Theor Appl Genet (1990) 80:296–304

Allozyme variation among the spontaneous species of *Sorghum* section Sorghum (Poaceae)

C. W. Morden [1], J. Doebley [2,*] and K. F. Schertz [3]

[1] Department of Biology, Indiana University, Bloomington, IN 47405, USA
[2] Department of Plant Biology, University of Minnesota, St. Paul, MN 55108, USA
[3] USDA-ARS in cooperation with the Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77843, USA

Received November 15, 1989; Accepted April 3, 1990
Communicated by A. R. Hallauer

Summary. A survey of allozyme variation among the spontaneous taxa of *Sorghum* section Sorghum was undertaken. Eight plants each of 90 accessions representing the diploid *S. bicolor* (ssp. *arundinaceum* and *drummondii*) and the tetraploids *S. almum* and *S. halepense* were analyzed for 17 enzyme systems encoded by 30 loci. Low levels of variation were found within and among accessions, although there was more variation than is typical of inbreeding species. We found an average of 3.2 alleles per locus in ssp. *arundinaceum*, with a mean expected heterozygosity for the accessions of 0.034 and total panmictic heterozygosity of 0.154. An analysis of the apportionment of genetic variation among accessions of ssp. *arundinaceum* indicated that 26% of the variation occurs within accessions and 74% among accessions. Cultivated sorghum contains far less allozymic variation than ssp. *arundinaceum*, its presumed progenitor. This is consistent with the prediction that cultivated sorghum experienced a loss of genetic variation during domestication. For the most part, cultivated sorghum contains a subset of the allozymes found in ssp. *arundinaceum*. Principal component analysis revealed continuous variation among the accessions and geographic regions, with accessions failing to segregate into discrete clusters. However, accessions of race virgatum of ssp. *arundinaceum* occupied one end of the continuum and were, in that sense, distinguished from the other accessions. Similarly, most accessions of *S. halepense* and *S. almum* occupied the central portion of the continuum. The allozymic data presented here are consistent with the hypothesized origin of *S. halepense* via autopolyploidy or segmental allopolyploidy.

* To whom correspondence and reprint requests should be addressed

Key words: *Sorghum* – Isozyme – Genetic diversity – Evolution – Systematics

Introduction

Allozyme analysis provides a powerful means of assessing the levels of genetic variation in plant populations (Gottlieb 1981). This method has been especially useful in addressing questions surrounding populations genetic structure and genetic conservation (Brown 1978). Furthermore, it has proven of immense value in delimiting the systematic relationships among plant species (Crawford 1983). It is the purpose of the present paper to survey allozymic variation among the spontaneous taxa of *Sorghum* section Sorghum and to employ this variation to assess systematic relationships among these taxa. We also compare levels of allozymic variation in the spontaneous taxa of section Sorghum to that in cultivated sorghum [*S. bicolor* (L.) Moench ssp. *bicolor*]. This work builds upon our previous paper (Morden et al. 1989), in which we examined the extent and nature of allozymic variation in the Old World races of cultivated sorghum.

Taxonomic background

The genus *Sorghum* is native to tropical and some temperate regions throughout much of the world. The species of *Sorghum* have been divided into five subgenera (Garber 1950; Celarier 1958b). One of these subgenera, *Eu-Sorghum*, contains cultivated sorghum and its nearest wild relatives. Recent authors (Doggett 1970; de Wet 1978) have referred to this subgenus as *Sorghum* section Sorghum, and we will follow this nomenclature in this paper. According to the most recent taxonomic treat-

ment (de Wet 1978), section Sorghum possesses three natural species: (1) *Sorghum bicolor*, an annual diploid (n=10) including ssp. *bicolor* (cultivated sorghum) and its wild or weedy relatives, ssp. *arundinaceum* (Desv.) de Wet et Harlan and spp. *drummondii* (Steud.) de Wet; (2) *S. halpense* (L.) Pers., a rhizomatous, tetraploid perennial native to southern Eurasia and eastward to India, and naturalized throughout the warm temperate regions of the world; and (3) *S. propinguum* (Kunth) Hitchc., a rhizomatous, diploid perennial of India and eastward to the islands of southeastern Asia (de Wet 1978). An additional species, *S. almum* Parodi, is a synthetic, tetraploid species resulting from crosses of cultivated sorghum and *S. halepense*. The latter species was treated as part of *S. halepense* by de Wet (1978).

De Wet's (1978) treatment of section Sorghum differs radically from that of Snowden (1936, 1955). The latter author recognized 31 species of cultivated sorghum and 17 species of wild sorghum, excluding *S. halepense* and *S. propinquum*. Several studies have shown that Snowden's species intergrade morphologically and are interfertile (Celarier 1958a, b; de Wet et al. 1970). As such, combining these 48 species into a single, morphologically diverse species, *S. bicolor*, with three subspecies seems preferable.

Most of the accessions assayed in this paper belong to *Sorghum bicolor* ssp. *arundinaceum*. This subspecies possesses considerable morphological and ecological diversity, and has been divided into four eco-geographic races (de Wet 1978). (1) Race arundinaceum is a robust grass with broad leaves and a large, pendulous inflorescence (when mature). It is predominantly found in tropical west Africa, although populations have been found in central Africa and as far south as South Africa. (2) Race verticilliflorum tends to be somewhat smaller than race arundinaceum, and its inflorescence is not pendulous when mature. Race verticilliflorum is a broadly distributed over most regions of Africa and grades into race arundinaceum in the tropical forests. (3) Race aethiopicum is a smaller desert grass with narrow leaves and typically smaller inflorescence than the previous two races. This race has large ovate-lanceolate spikelets that are usually densely tomentose. Race aethiopicum is found in eastern Africa from Chad to Sudan and as far north as Ethiopia and Egypt. (4) Race virgatum has narrowly linear leaf blades and inflorescence branches that are mostly erect. Populations of race virgatum are found along stream banks and irrigation ditches of northeastern Africa.

