Full Length Research Paper

The population structure of wild sorghum species in agro-ecological zones of Western Kenya

Titus O. Magomere[1,4]*, Eliud K. Ngugi[1], Silas D. Obukosia[2], Eunice Mutitu[1] and Solomon I. Shibairo[3]

[1]Plant Science and Crop Protection Department University of Nairobi, P.O. Box 29053-00625, Nairobi, Kenya.
[2]Africa Harvest Biotechnology Foundation International, P.O. Box 642, 00621-Village Market, Nairobi, Kenya.
[3]Kibabii University College, Masinde Muliro University of Science and Technology, P.O Box 1699-50200 Bungoma, Kenya.
[4]Department of Biochemistry and Biotechnology, School of Pure and Applied Sciences, Kenyatta University, P.O. Box 43844, Nairobi, Kenya.

There is need to understand the genetic structure of wild sorghums that grow alongside cultivated traditional sorghum varieties in order to assess the potential effect of crop genes in wild populations. In this study, 175 wild sorghum samples were collected from 13 agroecological zones (AEZs) from three counties in Western Kenya and genotyped using microsattelite markers. Crop alleles were observed in wild sorghum populations. The range of allelic frequencies varied from low (<0.4), to moderate (0.4-0.7) and to high (0.7) in the AEZs. Wild sorghum populations had moderate to high expected heterozygosity (H_E) values of between 0.453 in LM_1 to 0.715 in LM_2. Differences in the magnitude of diversity was significant in the counties (Busia H_E = 0.59 – 0.71; Homabay H_E = 0.58-0.68 and Siaya H_E = 0.45-0.59) but not distinct among the AEZs. Whole population F_{IS}, F_{ST} and F_{IT} values were low at 0.15, 0.16 and 0.29, respectively indicating low level of inbreeding, low genetic differentiation of the population and low to moderate deviation from Hardy–Weinberg (HW) equilibrium respectively. The deviation from HW equilibrium was significant in some wild populations from Siaya and Busia. Intra-population diversity (H_S) was larger than inter-population diversity (D_{ST}) in 13 populations from the sampled AEZs, indicating the importance of gene flow between populations of wild sorghums. Heterozygosity values under mutation drift equilibrium (H_{EQ}) varied under infinite allele model (IAM), two–phase model (TPM) and the step wise mutation model (SMM). However, significant population bottlenecks were absent in the wild sorghums. Presence of significant geographic county clusters and lack of significance on AEZ clusters indicate that human activities have had more influence on the distribution and diversity of wild sorghums than the prevailing climatic conditions. Efforts towards physical and genetic containment of crops genes need to be enhanced for successful ecologically sensitive confined field trials and future adoption of transgenics in cropping systems.

Key words: Diversity, *Sorghum bicolor*, *Sorghum halepense*, *Sorghum sudanense,* microsatellite loci.

INTRODUCTION

The sympatric nature of the members of the sorghum genus over time may have contributed to the spon- taneous occurrence of *Sorghum bicolor* alleles in *Sorghum halepense* and *Sorghum sudanense* populations

(Ejeta and Grenier, 2005; Mutegi et al., 2009). *S. sudanense*, a cultivated form of *S. bicolor* ssp. *drummondii*, has shown natural outcrossing ranging from 0-100% on individual plants with averages of 39 and 57% when most tillers were at anthesis (Pedersen et al., 1998). Morrell et al. (2005) found *S. bicolor* alleles in *S. halepense* weeds suggesting persistent natural and spontaneous out-crossing in the species. The potential for crop to wild hybridization in sorghum populations in Ethiopia (a center of origin) and in Niger was found to be widespread (Tesso et al., 2008; Adugna et al., 2013). Crop and wild sorghums have been found to be intermixed in farming communities where they were interfertile and had synchronized flowering with several putative crop wild hybrids being observed (Tesso et al., 2008). Ancient and recent cross hybridization events after speciation in the sorghum genus have maintained several crop genes in the wild sorghum species (Morrell et al., 2005). In some cases the spontaneous hybrids formed could have obtained features from both parents, whose heterosis and unique features may have resulted in formation of a bridge species within the sorghum genus. Crop alleles might have an impact on the fitness of the crop x weed hybrids and may enhance or diminish their weedy potential (Hokanson et al., 2010).

