Conserv Genet (2012) 13:1059–1071
DOI 10.1007/s10592-012-0353-y

RESEARCH ARTICLE

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

E. Mutegi · F. Sagnard · M. Labuschagne · L. Herselman · K. Semagn · M. Deu · S. de Villiers · B. M. Kanyenji · C. N. Mwongera · P. C. S. Traore · D. Kiambi

Received: 13 October 2011 / Accepted: 30 March 2012 / Published online: 11 April 2012

Abstract Little information is available on the extent and patterns of gene flow and genetic diversity between cultivated sorghum and its wild related taxa under local agricultural conditions in Africa. As well as expanding knowledge on the evolutionary and domestication processes for sorghum, such information also has importance in biosafety, conservation and breeding programmes. Here, we examined the magnitude and dynamics of crop–wild gene flow and genetic variability in a crop–wild–weedy complex of sorghum under traditional farming in Meru South district, Kenya. We genotyped 110 cultivated sorghum, and 373 wild sorghum individuals using a panel of ten polymorphic microsatellite loci. We combined traditional measures of genetic diversity and differentiation with admixture analysis, population assignment, and analyses of spatial genetic structure to assess the extent and patterns of gene flow and diversity between cultivated and wild sorghum. Our results indicate that gene flow is asymmetric with higher rates from crop to wild forms than vice versa. Surprisingly, our data suggests that the two congeners have retained substantial genetic distinctness in the face of gene flow. Nevertheless, we found no significant differences in genetic diversity measures between them. Our study also did not find evidence of isolation by distance in cultivated or wild sorghum, which suggests that gene dispersal in the two conspecifics is not limited by geographic distance. Overall our study highlights likely escape and dispersal of transgenes within the sorghum crop–wild–weedy complex if genetically engineered varieties were to be introduced in Africa's traditional farming systems.

Keywords Gene flow · Natural introgression · Hybridization · Genetic diversity · *Sorghum bicolor* · Traditional agro-ecosystem

Electronic supplementary material The online version of this article (doi:10.1007/s10592-012-0353-y) contains supplementary material, which is available to authorized users.

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

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

Present Address:
E. Mutegi (✉)
Department of Evolution, Ecology, and Organismal Biology,
Ohio State University, 318 W. 12th Ave, Columbus,
OH 43210-1293, USA
e-mail: e_mutegi@yahoo.com

F. Sagnard · M. Deu
CIRAD, UMR AGAP, Avenue d'Agropolis, 34398 Montpellier, France

M. Labuschagne · L. Herselman
Department of Plant Sciences, University of the Free State,
P.O. Box 339, Bloemfontein, South Africa

K. Semagn
International Maize and Wheat Improvement Center (CIMMYT), Village Market, P.O. Box 1041, Nairobi 00621, Kenya

B. M. Kanyenji
KARI-Embu Research Station, P.O. Box 27, Embu, Kenya

P. C. S. Traore
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT-Bamako), Station de Samanko, Bamako BP320 , Mali

Introduction

Gene flow between crops and their wild relatives is a widespread phenomenon that may have important genetic and evolutionary consequences in both kinds of populations (Ellstrand et al. 1999). Gene flow from wild to crop populations is considered an important avenue for broadening genetic diversity and/or creating new gene combinations in traditional farming systems, some of which may be of adaptive and/or agronomic value (Jarvis and Hodgkin 1999). Historically, breeders have also taken advantage of crop–wild cross-compatibility to transfer desirable traits such as pest and disease resistance from wild sources to cultivated varieties (for a recent review see Hajjar and Hodgkin 2007). However, gene flow can also negatively impact the evolutionary ecology of both domesticated and wild populations.

The advent of genetically modified (GM) crops in agriculture has been accompanied by concerns about the potential for transgene escape to sexually compatible wild populations via gene flow. It has been suggested that transgenes may confer a fitness advantage to recipient wild populations, potentially leading to the evolution of increased invasiveness and weediness (Snow and Moran-Palma 1997; Conner et al. 2003). Also, extensive gene flow from domesticated crops (GM or otherwise) may lead to reduced fitness and in extreme cases local extinction of recipient wild populations through demographic swamping (Wolf et al. 2001). Likewise, extensive gene flow may lead to genetic assimilation of locally rare taxa (cultivated or wild) into the more frequent one (Ellstrand et al. 1999; Haygood et al. 2003). Understanding the patterns of natural gene flow between domesticated crops and their wild relatives is therefore important, not only as a critical component of risk assessment for the release of transgenic crops, but also as a prerequisite for effective conservation and management of genetic resources. Centers of origin and diversity of crop plants provide interesting systems for evaluating patterns, extent and the consequences of natural introgression between crops and their wild relatives. Such centers are known to harbour diverse types of crop landraces, often in close proximity to inter-fertile wild and/or weedy relatives.

Sorghum, *Sorghum bicolor* (L.) Moench, is one of the world's most important cereals. It is a dietary staple in arid and semi-arid lands of Africa and Asia. Sorghum originated in Africa (Harlan and Stemler 1976) and consists of three subspecies: *S. bicolor* ssp. *bicolor*, which includes the domesticated forms, *S. bicolor* ssp. *verticilliflorum* (Steud.) De Wet, the wild progenitor of cultivated forms, and *S. bicolor* ssp. *drummondii* (Steud.) De wet, highly heterogeneous weedy sorghum types that are thought to be stabilized derivatives of natural hybridization between domesticated and wild sorghum types. The three subspecies of *S. bicolor* are fully interfertile and grow sympatrically in many traditional agroecosystems of Sub-Saharan Africa. They constitute an extremely variable crop–wild–weedy complex of *S. bicolor* in which a continuum of morphologically intermediate putative hybrid forms have been recorded in sorghum and other cereals fields, and in intermediate habitats such as fallows, field margins and roadsides (Doggett and Majisu 1968; Doggett 1988; Tesso et al. 2008; Mutegi et al. 2010; Sagnard et al. 2011).

Understanding the extent and patterns of gene flow and genetic diversity within the crop–wild–weedy complex is the first step in characterizing the potential environmental risks for testing and/or releasing transgenic sorghum in Africa's largely traditional agroecosystem. It is also essential for designing strategies for effective conservation and management of these important genetic resources. Yet, relatively few previous studies have investigated patterns of gene flow and diversity within the crop–wild–weedy complex of *S. bicolor*, especially under local agricultural field conditions in Africa. Doggett and Majisu (1968) used hybrid index scores based on morphological characters to infer hybridization between samples of cultivated and wild sorghum obtained from three eastern African countries. More recently, patterns of diversity and gene flow in the crop–wild–weedy complex of sorghum have been investigated at a broad-scale in Kenya (Mutegi et al. 2011) and in Mali and Guinea (Sagnard et al. 2011) using molecular markers. In both cases the authors observed close genetic proximity between cultivated and wild sorghum, which was attributed to historical and/or contemporary post-domestication gene flow between the two congeners. However, with the notable exception of the study of Barnaud et al. (2009), the extent and patterns of gene flow and genetic diversity between cultivated and wild sorghum at the local-scale in Africa is poorly understood. Barnaud et al. (2009) combined morphological and molecular markers to confirm the introgressed status of some intermediate weedy sorghum types in a traditional cropping system at village level in northern Cameroon. Further, the authors showed that farmer practices (both conscious and unconscious) such as seed selection and weed control may have had an impact on the magnitude and characteristics of the observed levels of gene flow and genetic diversity within the crop–wild–weedy complex. Similar studies that use molecular markers to characterize local-scale patterns of gene flow and genetic diversity are needed in more and different regions of Africa. As yet, no such studies have been reported for Eastern Africa, even though it is the primary centre of origin and diversity for sorghum.

