Genet Resour Crop Evol
DOI 10.1007/s10722-012-9921-8

RESEARCH ARTICLE

Population genetic structure of *in situ* wild *Sorghum bicolor* in its Ethiopian center of origin based on SSR markers

Asfaw Adugna · Allison A. Snow · Patty M. Sweeney · Endashaw Bekele · Evans Mutegi

Received: 3 May 2012 / Accepted: 1 October 2012

Abstract *In situ* population studies of wild relatives of crops are crucial for the conservation of plant genetic resources, especially in regions with high genetic diversity and a risk of local extinction. Ethiopia is the center of origin for sorghum, yet little is known about the genetic structure of extant wild populations. Using 9 Simple Sequence Repeat loci, we characterized 19 wild populations from five regions, 8 local cultivar populations from three regions, and 10 wild sorghum accessions from several African countries. To our knowledge, this is the most comprehensive study to date of *in situ* wild sorghum populations in Africa. Genetic diversity corrected for sample size was significantly greater in the wild populations *in situ* than in local cultivars or the accessions. Approximately 41 % of the genetic variation in the wild plants was partitioned among populations, indicating a high degree of differentiation and potential value for germplasm conservation, and the average number of migrants (N_m) per generation was 0.43. Cluster analyses showed that some wild populations were grouped by geographic region, whereas others were not, presumably due to long-distance seed movement. Four wild populations from disjunct regions formed a unique cluster with an Ethiopian accession of subsp. *drummondii* and probably represent a weedy race. STRUCTURE and other analyses detected evidence for crop-wild hybridization, consistent with previous molecular marker studies in Kenya, Mali, and Cameroon. In summary, *in situ* wild sorghum populations in Ethiopia harbor substantial genetic diversity and differentiation, despite their close proximity to conspecific cultivars in this crop/wild/weedy complex.

Keywords Diversity · Gene flow · Landrace · Microsatellites · *Sorghum bicolor* · Transgenic sorghum · Wild sorghum

Electronic supplementary material The online version of this article (doi:10.1007/s10722-012-9921-8) contains supplementary material, which is available to authorized users.

A. Adugna · E. Bekele
Department of Biology, Addis Ababa University, P.O. Box 1176, Addis Ababa, Ethiopia

Present Address:
A. Adugna (✉)
Melkassa Agricultural Research Center, P.O. Box 1085, Adama, Ethiopia
e-mail: asfaw123@rediffmail.com

A. A. Snow · P. M. Sweeney · E. Mutegi
Department of Evolution, Ecology and Organismal Biology, Ohio State University, Columbus, OH 43210, USA

Introduction

The primary gene pools of many crop plants are depleted in genetic variability and breeders often rely on wild relatives for sources of important agronomic traits (e.g., Langridge and Chalmers 2004). Therefore, wild relatives and locally adapted cultivars (landraces)

Published online: 12 October 2012

are often viewed as sources of "genetic insurance" against emerging pests or changing abiotic conditions (e.g., Tanksley and McCouch 1997). Thus, there has been continued interest in studying the wild relatives of domesticated species as reservoirs of genetic variation for crop improvement (Escalante et al. 1994).

Here, we focus on the population genetics of wild relatives of sorghum (*Sorghum bicolor* (L.) Moench), which is the fifth most important food crop in the world and a staple food crop in many regions of Africa and Asia (Ejeta and Grenier 2005). In Ethiopia, sorghum is the third most important food cereal after tef and maize (e.g., CSA 2010). Ethiopia is one of the centers of origin of sorghum (De Wet and Harlan 1971; Doggett 1988), and diversified forms of the crop and its wild relatives represent possible sources of germplasm for crop improvement. For example, genes that confer resistance to ergot and green bug have been obtained from wild sorghum species (Reed et al. 2002; Wu et al. 2006), and Singh and Axtell (1973) identified two high-lysine Ethiopian sorghum varieties. Moreover, recent studies identified sources of "stay-green" drought tolerance derived from sorghum lines native to Ethiopia (Borrell et al. 2000). As a center of origin for domesticated sorghum, Ethiopia may harbor unique wild germplasm that is worthy of further conservation efforts.

Population genetic studies are useful for determining how genetic diversity is partitioned within and among populations of native plants and for inferring levels of gene flow among populations. This baseline knowledge provides a starting point for designing *in situ* and *ex situ* conservation programs (e.g., Rice 2004; Ellstrand and Elam 1993). Determining the level of gene flow between crops and their wild relatives also is critical for conserving wild germplasm (Ellstrand et al. 1999; Jarvis and Hodgkin 1999). Because crops are typically genetically impoverished compared to their wild relatives (Ladizinsky 1985), overwhelming gene flow from the crop is expected to deplete genetic diversity in wild populations (Ellstrand et al. 1999). Hybridization has been implicated in the extinction of certain wild crop relatives, mainly due to outbreeding depression and genetic swamping (Small 1984; Ellstrand et al. 1999). Another practical consequence of crop-to-wild gene flow is related to the evolution of more aggressive weeds, perhaps facilitated by fitness-enhancing transgenes (e.g., Snow et al. 2005). Ellstrand et al. (1999) found that crop-to-weed gene flow has been implicated in the evolution of enhanced weediness in wild relatives of 7 of the world's 13 most important crops.

In Ethiopia, it is likely that natural cross-fertilization between cultivated sorghum and its wild relatives occurs, since the two forms coexist, have overlapping flowering times (Tesso et al. 2008; Ayana et al. 2001), and form putative crop-wild hybrid populations (see review by Ejeta and Grenier 2005). Cultivated sorghum is autogamous, but outcrossing rates can range from 2 to 40 % (Pedersen et al. 1998; Djé et al. 2004; Schmidt and Bothma 2006; Barnaud et al. 2008). Recent studies by Muraya et al. (2011) and Adugna et al. (unpublished data) documented variable and higher outcrossing rates in wild sorghum (~31–75 %), a feature that could facilitate crop-to-wild gene flow. Previous studies involving molecular markers document extensive crop-wild hybridization in wild or weedy sorghum populations in Kenya (Mutegi et al. 2011), Mali (Sagnard et al. 2011), and Cameroon (Barnaud et al. 2009).

Another consequence of crop-wild hybridization is the introgression of transgenic traits, which might affect the fitness of wild plants. Genetic transformation of sorghum has been undertaken using *Agrobacterium tumefaciens* (e.g., Zhao et al. 2000) and biolistic bombardment (e.g., Zhu et al. 1998). Liang and Gao (2001) successfully generated transgenic sorghum plants with a gene from rice, encoding chitinase for resistance to *Fusarium* stalk rot, and another gene, also from rice, encoding a thaumatin-like protein for drought tolerance and disease resistance. The Africa Biofortified Sorghum (ABS) project developed the first generation transgenic sorghum line (ABS #1) that possesses grain with a 50 % increase in lysine (Zhao et al. 2003). Also under development is transgenic sorghum (ABS #2) which improves grain digestibility and has other nutritional benefits (Zhao 2007).

The presence of wild and weedy types of sorghum (Ayana et al. 2001; Tesso et al. 2008) and heterogeneous environmental conditions in Ethiopia make the country an ideal place for conducting population genetic studies of wild sorghum. Few previous studies have focused on wild sorghum populations *in situ* in the crop's center of origin. Ayana et al. (2001) surveyed 4–10 plants from each of 11 populations in western and central Ethiopia in 1998 using RAPD markers. Unlike our current study, they reported low levels of genetic differentiation among populations and suggested that this was due to small population size. Mutegi et al.

(2011) sampled wild and cultivated plants in a country-wide survey of Kenya, but they did not attempt to characterize populations (they sampled one plant from each population). Other published studies focus on historical accessions of wild sorghum, typically from several countries or continents (e.g., Morden et al. 1990; Aldrich et al. 1992; Casa et al. 2005). Therefore, the objectives of the present study were to:

1. Determine the extent to which *in situ* wild sorghum populations in Ethiopia are genetically differentiated from representative accessions of wild *S. bicolor* subspecies and races obtained from ICRISAT.
2. Determine the distribution of genetic variation within and among *in situ* wild populations to infer levels of gene flow and aid efforts to conserve genetic diversity.
3. Evaluate indirect evidence for extensive crop-wild hybridization by testing for admixture of SSR markers in wild and landrace populations from the same regions.

