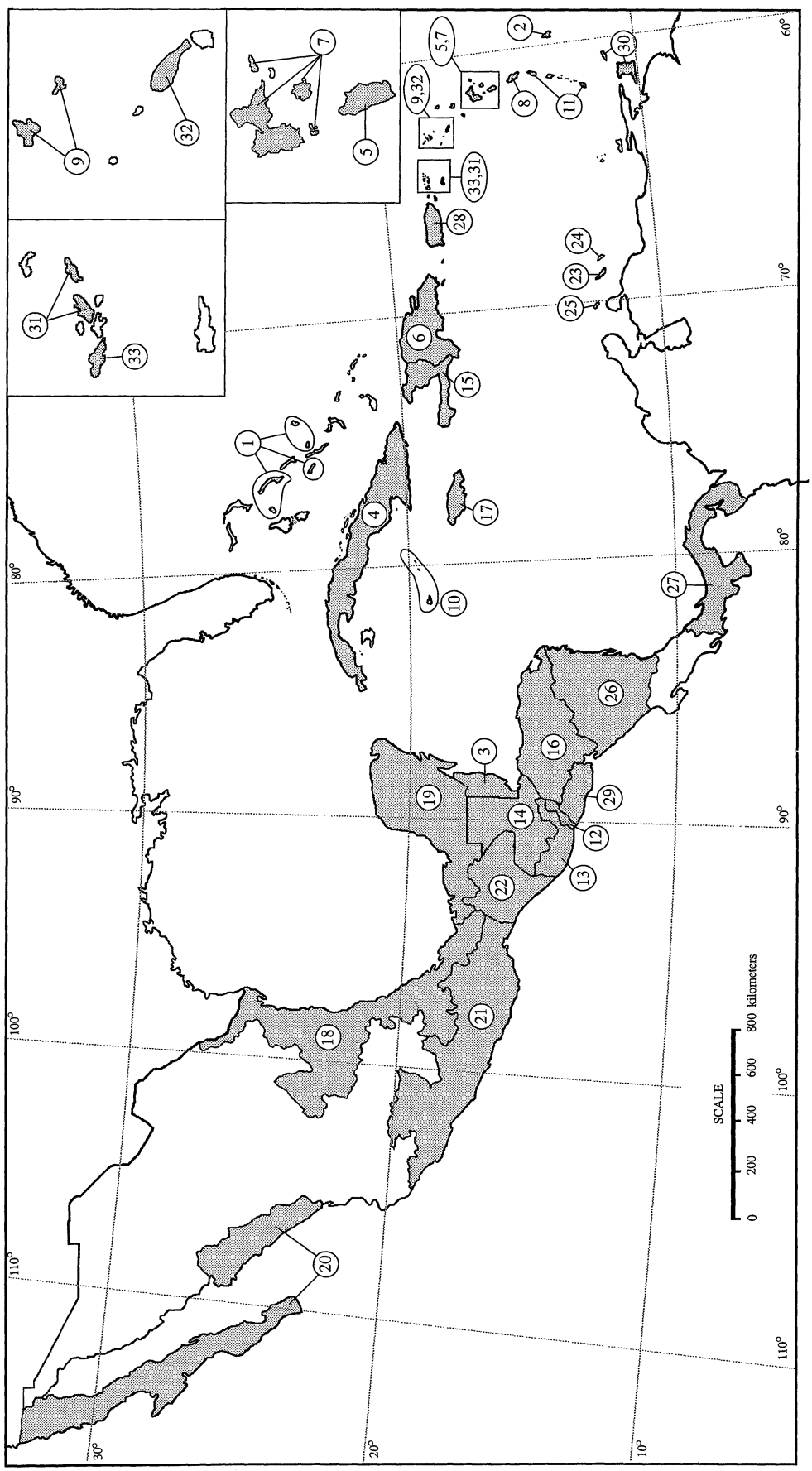

Fig. 1. Map of Mesoamerica and the Caribbean showing sampled areas for 538 accessions of *Gossypium hirsutum*. Boxed regions in main map are magnified in insets (upper right). Numbers correspond to regional groups in Table 1.

TABLE 4. *Geographic distribution[a] of alleles at 30 polymorphic allozyme loci in* Gossypium hirsutum

Locus	Allele	Regions of occurrence	Locus	Allele	Regions of occurrence
Aat1	2	6, 25–26	Idh2	4	1–34
	4	1–34		5	6, 32
Aat2	1	1–4, 6, 8–9, 12–14, 17–19, 21–22, 28		n	2, 4, 6–7, 9, 11, 17, 19, 23, 25, 28
	4	1–34	Leu1	2	1, 3–4, 12–13, 21–22, 26
	6	1–4, 6–23, 25–26, 28–30, 32, 34		4	1–34
	8	25		5	1, 3–4, 6–7, 11–16, 18–22, 25–26, 28, 30, 32, 34
Aat4	2	1–3, 6, 8–9, 12–14, 16, 18–19, 21–22, 28, 30		n	21
	4	1–34	Mdh4	4	1–34
	6	12, 34		6	1, 4, 8, 10, 18–19, 28, 30, 34
	n	26	Mdh6	4	1, 3–10, 12–23, 25–32, 34
Aco1	4	17		6	1–24, 26, 28, 29, 31–34
	8	1–34		n	1, 3–9, 13–17, 19–21, 23–32
	n	14, 16, 26	Nad1	4	1–34
Aco3	1	1–24, 26–32, 34		n	4, 7, 9, 21, 31, 34
	3	1–19, 21–23, 25–34	Pgd1	1	1, 3–10, 12–32, 34
	n	1, 10, 17, 19		3	1–34
Aco5	2	24	Pgd2	4	1–34
	3	6, 23, 24		8	6, 10, 17, 19, 23–24, 28, 32
	4	1, 3–34	Pgd3	4	1–34
	6	1–9, 11–23, 26–29, 32–34		6	4, 6, 9, 12, 14–16, 26–27
Adh1	2	6, 8, 15, 17, 24	Pgi3	4	1–34
	4	1–34		6	6
	6	20	Pgm2	1	27
	n	1, 30		2	3, 6, 17, 29
Adh2	4	1–34		4	1–4, 6–26, 28–29, 31–32, 34
	6	4, 6–7, 9, 13, 15–18, 20, 25–26		8	1–19, 21–34
	n	1, 30	Pgm6	1	1–2, 4–21, 23–26, 28, 31–33
Arg1	3	1, 21		2	1, 3–8, 10, 12–14, 16–24, 26, 28–29, 34
	4	1–30, 32, 34		3	1–3, 5–7, 9–10, 12–23, 26, 28, 30–32, 34
	5	1–11, 13–15, 17–19, 23–25, 28–33		4	7, 12, 26, 32
	n	1, 3–4, 6, 14–18, 21, 23–24, 26, 28		8	4–9, 11, 13, 15–17, 19, 26–28, 30, 32
Arg2	1	1, 5–14, 17–19, 21–25, 28–29, 32, 34	Pgm7	1	1, 3, 5–9, 11, 13, 15–17, 19, 23–25, 28–33
	2	1–34		2	1–24, 26–32, 34
	6	3, 17, 23		n	5, 7–9, 12–14, 16, 21, 28, 30, 32
	n	6, 23	Skd1	3	4
Enp1	3	1, 8, 12–13, 16, 18, 20–21, 32		4	1–34
	4	1–34		5	1, 4, 7, 10, 15, 17–19, 21, 24, 30, 32, 34
	5	6, 15		6	5, 7, 17, 21, 30
	n	28		n	12, 17, 26
Enp2	2	26	Tpi3	4	1–26, 28–32, 34
	3	17, 22, 24, 26		6	1–9, 11–19, 22–23
	4	1–34		n	17, 19
Gdh1	1	7, 32	Tpi4	4	3, 7–9, 12–16, 18–22, 26, 28–29, 34
	2	1–30, 32, 34		6	1–34
	6	1, 3, 5–13, 15–19, 23–26, 28, 30–33	Tpi7	4	1–34
	n	23, 26, 28		6	17
Idh1	2	1–4, 6–10, 12–22, 26, 28–29, 32, 34			
	4	6, 17, 19, 23–24, 28, 32			
	8	1–9, 11, 13–19, 23–33			