Here, we genotyped samples of cultivated and wild sorghum at ten microsatellite loci to uncover local-scale patterns of gene flow and genetic diversity in a traditional

farming system. The following two questions were more specifically addressed: (i) What, if at all, is the magnitude and pattern of gene flow between cultivated and wild sorghum? and (ii) Are there differences in the extent and patterns of genetic diversity and structure between cultivated sorghum and its wild relatives? de Wet et al. (1976), observed that, it is not easy to morphologically distinguish between members of subspecies *drummondii* and *verticilliflorum* as the former are thought to be derivatives of hybridization between the latter and members of the subspecies *bicolor*. Therefore, no attempt was made in this study to taxonomically discriminate between the wild subspecies. Rather, the term 'wild sorghum' (abbreviated as W) is henceforth used to refer to the entire non-cultivated pool, which was categorized in the field into either 'putative hybrid' (HM) or 'pure wild' (WM) based on morphological evidence of hybridization as described by Doggett (1988).

Materials and methods

Study site description

The study was conducted in an 8 km × 8 km site on the eastern slopes of Mt. Kenya, on the easternmost limit of Meru South district, near the town of Chuka, Kenya (Fig. 1). This site was selected based on the importance of sorghum to the farming communities and because wild sorghum occurs sympatrically with its cultivated counterpart within and around cropping fields. Furthermore, the area was easily accessible, and characterized by extensive environmental heterogeneity and variability of farmers' cropping systems and practices and, farmers were willing to participate in the study by providing information and samples. As is the case with traditional cropping systems in Africa, farming in the site is exclusively rain fed, small scale and largely for subsistence purposes. The study site is located within an altitudinal gradient ranging from 750 to 1,050 masl. It experiences a bi-modal rainfall regime, with the short, more reliable rain from October to December and long, less reliable rain from March to May. Subsequently, the short rains provide the major cropping season with planting around November and harvesting in late January/February, while the long rains correspond to the minor season whose planting is in late February/March and harvesting in July/August. Although sorghum is cultivated throughout the study site, it is more important in the drier lower altitudes, compared to the wetter higher altitudes where maize and tobacco are more important crops.

Sample collection

Sampling was conducted in 14 sites (each representing a farmer household) across the study site in February 2007 to correspond to the main harvesting season. The households were selected to include a wide range of environmental and cropping systems and, include the greatest possible diversity of cultivated sorghum as well as, to acquire populations of wild sorghum from a wide range of habitats within the agro-ecosystem. Farmers recognized and named cultivated varieties using different landrace names; nevertheless, they were mostly found to purposely grow assemblages of landraces in the same field. Five panicles of each landrace were randomly sampled in eight different farmers' fields. Preliminary surveys within the study site,

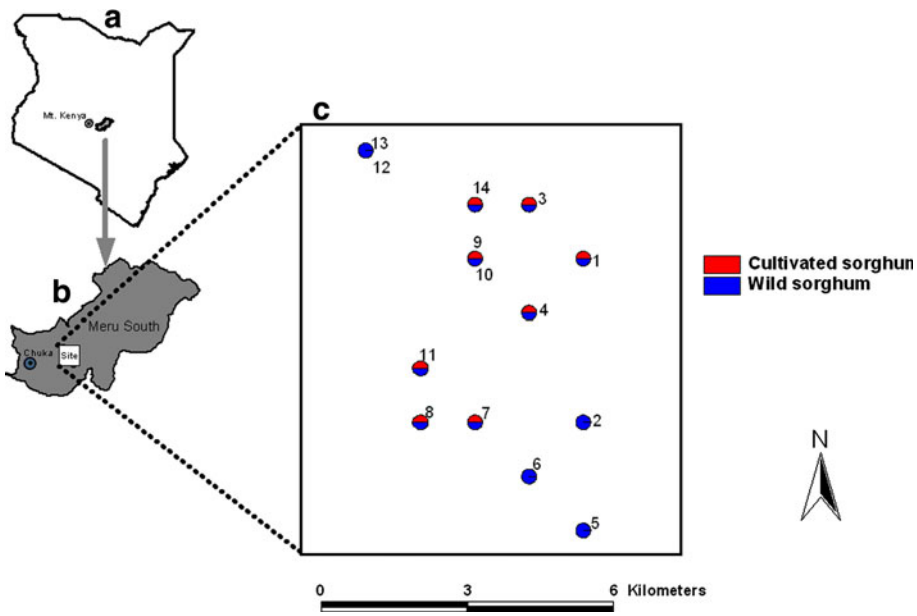

Fig. 1 Location of the study site: **a** Map of Kenya indicating the location of Meru south administrative district where the survey was undertaken (eastern slopes of Mt. Kenya). **b** Map of Meru south with the location of the 8 km × 8 km study site (*white square*) in the district indicated. **c** Detailed map of the study area showing the distribution of the sampling points (denoted by *numbered circles*). Each sorghum type is represented by a different colour: *red* cultivated sorghum and *green* wild sorghum. Points where populations of the two were sampled sympatrically are indicated by *mixed-colour* (*red* and *green*) *circles/pies*. (Color figure online)

Table 1 List of collected cultivated and wild sorghum populations

Site ID	Population ID	Types	Variety names[a]	Habitat (approx. size)	Altitude (masl)
1	2	W (HM)		F (<50)	841
	3	C	Ciamaguru		
	4	C	Muruge		
	5	C	Kaguru		
	6	C	Mucuri		
	7	C	Kathirigwa		
	8	C	Mugana		
2	20	W (WM)		FM (<50)	820
3	25	W (HM)		SF (50–100)	884
	26	C	Musalama		
	27	C	Kaguru		
	28	C	Muruge		
	29	C	Mugana		
4	30	W (HM)		F (>100)	846
	32	C	Mugeta		
	33	C	Mugana		
5	36	W (WM)		FM (50–100)	817
6	39	W (WM)		SF (>100)	840
7	43	W (WM)		SF (50–100)	869
	44	C	Kathirigwa		869
8	45	C	Kathirigwa		890
	46	C	Serena		
	47	W (WM)		SF (<50)	
9	51	W (WM)		F (<50)	874
10	53	C	Kathirigwa		869
	54	W (WM)		SF (<50)	
11	55	W (HM)		SF (50–100)	884
	56	C	Mugana		
	57	C	Munyerege		
	58	C	Kathirigwa		
12	60	W (WM)		SN-R (<50)	1,032
13	61	W (WM)		SF (50–100)	1,012
19	63	W (HM)		SF (>100)	936
	64	C	Mugana		
	66	W (WM)		SN-G (50–100)	
	65A	C	Kathirigwa		
	65B	C	Kathanta		

Sorghum type: *C* cultivated, *W* wild sorghum with morphotype in parentheses (*HM* putative hybrid, *WM* pure wild); Habitat: *F* fallow, *SF* sorghum field, *SN-R* Seminatural riverine, *SN-G* Seminatural grassland (with approximate total number of individuals per population indicated in parenthesis)