Materials and methods

Sample collection and acquisition

De Wet (1978) recognized three *Sorghum bicolor* (L.) Moench subspecies. These are *S. bicolor* subsp. *bicolor*, which is comprised of five basic races of cultivated sorghum and their 10 derivatives; *S. bicolor* subsp. *drummondii* (Steud.) De Wet (based on *S. drummondii* (Steud.) Millsp. et Chase), which includes stabilized derivatives of hybridization among grain sorghums and their closest wild relatives, and *S. bicolor* subsp. *arundinaceum* (Desv.) De Wet et Harlan (based on *S. arundinaceum* (Desv.) Stapf in Prain), which comprises four races, which are the closest wild relatives of the cultivated sorghum (aethiopicum, virgatum, arundinaceum, and verticilliflorum). In this study we refer to all non-cultivated sorghum as "wild" as in Aldrich et al. (1992) and Mutegi et al. (2012).

DNA samples from 19 *in situ* wild sorghum populations were collected in October and November 2008 in five major sorghum growing regions of Ethiopia (Table 1; Fig. 1). Each collection site was considered to represent a population. The distances among populations ranged from 3.5 to 828 km, based on GPS coordinates, and their elevations ranged from ~600–1,800 m asl. The populations were designated as intermixed with cultivated sorghum (≤500 m) *vs.* more isolated from the crop (>500 m), as indicated in Table 1. One population from Awash National Park was collected in a non-agricultural region, 6 others were >500 m away from the crop, and the remaining 12 populations were either intermixed with the crop or found within 500 m from cultivated sorghum. We obtained 20 individual plant samples from each population.

In November 2009, we also collected DNA samples from 20 plants in each of eight cultivated populations that were planted near wild populations shown in Table 1, in order to examine evidence for crop-wild hybridization. Four cultivars were local landraces of durra, caudatum or durra-caudatum, and one, WC2, was an improved cultivar of caudatum. Two of the durra cultivars, at WC3 and TC1, were known locally as "Degalit".

To compare wild sorghum populations with earlier accessions from several countries in Africa, we sampled DNA from 10 individual plants from each of 10 wild sorghum accessions obtained from ICRISAT (International Crops Research Institute for the Semi-Arid Tropics; Table 2). The goal of including these materials was to determine whether our Ethiopian wild sorghum collections clustered with any of the accessions according to the sub-species and races. Five of the ICRISAT accessions in our study were also used by Morden et al. (1990), Aldrich et al. (1992), and/or Casa et al. (2005). These were accessions IS-14300, IS-14315, IS-18810, IS-14232, and IS-14485).

DNA extraction and SSR fragment detection

DNA was extracted in two steps: *in situ* collection of leaf squashes from live plants using the Whatman FTA cards and Whatman chromatography paper in Ethiopia, and sample purification in the laboratory following the manufacturer's procedure and optimized for sorghum by Adugna et al. (2011). Polymerase chain reactions (PCR) were run using 12 sorghum microsatellite loci that were previously mapped (Brown et al. 1996; Taramino et al. 1997; Bhattramakki et al. 2000; Li et al. 2009) and represented all of the 10 sorghum linkage groups (Supplementary Table S1). These loci were selected based on high polymorphism and compatibility for multiplexing. Three loci (Sb6-342, Sb4-15, and

Table 1 Characteristics of sites where DNA samples were collected from 19 wild populations and 8 cultivated populations in five geographical regions of Ethiopia. Sample sizes are shown in Table 4

Region	Population code		Specific location	Longitude (N)	Latitude (E)	Wild: Isolated or Mixed[a]	Altitude (masl)
	Wild	Cultivated					
Ghibe	G1		Gibe Berha/Sateri	8° 11′49″	37° 33′90″	I	1,425
	G2	GC2	Ghibe river bridge	8° 13′45″	37° 34′34″	I	1,115
		GC1	GhibeILRI	8° 14′21″	37° 34′55″		1,149
	G3		Boqata and Serti	8° 15′43″	37° 37′68″	M	1,607
Awash/Hararghe	AW		Awash National Park	8° 56′57″	40° 05′49″	I	1,044
	H1		Bisidimo	9° 12′51″	42° 13′01″	M	1,386
	H2		Kara (Gerbi)	8° 52′53″	40° 42′14″	M	1,709
Pawe/Metekel	P1		Kanensan	11° 01′24″	36° 27′34″	I	1,682
		PC2	Village5	11° 16′95″	36° 23′29″		1,040
	P2		Village3	11° 13′65″	36° 21′93″	I	1,045
	P3		Village7	11° 21′19″	36° 23′32″	M	1,076
	P4	PC1	Mandura	11° 05′61″	36° 25′67″	M	1,404
	T1	TC1	AlamataGerjele	12° 26′28″	39° 36′03″	M	1,462
Tigray	T2		Subday	14° 21′09″	38° 10′16″	I	1,808
	T3		Humera Agric. Res. Ctr.	14° 15′80″	36° 37′67″	M	609
	T4		HumeraDonkey	14° 14′09″	36° 41′34″	M	631
	T5		HagereselamIdris	14° 05′66″	36° 58′71″	M	820
N. Wello/S. Tigray		WC1	Jara Kechema	10° 30′93″	39° 56′88″		1,433
	W1		Hijira	10° 38′53″	39° 55′20″	I	1,425
	W2	WC2	Abuare	12° 05′29″	39° 39′71″	M	1,426
	W3		Zobel	12° 11′41″	39° 46′12″	M	1,712
	W4	WC3	Kobo town side	12° 08′27″	39° 37′71″	M	1,500

[a] M = wild population mixed with cultivated sorghum or within 500 m; I = wild population isolated from cultivated sorghum (>500 m); *masl* meters above sea level

Sb5-236) failed to amplify in many samples from 4 of the 19 wild populations (populations AW, H2, T1, and W4) and hence were excluded from the final analysis (Supplementary Table S1; this omission did not change the conclusion based on all 12 loci for the remaining 15 populations; unpublished data).

PCR followed the QIAGEN® multi-master mix kit protocol for SSR multiplex, and forward primers were labeled with different fluorescent colors. PCR was carried out in 12 µl total volume of reaction mix containing 1 µM of each primer pair in a multiplex, 1 µl of template DNA, 2.6 µl of sterile ddH_2O, 6 µl of QIAGEN® Multiplex PCR 2X Master mix. PCR were run in a Master cycler (Eppendorf™) with an initial denaturation step of 15 min at 95 °C, followed by 35 cycles of 30 s at 94 °C, 90 s at 58 °C, 60 s at 72 °C, 30 min at 60 °C, and held at 4 °C following QIAGEN® protocol for microsatellite multiplexes.

For fragment-size determinations, 2 µl of the PCR product was diluted with 14 µl of ddH_2O and then 2 µl of the diluted PCR product was added to 14 µl of 36:1 Hi-Di-Formamide: GenScan™/350 Rox™ size standard in a 96 well microtiter plate and was denatured at 95 °C for 5 min and cooled on ice for at least 5 min. Allele size scoring of the PCR fragments was done by ABI 3100 Genetic Analyzer (DNA sequencer) and sizes were read using the associated GeneMapper 3.7 software (Applied Biosystems Inc., CA, USA). To verify the repeatability of allele scoring, we included samples of a well-studied sorghum inbred line, BTx623, by obtaining seeds from the Department of Agronomy, Purdue University. To exclude the possible effects of imprecise DNA fragment sizes due to stuttering, large allele drop out, or null alleles on genotyping, the software Allelobin (Prasanth et al. 2006) was used to classify observed SSR allele sizes into representative

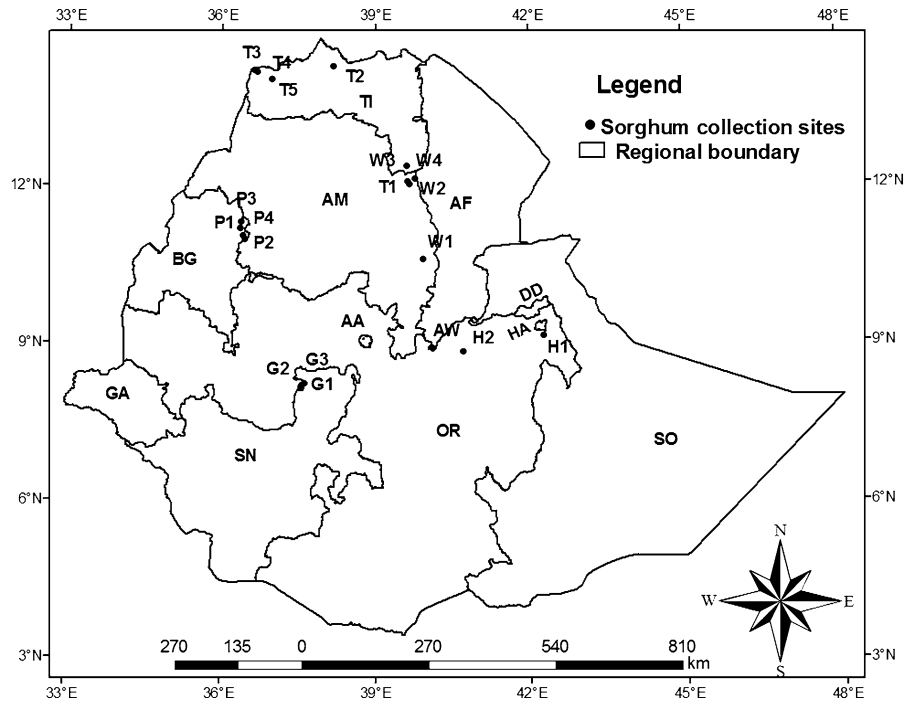

Fig. 1 Map of Ethiopia showing collection sites for wild sorghum populations. Key of abbreviations of the regional states: *AA* Addis Ababa, *AF* Afar, *AM* Amhara, *BG* Benishangul-Gumuz, *DD* Dire Dawa, *GA* Gambella, *HA* Harari, *OR* Oromiya, *SN* Southern Nations, *SO* Somali, and *TI* Tigray. See Table 1 for details and collection sites for cultivated sorghum. Note that Site T1 is located near W2, W3, and W4

