Theor Appl Genet (2011) 123:571–583
DOI 10.1007/s00122-011-1608-6

ORIGINAL PAPER

Genetic structure and diversity of wild sorghum populations (*Sorghum* spp.) from different eco-geographical regions of Kenya

Moses M. Muraya · Santie de Villiers · Heiko K. Parzies · Evans Mutegi · Fabrice Sagnard · Ben M. Kanyenji · Dan Kiambi · Hartwig H. Geiger

Received: 10 January 2011 / Accepted: 26 April 2011 / Published online: 4 June 2011
© Springer-Verlag 2011

Abstract Wild sorghums are extremely diverse phenotypically, genetically and geographically. However, there is an apparent lack of knowledge on the genetic structure and diversity of wild sorghum populations within and between various eco-geographical regions. This is a major obstacle to both their effective conservation and potential use in breeding programs. The objective of this study was to assess the genetic diversity and structure of wild sorghum populations across a range of eco-geographical conditions in Kenya. Sixty-two wild sorghum populations collected from the 4 main sorghum growing regions in Kenya were genotyped using 18 simple sequence repeat markers. The study showed that wild sorghum is highly variable with the Coast region displaying the highest diversity. Analysis of molecular variance showed a significant variance component within and among wild sorghum populations within regions. The genetic structure of wild sorghum populations indicated that gene flow is not restricted to populations within the same geographic region. A weak regional differentiation was found among populations, reflecting human intervention in shaping wild sorghum genetic structure through seed-mediated gene flow. The sympatric occurrence of wild and cultivated sorghums coupled with extensive seed-mediated gene flow, suggests a potential crop-to-wild gene flow and vice versa across the regions. Wild sorghum displayed a mixed mating system. The wide range of estimated outcrossing rates indicate that some environmental conditions may exist where self-fertilisation is favoured while others cross-pollination is more advantageous.

Communicated by X. Xia.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-011-1608-6) contains supplementary material, which is available to authorized users.

M. M. Muraya (✉) · H. K. Parzies · H. H. Geiger
Institute of Plant Breeding, Seed science and Population Genetics, University of Hohenheim, Fruwirthstrasse 21, 70599 Stuttgart, Germany
e-mail: mahugu2002@yahoo.com

M. M. Muraya
Leibniz Institute of Plant Genetics and Crop Plant Research, Corrensstrasse 3, 06466 Gatersleben, Germany

F. Sagnard
CIRAD, UMR Développement et Amélioration des Plantes, c/o ILRI, PO Box 30709, Nairobi, Kenya

S. de Villiers · E. Mutegi · F. Sagnard · D. Kiambi
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT-Nairobi), P.O Box 39063-00623, Nairobi, Kenya

E. Mutegi
Kenya Agricultural Research Institute (KARI), P.O. Box 30148, Nairobi, Kenya

B. M. Kanyenji
Kenya Agricultural Research Institute (KARI-Embu), P.O. Box 27-60100, Embu, Kenya

Introduction

Sorghum bicolor (L.) Moench has its centre of diversity in Africa. Species *S. bicolor* comprises the cultivated sorghum and its closely related wild relatives with a natural range throughout Africa (de Wet 1978; Duvall and Doebley 1990). The three recognised subspecies of *S. bicolor* (L.) Moench in Africa are: *S. bicolor* ssp. *bicolor* (L.) Moench (cultivated sorghum), *S. bicolor* ssp. *verticilliflorum* (Steud.) Piper (wild sorghum) and *S. bicolor* ssp. *drummondii* (steud.) de Wet (weedy sorghum). *Sorghum bicolor* ssp. *verticilliflorum* is the most widely distributed

wild sorghum in Africa. Four races are recognised within *S. bicolor* ssp. *verticilliflorum*: races *aethiopicum*, *vigartum*, *arundinaceum* and *verticilliflorum*, while two races are recognised in *S. bicolor* ssp. *drummondii*: races *drummondii* and *hewlsonni*. Several other wild sorghum species, e.g., *S. almum* Parodi, *S. purpureoserium* (Hochst. Ex A. Rich) Asch. & Schweinf, *S. halepense* (L.) Pers, and *S. vertisicolor* Andersson, are also recognised in Africa (Price et al. 2005), however, their morphological and ecological boundaries are not well defined. In Africa, cultivated and wild sorghums have overlapping geographic distributions and are interfertile (Doggett 1988), making it easier to utilise wild sorghum as a genetic resource in cultivated sorghum breeding programs.

Wild species of sorghum may constitute an important gene reservoir for crop sorghum improvement. For instance, *Striga*-resistance mechanisms such as low germination stimulant production, germination inhibition, and low haustorial initiation activity have been found in wild sorghum (Rich et al. 2004). Furthermore, wild sorghum has novel grain starch properties that could be used to improve digestibility of crop-sorghum for intensive livestock industries (Dillon et al. 2007). At the same time, however, crop-wild sorghum hybridisation can also introduce genes into wild populations as such introgression has played a role in structuring the genetic diversity of species (Rieseberg 1997), in the origin of new adaptations (Rieseberg 1991), in the transfer of new adaptations between species (Rieseberg et al. 2003), in formation of new ecotypes or species (Rieseberg 1997; Soltis and Soltis 1999) and in the evolution of invasiveness (Anttila et al. 1998; Ellstrand and Schierenbeck 2000). Despite these possibilities, the risk of "genetic pollution" due to crop-wild hybridisation is minimal because farmers select their own seeds at the end of the cropping season, choosing the best plants and panicles; therefore such selection is a basic part of informal seed systems that characterise subsistence agriculture (Muteg et al. 2010).

Studying genetic variation of wild sorghum populations in Kenya attracts special interest for several reasons. Kenya lies within the broad geographic range where sorghum is believed to have been domesticated and where the greatest genetic variation for both wild and cultivated sorghum is found (Mann et al. 1983; Doggett 1988; Muteg et al. 2010; Muraya et al. 2010). The genetic diversity existing in wild sorghum centres of origin represents one of the world's most important natural resources for future sorghum breeding efforts and global food security. Conserving wild diversity with and through the evolutionary processes that generate variability leaves them open to change via gene flow and introgression.

Gene flow from cultivated crops to wild relatives is a risk, among others, also from genetically modified plants. The main concern is that widespread cultivation of some transgenic cultivars could accelerate the evolution of undesirable and more invasive weeds, thereby leading to biodiversity erosion or to ecosystem disequilibrium effects (Conner et al. 2003; Stewart et al. 2003; Pilson and Prendeville 2004; Hails and Morley 2005). With regard to genetically modified (GM) crops, crop-to-wild gene flow and subsequent introgression of crop genes into wild populations raises concerns among conservationists. With appropriate DNA markers, it is possible to detect gene flow from transgenic crops, but it is difficult to predict the ecological effects of transgenes that are integrated into the different genetic backgrounds or expressed in different ecological contexts. Plants that acquire transgenes will continue to evolve, subject to natural and artificial selection pressures in the agricultural setting and beyond. A biofortified GM sorghum cultivar is being developed in Africa (http://biosorghum.org/project) and its commercial production may start in the foreseeable future. However, the risk of adverse genetic pollution from cultivated crops to wild relatives must primarily focus on the need to monitor the evolution of wild species in time and space and how the human intervention can modify their natural equilibrium. Therefore, the need to provide baseline data on the genetic diversity and structure of wild sorghum populations is both urgent and justified (Hokanson et al. 2010).

Genetic diversity and population structure in sorghum have been estimated using various types of molecular markers (Taramino et al. 1997; Djè et al. 2000; Uptmoor et al. 2003; Anas 2004; Menz et al. 2004; Folkertsma et al. 2005; Barnaud et al. 2007). In general, these studies showed greater genetic variability and significant differentiation among cultivated sorghum. Unfortunately, most of the documented studies only dealt with diversity in cultivated sorghum. Considering that gene flow is a key evolutionary factor affecting the structure of the genetic diversity of wild sorghum populations, this study assessed the genetic diversity and structure of wild sorghum populations across a range of eco-geographical conditions in Kenya. This was achieved by analysis of wild sorghum populations sampled in the four main sorghum growing regions with a set of 18 simple sequence repeat (SSR syn. microsatellite) markers.

