Chapter 2
The Gene Pool of *Sorghum bicolor* and Its Improvement

Clarissa T. Kimber, Jeff A. Dahlberg, and Stephen Kresovich

Abstract Sorghum, a genus having evolved across a wide range of environments in Africa, exhibits a great range of phenotypic diversity and numerous resistances to abiotic and biotic stresses. Sorghum is recognized as a highly productive, drought tolerant, C_4 cereal that provides humankind with food, feed, fuel, fiber, and energy, particularly in the semiarid tropics of the world.

Sorghum has been collected and conserved over the past 50 years and numerous international and national collections exist. The major collections have in excess of 40,000 accessions and much of the native diversity of cultivated sorghum is represented. However, much of the diversity of the wild races of sorghum is underrepresented in these same collections. Over the past decade, the major collections have benefited by efforts to better characterize these accessions but these efforts have not significantly increased use of the materials. Therefore, despite a significant number of collections and holdings, much of the diversity of sorghum remains untapped.

Over the past decade, tremendous progress has been made to build the molecular and genomic foundation required to increase our understanding of sorghum diversity in the genome and gene pool and, ultimately, to link this information to crop improvement. Sorghum represents the first crop genome of African origin to be sequenced (Paterson et al. Nature 457:551–556, 2009) and, through coordinated national and international efforts, high-density genetic and physical maps, extensive

C.T. Kimber (✉)
Department of Geography, Texas A&M University, 705 O&M Building,
College Station, TX 77843, USA
e-mail: ctk5617@geog.tamu.edu

J.A. Dahlberg
U.S. Sorghum Checkoff Program, Lubbock, TX, USA
e-mail: jeff@sorghumcheckoff.com

S. Kresovich
Department of Biology, University of South Carolina, Columbia, SC, USA
e-mail: sk@sc.edu

sets of RFLP and SSR markers, association (Casa et al. Crop Sci 48:30–40, 2008) and diversity panels (Deu et al.Genome 49:168–180, 2006), nested association mapping populations, and other resources are readily available for use for scientific investigations and breeding efforts. The generation and use of these genomic resources have added to our insights about sorghum domestication and diversity. Future studies will enrich our understanding and provide increasing resolution to quantify and use both wild and domesticated sources of diversity in crop improvement.

Keywords Bottleneck • Collections • Conservation • Diversity • Domestication • Evolution • Genome • Races • Selection

1   Properties of the Species

1.1   Introduction

Having been domesticated for a variety of useful products and cultivated in a broad range of environments, sorghum exhibits a great range of phenotypic diversity. Around the world, sorghum is grown for the production of dense grain panicles (for food, feed, and/or energy), tall, thick sweet stalks (for food, feed, and/or energy), and various forage types (for feed and fuel). Through advanced genetic manipulation, striking divergence among the different forms of sorghum can occur (see Fig. 2.1). In regions where mechanical harvesting predominates, grain types tend to be short in stature (0.5–1.0 m) with a large erect stem supporting a semicompact or compact panicle. In regions where manual harvesting is still practiced and multiple plant parts are desired, the plant may be tall (3–5 m) with an open panicle.

As a member of the grass family, sorghum represents a robust, cane-like species. It has the ability to tiller, regrow following harvest, and produce a fibrous, deep root system. Sorghum leaf blades, similar in appearance to those of its close relative maize, may be up to a meter in length with a width of 10–15 cm. As a drought tolerant, nutrient use efficient, C_4 species, it is highly productive and resilient.

The panicles and grains of the *Sorghum* species can vary widely in shape and size and represent a means for racial classification. Sorghum panicles are made up of perfect flowers, and it is considered a self-pollinated species; however, outcrossing can be as high as 70 % in certain races in particular environments.

Originating in Sudan and Ethiopia, sorghum in nature is a short-day plant and photoperiod sensitive. Elite germplasm, however, has been bred to be photoperiod insensitive with multiple maturity classes based on manipulation of various known maturity genes. Sorghum is predominantly propagated by seed, but some species (*S. halepense* and *S. propinquum*) expand their cover through the production of rhizomes. As such, *Sorghum* species are represented by both annual (maturing in approximately 60–180 days) and perennial types.
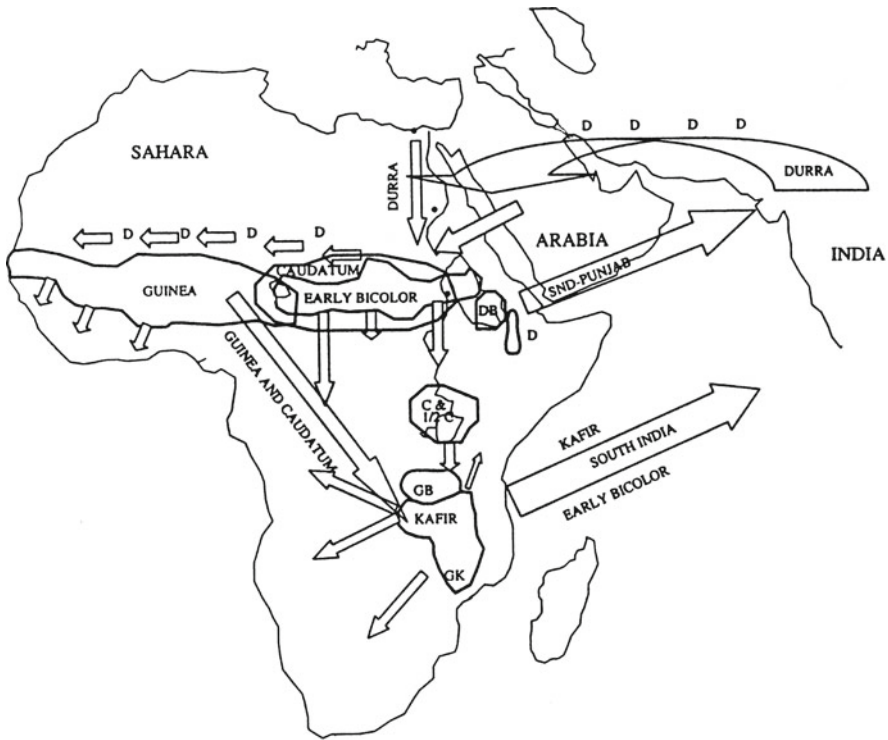
Fig. 2.1 Areas of origin and development for the domesticated races of *bicolor* and possible migration routes

