Theor Appl Genet (1992) 85:451–460

Patterns of allozyme variation in cultivated and wild *Sorghum bicolor*

P. R. Aldrich[1], J. Doebley[1], K. F. Schertz[2], and A. Stec[1]

[1] Department of Plant Biology, University of Minnesota, St. Paul, MN 55108, USA
[2] USDA-ARS in cooperation with the Department of Soil and Crop Sciences, Texas A & M University, College Station, TX 77843, USA

Received November 8, 1991; Accepted May 7, 1992
Communicated by A. R. Hallauer

Summary. Patterns of allozyme variation were surveyed in collections of cultivated and wild sorghum from Africa, the Middle East, and Asia. Data for 30 isozyme loci from a total of 2067 plants representing 429 accessions were analyzed. Regional levels of genetic diversity in the cultivars are greater in northern and central Africa compared to southern Africa, the Middle East, or Asia. The spatial distribution of individual alleles at the most variable loci was studied by plotting allele frequencies on geographic maps covering the distribution of sorghum. Generally, many of the alleles with frequencies below 0.25 are localized in specific portions of the range and are commonly present in more than one race in that region. Several alleles occur in both wild and cultivated sorghum of one region and are absent from sorghum elsewhere, suggesting local introgression between the wild and cultivated forms. Although the same most common allele was found in the wild and cultivated gene pools at 29 of the 30 loci, phenetic analyses separated the majority of wild collections from the cultivars, indicating that the two gene pools are distinct. Wild sorghum from northeast and central Africa exhibits greater genetic similarities to the cultivars compared to wild sorghum of northwest or southern Africa. This is consistent with the theory that wild sorghum of northeast-central Africa is ancestral to domesticated sorghum. Wild sorghums of race arundinaceum of northwest Africa and race virgatum from Egypt are shown to be genetically distinct from both other forms of wild sorghum and from the cultivars. Suggestions for genetic conservation are presented in light of these data.

Correspondence to: J. Doebley

Key words: Sorghum – Isozyme – Genetic diversity – Domestication – Conservation

Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is an important cereal crop worldwide. In Africa, it is used extensively for such products as grain, fodder, construction material, brooms, syrup, and beer. The primary gene pool of sorghum, which includes the cultivars and their wild and weedy relatives (Harlan and de Wet 1971), has been studied intensely because of its importance as a potential source of germ plasm. Until recently, information concerning the genetic diversity of sorghum and the relationships among the infraspecific taxa has come primarily from studies of comparative morphology and biogeography (de Wet et al. 1970, 1976; de Wet 1978; Harlan and de Wet 1972; Harlan and Stemler 1976). However, interpreting the morphological divergence of a cultivar from its predecessor can pose serious problems for the taxonomist or crop evolutionist. Humans have strongly selected crops for agronomically important traits, and the phylogenetic relationships between crops and their wild progenitors are often not obvious (Doebley 1989).

Isozymes are useful markers for studying genetic variation and evolution at the species level (Crawford 1990) and have probably not been under direct human selection in the domestication process (Doebley 1989). Although the number of loci available for study is limited, isozyme variants, which are codominantly inherited, are often quite abundant within a species. Furthermore, isozyme analysis is relatively simple and inexpensive, which permits the analysis of many populations or individuals from a species.

The investigation presented here builds upon the work of Morden et al. (1989, 1990), who used isozyme analysis to characterize the levels of genetic diversity in both the cultivated and the wild and weedy sorghums. Our study broadens the scope of the previous surveys. It utilizes data from additional cultivated collections representing a larger geographic sample, and it directly compares genetic patterns of diversity in wild and cultivated sorghums to elucidate evolutionary relationships within the species. This paper addresses four questions. (1) Are there centers of genetic diversity in sorghum, and if so, where are they located? (2) How is genetic diversity geographically distributed over the range? (3) Which populations of wild sorghum show the greatest genetic similarity to the cultivars? and (4) What would be a sound strategy for genetic conservation in sorghum?

Taxonomic background

The most widely accepted taxonomic treatment of cultivated sorghum and its wild relatives is that of de Wet (1978) and Harlan and de Wet (1972), who proposed a single species, *Sorghum bicolor*, and three subspecific categories: ssp. *arundinaceum* (Desv.) de Wet et Harlan, ssp. *bicolor*, and ssp. *drummondii* (Steud.) de Wet. Subspecies *arundinaceum* consists of four wild races: aethiopicum, arundinaceum, verticilliflorum, and virgatum. Morphological and ecological boundaries between these races are not well-defined, and thus de Wet (1978) suggested that they represent ecotypes or eco-geographic races rather than formal taxonomic categories. Race aethiopicum is a desert grass found in northern Africa, south of the Sahara (de Wet et al. 1970; de Wet 1978). Race arundinaceum is a forest grass primarily of the wet tropics of northwest Africa along the Guinea Coast and Congo (de Wet et al. 1976; de Wet 1978). Race verticilliflorum is common in most regions of Africa, principally the savannah regions (de Wet et al. 1976). Race virgatum occurs primarily in the arid regions of northeast Africa (de Wet 1978).

Subspecies *bicolor* contains the five cultivated races: bicolor, caudatum, durra, guinea, and kafir. These races have overlapping geographic ranges, although one race may be predominant in a particular region (Harlan and de Wet 1972; de Wet et al. 1976; Harlan and Stemler 1976). These five races readily cross with one another and in combination produce ten intermediate races (de Wet et al. 1976; de Wet 1978). Subspecies *drummondii* includes weedy forms that are thought to result from crosses between wild and cultivated sorghum growing sympatrically (de Wet et al. 1970; de Wet 1978). Subspecies *arundinaceum* and *drummondii* will be referred to collectively as the wild sorghum in the remainder of this paper.

Materials and methods

Isozymes were analyzed in 2067 plants from 429 accessions of sorghum. Accessions were selected to maximize coverage of racial diversity across the geographic range of sorghum, although some desired germ plasm was unavailable. Caryopses were obtained from the International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) and the USDA Regional Plant Introduction Station, Experiment, Georgia (RPIS). Accessions have been increased an unknown number of times since they were initially collected.

Some accessions have been analyzed previously for isozymes. Morden et al. (1989) assayed 8 plants each of 83 accessions of the cultivated sorghum. Data for all of these accessions are included in the present study with the exceptions of accessions IS12662, IS1309, IS2169, IS6832, and PI499306, which are of unknown geographic origin. Morden et al. (1990) assayed 8 plants each of 90 accessions of wild sorghum, all of which are included in this study except accessions IS18853, IS18852, IS18894, IS14212, IS14241, IS18843, IS18845, IS18927, PI302232, PI302233, PI228364, and PI255738, which either are of unknown geographic origin or are species other than *Sorghum bicolor*.