Diversification within the sorghum genus heavily draws from disruptive selection (Dogget and Majisu, 1968). This phenomenon favors extreme phenotypes to the inter-mediates leading to wide variability within the species on specific traits (Rueffler et al., 2006). This has led to the wide variability on the morphological features of sorghum and on important food and forage associated traits as a result of artificial and natural selection for more than one level of a particular character within populations of crop and weed sorghums (Doggett, 1988). This diverse variability is also observed on important microsatellite locus in the genus (Yonemaru et al., 2009). The polymorphism on the microsatellite loci shows main-tenance of extreme mutations showing losses or addition of several repeats on mononucleotide, di-nucleotide, tri-nucleotide and tetra-nucleotide SSR motifs. Individual species within the genus also show similar variability at the genotypic and phenotypic levels (Yonemaru et al., 2009).

Analysis of the population structure of crop and wild sorghum species is important in Africa and Ethiopia, which is a probable "center of origin" of the genus.

Therefore, wide genetic diversity is expected to be naturally maintained in East and Central Africa regions and other secondary centers of diversity where sorghum plays an important role in human diets. Recent studies have shown that there is heterozygosity excess and diversity in crop and wild sorghum populations in some parts of Kenya (Mutegi et al., 2011) Ethiopia (Adugna et al, 2013) and Africa (Barnaud et al., 2007; Sagnard et al., 2011) with the wild sorghums contributing higher magnitudes of diversity than the crops or landraces. North-West and South-East wild sorghum populations from Kenya had H_{EQ} of 0.34 and 0.56 each indicating no apparent loss of genetic variability from the expected diversity (Muraya et al., 2010). Mutation models are commonly used for the analysis of population bottlenecks in natural populations. The step wise mutation model (SMM) considers the size of microsatellite repeats, thus it may be subject to problems of homoplasy in its interpretation. The infinite allele model (IAM) assumes that each mutation can create any new allele randomly and considers any point mutation along a stretch of DNA within a locus to constitute a new allele (Goodman, 1998; Schlotterer, 2000). The two-phase model (TPM) is an intermediate model that provides better analysis of how DNA sequences evolve (Di Rienzo et al., 1994). Excess of H_{EQ} over the observed heterozygosity on loci evolving under the IAM, SMM, or TPM models would suggest environmental evolutionary pressures had minimal effect on reducing the population size and thus the allelic diversity (Cornuet and Luikart, 1996). Therefore, indicating absence of significant population bottlenecks in the recent past. This situation is sustained despite the pressures that the wild populations would have experienced as a result of agricultural practices like weeding and changes in biotic and abiotic conditions. If such pressures were significant, mutation and genetic drift would have been important in the populations thus an excess of observed heterozygosity over H_{EQ} would have been apparent. In some members of Poaceae spontaneous occurrence of crop alleles and crop transgenes in wild and landrace populations have been observed (Gepts and Papa 2003). In maize (*Zea mays*) distinct transgenes have been shown to exist in landraces in Mexico showing evidence of genetic contamination of the centers of origin (Piñeyro-Nelson et al., 2009). Hybridization between maize and teosinte (*Zea mays ssp. parviglumis*) (wild) has been shown to

*Corresponding author. E-mail: magomeretito@gmail.com.

exist naturally despite the distinct morphological differences that exist between the two (Doebley, 2004). In rice, natural hybridization between *O. rufipogon* and *O. sativa* has been reported in important rice cultivation regions globally (Song et al., 2002).

With the recent development and introduction of transgenic crops (James et al., 2010), it is vital to isolate and characterize the role of crop alleles within local wild populations of sorghum and study their spatial and temporal significance (Hokanson et al., 2010). Intra-population and inter-population statistics provide a means to elucidate the diversity of wild sorghum populations in the agro-ecological zones (AEZs) from three counties in Western Kenya. This study evaluated the effect of spontaneous hybridization in conspecific sorghum species on the spatial and temporal allelic composition of wild sorghum populations. The study also evaluated allelic differentiation among wild sorghum species across AEZs in Homabay, Siaya and Busia counties. The local population structure of wild sorghum species in AEZs from the three counties of Kenya was examined. In addition heterozygosity deficiency under the infinite allele model (IAM), under the two–phase model (TPM) and step wise mutation model (SMM) was evaluated.