1.2 Evolution and Domestication of the Species

The largest diversity of cultivated and wild sorghum is in Africa (Doggett 1970; de Wet and Harlan 1971; de Wet 1977). The great diversity of *S. bicolor* was created through disruptive selection and by isolation and recombination in the extremely varied habitats of northeast Africa and the movement of peoples carrying the species throughout the continent (Doggett 1970, cited in Miller 1982). There seems no argument that sorghum plants are African in origin, with the earliest known record of sorghum found in 8000 BP charred remains of sorghum at the Nabta Playa in Southern Egypt (Dahlberg and Wasylikowa 1996), but the domestication event(s) may also have taken place elsewhere and more than once.

Based on experiments and on the work of Snowden (1936), Harlan (1995) and his associates confirmed that "all of the races belong to the same biological species and are fully fertile when hybridized." For the cereal sorghums, they identify four wild races and five cultivated races (Harlan and Stemler 1976). The four wild races of *Sorghum bicolor* are arundinaceum, virgatum, aethiopicum, and verticilliflorum. They are now placed in *S. bicolor* subspecies *verticilliflorum*, formerly subspecies *arundinaceum*.

Using de Wet and Rao's interpretation (de Wet and Rao 1986), Doggett (1988) puts the four weedy races under *S. bicolor* subspecies *verticilliflorum* even though de Wet earlier (de Wet 1977) thought they were well-defined ecotypes and should not be given formal taxonomic status.

The cultivated races as presently conceived are (1) bicolor, the primitive type, (2) guinea, (3) kafir, (4) caudatum, and (5) durra. Intermediates that are caused by hybridization of races exhibit characters of both parents. All will also breed with wild species with which they are sympatric. These cultivated races are placed in *S. bicolor* subspecies *bicolor*. Comparison of the distribution of the wild and cultivated races indicates a narrower concentrated band of the cultivated races in the east–west line across and north–south line on the eastern part of the continent than for the wild races. Such a distribution is support for the dispersal of sorghum taxa by migrating peoples across the Sahel-Sudan grasslands and southward from the Nile Valley region along the Great Rift (Murdock 1959; Harlan 1995).

Snowden (1936) took the position that sorghum had separate centers of origins for different types. According to Snowden, wild race aethiopicum gave rise to races durra and bicolor, arundinaceum to guinea, and verticilliflorum to kafir. de Wet and Huckabay (1967) had much the same understanding except they proposed that durras came out of kafirs. Doggett (1965) suggested that the diversity seen in the wild forms might reflect human manipulation and intervention associated with the selection of domesticated types.

Today, bicolor is distributed widely but is nowhere dominant among the African regions with cultivated sorghums. Bicolor, however, is not only widely distributed in Africa but is also apparently ancient in Asia, coastwise from India to Indonesia to China (de Wet and Price 1976). How bicolor migrated out of Africa is unknown, as are the people who were responsible for its diffusion. Cultivated sorghum may have reached China from Indochina by way of the Mekong River or other river valleys. Hawkes (1973) asserts that Semitic speakers from Africa carried their culture to India before 3000 B.C. This may have been one avenue on which sorghum moved off the coast of Africa and into India.

Today, we see the four wild races distributed throughout the African continent, based on their biological traits. The wild race arundinaceum is distributed largely in wet and humid parts of forested central and west Africa along stream banks and in clearings, which are not suitable environments for the cultivated races. The Nile Valley proper is a region of seasonal flooding, with wild grasses such as *S. virgatum* among the first colonizers after the waters recede. It also is found in disturbed riparian habitats in the Sudan. Based on the biology of this wild species, Harlan's group rejects it as the source of the cultivated types. Race aethiopicum is found in the Kassala region of the Sudan, more sparsely to the west along the fringes of the Sahara, and in Ethiopia. Based on the ecology of the wild races, this leaves *S. verticilliflorum*, widely distributed throughout the sorghum-growing areas, including the savanna zone of eastern and southern Africa (although not well represented in Nigeria), and morphologically appropriate as the parent race for *S. bicolor* (Harlan 1992). This position is a change from the position 20 years previously when Harlan and de Wet considered aethiopicum/verticilliflorum as a complex that was the primitive primogeniture of the cultivated races.

Fig. 2.2 Sorghum phenotypic diversity for height (*left*) and panicle (*right*)

The cultivated races also are found in different regions of Africa according to their biological traits and also their histories of distribution (see Fig. 2.2). The guinea race is basically a West African race but also has a distribution in the mountains of eastern Africa that receive high rainfall amounts. In an earlier publication, Harlan and Stemler (1976) considered the guineas to be the oldest of the specialized races because of its relatively wide distribution and diversity. It is better adapted to atmospheric conditions that are more wet and humid than are other domesticated races. Its relatively lax panicle provides for the movement of air among the seeds on individual panicle branches. Other scientists have concluded that as bicolor moved west, it came into contact with wild *S. arundinaceum*, and is now found with it in mixed populations, thus picking up some adaptive genetic material through introgression, and from these the race guinea evolved. All three biotypes can be found in feral populations today, especially across the African savannas. Plants of this race can tolerate up to 5,000 mm of rainfall. Guineas moved into east Africa and humid, foggy southeastern Africa and then were transported, probably from east African ports to the Malabar coast of India (Vishnu-Mittre 1974; Harlan and Stemler 1976; Kimber 2000, Fig. 2.2).