In addition, 273 other accessions of the cultivars (Table 1) have been analyzed for this study. Three plants were assayed for each of these accessions. The accessions used in the analysis are divided among the taxonomic categories as follows: *S. bicolor* ssp. *arundinaceum* race aethiopicum (10 accessions), race arundinaceum (18 accessions), race verticilliflorum (28 accessions), race virgatum (12 accessions), *S. bicolor* ssp. *bicolor* race bicolor (65 accessions), race caudatum (89 accessions), race durra (62 accessions), race guinea (61 accessions), race kafir (28 accessions), intermediate races (25 accessions), identity of race unknown (21 accessions), *S. bicolor* ssp. *drummondii* (10 accessions).

Seedlings were grown, enzymes extracted, and electrophoresis conducted as previously described in Morden et al. (1987, 1989, 1990). Briefly, seeds were wrapped in wet paper towels and grown in the dark for approximately 7 days. All seedlings were immersed in water for 14 h prior to enzyme extraction to enhance the expression of anaerobic alcohol dehydrogenase. This short treatment has no differential effect on the expression of the other enzyme systems. Etiolated seedling tissue was ground in a chilled mortar with 2–4 drops of extraction buffer (Morden et al. 1987). Plant extracts were subsequently absorbed onto filter paper wicks and stored at −80 °C until electrophoresed.

Extracts of each plant were electrophoresed in four buffer systems: system L, lithium hydroxide-borate (Cardy et al. 1983); system M, morpholine-citrate (modified from Clayton and Tretiak 1972); system N, sodium hydroxide-borate (Wendel and Stuber 1983); and system T, TRIS-citrate (Meizal and Markert 1967). Following electrophoresis, the gels were trimmed, and the slab anodal to the origin was sliced and stained for enzyme activity. Seventeen enzyme systems encoded by 30 loci were analyzed (Table 2). Most of the loci are unlinked, indicating that they come from many different regions of the genome (Doebley, Schertz, Aldrich, and Morden unpublished results).

To analyze the geographic distribution of genetic variation, allele frequency data from both the wild and cultivated accessions were pooled separately by country, or, where sampling was less extensive, by groups of contiguous countries, since the country of origin is the only location information available for the accessions. For some analyses, larger geographic regions were recognized as follows: northwest Africa, northeast Africa, central Africa, southern Africa, Middle East, and Far East (Table 3).

Total panmictic heterozygosity for each locus per country was calculated as follows:

$$H = 1 - \Sigma x_i^2$$

where x_i is the average frequency of the i^{th} allele in the country.

Table 1. Country of origin, race, and identification number of the sorghum accessions included in this study but not analyzed by Morden et al. (1989, 1990). Seed from accessions with IS and PI prefixes were obtained from the International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) and USDA Regional Plant Introduction Station, Experiment, Georgia (RPIS), respectively

AFGHANISTAN-Bicolor: IS3081, IS3085. Caudatum: IS2354. Kafir: IS2425
ALGERIA-Durra: IS917
ALGOLA-Bicolor: IS14212, IS14240. Caudatum: IS14224, IS14238, IS14263
BANGLADESH-Caudatum: IS19389. Guinea: IS18239, IS18241
BENIN-Durra: IS12399
BOTSWANA-Bicolor: IS19460, IS22241. Caudatum: IS14282, IS14284. Durra: IS14286, IS14328. Guinea: IS22297, IS22313. Kafir: IS3399, IS3404
BURMA-Bicolor: IS22615, IS22618. Guinea: IS1194
CAMEROON-Bicolor: IS14863, IS16555. Caudatum: IS14774, IS15910. Durra: IS14790, IS14995. Guinea: IS14918, IS15285
CENTRAL AFRICAN REPUBLIC-Caudatum: IS2406, IS7086
CHAD-Bicolor: IS10747, IS10810. Caudatum: IS8556, IS8557. Durra: IS6842. Guinea: IS7171, IS12496
CHINA-Bicolor: IS12711, IS12712. Caudatum: IS3390, IS3392. Guinea: IS1293. Kafir: IS259. UNKNOWN: IS1, IS140, IS226, IS347, IS422, IS451, IS527, IS536, IS568, IS730, IS731, IS779, IS793, IS794, IS820, IS3046, IS3113, IS3335, IS3467, IS5981
CONGO-Caudatum: IS12599
EGYPT-Caudatum: IS2872, IS20857. Durra: IS2870, IS2874
ETHIOPIA-Caudatum: IS2662. Guinea: IS10018. Kafir: IS8539, IS12541
GAMBIA-Durra: IS23620. Guinea; IS23619, IS23650
GHANA-Bicolor: IS25061, IS25062. Caudatum: IS17622, IS17632. Durra: IS17616, IS17628. Guinea: IS17609, IS17613
INDIA-Caudatum: IS8187, IS8266. Durra-Bicolor: IS1387, IS3435. Guinea-Caudatum: IS6721. Kafir: IS2067, IS21963. Kafir-Caudatum: IS8166, IS8231
INDONESIA-Bicolor: IS22581, IS22582. Caudatum: IS2350, IS20952
IRAN-Bicolor: IS2373. Caudatum: IS2413. Durra: IS3082
IRAQ-Durra: IS13148, IS13149
ISRAEL-Caudatum: IS10352, IS10366
JAPAN-Bicolor: IS3606, IS8053. Caudatum: IS8013, IS8094. Durra: IS8061, IS8107. Guinea: IS8063, IS8117.
KENYA-Bicolor: IS8822, IS9143. Caudatum: IS12586. Durra: IS12577, IS12582. Guinea: IS8726, IS14541
KOREA-Caudatum: IS10343, IS20843
LEBANON-Caudatum: IS18056, IS18057. Durra: IS18085, IS24607. Guinea: IS24600, IS24680
LESOTHO-Bicolor: IS14438, IS14442. Caudatum: IS14434. Guinea: IS14430
MADAGASCAR-Caudatum: IS13515
MALAWI-Bicolor: IS1195, IS14357. Caudatum: IS14412, IS21499. Durra: IS21565, IS23788. Guinea: IS14337, IS14353
MALI-Bicolor: IS3845, IS21395. Caudatum: IS3867, IS22665. Durra: IS3819, IS3888 Guinea: IS3817, IS3829, IS7979, IS22669
MOZAMBIQUE-Caudatum: IS23680. Guinea: IS23675, IS23695
NEPAL-Durra: IS3960. Guinea: IS3956
NIGER-Bicolor: IS9602, IS20474. Caudatum: IS9586, IS9594. Durra-Caudatum: IS7148. Durra: IS18308. Guinea: IS7184, IS18303. Kafir-Caudatum: IS8003
NIGERIA-Kafir: IS2901, IS17126
PAKISTAN-Bicolor: IS8341, IS13253. Caudatum: IS8337. Durra: IS1401, IS8348. Guinea: IS8347
SENEGAL-Bicolor: IS20013, IS20073. Caudatum: IS19973. Durra: IS3425. Guinea: IS3424, IS3433
SIERRA LEONE-Guinea: IS21848, IS21849
SOMALIA-Caudatum: IS7124, IS7125. Durra: IS22681, IS22696
SOUTH AFRICA-Bicolor: IS1252, IS3150. Caudatum: IS2377, IS2381. Durra: IS2553, IS2864. Guinea: IS3175, IS9404. Kafir: PI409306. Kafir-Caudatum: IS2401, IS2419
SRI LANKA-Bicolor: IS22605. Caudatum: IS22592, IS22595. Durra: IS22596, IS22601. Guinea: IS19466
SUDAN-Kafir: IS9618, IS9624
SWAZILAND-Bicolor: IS14318. Caudatum: IS8670, IS8680. Guinea: IS14316, IS14317
SYRIA-Durra: IS13150
TAIWAN-Durra: IS3697
TANZANIA-Bicolor: IS1291, IS23085. Caudatum: IS1309, IS1337. Durra: IS23102. Guinea: IS7173
THAILAND-Caudatum: IS10302, IS10303. Guinea: IS3786. Kafir: IS10301
TURKEY-Bicolor: IS12784, IS12785. Caudatum: IS12828, IS20865. Durra: IS12859, IS12863
UGANDA-Bicolor: IS2668, IS2736. Caudatum: IS2729, IS7132. Durra: IS2740, IS8209. Guinea: IS2649, IS2724. Kafir: IS10400, IS12219
UPPER VOLTA-Caudatum: IS7088. Durra: IS7093
USSR-Bicolor: IS13196, IS13243. Caudatum: IS13227, IS13232. Durra: IS19399
YEMEN-Bicolor: IS23989, IS23924. Caudatum: IS17590, IS17601. Durra: IS17592
ZAIRE-Bicolor: IS1255. Caudatum: IS1311, IS3405
ZAMBIA-Bicolor: IS23222, IS23232. Caudatum: IS23229, IS23323. Guinea: IS8698, IS23185
ZIMBABWE-Bicolor: IS12294, IS13478, IS13482. Caudatum: IS2814, IS2836. Durra: IS12274. Guinea: IS5576, IS12305. Kafir: IS2694, IS2843. Kafir-Caudatum: IS1269