The present study also includes several accessions of *Sorghum bicolor* ssp. *drummondii*, a weedy form that occurs wherever cultivated sorghum and ssp. *arundinaceum* grow sympatrically (de Wet 1978). The fact that ssp. *drummondii* is morphologically intermediate between the other two subspecies suggests that ssp. *drummondii* is of hybrid origin. The frequent occurrence of ssp. *drummondii* in fields of cultivated sorghum is consistent with this proposed mode of origin (de Wet et al. 1970).

The origin of the rhizomatous tetraploid, *S. halepense*, has been extensively debated over the past several decades, and various models for its origin have been proposed. It is generally believed that *S. halepense* arose as a segmental allotetraploid derived from the cross of two diploid (n=10) "species" (Duara and Stebbins 1952; Hadley 1953; Endrizzi 1957; Celarier 1958a). According to some authors, the two species involved were the rhizomatous perennial, *S. propinquum*, and the annual, *S. bicolor* ssp. *arundinaceum* (Celarier 1958a; Doggett 1970). An alternative hypothesis was advanced by Bhatti et al. (1960), who observed small, inconspicuous rhizomes on plants of ssp. *arundinaceum* race virgatum. They noted that, when these plants were crossed to cultivated sorghum and the progeny's chromosome number was doubled by a colchicine treatment, strongly rhizomatous plants were produced. On this evidence, they suggested that *S. halepense* might have arisen from such a hybrid.

Sorghum almum is almost certainly a recent hybrid between *S. halepense* and *S. bicolor* (Doggett 1970). This species, which arose in Argentina, is cytologically similar to n=20 hybrids of *S. halepense* and *S. bicolor* (Celarier 1958a).

Materials and methods

Eight plants from each of 90 accessions representing the spontaneous taxa of *Sorghum* section Sorghum from Africa, India, and Thailand were analyzed (Table 1). Caryopses were obtained from the International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) and the USDA Regional Plant Introduction Station, Georgia (RPIS). Accessions for each taxon were selected to represent the complete geographic range of the taxon, with the constraint that accessions for some regions were not available from the germ plasm banks. The geographic locations were not available for four accessions (Table 1); however, these were included in the study because there were few available accessions for these taxa. RPIS record 's indicate that PI302232 and PI302233 were collected in the USSR, which is clearly incorrect. Attempts to clarify their origin have been unsuccessful. Another accession (PI302232) was erroneously identified as belonging to race virgatum in the RPIS records. We included this accession with race verticilliflorum based on our own determination of its identity. The 90 accessions are divided among the taxa as follows: *S. bicolor* ssp. *arundinaceum* (71 accessions), *S. bicolor* ssp. *drummondii* (9 accessions), *S. almum* (3 accessions), and *S. halepense* (7 accessions). Unfortunately, adequate samples of *S. propinquum* were unavailable for this study. The accessions have been increased an unknown number of times since they were first collected. Because of inadequate germination, we were unable to obtain eight seedlings for three accessions (IS18821, IS21340, and PI255738). These samples were subsequently excluded from analyses where balanced sampling was desired. Voucher specimens for all accessions have been collected, identified, and deposited at the S. M. Tracy Herbarium, Texas A & M University (TAES).

Table 1. Taxonomic identity, country of origin, and identification number for the 90 accessions analyzed[a]

Sorghum bicolor ssp. arundinaceum
Race aethiopicum – CHAD: IS18909. EGYPT: IS18820, IS18821. ETHIOPIA: PI302105. SUDAN: IS14485, IS14564, IS14564-1, IS14564-2, IS14567, IS18819

Race arundinaceum – ANGOLA: IS14216. BENIN: IS18876. EGYPT: IS18812. ETHIOPIA: PI302118. GHANA: IS18883. IVORY COAST: IS18825, IS18826, IS18881. KENYA: IS21340. LESOTHO: IS14467. MALAWI: IS14359. NIGERIA: IS18878, IS18879. SOUTH AFRICA: IS14301, IS18797, PI302111. SWAZILAND: IS14315. ZIMBABWE: PI156549, PI225905

Race verticilliflorum – ANGOLA: IS14219, IS14238, IS14257. CHAD: IS18867, IS18869. EGYPT: IS18808, IS18814, IS18815. ETHIOPIA: IS14583, IS14719, IS14766. KENYA: IS14569, IS21153, PI213900. MALAWI: IS14357, IS14414, IS18833. SOUTH AFRICA: IS18798, PI185573, PI185574, PI208190, PI300118, PI300119, PI300120. SUDAN: IS18908, IS18916, IS18928. UGANDA: IS14505, IS14515, IS14518. ZAIRE: PI247723. ZIMBABWE: PI213901. UNKNOWN: PI302232

Race virgatum – EGYPT: IS18809, IS18810, IS18813, IS18900. SUDAN: IS18816, IS18817. UGANDA: IS18803, IS18806. UNKNOWN: PI302233