[a] Farmer-named cultivar names

had established that on average, four different sorghum landraces are grown per household (Mwongera et al. unpublished data). Thus, the objective was to sample at least three different sorghum landraces per household. Ultimately, one to six landraces were obtained from each sampled household (Table 1). In each household, mature panicles of co-occurring wild sorghum types were also randomly collected from 25 distinct mother plants, located at least 2 m apart. Seven more populations of wild sorghum (25 panicles each) were sampled outside sorghum fields in three contrasting habitats; field margins (FM; 2 populations), abandoned/fallow fields (FF; 3 populations) and semi-natural habitat (SN; 2 populations) close to crop fields. Each population of wild sorghum was categorized in the field into either a 'pure wild' morphotype (WM) or a 'putative hybrid' morphotype (HM) based on visual evidence of overall level of hybridization as described by Doggett (1988). The author remarked, "on sites such as abandoned cultivation, evident hybrids may be seen, sometimes also in farmers' fields. These usually have closer panicles than the wild type, with broader leaves, larger grain and tight black glumes. The grains shed readily

when ripe, with the glume attached. These hybrids cannot be confused either with the cultivated range of material, or with the wild type" (Doggett 1988 p. 9). Farmers referred to members of WM and HM by the same local name: "Munya wa maguna", which translates to "Monkey sorghum". In total, 15 wild sorghum populations comprising 373 individuals and 22 landrace samples (12 distinct landrace names) comprising 110 individuals were analyzed (Table 1). A map representing the sampling sites of wild populations and cultivated landraces is provided in Fig. 1. For each landrace, information was also gathered on how it was first introduced in the household (supplementary data Figure S3), as a way of gaining some insight on the local seed exchange systems.

DNA isolation and SSR genotyping

For each collected population/sample of cultivated and wild sorghum, five seeds were randomly selected from every panicle and germinated. For each panicle, total genomic DNA was isolated from fresh leaves (4–6 cm) collected from one 2-week-old seedling following a CTAB protocol described by Mutegi et al. (2011). Ten highly polymorphic SSR markers (Supplementary Table S1) were analyzed using the M13-tailed primer method (Schuelke 2000) to facilitate visualization on an ABI 3730 (Applied Biosystems) capillary sequencer. The ten markers were a subset of 30 SSRs recently used to characterize and compare genetic diversity and structure between cultivated and wild sorghum in Kenya at national scale (Mutegi et al. 2011). The ten SSRs were selected based either on their high overall polymorphism information content (PIC $\geq$ 0.70), or on their ability to distinguish ($F_{ST} \geq 0.10$) between gene pools of the two congeners. Polymerase chain reaction (PCR) conditions and genotyping was performed as described in Mutegi et al. (2011). DNA extraction and genotyping was carried out at the Biosciences eastern and central Africa (BECA) hub located at the International Livestock Research Institute (ILRI), Nairobi, Kenya.

Data analysis

Genetic diversity

The total number of alleles (A^t), number of rare alleles (A^r, alleles with a frequency <5 % in a group), private alleles (A^p, alleles unique to a group), observed heterozygosity (H_o) and unbiased expected heterozygosity (H_e) for each locus and sorghum type were calculated using GENETIX 4.05 (Belkhir et al. 2004) to estimate diversity. Because the number of observed alleles is highly dependent on sample size, the program HP-RARE (Kalinowski 2005) was used to estimate allelic richness (R_s) and private allelic richness (PR_s) for cultivated and wild sorghum. This program uses a rarefaction procedure to compute the two diversity parameters using comparable sample sizes. Differences in R_s, PR_s and H_e between the two sorghum congeners were further evaluated for significance using the Wilcoxon signed rank test in the program R (R Development Core Team 2011).

Genetic differentiation and spatial structure

To further investigate genetic divergence between cultivated and wild sorghum, among populations of wild sorghum, among samples of cultivated sorghum landraces, and among the three morphotypes (C, HM and WM) based on F_{ST} (Weir and Cockerham 1984) the program GENETIX was used. The significance of F_{ST} values was tested by 10,000 permutations. To assess isolation-by-distance (IBD) for cultivated sorghum and its wild congener at local scale, two approaches, both implemented in the program SPaGeDi (Hardy and Vekemans 2002) were used. First, a matrix of pairwise genetic distance [$F_{ST}/(1 - F_{ST})$] (Rousset 1997) was computed and regressed against a matrix of the logarithmic distances to estimate population-level IBD. Secondly, individual-level IBD was assessed by correlating a matrix of relative kinship coefficient against logarithmic distance using the method of Ritland (1996). This individual-level approach results in lower sampling variance, hence, thought to be powerful for detecting genetic structure (Hardy and Vekemans 2002). Pairwise geographic distances between sampling sites for cultivated and wild sorghum were in both approaches calculated from the linear distances between latitude and longitude positions using the software Geographic Distance Matrix Generator (http://biodiversityinformatics.amnh.org). Significance of the regression slope was tested by permuting population/individuals among sampling sites 1,000 times under the null hypothesis of no correlation between genetic and geographic distances. Results were visualized by plotting estimates of genetic distance against logarithmic geographic distance, with the regression line and the 95 % confidence interval envelope shown, using R.

Extent and patterns of gene flow

Extent and patterns of introgression between cultivated and wild sorghum were assessed using three complimentary approaches: the Bayesian clustering-based method, STRUCTURE (Pritchard et al. 2000), the Bayesian population assignment test-based method, BAYEASS 1.3 (Wilson and Rannala 2003), and a multivariate ordination method, the principal component analysis (PCA).

STRUCTURE was used to investigate the level of differentiation and to detect probable introgression between cultivated and wild sorghum gene pools. STRUCTURE

uses a Bayesian probabilistic model to infer the number K of genetic clusters in a sample of individuals, and for each individual, estimate the proportion of its genes contributed by each cluster (q_{ik}). A model where the entire dataset was assigned to two populations (i.e., K = 2; allowing two genetic clusters) was assumed, as it was theoretically relevant to the study of crop–wild hybridization. This model allowed, without prior information, to potentially identify two sets of allele frequencies typical of parental gene pools (i.e. cultivated and wild sorghum), and to detect the hybrid index within each individual. To test how well this model was supported by the data, STRUCTURE was also run assuming populations (K) from 1 to 10, with 10 runs per K value. Subsequently, the method of Evanno et al. (2005) obtained a modal value of ΔK at K = 2 (Supplementary Figure S1), supporting a two-population model as the most likely genetic structure for the data. Each STRUCTURE run was performed with the 'admixture' model, with no a priori information (on sorghum type or morphotype), with correlated allele frequencies, with a burn-in of 500,000 MCMC iterations, and 10^6 MCMC iterations of data collection.

BAYEASS was used to explicitly estimate rates of gene flow or migration between cultivated and wild sorghum gene pools, and for each individual of the two congeners to determine its migration (gene flow) history. BAYEASS uses a Bayesian assignment test-based algorithm to estimate whether an individual is an immigrant, of immigrant ancestry one generation back, or of local origin. Subsequently, the proportion of immigrants in the present and previous generations is used to obtain directional estimates of migration rates among populations. BAYEASS makes relatively few demographic assumptions. The program allows genotype frequencies to deviate from HWE but assumes large enough populations for negligible genetic drift to take place over two or three generations. The program was run with the default settings for number of MCMC iteration (3 million, of which 999,999 were discarded as burn-in) thinning interval of 2,000 (frequency of sampling data from the chain), and starting parameters. Two independent runs were performed to confirm the convergence of the MCMC at these default settings.