According to Stemler et al. (1975), race caudatum is a later domesticate than bicolor and guinea, having been segregated out of bicolor in the ancestral home territory of that race. Caudatum is associated with peoples speaking the Chari-Nile language group. Groups in areas outside the territory of these peoples use the caudatums; however, such groups have only been within the trade area of the Chari, such as Ethiopia and Cameroon (Stemler et al. 1975). Stemler and her associates concluded that caudatums of the Ethiopian highlands are intrusive, probably acquired by trade. These plants, described as being adapted to harsh conditions, are found most commonly in areas receiving from 250 to 1,300 mm of rain annually. (Stemler et al. 1975, 1977).

De Wet (1978) and Harlan et al. (1976) concluded that race kafir was derived from an early bicolor race, which had been carried east and south from the Savanna belt. Electrophoresis data collected by Schecter and de Wet (1975) suggest that the kafirs are more closely associated with wild race verticilliflorum. The kafir sorghums are very much associated with the Bantu-speaking peoples of eastern and southeastern Africa. It is known that the Bantu entered this part of Africa from the

western forested regions to the north and may have migrated before the guinea race was segregated from the race bicolor. Kafirs may have been taken to the Indian plateau edge ports after discovery of the monsoon wind systems by traders in the western Indian Ocean. This route is different from the route via the northern ports, so it may be a later development than the movement of bicolors to the Indus Valley and Punt, or it may be a second migration into the southern plateau area, having been preceded by bicolor.

The compact panicle and predominantly white seeds of race durra are indications of adaptation to low-rainfall environments with a low risk of grain mold (Mann et al. 1983). The most important grain in Ethiopia, it is found strictly north of the equator in Africa. It is an important type in India and may have been domesticated there (Harlan and Stemler 1976). The name is derived from an Arabic root. Until recently, the durras were almost entirely cultivated by Muslim Africans and Arabic people in Ethiopia. Harlan et al. (1973) reported that the main growers of durra sorghum in Ethiopia are the Muslim Oromo (Gallo), who settled the fertile warm highland almost 500 years ago and have used race durra sorghum as the foundation of their agricultural system. Durras are presently distributed in the mid-altitude highlands of Ethiopia, the Nile Valley of Sudan and Egypt, and in a belt 10–15° north latitude from Ethiopia to Mauritania. They are grown also in the Islamic and Hindu areas of India and Pakistan.

The kaoliangs (galiangs) of China are also thought to be derived from races of bicolor introduced from India to China (Harlan 1995). Alternatively, they may be derived from wild diploid sorghums with which they were compatible (Harlan 1995). At one time they were considered native to Manchuria (Quinby 1974). According to Harlan (1995), the Chinese kaoliangs are more fibrous and have been selected for nongrain uses such as basketry, fencing materials, and house construction materials. Broomcorn sorghum is thought to be part of the story, as are the sorgos such as amber cane. In some regions, the Chinese use the grain in the preparation of a fiery whiskey called mai-tai.

In Southeast Asia and Indonesia, the sorghum are different as well. *S. propinquum* is found in southern China through Thailand, Cambodia, Malaya, and Burma to the Philippines (Burkill 1966). These sorghum are characterized by very large, loose, open panicles and may also have a history different from those of the African-based races (Harlan 1975; Doggett 1988). In fact, recent molecular studies show that *S. propinquum* exhibits approximately an 1.2 % nucleotide difference in coding regions of the genome from *S. bicolor*, suggesting a divergence of 1–2 million years between the two sorghum (Feltus et al. 2004).

1.3 Domestication and Its Effect on Genome and Gene Pool Diversity

Over the past decade, tremendous progress has been made to build the molecular and genomic foundation required to increase our understanding of sorghum diversity in

the genome and gene pool. In complement, these genomic resources currently are being deployed in sorghum improvement efforts across the world. Sorghum represents the first crop genome of African origin to be sequenced (Paterson et al. 2009) and, through coordinated national and international efforts, high-density genetic and physical maps, extensive sets of RFLP and SSR markers, association (Casa et al. 2008) and diversity panels (Deu et al. 2006), nested association mapping populations, etc., are readily available for use for scientific investigations and breeding efforts.

The generation and use of these genomic resources have added to our insights of sorghum domestication and diversity. Future studies will enrich our understanding and provide increasing resolution to quantify and use both wild and domesticated sources of diversity in crop improvement.

At the genomic level, diversity studies have focused on measure of neutral diversity (simple sequence repeats, single nucleotide polymorphisms, etc.), genic diversity, and linkage disequilibrium. In all cases, the domestication process has had a major effect on how much diversity is present and how it is organized in the genome and gene pool.

In recent years, a number of studies of genomic architecture and diversity of sorghum have been undertaken (Casa et al. 2005, 2006, 2008; Hamblin et al. 2004, 2005, 2006; de Alencar Figueiredo et al. 2008). Hamblin and her colleagues (2004) observed an average of one single nucleotide polymorphism (SNP) about every 120 nucleotides of a sample that included a survey of approximately 30 cultivated and wild sorghum accessions and 96 loci representing 29,186 bases of DNA. This is about one-fourth the frequency observed in a comparable sample of maize (Tenaillon et al. 2001). Subsequent studies, across a broader range of sorghum, have yielded similar results (de Alencar Figueiredo et al. 2008).

Studies of linkage disequilibrium in sorghum are of interest because they provide evidence of both equilibrium (e.g., mating system or long-term population structure) and nonequilibrium (e.g., demographic or selective) processes, and also because of their importance in strategies for identifying the genetic basis of complex traits of importance to agriculture. Hamblin and her associates (2005) randomly surveyed six unlinked genomic regions of sorghum and found that patterns of linkage disequilibrium ranged from a few thousand to tens of thousands of bases. An average value of approximately 15 kb suggests that sorghum may be well suited for association studies using a reasonable number of markers. This is in sharp contrast to maize where linkage disequilibrium decays in less than 2 kb in many instances. Moving from studies of genomic to gene pool diversity also has been rapid; this transition likely occurred because of the wealth of molecular tools accessible in sorghum.