Only countries represented by 4 or more accessions were used for this analysis. Average panmictic heterozygosity for a region was estimated by averaging all panmictic heterozygosities per country within that region, weighted by number of accessions per country.

Alleles were classified according to their frequencies in the wild and cultivated gene pools and their geographic distributions. Broadly distributed alleles occurring at overall frequencies greater than 0.25 are referred to as 'cosmopolitan'. Narrowly distributed alleles with frequencies below 0.25 are referred to as 'rare' or 'uncommon'.

The geographic distribution of alleles was studied by portraying allele frequencies (pooled by country) on maps covering the range of sorghum. Only loci that showed sufficient variability to be informative were mapped in this fashion.

Phenetic techniques were used to study the genetic relatedness among the geographically identified collections of wild and cultivated sorghum. Principal component analysis was performed using the variance-covariance matrix of allele frequencies. The operational taxonomic units used were wild and cultivated accessions pooled, separately, according to their country of origin. Modified Rogers' distances (Wright 1978) based on allele frequencies were calculated from the same data set (pooled by country) and used to construct a UPGMA phenogram (unweighted pair-group method using arithmetic averages). Principal component analysis was also used to study the genetic relations among individual accessions of wild sorghum using allele frequency data for all geographically identified accessions of ssp. *arundinaceum* and ssp. *drummondii*.

Table 2. Enzyme systems assayed, number of loci resolved, and the gel buffer systems

Enzyme	Abbreviation	Number of loci	Buffer system[a]
Aconitase	ACO	2	T
Adenylate kinase	ADK	1	T
Alcohol dehydrogenase	ADH	3	L
Aspartate aminotransferase	AAT	3	N
Catalase	CAT	1	N
Endopeptidase	ENP	1	N
Fructokinase	FRK	2	L, T
Glutamate dehydrogenase	GDH	1	N
Glutamate-pyruvate transaminase	GPT	2	L
Isocitrate dehydrogenase	IDH	2	M
Malate dehydrogenase	MDH	3	M
Mitochondrial MDH modifier	MMM	1	M
Phosphoglucose isomerase	PGI	2	L
Phosphoglucomutase	PGM	1	T
6-Phosphogluconate dehydrogenase	6-PGD	2	M
Shikimate dehydrogenase	SAD	1	M
Triosephosphate isomerase	TPI	2	L

[a] L, Lithium hydroxide-borate; M, morpholine-citrate; N, sodium hydroxide-borate; T, TRIS-citrate (Morden et al. 1987)

Results

Levels and distribution of diversity

Genetic similarities between wild and cultivated sorghums are substantial. Approximately half (57) of the total 110 alleles are shared by both taxa. The same allele was most common in both wild and cultivated at 29 of the 30 loci examined. *Enp1* is the exception, with *Enp1–5* at a higher frequency in the cultivars and *Enp1–2* most prevalent in the wild. Seven loci were invariant in the cultivated gene pool, while 5 were invariant in the wild gene pool. Four of these loci (*Aat1*, *Idh1*, *Mdh2*, and *Pgm1*) were invariant in both gene pools.

Levels of genetic diversity are greater in wild sorghum than in cultivated sorghum. Ninety-seven alleles were identified in the 78 wild accessions assayed, while only 70 alleles were found among the 351 cultivated accessions. Differences between the wild and cultivated sorghum are primarily due to the presence or absence of low frequency alleles. Of the alleles found in the wild taxa, 41% (40 alleles) were unique to the wild taxa. These 40 alleles had an average frequency of 0.025 and a range of 0.001–0.169 in the total sample of wild accessions. Of the alleles in the cultivated accessions 17% (12 alleles) were found only in those accessions. These alleles had an average frequency of 0.009 and a range of 0.001–0.042. Thus, all of the common alleles in the cultivated sorghum are represented in its wild relatives.