Table 2 List of wild sorghum accessions acquired from ICRISAT

No other information was available regarding the source populations or collection dates for these accessions; we suspect that the USA accession was originally collected in Africa (see text)

Subspecies	Race	Source of accession	Accession no.
verticilliflorum	*arundinaceum*	South Africa	IS-14300
verticilliflorum	*arundinaceum*	Swaziland	IS-14315
verticilliflorum	*virgatum*	USA	IS-18804
verticilliflorum	*virgatum*	Egypt	IS-18810
verticilliflorum	*arundinaceum*	Angola	IS-14232
verticilliflorum	*verticilliflorum*	Sudan	IS-18864
verticilliflorum	*aethiopicum-1*	Sudan	IS-14485
verticilliflorum	*aethiopicum-2*	Sudan	IS-18822
drummondii		Sudan	IS-18920
drummondii		Ethiopia	IS-11490

discrete allele sizes using a variation of the least-square minimization algorithm of Idury and Cardon (1997).

SSR polymorphism and analysis of genetic diversity

Different measures of diversity were analyzed using two or more software programs in order to verify consistency of the results. For estimating observed heterozygosity, total expected heterozygosity, polymorphism information content (PIC) (Botstein et al. 1980), and allele frequency based genetic distance analysis, we used PowerMarker software V3.25 (Liu 2005) and FSTAT software (Goudet 2002). Nei's heterozygosity estimates (Ho, Hs, and Ht) were computed using FSTAT software (Goudet 2002). Because the number of distinct alleles and the number of private alleles depend on sample size, they can be difficult to interpret when sample sizes differ across populations. The rarefaction method (first used by Hulbert (1971) for allelic richness and private allelic richness trims unequal samples to the same standardized sample size, a number less than or

equal to the smallest sample size across populations (Szpiech et al. 2008). This method, implemented in HP-Rare 1.1 software (Kalinowski 2005) was used to compute allelic richness (R_s) and private allelic richness (R_p). Significance of differences in gene diversity, allelic richness, and private allelic richness was compared among the three sorghum types (wild populations, cultivars and ICRISAT-accessions) using Wilcoxon's signed rank test. Kruskall-Wallis test was also used to compare differences in the diversity parameters (H_e, R_s and R_p) among populations in each of the three sorghum types. Shared alleles distance matrix (Chakraborty and Jin 1993) was used to construct UPGMA (unweighted pair-group method using arithmetic averages) clusters for the various populations using PowerMarker, and the resulting UPGMA dendrograms were viewed using TreeView 1.6.6 (Page 2001, available at http://taxonomy.zoology.gla.ac.uk/rod/rod.html). To study the distribution of the components of variance within and among the 19 Ethiopian wild sorghum populations, analysis of molecular variance (AMOVA) was performed using Arlequin v 3.1 (Excoffier et al. 2005).

Population structure and estimated gene flow

Wright's (1951) fixation index (F_{ST}) was computed using FSTAT software v.2.9.3.2 (Goudet 2002). Further, the pattern of population structure and detection of probable introgression was visualized using a Bayesian model based clustering method implemented in STRUCTURE software, Version 2.2 (Pritchard et al. 2000). For this, the admixture model with correlated allele frequencies was used. A burn-in period of 500,000 was used followed by 1,000,000 Markov Chain Monte Carlo (MCMC) replications for data collection for $K = 1$ to $K = 11$ groups. For each K value, five replicates were run. This procedure clusters individuals into populations and estimates the proportion of membership in each population for each individual (Falush et al. 2003). Moreover, the model does not assume a particular mutation process (Pritchard et al. 2000). The optimum number of clusters was predicted between $K = 1$ and $K = 11$ following the simulation method of Evanno et al. (2005) using the web based software STRUCTURE HARVESTER v0.6.8. (Earl and von Holdt 2011).

An indirect estimate of gene flow based on the number of migrants per generation (N_m) was computed using Wright's (1951) fixation index (F_{ST}) as $N_m = (1 - F_{ST})/4F_{ST}$. Further, to get additional information on gene flow, the whole data set was subjected to principal component analysis using R program package (R Development Core Team 2008) and a neighbor-joining dendrogram was constructed based on pair wise simple matching distance using DARwin 5 program software (Perrier and Jacquemoud-Collet 2006). The geographic distance matrix of the sampling sites was calculated from geographical coordinates marked with the aid of GPS using the Geographic Distance Matrix Generator (GDMG) version 1.2.3 software of the American Museum of Natural History, Center for Biodiversity and Conservation (http://biodiversityinformatics.amnh.org/open_source/gdmg/index.php). The correlation between the resulting geographic distance matrix of the collection sites and Rousset and Raymond (1997) genetic distance matrix, which is most convenient for isolation-by-distance models (Templeton 2006), was computed and the significance was tested by a Mantel test (Mantel 1967) using IBDWS v. 1.52 web-based program software (Bohonak 2002) to examine the geographic pattern of gene flow. Rousset and Raymond (1997) genetic distance was computed as $\frac{F_{ST}}{1-F_{ST}}$. In order to estimate the level of divergence between wild and cultivated sorghum, F_{ST} was computed by grouping the whole data set into wild and cultivated using PowerMarker (Liu 2005). For this purpose, only those wild and cultivated sorghum populations that were collected in the same region (see Table 1) were included.

Results

Genetic diversity of wild populations, accessions, and cultivars

The observed microsatellite allele size ranges were more or less similar to those reported in earlier studies using the same loci (Supplementary Table S1). The mean number of observed alleles and polymorphic information content (PIC) are presented in Table 3. On average the 9 SSR loci included 18.33, 10.78, and 12.67 alleles in the wild populations, ICRISAT-accessions, and cultivars, respectively. The PIC in all populations ranged from 0.66 (SbKAFGK1) to 0.92 (Sb1-1) with a mean of 0.81. Moreover, the proportion of loci without missing alleles in our data set ranged from 0.96 (Sb6-57) to 1.0 (Sb5-206, Sb5-256, Sb4-121, and SBKAFGK1).

Measures of genetic diversity for wild sorghum populations, cultivars and ICRISAT-accessions are presented in Table 4. Mean allelic richness over the nine SSR loci was 14.93 in the wild sorghum pool, which was significantly higher ($P \leq 0.01$) compared with cultivars (Rs = 10.0) and the ICRISAT accessions (Rs = 11.41). The wild sorghum populations were also characterized by significantly higher levels ($P \leq 0.05$) of private allelic richness (Rp = 3.79) than both the cultivars (Rp = 1.35) and the accessions (Rp = 1.59). Gene diversity was also relatively higher in the wild pool (He = 0.89) compared both with the cultivars (He = 0.68) and the accessions (0.79), although the difference was only significant for the wild *vs.* cultivar comparison ($P \leq 0.01$). The level of diversity varied significantly among the wild populations (Table 4) in terms of allelic richness (range: 1.87 in T1 to 4.98 in G2; $P < 0.001$) and in terms of gene diversity (range: 0.17 in T1 to 0.75 in G1; $P < 0.001$), but not in terms of private allelic richness (range: 0.02 in W4 to 0.50 in P2; $P > 0.05$). Observed heterozygosity (Ho) was > 0.50 in population T4, while populations H2 and T1 had the lowest observed heterozygosity (Ho = 0.01 and 0.04, respectively).