MATERIALS AND METHODS

Three counties (Homabay, Siaya and Busia) were sampled to represent Western Kenya sorghum producing regions around Lake Victoria (Figure 1). Sampling units or farms that had wild sorghum forms within the counties were clustered based on the mapped AEZs. Sorghum farms were randomly sampled and 175 samples were collected from 13 AEZs from three counties in July 2011 (Table 1). Sampling was done after seed set in farms that had both wild and crop sorghums. Some wild sorghums growing on road reserves and uncultivated land in the clusters were also sampled. The seeds and first two leaves next to the panicle were collected labelled and kept on ice. Sampled AEZs in the three counties represented the diversity of wild sorghums in varying ecological conditions (Table 2).

DNA extraction and electrophoresis analysis

Sampled leaves and seeds were maintained on ice during transportation to the laboratory located at the College of Agriculture and Veterinary Sciences (CAVS) (-1° 14' 59.72", +36° 44' 30.79") of the University of Nairobi. The leaves were stored at -20°C in the laboratory freezers while the seeds were dried in the greenhouse. The seeds were germinated and young leaves collected and maintained on ice for DNA extraction. Sampled leaves and young leaves were washed with detergent in running water to remove dust particles and other debris. Leaves were weighed and 0.3 g kept in clean and labelled polyethylene bags on ice until grinding. The seeds from sampled panicles were threshed and kept in labelled plastic bags. Two procedures were applied to obtain genomic DNA from young leaves (two to five weeks old plants), from old leaves (five weeks old to flowering plants) and from dried seeds. Genomic DNA was extracted from young leaves by using a modified CTAB extraction procedure (Doyle and Doyle, 1990; Barnaud et al., 2008). Total nucleic acid extraction from seed was done by a modified CTAB-based seed extraction protocol (Delobel et al., 2007).

Agarose gel electrophoresis for genomic DNA extracted from young leaf, old leaf and seed tissues was done before running multiplex PCR. The multiplex PCR products were separated and analysed using a 4% UltraPure ™ Agarose from Invitrogen ™.

Simple sequence repeat (SSR) marker assay

Primers were designed based on the polymorphic SSR sequences from crop sorghum physical map genomic clones on five linkage groups chromosomes found in Phytozome databases. To ensure the specificity of the primers prior to synthesis, e-PCR was performed for all primer combinations. Annealing sites were also confirmed by using BLASTn procedure against the NCBI GenBank database. The primers were selected on the basis of their polymorphism and ability to distinguish several wild sorghum accessions (Table 3). Three out of the seven primers analysed had high polymorphic index.

PCR was done in 0.5 ml reaction tubes with a hybaid® thermocycler. Reaction volumes of 11 µl were used in all experiments. The reaction conditions were set in conventional PCR format using Invitrogen™ PCR reagent system, Taq DNA polymerase with (W-1) Invitrogen™ and 10mM dNTP Mix, PCR grade Invitrogen™. The PCR components in the reaction mix included; nuclease free water, 1X Taq DNA polymerase reaction buffer [200 mM Tris-HCl (pH 8.4), 500 mM KCl], 1.5 mM magnesium chloride, 0.2 mM of each of the four dNTPs, 0.05% W1 detergent, 0.5mM of each of the two primers, 0.5-1 U Taq DNA polymerase enzyme and 50-100 ng template DNA depending on the PCR product size and number of alleles. Mineral oil overlay was added to control evaporation of the reaction. A master mix for 50 samples was prepared and placed on ice and the appropriate volumes aliquoted into each reaction tube before adding the DNA template.

The PCR products were separated and analysed using a 4% UltraPure ™ Agarose, Invitrogen™. PCR products were evaluated on 3-4% UltraPure ™ agarose gels depending on fragment size separation requirement. UltraPure ™ Agarose, Invitrogen ™ (3-4) was added in 300 ml of 1X TBE buffer (recommended rate with modification in run time and voltage) at room temperature and stirred to remove clumps. The agarose was melted by heating in the microwave oven till boiling. Ethidium bromide (15 µl) (10 mg/ml) was added after cooling to approximately 60 to 70°C. This was poured in gel casting equipment with 50 well combs and left to solidify, air bubbles were removed manually. Mineral oil was removed by pipeting from PCR products and 1 µl of loading dye (BlueJuice™ gel loading buffer) Invitrogen™ was added to 11 µl of the PCR products including a 100bp DNA ladder Invitrogen™ 1 µg/µl. These were loaded in the wells while the gel was submerged in electrophoresis buffer. Electrophoresis was done at 120 v for 2-4 h to ensure maximum separation of fragments with varying molecular weight. The gel was placed on a UV transilluminator and a photograph taken.