[a] Geographic regions and included accessions are listed in Table 1 and illustrated in Fig. 1.

relatively few alleles were detected (64), and panmictic heterozygosity (0.056) is only about half of the mean across all regions (0.109).

The data of Tables 2 and 4 also demonstrate that gene frequencies at most polymorphic loci are heterogeneous among regions. Statistical analysis (χ^2 contingency tests of homogeneity) demonstrates that this is the case (at the 0.05 level) for 28 of 30 polymorphic loci (all but the weakly polymorphic loci *Aco1* and *Tpi7*; P_1 of Table 3). Gene frequencies at most polymorphic loci are also heterogeneous when only mainland or Caribbean regions are included in the analysis (P_2 and P_3 of Table 3, respectively). Effects of this regional divergence on apportionment of genetic diversity were quantified by the gene diversity statistics of Nei (1987). The proportion of total genetic variation due to gene frequency differences among regions (G_{ST}) ranged from 0% for *Pgi3* and *Tpi7* (equivalent gene frequencies in all regions) to 58% for *Idh1* (Table 3). Averaged across the 30 polymorphic loci, one-third (33.6%) of the total genetic variation in *G. hirsutum* arises from regional differentiation.

A second principal component analysis was performed on the variance-covariance matrix of regional gene frequencies (dimensions = 34 × 95). Regional clusters were plotted along the first two components (Fig. 2), which resulted in a plot with a distinctive geographical interpretation. The first principal component, which accounts for nearly half (46.9%) of the total variance, separates most mainland Mesoamerican accessions (high negative PCA1 scores) from most Caribbean accessions (high pos-

TABLE 5. *Genetic diversity*[a] *by region in* Gossypium hirsutum *for 50 allozyme loci*

Region[b]	*N*	*A*	*P*	*H*	Total alleles
1. Bahamas	16	1.54	40	0.132	77
2. Barbados	3	1.20	18	0.101	60
3. Belize	10	1.42	32	0.121	71
4. Cuba	6	1.46	34	0.160	73
5. Dominica	7	1.30	24	0.092	65
6. Dominican Republic	30	1.70	46	0.147	85
7. French West Indies I	34	1.52	38	0.111	76
8. French West Indies II	12	1.44	36	0.112	72
9. French West Indies III	13	1.46	38	0.120	73
10. Cayman Islands	6	1.28	24	0.092	64
11. Grenada/St. Lucia	5	1.24	24	0.106	62
12. Guatemala I	36	1.46	34	0.082	72
13. Guatemala II	23	1.48	36	0.115	74
14. Guatemala III	16	1.42	34	0.125	71
15. Haiti	19	1.44	38	0.108	72
16. Honduras	9	1.46	38	0.145	73
17. Jamaica	29	1.66	42	0.140	83
18. Mexico I	23	1.44	38	0.117	72
19. Mexico II	28	1.54	40	0.094	77
20. Mexico III	7	1.24	20	0.084	62
21. Mexico IV	32	1.46	32	0.110	73
22. Mexico V	10	1.30	26	0.104	65
23. Curaçao	10	1.44	30	0.143	72
24. Bonaire	6	1.34	30	0.132	67
25. Aruba	5	1.26	24	0.095	63
26. Nicaragua	19	1.54	40	0.139	77
27. Panama	3	1.12	12	0.061	56
28. Puerto Rico	25	1.56	40	0.136	78
29. El Salvador	4	1.28	24	0.113	64
30. Trinidad & Tobago	23	1.34	30	0.050	67
31. UK Virgin Islands	5	1.18	16	0.080	59
32. St. Kitts	5	1.48	36	0.173	74
33. USA Virgin Islands	4	1.02	2	0.011	51
34. Upland Cultivars	50	1.28	28	0.056	64
Mean of 34 regions	15.7	1.39	31	0.109	70
Over all accessions	533	2.30	60	0.163	115
Mean of mainland[c]	16.9	1.40	31	0.108	70
Mainland overall[c]	220	2.00	56	0.136	100
Mean of Caribbean[d]	13.2	1.39	31	0.110	69
Caribbean overall[d]	263	2.20	60	0.148	110

[a] All estimates include 20 monomorphic loci. *N* = number of accessions; *A* = mean number of alleles per locus; *P* = percent polymorphic loci; *H* = mean panmictic heterozygosity.

[b] Accessions included in each geographic region are listed in Table 1.

[c] Includes accessions from regions 3, 12–14, 16, 18–22, 26, 27, and 29. Arithmetic means of estimates for the included regions are shown ("mean of mainland"), as well as estimates derived by treating all mainland Meso-America accessions as a single population ("mainland overall").

[d] Includes accessions from regions 1, 2, 4–11, 15, 17, 23–25, 28, and 30–33. Arithmetic means of estimates for the included regions are shown ("mean of Caribbean"), as well as estimates derived by treating all Caribbean accessions as a single population ("mean of Caribbean").

itive PCA1 scores); PCA1 therefore represents an approximate east-west axis. As expected, the loci and alleles with the highest eigenvector loadings for PCA1 (*Aat2, Aco3, Gdh1, Idh1, Mdh6, Pgm2, Pgm6, Pgm7, Tpi3*) are the same as those that maximally differentiate accessions from these two areas on the basis of allele frequencies (Table 2). No clear interpretation is evident for PCA2, which accounts for an additional 10.6% of the variance.

Multivariate relationships among regional groupings of accessions were also explored using average linkage cluster analysis (UPGMA). A phenogram produced using Rogers' genetic distance is presented in Fig. 3 (cophenetic correlation = 0.77; % standard deviation = 21.5). Relationships suggested by cluster analysis are largely consistent with those inferred from PCA.