Materials and methods

Sample collection

A field survey was conducted between June and August 2006 in the four main sorghum growing regions in Kenya (Fig. 1) along an east–west transect, stopping at every 50 m fall in altitude. The latitude, longitude and elevation of all sampling sites were recorded using a Global Positioning System (GPS). A questionnaire and direct

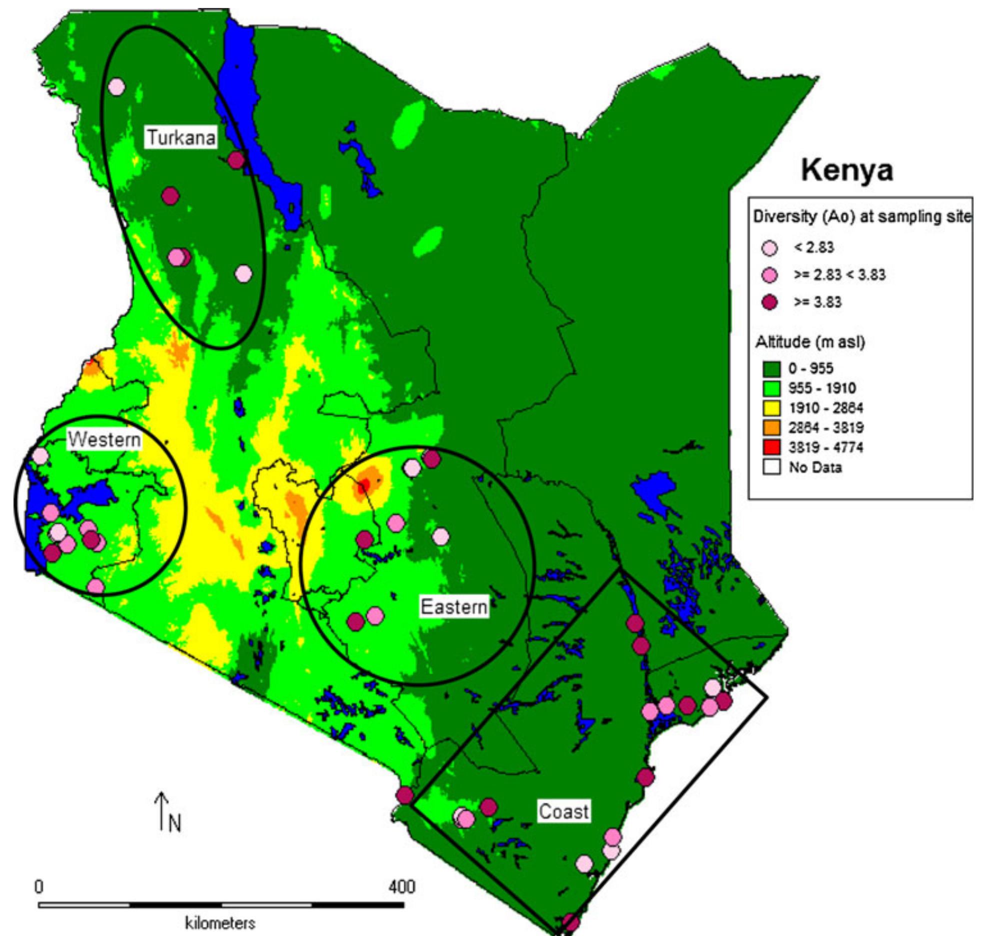

Fig. 1 Major growing areas (*encircled*) of sorghum in Kenya; allelic diversity of wild sorghum populations at respective sampling sites are indicated as *coloured dots* (see legend) (colour figure online)

observations were used to record additional data on each wild sorghum population. Panicle compactness and shape, inflorescence exertion, awn at maturity, glume colour, glume hairiness, glume hair colour, grain cover, grain colour, grain plumpness and shattering were recorded using published sorghum descriptors (IBPGR/ICRISAT 1993; Table S1). This data were used to compute phenotypic distance matrices. At each sampling site, 30 panicles of wild sorghum were collected from individual plants, at least 1 m apart. In total, 107 populations were sampled consisting of 28, 22, 33 and 24 populations from Turkana, Western, Coast and Eastern regions, respectively. However, they could not be classified clearly into races and/or species because wild sorghum has previously been classified according to phenotypic traits (de Wet and Harlan 1971; Harlan and de Wet 1972; Doggett 1988), which varies continuously and are not reliable in taxonomy, resulting in a loose classification. We also collected some 'weedy sorghums' (see Table S2), which could possibly have resulted from introgression of crop genes with selective advantages into a wild population (*S. bicolor* ssp. *drummondii*) or from stabilised populations resulting from different wild species crosses.

Selection of populations for molecular and phenotypic diversity studies

From the 107 wild sorghum populations collected, 62 were chosen for this study (17, 12, 19 and 14 populations from Turkana, Western, Coast and Eastern, respectively). Selection was based on population size at each sampling site and the proportionate coverage of the four sampling regions (Table S2). Populations with estimated $N > 500$ individual plants at the sampling site were selected. One offspring from each of 24 plants (panicles) per population was genotyped, giving a total of 1,488 individuals genotyped across all populations.

Molecular genetic studies

DNA extraction

For DNA extraction seedlings were raised in a greenhouse at the University of Hohenheim in April 2007. In the third leaf stage, leaf tissue was harvested and lyophilised. Total genomic DNA was extracted using a modified CTAB

protocol (Mace et al. 2003). Concentration and quality of the extracted DNA was assessed with electrophoresis of 1 μl of DNA on a 0.7% agarose gel followed by normalisation of the concentration to 50 ng/μl.

Polymerase chain reaction and fragment analyses

Eighteen sorghum SSR primers of known map location and distributed throughout the ten linkage groups were used for genotyping (Table S3). Forward primers were labelled with FAM, HEX or TET, allowing multiplexing of primer products into six groups of three each. PCR was performed in 20 μl containing 1× PCR buffer (20 mM Tris–HCl (pH 8.4), 50 mM KCl), 1.5 mM $MgCl_2$, 0.25 μM of each fluorescent labelled forward and unlabelled reverse primers, 0.2 mM dNTPs, 0.5 U of Taq polymerase and 100 ng template DNA.

The amplification reaction consisted of a denaturing step for 3 min at 94°C, followed by 40 cycles of 94°C for 1 min, annealing at 55°C or 62°C (depending on SSR primers) for 1 min, extension at 72°C for 1 min followed by terminal extension at 72°C for 10 min using a MJResearch icl PTC-100 thermocycler. Fragments were analysed on an automated sequencer (MegaBACE). Reference genotype BTX 623 was included as control on each plate to confirm the reproducibility of allele sizes.

Diversity analysis

The level of polymorphism at each locus was calculated according to Botstein et al. (1980) and polymorphic information content (PIC) determined as:

$$\mathrm{PIC} = 1 - \sum_{i=1}^{k} p_i^2 - \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} 2p_i^2 p_j^2$$

where p_i and p_j are the frequencies of alleles i and j, respectively.

PowerMarker v. 3.25 (Liu and Muse 2005) was used to calculate the PIC and inbreeding coefficient for each marker as well as the average across markers. The total number of alleles, rare alleles, allelic richness and private allelic richness were calculated for each locus, population and region using HP-RARE 1.2 (Kalinowski 2005). Nei's (1987) unbiased gene diversity or expected heterozygosity, and observed heterozygosity were estimated for each locus, population and region using the software GENETIX 4.05 (Belkhir et al. 2004). The significance of allelic richness, private allelic richness and gene diversity among populations and regions was explored using the Kruskal–Wallis test in R (R Development Core Team 2008). To explore differentiation among populations and regions, F statistics parameters (F_{IS}, F_{IT} and F_{ST}) were estimated in GENETIX 4.05 (Belkhir et al. 2004). F_{IS}, the inbreeding coefficient, describes the divergence of observed heterozygosity from expected heterozygosity within the population assuming panmixia. F_{IT}, the overall inbreeding coefficient, describes the reduction of heterozygosity within individuals relative to the total population due to non-random mating within subpopulation (F_{IS}) and population subdivision (F_{ST}). F_{ST}, the fixation index, describes the reduction in heterozygosity within population relative to total population due to selection or drift (Lowe et al. 2004).