Deu and her associates (2006) evaluated a diverse core collection of 210 accessions representing the cultivated races. Two major geographic poles for sorghum evolution and differentiation were established (northern and southern equatorial types). The absence of rare alleles in the southern equatorial accessions (kafir, guinea, and caudatum) fits with the classical view that southern equatorial sorghums evolved later from other African sorghum. Additionally, in support of the historical studies, Deu found that morphological race had a major effect on patterns of genetic diversity.

Additionally, Deu and her colleagues (2006) quantified amount of diversity within race (based on a measure of gene diversity, mean number of alleles, its presence within the defined clusters of accessions) noting that the greatest levels were found with the races bicolor and guinea. Kafir represented the race with the most limited representation of diversity, and this finding likely reflects the classical view of this race and its recent origin and restricted geographic distribution. The assessment of kafir's limited diversity (based on restriction fragment length polymorphisms) is consistent with a complementary simple sequence repeat analysis of wild and domesticated sorghum accessions conducted by Casa and her colleagues (2005).

Like in other crop species, the process of domestication has reduced the amount of diversity present in the cultivated gene pool of the species. Casa and her colleagues (2005) noted the reduction across loci (as measured by simple sequence repeat analysis) was equivalent to 86 % of the diversity as observed in the wild sorghums. Statistical methods for identifying genomic regions with patterns of variation consistent with selection yielded approximately 10 % of the screened loci possibly under some selection pressure. Interestingly, approximately two-thirds of these loci mapped in or near genomic regions associated with domestication-related QTLs (including seed shattering, seed weight, and rhizomatousness).

However, domestication and crop improvement is not always a "one-way" street representing an ever-present move toward reduced diversity. De Alencar Figueiredo and others (2008) studied diversity relationships in sorghum for six candidate genes associated with grain quality and established that in some cases, the genes *Waxy* and *Amylose extender 1*, novel variation was detected (via positive selection), suggesting that postdomestication mutations had occurred and were seen as desirable. Also, these variants subsequently were conserved and increased in frequency and range under human selection. These recent findings have important implications for genetic diversity conservation and use. With increasing access to and cost-effective use of molecular tools, curators now will be able to establish high-resolution studies of diversity in their collections and in natural settings.

2 Centers of Diversity: Current Status of Sorghum Collections

2.1 Global Status of Sorghum

Investigators continue to collect new and additional land races from isolated farmers' fields (Benor and Sisay 2003), and various collections have been assembled world-wide that represents much of the genetic diversity that can be found in sorghum. Major collections from the site of origin of sorghum exist for Ethiopia and Sudan, while other collections representing the major races and working groups of sorghum have been compiled in areas such as Mali, South Africa, India, and China. Internationally, two collections have worked to bring most, if not, all of these various collections together; the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) located in Andhra Pradesh, India and the National Center for Genetic Resources Preservation, located in Fort Collins, Colorado, USA.

Table 2.1 Major sorghum collections from various countries from around the world

Country	Institute	No. of accessions	% of total holdings (194,250 acc.)
USA	USDA-ARS-PGRCU	43,104	22.2
Global	ICRISAT	36,774	18.9
India	NBPGR	18,853	9.7
China	CAAS	18,250	9.4
Ethiopia	IBC	9,772	5.0
Brazil	EMBRAPA	8,017	4.1
Russia	VIR	7,335	3.8
Zimbabwe	NPGRC	7,009	3.6
Australia	DPI	5,403	2.8
Sudan	PGRU-ARC	4,191	2.2
Mali	IER	2,975	1.5
France	CIRAD	2,690	1.4
Kenya	NGBK	1,320	0.7
Zambia	NPGRC	1,005	0.5
South Africa	NPGRC	428	0.2
Malawi	NPGRC	401	0.2
Nigeria	NCGRB	159	0.1
Serbia	Inst. Field and Vegetable crops	152	0.1
Global	ILRI	52	0.0
	Total 19 institutes	167,890	86.0

(From the "Strategy for the Global Ex Situ Conservation of Sorghum Genetic Diversity")

In 2006, the Global Crop Diversity Trust initiated a review of the sorghum collections from the Germplasm Holding Database maintained by Bioversity International (formerly known as IPGRI). This became a major attempt to catalogue collections from around the world and review the condition of those collections. One hundred twenty-two collections were identified and based on input from various experts, a global survey was sent to 57 institutes to gather information on collection numbers, state of preservation, regeneration procedures, and other information regarding each collection. Nineteen institutes responded and from this the beginnings of a world-wide inventory began to form. From this initial review, the Global Crop Diversity Trust brought together sorghum experts from around the world in 2007 to formulate the "Strategy for the Global Ex Situ Conservation of Sorghum Genetic Diversity," which met at ICRISAT (see report at http://www.crop-trust.org/documents/cropstrategies/sorghum.pdf, verified May 5, 2010). Major collections are outlined in Table 2.1.

It became clear from this undertaking that many of these collections represent duplications of various sorghum accessions. One of the major challenges identified by this working group was the lack of relevant passport data available on these collections which could assist in grouping duplicate accessions and developing a more accurate picture of the true collection status of sorghum. Another major concern was the lack of wild relative species within these collections. Only 159 wild relative species were identified in the 19 collections that sent back their surveys, and it is questionable as to how much of that collection has been correctly identified and classified.

