Theor Appl Genet (1995) 90:787–796

Y. X. Cui · G. W. Xu · C. W. Magill · K. F Schertz
G. E. Hart

RFLP-based assay of *Sorghum bicolor* (L.) Moench genetic diversity

Received: 5 September 1994 / Accepted: 18 October 1994

Abstract Sixty-two single-copy sorghum DNA clones were used to compare restriction fragment patterns of 53 sorghum accessions from Africa, Asia and the United States. Included were accessions from five morphological races of the cultivated subspecies *bicolor*, and four races of the wild subspecies *verticilliflorum*. From two to twelve alleles were detected with each probe. There was greater nuclear diversity in the wild subspecies (255 alleles in ten accessions) than in the domestic accessions (236 alleles in 37 accessions). Overall, 204 of the 340 alleles (60%) that were detected occurred in both subspecies. Phylogenetic analysis using parsimony separated the subspecies into separate clusters, with one group of intermediate accessions. Though exceptions were common, especially for the race bicolor, accessions classified as the same morphological race tended to group together on the basis of RFLP similarities. Selection for traits such as forage quality may have led to accessions genetically more similar to other races being classified as bicolors, which have a loose, small-grained panicle similar to wild races. Population statistics, calculated using four nuclear and four cytoplasmic probes that detect two alleles each, revealed a low but significant amount of heterozygosity, and showed little differentiation in alleles in the wild and cultivated subspecies. Outcrossing with foreign pollen appears to have been more important than migration via seed dispersal as a mechanism for gene flow between the wild and domestic accessions included in this study.

Key words *Sorghum* · RFLP · Phylogeny
Gene flow

Communicated by J. Mac Key

Y. X. Cui[1] · G. W. Xu · K. F. Schertz · G. E. Hart (✉)
Department of Soil and Crop Sciences, Texas A&M University, College Station, Texas 77843, USA

C. W. Magill
Department of Plant Pathology and Microbiology, Texas A&M University, College Station, Texas 77843, USA

K. F. Schertz
USDA-ARS, Texas A&M University, College Station, Texas 77843, USA

Present address:
[1] Department of Plant Pathology and Microbiology, Texas A&M University, College Station, Texas 77843, USA

Introduction

Sorghum bicolor (L.) Moench (2n = 20) is a highly diverse species belonging to the genus *Sorghum* of the tribe *Andropogoneae*. It is the fifth most important cereal crop worldwide (Doggett 1988). It includes all annual taxa and some perennials of the section *Sorghum*. It consists of an extremely variable complex of cultivated taxa, a widely distributed and ecologically variable wild complex, and stabilized weedy derivatives that have resulted from introgression between domesticated grain sorghums and their close wild relatives. The species has been divided into three subspecies, namely *bicolor*, *verticilliflorum* and *drummondii* (Doggett 1988), among which *bicolor* and *verticlliflorum* are the two most important to modern agriculture. Subspecies *bicolor* is recognized as consisting of five basic morphological races, namely, bicolor, kafir, caudatum, durra and guinea and ten intermediate races as well (Harlan and de Wet 1972). Four races, namely, verticilliflorum, arundinaceum, virgatum and aethiopicum, are recognized in the subsp. *verticilliflorum* (de Wet and Prasada 1986).

The origins and relationships of modern sorghums have been the subject of morphological, cytogenetic, cytoplasmic and isozyme analyses (Liang et al. 1966; de Wet 1978; de Wet and Huckabay 1967; de Wet et al. 1970; Mann et al. 1983; de Wet and Prasada 1986; Morden et al. 1989, 1990; Aldrich et al. 1992). However, these techniques have not provided unequivocal answers to questions concerning the amount of genetic diversity and the phylogenetic relationships among populations, accessions and species. RFLP analysis, as applied to other crops (Song et al. 1988, 1990; Bernatzky

and Tanksley 1989; Miller and Tanksley 1990; Lubbers et al. 1991), as well as to sorghum (Aldrich and Doebley 1992), has proven to be an additional, and more sensitive, tool for studying these questions. Point mutations, deletions or additions which do not lead to changes at the level of proteins or phenotype may still be detected using DNA probes cloned from the same or related species. Because RFLPs directly reflect heritable changes in nucleotide sequence of both coding and noncoding regions of the DNA, the number of possible comparisons is practically unlimited. In addition to clones from nuclear DNA that reveal Mendelian segregation patterns when used as hybridization probes, clones from organellar DNA that show maternal inheritance patterns are also available, so that both nuclear and cytoplasmic diversity can be examined and compared.

A large number of sorghum accessions, including cultivars and their wild relatives, have been collected from all over the world. Some of them may be valuable sources of genes for crop improvement, such as disease and insect resistance, stress tolerance, and male sterility. Knowledge of genetic variation can be used to determine the phylogenetic relationships among accessions, races and subspecies, which has practical implications for germplasm collection and gene-introgression programs.

Two prior studies have utilized heterologous probes to examine RFLP diversity in sorghum. Tao et al. (1993) found low frequencies of polymorphisms for 27 genotypes examined, but Aldrich and Doebley (1992) found much greater allelic diversity among RFLPs detected by maize probes than when isozymes were used to compare a set of 56 geographically and racially diverse accessions.

Several RFLP maps of the sorghum genome have been constructed (Whitkus et al. 1992; Berhan et al. 1993; Chittenden et al. 1994; Xu et al. 1994). Consequently, it is now possible to study genetic diversity in sorghum using sorghum DNA clones selected from throughout the genome. This paper reports the results of a study to utilize unique-sequence DNA probes from both nuclear and cytoplasmic sources to examine the relationships of accessions that have been classified into nine races in two subspecies, primarily based on morphological characters.