Relationship of wild populations to ICRISAT accessions

This comparison was made to investigate whether the unidentified Ethiopian *in situ* collections cluster with the already identified accessions from ICRISAT according to their subspecies and races. In the dendrogram (Fig. 2) three clusters joined ICRISAT-accessions with the Ethiopian wild sorghum populations. One cluster (V) joined the two Ghibe (G1 and G2) populations with Sudan *aethiopicum*-1. Cluster XI joined W3 with the Sudan *drummondii* accession. Cluster XII joined T1, H2, AW, and W4 with the Ethiopian *drummondii* accession. Below, we present additional evidence showing that these four wild populations are genetically distinct from the others. The four wild populations from Pawe grouped together (Cluster IV). The accession labeled "USA-*virgatum*" clustered closely with "Egypt-*virgatum*" (Cluster I). Sudan *aethiopicum*-2 remained single in Cluster II. All of the three *arundinaceum* accessions from the Southern Africa region (South Africa, Swaziland, and Angola) and the only *verticilliflorum* accession from Sudan joined together in Cluster III.

Table 3 Indices of diversity of wild and cultivated sorghum for the 9 microsatellite loci

SSR Locus	NA	PIC	R_s	H_o	H_s	H_t	F_{ST}
Cultivated (N = 160)							
Sb5-206	14.0	0.75	4.29	0.16	0.49	0.78	0.41
Sb1-1	27.0	0.85	5.77	0.15	0.53	0.87	0.42
Sb6-34	8.0	0.50	2.99	0.11	0.31	0.54	0.47
Sb5-256	3.0	0.54	2.24	0.32	0.38	0.61	0.41
Sb4-72	9.0	0.61	3.41	0.13	0.41	0.66	0.42
Sb6-84	12.0	0.72	3.88	0.13	0.45	0.75	0.44
Sb4-121	8.0	0.62	3.42	0.05	0.41	0.68	0.42
Sb6-57	8.0	0.48	2.58	0.12	0.38	0.57	0.36
SbKAFGK1	8.0	0.63	2.79	0.07	0.38	0.69	0.49
Mean	10.78	0.63	3.48	0.13	0.41	0.68	0.43
Ethiopian Wild (N = 380)							
Sb5-206	27.0	0.92	4.48	0.21	0.63	0.92	0.33
Sb1-1	32.0	0.92	3.85	0.17	0.53	0.93	0.44
Sb6-34	15.0	0.85	3.97	0.33	0.62	0.87	0.30
Sb5-256	11.0	0.71	2.94	0.07	0.48	0.76	0.38
Sb4-72	20.0	0.84	4.00	0.16	0.60	0.87	0.32
Sb6-84	21.0	0.82	3.39	0.20	0.52	0.83	0.38
Sb4-121	14.0	0.85	3.44	0.14	0.57	0.87	0.35
Sb6-57	15.0	0.73	2.85	0.14	0.42	0.76	0.46
SbKAFGK1	10.0	0.61	2.43	0.23	0.37	0.65	0.44
Mean	18.33	0.81	3.48	0.18	0.53	0.83	0.38
ICRISAT wild (N = 100)							
Sb5-206	17.0	0.82	3.02	0.26	0.43	0.84	0.51
Sb1-1	20.0	0.89	3.32	0.16	0.52	0.90	0.45
Sb6-34	8.0	0.77	2.13	0.09	0.33	0.80	0.62
Sb5-256	4.0	0.50	1.88	0.12	0.32	0.56	0.46
Sb4-72	16.0	0.85	2.93	0.13	0.44	0.87	0.53
Sb6-84	11.0	0.83	2.21	0.16	0.38	0.85	0.58
Sb4-121	9.0	0.78	2.46	0.14	0.45	0.81	0.47
Sb6-57	10.0	0.78	2.72	0.10	0.43	0.81	0.50
SbKAFGK1	8.0	0.58	2.18	0.12	0.30	0.61	0.53
Mean	11.44	0.76	2.54	0.14	0.40	0.78	0.52

NA Number of alleles, *PIC* polymorphism information content, R_s allelic richness, H_o observed heterozygosity, H_s the mean expected heterozygosity in the sub populations, H_t expected heterozygosity in the total population, F_{ST} Wright (1951) fixation index

Genetic structure and gene flow in wild populations

Wild populations displayed considerable genetic differentiation. Overall value of F_{ST} for the 19 wild populations was 0.38 (Table 3). Likewise, AMOVA

Table 4 Genetic diversity indices of the Ethiopian wild and cultivated sorghum populations, and ICRISAT wild sorghum accessions over all 9 loci (R_s allelic richness, R_p private allelic richness, H_e gene diversity, H_o average observed heterozygosity, F_{IS} inbreeding coefficient)

Population	R_s^a	R_p^a	H_e	H_o	F_{is}
Ethiopian wild *in situ* (N = 380; 20 per population)					
AW	2.95	0.12	0.43	0.11	0.80
G1	4.87	0.32	0.75	0.25	0.71
G2	4.98	0.39	0.77	0.31	0.59
G3	4.01	0.26	0.56	0.17	0.68
H1	4.29	0.28	0.60	0.43	0.27
H2	2.63	0.17	0.29	0.04	0.89
P1	3.40	0.39	0.54	0.10	0.77
P2	4.12	0.50	0.69	0.13	0.77
P3	3.20	0.29	0.59	0.23	0.69
P4	3.19	0.28	0.57	0.16	0.78
T1	1.87	0.09	0.17	0.01	0.93
T2	2.94	0.18	0.47	0.20	0.64
T3	3.85	0.22	0.64	0.18	0.73
T4	4.81	0.34	0.65	0.54	0.13
T5	4.28	0.22	0.69	0.17	0.80
W1	2.89	0.05	0.39	0.19	0.51
W2	3.57	0.04	0.67	0.14	0.78
W3	1.99	0.14	0.20	0.10	0.53
W4	2.35	0.02	0.34	0.14	0.58
Overall	14.93	3.76	0.83	0.19	0.66
Ethiopian cultivars *in situ* (N = 160; 20 per population)					
GC1	1.93	0.11	0.15	0.12	0.25
GC2	3.24	0.17	0.50	0.25	0.55
PC1	4.14	0.80	0.41	0.10	0.76
PC2	4.43	0.77	0.52	0.24	0.55
TC1	2.82	0.56	0.31	0.01	0.95
WC1	5.57	1.05	0.74	0.18	0.71
WC2	2.88	0.53	0.33	0.05	0.80
WC3	2.87	0.11	0.34	0.10	0.68
Overall	9.95	1.35	0.68	0.13	0.65
Wild ICRISAT accessions (N = 100; 10 per accession)					
South Africa *arundinaceum*	2.01	0.00	0.22	0.02	0.90
Swaziland *arundinaceum*	3.08	0.44	0.57	0.08	0.84
USA *virgatum*	2.92	0.55	0.55	0.32	0.46
Egypt *virgatum*	1.76	0.32	0.27	0.07	0.79
Angola *arundinaceum*	3.36	0.76	0.55	0.18	0.70
Sudan *verticilliflorum*	2.71	0.36	0.44	0.06	0.72
Sudan *aethiopicum*1	3.62	1.15	0.63	0.35	0.45
Sudan *aethiopicum*2	2.29	0.96	0.27	0.09	0.61
Sudan *drummondii*	1.64	0.44	0.18	0.13	0.46
Ethiopia *drummondii*	2.02	0.38	0.32	0.14	0.62
Overall	11.41	1.59	0.79	0.14	0.65

[a] For allelic richness and private allelic richness, overall values were computed with corrected sample sizes among the three sorghum groups (wild, cultivar and accessions; N = 96 for each) and for individual populations the minimum sample sizes were taken from each of the groups (N = 7 in wild, N = 15 in cultivars, and N = 7 in accessions)

showed that 41.2 % of the variation occurred among populations and 58.8 % found within populations (Table 5). An indirect estimate of gene flow using Wright's F_{ST} showed that the average number of migrants per generation across wild populations was $N_m = 0.43$, with higher gene flow in some adjacent wild populations. For instance, the largest number of migrants per generation (based on F_{ST}) was between G1 and G2 ($N_m = 3$). Similarly, $N_m = 2$ between P1 and P2, and $N_m = 1$ between P2 and P4. In general, gene flow was evident in 18 population pairs ($\sim$10.53 % of the total population combinations) (with $N_m = 1$ to $N_m = 3$) (Supplementary Table S2) computed from Weir and Cockerham (1984) pair wise F_{ST} (Supplementary Table S3).