2.2   Major Collections

2.2.1   The Americas

The single largest collection of sorghum germplasm resides in the USA, which began collecting various forms of sorghum as early as 1757, when Benjamin Franklin mentioned it in a letter to a Mr. Ward. He was interested in its unique panicle formation, used for making brooms and had brought seed back from Europe to the USA to increase and to share with friends. The Philadelphia Agricultural Society mentioned Guinea Corn in 1810 (Quinby 1974) and references to sorghums such as sorgo, Chinese amber cane, white and brown durras, milo, feterita, and hegari can be found in various publications between 1853 and 1908 (Doggett 1988). Formal collections and distributions by the United States Department of Agriculture began around 1905 when it bought, increased and distributed Dwarf Yellow Milo; further introductions and research began at the Texas Agricultural Experiment Station at Chillicothe, Texas (Quinby 1974). Prior to the introduction of hybrids in 1958, approximately 13,611 accessions of sorghum had been introduced into the USA; however, serious curation of the crop did not take place until the early 1980s. Since then a total of 31,163 accessions have been added to the collection for a total of 44,774 accessions (Table 2.1). Several groups have reviewed the status of the USA collection (Duncan, Bramel-Cox, and Miller 1991; Dahlberg and Spinks 1995). The second largest collection is housed with EMBRAPA in Brazil and is used extensively in breeding programs used to search for abiotic and biotic source of resistance to things such as anthracnose and acid-soil tolerance, two major issues facing sorghum producers in this country. Both Mexico and Argentina maintain collections of over 3,000 accessions for use in their national breeding programs and for research purposes.

2.2.2   Africa

Several collections exist in Africa, the site of sorghum domestication and early distribution (Mann et al. 1983). The Ethiopian and Sudanese collections are two of the most important worldwide, since these two regions are considered the primary sites of sorghum domestication. The discovery of the 8000 BP charred remains of sorghum at the Nabta Playa in Southern Egypt (site E-75-6) were identified by de Wet and Harlan (sorghum taxonomy according to de Wet 1978) as wild *Sorghum bicolor* (L.) Moench. subsp. *arundinaceum* (Desv.) based on spikelet morphology and the small size and shape of the grains (Dahlberg and Wasylikowa 1996). The Jimma Agricultural Technical School began to centralize collections in Ethiopia between 1958 and 1960, and the collection was grown out and characterized in an Experiment Station Bulletin entitled "The Cultivated Sorghums of Ethiopia." This activity was taken over by the Ethiopian Sorghum Improvement Project and continued collections have taken place since the 1970s. Sudan collected a large collection of

landraces in the 1950s which was maintained by the Tozi Research station. This collection was turned over to the Rockefeller Foundation Project in India (Rosenow and Dahlberg 2000). In 1992, collections with known Sudanese origins were grown in Wad Medani for increase, characterization, and distribution. This collection was shared with the USA and 3,182 of these accessions were grown out in St. Croix, USA Virgin Islands in 1993 for verification, seed increase and characterization. The Sudanese collection may be one of the most fully characterized collections in the world to date and contains representatives of most of the major working groups (Dahlberg and Spinks 1995; Dahlberg and Madera-Torres 1997; Rosenow and Dahlberg 2000; Dahlberg, Burke, and Rosenow 2004, Grenier, Bramel, Dahlberg, El-Ahmadi, Mahmoud, Peterson, Rosenow, and Ejeta 2004). Uganda also maintains approximately 2,600 accessions for use in its research programs.

The largest collection outside of Eastern Africa is located with NPGRC in Zimbabwe, with over 7,000 accessions. Western Africa is represented by collections from Mali, Niger, Nigeria, and Burkina Faso, but their numbers are more difficult to pin down. The Mali collection was shared with the USA and a joint project involving the USA, ICRISAT, CIRAD, ORSTOM, and Mali was undertaken in 1997, to grow out, increase, characterize, and distribute the collection in Mali. The collection was also regrown in St. Croix and data sets from both countries were incorporated into the GRIN database. Other smaller collections exist in Africa and represent some of the unique diversity that is found in the world collection.

2.2.3 Asia

The second largest collection worldwide is held in trust at ICRISAT. K.O. Richie transported a Rockefeller Foundation collection of sorghum from Mexico to India in 1957, thus starting the international collection of sorghum. The collection was a cooperative effort between the Rockefeller Foundation and the All-India Sorghum Improvement Program. Comprised of various collections from around the world, this collection began an intensive effort to collect indigenous sorghums from places not fully represented in those early collections (House 1980). International Sorghum numbers were given to the collection, and it grew to its current status of approximately 36,774 accessions. Both China (Qingshan and Dahlberg 2001) and India have collections of over 18,000 accessions; however, the Chinese collection is not well represented in either the ICRISAT or US collection. Much of the diversity of the Indian accessions has been collected in the International collection. Japan and the Philippines have small research collections.

2.2.4 Other Regions

Various other collections exist worldwide; however, it is difficult to know how much of these collections represent unique accessions of sorghum or duplicates of germplasm from various regions from around the world. Russia has approximately

7,000 accessions of sorghum, while Australia reported a collection size of 5,403. Many of the other collections are less than 2,000 and most have been obtained for research purposes in national sorghum improvement programs. With over 160,000 accessions reported in various collections, it is difficult without better phenotypic and genotypic information to fully understand how much diversity of sorghum has been duplicated in the world collections.

3 Assessing Useful Diversity, Breeding, and Race and Working Groups

3.1 Assessing Phenotypic Diversity

The international sorghum community has shown a tremendous amount of interest in evaluating germplasm collections. One of the many difficulties that the working group identified as they worked on the Strategy for the Global Ex Situ Conservation of Sorghum Genetic Diversity was how to bring those data points together in a useful database that could be available for worldwide use.

Several issues were identified that limit the overall utility of the various collections. Passport data was available in some electronic form, but different nomenclature use by the various institutes or collecting agencies made comparison of collections problematic, especially as to the when, where, and how accessions were collected. Different databases, ratings, evaluation techniques, and other issues have made a simple combining of the various data points difficult. Few of the collections have longitude and latitude designations, making spatial evaluation of the collections almost impossible. Many of the collections were done pre-GIS technology and relied on local names and/or villages as collection points. These same concerns are also true for both characterization and evaluation data.

It is clear that there is tremendous diversity within the sorghum collections. The USA has clearly shown this in its descriptors and through various evaluations that they have undertaken on their national collection; however approximately 33 % of the collection has extensive phenotypic descriptor data, while roughly 60 % of the collection has race and working designation, and 40–60 % of the collection has agronomic characteristics (Table 2.2). The collection has been evaluated for photoperiod response in a temperate zone (Table 2.3).