Estimation of equilibrium outcrossing rates

Twenty-four individuals per population were used to compute estimates of equilibrium outcrossing rate (t_e) as follows:

$$t_e = \frac{1 - F}{1 + F}$$

where F is the inbreeding coefficient which was computed according to Weir (1996). Correlation between t_e and panicle compactness and shape were also computed. In our case the "inbreeding" coefficient is an estimate which relates to the relative equilibrium heterozygosity of a single local population and the mean across several such populations within a region (Table 2) corresponds to the "fixation" index F_{IS} in Table 3.

Analysis of molecular variance (AMOVA)

The molecular genetic variation between and within populations was analysed by AMOVA using the Arlequin v 3.1 (Excoffier et al. 2006) with 10,000 permutations to obtain reliable variance components. Two types of AMOVA were conducted based on (a) grouping populations according to the regions where they were collected, and (b) the population clusters obtained by STRUCTURE analysis.

Population structure

The software STRUCTURE (Pritchard et al. 2000, 2007) was used to analyse the population structure based on the admixture model where each individual draws some fraction of its genome from each of the K populations. This method is useful to identify gene flow events since individuals whose genotypes indicate admixture are assigned jointly to two or more populations. The correlated allele frequencies model (Falush et al. 2003) which often improves clustering for closely related populations was used. Twenty-five runs of STRUCTURE were carried out for each set of K sub-populations, with K values from 2 to 10. The choice of the number of K (2–10) was based on

Table 1 Diversity statistics of the SSR markers combined across 62 wild sorghum populations

Marker	Motif	A_t	A_r	PIC	ASOP	ASOS
Genomic markers						
Xtxp057	(GT)21	34	26	0.91	223–257	191–283
Xtxp136	(GCA)5	17	14	0.42	240–243	217–273
Xtxp145	(AG)22	24	18	0.77	208–244	196–254
Xtxp273	(TTG)20	36	26	0.92	169–199	171–265
mSbCIR262	(CATG)3.25	18	10	0.85	208–220	200–234
sb6-84	(AG)14	32	23	0.91	183–217	145–225
Xtxp012	(CT)22	31	23	0.91	161–205	163–229
Xtxp021	(AG)18	31	23	0.89	169–199	147–209
Xtxp141	(GA)23	29	23	0.91	135–167	121–179
SbAGB02	(AG)35	40	33	0.91	96–154	82–178
xtxp040	(GGA)7	25	17	0.84	129–141	111–159
Xtxp015	(TC)16	30	21	0.87	199–223	181–239
Submean		28.92	21.42	0.84		
EST-derived markers						
Xcup63	(GGATGC)4	27	25	0.48	133–145	123–185
Xcup14	(AG)10	29	14	0.84	211–225	183–257
Xcup61	(CAG)7	27	15	0.58	198–201	170–228
Xcup02	(GCA)6	16	12	0.59	192–204	162–230
Xcup53	(TTTA)5	17	14	0.51	186–198	180–222
Xcup62	(GAA)6	18	14	0.32	190–193	170–208
Submean		22.33	15.67	0.55		
Overall mean		26.72	19.50	0.75		

$N = 24$ individuals per population

A_t total number of alleles, A_r rare alleles, *PIC* polymorphic information content, *ASOP* allele size range in original publication and *ASOS* allele size range observed in this study

information obtained from the National Gene Bank of Kenya (NGBK), which indicated that there exist about eight distinct wild morphotypes in Kenya. For each run, both the burn-in length and number of replications were set at 100,000 iterations according to Pritchard et al. (2000). The number of sub-populations was determined according to Evanno et al. (2005).

The Bayesian genotypic clustering method InStruct (Gao et al. 2007) was used to validate population-based approaches and to infer population structure using an extended Bayesian clustering approach of STRUCTURE (Pritchard et al. 2000) that absorbs inbreeding or selfing rate for population inference. It quantifies the contribution of two forms of non-random mating—inbreeding and population substructure—when determining the pattern of existing genetic variation (Gao et al. 2007). InStruct was run for $K = 2$ to $K = 5$ in mode 2 for joint inference of population selfing rate and population substructure for five independent chains, each with 200,000 iteration steps, 100,000 burn-ins, and a thinning interval of ten steps, assuming different starting points. The K value was allowed to vary between two and five since the ad hoc Evanno et al. (2005) criteria has indicated an optimum value of $K = 5$ using the STRUCTURE. Cluster analysis of both molecular and phenotypic data was based on the neighbour-joining method implemented in MEGA v 4.0 (Tamura et al. 2006).

Spatial genetic structure

Country-scale spatial genetic structure of wild sorghum was analysed SPAGeDi v 1.2 (Hardy and Vekemans 2002) using two approaches: (1) the method of Rousset (1997), based on the computation of a linear regression of pairwise $F_{ST}/(1-F_{ST})$ estimates to the natural logarithm of geographic distances between pairs of populations, and (2) the method of Loiselle et al. (1995), based on the linear regression of pairwise kinship (Fij) to the natural logarithm of geographic distances between pairs of individuals. The significance of the regression slope was computed using 1,000 permutations under the null hypotheses of random relationship among populations and individuals, respectively, for each method.

Isolation by Distance Web Service software (IBDWS; Bohonak 2002) was used to determine whether there is a statistically significant relationship between the genetic and geographic distance matrix. IBDWS calculates the slope and intercept of the relationship using reduced major axis (RMA) regression. The deviation from the slope and intercept were estimated using standard linear approximations with a one-delete jackknife across population pairs (Bohonak 2002).

Correlation between genetic, phenotypic diversity and geographic distance matrices

Mantel (1967) test was used to test for significance of correlation between the genetic and phenotypic distance matrices. It is implemented in the Tools for Population Genetic Analysis (TFPGA) v 1.3 (Miller 2000). In addition, a Mantel test (Manly 1994) implemented in FSTAT v. 2.9.3.2 (Goudet 2002) was used to compare genetic and geographical distances matrices.

Results

Polymorphism of SSR markers

In total, 481 alleles were observed for the 18 SSR markers across all 62 populations, of which 351 (73%) were rare (Table 1). The number of alleles per locus ranged from 16 to 40 with a mean value of 27. All SSR loci scored were

highly polymorphic, displaying PIC values ranging from 0.32 to 0.92 with a mean of 0.75. In all cases the PCR amplification product size range was greater than reported for other sorghum genetic materials.

Genetic diversity of wild sorghum populations

The average number of alleles per SSR marker ranged among populations from 1.35 to 8.82 with a mean of 3.47 (Table 2). Highly significant differences ($p \leq 0.001$) were observed among populations for allelic richness, private allelic richness and gene diversity. Among populations, observed heterozygosity ranged from 0.02 to 0.61 (mean 0.27) and was generally lower than the expected values (gene diversity) which varied from 0.02 to 0.69 (mean 0.32). Eastern region had the highest mean for number of alleles per loci, private allelic richness, gene diversity and observed heterozygosity of 4.01, 0.14, 0.38 and 0.31, respectively. Turkana had the lowest mean for number of alleles per loci (2.96), while Western had the lowest private allelic richness (0.03) and gene diversity (0.29). Coast had the lowest observed heterozygosity (0.14). Overall, significant inter-regional differences ($p \leq 0.001$) were observed for allelic richness and private allelic richness but not for gene diversity ($p = 0.237$).

Genetic structure of wild sorghum populations

Estimates of F_{IS}, F_{IT} and F_{ST} were significant ($p < 0.001$) at both population and region level (Table 3). AMOVA also indicated significant genetic differences among regions and among populations within regions (Table 4). The pertinent variance components within and among populations within regions were more or less equal. Genetic variation attributable to regions was low but also significant. Populations were better discriminated using STRUCTURE analysis resultant clusters (12.52%) than region where they were collected (7.58%; Table 4).