A centered PCA was performed in the software R using the 'packages' ADEGENET (Jombart 2008) and ADE4 (Dray and Dufour 2007). ADEGENET was used to code alleles as 0 if absent in an individual, 0.5 if heterozygous, and 1 if homozygous. The resulting matrix of 483 individuals and 134 alleles was subjected to PCA analysis using ADE4. PCA has the important advantage over both STRUCTURE and BAYEASS that it does not require strong assumptions about an underlying genetic model; rather, it associates individuals only on the basis of their genotypes. It provided a good option for corroborating inferences from the two Bayesian model-based analytical approaches.

Results

Extent and patterns of genetic diversity

The ten SSR loci used in this study were polymorphic in both cultivated and wild sorghum, and revealed a total of 134 alleles. Overall, 103 and 115 alleles were amplified in cultivated and wild sorghum, respectively (Table 2). The number of alleles per locus (A^t) ranged from 2 (msbCIR246, xcup53 and xtxp278) to 21 (msbCIR238) in cultivated, and from 3 (msbCIR246) to 18 (xtxp057, xgap206

Table 2 Summary of diversity indices for cultivated and wild sorghum

Marker	Cultivated sorghum (n = 110)							Wild sorghum (n = 373)						
	A^t	A^p	A^r	R_s	PR_s	He	Ho	A^t	A^p	A^r	R_s	PR_s	He	Ho
msbCIR238	21	6	16	20.7	9.0	0.867	0.140	16	1	11	12.6	0.9	0.790	0.140
msbCIR246	2	0	0	2.0	0.0	0.127	0.027	3	1	1	2.5	0.5	0.102	0.032
msbCIR248	7	1	1	7.0	1.0	0.727	0.359	9	3	4	7.5	1.6	0.700	0.241
sbAGB02	12	2	7	12.0	3.4	0.671	0.235	18	8	14	14.2	5.6	0.618	0.135
xcup53	2	0	0	2.0	0.0	0.195	0.029	4	2	2	3.3	1.3	0.370	0.110
xgap206	18	6	12	18.0	8.3	0.848	0.242	18	6	12	14.2	4.6	0.813	0.275
xtxp012	14	2	5	14.0	3.2	0.886	0.374	16	4	11	12.7	1.9	0.787	0.251
xtxp057	17	1	12	16.8	3.7	0.822	0.216	18	2	13	14.6	1.5	0.687	0.278
xtxp278	2	0	1	2.0	0.0	0.028	0.009	4	2	2	3.1	1.1	0.366	0.119
xtxp320	8	1	3	8.0	3.1	0.695	0.061	9	2	5	6.0	1.1	0.470	0.123
Overall	103	19	57	10.2	3.2	0.587	0.169	115	31	75	9.1	2.0	0.570	0.170

n number of samples, A^t total number of alleles, A^p number of private alleles, A^r number of rare alleles, R_s allelic richness, PR_s private allelic richness, *He* expected heterozygosity, *Ho* observed heterozygosity

and sbAGB02) alleles in wild sorghum. Overall, 57 (55.3 %) and 75 (65.2 %) alleles were rare (A^r) in cultivated and wild sorghum, respectively. The average allelic richness (R_s) was 10.2 (range: 2.0–20.7 per locus) in cultivated and 9.1 (range: 2.5–14.6 per locus) in wild sorghum. The number of private alleles (A^p) was 19 (18.4 %) and 31 (27.0 %) in cultivated and wild sorghum, respectively, and the average private allelic richness (PR_s) was 3.2 and 2.0, respectively. However, there were no statistical differences in the overall or private allelic richness between cultivated and wild sorghum (Wilcoxon signed rank test; $P \geq 0.05$) after correction for uneven sample sizes. Observed heterozygosity (Ho) ranged from 0.009 (xtxp278) to 0.374 (xtxp012) with an average of 0.169 in cultivated sorghum, whereas in wild sorghum it ranged from 0.032 (msbCIR246) to 0.278 (xtxp057) with an average of 0.170. Gene diversity (He) averaged over the ten loci was 0.587 (range: 0.028–0.886) in cultivated sorghum and 0.570 (range: 0.366–0.813) in wild sorghum. Wilcoxon signed rank test however, revealed no significant differences ($P \geq 0.05$) between cultivated and wild sorghum with respect to the extent of gene diversity (He).

When wild sorghum morphotypes were considered (i.e. WM and HM), the 'putative hybrid' group (HM) showed significantly higher levels of diversity (Wilcoxon signed rank test; $P \leq 0.05$) than the 'pure wild' group (WM), as estimated by allelic richness ($P = 0.037$) and gene diversity ($P = 0.004$) (Supplementary Figure S2 a and b). Notably, levels of diversity were not significantly different ($P \geq 0.05$) between the HM and cultivated sorghum (C) gene pools. Furthermore, WM revealed less diversity than C, even though the difference was significant only in terms of gene diversity ($P = 0.014$) but not in terms of allelic richness ($P = 0.084$). When the analysis was performed separately within the wild sorghum pool (both HM and WM populations), highly significant differences (Kruskal–Wallis test; $P \leq 0.0001$) were found among the 15 populations in the amount diversity, based both on allelic richness (range R_s: 1.51–3.61) and gene diversity (range H_e: 0.141–0.617). Notably, HM populations were characterized by higher levels of diversity than WM populations (Supplementary Figure S2 c and d).

Genetic differentiation and spatial genetic structure

The F_{ST} value obtained across all loci between cultivated and wild sorghum was 0.27 ($P < 0.001$), which indicated substantial genetic divergence between them. Individual locus estimates of F_{ST} were variable (Supplementary Table S2), ranging from 0.003 (in msbCIR246) to 0.663 (in xtxp278) and all except one were statistically significant ($P < 0.001$). Genetic differentiation was high among wild sorghum populations ($F_{ST} = 0.354$; $P \leq 0.001$), as well as among samples of cultivated sorghum landraces ($F_{ST} = 0.392$; $P \leq 0.001$). However, when the relative effect of geographic distance on genetic structure was investigated, no evidence was found of IBD patterns in cultivated sorghum landraces (population-level: Slope $= -0.142$ ($P = 0.12$), Fig. 2a; individual level: Slope $= -0.014$ ($P = 0.41$), Fig. 2c) or in wild sorghum (population-level: Slope $= 0.026$ ($P = 0.69$), Fig. 2b; individual-level: Slope $= -0.001$ ($P = 0.92$), Fig. 2d). Pairwise estimates of F_{ST} among C, WM and HM were also statistically significant (Supplementary Table S3) with the highest value ($F_{ST} = 0.42$) observed between C and WM and the smallest value ($F_{ST} = 0.09$) occurring between C and HM. Notably, the HM appeared to be genetically closer to C than to WM.