Materials and methods

Plant materials

Fifty-three accessions of *S. bicolor* (L.) Moench of diverse geographic origin from two subspecies and nine races were studied (Table 1). The accessions have been maintained by bulking seed from self-fertilizations. The number of plants grown and the number of generations since collection varied among the accessions. Two duplicated accessions identified by different codes were independantly carried through each step of the analysis to provide a check for consistency. Young leaves were collected from greenhouse-grown seedlings of each accession, their midribs removed, and the leaves were either stored at −20 °C or were lyophilized, then ground to a fine powder and stored at −20 °C. In most cases for cultivated accessions, approximately five plants contributed to the pooled leaf sample, while for wild accessions more than ten were combined.

Probe source and preparation

Sixty-one sorghum nuclear DNA clones and a primer dimer (Xu et al. 1993) were used in this study. Fifty-two of the nuclear DNA clones and the primer dimer hybridize to single-copy loci scattered throughout the *S. bicolor* ssp. *bicolor* genome in different linkage groups (Table 2). The fragments to which the other nine nuclear DNA clones hybridize have not been mapped but they are likely also to be single-copy sequences, based on the observation that each detects only one fragment when hybridized to *Eco*RI- or *Eco*RV-digested genomic DNA from each of the parents of the mapping population studied by Xu et al. (1994). The insert DNA fragments were released by *Pst*I-digestion of recombinant plasmids and were isolated from low-melting-point agarose following electrophoresis (Xu et al. 1994).

DNA extraction, digestion, Southern blotting and RFLP probing

DNAs from the 53 sorghum accessions (Table 1) were extracted from 0.6 g of dried powder or from 6 g of fresh leaf tissue that had been powdered by grinding in liquid nitrogen (Murray and Thompson 1980). Aliquots of purified DNA were digested with *Eco*RI and *Eco*RV and 8–10 μg of digested DNA was added to each well for electrophoresis.

DNA electrophoresis, Southern blotting, radio labeling of probes, hybridization and autoradiography followed standard procedures (Maniatis et al. 1982; Feinberg and Vogelstein 1983; Reed and Mann 1985; Helentjaris et al. 1986).

Data analysis

Parsimony analysis of RFLP data was carried out using PAUP, a Macintosh computer software program developed by Swofford (1993). Because each of the probes used in believed to identify only one locus, each RFLP fragment detected by Southern analysis was treated as a unique character. All RFLP fragments detected by all probes for each accession were used to make comparisons among accessions, races and subspecies. All fragments of the same size for a particular probe and restriction-digest combination were assumed to identify the same allele. The 55 taxa and 340 alleles (characters) were organized into a 1–0 matrix (1 – present, 0 – absent). The heuristic option of the PAUP program, using stepwise addition, was employed to generate phylogenetic trees.

The coefficients of relationship, F_{IS} and F_{ST}, developed by Wright (1931, 1932, 1963) were used to describe the genetic structure of the sorghum subspecies populations using four nuclear and four cytoplasmic loci for which simple allelic differences are present in the 53 accessions. The data for the four nuclear loci were obtained in this study and the patterns for the four organellar loci are reported in Xu et al. (1995). For seven of the loci, only two alleles were detected. For locus *txs349*, a third allele that was detected in only one accession was pooled with the rarer common allele for analysis. For the nuclear loci, F_{IS} was calculated from $(\bar{H}_S - H_1)/(\bar{H}_S)$, where $\bar{H}_S$ is the expected heterozygosity of an individual in a subpopulation based on Hardy-Weinberg equilibrium and H_1 is the observed frequency of heterozygosity in the subpopulation. Entries classified as cultivated accessions (RFLP clusters I–V; see below) were considered to be one subpopulation and those classified as wild accessions (RFLP clusters VI and VII; see below) another. For these loci, the fixation index F_{ST} was calculated from the formula $F_{ST} = (H_T - \bar{H}_S)/(H_T)$, where H_T is the expected heterozygosity of an individual assuming all entries are drawn from the same random-mating population. For the four organellar loci, F_{st} was calculated

Table 1 Accessions of *S. bicolor* included in this study

Accession code or name	Race[a]	Origin	Cytotyp[e]	Lane in Fig. 1
S. bicolor ssp. *bicolor*[a]				
Chinese Amber FC8728	bicolor	China	B	16
Chinese Nat. Acc. No. 422	bicolor	China	B	1
CI171	bicolor	China	B	22
EBA-9[b]	bicolor	Sudan	B	18
IS1598C	bicolor	India	B	26
IS6964	bicolor	Sudan	B	28
IS7542C	bicolor	Nigeria	B	27
Shanqui Red[c]	bicolor	China	E	23
Standard Broomcorn CI556[b]	bicolor	Italy	B	10
Sweet Sudangrass SA372[b]	bicolor	Sudan	A	12
IS12568C	caudatum	Sudan	B	29
IS12608C	caudatum	Ethiopia	B	30
IS6710C	caudatum	Senegal	B	31
M91051	caudatum	USA	B	24
BTx623	caudatum/kafir	USA	B	2
3-Dwarf White Sooner Milo	durra	unknown	A	7
IS1022C	durra	India	B	32
IS12570C	durra	Sudan	B	34
IS7333C	durra	Nigeria	B	33
SA7078	durra	unknown	B	8
Tx403	durra	USA	A	25
IS3620C[b]	guinea	Nigeria	A	15
IS5332C	guinea	India	B	36
IS7173C	guinea	Tanzania	G	20
IS7920	guinea	Nigeria	B	9
IS7920C[c]	guinea	Nigeria	C	35
IS3477C	guinea	Sudan	B	38
IS3614C	guinea	Nigeria	B	37
IS3955C	guinea	Nepal	A	39
IS7419C	guinea	Nigeria	B	17
87-4325	kafir	S. Africa	B	40
87-4326	kafir	S. Africa	B	41
87-4327	kafir	S. Africa	B	42
87-4328	kafir	S. Africa	B	43
87-4329	kafir	S. Africa	B	44
BTx378	kafir	USA	B	3
BTx398	kafir	USA	B	4
BTx3197	kafir	USA	B	5
B35	unknown	USA	B	21
QL3-India	unknown	India	B	11
RTx430	unknown	USA	B	6
S. bicolor ssp. *verticilliflorum*				
IS14564[d]	aethiopicum	Sudan	B	14 and 45
IS18820	aethiopicum	Egypt	B	47
PI302105	aethiopicum	Ethiopia	B	46
IS18826[d]	arundinaceum	Ivory Coast	F	13 and 48
PI156549	arundinaceum	Rhodesia	A	49
PI186570	arundinaceum	Nigeria	A	50
IS14505	verticilliflorum	Uganda	B	52
IS18798	verticilliflorum	S. Africa	B	51
PI208190	verticilliflorum	unknown	B	19
PI267331	verticilliflorum	India	B	53
IS18803	virgatum	unknown	B	55
IS18809	virgatum	Egypt	B	54