The ICRISAT collection's database contains 24 data points on each of its accessions. Most of the evaluation points reflect agronomic phenotypic characterization and is available from the SINGER database. The Chinese collection has approximately 50 % of its collection characterized, but most of that is not available electronically (see Qingshan and Dahlberg 2001). CIRAD has significant evaluation data, but until datasets can be electronically merged and validated, accessing the phenotypic value of the various collections will be difficult.

Table 2.2 Phenotypic characteristics and evaluation data points taken on various accessions within the US National Sorghum Collection (source GRIN database)

Descriptor	Distinct accessions	Total observations
Chemical		
Acid detergent fiber %	2882	2,914
Brix measurements	1229	1,260
Crude protein %	2882	2,914
In vitro dry matter digestibility	2881	2,914
Fat %	2882	2,913
Phosphorous %	2882	2,914
Starch %	2882	2,914
Sucrose %	1198	1,229
Cytological		
Normal ploidy level	49	49
Disease resistance		
Colletotrichum graminicola	15670	16,399
Peronosclerospora sorghi	4674	6,214
Peronosclerospora sorghi P3	3937	3,938
Claviceps africana	2022	2,022
Cercospora sorghi ELL. and Ex.	306	306
Cercospora fusimaculans	1437	1,470
Helminthosporium turcicum	340	340
Puccinia purpurea	15819	17,402
Sugarcane mosaic virus	427	427
Sorghum yellow banding virus	210	210
Gloeocercospora sorghi	1437	1,470
Growth		
Height uniformity of plant	15651	16,149
Plant height (to top of panicle)	20195	21,577
Seedling vigor	3984	3,987
Insect resistance		
Pseudaletia unipuncta (Fall armyworm)	8940	8,942
Taxoptera graminum (Greenbug)	15990	16,035
Sipha flava (Yellow sugarcane aphid)	5564	5,564
Restorer		
B/R line reaction to A1, A2, and A3	620	1,828
Morphological screening		
% Glume covering of kernel	14655	15,123
Type of awns at maturity	15615	16,101
# of basal tillers per plant	14690	15,158
Panicle branch angle	14637	15,106
Color of endosperm	14540	15,003
Texture of endosperm	14537	15,000
Type of endosperm	14547	15,008
Color of glume	14659	15,128
Glume pubescence	14657	15,127
Inflorescence exsertion	14674	15,139
Phenotypic seed color	18745	19,135

(continued)

Table 2.2 (continued)

Descriptor	Distinct accessions	Total observations
Plumpness of kernel	14532	14,992
Shape of kernel	14540	15,003
% lodging	965	965
Mesocarp thickness	14651	15,119
Leaf midrib color	19081	19,770
Nodal tillering	14680	15,147
Compactness of panicle	15610	16,109
Panicle erectness	15637	16,132
Length of panicle	14641	15,105
Shape of panicle	14663	15,132
Color of pericarp (genetic)	14362	14,791
Plant color	14743	15,210
Form of seed (single, twin)	14642	15,111
Seed shattering	14651	15,121
Spreader	13150	13,479
Seed sprouting tendency	14687	15,155
Juicy/dry midrib	19054	20,401
Waxiness of the stem	14664	15,131
Presence/absence of testa	14713	15,182
Transverse wrinkle	14201	14,666
Other		
Cold tolerant population	173	174
Core subset	2438	2,438
Images present in GRIN	6880	7,089
Sorghum association panel	386	387
Phenology		
Flowering rating 65 and 90 days after planting	32680	32,718
Long day anthesis rating	739	742
Photoperiod sensitivity rating	1533	1,533
Short day anthesis rating	12437	12,763
Production		
Primary plant usage	1390	1,390
Yield potential (1=high, 5=low)	14665	15,130
Overall plant desirability ratings	14729	15,196
Quality		
Grain weathering	14657	15,126
Metabolizable energy (swine)	2882	2,914
Net energy for gain (cattle)	2882	2,914
Net energy for lactation (cattle)	2882	2,914
Net energy for maintenance (cattle)	2882	2,914
% Total digestable nutrients	2882	2,914
Stress		
Aluminum toxicity tolerance	10332	10,384
Manganese toxicity tolerance	7302	7,339
Taxonomy		
Race	22319	23,011
Working group designation	14805	15,262
Total number of observations		7,88,302

Table 2.3 Photoperiod rating of US National Sorghum Collection (source GRIN database)

Definition	Number of accessions
Very early (50 days or less)	197
Early (50–60 days)	1,473
Medium (60–70 days)	3,796
Late (75–90 days)	3,498
Very late (90+ days)	925
Photoperiod sensitive	22,762
Mixed flowering rating	66
Total number of accession screened	32,717

3.2   Breeding Collections

Little work has been done on establishing a robust acquisition system for breeding collections as sorghum breeders retire or pass away. Several collections have been either discarded or archived with little recorded evidence of their existence. This is also true within the private industry as collections have been either neglected or destroyed when a program has ended or a breeder has retired. Several important collections have been noted in research articles and other publications, but are no longer available for distribution. The USDA-ARS tried reviving Dr. Keith Schertz's genetic stocks after he had passed away; however, little of the collection was maintained properly and only a few of his genetic stocks were recovered and placed into GRIN. No concerted effort worldwide has been made to preserve these potentially valuable collections.

Probably one of the most useful breeding collections has been the Sorghum Conversion Program, which has recently been relaunched through the support of the United Sorghum Checkoff Program and MMR Genetics. The program was first described by Stephens, Miller, and Rosenow in 1967, though it was initiated in 1963 with the first planting taking place in Mayagüez, Puerto Rico. They recognized the need to convert photoperiod sensitive sorghums, to combine height, photoperiod insensitive germplasm that would flower in temperate regions. This was partly driven by the need to expand the germplasm base that was currently in use in the USA. As noted earlier, sorghum is unique in that it is grown from sea level to elevations of 2,500 m or more, in both extremely wet and dry environments, and on both poor and rich soils. This has created a tremendously diverse and genetically rich germplasm base that has been mostly ignored because of its photoperiod sensitivity and the difficulty of using these germplasm sources in temperate zones.