Data analysis

Genotypic and allelic data was recorded after analysis of PCR agarose gels. The data was loaded into analysis software POPGENE version 32 (http://www.ualberta.ca/~fyeh/popgene_download.html) to evaluate intra-population and inter-population diversity on study loci. Analysis of molecular variance and test for HW equilibrium in the populations was done in ARLEQUIN version 3.5 (http://cmpg.unibe.ch/software/arlequin3). Dissimilarity among populations were analysed in DARwin version 5.0.158 (http://darwin.cirad.fr/darwin). Population bottlenecks among populations of wild sorghum growing in three counties in Western
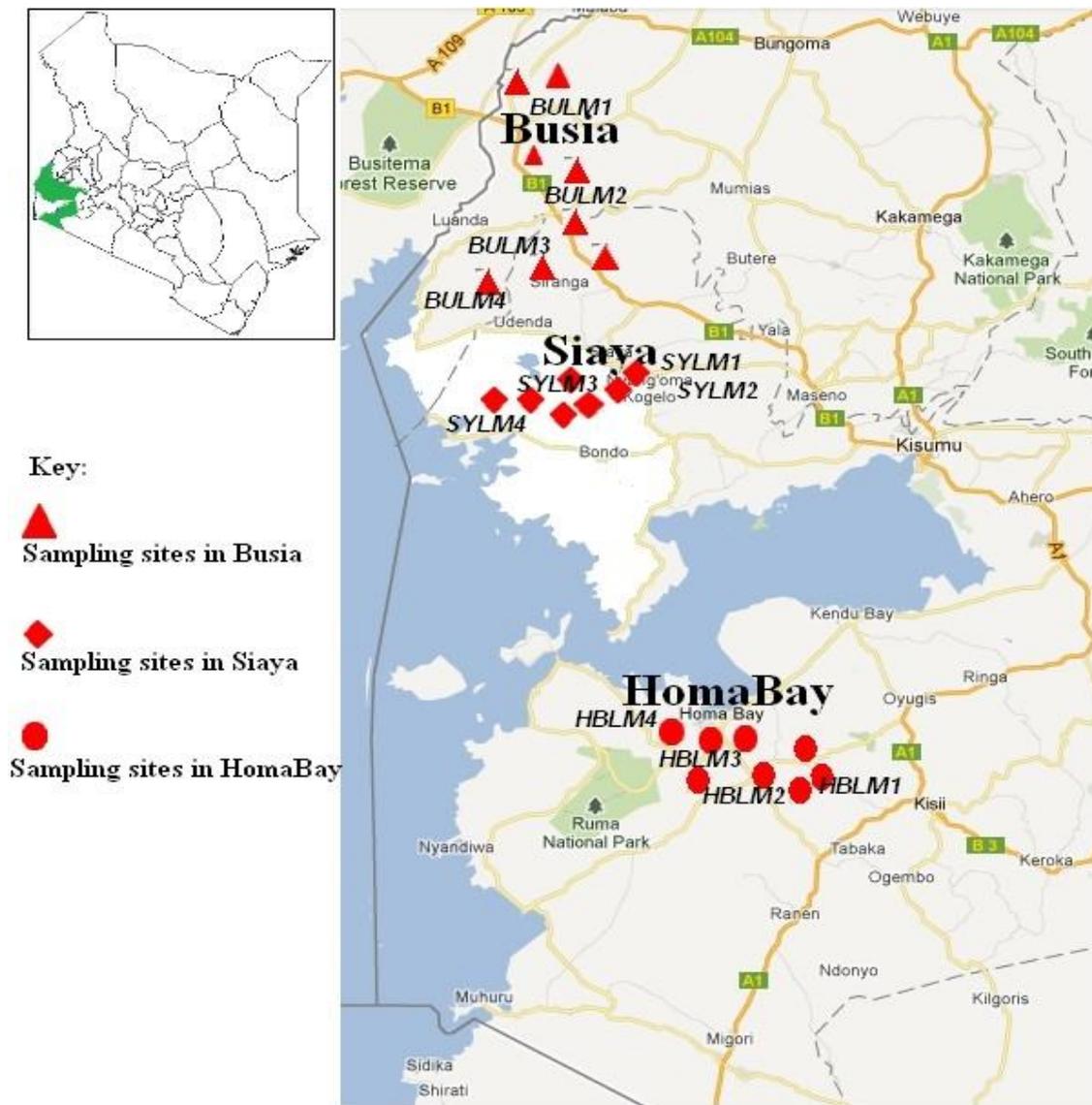
Figure 1. Sampling of wild sorghums in agro-ecological zone characteristics in survey sites in Homabay, Siaya and Busia counties of Kenya.