STRUCTURE analysis predicted $K = 5$ as the optimum number of sub-populations, revealing that at least five distinct wild sorghum groups exist in the studied regions (Fig. 2). Most populations from Turkana belonged to one of these groups. In contrast, Coast populations were highly heterogeneous comprising sizable portions of each of the five groups. Western populations belong mainly to two groups and Eastern populations to three. Analysis with InStruct did not display much deviation in genetic structure patterns of wild sorghum populations compared to that of STRUCTURE (Fig. S4).

Cluster analysis based on modified Rogers distances estimated from the genotyping data revealed five groups, consisting of three major groups (designated i, iii and iv) and two minor groups (designated ii and v; Fig. 3a).

Table 2 Diversity statistics of 62 wild sorghum populations averaged across 18 SSR loci

Population	Region	A_r	A_{pr}	D (SD)	H_o (SD)	F	t_e
1	Turkana	2.38	0.05	0.42 (0.12)	0.31 (0.24)	0.16	0.72
2	Turkana	4.11	0.06	0.34 (0.27)	0.25 (0.26)	0.39	0.44
3	Turkana	2.51	0.00	0.36 (0.22)	0.26 (0.23)	0.22	0.64
4	Turkana	1.98	0.00	0.19 (0.21)	0.15 (0.12)	0.30	0.54
5	Turkana	2.36	0.05	0.23 (0.20)	0.22 (0.22)	0.56	0.28
6	Turkana	2.60	0.01	0.27 (0.25)	0.35 (0.26)	0.26	0.59
7	Turkana	3.18	0.00	0.32 (0.21)	0.25 (0.22)	0.35	0.48
8	Turkana	1.85	0.00	0.20 (0.21)	0.17 (0.21)	0.28	0.56
9	Turkana	4.52	0.05	0.39 (0.18)	0.37 (0.23)	0.54	0.30
10	Turkana	3.83	0.10	0.44 (0.19)	0.56 (0.16)	0.02	0.96
11	Turkana	1.89	0.00	0.10 (0.14)	0.14 (0.16)	0.34	0.49
12	Turkana	3.27	0.06	0.37 (0.24)	0.28 (0.18)	0.42	0.41
13	Turkana	4.68	0.01	0.54 (0.21)	0.44 (0.26)	0.26	0.59
14	Turkana	2.68	0.05	0.25 (0.26)	0.18 (0.22)	0.72	0.16
15	Turkana	3.34	0.05	0.22 (0.16)	0.22 (0.20)	0.29	0.55
16	Turkana	2.25	0.00	0.30 (0.19)	0.29 (0.18)	0.46	0.37
17	Turkana	2.95	0.15	0.36 (0.18)	0.24 (0.22)	0.38	0.45
Mean	Turkana	2.96	0.04	0.31 (0.13)	0.28 (0.21)	0.35	0.50
18	Western	3.43	0.10	0.44 (0.23)	0.38 (0.22)	0.36	0.47
19	Western	2.76	0.00	0.32 (0.25)	0.36 (0.21)	0.28	0.56
20	Western	3.71	0.00	0.33 (0.15)	0.32 (0.22)	0.62	0.23
21	Western	3.00	0.05	0.39 (0.24)	0.32 (0.27)	0.62	0.23
22	Western	3.35	0.00	0.23 (0.16)	0.22 (0.16)	0.63	0.23
23	Western	3.95	0.01	0.33 (0.22)	0.26 (0.24)	0.69	0.18
24	Western	3.28	0.01	0.29 (0.22)	0.29 (0.27)	0.66	0.20
25	Western	1.35	0.00	0.02 (0.03)	0.02 (0.03)	0.21	0.65
26	Western	2.12	0.06	0.12 (0.15)	0.14 (0.19)	0.16	0.72
27	Western	3.07	0.00	0.38 (0.23)	0.22 (0.25)	0.49	0.34
28	Western	2.38	0.11	0.13 (0.15)	0.09 (0.09)	0.76	0.14
29	Western	4.52	0.01	0.47 (0.23)	0.49 (0.17)	0.27	0.57
Mean	Western	3.08	0.03	0.29 (0.19)	0.26 (0.19)	0.48	0.38
30	Coast	2.88	0.11	0.28 (0.22)	0.19 (0.17)	0.35	0.48
31	Coast	2.75	0.00	0.25 (0.21)	0.21 (0.22)	0.74	0.15
32	Coast	1.63	0.01	0.05 (0.10)	0.08 (0.14)	0.88	0.06
33	Coast	4.55	0.11	0.36 (0.20)	0.25 (0.21)	0.79	0.12
34	Coast	4.66	0.07	0.36 (0.27)	0.28 (0.25)	0.69	0.18
35	Coast	4.32	0.05	0.29 (0.19)	0.29 (0.21)	0.67	0.20
36	Coast	3.58	0.00	0.34 (0.15)	0.25 (0.18)	0.47	0.36
37	Coast	2.34	0.01	0.20 (0.22)	0.14 (0.22)	0.84	0.09
38	Coast	3.35	0.10	0.26 (0.27)	0.18 (0.22)	0.38	0.45
39	Coast	2.56	0.10	0.23 (0.26)	0.10 (0.18)	0.48	0.35
40	Coast	5.25	0.23	0.35 (0.17)	0.40 (0.26)	0.55	0.29
41	Coast	4.66	0.26	0.36 (0.29)	0.23 (0.23)	0.55	0.29
42	Coast	3.12	0.16	0.19 (0.18)	0.10 (0.11)	0.62	0.23
43	Coast	2.03	0.00	0.21 (0.22)	0.15 (0.22)	0.52	0.32
44	Coast	3.25	0.06	0.31 (0.23)	0.26 (0.35)	0.82	0.10
45	Coast	7.23	0.23	0.63 (0.15)	0.61 (0.22)	0.29	0.55
46	Coast	4.37	0.26	0.40 (0.23)	0.29 (0.28)	0.71	0.17

Table 2 continued

Population	Region	A_r	A_{pr}	D (SD)	H_o (SD)	F	t_e
47	Coast	6.23	0.43	0.59 (0.25)	0.41 (0.24)	0.68	0.19
48	Coast	3.03	0.12	0.29 (0.23)	0.12 (0.14)	0.52	0.32
Mean	Coast	3.78	0.12	0.31 (0.21)	0.24 (0.21)	0.61	0.26
49	Eastern	4.62	0.07	0.44 (0.25)	0.44 (0.23)	0.12	0.79
50	Eastern	4.32	0.10	0.50 (0.24)	0.49 (0.27)	0.09	0.83
51	Eastern	3.24	0.11	0.41 (0.22)	0.40 (0.26)	0.33	0.50
52	Eastern	3.28	0.05	0.33 (0.21)	0.24 (0.21)	0.49	0.34
53	Eastern	2.82	0.05	0.32 (0.21)	0.20 (0.19)	0.41	0.42
54	Eastern	2.73	0.00	0.32 (0.21)	0.27 (0.23)	0.37	0.46
55	Eastern	2.52	0.06	0.21 (0.23)	0.17 (0.20)	0.52	0.32
56	Eastern	4.89	0.12	0.47 (0.22)	0.30 (0.30)	0.68	0.19
57	Eastern	3.43	0.17	0.32 (0.26)	0.24 (0.30)	0.58	0.27
58	Eastern	2.67	0.07	0.19 (0.21)	0.15 (0.23)	0.64	0.22
59	Eastern	2.56	0.11	0.15 (0.17)	0.16 (0.31)	0.51	0.32
60	Eastern	3.57	0.07	0.40 (0.15)	0.36 (0.25)	0.44	0.39
61	Eastern	6.71	0.18	0.52 (0.27)	0.39 (0.28)	0.67	0.20
62	Eastern	8.82	0.82	0.69 (0.23)	0.50 (0.29)	0.65	0.21
Mean	Eastern	4.01	0.14	0.38 (0.22)	0.31 (0.25)	0.46	0.39
Overall mean		3.47	0.09	0.32 (0.19)	0.27 (0.22)	0.48	0.38

$N = 24$ individuals per population

A_r allelic richness, A_{pr} private allelic richness, D gene diversity, H_o observed heterozygosity, F inbreeding coefficient [computed according to Weir (1996) pp 77–78] and t_e indirect equilibrium estimate of outcrossing rate [$t_e = (1-F)/(1+F)$]