Extent and patterns of gene flow

Genetic structure of the 110 cultivated and 373 wild sorghum individuals based on STRUCTURE analysis at $K = 2$ without a priori information (on either taxonomic or morphotype classification) is presented in Fig. 3. Cultivated sorghum individuals were largely assigned to a single genetic cluster ($Cluster_I = 0.97$, $Cluster_{II} = 0.03$). Contrastingly, a substantial number of wild individuals appeared to jointly assign in the two genetic clusters ($Cluster_I = 0.28$, $Cluster_{II} = 0.72$). This was particularly evident in populations of wild categorized a priori in the field as HM. Consistent with the suspected hybrid origin; HM types appeared to contain substantial amount of genetic material from cultivated sorghum. Also, as anticipated, members of WM appeared to exhibit fewer admixtures with up to 92 % of its genome assigned to cluster II.

BAYEASS analysis of gene flow (recent migration history) for the 110 cultivated and 373 wild sorghum individuals is presented in Fig. 4. Overall, analysis revealed asymmetric gene flow from cultivated to wild sorghum because a higher rate of migration was indicated from the former to the latter (crop-to-wild; $m = 0.080$; $SD = 0.008$) than vice versa (wild-to-crop; $m = 0.003$; $SD = 0.003$). Based on the probability distribution of individual migrant ancestries, all cultivated individuals were correctly assigned to their source population (cultivar) with posterior probabilities (P) greater than 99 %. Contrastingly, at least 109 wild individuals (29.2 %) were indicated to be of migration/gene flow origin as they could not be unambiguously assigned ($P < 0.95$) to their population of origin ($P > 0.05$). Of these, 69 individuals (63.3 %) were assigned to cultivated sorghum whereas the remaining 40 individuals (36.7 %) could not be assigned to either of the two types ($P < 0.95$). Interestingly, 89 (81.6 %) of the 109 wild sorghum individuals indicated by BAYEASS to be of migration origin belonged to populations of wild sorghum categorized in the field as 'putative hybrid' on the basis of their intermediate morphology. Also, among

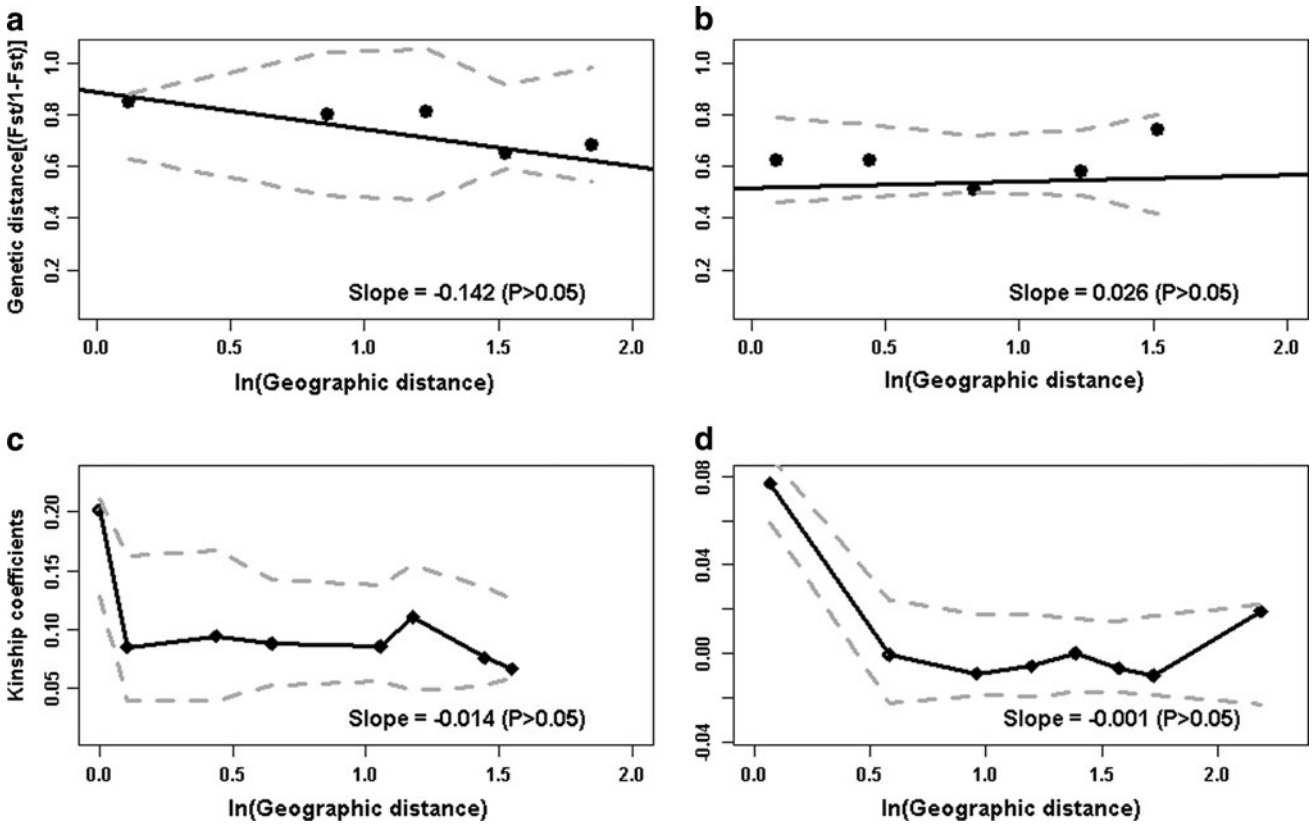

Fig. 2 Plots illustrating local scale patterns of isolation-by-distance (IBD) in cultivated and wild sorghum. Population-level IBD based on pairwise F_{ST} [$F_{ST}/(1 - F_{ST})$] and logarithmic distance is shown in **a** for cultivated sorghum landraces and in **b** for wild sorghum populations. Individual-level IBD based on relative kinship coefficient is shown in **c** and **d** for cultivated and wild sorghum, respectively. In each case, *dashed* lines demarcate the upper and lower limits for the 95 % confidence interval of the regression slope (*solid line*) under the null hypothesis of no spatial structure

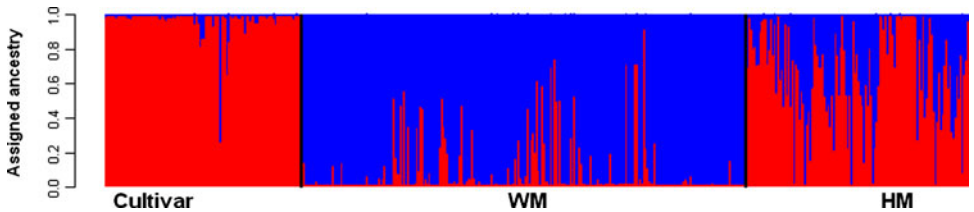

Fig. 3 Bar plot of the estimated genetic structure at $K = 2$ using the default STRUCTURE parameters with the individuals ordered by sorghum morphotype. Each individual is represented by a *vertical line* which is partitioned into coloured segments that represent its proportion of genome in K (*coloured*) clusters. (Color figure online)

the wild populations, those categorized as HM (i.e. 55, 2, 30, 25, 63) appeared to be the most introgressed as they had the smallest proportion of individuals assigned into their source (wild sorghum) population (Fig. 4). Overall, these results are in agreement with those obtained with STRUCTURE analysis.