The UPGMA cluster analyses (Fig. 2) and STRUCTURE (Fig. 3) show interesting and consistent patterns of genetic differentiation among the 19 wild populations (similar patterns were observed when the cultivars were omitted from these analyses). The Evanno et al. (2005) simulation method predicted K = 4 to be the optimum number of clusters for the STRUCTURE analysis. Four wild populations clustered together in the dendrogram and in each level of K in the STRUCTURE analysis (Figs. 2, 3). These were W4 and T1 in the north and AW and H2 in east-central Ethiopia. Interestingly, these populations were grouped with the Ethiopian *drummondii* accession in the dendrogram, and did not show affinity with the cultivars in either type of cluster analysis (Figs. 2, 3). Thus, we consider them to represent a distinct subgroup. The four populations in Pawe consistently clustered together, as did two in Ghibe (G1 and G2). Otherwise, there was little evidence of geographic structuring in the wild populations and the Mantel test of the correlation between genetic and geographic distance matrices was not significant ($r = -0.152$, $P = 0.07$). Altitude had significant negative

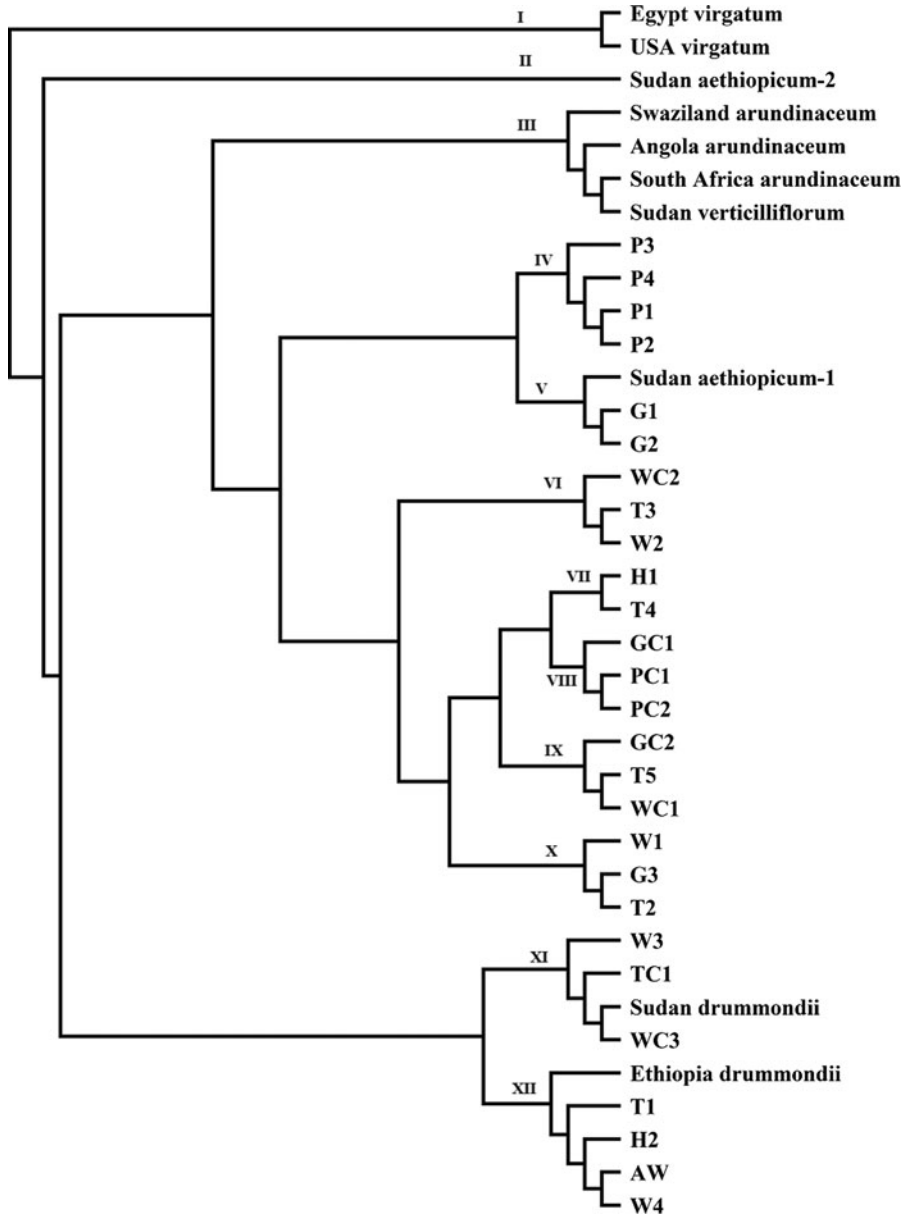

Fig. 2 Dendrogram showing relationship of Ethiopian wild and cultivated sorghum populations with ICRISAT accessions. See Tables 1 and 2 for collection site information and source of accessions (*G* Ghibe, *AW* Awash, *H* Hararghe, *P* Pawe, *W* Wello, and *T* Tigray; T1 and TC1 are located near the Wello sites)

Table 5 Analysis of molecular variance (AMOVA) of the 19 wild sorghum populations sampled in Ethiopia in 2008

Source of variation	*df*	Sum of squares	Variance components	Variation (%)	*P* value
Among populations	18	9099.1	1.20	41.2	<0.001
Within populations	7,961	13629.6	1.71	58.8	<0.001
Total	7,979	22728.7	2.91		

correlation with gene diversity ($r = -0.529$, $P = 0.02$), average observed heterozygosity ($r = -0.448$, $P = 0.054$) and with allelic richness ($r = -0.526$, $P = 0.021$) of wild sorghum populations.

Evidence for crop-wild gene flow

Many wild plants had morphological traits similar to crop-wild hybrids (Adugna 2012). To investigate

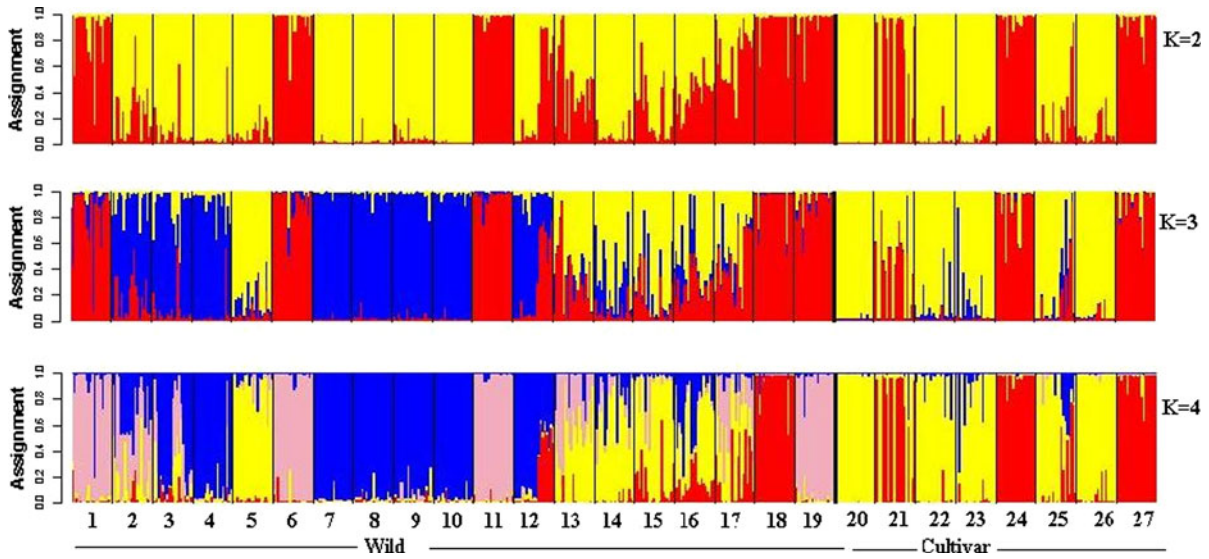

Fig. 3 STRUCTURE *bar* graph of 19 wild and 8 cultivated sorghum predefined populations (*x-axis*) with k = 2 to k = 4. Figures in the *y-axis* show coefficient of membership/ assignment. Each of the 540 plants is represented by a *single vertical line* of a *color* according to the estimated membership fraction averaged over all the 9 loci and each color group represents a cluster (K). Wild: *1* AW, *2* G1, *3* G2, *4* G3, *5* H1, *6* H2, *7* P1, *8* P2, *9* P3, *10* P4, *11* T1, *12* T2, *13* T3, *14* T4, *15* T5, *16* W1, *17* W2, *18* W3, *19* W4; Cultivated: *20* GC1, *21* GC2, *22* PC1, *23* PC2, *24* TC1 [Degalit cultivar], *25* WC1, *26* WC2 [improved cultivar 76T1#23], *27* WC3 [Degalit cultivar]. See Table 1 for collection site information (*G* Ghibe, *AW* Awash, *H* Hararghe, *P* Pawe, *W* Wello, and *T* Tigray; T1 and TC1 are located near the Wello sites)

admixture with wild populations, we sampled 8 locally grown cultivars in 2009 from areas where wild populations had been sampled in 2008 (Table 1). All of the cultivars had compact to semi-compact panicles typical of the durra and caudatum landraces, which are common throughout Ethiopia, and one, WC2, was an improved caudatum variety, $76T_1$#23. Two durra populations, WC3 and TC1, were collected in the same Kobo-Alamata plain located in adjacent districts ($\sim$ 35 km apart) and had the same local name, "Degalit".