Sorghum bicolor ssp. drummondii
ANGOLA: IS14240-1. ETHIOPIA: IS14717-1. INDIA: PI267331. KENYA: PI226906. NIGERIA: IS18836, PI186570. SOUTH AFRICA: PI199869, PI302113. SUDAN: IS18838

Sorghum almum
THAILAND: IS18853. UNKNOWN: IS18852, IS18894

Sorghum halepense
ANGOLA: IS14212, IS14241. INDIA: IS18843, IS18845. IRAN: PI228364. SUDAN: IS18927. TURKEY: PI255738

[a] Seeds from accessions with IS and PI prefixes were obtained from the International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) and the USDA Regional Plant Introduction Station, Georgia (RPIS), respectively

Growth of the seedlings and extraction of enzymes for electrophoresis was done as previously described (Morden et al. 1987, 1989). Etiolated seedlings were submerged in water for 15–24 h prior to extraction, to enhance expression of anaerobic alcohol dehydrogenase isozymes. Extracts were absorbed into filter paper wicks (Whatman no. 3, 2 × 11 mm) and stored at −80 °C until electrophoresed. Electrophoresis of enzymes was carried out on four buffer systems described by Morden et al. (1987, 1989). Following electrophoresis, the gels were trimmed and the slab anodal to the origin was sliced and stained for enzymatic activity. Seventeen enzyme systems encoded by 30 loci were analyzed (Table 2). Loci were identified in the same manner as that described by Morden et al. (1987, 1989). All gels were scored initially by the first author, and the scoring was verified by the second author. Because of potential variation in electrophoretic conditions, alleles from different accessions with similar migration were rerun in side-by-side tests to verify scoring. The cultivar BTx623 was included on all gels as a standard. The number of copies of each allele at heterozygous loci in the tetraploid (*S. halepense*) was determined by scoring differences in band intensities.

Table 2. Enzyme systems assayed, the number of loci scored and the gel buffer systems on which they were assayed. A complete description of assay conditions is given in Morden et al. (1987)

Enzyme	Abbreviation	No. of loci	Buffer system[a]
Aconitase	ACO	2	T
Adenylate kinase	ADK	1	M, 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	T
Phosphoglucomutase	PGM	1	T
6-Phosphogluconate dehydrogenase	6-PGD	2	M
Shikimate dehydrogenase	SAD	1	M
Triosephosphate isomerase	TPI	2	L

[a] Buffer systems: L, lithium hydroxide-borate; M, morpholine-citrate; N, sodium hydroxide-borate; T, tris-citrate.

Expected panmictic heterozygosity was calculated for accessions (H_s), country (H_c), and total sample (H_t) for each locus as follows:

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

where x_i is the frequency of the i^{th} allele. Because sampling for loci and within accessions was balanced, pooling over loci and accessions was unweighted. Pooling over subspecies and country was weighted by the number of accessions assayed for each subspecies or country. Relative genetic differentiation (G_{st} or G_{sc}) was calculated by the methods of Nei (1973, 1977).

The degree of genetic similarity among accessions, races, subspecies, and species was assessed using Nei's genetic identify, I (Nei 1972). In accordance with the recommendation of M. Nei (personal communication), I was calculated for the tetraploid (*S. halepense*) in precisely the same manner as for the diploids. Principal component analysis for the accessions was performed using the variance-covariance matrix of allele frequencies.

Results

Genetic diversity within the taxa

Of the 30 loci examined, 4 (*Aat1, Idh1, Mdh2,* and *Tpi2*) were monomorphic among all accessions in all taxa. The 26 polymorphic loci differed considerably in their degree of variation from weakly polymorphic loci such as *Gdh1* and *Mmm1* to more variable loci such as *Aat3* and *Frk2*. Table 3 lists the number of accessions in which each allele was present and the frequency of each allele (for the variable loci) within the four taxa studied. Among the 30

Table 3. List of alleles observed in *Sorghum* sect. Sorghum and their frequency within each of the taxa analyzed. *N* is the number of accessions in which the allele was present among the 90 accessions analyzed. ARUN = *S. bicolor* ssp. *arundinaceum*; DRUM = *S. bicolor* ssp. *drummondii*; ALMU = *S. almum*; and HALE = *S. halepense*