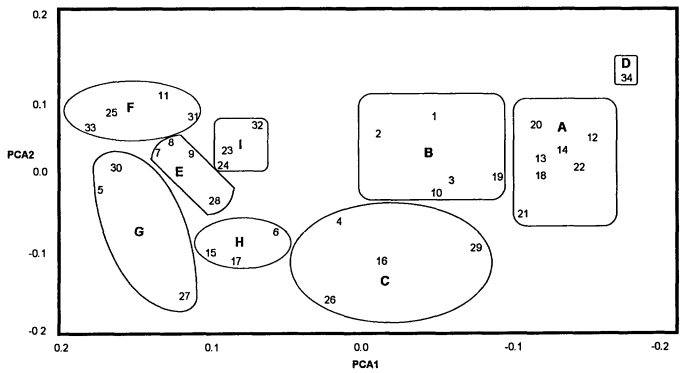

Fig. 2. Principal component analysis of *Gossypium hirsutum* accessions from 34 geographic regions (see Fig. 1; Table 1), based on a variance-covariance matrix of regional gene frequencies at 30 polymorphic allozyme loci. The first two axes account for 46.9% and 10.6% of the total variance, respectively. More inclusive geographical "areas" (see Fig. 3) are indicated.

Overall similarity among the 34 regional groupings of accessions was summarized by the unbiased genetic identity coefficient (*I*) of Nei (1987). Among the 561 pairwise comparisons (data not presented), estimates range from 0.779 (between Upland cultivars and accessions from St. Thomas) to near identity (for many region-pairs). These quantitative data parallel the depictions of relationships revealed by both principal component and cluster analysis. Estimates of *I* involving Upland cotton are all below 0.950, except in comparisons with Guatemalan and Mexican groups of accessions, which ranged from 0.933 to 0.989.

Principal component and cluster analysis results provided the rationale for a secondary grouping of accessions, namely, merging closely related regions into more inclusive geographical "areas," as illustrated in Fig. 3 ("A" through "I"). Although detail is lost in this process, it was motivated by an attempt to clarify overall geographical patterns. Similarity between these areas is summarized by the genetic identity estimates of Table 6, which lists the mean and range of *I* for all interregional comparisons between areas. Most diagonal estimates (intra-area means) are considerably higher than off-diagonal estimates, as expected if grouping into geographical areas was justified. As anticipated from results already discussed, some areas (e.g., area "B") were more heterogeneous than others, as indicated by the range of estimates of *I*. Table 6 further quantifies what is evident from cluster analysis, i.e., that areas A–D and E–I represent two relatively distinct "super-clusters": interarea means within each of these two super-clusters are 0.950 and 0.963, respectively; between the two the mean *I* is 0.904.

DISCUSSION

Genetic diversity in Gossypium hirsutum—Allozyme variability in crop species and their wild progenitors was reviewed by Doebley (1989). Although allozyme surveys have varied widely in important experimental details, such as number of loci and populations (or accessions)

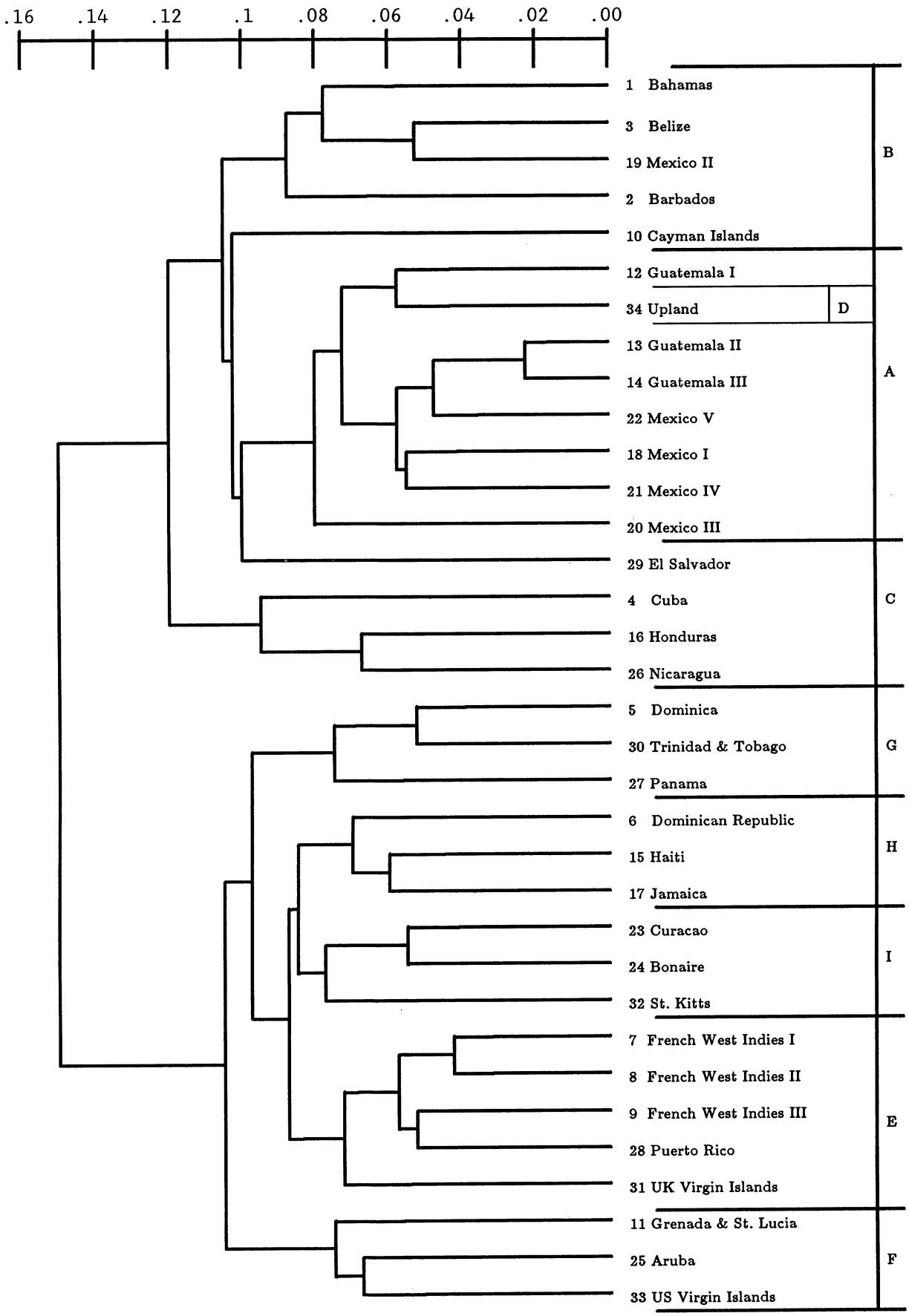

TABLE 6. *Genetic identity estimates*[a] *for* Gossypium hirustum *from different geographical areas*[b]