Table 3 Estimates of F_{IS}, F_{IT} and F_{ST} at population and region level, and estimates of pairwise F_{ST} among wild populations and different geographical regions

F statistics	F_{IS}	F_{IT}	F_{ST}
Populations level	0.47*	0.77*	0.55*
Regional level	0.74*	0.77*	0.11*
Pairwise F_{ST}	**Eastern**	**Turkana**	**Western**
Coast	0.08	0.13	0.10
Eastern	*	0.09	0.09
Turkana		*	0.13
			*

* p value < 0.001

The Neighbour-Joining dendrogram revealed that most of the populations clustered closely according to their regions of origin with few overlaps. It confirmed that Turkana populations were strongly differentiated. Whereas populations from Turkana, Western and Coast regions showed relatively close groupings, those from the Eastern region were scattered across the three clusters (Fig. 3a). The dendrogram of the hierarchical cluster analysis using ten phenotypic traits deviated considerably from the marker-based dendrogram (Fig. 3b). Turkana populations clustered into three groups with a few populations scattered across the other clusters. Coast populations were scattered across all clusters. Western and Eastern populations clustered into two and three groups, respectively, with a few populations of each scattered across all the other clusters. The correlation between molecular and phenotypic clustering (genetic matrices), was low but significant ($r = 0.1344$; $p = 0.003$).

Spatial genetic structure

Rousette distance measure, $F_{ST}/(1-F_{ST})$ ratio, increased with logarithmic distance with a significant slope (slope $= 0.06$, $p < 0.001$; Fig. 4a, b). Likewise, Loiselle pairwise kinship coefficient at individual level show strong spatial structure (Fig. 4c). Wild sorghum populations had a mean regression slope (blog) value of -0.023 ($p < 0.001$) and coefficient of determination value (r^2) of 0.091 for Loissele pairwise kinship. Furthermore, isolation by distance displayed a loose relationship between the Modified Roger's Distance and the geographic distance of population pairs (Fig. 5). The slope of the $\log_{10}$ transformed regression was 0.13 (SE $= 0.003$) and the regression line explained 18% of the $\log_{10}$(MRD) variation. The slope was not significantly different ($p = 0.05$) from zero. A mantel test indicated a significant relationship between genetic distance and geographic distance (in kilometre) matrices ($r^2 = 0.053$, $p = 0.0005$).

Mating systems in wild sorghum populations

Inbreeding coefficients ranged from 0.02 to 0.88 (Table 2). Coast populations displayed the highest mean value for inbreeding depression (0.61) and Turkana the lowest (0.35). The indirect equilibrium estimates of the outcrossing rate (t_e) ranged from 0.06 to 0.96. Turkana population had the highest t_e (0.50) and Coast had the lowest (0.26). There was significant correlation between altitude and t_e ($r = 0.28$, $p > 0.0001$).

Discussion

Genetic diversity of wild sorghum populations

To the best of our knowledge this study describes for the first time in Kenya, genetic diversity and population structure of wild sorghum. The PCR product size ranges of the 18 SSR markers were greater than those reported previously (Brown et al. 1996; Taramino et al. 1997; Kong et al. 2000; Schloss et al. 2002) and may be because most

Table 4 AMOVA of 62 wild sorghum populations genotyped using 18 highly polymorphic SSR markers and based on (a) four geographical regions and (b) clusters obtained by STRUCTURE analysis

Source of variation	*df*	Variance component	Percentage of variation	*p* value
(a) Grouped by region of collection				
Among regions	3	0.53	7.58	<0.001
Among populations within regions	58	3.47	49.20	<0.001
Within populations	2,912	3.04	43.22	<0.001
(b) Group by structure analysis				
Among clusters	4	0.89	12.52	<0.001
Among populations within cluster	57	3.18	44.69	<0.001
Within populations	2,912	3.04	42.79	<0.001

df degrees of freedom

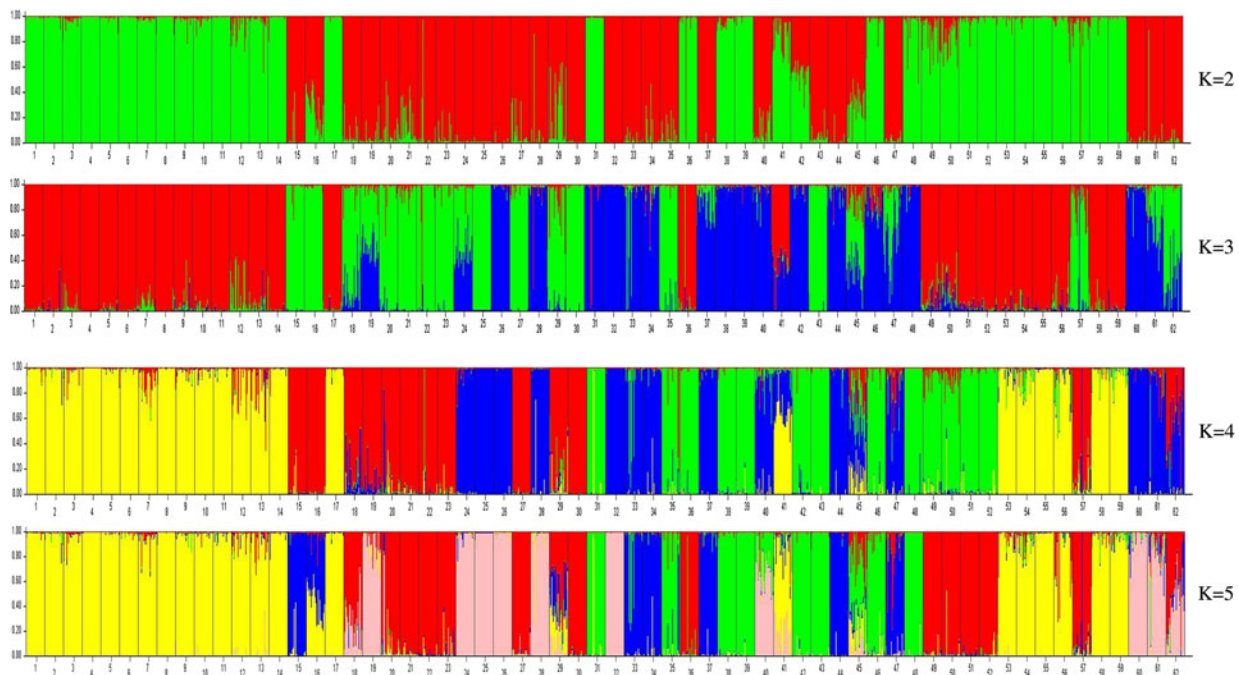

Fig. 2 *Bar plots* of the STRUCTURE analysis. Each of the 62 wild sorghum populations is represented by a *vertical bar* being partitioned in $K = 2$ up to $K = 5$ coloured segments that designate the population's estimated membership fraction in the inferred subgroups. Populations are sorted according to regions of origin and were collected from Turkana (1-17), Western (18-29), Coast (30-48) and Eastern (49-62) regions of Kenya, respectively (colour figure online)

earlier studies concentrated on cultivated sorghum and different geographic origins. Djè et al. (2000) and Manzelli et al. (2007) reported different size ranges of PCR products even in cultivated sorghum when using accessions from various countries. Kenyan wild sorghum displayed high genetic variability both among and within regions (Tables 2, 3, 4). Muraya et al. (2010) also found that Kenyan wild sorghum displays high phenotypic variation. However, the high degree of diversity observed at the molecular level among and within populations had a low correlation with observed phenotypic data. Genetic diversity was lower than found earlier by Casa et al. (2005). However, comparisons of genetic diversity between studies are difficult, since diversity statistics depend on the sampling schemes (single plant or DNA bulk), the number and type of surveyed markers, the polymorphic information content (PIC) of the markers, and the location of the markers in the genome.