Locus-allele	*N*	ARUN	DRUM	ALMU	HALE
Aat2-1	26	0.194	0.000	0.917	0.478
5	78	0.789	1.000	0.083	0.522
5L	2	0.015	0.000	0.000	0.000
8	1	0.003	0.000	0.000	0.000
Aat3-1	2	0.011	0.000	0.000	0.010
2	2	0.017	0.000	0.000	0.000
3	1	0.003	0.000	0.000	0.000
4	1	0.003	0.000	0.000	0.000
5	89	0.954	1.000	1.000	0.989
9	1	0.013	0.000	0.000	0.000
Aco1-1	2	0.020	0.000	0.000	0.000
2	5	0.026	0.000	0.000	0.021
5	82	0.831	0.831	0.917	0.656
8	20	0.120	0.169	0.083	0.323
9	1	0.003	0.000	0.000	0.000
Aco2-3	1	0.013	0.000	0.000	0.000
5	86	0.910	0.762	0.812	0.739
6	5	0.017	0.006	0.000	0.208
7	2	0.014	0.000	0.000	0.000
8	15	0.047	0.231	0.188	0.052
Adh1-3	24	0.147	0.175	0.063	0.024
5	82	0.792	0.787	0.937	0.934
5L	2	0.029	0.000	0.000	0.000
6L	2	0.015	0.000	0.000	0.000
7	3	0.017	0.000	0.000	0.042
N	1	0.000	0.037	0.000	0.000
Adh2-1	2	0.006	0.043	0.000	0.000
2	2	0.017	0.000	0.000	0.000
5	89	0.960	0.956	1.000	1.000
N	2	0.017	0.000	0.000	0.000
Adh3-1	1	0.000	0.012	0.000	0.000
5	89	0.937	0.987	0.917	1.000
5N	2	0.017	0.000	0.000	0.000
8	10	0.046	0.000	0.083	0.000
Adk1-2	33	0.247	0.037	0.229	0.589
5	80	0.753	0.962	0.771	0.411
Cat1-2	2	0.007	0.000	0.000	0.073
5	81	0.845	0.856	1.000	0.927
8	3	0.021	0.044	0.000	0.000
9	11	0.127	0.100	0.000	0.000
Enp1-0.5	7	0.000	0.000	0.292	0.260
2	74	0.754	0.631	0.708	0.573
5	25	0.182	0.369	0.000	0.167
N	6	0.064	0.000	0.000	0.000
Frk1-4	4	0.019	0.137	0.000	0.000
5	88	0.966	0.862	1.000	1.000
7	2	0.015	0.000	0.000	0.000
Frk2-3	15	0.046	0.169	0.000	0.187
4	3	0.016	0.000	0.000	0.000
5	79	0.773	0.831	1.000	0.802
5N	1	0.015	0.000	0.000	0.000
6	3	0.018	0.000	0.000	0.000
7	10	0.088	0.000	0.000	0.000
8	5	0.034	0.000	0.000	0.010
N	1	0.011	0.000	0.000	0.000
Gdh1-4	10	0.063	0.000	0.000	0.000
5	88	0.937	1.000	1.000	1.000
Gpt1-4	1	0.005	0.000	0.000	0.000
5	87	0.950	1.000	1.000	1.000
6	3	0.044	0.000	0.000	0.000
Gpt2-2	20	0.178	0.087	0.000	0.219
5	79	0.807	0.912	1.000	0.781
N	1	0.014	0.000	0.000	0.000
Idh2-5	84	0.867	0.900	1.000	0.989
7	14	0.133	0.100	0.000	0.010
Mdh1-4	2	0.015	0.000	0.000	0.000
5	88	0.962	1.000	1.000	0.979
8	2	0.009	0.000	0.000	0.021
9	1	0.014	0.000	0.000	0.000
Mdh3-2	2	0.009	0.000	0.000	0.000
5	89	0.961	1.000	1.000	1.000
8	3	0.029	0.000	0.000	0.000
Mmm1-M	89	0.949	1.000	1.000	1.000
m	9	0.051	0.000	0.000	0.000
6Pgd1-3	1	0.003	0.000	0.000	0.000
5	88	0.950	1.000	0.958	0.958
7	5	0.029	0.000	0.000	0.021
8	3	0.003	0.000	0.042	0.020
10	1	0.015	0.000	0.000	0.000
6Pgd2-2	1	0.006	0.000	0.000	0.000
5	88	0.897	1.000	0.812	0.760
7	4	0.000	0.000	0.042	0.182
8	3	0.000	0.000	0.146	0.036
9	11	0.097	0.000	0.000	0.021
Pgi1-3	1	0.002	0.000	0.000	0.000
4	2	0.029	0.000	0.000	0.000
5	88	0.967	1.000	1.000	1.000
N	1	0.002	0.000	0.000	0.000
Pgi2-1	2	0.007	0.000	0.000	0.000
2	1	0.013	0.000	0.000	0.000
5	90	0.979	1.000	1.000	0.989
7	2	0.001	0.000	0.000	0.010
Pgm1-5	90	1.000	1.000	0.708	0.917
8	5	0.000	0.000	0.292	0.083
Sad1-5	89	0.897	1.000	1.000	0.958
7	12	0.103	0.000	0.000	0.042
Tpi1-5	90	0.996	1.000	1.000	1.000
8	1	0.005	0.000	0.000	0.000

loci, we detected 102 alleles or an average of 3.40 alleles per locus. For each locus, there was a single predominant allele found in 74 or more of the 90 accessions. The predominant allele for each locus was the same in all four taxa for all loci except *Adk1* and *Enp1*. Thus, the taxa and populations tend to be differentiated from one another mostly by the presence/absence of low-frequency alleles. Most of these alleles occur at frequencies below 0.05, and 24 of the low-frequency alleles were restricted to a single accession.

For subspecies *arundinaceum*, a total of 96 different alleles was present among the 30 loci examined (Table 3), for an average of 3.2 alleles per locus (Table 4). For ssp. *drummondii*, *S. halepense*, and *S. almum*, we found fewer alleles than in ssp. *arundinaceum* as expected because of the smaller number of accessions assayed (Table 3). There were only six alleles found in these three taxa that were not found in ssp. *arundinaceum*. Two of these (*Adh1-N* and *Adh3-1*) are restricted to ssp. *drummondii*, while the remaining four (*Enp1-0.5*, *6Pgd2-7*, *6Pgd2-8*, and *Pgm1-8*) occurred in both *S. halepense* and *S. almum*. The mean proportion of polymorphic loci per population (PLP) is 0.19, 0.08, 0.18, and 0.18 in ssp. *arundinaceum*, ssp. *drummondii*, *S. halepense*, and *S. almum*, respectively. Curiously, PLP for the tetraploids is not higher than it is for ssp. *arundinaceum*. Overall, 25 of the 30 loci (83%) were variable in ssp. *arundinaceum*; however, using the criterion that a locus is polymorphic if the frequency of the most common allele is less than 0.99 (Gottlieb 1981), 24 of 30 (80%) loci examined were polymorphic. Again, because of more limited sampling, a smaller portion of the loci was polymorphic in the other three taxa.