	A	B	C	D	E	F	G	H	I
A	0.98 (0.96–0.99)								
B	0.96 (0.92–0.99)	0.97 (0.94–0.99)							
C	0.96 (0.92–0.98)	0.95 (0.92–0.98)	0.97 (0.94–0.99)						
D	0.97 (0.96–0.99)	0.94 (0.93–0.95)	0.93 (0.90–0.94)	—					
E	0.91 (0.86–0.94)	0.93 (0.87–0.96)	0.94 (0.89–0.97)	0.86 (0.84–0.88)	0.99 (0.97–0.99)				
F	0.86 (0.80–0.90)	0.88 (0.84–0.93)	0.90 (0.85–0.95)	0.82 (0.78–0.86)	0.97 (0.95–0.99)	0.98 (0.97–0.99)			
G	0.89 (0.86–0.91)	0.90 (0.88–0.93)	0.93 (0.90–0.97)	0.84 (0.83–0.85)	0.96 (0.91–0.99)	0.95 (0.90–0.99)	0.98 (0.97–0.99)		
H	0.93 (0.88–0.96)	0.94 (0.92–0.97)	0.96 (0.94–0.97)	0.88 (0.87–0.90)	0.97 (0.95–0.99)	0.94 (0.92–0.96)	0.97 (0.95–0.98)	0.99 (0.98–0.99)	
I	0.93 (0.89–0.97)	0.94 (0.90–0.98)	0.95 (0.92–0.97)	0.89 (0.87–0.92)	0.98 (0.96–0.99)	0.95 (0.93–0.97)	0.95 (0.93–0.98)	0.98 (0.96–0.99)	0.99 (0.99–0.99)

[a] Means and ranges (in parentheses) of Nei's unbiased genetic identity (Nei, 1987).

[b] Geographical areas represent combinations of the 34 regions defined in Table 1 and illustrated in Fig. 1, as follows: A—12, 13, 14, 18, 20, 21, 22; B—1, 2, 3, 10, 19; C—4, 16, 26, 29; D—34; E—7, 8, 9, 28, 31; F—11, 25, 33; G—5, 27, 30; H—6, 15, 17; I—23, 24, 32. See text for justification of division into geographic areas.

sampled, these data provide a useful comparative framework. Comparisons indicate that overall levels of diversity in *G. hirsutum* are unexceptional, in that it has a slightly larger percentage of polymorphic loci (60%) and more alleles per locus (2.30) than an "average" crop species (P = 49%; A = 2.15), but a slightly lower than average total diversity (H_T = 0.16 vs. 0.19 for other crop species). By all measures, *G. hirsutum* possesses much greater genetic diversity than the other three species of cultivated cotton, i.e., *G. arboreum* (P = 28%; A = 1.38; H_T = 0.07; Wendel, Olson, and Stewart, 1989), *G. herbaceum* (P = 25%; A = 1.20; H_T = 0.04; Wendel, Olson, and Stewart, 1989), and *G. barbadense* (P = 41%; A = 1.69; H_T = 0.06; Percy and Wendel, 1990).

Domestication of and relationships in Gossypium hirsutum—The present genetic structure of *G. hirsutum* has been shaped by both natural forces and several millennia of pre- and post-Columbian human influence. Unraveling the effects of each is a formidable task, given our limited understanding of the species' native range and its patterns of genetic relationships prior to domestication. A central question is whether *G. hirsutum* achieved widespread distribution and regional differentiation as a wild plant prior to domestication, or if it was widely distributed as a perennial semidomesticate by pre-Columbian peoples from a much smaller native range.

Complications arise from the possibility that *G. hirsutum* may have been domesticated more than once, in more than one part of its native range, and at different times. Hutchinson (1951), for example, considered all six cultivated races to have arisen independently in different cultural and geographical foci. In contrast, Fryxell (1979) argues that the original domestication involved a morphological transition that is unlikely to have occurred more than once. Most who have considered the problem believe that Upland cottons were developed from local, semidomesticated *G. hirsutum* race 'latifolium' progenitors in a center of diversity near the Mexican-Guatemalan border (Hutchinson, Silow, and Stephens, 1947; Hutchinson, 1951; Stephens, 1975). An independent domestication for the morphologically distinctive, "Hopi cottons" of the American Southwest is suggested by Lee (1984), who traces these to wild northern Mexican ancestors. Cultivated forms of *G. hirsutum* race 'marie-galante', from the northern part of South America and many Caribbean islands, are suggested to have been derived either from northern Colombian stock or from introgression between West Indian wild forms of *G. hirsutum* and introduced *G. barbadense* (Stephens, 1967; Lee, 1984).

Given the complex cultural history of *G. hirsutum,* it is necessary to consider the types of evidence that can address its time and place of original domestication, i.e., morphological and geographical comparisons with ancestral populations, in conjunction with studies of comparative diversity. A complication arises from the difficulty of distinguishing truly wild populations from those that are secondarily wild (feral) as a consequence of escape from cultivation followed by reestablishment in native vegetation. In fact, it is unclear whether any truly wild *G.*

←

Fig. 3. Average linkage cluster analysis of relationships among *Gossypium hirsutum* accessions from 34 geographic regions (see Fig. 1; Table 1), based on Rogers' genetic distance estimated from allele frequencies at 30 polymorphic allozyme loci. More inclusive geographical "areas" are indicated (see text). Cophenetic correlation = 0.77; standard deviation = 21.5%.

hirsutum populations exist, although naturally occuring wild or feral forms are found in beach strand and other littoral environments in many parts of the species' range (Hutchinson, 1951; Stephens, 1958; Fryxell, 1979). Many of these populations, however, exhibit one or more features suggestive of human selection (e.g., in lint characteristics), implying that they represent feral, self-seeding escapes from some earlier period in domestication (Hutchinson, 1951). This view was held by Stephens, who stated that "No primitive forms of the cultivated species of cotton have so far been recorded and in all probability they no longer exist" (Stephens, 1958, p. 19). Others have viewed at least some of the littoral populations as representing the unimproved, natural condition, especially in race 'yucatanense', which is never cultivated, and in race 'punctatum', which includes many of these putatively wild populations. Fryxell, for example, contends that many of the "outpost shrubs" rimming the Gulf of Mexico represent wild *G. hirsutum* (Fryxell, 1979).

Because he sees no selective advantage for long seed hairs in natural populations, Fryxell (1979) envisions wild *G. hirsutum* as having relatively short, sparse fiber (other traits characteristic of putatively wild populations are small fruits, small seeds, and a low lint percentage). These putatively wild, littoral cottons may resemble the populations originally domesticated by humans. The alternative, i.e., that *G. hirsutum* with a higher lint percentage was originally domesticated, is more problematic, in that a hypothetical ancestor is required with a seed hair morphology that is unknown in a demonstrably wild plant.

If the wild littoral cottons represent descendants of the type of *G. hirsutum* that was originally domesticated, there are few clues to the geographic locality of initial domestication, because such populations are geographically widespread. The ten accessions studied here with lint percentages of 11% or below come from Martinique, Baja California, Michoacán (Mexico), Dominican Republic, Belize, Puerto Rico, Grenada, and Honduras. Other accessions in the National Collection of *Gossypium* Germplasm with lint percentages lower than 12% come from Yucatán (Mexico), Veracruz (Mexico), Colombia, Jamaica, Seychelles, and Brazil.

This lack of geographical coherence in putatively wild cottons is reflected in the allozyme data: the ten accessions included in the present study with the lowest lint percentage ($\leq 11\%$) were neither genetically distinct (no unique alleles), nor genetically coherent as a group. Thus, neither the geographic distribution of low lint percentage populations nor their genetic composition, as determined by allozyme data, provide useful clues to the geographic origin of domesticated *G. hirsutum,* nor to the question of whether it was domesticated more than once.