The eight cultivars were contained in four of the 12 clusters in the UPGMA dendrogram (Fig. 2), with the two "Degalit" cultivars grouped together in Cluster XI. The caudatum and durra-caudatum landraces from Pawe and Ghibe clustered together, as did the durra landraces from Wello and neighboring Tigray (TC1).

Overall, our results suggest that wild populations are closely related to the cultivars (Fig. 2), and a Principal Component Analysis (data not shown). In Fig. 2, three of the 12 clusters (VI, IX, XI) joined Ethiopian wild and cultivated sorghum. The two "Degalit" cultivars clustered with the adjacent wild sorghum population, W3 (Figs. 2, 3). The improved variety (76T1#23, WC2) and an intermixed weedy phenotype (W2) were in one cluster (VI). Many populations show evidence of admixture at K = 4 (Fig. 3). Wild populations that were closest to the sampled cultivars were W2, T3, T5, and W3.

Discussion

Comparisons of genetic diversity among wild populations, accessions, and cultivars

The range of the observed microsatellite allele sizes were more or less similar to those reported in earlier studies using the same loci (Brown et al. 1996; Dean et al. 1999; Bisrat Ghebru et al. 2002; Agrama and Tuinstra 2003; Abu-Assar et al. 2005). Our study revealed significantly higher levels of diversity in wild population compared to cultivars and accessions, which were similar to each other, in terms of overall allelic richness, private allelic richness, and other parameters. Although we did not sample the cultivars or the available accessions very extensively, our findings have interesting implications for germplasm

conservation, as discussed below. Previous studies also reported greater genetic diversity in wild accessions (Aldrich et al.1992; Casa et al. 2005) or wild extant populations (Mutegi et al. 2011) than in sorghum cultivars, consistent with the expectation that cultivars have undergone genetic bottlenecks during domestication and further breeding. However, studies of *in situ* populations at a local scale in Cameroon (Barnaud et al. 2009), Mali (Sagnard et al. 2011), and Kenya (Mutegi et al. 2012) reported similar levels of genetic diversity between cultivated and wild sorghum, which they attributed to crop-to-wild gene flow. These authors also reported higher levels of diversity in a pool of wild accessions compared both to *in situ* wild and cultivated sorghum, which is not consistent with what we observed in this study. The significantly lower genetic diversity that we observed in ten ICRISAT accessions originating from at least six different African countries suggests that extant wild populations from Ethiopia are more diverse than these *ex situ* collections. This result also corresponds well with the suggestion that Ethiopia is within the center of origin for sorghum (e.g., Doggett 1988; Aldrich et al. 1992).

For the accessions in our study, consistent patterns of clustering according to the subspecies and races were not found (Fig. 2). Aldrich et al. (1992) also reported the failure of clustering according to races in accessions of wild sorghum, and many authors have noted the difficulty of assigning *in situ* wild or weedy sorghum plants to discrete taxonomic groups (e.g., Mutegi et al. 2012). This may indicate that the botanical classification into subspecies cannot be confirmed by genetic analysis. The "USA" accession in our study, which clustered with the *virgatum* accession from Egypt, may have originated in North Africa. Doggett and Majisu (1968) also used USA accessions of race *virgatum* which they reported to have originally collected from North Africa.

Patterns of genetic structure in wild sorghum populations

In addition to being highly diverse, the *in situ* wild populations showed interesting patterns of differentiation, with an average F_{ST} value of 0.37. In comparison, Ayana et al. (2001) reported that 25 % of the variation in RAPD markers was found among 10 *in situ* populations in central and western Ethiopia, and levels of genetic diversity within populations that were considered to be low. Although we sampled from some of the same areas a decade later than Ayana et al. (2001), our use of SSR markers *vs*. RAPDs, larger sample sizes per population (20 vs. 4–10), and more populations (19 vs. 11) may account for differences between the two studies.

We found that four widely separated populations (W4, T1, AW, and H2) were closely related to each other and to the ICRISAT-accession of subspecies *drummondii* from Ethiopia. Three of these populations were collected in tef fields, close to cultivated sorghum, while the AW population was found in a national park and was isolated from any cultivated fields. Subspecies *drummondii* is often described as a stabilized crop-wild hybrid lineage (e.g. Doggett 1988; Harlan and De Wet 1972), and our data suggest that these populations form a fairly distinct group. The four wild populations shared morphological characteristics such as short stature (91-230 cm), weak stalks, profuse tillering, dense racemes, the peduncle well exserted, narrow leaves (2-5 cm wide), the grain fully covered, and awned glumes (Adugna 2012). Another shared feature is that three of the original 12 SSR primers that we screened failed to amplify for most of these samples (see "Methods"). We also found that two of the three *in situ* Ghibe populations (G1 and G2) clustered with race *aethiopicum*-1 from Sudan and another population from Wello (W3) clustered with Sudan *drummondii* (Fig. 2).

Some neighboring wild populations clustered together, especially in Pawe, while others showed greater affinity with distant populations (Figs. 2, 3). The latter could be related to long-distance seed movement, perhaps as contaminants of crop seed. For example, farmers at site H1 (Bisidimo) told us that wild sorghum was a recent introduction to their area and its appearance coincided with food grain sorghum aid. Therefore, clustering of the H1 population in the east with Tigray populations (T4) near Sudan may occur because grain that was given as relief to farmers was probably purchased in Tigray, and the wild sorghum seeds might have moved with it. Moreover, the clustering of Ghibe cultivar (GC2) with South Wello cultivar (WC1) could be due to long distance movement of seed. During our interviews with farmers, they told us that they came from South Wello and North Shewa areas as settlers.

In Kenya, Mutegi et al. (2011) reported consistently strong spatial genetic structure in a study of wild

sorghum plants collected from throughout the country. These authors attributed correlations between genetic and geographic distances to patterns of seed movement within regions. Interestingly, sorghum landraces showed a similar pattern of spatial structure from these regions (Mutegi et al., 2011). In our study, there was no significant pattern of gene flow by isolation-by-distance presumably because of other barriers including the mountains separating some of the collections sites.

Evidence for crop-wild gene flow

The relatively high F_{ST} value (0.37, $P < 0.001$) observed in wild sorghum in this study shows high differentiation among various populations, while the divergence between cultivated and wild sorghum from the same areas was moderate ($F_{ST} = 0.105$), a possible indication of crop-wild gene flow and introgression. Similar to Mutegi et al. (2011), STRUCTURE analyses at $K = 2$ (Fig. 3) did not differentiate between wild and cultivated plants, nor did Principal Component Analyses. Moreover, a neighbor joining cluster dendrogram based on pair-wise simple matching distance among the 540 individuals of cultivated and wild populations showed some level of admixture (Supplementary Fig. S1). These patterns could reflect extensive hybridization and/or common ancestry of the cultivars from their wild progenitors in Ethiopia. In Kenya, Mutegi et al. (2011) found evidence that repeated gene flow and/or common ancestry have largely blurred distinctions between wild and cultivated sorghum when compared among regions.

Where wild populations did not cluster with cultivars, possible reasons include reduced levels of crop-wild gene flow, including effects of hand-weeding to remove hybrids, or the fact that we did not sample a very large number of cultivar populations. As noted above, four Ethiopian wild populations that clustered with an accession of subsp. *drummondii* were very distinct from the landraces in this study. Likewise, four populations from Pawe and one from Ghibe were also differentiated from the crop. Overall, differences between our report of some wild populations that are differentiated from cultivars and those of Mutegi et al. (2011) could be attributed to sampling designs, differences in SSR markers, and variation in farmers' practices of weeding out the wild or hybrid sorghum before it is able to set seed.

In agreement with our results, Casa et al. (2005) surveyed many African accessions and reported that the degree of differentiation between wild *vs*. cultivated was moderate ($F_{ST} = 0.13$). Similarly, Cui et al. (1995) reported moderate differentiation among wild and cultivated sorghum. However, these findings are not inconsistent with the expectation that a limited amount of gene flow has occurred, as many previous investigators have speculated (e.g., Harlan and De Wet 1972; Doggett 1988; Aldrich et al. 1992). Analyses of progeny segregation, allozymes, and RFLPs revealed crop-specific alleles in wild *S. bicolor* when it co-occurred with the crop in Africa, suggesting that intraspecific hybridization and introgression are common (Aldrich and Doebley 1992).