[a] Subspecies and race classification is based on morphological classification according to Harlan and de Wet (1972) and de Wet (1978)
[b] Based on the RFLP patterns in this study, this accession has greater similarity to subspecies *verticilliflorum*
[c] This accession has a unique fragment pattern detected with one or more probes of organellar-DNA origin
[d] Duplicate samples
[e] Cytotypes from Xu et al. (1995)

Table 2 pSbTXS clones used in this study, linkage groups in which the clones detect loci, and number of alleles detected in subsets of sorghum accessions

Clone	Linkage group[b]	# of alleles	No. of common and rare[a] alleles, respectively		
			Cultivated[c]	Wild[d]	Shared[e]
284	G	8	4 and 1	5 and 2	3
309	unk[f]	3	2	2 and 1	2
338	unk	5	3 and 1	2 and 1	2
443	C	2	2	2	2
463	unk	4	2	4	2
503	C	5	2	2 and 3	2
513	A	2	2	2	2
545	C	6	3 and 1	3 and 3	4
547	D	6	2 and 1	3 and 3	3
578	C	4	4	4	4
584	C	4	2 and 1	1 and 2	2
585	H	4	3	4	3
604	H	4	2 and 2	1 and 1	2
1015	M	12	7 and2	8 and 2	7
1019	unk	9	3 and 1	5 and2	3
1030	E	5	2 and 2	3	2
1033	L	4	2	1 and 2	1
1077	H	5	5	4 and 1	5
1085	E	4	2 and 1	2 and 1	2
1090	H	6	5	4 and 2	5
1106	B	8	3 and 2	3 and 3	3
1111	G	6	1 and 2	4 and 2	3
1113	H	5	4	3 and 1	3
1114	unk	6	2	5 and 1	2
1116	unk	8	4 and 2	5 and 1	4
1126	D	6	2 and 1	5 and 1	3
1128	I	2	2	2	2
1129	J	6	3 and 2	5 and 1	5
1139	E	7	3 and 2	6 and 1	5
1143	G	5	4 and 1	4 and 1	5
1159	I	6	3 and 1	3 and 2	3
1164	I	6	3 and 1	4 and 1	3
1173	E	3	2	3	2
1175	C	7	3 and 1	6 and 1	4
1178	C	4	2	4	2
1183	E	4	2 and 1	3 and 1	3
1185	C	4	2 and 1	3	2
1197	B	4	1 and 1	1 and 3	2
1219	H	6	3 and 1	5 and 1	4
1249	unk	6	4	4 and 1	3
1323	unk	6	4	2 and 2	2
1433	I	9	5 and 1	3 and 4	6
1449	B	7	5 and 1	6	5
1451	unk	7	4 and 1	3 and 3	4
1527	G	3	3	3	3
1537	G	5	3	2 and 2	2
1563	D	7	3 and 1	3 and 2	3
1610	G	6	4 and 1	5 and 1	5
1611	B	6	2 and 1	2 and 4	3
1625	J	3	2 and 1	2 and 1	3
1674	D	6	3 and 1	4 and 1	4
1714	H	2	2	2	2
1765	J	4	3	3 and 1	3
1845	B	3	2	3	2
1851	A	9	6	4 and 5	6
1855	G	4	3 and 1	3 and 1	4
1858	B	10	4 and 2	5 and 3	5
1901	H	11	5 and 3	6 and 2	7
1925	G	6	5 and 1	4 and 1	5
1927	C	4	2 and 1	3 and 1	3
1929	B	6	4	3	2
pd[g]	A	5	2	2 and 3	2
Total (62)		340	188 48	213 85	204

[a] Allele is present in only one accession
[b] Xu et al. 1994
[c] RFLP clusters I to V; 37 accessions from subspecies *bicolor*
[d] RFLP clusters VI and VII; 16 accessions primarily considered to belong to *verticilliflorum* subspecies.
[e] Allele is present in both cultivated and wild clusters
[f] Fragment(s) to which the clone hybridized have not been mapped
[g] A primer dimer (Xu et al. 1993)

from the relationship

$$F_{st} = \frac{\sum_{i=1}^{n} \sigma_i^2}{\sum_{i=1}^{n} \bar{p}_i(1-\bar{p}_i)}, \quad \text{where } \sum_{i=1}^{n} \sigma_i^2 = \sum_{i=1}^{n} w_i(p_i - \bar{p}_i).$$

The parameter N_em was used to estimate the rate of gene flow between subpopulations (Hartl and Clark 1989). N_em was calculated based on the relationship:

$$F_{ST} = \frac{1}{4N_em + 1} \quad \text{(applied for nuclear loci) or,}$$

$$F_{ST} = \frac{1}{2N_em + 1} \quad \text{(applied for organellar loci).}$$

Results and discussion

Nuclear DNA diversity among accessions

Polymorphism

Although each of the probes that were used hybridized to a unique sequence, many polymorphisms were detected (Table 2). The range in the number of alleles detected per probe among the 53 accessions was 2–12 with an average of 5.24 for *Eco*RV-digested sorghum genomic DNA and 2–11 with an average of 5.63 for *Eco*RI-digested DNA (determined over all accessions). An *Eco*RI allele for each locus was thus present in 10 of the 53 accessions on average and an *Eco*RV allele was present in 9.4 of the 53 accessions on average. These figures are higher than for some other plant species, such as maize and barley (Melchinger et al. 1990; Lubbers et al. 1991), and much higher than for wheat. This may in part reflect the fact that monomorphic loci were unlikely in this study because the probes were selected among those that revealed polymorphisms for one or both restriction enzymes between BTx623 and IS3620C, both of which are included. In addition, the 53 accessions examined were deliberately chosen to include racial and geographic diversity. Even so, as has been observed in morphological and isozyme studies, sorghum has an unusual amount of diversity for a predominately self-pollinating species (de Wet et al. 1970; de Wet 1978; Mann et al. 1983; House 1985; Aldrich et al. 1992). Multiple origins for domesticated sorghums, cross pollination between selected races, and outcrossing between domestic cultivars and highly variable wild species, are all considered to be factors contributing to the high rate of polymorphism in sorghum (Doggett 1988). The relatively high frequencies of different RFLP alleles should be helpful in phylogenetic analysis.