Kenya were analysed in BOTTLENECK version 1.2.02 (www.montpellier.inra.fr/URLB/bottleneck/bottleneck.html).

Total gene diversity (H_T) was partitioned into inter-population and intra-population components (Nei, 1973) to derive intra-population gene diversity (H_S) and inter-population gene diversity ($D_{ST} = H_T - H_S$) values. Furthermore, the coefficient of gene differentiation (G_{ST}), which represents the relative magnitude of gene differentiation among subpopulations, was calculated as the proportion of inter-populational gene diversity to total gene diversity as ($G_{ST} = D_{ST}/H_T$). Expected heterozygosity and diversity (Mondini et al., 2009) indices were estimated using the ARLEQUIN version 3.5 software (Excoffier et al., 2005). Analysis of molecular variance (AMOVA) was done using ARLEQUIN software to estimate the significance of the covariance components allocated for within individuals, within populations, within groups of populations and among groups using non-parametric permutation procedures (Excoffier et al., 1992). The Ewens-watterson neutrality tests were

performed with ARLEQUIN software to test for the selective neutrality based on Ewens sampling theory in a population at equilibrium (Ewens, 1972).

Allelic variations on all loci, Shannon index on the polymorphic information on each locus and the overall allele frequency were calculated. F-statistics (fixation index -F_{IS}, Index of deviation from HW equilibrium - F_{IT} and the degree of population differentiation F_{ST}) were estimated for multiple populations (Hartl and Clark, 1989; Weir, 1990; Mohammadi and Prasanna, 2003). Test of Hardy Weinburg (HW) equilibrium was done by computing expected genotypic frequencies under random mating using the algorithm by Levene (1949), and performing chi-square (X^2) tests. Probability values were used to determine the populations under HW equilibrium. Nei's unbiased genetic distance and genetic identity was estimated with the POPGENE software version 32 (Yeh and Boyle, 1997; Yang and Yeh, 1993).

The classical linkage disequilibrium coefficient measuring

Table 1. Number of wild sorghums collected from various agro-ecological zone sites in Homabay, Siaya and Busia counties of Kenya.

County	AEZs	Number of samples
Homabay	LM 1	9
	LM 2	8
	LM 3	12
	LM 4	21
	UM 1	11
Siaya	LM 1	11
	LM 2	12
	LM 3	11
	LM 4	19
Busia	LM 1	18
	LM 2	13
	LM 3	16
	LM 4	14
Total		175

Table 2. Agro-ecological zone characteristics in Homabay, Siaya and Busia counties of Kenya where wild sorghums were collected.

Agro-ecological zone	Altitude (m)	Annual mean temperature (^{o}C)	Annual mean rainfall (mm)
Busia			
(+0° 28' 45.07", +34° 7' 11.39")			
Lower midlands (LM)			
LM1	1200-1440	22.2-21.0	1650-2000
LM2	1200-1350	22.3-21.4	1420-1800
LM3	1140-1500	22.7-21.0	1100-1450
LM4	1135-1200	22.7-22.3	900-1100
Homabay			
(-0° 36' 15.25", +34° 29' 57.86")			
Upper midlands (UM)			
UM1	1500-1570	21.0-20.5	1600-1900
Lower Midlands (LM)			
LM1	1300-1500	21.8-20.9	1500-1900
LM2	1200-1350	22.3-21.5	1400-1600
LM3	1140-1250	22.7-22.0	1020-1390
LM4	1135-1200	22.7-22.3	890-1020
LM5	1135-1180	22.7-22.4	-
Siaya			
(+0° 2' 18.43", +34° 12' 19.88")			
Upper midlands (UM)			
UM1	1500-1620	20.5-19.9	1600-1800
UM2-3	1500-1800	20.5-19.3	1450-1600
UM3-4	1500-1700	20.5-19.3	1200-1450

Table 2. Contd.