A broad survey involving more cultivar samples and crop-specific molecular markers might provide stronger evidence for crop-wild hybridization in Ethiopia, but data from the current study do not suggest that such gene flow has been extensive to blur genetic diversity in the wild. This is rather surprising because cultivated and wild sorghum frequently co-occur and have overlapping flowering times (e.g., Tesso et al. 2008; Mutegi et al. 2010). In addition, F_1 crop-wild sorghum hybrids do not show obvious fitness penalties (Muraya et al. 2011) and may even exhibit heterosis (Adugna 2012). As noted earlier, outcrossing rates of wild sorghum ranged from 0.33 to 0.65 in a survey of six populations in Ethiopia (Adugna 2012) and from 0.09 to 0.70 in 12 populations in Kenya (Muraya et al. 2011). Populations with relatively high outcrossing rates should be more likely to hybridize with the crop, depending on the likelihood of pollen flow and the likelihood of being weeded out by farmers. For instance, Teshome et al. (1997) suggested that the morphologically intermediate forms in north Shewa and south Wello appear as a result of farmers' tolerance to leave the wild relatives in and around their fields to effect gene flow through outcrossing.

Implications for germplasm conservation and use

We found that considerable genetic diversity remains in extant wild populations, some of which (the isolated types) were thought to be threatened by human overpopulation and expansion of modern agriculture (Ayana et al. 2001). The existence of such diversity in the wild and weedy populations could be due to local adaptation, geographic isolation, and perhaps

co-evolution with local cultivars due to farming practices. The weedy sorghums are intimately associated with the crop whereby they resemble the crop before flowering and, as a result of misidentification, escape eradication (i.e., mimicry, as in Barrett 1983). In addition, most of them shatter their seeds before harvest of the crop. Our results also show that some crop-to-wild gene flow is likely. This implies that transgenes could also spread to wild populations.

The significantly higher level of private allelic richness that we observed in wild sorghum populations relative to cultivars may be beneficial to sorghum breeding as it may be linked to unique traits. Further conservation effort may be necessary to safeguard this unique diversity. As with many crop species, rigorous comparisons between *in situ* and *ex situ* populations are generally needed to fully assess genetic resources for germplasm preservation (Heywooda and Iriondob 2003; Maxted 2003). Sampling genetic diversity and storing or propagating the collected material outside the natural environments using *ex situ* conservation methods is useful to safeguard endangered species (Heywooda and Iriondob 2003). For instance, after reintroduction of 20 cultivated sorghum accessions managed in gene bank for an average of 17.5 years to their original places in Wello, Ethiopia, Seboka and van Hintum (2006) found that 5 of them (those collected 22 years ago) were lost or not remembered by the farmers. *Ex situ* conservation saved them from extinction. Moreover, Rice (2004) found that the diversity of older Jala maize gene bank populations (*ex situ* conserved) was less and more differentiated from recent Jala (*in situ* conserved) due to gene bank regeneration methods. However, *ex situ* conservation has limitations such as poor genetic or demographic management almost inevitably resulting in genetic erosion and spontaneous hybridization (Volis et al. 2009) in addition to the costly operation (Maxted 2003).

On the other hand, *in situ* conservation allows evolutionary change to continue in the component species and populations (Heywooda and Iriondob 2003) and provides better representation of genetic diversity (Volis et al. 2009). In our study, the reduced gene diversity and allelic richness in the ICRISAT accessions compared to in situ samples was perhaps because they have been under stringent selection pressure during maintenance. Sampling accessions from seed banks that have been increased to maintain the germplasm was suggested to be one reason for low variation in sorghum accessions by Morden et al. (1989). Maintaining the high outcrossing rate and other life history traits may help sustain the high diversity.

Our project was aimed at studying the relationships between cultivated and wild sorghum in major sorghum growing regions in Ethiopia. However, wild sorghums also occur along irrigation ditches and in areas where sorghum is not grown. Therefore, future studies should focus on the complete geographical distribution of the various sub-species and races of wild sorghums in Ethiopia and elsewhere. Also, more extensive surveys could be carried out to better understand the extent and consequences of ongoing crop-to-wild gene flow in Ethiopia.

Acknowledgments This study was supported by a grant from the Biotechnology-Biodiversity Interface (BBI) program of the United States Agency for International Development (USAID) to A. A. Snow and serves as part of the dissertation research for the partial fulfillment of the requirements for a PhD degree to the first author. We also thank G. Ejeta, C. Grenier, J. Pederson, T. Tesso, and Y. Teklu for helpful advice and assistance with this research. We extend our sincere thanks to ICRISAT for providing the seeds of wild sorghum accessions.

References

Abu-Assar AH, Uptmoor R, Abdelmula AA, Salih M, Ordon F, Friedt W (2005) Genetic variation in sorghum germplasm from Sudan, ICRISAT, and USA assessed by simple sequence repeats (SSRs). Crop Sci 45:1636–1644

Adugna A (2012) Population genetics and ecological studies in wild sorghum [*Sorghum bicolor* (L.)] in Ethiopia: implications for germplasm conservation. PhD Thesis, Addis Ababa University

Adugna A, Snow AA, Sweeney P (2011) Optimization of a high throughput, cost effective and all-stage DNA extraction protocol for sorghum. J Agric Sci Tech 5:243–250

Agrama HA, Tuinstra MR (2003) Phylogenetic diversity and relationships among sorghum accessions using SSRs and RAPDs. Afr J Biotech 2:334–340

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

Ayana A, Bekele E, Bryngelsson T (2001) Genetic variation in wild sorghum (*Sorghum bicolor* subsp. *verticilliflorum* (L.) Moench) germplasm from Ethiopia assessed by random amplified polymorphic DNA (RAPD). Hereditas 132:249–254

Barnaud A, Trigueros G, McKey D, Joly H (2008) High outcrossing rates in fields with mixed sorghum landraces: how are landraces maintained? Heredity 101:445–452

Barnaud A, Deu M, Garine E, Chantereau J, Bolteu J, Koıda EO, McKey D, Joly HI (2009) A weed–crop complex in sorghum: the dynamics of genetic diversity in a traditional farming system. Am J Bot 96:1869–1879

Barrett SCH (1983) Crop mimicry in weeds. Econ Bot 37:255–282

Bhattramakki D, Dong J, Chhabra AK, Hart GE (2000) An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. Genome 43:988–1002

Bohonak AJ (2002) IBD (Isolation By Distance): a program for analyses of isolation by distance. J Heredity 93:153–154

Borrell AK, Hammer GL, Douglas ACL (2000) Does maintaining green leaf area in sorghum improve yield under drought? I. Leaf growth and senescence. Crop Sci 40:11026–11103

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Human Genet 32:314–331

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzalez-Candelas F, Kresovich S (1996) Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. Theor Appl Genet 93:190–198

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. Theor Appl Genet 111:23–30

Central Statistical Agency (CSA) (2010) Agricultural sample survey, report on area and production of crops (private peasant holdings, Meher season), vol IV. Addis Ababa, Ethiopia

Chakraborty R, Jin L (1993) Determination of relatedness between individuals by DNA fingerprinting. Human Biol 65:875–895

Cui YX, Xu GX, Magill CW, Schertz KF, Hart GE (1995) RFLP-based assay of *Sorghum bicolor* (L.) Moench genetic diversity. Theor Appl Genet 90:787–796

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 25:128–131

Dean RE, Dahlberg JA, Hopkins MS, Mitchell SE, Kresovich S (1999) Genetic redundancy and diversity among 'orange' accessions in the U.S. national sorghum collection as assessed with simple sequence repeat (SSR) markers. Crop Sci 39:1215–1222

Djé Y, Heuertz M, Lefebvre C, Vekemans X (2000) Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers. Theor Appl Genet 100:918–925

Djé Y, Heuertz M, Ater M, Lefebyre C, Vekemans X (2004) In situ estimation of outcrossing rate in sorghum landraces using microsatellite markers. Euphytica 138:205–212

Doggett H (1988) Sorghum, 2nd edn. Longman, London

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

Earl and von Holdt (2011) STRUCTURE HARVESTER v0.6.8. http://taylor0.biology.ucla.edu/struct_harvest). Accessed 2 Nov 2011

Ejeta G, Grenier C (2005) Sorghum and its weedy hybrids. In: Gressel J (ed) Crop ferality and volunteerism. Taylor and Francis, Florida, pp 123–135

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. Ann Rev Ecol Syst 24:217–242

Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. Ann Rev Ecol Evol Syst 30:539–563

Escalante A, Coello MG, Eguiarte LE, Piñero D (1994) Genetic structure and mating systems in wild and cultivated populations of *Phaseolus coccineus* and *P. vulgaris* (Fabaceae). Am J Bot 81:1096–1103

Evanno S, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinfo Online 1:47–50

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Ghebru B, Schmidt RJ, Bennetzen JL (2002) Genetic diversity of Eritrean sorghum landraces assessed with simple sequence repeat (SSR) markers. Theor Appl Genet 105: 229–236