The extensive diversity among populations within eco-geographic regions (Table 2) as was found in this study indicate the likely co-existence of different wild types in the same region. Muraya et al. (2010) reported that two or more

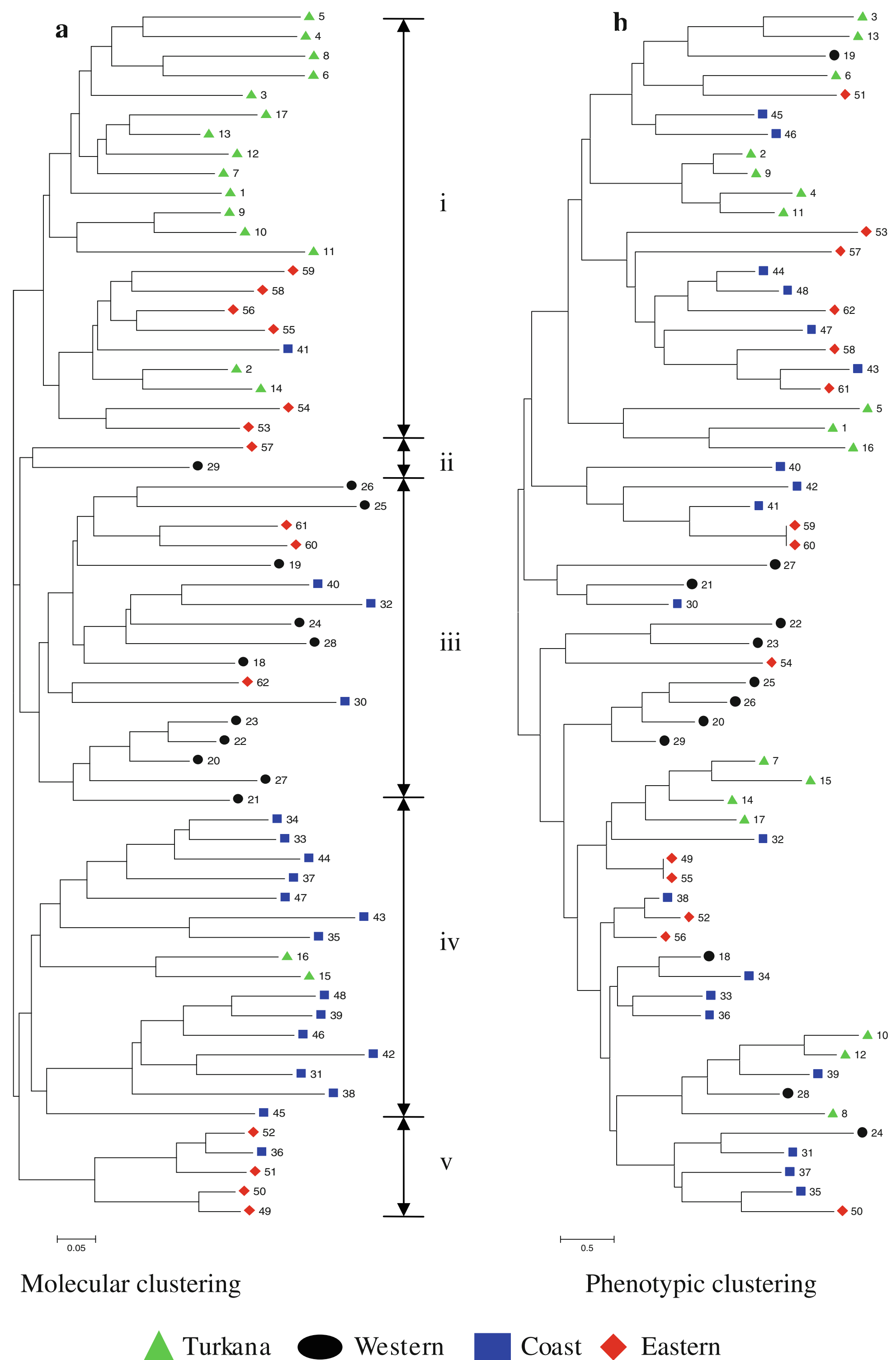

Fig. 3 Neighbour-joining (NJ) tree based on (**a**) modified Roger distances of allelic data from 18 SSR loci and (**b**) ten phenotypic traits among 62 wild sorghum populations collected from four regions of Kenya. Populations are identified by *number* and each region is identified by a different *symbol* and *colour* (colour figure online)

species of wild sorghum can be found in the same region. Furthermore, more than one race of *S. bicolor* spp. *verticilliflorum*, *S.* (races *aethiopicum*, *verticilliflorum*, and *arundinaceum* but not virgatum) have been found in the same region where this study collected samples (Muraya et al. 2010).

Genetic structure of wild sorghum populations

Analysis of molecular variance showed moderate differentiation between geographic regions but high among populations within regions and among individuals within populations (Table 3). Clustering patterns of populations

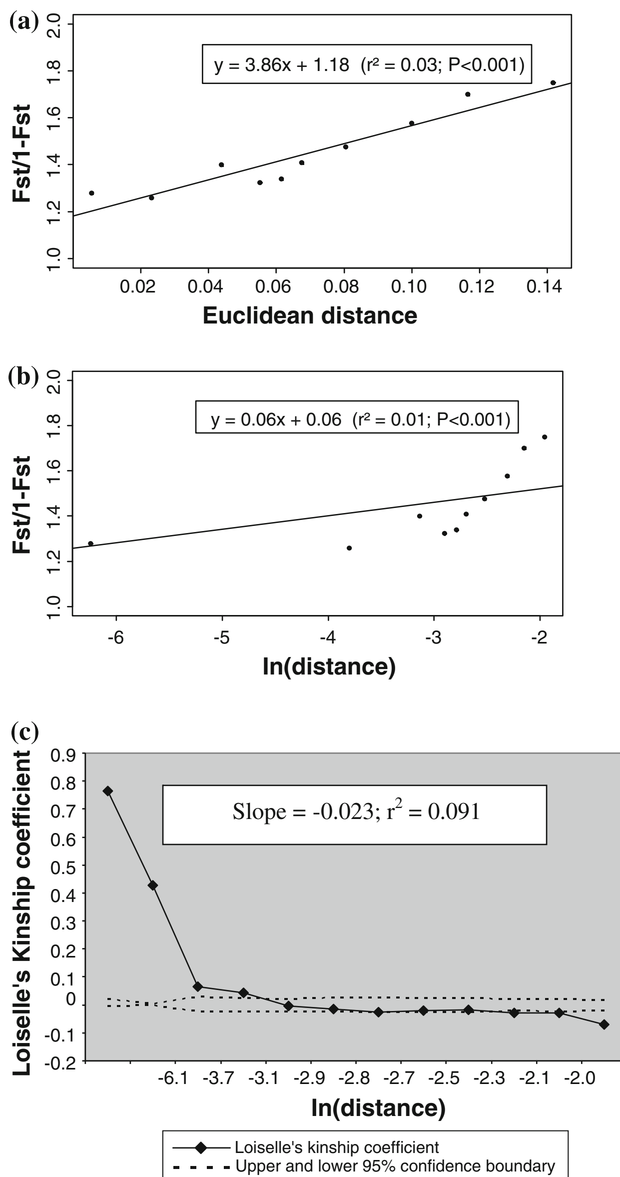

Fig. 4 Spatial patterns of genetic diversity among wild sorghum populations. A plot of the regress of **a** and **b** individual pairwise kinship coefficient analysis and **c** correlograms of pairwise relatedness (Loissele kinship coefficient) among individuals of wild sorghum populations with distance (Euclidean). *Dashed lines* represent the upper and lower 95% confidence limit envelopes around the null hypothesis of no spatial structure

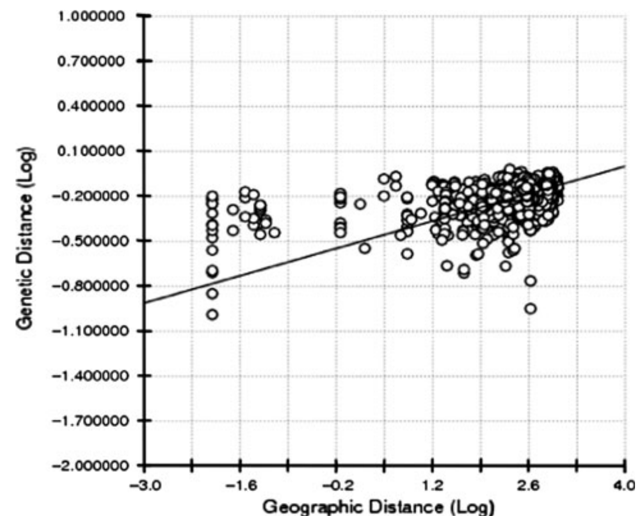

Fig. 5 Log_{10} values of modified Roger's distance [$\log_{10}$(MRD)] plotted against $\log_{10}$ values of geographic distance (km) among 62 wild sorghum populations sampled in four eco-geographic regions of Kenya (long rains, 2006). $Y = -0.52 + 0.13X$; $r^2 = 0.18$; $p < 0.0010$

(Fig. 2) suggest that there might be two centres of diversity of wild sorghum in Kenya, namely Turkana and Western regions and that gene flow from these regions considerably contributed to the diversity of the Eastern and Coast populations (Fig. 2).