PCA analysis provided further evidence of gene flow between cultivated and wild sorghum (Fig. 5a and b). Overall, the PCA bi-plot revealed substantial but incomplete divergence between cultivated and wild sorghum gene pools along the first axis (Fig. 5a). This finding is congruent with than obtained with STRUCTURE and F_{ST} analyses. When the wild sorghum morphotypes were considered, the cultivated sorghum gene pool (C) was clearly separated from WM along the first axis of the PCA bi-plot with putative crop–wild hybrid lineage (HM) forming a continuum between the two forms (Fig. 5b).

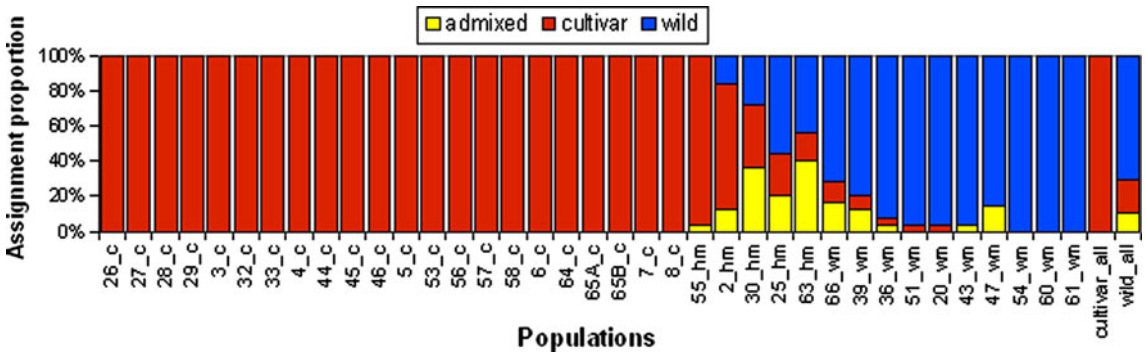

Fig. 4 Summary of individual migratory history assignment for cultivated and wild sorghum based on BAYEASS analysis. The proportion of individuals assigned to cultivated and/or wild sorghum gene pools are shown for each population/sample and overall for each sorghum type. The number below each *bar* on the *x*-axis denotes the sample/population id, whereas the letters *C*, *WM* and *HM* represent cultivar, wild and 'putative hybrid' morphotypes, respectively

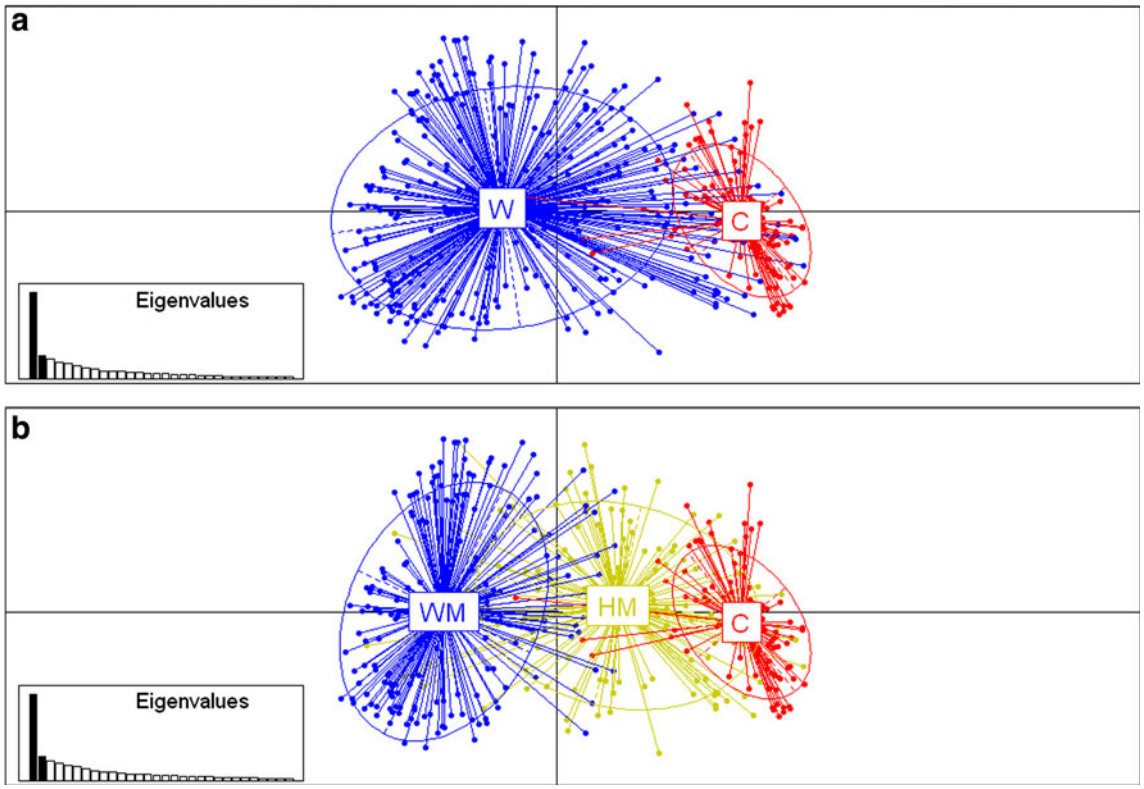

Fig. 5 Plot of first and second components of a principle component analysis (PCA) based on 10 SSR markers, 110 cultivated, and 373 wild sorghum individuals. **a** Shows a PCA plot of the cultivated sorghum (*C*) individuals against the entire pool of wild sorghum (*W*), whereas in **b**, the PCA plot shows the clustering pattern for cultivated sorghum and its wild relatives with the morphological classification of wild types (*HM* putative hybrid and *WM* pure wild) considered. Eigen values corresponding to the two components are filled in *black*. *Each point* represents an individual genotype, and is connected to the mean point of its group by a *line* of similar color. The *ellipses* are used to point out the 95 % confidence limit around the mean of the group

Discussion

Extent and patterns of gene flow

Our analysis at local scale in a traditional farming system indicates that gene flow between cultivated and wild sorghum was asymmetric, with higher rate from crop-to-wild than vice versa. STRUCTURE analysis (Fig. 3) revealed evidence of admixture between the two sorghum congeners, with up to 28 % of wild sorghum genome estimated to have origin in cluster I, the cluster associated with its cultivated counterpart. Also, BAYESASS analysis

detected a higher rate of gene flow (recent migration history) from crop-to-wild (m = 0.080) than vice versa (m = 0.003). Both of these analyses (Figs. 3 and 4), as well as PCA (Fig. 5) provided strong evidence of crop–wild introgression in populations of wild sorghum categorized in the field as 'putative hybrid' morphotypes (HM). Strikingly, the HM lineage appeared to form a continuum between the C and WM morphotypes in the PCA bi-plot (Fig. 5b), consistent with their presumed hybrid origin. This strongly points to concordance between molecular and morphological evidence for crop–wild hybridization in the crop–wild–weedy complex of *S. bicolor*. Barnaud et al. (2009) demonstrated similar trends in their investigations on crop–weed complex of sorghum in a traditional farming system in Cameroon although in that case no 'pure' wild sorghum populations were encountered. Our finding in a primary centre of origin and diversity for sorghum is also congruent with Papa and Gepts (2003) and Martinez-Castillo et al. (2007) who reported asymmetric crop-to-wild gene flow in crop–wild–weedy complexes of common beans (*Phaseolus vulgaris* L.) and lima beans (*Phaseolus lanutus* L.), respectively. Three non-mutually exclusive factors may explain this outcome.