Table 3. Total panmictic heterozygosity per country in cultivated sorghum averaged over all loci

Region	Country	N	Hc
Northwest Africa	Benin/Ghana/Ivory Coast	12	0.078
	Gambia/Senegal/ Sierra Leone	13	0.113
	Mali	10	0.106
	Nigeria	18	0.090
	Niger	9	0.085
	Upper Volta	4	0.097
		Ntot = 66	Ave = 0.095
Northeast Africa	Chad	7	0.081
	Egypt	4	0.042
	Ethiopia	20	0.085
	Somalia	4	0.072
	Sudan	17	0.085
		Ntot = 52	Ave = 0.080
Central Africa	Cameroon	8	0.093
	Central African Republic/Congo/ Zaire	6	0.063
	Kenya	8	0.069
	Uganda	11	0.070
		Ntot = 33	Ave = 0.074
Southern Africa	Angola	5	0.062
	Botswana	10	0.041
	South Africa/ Swaziland/Lesotho	31	0.042
	Malawi	8	0.059
	Tanzania	6	0.046
	Zimbabwe	12	0.065
	Zambia	6	0.083
		Ntot = 78	Ave = 0.052
Middle East	Afghanistan/ Pakistan/Iran	13	0.065
	Lebanon/Iraq/ Israel/Syria	11	0.083
	Turkey	6	0.058
	Yemen	5	0.053
		Ntot = 35	Ave = 0.068
Far East	Bangladesh/India/ Nepal	28	0.070
	Burma/Thailand	7	0.072
	China	26	0.061
	Indonesia	4	0.059
	Japan	8	0.081
	Sri Lanka	6	0.044
		Ntot = 79	Ave = 0.066

Abbreviations: N, Sample size; Hc, total panmictic heterozygosity; Ntot, total number of accessions in a region; ave, average panmictic heterozygosity per region weighted by number of accessions per country

Estimates of total panmictic heterozygosity are higher on average in the cultivars of northern and central Africa than in those of Asia and southern Africa (Table 3). Southern Africa has the lowest observed regional average (0.052). Northwest African cultivars exhibit the greatest genetic diversity (0.095), particularly in Gambia/Senegal/Sierra Leone (0.113).

Geographic patterns of diversity

Maps depicting the spatial distribution of alleles at four loci (*Aco1*, *Adk1*, *Cat1*, *Frk2*) are shown (Figs. 1–4). Aldrich (1991) presents maps for 10 additional loci. Allele frequencies for wild and cultivated accessions have been pooled separately by country in these maps. Large pie diagrams on the maps show allele frequencies from wild collections, and small pie diagrams indicate their frequencies in cultivars. Pie diagrams that overlap national boundaries represent data from countries that have been pooled together for better sampling of the area.

Genetic relationships

Genetic relationships among the wild and cultivated sorghum collections were visualized using principal component analysis (Fig. 5). Allele frequency data were pooled across accessions according to the country of origin, averaging wild and cultivated gene pools separately. The shapes of the symbols in Fig. 5 indicate regional categories and shaded symbols represent wild collections.

The first two principal components in Fig. 5 captured 43% of the variation. The wild and cultivated collections

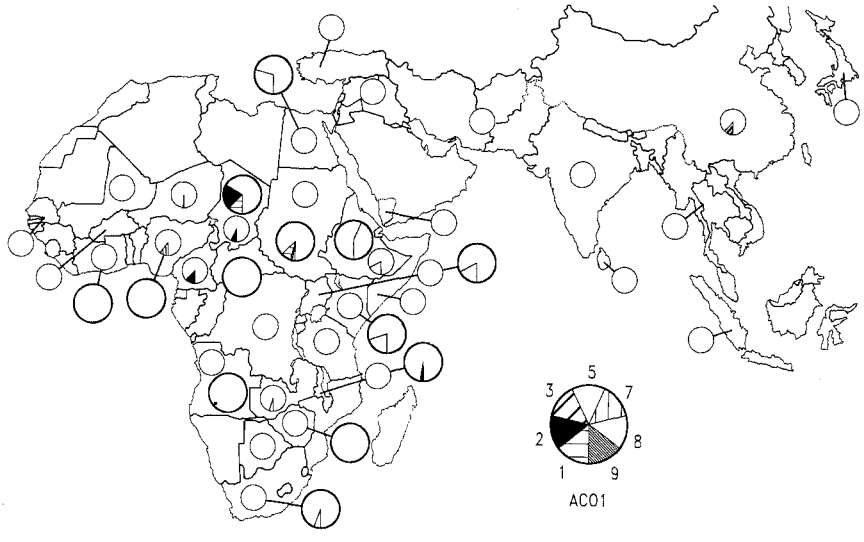

Fig. 1. Geographic map representing diversity at *Aco1*. *Large and small pie diagrams* indicate isozyme allele frequency data that has been pooled by country of origin for wild and cultivated sorghum accessions, respectively

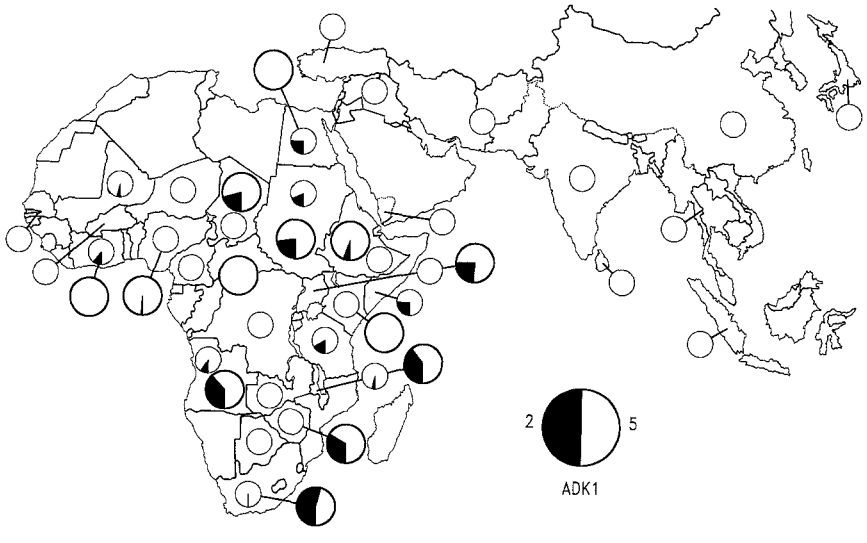

Fig. 2. Geographic map representing diversity at *Adk1*. *Large and small pie diagrams* indicate isozyme allele frequency data that has been pooled by country of origin for wild and cultivated sorghum accessions, respectively