Agro-ecological zone	Altitude (m)	Annual mean temperature (^{o}C)	Annual mean rainfall (mm)
Lower midlands (LM)			
LM1	1300-1550	21.7-20.4	1600-1800
LM2	1300-1550	21.7-20.4	1300-1600
LM3	1140-1550	22.7-20.4	900-1500
LM4	1140-1450	22.7-20.9	800-1300
LM5	1135-1300	22.7-21.7	700-900

Farm management hand book of Kenya (2005, 2009).

Table 3. SSR loci used to analyze polymophism in the wild sorghum populations.

Locus	Chromosome	Forward primer (5' -> 3')	Reverse primer (5' -> 3')
SB1764	3	CTTGTGCTTGCTTGCACCATATTC	GTCGATGAGGAGCTTCATGCTCAG
SB3420	6	GAGCCAGCATGCATGATAATTGTT	CACAAAGGCATGACAGTCAATCAA
SB4688	9	CTGTAAGCATGATGAAGGTCGTGG	AAGAAGGTGATGACAGGGATGGAG
SB5058	9	GAGAATTGGAAGAAAGCCTCGGTT	CAGAGCTCCTAAACGGTCCTCAAA
SBI000X	10	CTAGAGGATTGCTGGAAGCG	CTGCTCTGCTTGTCGTTGAG
SB858	1	TTTCGTCTTCTTCCACATGACCAC	AGAAGATGGCGACAATCAGAAAGG
SB5293	10	TGAATAATGCACGCAGTAGCGTCT	TATTCCCACGGCTCGCTAGCTACT

deviation from random association between alleles at different loci (Lewontin and Kojima, 1960) D and its standardization by the maximum value it can take (D max), given the allele frequencies (Lewontin, 1964) were estimated in POPGENE. Dendogram for populations from AEZs were based on Nei's genetic distances using UPGMA, which is an adoption of the program NEIGHBOR of PHYLIP version 3.5c.

The Wilcoxon test was applied in the Bottleneck software to test for heterozygosity deficiency under the infinite allele model (IAM), the two–phase model (TPM) and the step wise mutation model (SMM). Estimates of H_{EQ} from different mutation models (IAM, SMM and TPM) and analysis for presence of population bottlenecks were performed in BOTTLENECK software. Dissimilarity matrixes were developed in DARwin software using the DICE procedure for presence or absence data below:

$$dij = \frac{b+c}{2a+(b+c)} \tag{1}$$

Where, *dij* is the dissimilarity between units *i* and *j; a* is the number of variables where *xi* = presence and *xj* = presence*; b* is the number of variables where *xi* = presence and *xj* = absence*; c* is the number of variables where *xi* = absence and *xj* = presence

Axial and radial phylogenetic trees were developed using the weighted neighbourhood joining Saitou and Nei (1987) in DARwin below.

$$\partial_{(S,K)} = \frac{(C_i + C_j)\, \partial_{(i, K)} + (C_j + C_k)\, \partial_{(j, k)} - C_k\, \partial_{(i, j)}}{C_i + C_i + C_k} \tag{2}$$

Where, i, j and k are elements units or groups of units; C_i, C_j and C_k are the unit numbers of these elements; $\partial_{(S,K)}$, weighted average of dissimilarities between k and elements i and j; i, j and k - elements units or groups of units.

RESULTS

Allelic variation for microsatellite loci and allele frequency in populations of wild species

Microsatellite loci SB1764, SB3420 and SB4688 showed polymorphisms within the populations from Homabay, Siaya and Busia counties and in sub-populations clustered along AEZs (UM$_1$, LM$_1$, LM$_2$, LM$_3$, and LM$_4$). Locus SB1764 showed 5 alleles from which 3.27 alleles were effective. A Shannon information index of 1.325 was observed (Table 4). Five alleles were seen from the analysis of locus SB3420 in all populations, where 3.17 out of the five alleles were effective. A Shannon information index of 1.3 was observed for this locus. For locus SB4688, five alleles were observed in the populations but only 2.92 of the alleles were effective, thus a Shannon information content of 1.26 was recorded (Table 4).