A second, often useful criterion for determining the origin of domestication is the location of the center of diversity, although it is clear that centers-of-diversity and centers-of-origin are not necessarily geographically congruent (Harlan, 1971, 1975). This criterion helped to clarify the origin and diffusion pathways of *G. barbadense,* for example (Percy and Wendel, 1990). In *G. hirsutum,* however, the allozyme data fail to implicate any particular region as being exceptionally diverse relative to other areas. Both multivariate depictions of relationships (Figs. 2, 3) and diversity estimates (Table 5) clearly indicate that there are two, equally diverse, geographically broad foci of genetic variability in *G. hirsutum,* one in southern Mexico and Guatemala, and the other in the Caribbean. This conclusion contrasts with current opinion, which is that there is a single center of diversity in southern Mexico and Guatemala (Hutchinson, Silow, and Stephens, 1947; Hutchinson, 1951; Stephens, 1975).

Unfortunately, the allozyme data fail to discriminate between alternative scenarios for the geographical location of original domestication. It may be argued that the existence of two centers of diversity reflects two independent domestications and regions of development, one in the Caribbean and the other in southern Mexico and Guatemala. The possibility remains, however, that *G. hirsutum* was domesticated only once, and that the present genetic structure represents the superimposed effects of pre- and post-Columbian human migration and trade. Evaluation of these two alternatives, and indeed others, such as a multiple domestication scenario, would seem to require information that is, at present, unavailable, i.e., geographic and genetic data on demonstrably wild progenitor populations.

Despite the inability of the data to critically address questions concerning domestication events, they do provide relatively detailed information on present genetic structure. That relationships among accessions have been influenced by geographic factors and cultural history is evident in multivariate depictions of infraspecific relationships (Figs. 2, 3). Nearly half of the total variance in principal component analysis is accounted for by PCA1, which is interpreted as a crude east-west axis separating mainland Mesoamerican accessions from Caribbean accessions. Consideration of both PCA axes leads to the recognition of several more-or-less discrete clusters among the 34 regions: 1) Most Mexican and Guatemalan regions form a discrete group with high negative PCA1 scores (area "A," Fig. 2); 2) Upland cultivars are clearly distinguished from all other cottons except those from Mexico and Guatemala, with their closest relationship being to accessions from region #12, the Guatemalan Departments of Jutiapa and Chiquimula (see below); 3) Most regions that are former British colonies (Cayman Islands, Barbados, Belize, Bahamas) cluster together near the center of the plot (area "B," Fig. 2). Inasmuch as these regions are geographically disjunct, the most plausible explanation is that germplasm exchange was widespread during the period of British colonial rule. In contrast, accessions from the Yucatán Peninsula of Mexico (Spanish colonial history) are genetically more similar to those from Belize (British colonial history) than to those from other Mexican regions (Spanish colonial history); this relationship is probably based on geographical proximity rather than colonial history; 4) Most regions with a strong Spanish colonial influence (other than Mexico and Guatemala) form a loose cluster near the center of the plot (El Salvador, Honduras, Nicaragua, Cuba, Dominican Republic, Puerto Rico), with relatively low PCA2 scores (area "C," Fig. 2); 5) Most regions representing islands whose colonial influences were primarily Dutch (Curaçao, Bonaire, Aruba) or French (French West Indies I, II, and III of Table 1), cluster together with high PCA1 scores.

These patterns provide partial support for Stephens' (1967, p. 124) contention that "the variation found in the

cottons of the Caribbean area is a product of European colonization and trade. . . ." Allozyme data, nevertheless, demonstrate some relationships that appear to be based on geographical proximity rather than colonial history, presumably reflecting earlier movements of Amerindian peoples superimposed on the preexisting natural distribution pattern (as is the case with Caribbean maize; Bretting, Goodman, and Stuber, 1987). These complex relationships are exemplified by area "I" of Figs. 2 and 3, which includes Hispaniola and Jamaica. Jamaica is one of the few former British colonies (although under Spanish influence from circa 1493 to the mid 1500s) that did not fall into area "B" in multivariate analyses. Rather, its cottons show a close genetic affinity to those from a former Spanish colony, the Dominican Republic, and to accessions from Haiti, a former French colony that together with the Dominican Republic constitute the neighboring island of Hispaniola. In fact, an allele that is otherwise restricted to Hispaniola (*Enp1-5*) was detected in some Jamaican accessions. Hence, genetic affinities arising from geographical proximity may remain evident in accessions from these regions.

Interspecific introgression from Gossypium barbadense—The allele frequency data of Table 2 may be compared with similar data for *G. barbadense* (Percy and Wendel, 1990) to reveal evidence of interspecific introgression. *Gossypium hirsutum* and *G. barbadense* are fixed, or nearly fixed, for alternate alleles at 11 loci that were included in both Percy and Wendel (1990) and the present study: *Aco1, Aco3, Adh2, Arg2, Gdh1, Idh1, Nad1, Pgd1, Pgm6, Pgm7, Tpi7*. Because *G. hirsutum* and *G. barbadense* are sympatric over much of their ranges in the Caribbean, and co-occur to a limited extent in Central America, it was expected that interspecific introgression would be most frequently detected in accessions from these regions. However, only one of 36 *G. barbadense* accessions from these regions showed evidence of *G. hirsutum* introgression (Percy and Wendel, 1990). In contrast, introgression of *G. hirsutum* germplasm into *G. barbadense* was determined to be widespread in modern commercial stocks, where 22% of accessions examined contained one or more introgressed *G. hirsutum* alleles.

The strongest evidence of *G. barbadense* introgression into *G. hirsutum* involves alleles that are fixed or nearly fixed in the former species and are rare in the latter, i.e., *Aco1-4, Enp1-5, Idh2-5, Nad1-n,* and *Pgm6-4* (Table 2). Additional evidence is supplied by shared, lower-frequency alleles (e.g., *Arg2-6*), although explanations other than introgression, such as shared ancestry or convergence, cannot be ruled out. Based on these alleles, 21 *G. hirsutum* accessions (in addition to two Upland cultivars, as discussed below) were diagnosed as putatively introgressant, and in every case, only a single introgressant locus was involved. The latter observation effectively precludes the possibility that allelic transfer occurred as a consequence of contamination in field nurseries subsequent to collection. The putatively introgressant accessions are predominantly Caribbean (16 of 21): *Aco1-4*—TX1229 (Jamaica); *Arg2-6*—TX794 (Belize); TX2219 (Jamaica); TX 2190 (Curaçao); *Enp1-5*—TX1588 (Haiti); TX1822 (Dominican Republic); *Idh2-5*—TX2025 (St. Kitts); TX2267 and TX2276 (Dominican Republic); *Nad1-n*—TX1779 ('marie-galante'); TX1795 and TX1797 (St. Barthelemy); TX802 (Cuba); TX226 (Mexico); TX1367 (Tortola, UK Virgin Islands); *Pgm6-4*—TX1843 ('marie-galante'); TX712 and TX956 (Nicaragua); TX 2024 (St. Kitts); and TX111 and TX141 (Guatemala). In addition, two of the five putatively introgressant Central American accessions have been classified as race 'marie-galante', suggesting a derivation from introduced Caribbean stocks; a third accession (TX226 from Guerrero, Mexico) is a domesticated accession of race 'latifolium'. Little is known regarding the history of the two putative introgressants from Nicaragua. Additional support for introgression being largely restricted to the Caribbean is offered by the geographical occurrence of *Idh1-4,* which is nearly fixed in *G. barbadense,* is rare in *G. hirsutum* outside of the Caribbean, and is relatively common in accessions from the Dominican Republic, Jamaica, Curacao, Bonaire, Puerto Rico, and St. Kitts (Table 2).