The weak regional differentiation can be explained by the traditional agricultural practices in sorghum growing areas where wild sorghum is frequently used for fodder or for grazing. Seed dispersal of wild sorghum may have occurred through movement of fodder, manure, animals and/or exchange of cultivated-wild seed admixtures, enabling wild seeds to travel over long distances and even overcome geographic boundaries. Questionnaires completed by farmers revealed a frequent exchange of seed and other farming related inputs (such as manure and fodder) between relatives (Muteg et al. 2010). A weak regional differentiation of cultivated sorghum in the centre of origin that has been reported using qualitative morphological traits (Ayana and Bekele 1998) and molecular markers (Ayana et al. 2000), support this theory.

In our study, the genetic structure of wild sorghum populations could most plausibly be attributed to seed-mediated gene flow from region to region since structure analysis showed that Coast populations contained alleles from all other regions (Fig. 2). This was confirmed by personal communication with sorghum farmers during our survey in Coast region. The majority of farmers are immigrants, mainly from the Western and lower Eastern regions of Kenya, who brought their sorghum seeds with them from as far as districts neighbouring the Turkana region, a distance of almost 1,000 km (Fig. 1). This may also explain why Coast region showed the highest diversity (Table 2). In addition, the high levels of diversity found within populations are also most likely caused by seed-mediated gene flow. Muteg et al. (2010) also found substantial exchange of cultivated sorghum seeds between adjacent villages, which could lead to long distance seed dispersal as seeds are moved from one village to the next, which can also be coupled with pollen-mediated gene flow. Though not investigated in this study, we cannot rule out pollen-mediated gene flow between wild and cultivated

sorghums as a factor shaping genetic diversity and structure of wild sorghum populations. In fact, we observed many wild-cultivated-weedy complexes in both cultivated and fallow sorghum fields.

Cluster analysis illustrated how difficult it is to classify wild sorghum. Using phenotypic traits, wild sorghum can somehow be resolved into races and/or species on a regional basis, but at the national level a poor agreement between phenotypic and molecular distance measures make it difficult to resolve wild sorghum populations into races and/or species. The most obvious reason is that most of the traits that were used to discriminate wild sorghum in the past, such as plant height, leaf and panicle size, are quantitative and generally not reliable as they are influenced by the environment. Another possible reason for the classification problem is the vast amount of variation manifested within wild sorghum populations and the possibility of finding several taxa within a single population (Snowden 1936), implying that distinct ecotypes of the same race may occur in different regions due to regional adaptation.

Spatial genetic structure

Wild sorghum populations displayed a strong spatial genetic structure, which can be due to several factors, including gene flow—both seed- and pollen-mediated. In plant populations, spatial distribution of genetic variation will primarily be determined by seed and pollen dispersal, habitat distribution, micro-environment mediated selection and genetic drift (Levin and Kerster 1974; Epperson 1993). Seed and pollen dispersal causes relatedness among populations within sympatric range, while distant populations will be differentiated (Sokal and Oden 1978; Epperson 1993, 2004). Our study revealed long distance seed-mediated gene flow. Muteg et al. (2010) also showed incidences of medium to long distance seed exchanges in cultivated sorghum, mainly through inter-ethnic marriage relationships. A near identical pattern of spatial genetic structure in cultivated and wild sorghum would be expected due to inadvertent establishment of wild sorghum seed via cultivated sorghum seed systems and pollen-mediated wild-cultivated genes in areas of sympatric occurrence.

Moreover, a pattern of isolation by distance was observed with a large amount of variation over short distances, suggesting that the populations are at disequilibrium (Slatkin 1993). In addition, the lack of a strong pattern of isolation by distance may imply that the populations have only recently been introduced. Furthermore, gene flow that occurred after subpopulations were in place would have obscured patterns that might have existed before. Though isolation by distance may be caused by drift and/or regional adaptation, in our case we believe that it was caused by regional adaptation.

Mating systems in wild sorghum populations

Indirect equilibrium estimates of outcrosssing rates for the 62 wild sorghum populations collected in the four distinct regions indicated that wild sorghum exhibits a mixed mating system. Muraya et al. (2011) observed a similar trend in multi-locus outcrossing rates among 12 populations (250 progenies per population) collected from the same regions. The outcrossing rates were higher than those reported for cultivated sorghum. For example, Djé et al. (2004) reported outcrossing rates in cultivated sorghum in Morocco of 7–16%, whereas Barnaud et al. (2008) reported 5–40% for sorghum landraces in Cameroon. In Kenya, outcrossing rates of 5–7% have been reported in cultivated sorghum (Rabbi et al. 2010). The high values of t_e indicate that pollen-mediated gene flow within and among populations is not restricted since hybridisation is possible between interfertile wild sorghum species (Doggett 1988). The wide range of estimated outcrossing rates in wild sorghum populations suggest that environmental conditions may exist under which fitness is favoured by outcrossing and others under which selfing is more advantageous. In general, a flexible mating system should be more favourable for natural populations which are subdivided into small isolated units. Wild sorghum populations are often small and may benefit from cross-pollination as a way of mitigating genetic drift. Selfing is advantageous for solitary plants if no other pollen is around; in contrast, outcrossing allows for heterosis and the production of genetically variable offspring. A flexible mating system may enable self-compatible plants to take advantage of self-fertilisation or cross-pollination depending on the existing environmental conditions.

Conclusion

Wild sorghum is widely distributed in Kenya and display extensive phenotypic and genetic diversity. It was found to show a weak regional differentiation, reflecting the importance of human intervention in shaping the population structure through seed-mediated gene flow, which can render geographical boundaries irrelevant. The sympatric occurrence of cultivated and wild sorghums (Table S2) coupled with the extensive seed-mediated gene flow and high outcrossing rates of wild sorghum found in this study indicate that there is potential of crop gene flow into wild sorghum populations and vice versa. Therefore, introduction of genetically modified cultivars with transgenes that can confer a selective advantage in the wild through pollen-mediated gene flow into wild relatives may have adverse ecological effects via enhanced weediness and/or eventual extinction of wild types. It is worth noting that studies on the importance of farmer practices in shaping their living

environment are few, despite their significance for strategies to be implemented in the future to strengthen subsistence agriculture and, as consequence, food security.

Acknowledgments This study was funded by the United States Agency for International Development (USAID) Biotechnology and Biodversity Interface Program (BBI; Dr. Fabrice Sagnard), the Institute of plant Breeding and Population Genetics at the University of Hohenheim, Germany, and German Academic Exchange Service (DAAD: A0523923). USAID-BBI funded field experiments and collection trips. Costs of laboratory infrastructure and consumables at the Institute of plant Breeding and Population Genetics at the University of Hohenheim, Germany, were shared by BBI and the University of Hohenheim. We acknowledge the Kenya Agricultural Research Institute and Ben Kanyenji who supervised the collection of genetic materials in full compliance with regulations according to the Convention on Biological Diversity (CBD).