Mean expected heterozygosity for accessions of ssp. *arundinaceum* (H_s) varied from 0.0 for the five monomorphic loci to 0.114 for *Frk2*, with an average over all loci of 0.034 (Table 4). Total panmictic heterozygosity (H_t) for this subspecies varied from 0.0 to 0.423 at *Frk2* (Table 4) and averaged 0.154 over all loci. Mean expected heterozygosity in ssp. *drummondii* (0.029) is similar to that in ssp. *arundinaceum*. H_t for ssp. *drummondii* (0.092) may not be comparable to that for ssp. *arundinaceum* because of the large difference in the numbers of accessions that we assayed for each.

Relative genetic differentiation (G_{st}) as defined by Nei (1973, 1977) equals $(H_t - H_s)/H_t$. G_{st} provides a measure of the proportion of the total genetic variation distributed among subpopulation (accessions). In ssp. *arundinaceum*, G_{st} ranges from 0.310 for *Tpi1* to 0.927 for *Gpt1*, with a weighted mean among all polymorphic loci of 0.744 (Table 4). G_{st} for ssp. *drummondii* ranges from 0.114 for *Adh3* to 1.000 for *Idh2*, with an average of 0.568. These values indicate that approximately 74% of the variation in ssp. *arundinaceum* and 57% of the variation in ssp. *drummondii* occurs among accessions and the remaining variation occurs within accessions. Large values (>0.5) for G_{st} are typical of self-pollinating species.

Table 4. Measures of genetic diversity within and among accessions of *Sorghum bicolor* ssp. *arundinaceum* for 25 variable loci. A = number of alleles per locus; $\bar{H}_s$ = expected mean heterozygosity; H_t = total panmictic heterozygosity; G_{st} = relative genetic differentiation

Locus	A	$\bar{H}_s$	H_t	G_{st}
Aat2	4	0.071	0.355	0.800
Aat3	6	0.018	0.087	0.797
Aco1	5	0.074	0.287	0.744
Aco2	5	0.048	0.165	0.710
Adh1	5	0.088	0.377	0.768
Adh2	4	0.023	0.076	0.693
Adh3	3	0.052	0.116	0.546
Adk1	2	0.083	0.365	0.773
Cat1	4	0.029	0.263	0.891
Enp1	3	0.059	0.410	0.856
Frk1	3	0.012	0.091	0.874
Frk2	8	0.114	0.423	0.730
Gdh1	2	0.051	0.140	0.638
Gpt1	3	0.007	0.092	0.927
Gpt2	3	0.039	0.310	0.874
Idh2	2	0.040	0.225	0.823
Mdh1	4	0.008	0.098	0.915
Mdh3	3	0.014	0.073	0.812
Mmm1	2	0.045	0.095	0.530
6Pgd1	5	0.019	0.106	0.819
6Pgd2	3	0.057	0.181	0.685
Pgi1	4	0.006	0.062	0.900
Pgi2	4	0.015	0.040	0.638
Sad1	2	0.037	0.179	0.794
Tpi1	2	0.006	0.009	0.310
MEAN	3.2	0.034	0.154	0.744

Nei's (1972) genetic identity (I) provides a measure of genetic similarity among populations or taxa. We used this as a measure of similarity among accessions both within and among the races of ssp. *arundinaceum* (Table 5). Nei's I among the races of ssp. *arundinaceum* ranges from 0.862 to 0.895 (Table 5), which is slightly lower than that expected among subspecific categories of plant species (Crawford 1983). If accessions of a single race are more similar to one another than they are to accessions of other races, then I among accessions within races should be higher than that among races. This was marginally true for races verticilliforum and virgatum; however, accessions of race aethiopicum showed greater similarity to accessions of the other races than they did to accessions of their own race (Table 5). This suggests that isozymic similarity is not well-correlated with taxonomic identity for ssp. *arundinaceum*.

Genetic identities among all the spontaneous taxa indicate that ssp. *arundinaceum* has a comparatively low level of similarity to the other taxa (Table 6). The mean value of I between accessions of ssp. *drummondii* and *S. halepense* (0.915) is higher than expected in that this

Table 5. The average Nei's coefficients of genetic identity (*I*) among races of *Sorghum bicolor* ssp. *arundinaceum*. The range for each comparison is shown parenthetically

Race	1.	2.	3.	4.
1. aethiopicum	0.870 (0.725–1.000)			
2. arundinaceum	0.873 (0.667–1.000)	0.894 (0.781–1.000)		
3.verticilliflorum	0.892 (0.667–1.000)	0.895 (0.781–1.000)	0.904 (0.794–1.000)	
4. virgatum	0.862 (0.651–1.000)	0.863 (0.631–1.000)	0.880 (0.651–1.000)	0.888 (0.696–1.000)

Table 6. The average Nei's coefficients of genetic identity (*I*) among spontaneous taxa of *Sorghum* section Sorghum. The range for each combination is shown parenthetically. ALMU = *S. almum*; HALE = *S. halepense*; DRUM = *S. bicolor* ssp. *drummondii*; and ARUN = *S. bicolor* ssp. *arundinaceum*