These data indicating potential widespread transfer of *G. barbadense* alleles into *G. hirsutum* in a broad area of sympatry in the Caribbean are not without precedent. Stephens noted both morphological "shifts" of Caribbean *G. hirsutum* toward *G. barbadense* (Stephens, 1967), and a *G. barbadense*-specific anthocyanin allele that was common in *G. hirsutum* race 'marie-galante' and otherwise unknown in *G. hirsutum* (Stephens, 1974). Stephens (1967) listed race 'marie-galante' as an entity that may have actually originated as an introgressant product. Although the morphological, genetic, and allozyme data demonstrate introgression from *G. barbadense* into *G. hirsutum* race 'marie-galante', definitive evidence for the introgressive origin of this race is, at present, lacking.

The interpretation that accessions of Caribbean *G. hirsutum* are often introgressed with *G. barbadense* germplasm leads to the unexpected observation that patterns of interspecific introgression between *G. barbadense* and *G. hirsutum* are not reciprocal: in *G. barbadense,* introgression is rare in areas of sympatry and common in modern cultivars, whereas in *G. hirsutum,* introgression is relatively common in the Caribbean and rare in Upland cotton (two of 50 cultivars). Stephens reached a similar conclusion regarding reciprocal introgression in the Caribbean: "The evidence from gross morphology is consistent with the hypothesis of a 'one-way' introgression from *barbadense* into *hirsutum*. . . ." (Stephens, 1967, p. 126). Percy and Wendel (1990) speculated that the relative infrequency of *G. hirsutum* introgression into sympatric *G. barbadense* reflects the operation of one or more reproductive isolation mechanisms, and that persistence of introgressed alleles in modern cultivars results largely from human selection and maintenance. Results of the present study suggest that: 1) these isolation mechanisms (see Percy and Wendel, 1990 for discussion) are either unidirectional or at least asymmetric in their effectiveness; and 2) retention of *G. barbadense* germplasm in improved cultivars requires the circumvention of relatively strong, negative barriers to gene flow (see additional discussion, below).

Implications of allozyme data for infraspecific classification—Principal component and cluster analyses show that patterns of genetic relationships in *G. hirsutum* have

a strong underlying geographical component (Figs. 2, 3). To this extent, allozyme data provide only limited support for Hutchinson's (1951) division of *G. hirsutum* into seven geographical races. In particular, the primary division of *G. hirsutum* into two foci of development, one centered in the Caribbean and the other in mainland Mesoamerica, corresponds to Hutchinson's separation of the Caribbean race 'marie-galante' from the remaining, primarily mainland races. Beyond this, though, it is questionable whether the existing racial classification system is justified. First, the races are discriminated on the basis of only a few variable and often phenotypically plastic morphological characters (e.g., size, habit, leaf shape). Apart from race 'marie-galante', easily identified by its strongly arborescent habit, and race 'palmeri', distinguished by its deeply laciniate leaves, the races are difficult to classify in the field and nearly impossible to identify from herbarium sheets. Second, although Hutchinson makes a strong case for the usefulness of recognizing seven races, the classification has not found wide applicability. This is underscored by the fact that only 24% of the accessions in the National Collection of *Gossypium* Germplasm have been classified to race, and an even lower percentage of collections from the last couple of decades have been assigned to race. Third, although our sampling of several races was limited, allozyme data failed to provide evidence of either genetic homogeneity within races or discontinuities between them, with the notable exception of race 'marie-galante'; rather, racially classified accessions tended to cluster according to their geographical origins. Fourth, classifying according to race provides a lower level of taxonomic resolution than is provided by the allozyme data. For example, several clusters are evident within the Caribbean accessions, most of which would be classified simply as race 'marie-galante' by their morphology. Clearly, the complicated human and natural history of *G. hirsutum* generated patterns of relationships that are not easily incorporated into neat infraspecific units.

Origin of Upland cotton — The origin and development of the highly productive, cotton-belt cultivars entailed several stages. The earliest period (until approximately 1750) must have involved transforming perennial, subtropical plants into day-length neutral annuals, but this earliest stage in the development of Upland cotton is enshrouded in mystery. The early introductions originated from the Caribbean, Mexico, or Central America, either directly, or indirectly via trans-Atlantic reintroduction of Asian, Mediterranean, or Levantine stocks collected by early European explorers (Ware, 1951; Fryxell, 1968). Historical records, however, fail to provide the detail necessary to specify the ultimate New World geographic origins of both the direct and indirect introductions (Ware, 1951; Ramey, 1966; Niles and Feaster, 1984; Meredith, 1991). Because of this, virtually nothing is known regarding the proportional contribution of germplasm from various potential source areas (Caribbean, Mexico, Central America) into the stocks that represented this earliest stage of Upland cotton development.

Historical records provide considerably more detail about the rise of cotton cultivation in the southeastern United States during the late 18th and early 19th centuries. Commercial-scale plantings did not begin in earnest until the revolutionary war, when two categories of cultivars, "green seed" and "black seed," were predominant. The green-seeded stocks had longer, finer lint, a higher yield and better disease resistance than the black-seeded stocks, but their adherent lint was difficult to gin (Affleck, 1851; Wailes, 1854; Stephens, 1958). The invention of Whitney's saw gin allowed the green-seeded stocks to predominate until the day-length neutral Mexican highland stocks were introduced, beginning in the early 1800s (Wailes, 1854; Ware, 1951; Moore, 1956; Brown and Ware, 1958; Niles and Feaster, 1984). These new Mexican stocks were allowed to introgress with the local green seed and black seed stocks, leading to the development of hybrids that were vastly improved in many agronomic features, including longer and finer fiber, higher yield, shorter growing season, increased disease resistance, and relative ease of harvest (Wailes, 1854; Moore, 1956). This second stage in the development of Upland cotton involved the deliberate introduction and introgression of imported germplasm into U.S. breeding populations, as well as considerable effort directed to the development of locally adapted cultivars over an expanding agricultural range.