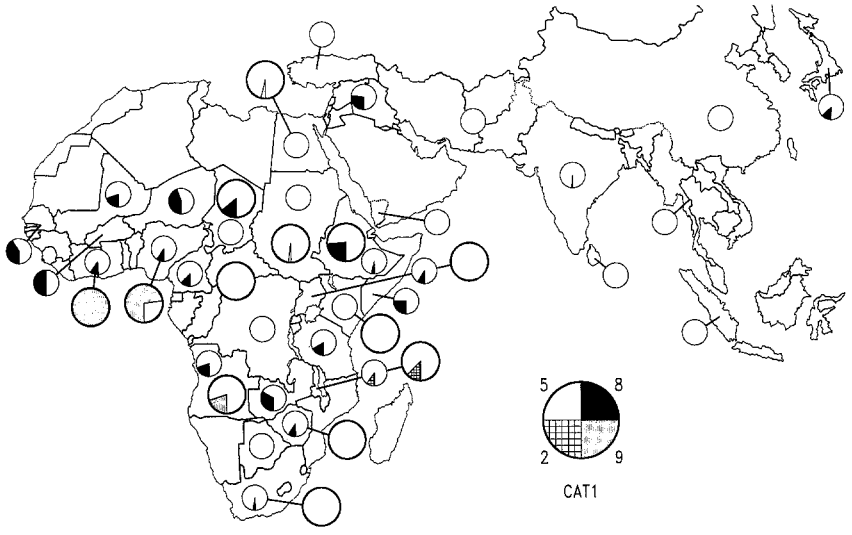

Fig. 3. Geographic map representing diversity at *Cat1*. *Large and small pie diagrams* indicate isozyme allele frequency data that has been pooled by country of origin for wild and cultivated sorghum accessions, respectively

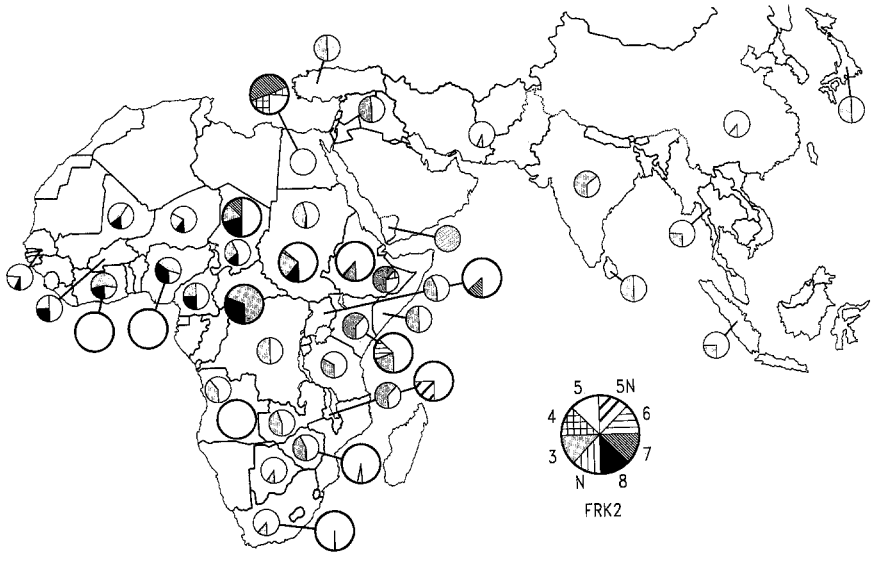

Fig. 4. Geographic map representing diversity at *Frk2*. *Large and small pie diagrams* indicate isozyme allele frequency data that has been pooled by country of origin for wild and cultivated sorghum accessions, respectively

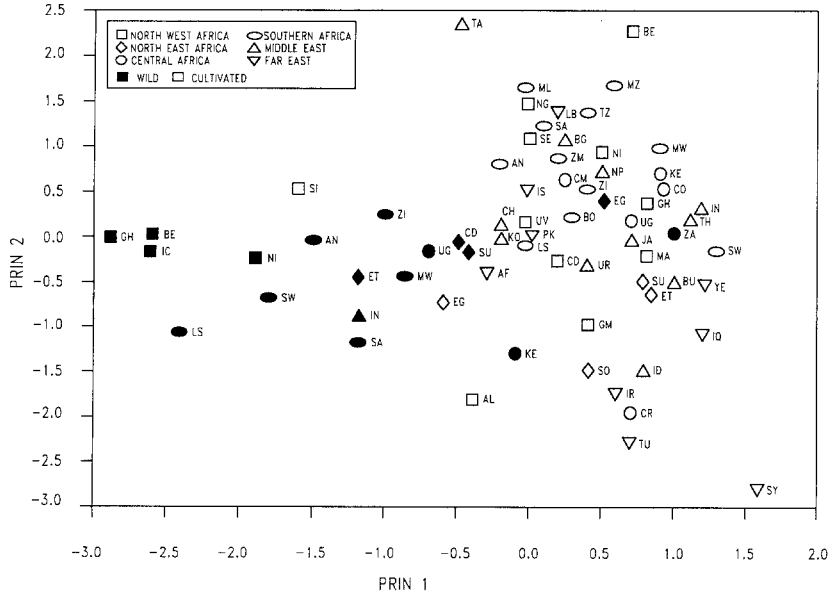

Fig. 5. Graph of the first two components of a principal component analysis based on wild and cultivated isozyme allele frequency data pooled by country of origin. Countries are abbreviated as follows: *AF* Afghanistan, *AL* Algeria, *AN* Angola, *BE* Benin, *BG* Bangladesh, *BO* Botswana, *BU* Burma, *CD* Chad, *CH* China, *CM* Cameroon, *CO* Congo, *CR* Central African Republic, *EG* Egypt, *ET* Ethiopia, *GH* Ghana, *GM* Gambia, *IC* Ivory Coast, *ID* Indonesia, *IN* India, *IQ* Iraq, *IR* Iran, *IS* Israel, *JA* Japan, *KE* Kenya, *KO* Korea, *LB* Lebanon, *LS* Lesotho, *MA* Mali, *ML* Malagasky Republic, *MW* Malawi, *MZ* Mozambique, *NG* Niger, *NI* Nigeria, *NP* Nepal, *PK* Pakistan, *SA* South Africa, *SE* Senegal, *SI* Sierra Leone, *SL* Sri Lanka, *SO* Somalia, *SU* Sudan, *SW* Swaziland, *TA* Taiwan, *TH* Thailand, *TU* Turkey, *TZ* Tanzania, *UG* Uganda; *UR* USSR, *UV* Upper Volta, *YE* Yemen, *ZA* Zaire, *ZI* Zimbabwe, *ZM* Zambia

are separated along the first principal component, which represents 25% of the variation. Wild collections from northwest and southern Africa are moved farthest to the left of the plot, along with the cultivated collections from Sierra Leone, while wild collections from northeast and central Africa (Chad, Sudan, and Uganda) fall along the interface between the wild and cultivated sorghum toward the center of the plot. The wild sorghum from Egypt, Kenya, and Zaire cluster with the majority of the cultivars. The third principal component (not shown) separates the wild sorghum of northwest and southern Africa. Other samples show little dispersion along this axis.