Phylogenetic relationships among accessions

Sixty-two sorghum nuclear DNA clones were used as hybridization probes on Southern blots containing DNA from each of the 53 accessions included in this study. Twenty-one were hybridized to both *Eco*RI and *Eco*RV restriction-enzyme-digested genomic DNA, an additional 17 were hybridized only to *Eco*RI-digested DNA and another 24 were hybridized only to *Eco*RV-digested DNA. All hybridization patterns suggested that the clones hybridized to single-copy loci (Fig. 1). Therefore, fragments of different length detected on Southern blots were believed to represent variant alleles, and each fragment was treated as a unique character. PAUP data matrices were generated in which each accession was treated as an operational taxonomic unit making no assumption about species assignment. From these matrices, dendrograms were constructed. There were no significant differences between the two dendrograms that were generated using *Eco*RI and *Eco*RV RFLP patterns for the 21 clones that were hybridized to both digests (data not shown). Several studies have demonstrated that polymorphism detected with different restriction enzymes by the same probe is likely to be due to the same deletion or insertion mutation (McCouch et al. 1988; Wang and Tanksley 1989). To avoid using the same information twice, we employed the data from all 45 probe-*Eco*RV enzyme combinations and the 17 probes that hybridized only to *Eco*RI digests to generate phylogenetic trees. Two trees of length 1633 were obtained. A consensus tree was generated using the strict consensus method which depicts the genetic relationship among the accessions (Fig. 2). Two pairs of accessions independently processed as blind unknowns were not separated in the tree, adding to our confidence that the banding patterns are reproducible and can be reliably scored.

Relationships among subspecies

Branches in the dendrogram maximally separated wild types, then mixtures of wild and domestic accessions, and to a lesser degree domestic types. Isolates in the tree have been grouped into seven clusters for discussion purposes (Fig. 2). The upper branch of the tree contains clusters I to V which include 37 accessions from five races all previously classified as members of the domesticated subspecies *bicolor*. All accessions in this group are cultivated types and their sources include all of the major sorghum-growing regions. The plants are annuals, often branched, and frequently tillered. Leaf blades are up to 90 cm long and 12 cm wide. Overall, the accessions in this group can be considered to be cultivated types of the subspecies *bicolor*. The 62 nuclear probes detected 236 alleles within these 37 accessions (3.8 alleles per locus), so that an allele was present on average in approximately 14 accessions.

On the other side of the tree, a group of ten accessions formed cluster VII. These accessions were relatively more variable; the 62 probes detected 255 alleles, so that an allele was present on average in only 2.4 accessions. In the classical sorghum classification, all accessions in

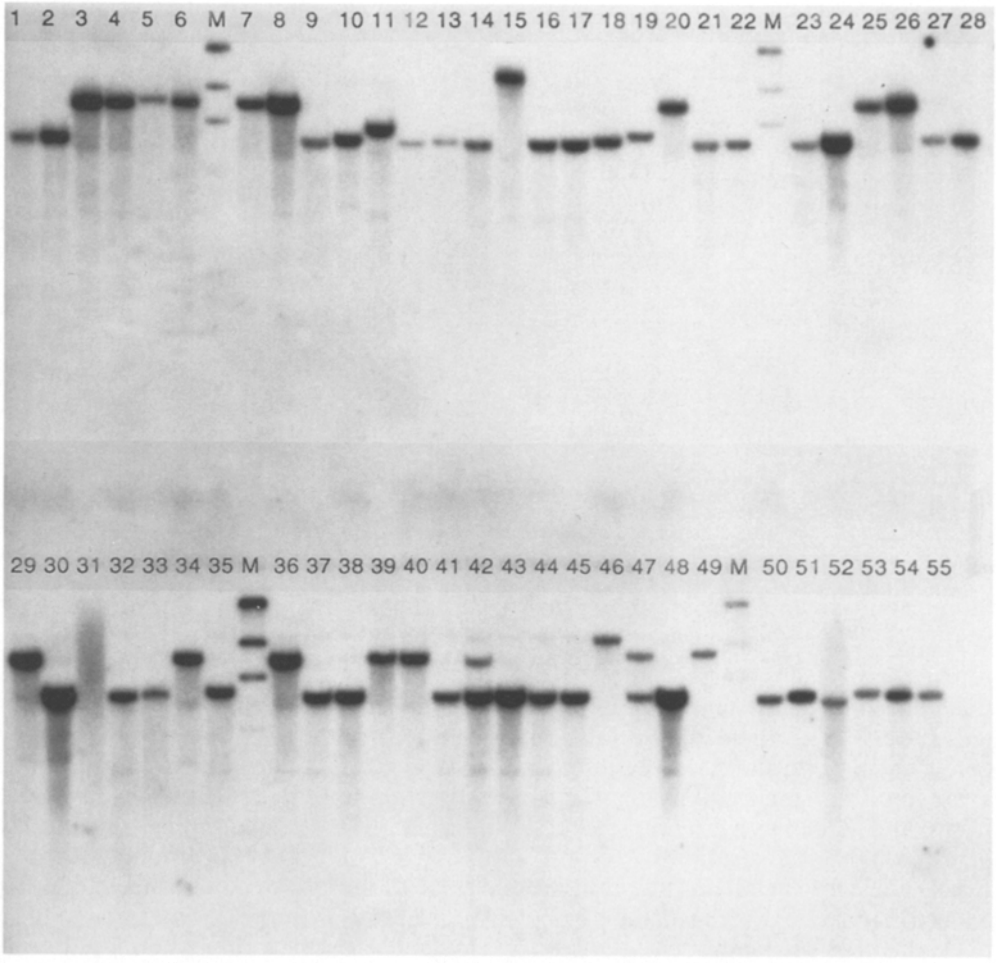

Fig. 1 A nuclear clone TXS1611 hybridized to a single-copy locus of *Eco*RV-digested DNA. Identification of accessions in *numbered lanes* is given in Table 1, and lanes labeled *M* contain a size marker