Firstly, asymmetric crop-to-wild gene flow may reflect on-farm population size differences between crops and their wild relatives. In agricultural lands, crop population sizes are usually much higher than those of their sympatric wild/weedy relatives. This was the case in our study site largely because farmers consider wild sorghum as weeds and control them through hand weeding, once or twice in the crop's growing cycle. However, sometimes it is difficult for farmers to tell apart wild sorghum from their cultivated counterpart before maturity, and some plants may also remain due to the high labour requirement (Barnaud et al. 2009). Under such circumstances, cultivated sorghum populations are predicted to produce larger pollen densities relative to remnant populations of their wild relatives, thus leading to higher rates of pollen flow from crop-to-wild.

Secondly, asymmetric crop-to-wild gene flow may reflect differences in mating systems between the two conspecifics. Members of *S. bicolor* are considered predominantly autogamous, as supported by our finding of low observed heterozygosity in both cultivated and wild sorghum. Nevertheless, occasional cross-pollination does occur at rates that may be different for cultivated and wild sorghum. For example, Pedersen et al. (1998) reported higher rates of natural outcrossing (up to 61 %) in the weedy Sudan grass compared to grain type sorghum (up to 26 %) in the USA. Recently, Muraya et al. (2011a) reported up to 75 % outcrossing rate in an analysis of 12 populations of wild sorghum from different agro-ecological zones of Kenya. It was argued that loose panicles, such as those typical of the wild sorghum, may favour outcrossing, whereas the characteristically more compact panicles of cultivated sorghum, may restrict pollen movement by wind thereby impeding outcrossing (Dje et al. 2004; Barnaud et al. 2008). We acknowledge, however, that this hypothesis needs validation by comparative outcrossing studies among populations of the two sorghum types. To our knowledge, no systematic comparisons on outcrossing rate have been conducted between cultivated sorghum and its proposed wild progenitor.

Finally, seed selection by farmers has been invoked to explain, at least in part, the asymmetry of gene flow from crops to their wild/weedy relatives under traditional agriculture (Zizumbo-Villarreal et al. 2005; Martinez-Castillo et al. 2007; Barnaud et al. 2009). When selecting seed for the subsequent season, it was suggested that farmers can easily recognize and select against early generations of crop x wild hybrids, thereby reducing the probability of wild gene introgression into the cultivated gene pool. This may be the case in the current study site as farmers were found to grow mostly local varieties of sorghum, each season using seed from panicles selected from the previous season's harvest.

Evolutionary relationships within the crop–wild–weedy sorghum complex

Cultivated sorghum (*S. bicolor* ssp. *bicolor)* is thought to have been domesticated from the wild form (*S. bicolor* ssp. *verticilliflorum*) in the north eastern quadrant of Africa approximately 5,000 years ago (Doggett 1988). Doggett (1965) proposed disruptive selection, the simultaneous selection for more than one level of a particular character within a population, as the mechanism through which cultivated sorghum arose from its wild progenitor under sympatric occurrence. The author suggested that, the balance between farmer selection for cultivated traits and natural selection for wild characteristics resulted in both improved types and wild types of sorghum, and gene flow between the two generated intermediate forms. Doggett (1988) speculated that disruptive selection still exist in the plots of many African smallholders, where wild sorghum often co-occurs with its cultivated counterpart (Tesso et al. 2008; Barnaud et al. 2009; Mutegi et al. 2010). Our analysis in a traditional sorghum farming system in Kenya provides molecular evidence to support this hypothesis. Clear genetic divergence was found between cultivated and "pure" populations of wild sorghum (WM), whereas the "putative hybrid" (HM) populations showed evidence of crop–wild admixture and occupied an intermediate position between the C and WM morphotypes (Fig. 5b). Thus, we hypothesize that the HM morphotypes encompass subspecies *drummondii* and products of hybridization between cultivated sorghum (*S. bicolor* ssp. *bicolor*) and the

remnants of 'true' subspecies *verticilliflorum.* One important caution concerning our analysis, however, is that no natural allopatric populations of *S. bicolor* ssp. *verticilliflorum* were available within the study area to compare with populations of cultivated and wild sorghum sampled from farmers' fields. Further studies on evolutionary relationships between cultivated and wild sorghum could benefit from additional sampling of allopatric populations of wild sorghum from natural habitats.

Extent and partitioning of genetic diversity

Our study found similar levels of diversity in cultivated sorghum and its wild relatives at local scale, as estimated with gene diversity, allelic richness and private allelic richness. The results presented here are contrary to recent findings on a broader scale (national-level) in Kenya (Mutegi et al. 2011). The authors found that wild sorghum harbour significantly higher levels of diversity than their cultivated counterpart, which was attributed to a genetic bottleneck due to domestication and the associated evolutionary phenomena such as founder effect and artificial selection (Ladizinsky 1999). Our results are, however, congruent with those obtained by Barnaud et al. (2009) at a local scale in Cameroon, and by Sagnard et al. (2011) at a national scale in Mali and Guinea. These apparent deviations from theoretical expectations within the evolutionary framework for the domestication of plants may be explained at least in part by genetic drift due to smaller effective population sizes in wild relative to cultivated sorghum; and by asymmetric crop-to-wild gene flow with the amount of variation in wild sorghum predicted to approach that of the cultivated congener (Barrett and Kohn 1991; Ellstrand and Elam 1993). Most wild sorghum populations in the present study were found in sorghum fields and in transitory habitats such as field margins and fallows, so that farmer practices such as weeding, fallowing, and establishment and/or expansion of fields are likely to lead to rapid and frequent reductions in their sizes. The second hypothesis is supported by the results of admixture and gene flow analysis, which showed molecular evidence of cultivar alleles in wild sorghum, especially in 'putative hybrid' populations of wild sorghum (HM; admixture analysis, Fig. 3). Introgressive populations may show higher genetic variability than 'pure' ones (Riesberg and Wendel 1993), which may explain the higher allelic richness and gene diversity indicated in HM relative to WM (Supplementary data Figure S2 a and b).

Our study found some alleles present only in cultivated sorghum (19 out of 103), and others only in its wild relatives (31 out of 115). This finding suggests that the two gene pools have retained some degree of genetic distinctness, even in the face of ongoing gene flow between them. A large proportion of rare alleles were observed in both cultivated (55.3 %) and wild sorghum (65.2 %). These results are consistent with previous analysis of the two conspecifics at a national scale in Kenya (Mutegi et al. 2011). It has been argued that private or rare alleles are of adaptive or evolutionary significance, particularly if representing loci that are associated with adaptation to unusual conditions (Huenneke 1991). In the case of wild sorghum, such private and/or rare alleles may be linked to genes for novel traits such resistance to pests, diseases, and/or drought.