As these accessions were converted, many were screened for various diseases and insects and several accessions were identified containing key genes of valuable agronomic characteristics (Duncan et al. 1991). Currently, 702 converted lines have been released. These releases ranged from exotic germplasm that was initially placed into the program for diversity, to agronomically important, high yielding white, tan food grade sorghums (see discussion by Rosenow and Dahlberg 2000). In 2009, the sorghum conversion program was reinitiated and is exploiting the

sorghum sequence to rapidly convert sorghum exotic germplasm sources. Typically, the conversion program took between 8 and 12 years to fully convert a photoperiod sensitive sorghum to a converted insensitive line. Using DarT technology, and markers for both the Ma_1 gene, which is highly linked to photoperiodism, and several of the dwarfing genes, the program hopes to achieve adequate conversion within 2–3 years. This will allow for a much more rapid distribution of needed exotic germplasm to both private and public breeding programs.

3.3 *Working Groups*

The cultivated sorghums fall within the subgenera *Sorghum* (see discussion by Dahlberg 2000). Snowden in 1936 proposed 31 races of sorghum, with varieties within each race, which made classification of sorghum quite difficult. As noted previously, Harlan and de Wet proposed a simplified classification scheme for the cultivated sorghum in 1972, which delineated five major races and ten intermediate races, which is used by many plant breeders today to classify sorghum; however, this simplified classification did not adequately explain the variation found within sorghum. Murty and Govil (1967) proposed a set of Working Groups to try to address this. Harlan (personnel communication) indicated that he and de Wet had wanted to integrate their proposed scheme with that of Murty and Govil to create a system for classification that was both easy to use and explicative of the diversity within the crop. Dahlberg (2000) proposed an integrated system of sorghum classification with this goal.

Little work has been done to evaluate and relate important genetics with various races and working groups, though it is clear that such an endeavor would be useful to sorghum researchers. Working groups such as Zerazeras contain useful sources of tan plant and white seed for use in food systems, while kafirs offer good sources of yield. Some work has been done on heterotic pools, with caudatum and kafir combinations being the most widely used. However, the work by Yang and colleagues (Yang et al. 1996) clearly indicates that Chinese germplasm may offer new sources of heterosis that have not been exploited within sorghum. Other potential sources of new heterotic pools between durras and bicolors offer some yield potential that would be new to sorghum as well (personal observations by author of diversity within the Malian collection). New genomic tools may offer the potential to exploit these pools in the future as more of the world collection becomes fingerprinted and evaluated for genetic diversity.

4 The Future

As competition for global water resources becomes more intense in the twenty-first century, sorghum will increase in importance as a source of food, feed, fiber, and fuel. However, key goals must be accomplished to support the wider use of sorghum.

First, wild and weedy relatives of cultivated sorghum must be effectively collected and conserved. Current collections across the world are void of the exceptional range and diversity of the genus. Particularly wild and weedy materials from Africa and Asia merit inclusion into international and national collections. Second, entries in the collection require more detailed characterization and evaluation if elite germplasm is to be effectively developed. In particular, photoperiod sensitive materials require conversion for appropriate agronomic evaluation. An expanded effort for sorghum conversion also should be a priority.

In complement with the agronomic benefits of increased sorghum production for multiple end-products, sorghum will increase in value for studies of comparative genomics in cereals and grasses (Paterson et al. 2009). Sorghum, because of its smaller and simpler genome, its wealth of readily accessible genomic tools and information, and its breadth of diversity for abiotic and biotic stress resistances will become a cornerstone for integrated studies linking gene pool diversity and crop improvement in the future.

References

Benor S, Sisay L (2003) Folk domestication of sorghum (Sorghum bicolour (L) Moench) land races and its ethnobotanical implications: a case study in northeastern Ethiopia. Ethnobiologia 3:2941

Burkill IH (1966) A dictionary of the economic products of the Malay Peninsula, vol 1, 2nd edn. A.H. Government of Malaysia and Singapore, Kuala Lumpur, Malaysia

Casa AM, Mitchell SE, Hamblin MT, Sun H, Bowers JE, Paterson AH, Aquadro CF, Kresovich S (2005) Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats (SSRs). Theor Appl Genet 111:23–30

Casa AM, Mitchell SE, Jensen JD, Hamblin MT, Paterson AH, Aquadro CF, Kresovich S (2006) Evidence for a selective sweep on Chromosome 1 of cultivated sorghum. Crop Sci 46:S27–S40

Casa AM, Pressoir G, Brown P, Mitchell SE, Rooney WL, Tuinstra MR, Franks CD, Kresovich S (2008) Community resources and strategies for association mapping in sorghum. Crop Sci 48:30–40

Dahlberg JA (2000) Classification and characterization of sorghum. In: Smith CW, Frederiksen RA (eds) Sorghum: origin, history, technology, and production. Wiley, New York, pp 99–130

Dahlberg JA, Burke JJ, Rosenow DT (2004) Development of a sorghum core collection: refinement and evaluation of a subset from Sudan. Econ Bot 58:556–567

Dahlberg JA, Madera-Torres P (1997) Restorer reaction in A_1 (ATx623), A_2 (A_2Tx632), and A_3 (A_3SC103) cytoplasms to selected accessions from the Sudan sorghum collection. ISMN 38:43–58

Dahlberg JA, Spinks M (1995) Current status of the U. S. sorghum germplasm collection. Sorghum Newsl 36:4–11

Dahlberg JA, Wasylikowa K (1996) Image and statistical analyses of early sorghum remains (8000 B. P.) from the Nabta Playa archaeological site in the Western Desert, southern Egypt. Veg Hist Archaeobot 5:293–299

De Alencar Figueirdeo LF, Calatayud C, Dupuits C, Billot C, Rami JF, Brunel D, Perrier X, Courtois B, Deu M, Glaszmann JC (2008) Phylogeographic evidence of crop neodiversity in Sorghum. Genetics 179:997–1008