this group, except IS3620C, belong to one of the five races derived from the wild subspecies *verticilliflorum*. Specifically, IS14505, PI208190 and PI267331 are placed in race verticilliflorum, IS18803 and IS18809 in race virgatum, IS18820 and PI302105 in race aethiopicum, and IS18826 and PI156549 in race arundinaceum. Though IS3620C is recorded as belonging to subspecies *bicolor*, race guinea, it has an extremely grassy phenotype (margaritiforum) more typical of the wild subspecies. The cross between IS3620C and BTx623, a parent of grain sorghum hybrids, was chosen over several other crosses for use in producing progeny for an RFLP map because it had the highest frequency of polymorphism (Xu et al. 1994). Data presented here show that IS3620C has many alleles in common with the wild subspecies, suggesting that this accession may represent a case of introgression.

The ability to separate the wild subspecies from cultivated types on the basis of RFLPs, and the relative increase in diversity in the wild subspecies over the cultivated types, are in agreement with the prior report by Aldrich and Doebley (1992).

Seven of the ten accessions in group VII are known to originate in Africa. The two accessions of uncertain origin are thought also to come from Africa, but this cannot be verified by available records. Accession PI1267331 of race verticilliflorum was collected in India, but it may well have been an early introduction from Africa, in accordance with Doggett's (1988) proposal. In addition, all accessions in this group are annuals or biennials with slender to stout culms and phenotypes typical of wild species. This cluser can be considered to include representatives of the subspecies *verticilliflorum*.

The results presented above agree with the classical taxonomy of *S. bicolor* species on the subspecies level. However, there was another group, cluster VI, that falls between the cultivated and wild types. Of the six accessions in this group, three are considered to belong to subspecies *verticilliflorum* (IS14564, race aethiopicum, IS18798, race verticilliflorum and PI186570, race arundinaceum) and three to ssp. *bicolor* race bicolor (Standard Broomcorn CI556, Sudangrass EBA-9, and Sweet Sudangrass SA372). The latter three accessions are either extremely grassy types or old established lines that are distinct from other domestic accessions. All accessions in this group except Standard Broomcorn CI556 are from Africa. Standard Broomcorn was introduced into the United States from Italy, but most likely also came from Africa originally.