Taxon	ALMU	HALE	DRUM	ARUN
ALMU	0.983 (0.980–1.000)			
HALE	0.926 (0.800–1.000)	0.910 (0.800–1.000)		
DRUM	0.927 (0.855–1.000)	0.915 (0.800–1.000)	0.929 (0.855–1.000)	
ARUN	0.876 (0.631–1.000)	0.876 (0.631–1.000)	0.880 (0.631–1.000)	0.874 (0.631–1.000)

subspecies is thought to have been derived from hybrids between cultivated sorghum and ssp. *arundinaceum*. The mean values of *I* among accessions within *S. halepense* (0.910) and ssp. *drummondii* (0.929) are equivalent to those that have been reported for other species or subspecies of flowering plants (Crawford 1983). In contrast, the mean value of *I* among accessions of ssp. *arundinaceum* (0.874) is lower than expected and indicates that the accessions of this subspecies are rather heterogeneous.

Geographic distribution of genetic diversity

Genetic diversity statistics for countries of origin for the accessions of ssp. *arundinaceum* are presented in Table 7. The mean number of alleles per accession (APA) in a country was generally higher in countries from northeastern Africa (notably Chad and Egypt) and lower in countries from southern and western Africa (Angola, Benin, Ghana, and the Ivory Coast). The mean proportion of polymorphic loci per accession (PLP) shows a similar trend. PLP was highest for Egypt and Chad (19.6% and 18.9% of the loci were polymorphic, respectively) and lowest for Benin, Ghana, and the Ivory Coast (no polymorphic loci). Mean expected heterozygosity for the accessions of each country (H_s) was again higher in

Table 7. Measures of genetic diversity within and among countries of origin for accessions of *Sorghum bicolor* ssp. *arundinaceum*. N = number of accessions; APA = average number of alleles per accession within each country; PLP = mean proportion of polymorphic loci per accession; $\bar{H}_s$ = mean expected heterozygosity per country; H_c = total panmictic heterozygosity per country; G_{sc} = relative genetic differentiation among accessions within countries

Country	N	APA	PLP[a]	$\bar{H}_s$	H_c	G_{sc}
Angola	5	30.8	0.025	0.003	0.112	0.974
Benin	1	30.0	0.000	0.000	0.000	–
Chad	3	48	0.189	0.080	0.132	0.397
Egypt	9	36.1	0.196	0.080	0.169	0.523
Ethiopia	5	31.8	0.053	0.017	0.071	0.769
Ghana	1	30.0	0.000	0.000	0.000	–
Ivory Coast	3	30.0	0.000	0.000	0.044	1.000
Kenya	2	32.5	0.083	0.022	0.068	0.673
Lesotho	1	33.0	0.100	0.015	0.015	–
Malawi	4	34.0	0.133	0.049	0.112	0.561
Nigeria	2	33.5	0.117	0.017	0.035	0.506
South Africa	10	31.6	0.057	0.021	0.098	0.787
Sudan	11	33.5	0.099	0.043	0.113	0.619
Swaziland	1	33.0	0.100	0.034	0.034	–
Uganda	5	32.0	0.060	0.023	0.106	0.779
Zaire	1	33.0	0.100	0.046	0.046	–
Zimbabwe	3	31.0	0.033	0.007	0.061	0.882

[a] 99% criterion

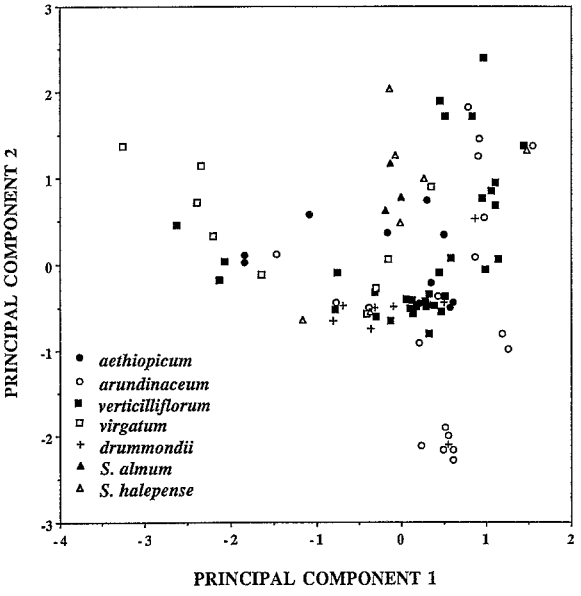

Fig. 1. Graph of the first two components from a principal component analysis based on allele frequencies from 90 accessions of Sorghum section Sorghum

Egypt and Chad and generally lower in more southern and western African countries. Relative genetic differentiation among accessions within countries (G_{sc}) ranges from 0.397 for Chad to 1,000 for the Ivory Coast. Most of these values are rather large (>0.50), suggesting that what genetic variation exists within a country is dispersed among the accessions (Table 7). As such, it would be necessary to sample numerous accessions from any one country to adequately represent the isozyme variation present within that country.

Systematic relationships within section Sorghum

Seventy-four percent of the variation in ssp. *arundinaceum* and 57% in ssp. *drummondii* occurred among accessions. This portion of the genetic variation can be employed to assess the systematic relationships among the accessions and taxa. To do this, we performed a principal component analysis using the variance-covariance matrix of allele frequencies among the accessions. Figure 1 illustrates the relationships among accessions of ssp. *arundinaceum*, ssp. *drummondii*, *S. almum*, and *S. halepense* for the first two principal components. Deliminations between taxa are, for the most part, not clear. Accessions from race verticilliflorum are broadly scattered throughout the figure. Accessions of race aethiopicum are restricted to the central portion of the figure, although they overlap substantially with all other taxa. Accessions of race *arundinaceum* are concentrated in the right side of Fig. 1 while accessions of *S. halepense* and *S. almum* are mostly in the upper central portion of the figure. Accessions of race virgatum tend to be located toward the left side of the figure. Accessions of ssp. *drummondii* are also broadly distributed.