Average linkage cluster analysis on the same pooled data set is shown in Fig. 6. The most salient aspect of this phenogram is the separation of the wild sorghum into three fairly well-defined groups. The wild group at the base of Fig. 6 is farthest removed from all other accessions. It consists of all the wild northwest African collections, along with those from Egypt and Lesotho. Of the accessions assayed from these countries 42% are of race virgatum and another 42% are of race arundinaceum. No members of race verticilliflorum occur in this group. At the very top of the tree is a cluster of five collections, consisting of three collections of wild sorghum from southern Africa, the wild collection from Ethiopia, and one cultivated collection from northwest Africa. These collections consist primarily of race verticilliflorum (58%), with race arundinaceum next most abundant at 21%. Race virgatum is absent from this group. Clustering closer to the majority of cultivars is a third group of wild collections. These are primarily of race verticilliflorum (41%) and race aethiopicum (21%). Wild collections of this part of the phenogram originate from northeast Africa (Chad and Sudan), southern Africa (Malawi and Zimbabwe), and the Far East (ssp. *drummondii* from India). This group also includes the wild sorghum most isozymically similar to the cultivars, the three diffusely situated collections from central Africa (Kenya, Uganda, and Zaire).

Principal component analysis was conducted on accessions of *S. bicolor* ssp. *arundinaceum* and ssp. *drummondii* (Fig. 7) in order to determine whether any correspondence exists between geographic origin and isozymic constitution in the wild gene pool. The first two principal components captured 33% of the total variation. Symbol shapes indicate races, and shading indicates regional class.

Wild accessions do not cluster into distinct racial groups. The largest cluster, in the upper left of Fig. 7, is a heterogeneous group consisting primarily of races verticilliflorum and aethiopicum. It is the accessions from this group of wild sorghum that clustered most closely with the cultivars in the phenogram. The cluster in the right of Fig. 7 consists primarily of race virgatum from Egypt, along with a virgatum from Uganda and an aethiopicum from Egypt. Virgatum appears to be

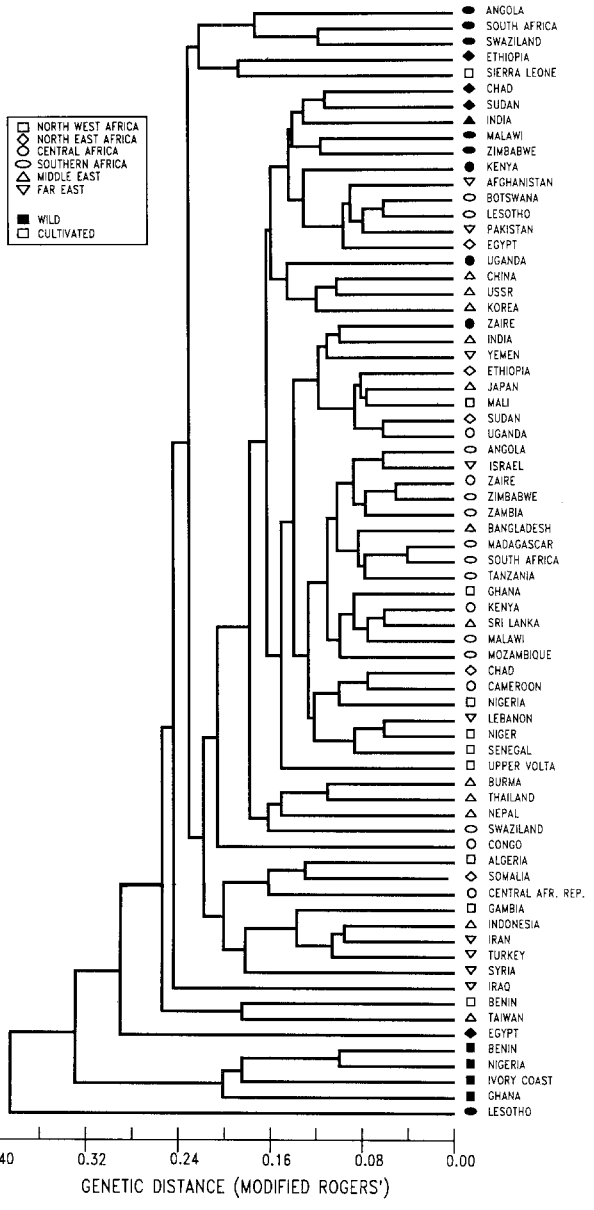

Fig. 6. Average linkage cluster analysis based on isozyme allele frequency data for wild and cultivated accessions pooled by country of origin using modified Rogers' distance (Wright 1978)

the most genetically distinct race since only 3 of the 12 virgatum accessions, 2 from Sudan and 1 from Uganda, group with the majority of wild accessions in the upper left. Another relatively distinct group appears in the lower left of the plot. All 7 accessions of race arundinaceum from northwest Africa occur here, along with a northwest African ssp. *drummondii* and a southern African collection of race arundinaceum. The 9 other accessions of race arundinaceum are scattered among the heterogeneous group in the upper left of Fig. 7 and originate in south, central and northeast Africa. Although, according to these marker loci, race arund-

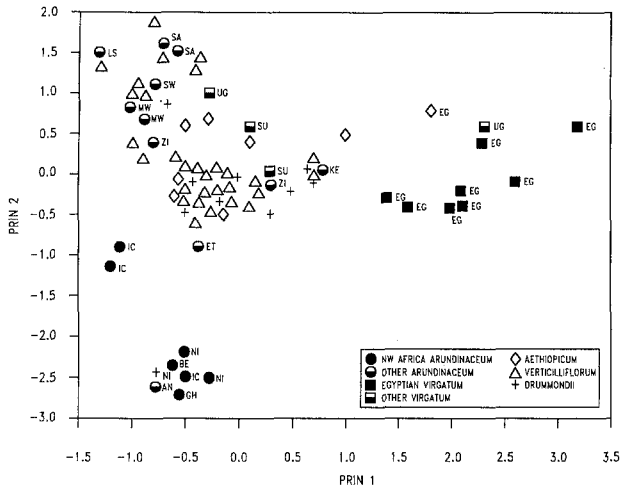

Fig. 7. Graph of the first two components of a principal component analysis based on isozyme allele frequency data from individual accessions of ssp. *arundinaceum* and ssp. *drummondii*. Abbreviations for countries of origin are listed in the caption of Fig. 5

inaceum does not exhibit a consistent genetic profile, accessions of race arundinaceum from northwest Africa are distinct from the other accessions of wild sorghum.