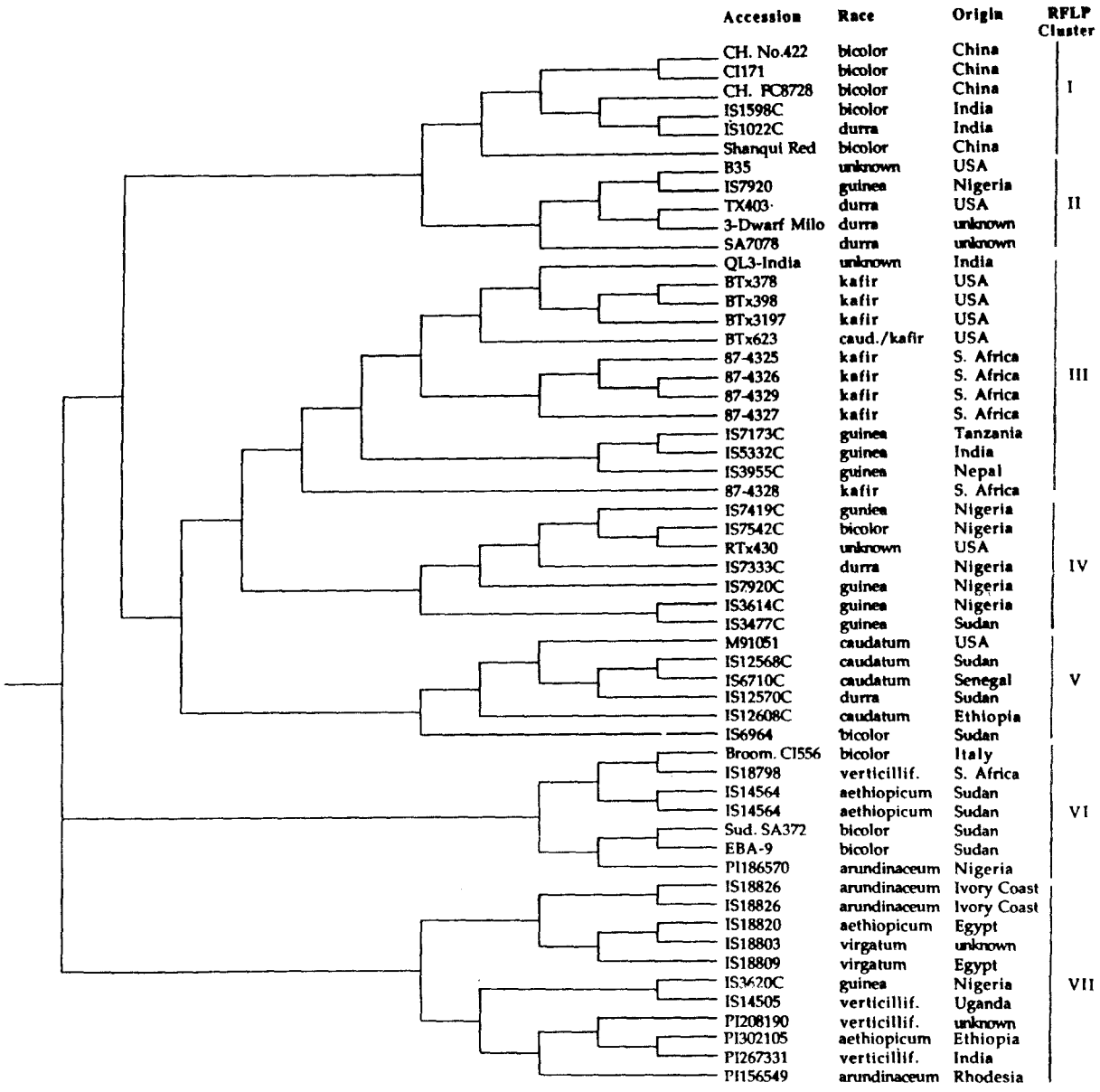

Fig. 2 A PAUP computer program generated dendrogram of all 53 accessions of *Sorghum bicolor* based on RFLPs detected with 62 sorghum genomic clones. Tree length = 1633. CI = 0.206. HI = 0.794. *First column* – accession code or name; *second column* – original classification; *third column* – Geographical origin; *fourth column* – RFLP cluster

Unless the PAUP program is instructed to use a specific isolate or group as an outgroup, the relative number of differences between one isolate and all others is used to build the tree; that is, the most divergent isolates will fall at the extremes, with intermediates between. In the comparison of RFLP alleles made here, the clusters remain the same when a member of either the cultivated or wild groups forms the basis for comparison, or if either group is defined as the out-group, and in all cases cluster VI remained intermediate. (No true outgroup, such as a known progenitor to all of the accessions, was available). It is possible that the isolates in this cluster actually are true intermediates that arose from natural crosses between wild and domestic species. Doggett and Majisu (1968) found that natural hybrids could be identified and some of these gave remarkably uniform progeny when planted in progeny rows. Doggett has speculated that the typical cultural practice in Africa of abandoning old fields in favor of new has permitted wild × cultivated hybrids to reproduce for several years on fallow ground, and has thus led to stable intermediate accessions (Doggett 1988).

It has been proposed that sorghum has diverse origins with the wild subspecies *verticilliflorum* giving rise to the cultivated subspecies *bicolor*, the wild race aethiopicum to the races durra and bicolor, and the wild races arundinaceum and verticilliflorum to the races guinea and kafirs, respectively (Mann et al. 1983). All races in wild subspecies *verticilliflorum* are widely

distributed in Africa. The race verticilliflorum is the most widely distributed and morphologically the most variable. It might have exchanged genetic material with other races currently under cultivation by man (Mann et al. 1983). Therefore, it is reasonable that the six accessions in cluster VI form an intermediate group.

Another interpretation that cannot be ruled out by parsimony analysis, since clusters VI and VII are separated at the same level from the clusters of cultivated accessions, is that cluster VI represents accessions derived from a separate origin. This interpretation would suggest that, although the current races represent distinct morphological types, the race classifications are not genetically valid. Under this interpretation, the accessions in cluster VI would be classified as a second wild subspecies.

Relationship among races

Across the whole dendrogram, it can be seen that morphologically defined races did not fall into unique clusters. Within the *bicolor* subspecies, all accessions from China and two accessions from India were grouped into the first cluster. Five of the six accessions were originally classified as race bicolor, but IS1022C from India is classified as a durra. Race bicolor accessions have a panicle more like wild species than other cultivated races. As defined by Harlan and de Wet (1972), bicolor includes sorghum types which were selected for characters other than grain and glume, such as broomcorn and the sweet-stemmed sorghums used for forage. Doggett (1988) argues that bicolor should not be considered a race in the taxonomic sense, which could account for the presence of bicolors in clusters I, IV, V and VI.

There were three durras, one guinea and one accession of an unknown race in the second cluster, which includes several accessions of agronomic importance. B35, a United States line originally from Nigeria, is being used as a source of drought tolerance (Darrell T. Rosenow, personal communication), SA7078 is often used as a susceptible host for several diseases in the US, and 3-Dwarf White Sooner Milo is a typical early-maturing milo.

Clusters I and II share many alleles, and could be considered to be a common cluster. Mann et al. (1983) suggest that races bicolor and durra arose from the same wild race, aethiopicum. Therefore, the two races would be expected to have a higher affinity to each other than to any other race.

All tested accessions which have presently been classified morphologically as race kafir fell into the third cluster. This cluster includes several very important agronomic kafir lines, namely, BTx378, a very high-yielding US female parent, BTx398, an early maturing female parent, BTx3197, formerly used as a female parent in the US, and BTx623, a widely used, high-yielding and disease-resistant US parent. The five other kafir accessions, 87-4325, 87-4326, 87-4327, 87-4328 and 87-4329, were introduced from S. Africa into the United States 10 years ago. In addition to kafir accessions, cluster III includes three guinea accessions, IS3955C, IS5332C and IS7173C, which originated from Nepal, India and Tanzania, respectively. QL3-India, a distinctive line derived from a cross of an unknown race with wild sorghum, also fell into this cluster. It was somewhat surprising that the subgroup of three guinea lines is in this cluster and separated kafir accession 87-4328 from the other South African kafir accessions. Other studies have shown that there is an overlapping area where both guineas and kafirs are distributed, and that guinea-kafir hybrid forms are frequent (Mann et al. 1983; House 1985). It is probable that there was a high degree of genetic introgression from kafir into these three guinea accessions.

The fourth cluster includes four guinea accessions, one durra accession, one bicolor accession, and one accession whose race is unknown. All accessions in this cluster have origins in Africa; five are from Nigeria, one was from Sudan and RTx430, a well known breeding cultivar, was developed in the US from a cross between a local cultivated line and a line from Ethiopia. RFLP analysis suggests that the accessions in this cluster have a common origin from what may be considered guinea progenitors.