One notable trend in Fig. 1 concerns accessions from Egypt. All accessions except one (IS18806 from Uganda) with principal component 1 values less than −1.4 were collected in Egypt. Included are representatives of each of the four races of ssp. *arundinaceum*. This result suggests that, in *Sorghum* section Sorghum, geographic proximity may in some cases be a better predictor of genetic relatedness than taxonomy.

Discussion

A primary goal of this research was to characterize the amount and distribution of genetic variation in the wild relatives of cultivated sorghum and to compare these data to similar data from other taxa. This aspect of our work has focused on ssp. *arundinaceum*, for which a substantial number (71) of accessions were available. In general, this subspecies contains higher levels of genetic variation than has been found in other inbreeding species (Gottlieb 1981). The proportion of polymorphic loci per population in ssp. *arundinaceum* (0.19) is about four times higher than the average value for inbreeders or about one-half the average value for outcrossers (0.37) (Gottlieb 1981). Similarly, mean expected heterozygosity per population for ssp. *arundinaceum* (0.034) is 30 times greater than the average values for inbreeders, although still less than half the average value reported for outcrossers (Gottlieb 1981). Our value for the average number of alleles per locus (3.2) exceeds the reported value for both inbreeders (2.26) and outcrossers (2.90) (Gottlieb 1981); however, the fact that we have performed more extensive sampling than most authors biases these results. The mean value of G_{st} for ssp. *arundinaceum* is 0.74. This indicates that most genetic variation is distributed among rather than within populations, a fact that is consistent with a mating system that this predominantly self-pollinating.

The variation observed in ssp. *arundinaceum* was not uniformly distributed among loci. Four of the 30 loci were monomorphic among all accessions analyzed. In general, the most variable loci in ssp. *arundinaceum* are also quite variable in ssp. *bicolor* (Morden et al. 1989). *Aco2*, *Adh1*, *Enp1*, and *Frk2* were among the most variable loci in both ssp. *arundinaceum* and ssp. *verticilliflorum*. Similarly, *Mdh1*, *Pgi1*, *Sad1*, and *Tpi1* were among the least variable of the polymorphic loci in both subspecies.

For the most part, the allozymes found in ssp. *bicolor* are also known in ssp. *arundinaceum*. Of the 49 alleles observed in ssp. *bicolor*, 41 were also found in ssp. *arundinaceum*. Six of the remaining alleles are all rare in ssp. *bicolor*, and the other two occur predominantly or strict-

ly in Asia where ssp. *arundinaceum* does not occur (Morden et al. 1989). This pattern of variation is consistent with the view that ssp. *arundinaceum* is ancestral to ssp. *bicolor* (de Wet and Harlan 1971).

The estimates of genetic variation presented in this paper can be compared to those available for the cultivated races of *S. bicolor* ssp. *bicolor* (Morden et al. 1989). By every measure, ssp. *arundinaceum* exceeds ssp. *bicolor* in its levels of genetic variation. The values for mean expected and total panmictic heterozygosity in ssp. *arundinaceum* (0.034 and 0.154, respectively) are considerably higher than those found among landraces of ssp. *bicolor* (0.008 and 0.093, respectively). The average number of alleles per locus among the 71 accessions of ssp. *arundinaceum* (3.2) is much higher than that observed for ssp. *bicolor* (1.8). Also, in our sample of 71 accessions of ssp. *arundinaceum*, 25 of the 30 loci (83%) were polymorphic, whereas only 16 of 27 loci (59%) were polymorphic in a sample of 83 accessions of ssp. *bicolor* (Morden et al. 1989). This reduction in the amount of genetic variation in the cultigen as compared to its presumed wild ancestor agrees with theoretical expectations (Doebley 1989). The domestication process places the incipient crop under strong selection for agronomic traits. This, in effect, produces a genetic "bottleneck" that results in a loss of genetic variation. This loss of genetic variation has been shown to be a general feature of plant domestication (Doebley 1989). For sorghum as for other crops, this feature of the domestication process indicates that the progenitors of crop species represent a pool of new genetic variation for crop improvement.

Another issue surrounding crop domestication is the relative apportionment of genetic variation in crops as compared to their ancestors. Doebley (1989) summarized data for four crops that suggested that genetic variation tends to be more dispersed among categories (populations) in crops as compared to their progenitors. This is also true for sorghum as shown by the values of G_{st} for ssp. *arundinaceum* (0.74) and ssp. *bicolor* (0.91) (Morden et al. 1989). The larger value for ssp. *bicolor* indicates that a greater proportion of the genetic variation is found among accessions rather than within accession of ssp. *bicolor* as compared to ssp. *arundinaceum*. Because the manner in which these two subspecies were sampled may not have been equivalent, these numbers should be viewed cautiously. However, if sampling bias is not a significant factor, then one must conclude that the genetic structure of sorghum has been altered during or after domestication. This topic needs further investigation with more extensive sampling.