Further development of the modern crop involved a series of additional introductions, beginning in the early 1900s, in response to the devastation brought on by the boll weevil (Niles and Feaster, 1984). Circumventing boll weevil-induced crop losses became a priority, so earlier maturing varieties became highly desired (Ware, 1951; Niles and Feaster, 1984), as were cultivars adapted to the specific ecological conditions of the four main cotton growing regions of the United States. Consequently, most modern Upland cultivars can be categorized as one of the following four types: Acala, Plains, Eastern, or Delta (Niles and Feaster, 1984; Meredith, 1991). Acala cultivars, grown primarily in irrigated regions of western Texas, New Mexico, and the San Joaquin Valley of California, have a complicated breeding history involving several introductions. They are based on Mexican stocks from the area surrounding Acala and Tuxtla, Mexico (Chiapas), with subsequent inputs of germplasm from the so-called "triple hybrid" (doubled [*G. thurberi* Todaro × *G. arboreum*] × *G. hirsutum*) and perhaps *G. barbadense* (Ware, 1951; Niles and Feaster, 1984; Meredith, 1991). Delta cottons (e.g., 'Deltapine' and 'Stoneville'), grown in the Mississippi Delta, Arizona, southern California, and elsewhere, are based primarily on pre-1900 cultivars, tracing to the 'Lone Star' group of cultivars which, in turn, can be traced back to a mid-1880 Mexican introduction (Ware, 1951; Ramey, 1966; Niles and Feaster, 1984). The Deltapine series apparently derives from Mexican introductions during the 1860s (Niles and Feaster, 1984). Plains cottons, grown largely in northern Texas and Oklahoma, are predominantly based on 'Big Boll Stormproof' stocks that were collected in Mexico about 1850. One prominent family of cultivars, the 'Paymaster' group, putatively includes 'Kekchi' germplasm collected in Guatemala in 1904 (Ware, 1951; Niles and Feaster, 1984). The last group of modern cultivars, the Eastern group (e.g., the 'Coker' family), is a heterogeneous assemblage of largely unknown pedigree. They are thought to consist primarily of selections from 19th and 20th century Mexican introductions, perhaps with introgressed *G. barbadense* germplasm (Niles and Feaster, 1984).

This abbreviated history suggests that the modern Upland cotton gene pool was derived from a complex admixture of a relatively large number of introductions from a variety of sources. It is widely believed, however, that much of the early germplasm (e.g., "green seed" and "black seed") was supplanted by a relatively limited number of later Mexican introductions during the 19th and 20th centuries. This is reflected in the general belief, previously unsubstantiated by data, that Upland cotton has a relatively narrow germplasm base (Anonymous, 1972; Endrizzi, Turcotte, and Kohel, 1985). The allozyme data provide convincing evidence that this is so. Although a diverse set of 50 Upland cultivars was analyzed, including representatives of all four major cultivar types (Acala, Plains, Eastern, and Delta), all measures of genetic diversity are low relative to the species as a whole: 1) only 14 of the 30 loci that are polymorphic in the species are variable in Upland cotton, and only six of these (*Aco3, Aco5, Arg2, Mdh4, Pgd1, Tpi4*) are polymorphic using the 0.90 criterion (Table 2); 2) allelic diversity is minimal in Upland cotton, as demonstrated by the absence of any triallelic or multiallelic loci and the correspondingly low estimate of mean number of alleles per locus (1.28 vs. 2.30 for the species); 3) mean panmictic heterozygosity in Upland cotton ($H_T = 0.056$) is lower than in all but one other region sampled (Trinidad and Tobago; Table 5) and is only one-third the level in the species as a whole ($H_T = 0.163$). These data demonstrate that a relatively severe genetic bottleneck accompanied the development of modern Upland cultivars, although relative to other crop plants (Doebley, 1989), modest levels of genetic variation remain in commercially important breeding populations.

Allozyme data also provide several insights into the origin and composition of the modern Upland cotton gene pool. First, the allelic composition of the group of 50 Upland cultivars represents a subset of the allelic profile of accessions from southern Mexico and Guatemala. In contrast, there is no evidence of inputs from the Caribbean or from other sources. These data indicate that the genetic base that existed in the United States prior to the 19th century was nearly completely supplanted by later Mexican introductions.

Second, although allozyme data clearly indicate that the modern Upland cotton gene pool is relatively narrow, the data raise uncertainty regarding the original provenance(s) of the contributing stocks. In contrast to historical records, which clearly indicate that a significant proportion of modern Upland germplasm stems from the introduction of stocks from the central plateau of Mexico, allozyme data suggest that Upland cotton resembles accessions from southeastern Guatemala and Chiapas, Mexico more than it does accessions from the central Mexican plateau.

Perhaps this is most readily appreciated by comparing Nei's genetic identity estimates with the results of multivariate analysis. The highest genetic identity for pairwise comparisons involving Upland cultivars is with accessions from Chiapas, Mexico (0.989); comparisons with other Mexican regions ranged from 0.933 to 0.962, while pairwise comparisons with the three Guatemalan regions (#12, 13, 14) resulted in nearly identical genetic identity estimates (0.976–0.980) that are almost as high as the estimate obtained for accessions from Chiapas. Cluster analysis based on Rogers' genetic distance shows Upland cotton to be embedded within a phenetic grouping consisting of accessions from the three Guatemalan regions and those from southern and western Mexico (Fig. 3), with Upland cottons resembling most closely accessions from the Guatemalan Departments of Jutiapa and Chiquimula (region #12 of Table 1 and Fig. 1). Furthermore, the allele *Aat4-6* was detected only in accessions from Jutiapa and Chiquimula and in Upland cultivars. Principal component analysis resulted in a similar depiction of genetic relationships (Fig. 2).

A possible explanation for the discrepancy between breeding history and multivariate depictions based on allozyme data is that the stocks introduced from the Mexican highlands were based on material originally developed, and perhaps originally domesticated, in southern Guatemala. If this were so, it is unclear whether their subsequent transfer to Mexico resulted from migrations of pre-Columbian Indians or from the activites of later Spanish colonists. In either case, the Mexican highland stocks probably had undergone fairly extensive domestication prior to their introduction into the southeastern United States. Mexican highland stocks were particularly interesting to breeders because they were already nearly day-length neutral when introduced into the United States, a feature crucial to the development of successful Upland cultivars, particularly after the introduction and subsequent spread of the boll weevil into the southern United States.

A third and somewhat surprising observation is that not a single unique allele was detected in the 50 Upland cultivars. This was unexpected because modern cultivars of the other domesticated tetraploid species, *G. barbadense,* are heavily introgressed with *G. hirsutum* germplasm (Percy and Wendel, 1990), leading to the expectation that the reverse would also be observed. In addition, Upland cultivar development is known to have involved deliberate introgression of transspecific sources of germplasm (Ramey, 1966; Fryxell, 1976; Meredith, 1991). In his review of the contributions of transspecific germplasm to Upland cotton development, Meredith (1991) discussed several suspected and documented cases of introgression from *G. barbadense,* as well as breeding programs that intentionally introduced germplasm from the Hawaiian tetraploid *G. tomentosum* Nuttall ex Seemann, the African and Asian diploids *G. anomalum* Wawra ex Wawra & Peyritch, *G. arboreum,* and *G. herbaceum,* and the Sonoran diploid *G. thurberi.* Thus, intentional interspecific introgression into *G. hirsutum* has involved a minimum of two tetraploid and four diploid *Gossypium* species. Except for the genetically similar *G. tomentosum* (DeJoode and Wendel, 1992), each of these species is readily distinguished from *G. hirsutum* by a relatively large number of diagnostic allozyme markers (data not presented). That only one diagnostic allele (the *G. barbadense* allele *Nad1-n*) was detected in the 50 Upland cultivars (in Cascot L-7 and Lankart PR-75), each examined for 50 loci, demonstrates that representation of exotic sources of genes is minimal in modern Upland cultivars. The suggestion to emerge from these data is that most exotic germplasm has been selectively eliminated through postintrogression breeding programs. Ongoing

research using nuclear, low-copy number restriction fragment length polymorphism (RFLP) loci (Brubaker and Wendel, unpublished data) should shed additional light on this problem.