Goudet J (2002) Fstat, version 2.932. In: Institute of Ecology, Biology Building, UNIL, Switzerland, Lausanne

Harlan JR, de Wet JMJ (1972) A simplified classification of cultivated plants. Taxon 20:509–517

Heywooda VH, Iriondob JM (2003) Plant conservation: old problems, new perspectives. Biol Conserv 113:321–335

Hulbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. Ecology 52:577–586

Idury RM, Cardon LR (1997) A simple method for automated allele binning in microsatellite markers. Genome Res 11:1104–1109

Jarvis DI, Hodgkin T (1999) Wild relatives and crop cultivars: detecting natural introgression and farmer selection of new genetic combinations in agroecosystems. Mol Ecol 8:5159–5173

Kalinowski S (2005) HP-RARE 10: a computer program for performing rarefaction on measures of allelic richness. Mol Ecol 5:187–189

Ladizinsky G (1985) Founder effect in crop-plant evolution. Econ Bot 39:191–199

Langridge P, Chalmers K (2004) The principle: identification and application of molecular markers. In: Lörz H, Wenzel G (eds) Molecular marker systems in plant breeding and crop improvement. Biotechnology in Agriculture and Forestry, vol 55. Springer, Berlin, pp 3–22

Li M, Yuyama N, Luo L, Hirata M, Cai H (2009) In silico mapping of 1758 new SSR markers developed from public genomic sequences for sorghum. Mol Breed 24:41–47

Liang H, Gao ZS (2001) Phylogenetic analysis and transformation of sorghum [*Sorghum bicolor* (L.) Moench]. Recent Res Dev Plant Biol 1:17–33

Liu J (2005) PowerMarker V3.125. http://www.powermarker.net. Accessed 21 Aug 2009

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Maxted N (2003) Conserving the genetic resources of crop wild relatives in European Protected Areas. Biol Conserv 113:411–417

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

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

Muraya MM, Mutegi E, Geiger HH, de Villiers SM, Sagnard F, Kanyenji BM, Kiambi D, Parzies HK (2011) Wild sorghum from different eco-geographic regions of Kenya display a mixed mating system. Theor Appl Genet 122:1631–1639. doi:10.1007/s00122-011-1560-5

Mutegi E, Sagnard F, Muraya M, Kanyenji B, Rono B, Mwongera C, Marangu C, Kamau J, Parzies H, de Villiers S, Semagn K, Traore' PS, Labuschagne M (2010) Eco-geographical distribution of wild weedy and cultivated *Sorghum bicolor* (L.) Moench in Kenya: implications for conservation and crop-to-wild gene flow. Genet Resour Crop Evol 57:243–253

Mutegi E, Sagnard F, Semagn K, Deu M, Muraya M, Kanyenji B, de Villiers S, Kiambi D, Herselman L, Labuschagne M (2011) Genetic structure and relationships within and between cultivated and wild sorghum (*Sorghum bicolor* (L.) Moench) in Kenya as revealed by microsatellite markers. Theor Appl Genet 122:989–1004

Mutegi E, Sagnard F, Labuschagne M, Herselman L, Semagn K, Deu M, de Villiers S, Kanyenji BM, Mwongera CN, Traore PCS, Kiambi D (2012) Local scale patterns of gene flow and genetic diversity in a crop–wild–weedy complex of sorghum (*Sorghum bicolor* (L.) Moench) under traditional agricultural field conditions in Kenya. Conserv Genet. doi:10.1007/s10592-012-0353-y

Page RDM (2001) TreeView 1.6.6. http://taxonomy.zoology.gla.ac.uk/rod/rod.html

Pedersen JF, Toy JJ, Johnson B (1998) Natural outcrossing of sorghum and sudangrass in the central great plains. Crop Sci 38:937–939

Perrier X, Jacquemoud-Collet JP (2006) DARwin software. http://darwin.cirad.fr/darwin

Prasanth V, Chandra S, Jayashree B, Hoisington D (2006) AlleloBin-a program for allele binning of microsatellite markers based on the algorithm of Idury and Cardon (1997). International Crops Research Institute for the Semi, Arid Tropics (ICRISAT), Patancheru, AP, India

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN:3-900051-07-0, URL http://www.R-project.org

Reed JD, Ramundo BA, Claflin LE, Tuinstra MR (2002) Analysis of resistance to ergot in sorghum and potential alternate hosts. Crop Sci 42:1135–1138

Rice EB (2004) Conservation and change: a comparison of in-situ and ex-situ conservation of Jala maize germplasm in Mexico. PhD Dissertation, Cornell University

Rousset F, Raymond M (1997) Statistical analyses of population genetic data: new tools, old concepts. Trends Ecol Evol 12:313–317

Sagnard F, Deu M, Dembe'le' D, Leblois R, Toure'L, Diakite' M, Calatayud C, Vaksmann M, Bouchet S, Malle' Y, Togola S, Traore PCS (2011) Genetic diversity, structure, gene flow and evolutionary relationships within the *Sorghum bicolor* wild–weedy–crop complex in a western African region. Theor Appl Genet 123:1231–1246

Schmidt MR, Bothma GC (2006) Risk sssessment for transgenic sorghum in Africa: crop-to-crop gene flow in *Sorghum bicolor* (L.) Moench. Crop Sci 46:790–798

Seboka B, van Hintum T (2006) The dynamics of on-farm management of sorghum in Ethiopia: implication for the conservation and improvement of plant genetic resources. Genet Resour Crop Evol 53:1385–1403

Singh R, Axtell JD (1973) High lysine mutant gene (hl) that improves protein quality and biological value of grain sorghum. Crop Sci 13:535

Small E (1984) Hybridization in the domesticated-weed-wild complex. In: Grant WF (ed) Plant biosystematics. Academic Press, Toronto, pp 195–210

Snow AA, Andow DA, Gepts P, Hallerman EM, Power A, Tiedje JM, Wolfenbarger LL (2005) Genetically engineered organisms and the environment: current status and recommendations. Ecol Appl 15:377–404

Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. Bioinformatics. doi:10.1093/bioinformatics/btn478

Tanksley SD, McCouch R (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063–1066

Taramino G, Tarchini R, Ferrario S, Lee M, Pe ME (1997) Characterization and mapping of simple sequence repeats (SSRs) in Sorghum bicolor. Theor Appl Genet 95:66–72

Templeton AR (2006) Population genetics and microevolutionary theory. Wiley, New Jersey

Teshome A, Baum BR, Fahrig L, Torrance JK, Arnason TJ, Lambert JD (1997) Sorghum [*Sorghum bicolor* (L.) Moench] landrace variation and classification in North Shewa and SouthWelo, Ethiopia. Euphytica 97:255–263

Tesso T, Kapran I, Grenier C, Snow AA, Sweeney P, Pedersen J, Marx D, Bothma G, Ejeta G (2008) The potential for crop-to-wild gene flow in sorghum in Ethiopia and Niger: a geographic survey. Crop Sci 48:1425–1431

Weir S, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370

Wright S (1951) The genetical structure of populations. Ann Eugen 15:323–354

Wu YQ, Huang Y, Tauer CG, Porter DR (2006) Genetic diversity of sorghum accessions resistant to greenbugs as assessed with AFLP markers. Genome 49:143–149

Volis S, Blecher M, Sapir Y (2009) Complex ex situ—in situ approach for conservation of endangered plant species and its application to Iris atrofusca of the Northern Negev. In: Krupp F, Musselman LJ, Kotb MMA, Weidig I (Eds)

Environment, biodiversity and conservation in the middle east, vol 3. BioRisk, pp 137–160

Zhao Z (2007) The Africa biofortified sorghum project –applying biotechnology to develop nutritionally improved sorghum for Africa. In: Xu Z et al (eds) Biotechnology and sustainable agriculture 2006 and beyond. Springer, Berlin, pp 273–277

Zhao ZY, Cai T, Tagliani L, Miller M, Wang N, Pang H, Rudert M, Schroeder S, Hondred D, Seltzer J, Pierce D (2000) Agrobacterium mediated sorghum transformation. Plant Mol Biol 44:789–798

Zhao ZY, Glassman K, Sewalt V, Wang N, Miller M, Chang S, Thompson T, Catron S, Wu E, Bidney D, Kedebe Y, Jung R (2003) Nutritionally improved transgenic sorghum. In: Vasil IK (ed) Plant biotechnology 2002 and beyond. Kluwer Academic Publishers, Florida, pp 413–416

Zhu H, Muthukrishnan S, Krishnaveni S, Wilde G, Jeoung JM, Liang GH (1998) Biolistic transformation of sorghum using a rice chitinase gene. J Genet Breed 52:243–252