References

Anas, Yoshida T (2004) Sorghum diversity evaluated by simple sequence repeat (SSR) markers and phenotypic performance. Plant Prod Sci 7:301–308

Anttila CK, Daehler CC, Rank NE, Strong DR (1998) Greater male fitness of a rare invader (*Spartina alterniflora*, Poaceae) threatens a common native (*Spartina foliosa*) with hybridization. Am J Bot 85:1597–1601

Ayana A, Bekele E (1998) Geographical patterns of morphological variation in sorghum (*Sorghum bicolor* (L.) Moench) germplasm from Ethiopia and Eritrea: qualitative characters. Hereditas 129:195–205

Ayana A, Bryngelsson T, Bekele E (2000) Genetic variation of Ethiopian and Eritrean Sorghum (*Sorghum bicolor* (L.) Moench) germplasm assessed by random amplified polymorphic DNA (RAPD). Genet Resour Crop Evol 47:471–482

Barnaud A, Deu M, Garine E (2007) Local genetic diversity of sorghum in a village in northern Cameroon: structure and dynamics of landraces. Theor Appl Genet 114:237–248

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

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France

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

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

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzales-Candelas F, Kresovich S (1996) Multiple methods for 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

Conner AJ, Glare TR, Nap JP (2003) The release of genetically modified crops into the environment: Part II. Overview of ecological risk assessment. Plant J 33:19–46

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

Dillon SL, Shapter FM, Henry RJ, Cordeiro G, Izquierdo L, Lee LS (2007) Domestication to crop improvement: genetic resources for *Sorghum* and *Saccharum* (*Andropogoneae*). Ann Bot 100:975–989

Djè Y, Heuertz M, Lefèbvre 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, Lefebvre 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

Duvall MR, Doebley JF (1990) Restriction site variation in the chloroplast genome of *Sorghum* (*Poaceae*). Syst Bot 15:472–480

Ellstrand NC, Schierenbeck KA (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? Proc Natl Acad Sci USA 97:7043–7050

Epperson BK (1993) Recent advances in correlation analysis of spatial patterns of genetic variation. Evol Biol 27:95–155

Epperson BK (2004) Multilocus estimation of genetic structure within populations. Theor Popul Biol 65:227–237

Evanno G, 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 (2006) An integrated software package for population genetics data analysis. Arlequin v3.1. http://cmpg.unibe.ch/software

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

Folkertsma RT, Frederick H, Rattunde W (2005) The pattern of genetic diversity of Guinea-race *Sorghum bicolor* (L.) Moench landraces as revealed with SSR markers. Theor Appl Genet 111:399–409

Gao H, Williamson S, Bustamante CD (2007) An MCMC approach for joint inference of population structure and inbreeding rates from multi-locus genotype data. Genetics (online)

Goudet J (2002) FSTAT, a program to estimate and test gene diversity and fixation indices. (version 2 9 3 2)

Hails RS, Morley K (2005) Genes invading new populations: a risk assessment perspective. Trends Ecol Evol 20:245–252

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol 2:618–620

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

Hokanson KE, Ellstrand NC, Ouedragado JT, Olwenya PA, Schaals B, Reybould AF (2010) Biofortified sorghum in Africa: using problem formulation to inform risk assessment. Nat Biotech 28:900–903

IBPGR/ICRISAT (1993) Descriptors for sorghum (*Sorghum bicolor* (L.) Moench). International Board of Plant Genetic Resources, Rome, Italy. International Crop Research Institute for Semi-Arid Tropics, Patancheru

Kalinowski S (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Mol Ecol 5:187–189

Kong L, Dong J, Hart GE (2000) Characteristics, linkage-map positions and allelic differentiation of Sorghum bicolor (L.) Moench DNA simple-sequence repeats (SSRs). Theor Appl Genet 101:438–448

Levin DA, Kerster HW (1974) Gene flow in seeds plants. Evol Biol 7:139–220

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics Research Center, North Carolina State University, Raleigh

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). Am J Bot 82:1420–1425

Lowe A, Harris S, Ashton P (2004) Ecological genetics: design analysis and application. Blackwell Publishing, USA, pp 70–72

Mace ES, Buhariwalla HK, Crouch JH (2003) A high throughput DNA extraction protocol for molecular breeding programs. Plant Mol Biol Report 21:459a–459h

Manly BFJ (1994) Multivariate statistical methods: a primer. Chapman and Hall, New York

Mann JA, Kimber CT, Miller FR (1983) The origin and early cultivation of sorghums in Africa. Texas Agricultural Experiment Station, Bulletin 1454

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

Manzelli M, Pileri L, Lacerenza N, Benedettelli S, Vecchio V (2007) Genetic diversity assessment in Somali sorghum (*Sorghum bicolor* (L.) Moench) accessions using microsatellite markers. Biodivers Conserv 16:1715–1730

Menz MA, Klein RR, Unruh NC, Rooney WL, Klein PE, Mullet JE (2004) Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. Crop Sci 44:1236–1244

Miller MP (2000) Tools for Population Genetic Analyses version 1.3. A Windows program for the analysis of allozyme and molecular population genetic data. Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ

Muraya MM, Hartwig HH, Mutegi E, Kanyenji BM, Sagnard F, de Villiers S, Kiambi D, Heiko K, Parzies HK (2010) Geographical patterns of phenotypic diversity and structure of Kenyan wild sorghum populations (*Sorghum* spp.) as an aid to germplasm collection and conservation strategy. Plant Genet Resour 8:217–224

Muraya MM, Mutegi E, Geiger HH, de Villiers S M, 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

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

Nei M (1987) Molecular evolutionary genetics. Columbia University press, New York

Pilson D, Prendeville HR (2004) Ecological effects of transgenic crops and the escape of transgenes into wild populations. Ann Rev Ecol Evol Syst 35:149–174

Price HJ, Dillon SL, Hodnett G, Rooney WL, Ross L, Johnston JS (2005) Genome Evolution in the Genus *Sorghum* (*Poaceae*). Ann Bot 95:219–227

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

Pritchard JK, Wen X, Falush D (2007) Documentation for structure software: Version 2.2. http://www.biovip.org/UploadFiles/2009-12/1219737299.pdf

Rabbi IY, Parzies HK, Kiambi D, Haussmann BIG, Folkertsma R, Geiger HH (2010) Experimental studies on pollen-mediated gene flow in *Sorghum bicolor* (L.) Moench using malesterile bait plants. Plant Breed doi:10.1111/j.1439-0523.2010.01775.x

Rich PJ, Grenier C, Ejeta G (2004) *Striga* resistance in the wild relatives of sorghum. Crop Sci 44:2221–2229

Rieseberg LH (1991) Homoploid reticulate evolution in *Helianthus*: evidence from ribosomal genes. Am J Bot 78:1218–1237

Rieseberg LH (1997) Hybrid origins of plant species. Annu Rev Ecol Syst 28:359–389

Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, Lexer C (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301:1211–1216

Rousset F (1997) Genetic differentiation and estimation of gene flow from *F* statistics under isolation by distance. Genetics 145:1219–1228

Schloss SJ, Mitchell SE, White GM, Kukatla R, Bowers JE, Paterson AH, Kresovich S (2002) Characterization of RFLP probe sequences for gene discovery and SSR development in *Sorghum bicolor* (L.) Moench. Theor Appl Genet 105:912–920

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264–279

Snowden JD (1936) The cultivated races of *Sorghum*. Adlard and Son, London

Sokal RR, Oden NL (1978) Spatial autocorrelation in biology 2 Some biological implications and four applications of evolutionary and ecological interest. Biol J Linn Soc 180(10):249

Soltis DE, Soltis PS (1999) Polyploidy: recurrent formation and genome evolution. Trends Ecol Evol 14:348–352

Stewart CN, Halfhill MD, Warwick SI (2003) Genetic modification: transgene introgression from genetically modified crops to their wild relatives. Nat Rev Genet 4:806–817

Tamura K, Dudley J, Nei M, Kumar S (2006) Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Version 4.0 (Beta release) Center for Evolutionary Functional Genomics Bio design Institute, Arizona State University

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

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. http://www.R-project.org

Uptmoor R, Wenzel W, Friedt W, Donaldson G, Ayisi K (2003) Comparative analysis on the genetic relatedness of sorghum bicolor accessions from Southern Africa by RADPs, AFLPs and SSRs. Theor Appl Genet 106:1316–1325

Weir BS (1996) Genetic data analysis II. Sinauer Associates, Sunderland