The four caudatum accessions tested, including accessions from Sudan, Ethiopia and Senegal, fell into the fifth cluster. The accession M91051 has been widely used as a parental line in the US and probably was also derived from African lines. Two accessions previously classified as non-caudatum also fell into this cluster. IS12570C classified as a durra from Sudan had high RFLP affinity to the caudatum accessions. As indicated previously, it is possible that IS6964, which is listed as a bicolor, had an origin in common with the caudatums, but was selected at some time for a trait other than those that affect panicle morphology.

Sixty-two polymorphic loci from sites scattered around the genome have been used to examine 53 diverse accessions of *S. bicolor*. Overall, the RFLP-based classification was consistent with the classical morphological classification, especially when geographic origins and the likelihood of selection for traits other than grain production are considered. It is not surprising that accessions from different origins or race classification fall into the same RFLP cluster. Because no mating barriers existed among the races, gene flow could have occurred between races that were grown together, or nearby, to generate apparent changes in taxonomic position. Crosses undoubtedly occurred between wild races, and have been documented to occur between wild and cultivated races, as well as between cultivated races used in breeding programs (Mann et al. 1983; Doggett 1988). RFLP analysis can help to distinguish the different subspecies, and can also help define or monitor the origins of existing intermediate subspecies and overlapping races. The relatively high frequency of polymorphism will make it possible to measure progress

during backcrosses designed to introgress a gene into a cultivated accession from a wild race.

Gene flow between subspecies

The genetic structure of and gene flow between the cultivated-race subspecies and the wild subspecies were tested using data from eight loci that displayed only two alleles per locus (the population size is too small to test multiple-allele loci). Allelic data used in this analysis are shown in Tables 2 and 3 and F_{IS}, F_{ST} and N_em values in Table 4. Based on the RFLP results, clusters I, II, III, IV and V were treated as one subpopulation and clusters VI and VII as another subpopulation. By checking hybridization blots of all probe-enzyme combinations, it was found that 0–10 accessions showed hybrid patterns (two or more fragments) at each locus. The accessions have been maintained by selfing since collection, which should lower the levels of heterozygosity; and even though the seeds produced were bulked, and the DNA samples were derived from mixed samples, some loss of natural variation is likely. Thus the overall estimate for heterozygosity of 6.33% is at best an underestimate of natural heterozygosity. The range of F_{IS} was 0.66 to 0.91 as compared to the near zero expected value for a random mating, non-inbreeding population. The range of F_{ST} was 0.0061 to 0.01446 at four nuclear loci and 0.0472 to 0.0662 at four cytoplasmic loci. These relatively small values suggest there has been little differentiation of the two subpopulations by the fixation of different alleles. (The same conclusion can be drawn from Table 2, where it can be seen that for these particular loci, both alleles were present in both subsets of accessions.) While Aldrich and Doebley (1992) concluded that the cultivated sorghums are derived from subspecies *arundinaceum* based on the high degree of genetic similarity, this does not conflict with our results, since we are following the recommendation that ssp. *arundinaceum* should now be called ssp. *verticilliflorum* (Doggett 1988). The F_{IS} figures are consistent with the fact that sorghum is predominantly self pollinating. However, evidence presented here also supports previous conclusions that outcrossing has played a very important role in sorghum evolution. It has been estimated that without controlled pollination there would be 5–50% outcrossing, with higher rates among wild sorghums than among cultivated sorghums (Maunder and Sharp 1963; Doggett 1988).

The rates of gene flow (N_em) as calculated from F_{ST} values for nuclear loci were two to four times higher than for cytoplasmic loci (Table 4). A possible explanation for this observation is that pollen from neighbouring wild accessions more often contributed to variation (gene flow) in the collected accessions than did migrants originating from seeds dispersed from a population with a distinct cytotype.

Table 3 Maternally inherited RFLP alleles[a] present in cultivars or accessions that differ in cytoplasmic-male-sterility cytotype (Data and cytotypes from Xu et al. 1995)

txs clone	Restriction endonuclease	Cytotype (accessions)			
		A (44)	B (7)	C (IS7920C)	E (Shanqui Red)
−349	*Eco*RV	2	2	1	2
	*Eco*RI	2	1	1	1
−1027	*Eco*RV	1	1	1	1
	*Eco*RI	3+4+5	1+2	3+4+5	3+4+5
−1058	*Eco*RV	2	1	2	1
	*Eco*RI	2	1	2	1+2
−1168	*Eco*RV	2	1	1	1+2
−1177	*Eco*RV	2	1	2	2
	*Eco*RI	2	1+2	2	2

[a] The largest fragment detected by a clone for a specific restriction enzyme is considered allele #1, the next largest is #2, etc

Table 4 F_{IS}, F_{ST} values and estimated gene flow (N_em) between the two sorghum subspecies at four nuclear loci and four cytoplasmic loci

Locus	F_{IS}	F_{ST}	N_em
*txs*443	0.83	0.0061	40.7
*txs*513	0.76	0.0064	38.8
*txs*1128	0.91	0.0146	16.9
*txs*1714	0.66	0.0077	32.2
*txs*349[a]		0.0662	7.1
*txs*1058[a]		0.0472	10.1
*txs*1168[a]		0.0662	7.1
*txs*1177[a]		0.0472	10.1

[a] Cytoplasmic loci

RFLP data support previous concepts of sorghum evolution; namely, that multiple origins, diverse environments and human involvement have contributed to the existence of different types of wild and cultivated sorghum. Outcrossing has led to gene introgression and gene flow among the natural populations. Then polymorphic subpopulations develop, and disruptive selection starts. Intermediate types may exist for a period, but differentiation continues until a number of distinct, separate and adaptive populations are formed. In summary, the population structure of modern sorghums seems to fit well into Wright's "shifting balance" theory of adaptation, which assumes that genetic drift and

selection operating on subpopulations leads to a number of genotypes occupying different adaptive peaks, even though gene flow can occur between the subpopulations (Wright 1931, 1932, 1978). Wright's theory has been widely accepted to explain plant evolution and speciation (Hartl and Clark 1989), including applications to the evolution of sorghum (Doggett and Majisu 1968; Doggett 1988).