LITERATURE CITED

AFFLECK, T. 1851. The early days of cotton growing in the South-West. *De Bow's Review* 10: 668–669.

ANONYMOUS. 1972. Genetic vulnerability of major crops. Cotton, 269–281. National Academy of Sciences, Washington, DC.

———. 1990. Agriculture Statistics Board, Annual Crop Summary, NASS, USDA, Washington, DC.

BRETTING, P. K., M. M. GOODMAN, AND C. W. STUBER. 1987. Karyological and isozyme variation in West Indian and allied American mainland races of maize. *American Journal of Botany* 74: 1601–1613.

BROWN, H. B., AND J. O. WARE. 1958. Cotton, 3d ed. McGraw-Hill, New York, NY.

DEJOODE, D. R., AND J. F. WENDEL. 1992. Genetic diversity and origin of the Hawaiian Islands cotton, *Gossypium tomentosum*. *American Journal of Botany* 79: 1311–1319.

DOEBLEY, J. F. 1989. Isozymic evidence and the origin of crop plants. *In* D. E. Soltis and P. S. Soltis [eds.], Isozymes in plant biology, 165–186. Dioscorides Press, Portland, OR.

ENDRIZZI, J. E., E. L. TURCOTTE, AND R. J. KOHEL. 1985. Genetics, cytology, and evolution of *Gossypium*. *Advances in Genetics* 23: 271–375.

FRYXELL, P. A. 1968. The typification and application of the Linnaean binomials in *Gossypium*. *Brittonia* 20: 378–386.

———. 1976. A nomenclator of *Gossypium*—the botanical names of cotton. USDA Agricultural Research Service Technical Bulletin 1491, Washington, DC.

———. 1979. The natural history of the cotton tribe. Texas A&M University Press, College Station, TX.

HARLAN, J. R. 1971. Agricultural origins: centers and non-centers. *Science* 174: 468–473.

———. 1975. Geographic patterns of variation in some cultivated plants. *Journal of Heredity* 66: 182–191.

HUTCHINSON, J. B. 1951. Intra-specific differentiation in *Gossypium hirsutum*. *Heredity* 5: 161–193.

———. 1959. The application of genetics to cotton improvement. Cambridge University Press, Cambridge.

———, R. A. SILOW, AND S. G. STEPHENS. 1947. The evolution of *Gossypium*. Oxford University Press, London.

JOHNSON, B. L. 1975. *Gossypium palmeri* and a polyphyletic origin of the New World cottons. *Bulletin of the Torrey Botanical Club* 102: 340–349.

LEE, J. A. 1984. Cotton as a world crop. *In* R. J. Kohel and C. L. Lewis [eds.], Cotton, Agronomy Monograph No. 24, 1–25. Crop Science Society of America, Madison, WI.

MEREDITH, W. R. 1991. Contributions of introductions to cotton improvement. *In* H. L. Shands and L. E. Wiesner [eds.], Use of plant introductions in cultivar development, part 1. Crop Science Society of America, Madison, WI.

MOORE, J. H. 1956. Cotton breeding in the old south. *Agricultural History* 30: 95–104.

NEI, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583–590.

———. 1987. Molecular evolutionary genetics. Columbia University Press, New York, NY.

NILES, G. A., AND C. V. FEASTER. 1984. Breeding. *In* R. J. Kohel and C. F. Lewis [eds.], Cotton, Agronomy Monograph No. 24, 201–231. Crop Science Society of America, Madison, WI.

PERCIVAL, A. E. 1987. The national collection of *Gossypium* germplasm. Southern Cooperative Series Bulletin 321, College Station, TX.

PERCY, R. G., AND J. F. WENDEL. 1990. Allozyme evidence for the origin and diversification of *Gossypium barbadense* L. *Theoretical and Applied Genetics* 79: 529–542.

RAMEY, H. H. 1966. Historical review of cotton variety development. Proceedings of the 18th Cotton Improvement Conference. Memphis, TN.

ROGERS, J. S. 1972. Measures of genetic similarity and genetic distance. Studies in genetics VII, University of Texas Publication 7213, 145–153. University of Texas, Austin, TX.

SMITH, C. E., AND S. G. STEPHENS. 1971. Critical identification of Mexican archaeological cotton remains. *Economic Botany* 25: 160–168.

SNEATH, P. H. A., AND R. R SOKAL. 1973. Numerical taxonomy. W. H. Freeman, San Francisco, CA.

STEPHENS, S. G. 1958. Factors affecting seed dispersal in *Gossypium* and their possible evolutionary significance, Technical Bulletin No. 131. North Carolina Agricultural Experimental Station, Raleigh, NC.

———. 1967. Evolution under domestication of the New World cottons (*Gossypium* spp.). *Ciencia e Cultura* 19: 118–134.

———. 1974. Geographic and taxonomic distribution of anthocyanin genes in New World cottons. *Journal of Genetics* 61: 128–141.

———. 1975. Some observations on photoperiodism and the development of annual forms of domesticated cottons. *Economic Botany* 30: 409–418.

WAILES, B. L. C. 1854. Report on the agriculture and geology of Mississsippi, embracing a sketch of the social and natural history of the state. Philadelphia, MS.

WARE, J. O. 1951. Origin, rise and development of American Upland cotton varieties and their status at present. Cotton and other fiber crops and diseases, 1–173. University of Arkansas, Fayetteville, Fayetteville, AR.

WEEDEN, N. F., AND J. F. WENDEL. 1989. Genetics of plant isozymes. *In* D. E. Soltis and P. S. Soltis [eds.], Isozymes in plant biology, 46–72. Dioscorides Press, Portland, OR.

WENDEL, J. F., P. D. OLSON, AND J. M. STEWART. 1989. Genetic diversity, introgression, and independent domestication of old world cultivated cottons. *American Journal of Botany* 76: 1795–1806.

———, AND R. G. PERCY. 1990. Allozyme diversity and introgression in the Galapagos Islands endemic *Gossypium darwinii* and its relationship to continental *G. barbadense*. *Biochemical and Systematic Ecology* 18: 517–528.

———, AND N. F. WEEDEN. 1989. Visualization and interpretation of plant isozymes. *In* D. E. Soltis and P. S. Soltis [eds.], Isozymes in plant biology, 5–45. Dioscorides Press, Portland, OR.

WORKMAN, P. L., AND NISWANDER, J. D. 1970. Population studies on southwestern Indian tribes. II. Local genetic differentiation in the Papago. *American Journal of Human Genetics* 22: 24–49.