Discussion

Genetic diversity

A primary goal of this research was to characterize the geographic distribution of genetic diversity in wild and cultivated sorghums. This study examines a substantial number of isozyme loci compared to other studies (Gottlieb 1981) and utilizes data from 78 wild accessions (18 countries) and 351 cultivated accessions (52 countries). This extensive data set allows a thorough overview of isozyme diversity across the species' geographic range.

Levels of isozyme diversity in cultivated sorghum (Morden et al. 1989; Ollitrault et al. 1989a, b) and wild sorghum (Morden et al. 1990) have been studied previously. Since Morden et al. (1990) directly compared levels of diversity in the two subspecies, this will not be discussed in detail here. Morden et al. (1990) report that the wild sorghums exceed the cultivars in all measures of diversity that they assessed. Our larger data set supports this observation. Ninety-seven alleles were found among the 78 wild accessions while only 70 alleles were found in the 351 cultivated accessions assayed. This loss of genetic variation in the cultigen agrees with theory predicting that selection for agronomic traits from a limited portion of the wild gene pool will produce a genetic bottleneck and an erosion of diversity over time (Brown 1978; Frankel and Soulé 1981; Doebley 1989).

The extent of genetic differentiation between the wild and cultivated sorghum is limited. Of the 30 loci examined, only *Enp1* carries a different most common allele in the two groups. The majority of differences between the wild and the cultivated sorghum can be attributed to the presence or absence of low frequency alleles (Morden et al. 1990). Wild collections possess 40 unique low-frequency alleles, while cultivated sorghum contains only 12. These observations are consistent with the theory that cultivated sorghum is a recent derivative of ssp. *arundinaceum* (de Wet et al. 1970; Harlan and Stemler 1976; Morden et al. 1989).

The existence of localized centers of diversity is of particular interest to conservation geneticists (Brown 1978; Marshall 1990). Our analyses and that of Morden et al. (1989) indicate that northern and central African cultivated sorghum contains greater isozyme genetic diversity than cultivated sorghum of southern Africa or Asia. Panmictic heterozygosity per country averaged over all of northern and central Africa (0.082) is higher than the average of all countries in Asia (0.066) and southern Africa (0.052). Very few uncommon or rare alleles were found in the cultivated sorghum of Asia and only 1 of these alleles is entirely limited to Asia (*Aco1–3*).

Geographic distributions of alleles

The geographic distributions of individual alleles at the 4 most variable loci are presented in Figs. 1–4. Geographic maps for alleles at 10 additional loci are reported elsewhere (Aldrich 1991). Analysis of these maps reveals several features concerning the distribution of genetic variation in *Sorghum bicolor*.

Most loci possess a single predominant allele that occurs throughout the range of the species (Aldrich 1991): examples include *Aco1–5* (Fig. 1), *Adk1–5* (Fig. 2), and *Cat1–5* (Fig. 3). These cosmopolitan alleles are presumably pleisiomorphic characters shared by most accessions through common descent from the progenitor of *Sorghum bicolor*. Support for this interpretation is provided by the occurrence of these cosmopolitan alleles in *Sorghum halapense* (L.) Pers., a closely related species (Morden et al. 1990).

Certain alleles at several loci are distributed predominantly in western-central Africa and are absent or nearly absent elsewhere. This pattern is seen for *Frk2–8* (Fig. 4), as well as for *Mdh3–2* and *Pgi1–4* (Aldrich 1991). Correspondingly, another suite of alleles occurs predominantly throughout eastern Africa from Egypt to South Africa and is absent or at much lower frequencies elsewhere. This pattern is seen for *Aco1–8* (Fig. 1) and *Adk1–2* (Fig. 2), as well as for *Aat2–1* and *Sad1–7* (Aldrich 1991). These patterns indicate that gene flow between these broad geographic regions may have been restricted.

The taxonomic-geographic distribution of several alleles indicates that race arundinaceum of northwestern Africa is genetically isolated from the cultivars of that region. For example, *Cat1–8* is found throughout the range of cultivated sorghum, although it is most common in northwestern Africa (Fig. 3). Despite its high frequency in cultivars of northwestern Africa, it is absent from the wild race arundinaceum of this region. It is present in wild collections from northeastern Africa (Chad and Ethiopia). This would indicate that the cultivars of the northwest are not as genetically similar to the indigenous wild sorghum (race arundinaceum) as they are to the wild sorghum of northeastern Africa. This genetic isolation of race arundinaceum of northwestern Africa is further supported by *Cat1–9*, which occurs in this race at a high frequency, but is absent in sympatric cultivars. Other alleles with distributions suggesting that the cultivars of the northwest are more closely related to wild forms from the northeast than they are to race arundinaceum of the northwest include *Frk2–3* and *Frk2–8* (Fig. 4), as well as *Mdh3–2* (Aldrich 1991).

Alleles at several loci show distributions that are suggestive of localized introgression between cultivated and wild populations. This pattern can be seen clearly for *Adh2–2* (Aldrich 1991) and *Cat1–2* (Fig. 3), which occur in both wild and cultivated populations from Malawi, but are found in no other countries. Introgression may also help explain the distribution of *Aco1–8*, which is found in both wild and cultivated populations from northeastern Africa (Fig. 1). This allele may have arisen in the wild sorghum of this region and subsequently spread to the cultigen by introgression. The distribution of *Aco1–2*, which is found in both wild and cultivated populations of central Africa (Fig. 1), also suggests localized introgression. Similarly, *Adh1–5L* is restricted to the wild and cultivated populations of South Africa (Aldrich 1991). Introgression between wild and cultivated populations may be an important factor in the evolution of *Sorghum bicolor*.

Geographic and racial origin of cultivated sorghum

Our data suggest that cultivated sorghum was initially domesticated in northeast-central Africa. These results are consistent with those of Ollitrault et al. (1989a), which indicate that cultivated sorghum originated from a primitive bicolor race in eastern, central Africa. One can hypothesize that other domestication events occurred independently of the proposed primary domestication in northeast-central Africa. However, if the wild sorghum from northwest or southern Africa had contributed substantially to the genetic composition of the extant crop's gene pool, then the wild sorghum of these regions should show the greatest genetic similarity to the cultivars of the same region, which is not supported by our data (Figs. 5, 6). This does not rule out small-scale independent domestications in these regions, but it is unlikely that northwest or southern Africa were as active in the domestication of sorghum as was northeast-central Africa.