Acknowledgements We are grateful to The Rockefeller Foundation for providing a fellowship to Y. X. Cui. Funds for this research were provided by USDA-ARS as recommended by the Sorghum Crop Advisory Committee and by TAES. This paper is a technical article of the Texas Agricultural Experiment Station.

References

Aldrich PR, Doebley J (1992) Restriction fragment variation in the nuclear and chloroplast genomes of cultivated and wild *Sorghum bicolor*. Theor Appl Genet 85:293–302

Aldrich PR, Doebley J, Schertz KF, Stec A (1992) Patterns of allozyme variation in cultivated and wild *Sorghum bicolor*. Theor Appl Genet 85:451–460

Berhan AM, Hulbert SH, Bulter LG, Bennetzen JL (1993) Structure and evolution of the genomes of *Sorghum bicolor* and *Zea mays*. Theor Appl Genet 86:598–604

Bernatzky R, Tanksley SD (1989) Restriction fragments as molecular markers for germplasm evaluation and utilization. In: Williams JT (ed) The use of plant genetic resources. Cambridge University Press, Cambridge, pp 353–363

Chittenden LM, Schertz KF, Lin YR, Wing RA, Paterson AH (1994) A detailed RFLP map of *Sorghum bicolor* × *S. propinquum* suitable for high-density mapping suggests ancestral duplication of *Sorghum* chromosomes. Theor Appl Genet 87:925–933

Doggett H (1988) Sorghum, 2nd ed. John Wiley and Sons, Inc., New York

Doggett H, Majisu BN (1968) Disruptive selection in crop development. Heredity 23:1–23

Feinberg AP, Vogelstein B (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Bioch 123:6–13

Harlan JR, de Wet JMJ (1972) A simplified classification of cultivated sorghum. Crop Sci 12:172–176

Hartl DL, Clark AG (1989) Principles of population genetics, 2nd ed. Sinauer Associates Inc., Massachusetts, pp 293–323

Helentjaris D, Slocum M, Wright S, Schaefer A, Nienhuis J (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. Theor Appl Genet 72:761–769

House LR (1985) A guide to sorghum breeding. 2nd Int Crop Research Institute for the Semi-Arid Tropics, India, pp 3–9

Liang GHL, Casady AJ (1966) Quantitative presentation of the systematic relationship among twenty-one sorghum species. Crop Sci 6:76–79

Lubbers EL, Gill KS, Cox TS, Gill BS (1991) Variation of molecular markers among geographically diverse accessions of *Triticum tauschii*. Genome 34:354–361

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Mann JA, Kimber CT, Miller FR (1983) The origin and early cultivation of sorghums in Africa. Bulletin, no. 1454. Texas A&M University, College Station, Texas, USA

Maunder AB, Sharp GI (1963) Localization of outcrosses within the panicle of fertile sorghum. Crop Sci 3:449

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of rice chromosomes. Theor Appl Genet 76:815–829

Melchinger AE, Lee M, Lamkey KR, Hallauer AR, Woodman WL (1990) Genetic diversity for restriction fragment length polymorphisms and heterosis for two diallel sets of maize inbreds. Theor Appl Genet 80:488–496

Miller JC, Tanksley SD (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor Appl Genet 80:437–448

Morden WC, Doebley J, Schertz KF (1989) Allozyme variation in old world races of *Sorghum bicolor* (Poaceae). Am J Bot 76:245–255

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

Murray MG, Thompson WF (1980) Rapid isolation of high-molecular-weight plant DNA. Nucleic Acids Res 8:4321–4325

Reed KC, Mann AD (1985) Rapid transfer of DNA from agarose gels to nylon membranes. Nuclic Acids Res 13:7207–7221

Song KM, Osborn TC, William PH (1988) Brassica taxonomy based on nuclear restriction fragment polymorphisms (RFLPs). Theor Appl Genet 75:784–794

Song KM, Osborn TC, William PH (1990) Brassica taxonomy based on nuclear restriction fragment polymorphisms (RFLPs). 3 Genome relationships in *Brassica* and related genera and the origin of *B. oleracea* and *B. rapa* (syn. *campestris*). Theor Appl Genet 79:497–506

Swofford DL (1993) Phylogenetic Analysis Using Parsimony, version 3.1 Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois

Tao Y, Manners JM, Ludlow MM, Henzell RG (1993) DNA polymorphisms in grain sorghum [*Sorghum bicolor* (L.) Moench]. Theor Appl Genet 86:679–688

Wang ZY, Tanksley SD (1989) Restriction fragment length polymorphism in *Oryza sativa* L. Genome 32:1113–1118

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

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

Wet JMJ de, Prasada RKE (1986) Wild sorghums and their significance in crop improvement. Proc Sorghum Conf Shenyang China, pp 28–33

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

Whitkus R, Doebley J, Lee M (1992) Comparative genome mapping of sorghum and maize. Genetics 132:1119–1130

Wright S (1931) Evolution in Mendelian populations. Genetics 16:97–159

Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proc 6th Int Cong Genet 1:356–366

Wright S (1963) Genic Interaction. In: Burdette WJ (ed) Methodology in mammalian genetics. Holden-Day, San Francisco, pp 159–192

Wright S (1978) Evolution and the genetics of populations, vol 4. Variability within and among natural populations. University of Chicago Press, Chicago 580 pp

Xu GW, Magill CW, Hart GE (1993) Primers that amplify inserts in a multi-cloning site also hybridize to *Sorghum bicolor* (L.) Moench DNA. Genome 36:198–201

Xu GW, Magill CW, Schertz KF, Hart GE (1994) An RFLP linkage map of *Sorghum bicolor* (L.) Moench. Theor Appl Genet 89:139–145

Xu GW, Cui YX, Schertz KF, Hart GE (1995) Isolation of mitochondrial DNA sequences that distinquish male-sterility inducing cytoplasms in *Sorghum bicolor* (L.) Moench. Theor Appl Genet (in press)