The data presented in this paper also provide a measure of the degree of genetic differentiation among the taxa of section Sorghum. At every locus, we observed a single predominant allele in ssp. *arundinaceum*, the frequency of which was greater than 0.75. For all but two (*Adk1* and *Enp1*) of the 30 loci assayed, the other taxa we surveyed possessed these same predominant alleles. This pattern contrasts with that observed among the wild relatives of maize, the teosintes (Doebley et al. 1984). Among these plants, which include three species and several subspecific categories, the occurrence of a single predominant allele at a locus is less frequent and the predominant alleles tend to differ between the taxa. This suggests that the taxa of *Sorghum* section Sorghum are less well differentiated from one another than are the taxa of *Zea*. This conclusion is supported by the fact that the values for Nei's *I* among taxa of *Sorghum* section Sorghum (0.87–0.98) are considerably larger than those among the taxa of *Zea* (0.69–0.98).

Principal component analysis generally failed to separate the accessions into discrete taxonomic categories. This result reinforces our conclusion that the taxa of the section are closely related to one another. As such, our results generally support the taxonomic treatment of the section proposed by de Wet (1978), which recognized only three species in section Sorghum, as opposed to Snowden's (1936, 1955) treatment, which recognized fully 17 wild species in the section. Nevertheless, variation in the frequencies of alleles did not separate the accessions into any recognizable taxonomic groupings, and thus we can neither substantiate nor negate the specific details of the classification for the section proposed by de Wet (1978). Our analyses do not appear to support de Wet's (1978) division of ssp. *arundinaceum* into races aethiopicum, arundinaceum, verticilliflorum, and virgatum, although race virgatum does appears to be allozymically distinct from the other races. This race is also the morphologically most distinct of the four races (de Wet and Huckabay 1967; de Wet et al. 1976; de Wet 1978).

Some authors have proposed that *S. halepense* is the segmental allotetraploid hybrid of *S. propinquum* and *S. bicolor* (Celarier 1958a; Doggett 1970). Alternatively, Bhatti et al. (1960) hypothesized that the parental taxa were ssp. *arundinaceum* race virgatum and cultivated sorghum. Our data provide some information bearing on this question. In addition to the data presented above, we obtained data for a single accession (three plants) of *S. propinquum* after the completion of this study. First, our analyses showed no evidence for fixed heterozygosity in *S. halepense*. This is consistent with a segmental allotetraploid or autoploid origin. Second, four alleles (*Enp1-0.5, 6Pgd2-7, 6Pgd2-8,* and *Pgm1-8*), found repeatedly in accessions of *S. halepense* and *S. almum*, were not found in any accessions of *S. bicolor* (spontaneous or cultivated races) or *S. propinquum*. A more exhaustive survey of ssp. *arundinaceum* and *S. propinquum* may uncover the source of these alleles and shed important new light on the origin of the tetraploids. Third, our data show that *S. halepense* is not well differentiated from *S. bicolor*

(Fig. 1), suggesting that *S. halepense* is of relatively recent origin and that *S. bicolor* (sensu lato) was one of the parental species.

Acknowledgements. This research was supported in part by a grant from Pioneer Hi-Bred International of Johnston/IA, and by the U.S. Department of Agriculture, Grant no. 86-CRCR-1-2161.

References

Bhatti AG, Endrizzi JE, Reeves RG (1960) Origin of Johnson grass. J Hered 51:106–110

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

Celarier RP (1958a) Cytotaxonomic notes on the subsection Halepensia of the genus *Sorghum*. Bull Torrey Bot Club 85:49–62

Celarier RP (1958b) Cytotaxonomy of the Andropogoneae. III. Subtribe Sorgheae, Genus *Sorghum*. Cytologia 23:395–418

Crawford DJ (1983) Phylogenetic and systematic inferences from electrophoretic studies. In: Tanksley SD, Orton TJ (eds) Isozymes in plant genetics and breeding, part A. Elsevier, Amsterdam, pp 257–287

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

Doebley JF, Goodman MM, Stuber CW (1984) Isoenzyme variation in *Zea* (Gramineae). Syst Bot 9:203–218

Doggett J (1970) Sorghum. Longmans, Green and Co., London

Duara BN, Stebbins GL (1952) A polyhaploid obtained from a hybrid derivative of *Sorghum halepense* × *S. vulgare* var. *sudanense*. Genetics 37:369–374

Endrizzi JE (1957) Cytological studies of some species and hybrids in the Eu-sorghums. Bot Gaz 119:1–10

Garber ED (1950) Cytotaxonomic studies in the genus *Sorghum*. Univ Calif Pub Bot 23:283–362

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

Hadley HH (1953) Cytological relationships between *Sorghum vulgare* and *S. halepense*. Agon J 45:139–143

Morden CW, Doebley J, Schertz KF (1987) A manual of techniques for starch gel electrophoresis of *Sorghum* isozymes. Texas Agricultural Experiment Station, College Station/TX, MP-1635

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

Nei M (1972) Genetic distance between populations. Am Nat 106:283–292

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Nei M (1977) *F*-statistics and analysis of gene diversity in subdivided populations. Ann Hum Genet 41:225–233

Snowden JD (1936) The cultivated races of sorghum. Allard and Son, London

Snowden JD (1955) The wild fodder sorghums of the section Eu-sorghum. J Linn Soc London Bot 55:191–260

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

Wet JMJ de, Harlan JR (1971) The origin and domestication of *Sorghum bicolor*. Econ Bot 25:128–135

Wet JMJ de, Huckabay JP (1967) The origin of *Sorghum bicolor*. II. Distribution and domestication. Evolution 21:787–802

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

Wet JMJ de, 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 451–463