Our data also support the view (Harlan and Stemler 1976) that cultivated sorghum originally was selected from a complex consisting primarily of races verticilliflorum and/or aethiopicum. In regions where races arundinaceum and virgatum can be identified as biologically distinct entities, i.e., in the genetically 'purest' sections of their range (Fig. 7), they are genetically distinct from most cultivated collections (Figs. 5, 6). Thus, races aethiopicum and arundinaceum probably have not contributed as greatly to the domestication event that produced the majority of cultivated sorghum as did races aethiopicum and verticilliflorum, although they may have contributed to the character of the extant crop through introgression (Doggett 1965).

It should be noted, however, that our seed collections could carry alleles introduced from other lines due to improper control of outcrossing during maintenance at the seed bank. Such occurrences might explain the genetic similarities among some of our collections independently of ancestor-descendent relationships, and our results should be viewed in light of this.

Conservation

The development of useful and efficient sampling strategies for the conservation of crop genetic resources requires an understanding of how the diversity is distributed taxonomically and geographically in the species (Brown 1978; Chapman 1989; Marshall 1990). Molecular markers can be used to estimate which portions of the primary gene pool carry the greatest amount of genetic diversity and endemism. Nonetheless, results should be viewed with caution since isozyme variation may not consistently reflect actual patterns of genetic diversity elsewhere in the genome. In addition, it is uncertain to what extent the seed bank lines utilized here are representative of the genetic variation maintained in natural populations.

Our work does reveal certain trends which might not be evident in smaller scale studies and which deserve further attention. Morden et al. (1990) previously identified race virgatum as the genetically most distinct of the wild taxa in sorghum. Our data support this and conclude that race virgatum, along with a geographic portion of race arundinaceum in northwest Africa, contain genetic variation not represented in the cultivated sorghum. Thus, these are potential new sources of variation for sorghum improvement. Our study also shows that diversity in the cultivars is greatest in northern and central Africa; therefore these regions might be sampled more heavily than southern Africa or Asia in conservation efforts.

Acknowledgements. Grateful acknowledgement is given to Glenn Furnier, Ed Cushing, and two anonymous reviewers for helpful comments on the manuscript. This research was supported in part by a grant from Pioneer Hi-Bred International of Johnston, Iowa, and by the U.S. Department of Agriculture, Grant no. 86-CRCR-1-2161.

References

Aldrich PR (1991) Molecular population genetics of cultivated and wild sorghum. MSc thesis, University of Minnesota, St. Paul, Minn.

Brown AHD (1978) Isozymes, plant population genetic structure and genetic conservation. Theor Appl Genet 52:145–157

Cardy BJ, Stuber CW, Wendel JF, Goodman MM (1983) Techniques for starch gel electrophoresis of enzymes from maize (*Zea mays* L.) Inst Stat Mimeo Ser No 1317, N.C. State Univ, Raleigh, N.C.

Chapman CGD (1989) Collection strategies for the wild relatives of field crops. In: Brown AHD, Frankel OH, Marshall DR, Williams JT (eds) The use of plant genetic resources. Cambridge Univ Press, Cambridge, pp 263–279

Clayton JW, Tretiak DN (1972) Amine-citrate buffers for pH control in starch gel electrophoresis. J Fish Res Board Can 29:1169–1172

Crawford DJ (1990) Plant molecular systematics: macromolecular approaches. John Wiley and Sons, New York

de Wet JMJ (1978) Systematics and evolution of *Sorghum* sect. *Sorghum* (Gramineae). Am J Bot 65:477–484

de Wet JMJ, Harlan JR, Price EG (1970) Origin of variability in the Spontanea complex of *Sorghum bicolor*. Am J Bot 57: 704–707

de Wet JMJ, Harlan JR, Price EG (1976) Variability in *Sorghum bicolor*. In: Harlan JR, de Wet JMJ, Stemler ABL (eds) Origins of African plant domestication. Mouton Press, The Hague, pp 453–463

Doebley JF (1989) Isozymic evidence and the evolution of crop plants. In: Soltis DE, Soltis PS (eds) Isozymes in plant biology. Dioscorides Press, Portland, Ore., pp 165–191

Doggett H (1965) The development of cultivated sorghums. In: Hutchinson J (ed) Crop plant evolution. Cambridge Univ Press, London, pp 50–69

Frankel OH, Soulé ME (1981) Conservation and evolution. Cambridge Univ Press, Cambridge

Gottlieb LD (1981) Electrophoretic evidence and plant populations. Prog Phytochem 7:1–46

Harlan JR, de Wet JMJ (1971) Toward a rational classification of cultivated plants. Taxon 20:509–517

Harlan JR, de Wet JMJ (1972) A simple classification of cultivated sorghum. Crop Sci 2:172–176

Harlan JR, Stemler ABL (1976) The races of sorghum in Africa. In: Harlan JR, de Wet JMJ, Stemler ABL (eds) Origins of African plant domestication. Mouton Press, The Hague, pp 465–478

Marshall DR (1990) Crop genetic resources: Current and emerging issues. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding, and genetic resources. Sinauer Assoc, Sunderland, Mass., pp 367–388

Meizel S, Markert CL (1967) Malate dehydrogenase isozymes of the marine snail *Ilyanassa obsoleta*. Arch Biochem Biophys 122:753–765

Morden CW, Doebley JF, Schertz KF (1987) A manual of techniques for starch gel electrophoresis of Sorghum isozymes. Tex. Agric Exp Stn, MP-1635, College Station

Morden CW, Doebley JF, Schertz KF (1989) Allozyme variation in old world races of *Sorghum bicolor* (Poaceae). Am J Bot 76:247–255

Morden CW, Doebley JF, Schertz KF (1990) Allozyme variation among the spontaneous species of *Sorghum* section *Sorghum* (Poaceae). Theor Appl Genet 80:296–304

Ollitrault P, Arnaud M, Chanterau J (1989a) Polymorphisme enzymatique des sorghos. II. Organisation genetique et evolutive des sorghos cultives. Agron Trop 44:211–222

Ollitrault P, Escoute J, Noyer JL (1989b) Polymorphisme enzymatique des sorghos. I. Description de 11 systemes enzymatiques determinisme et liaisons genetiques. Agron Trop 44:203–210

Wendel JF, Stuber CW (1983) Plant isozymes: enzymes studied and buffer systems for their electrophoretic resolution in starch gels. Isozyme Bull 17:4–11

Wright S (1978) Evolution and the genetics of populations, vol 4. Variability within and among natural populations. University of Chicago Press, Chicago