Gene dispersal

Isolation-by-distance theory (Wright 1943) predicts that genetic distance and geographic distance will be correlated with one another largely due to limited mating among individuals and/or limited dispersal of propagules. In our study at local-scale, no relation was found between genetic differentiation and geographic distance among cultivated sorghum landraces or among wild sorghum populations. This finding suggests that seed dispersal for the two sorghum conspecifics is not effectively limited by geographical distance. Although the result in cultivated sorghum landraces should be interpreted with caution given the small sample size (five individuals each), it is consistent with observations that seed exchanges among farmers can occur at a large spatial scale, and not necessarily just among neighbours. For example, local markets were indicated as the most important sources of introducing seed in households, followed by relatives from other villages (Supplementary Figure S3). Both of these seed exchange practices have the potential to spatially randomize gene dispersal within the study site. In wild sorghum, lack of an isolation-by-distance pattern may be explained by human-mediated secondary seed dispersal in seed lots (as contaminants) and/or dispersal of seed by grazing animals. These hypotheses need to be validated by the incorporation of more empirical data on local seed systems and other related farmer practices.

Practical implications

An overriding motivation of this study was to generate empirical data to guide the process of formulating biosafety regulations and guidelines for testing and/or commercially releasing transgenic sorghum in Eastern Africa, the primary centre of origin and diversity for cultivated sorghum and its wild relatives. Our data obtained at local scale in a traditional farming system in Kenya suggests that widespread gene flow takes place between cultivated and wild sorghum, with the prevalent direction being from crop-to-wild. The biosafety implication of our findings is that deployment of GM sorghum in farmers fields will most

likely lead to the escape and dispersal of the transgene into existing populations of the crop–wild–weedy complex of sorghum. This conclusion is reinforced by our finding that, local scale gene dispersal in cultivated and wild sorghum may not be limited by geographic distance due to among other, seed exchanges among farmers, contamination of seed lots, and/or use of local markets as sources of seed. Whether or not transgenic escape into wild/weedy populations will lead to negative effects such as extinctions, increased weediness and/or invasiveness will however depend on the nature of the transgene trait: whether neutral, detrimental, or beneficial in the ecological and genomic environment of the recipient population (Ellstrand et al. 1999). For example, Muraya et al. (2011b) recently found no fitness penalties in wild × crop hybrids of sorghum which led them to speculate that neutral or beneficial transgenes can persist in wild/weedy populations. Such hybrid fitness studies however need to be extended to GM sorghum in order to generate empirical data for the possible effect of particular transgenes in ecological and/or genomic environment of wild–weedy sorghum relatives.

Finally, sorghum landraces and their sympatric wild–weedy relatives constitute important genetic resources for sorghum breeding programmes and deserve conservation attention. Information on the amount and organization of genetic diversity within the crop–wild–weedy complex of sorghum can assist conservationists and/or breeders in focusing effort and resources on truly distinctive groups. Despite the observation of gene flow between cultivated and wild sorghum, our study found private alleles in the two congeners. We also found a large proportion of rare alleles in the two congers. Some of these private and/or rare alleles may be linked to important adaptive traits some that can be exploited in breeding programs to respond to existing and future biotic and biotic stresses.

Acknowledgments This work was supported by the USAID Biotechnology and Biodiversity Interface (BBI) and Plant Biosafety Systems (PBS) Programmes within the project: "Environmental Risk Assessment of Genetically Engineered Sorghum in Mali and Kenya.'' The authors owe gratitude to farmers within the study site for providing seed and information for this study.

Conflict of interest The authors declare that they have no conflict of interest.

References

Barnaud A, Trigueros G, Mckey D, Joly HI (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 et al (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, Kohn JR (1991) Genetic and evolutionary consequences of small population size in plants: implications for conservation. In: Falk DA, Holsinger KE (eds) Genetics and conservation of rare plants. Oxford University Press, New York, pp 3–30

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

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, Harlan JR, Price EG (1976) Variability in Sorghum bicolor. In: Harlan JR, de Wet JMJ, Stemler ABL (eds) Origins of African plant domestication. Mouton, The Hague, pp 453–463

Dje 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 (1965) Disruptive selection in crop development. Nature 206:279–280

Doggett H (1988) Sorghum. Longman Scientific and Technical Essex, New York

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

Dray S, Dufour A (2007) The ade4 package: implementing the duality diagram for ecologists. J Stat Softw 22:1–20

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

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

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

Hajjar R, Hodgkin T (2007) The use of wild relatives in crop improvement: a survey of developments over the last 20 years. Euphytica 156:1–13

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, Stemler ABL (1976) The races of sorghum in Africa. In: Harlan JR, De Wet JMJ, Stemler ABL (eds) Origins of African plant domestication. Mouton, The Hague

Haygood R, Ives AR, Andow DA (2003) Consequences of recurrent gene flow from crops to wild relatives. Proc Biol Sci 270: 1879–1886. doi:10.1098/rspb.2003.2426

Huenneke LF (1991) Ecological implications of genetic variation in plant populations. In: Falk DA, Holsinger KE (eds) Genetics and conservation of rare plants. Oxford University Press, New York, pp 31–44

Jarvis DI, Hodgkin T (1999) Wild relatives and crop cultivars: detecting natural introgression and farmer selection of new genetic combinations in agroecosystems. Mol Ecol 8:15–173

Jombart T (2008) *a*degenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405

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

Ladizinsky G (1999) Plant evolution under domestication. Kluwer, London

Martinez-Castillo J, Zizumbo-Villarreal D, Gepts P, Colunga-GarciaMarin P (2007) Gene flow and genetic structure in the wild–weedy-domesticated complex of *Phaseolus lunatus* L. in its Mesoamerican center of domestication and diversity. Crop Sci 47:58–66

Muraya MM, Mutegi E, Geiger HH, de Villiers SM, Sagnard F, Kanyenji BM et al (2011a) Wild sorghum from different eco-geographic regions of Kenya display a mixed mating system. Theor Appl Genet 122(8):1631–1639. doi:10.1007/s00122-011-1560-5

Muraya MM, Geiger HH, Sagnard F, Toure L, Traore PS, Togola S et al (2011b) Adaptive values of wild x cultivated sorghum (*Sorghum bicolor* (L.) Moench) hybrids in generations F1, F2 and F3. Genet Resour Crop Evol 59:83–93. doi:10.1007/s10722-011-9670-0

Mutegi E, Sagnard F, Muraya M, Kanyenji B, Rono B, Mwongera C et al (2010) Ecogeographical 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 S et al (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

Papa R, Gepts P (2003) Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. Theor Appl Genet 106:239–250

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

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

R Development Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, Austria (http://www.r-project.org)

Riesberg LH, Wendel JF (1993) Introgression and its consequences in plants. In: Harrison RG (ed) Hybrid zones and the evolutionary processes. Oxford University Press, New York, pp 70–109

Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. Genet Res 67:175–185

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

Sagnard F, Deu M, Dembe′le′ D, Leblois R, Toure′ L, Diakite′ M et al (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(7): 1231–1246. doi:10.1007/s00122-011-1662-0

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. A poor man's approach to genotyping for research and high throughput diagnostics. Nat Biotechnol 18: 233–234

Snow AA, Moran-Palma P (1997) Commercialization of transgenic plants: potential ecological risks. Bioscience 47:86–96

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

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

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163:1177–1191

Wolf DE, Takebayashi N, Rieseberg LH (2001) Predicting the risk of extinction through hybridization. Conserv Biol 15:1039–1053

Wright S (1943) Isolation by distance. Genetics 28:114–138

Zizumbo-Villarreal D, Colunga-GarciaMarin P, de la Cruz EP, Gado-Valerio P, Gepts P (2005) Population structure and evolutionary dynamics of wild–weedy domesticated complexes of common bean in a Mesoamerican region. Crop Sci 45:1073–1083