Deu M, Rattunde F, Chantereu JA (2006) A global view of genetic diversity in cultivated sorghums using a core collection. Genome 49:168–180

de Wet JMJ (1977) Domestication of African cereals. Afr Econ Hist 3:15–32

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

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

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

de Wet JMJ, Price EG (1976) Plant domestication and indigenous African agriculture. In: Harlan JR, De Wet JMJ, Stemler A (eds) Origins of African plant domestication. Mouton, The Hague, pp 453–464

de Wet JMJ, Rao KEP (1986) Wild sorghums and their significance in crop improvement. In: Proceedings of sorghum conference, 20–23 August 1986, Shenyang, China. Cited in J. Dahlberg (2000)

Doggett H (1965) Disruptive section in crop development. Nature 4981:279–280

Doggett H (1970) *Sorghum* (Tropical Agriculture Series). Longmans, London

Doggett H (1988) Sorghum. Wiley, New York

Duncan RR, Bramel-Cox PJ, Miller FR (1991) Contributions of introduced sorghum germplasm to hybrid development in the USA. In: Shands HL, Wiesner LE (eds) Use of plant introductions in cultivar development: part 1. CSSA Spec. Publ. No. 17. CSSA, Madison, WI, pp 69-102

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) A SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. Genome Res 14:1812–1819

Grenier C, Bramel PJ, Dahlberg JA, El-Ahmadi A, Mahmoud M, Peterson GC, Rosenow DT, Ejeta G (2004) Sorghums of the Sudan: analysis of regional diversity and distribution. Genet Resour Crop Evolut 51:489–500

Hamblin MT, Casa AM, Sun H, Murray SC, Paterson AH, Aquadro CF, Kresovich S (2006) Challenges of detecting directional selection after a bottleneck: lessons from *Sorghum bicolor*. Genetics 173:953–964

Hamblin MT, Mitchell SE, White GM, Gallego J, Kukatla R, Wing RA, Paterson AH, Kresovich S (2004) Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium, and selection in a diverse sample of *Sorghum bicolor*. Genetics 167:471–483

Hamblin MT, Salas Fernandez MG, Casa AM, Mitchell SE, Paterson AH, Kresovich S (2005) Equilibrium processes cannot explain high levels of short- and medium-range linkage disequilibrium in the domesticated grass *Sorghum bicolor*. Genetics 171:1247–1256

Harlan JR (1975) Crops and man. American Society of Agronomy, Madison, WI

Harlan JR (1992) Indigenous African agriculture. In: Cowan CW, Watson PJ (eds) The origins of agriculture: an international perspective. Smithsonian Institution Press, Washington, DC, pp 59–70

Harlan JR (1995) The living fields: our agricultural heritage. Cambridge University Press, Cambridge

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

Harlan JR, de Wet JMJ, Price EG (1973) Comparative evolution of cereals. Evolution 27:311–351

Harlan JR, de Wet JMJ, Stemler ABL (eds) (1976) Origins of African plant domestication. Mouton, The Hague

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

Hawkes J (1973) The First Great Civilizations. Penguin Books, Hammondsworth, Middlesex

House LR (1980) A guide to sorghum breeding. International Crops Research Institute for the Semi-Arid Tropics, Patancheru

Kimber C (2000) Origins of domesticated sorghum and its early diffusion to India and China. In: Smith CW, Frederickson RA (eds) Sorghum: origin, history, technology, and production. Wiley, New York, pp 3–98

Mann JA, Kimber CT, Miller FR (1983) The origin and early cultivation of sorghums in Africa. Texas Agric. Exp. Stn. Bull. 1454

Miller FR (1982) Genetic and environmental response characteristics of sorghum. International Crops Research Institute for the Semi-Arid Tropics Symposium on Sorghum, 2–7 November 1981, Patancheru, India, pp 393–446

Murdock G (1959) Africa: its peoples and their culture history. McGraw Hill, New York

Murty BR, Govil GN (1967) Description of 70 groups in genus Sorghum based on a modified Snowden's classification. Indian J Genet Plant Breed 27:75–90

Qingshan L, Dahlberg JA (2001) Chinese sorghum genetic resources. Econ Bot 55:401–425

Quinby JR (1974) Sorghum improvement and the genetics of growth. Texas A&M University Press, College Station, TX

Paterson A, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellston U, Mitros T, Poliakov A, Schmutz J, Spannag M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, ur-Rahman M, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457:551–556

Rosenow DT, Dahlberg JA (2000) Collection, conversion, and utilization of sorghum. In: Smith CW, Frederiksen RA (eds) Sorghum: origin, history, technology, and production. Wiley, New York, pp 309–328

Schecter Y, de Wet JMJ (1975) Comparative electrophoresis and isozyme analysis of seed proteins from cultivated races of sorghum. Am J Bot 62:254–261

Snowden JD (1936) The cultivated races of sorghum. Adlard & Sons, Ltd., London

Stemler ABL, Harlan JR, de Wet JMJ (1975) Evolutionary history of cultivated sorghums (*Sorghum bicolor* [Linn.] Moench) of Ethiopia. Bull Torrey Bot Club 102:325–333

Stemler ABL, Harlan JR, de Wet JMJ (1977) The sorghums of Ethiopia. Econ Bot 31:446–450

Stephens JC, Miller FR, Rosenow DT (1967) Conversion of alien sorghums to early combine genotypes. Crop Sci 7:396

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc Natl Acad Sci U S A 98:9161–9166

Vishnu-Mittre A (1974) The beginning of agriculture, paleobotanical evidence in India. In: Hutchinson JB (ed) Evolutionary studies in world crops: diversity and change in the Indian subcontinent. Cambridge University Press, Cambridge, pp 3–30

Yang W, de Oliveira AC, Godwin I, Schertz K, Bennetzen JL (1996) Comparison of DNA marker technologies in characterizing plant genome diversity: variability in Chinese Sorghums. Crop Sci 36:1669–1676