PCR amplification of three loci gave bands ranging from 250bp to 520bp in the wild populations from the three Western Kenya counties (Figure 2a, 2b, 2c). Amplifications were obtained on loci SB4688 and SB3420 in wild sorghums collected from Siaya and Homabay. Some wild materials did not amplify on loci SB4688 (lanes 34 to 38, Figure 2a) and SB3420 (lanes 14, 15, 20, 21, 22, 25, 40, 41 and 42, Figure 2b), these were different from the expected alleles in *S. halepense, S. verticilliflorum, S. bicolor* and *S. sudanense* (Figure 2c). A 550bp band from locus SB3420 (Lanes 5, 9, 12, 31 and 32, Figure 2b) demonstrated a *S. verticilliflorum* origin, while 300 and 290 bp from the same locus demonstrated

Table 4. Observed number of alleles and Shannon Index of the for SSR loci assayed.

Locus	Chromosome	Sample size	na	ne	I
SB1764	3	332	5.0000	3.2686	1.3247
SB3420	6	332	5.0000	3.1734	1.2899
SB4688	9	332	5.0000	2.9202	1.2551
SB5058	9	112	3.0000		0.5553
SBI000X	10	64	2.0000		0.2338
SB858	1	124	4.0000		0.9824
SB5293	10	152	3.0000		0.8092

na = Observed number of alleles; ne = effective number of alleles [Kimura and Crow (1964)] (only for primers with above one I); I = Shannon's Information index [Lewontin (1972); Sherwin et al., 2006].
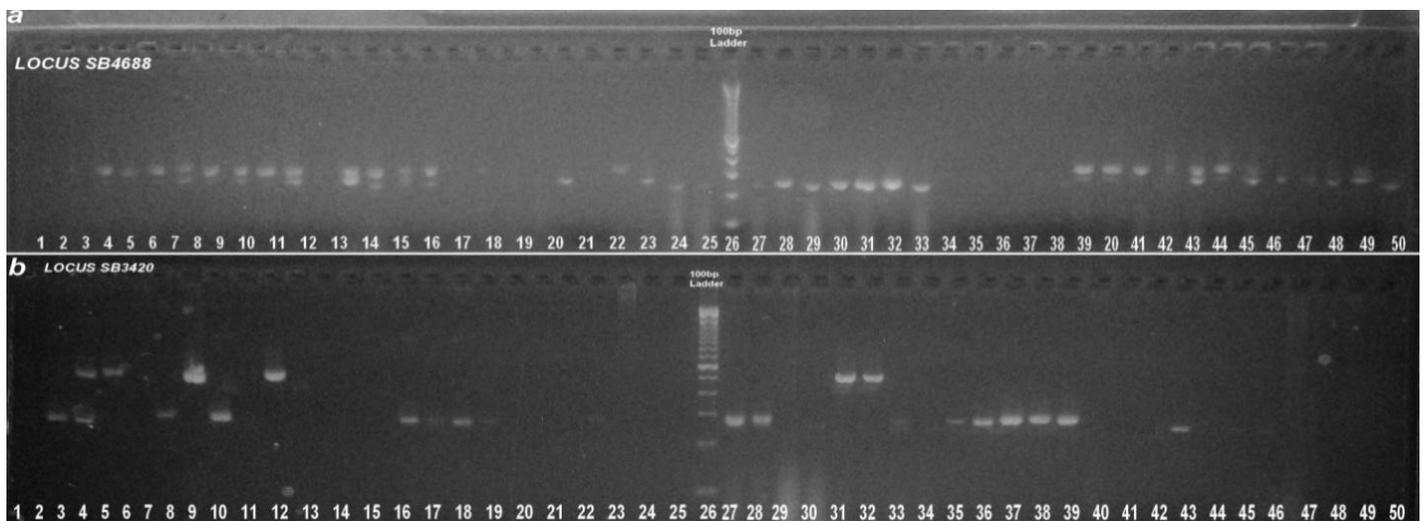
Figure 2. Agarose gel electrophoresis showing alleles 50 representative samples of wild populations from Western Kenya counties and AEZS. **(a)** Locus SB4688, **(b)** locus SB3420. *S. bicolor* is used as a positive control in lane 50 and in lane 3 (a) and (b) respectively; water was used as a negative control and loaded with the 100 bp ladder in lane 26.

Figure 2c. Agarose gel electrophoresis showing expected alleles from sorghum species on loci SB4688 and SB3420.

a *S. bicolor* and *S. sudanense* origin (Figure 2b). Some materials had two alleles showing recent hybridization events between *S. bicolor* and *S. verticilliflorum* on locus SB3420 (Lanes 4, Figure 2b).

Figure 3. Wild species in crop sorghum farms in Western Kenya.

This is a result of their proximity in crop sorghum stands (Figure 3).

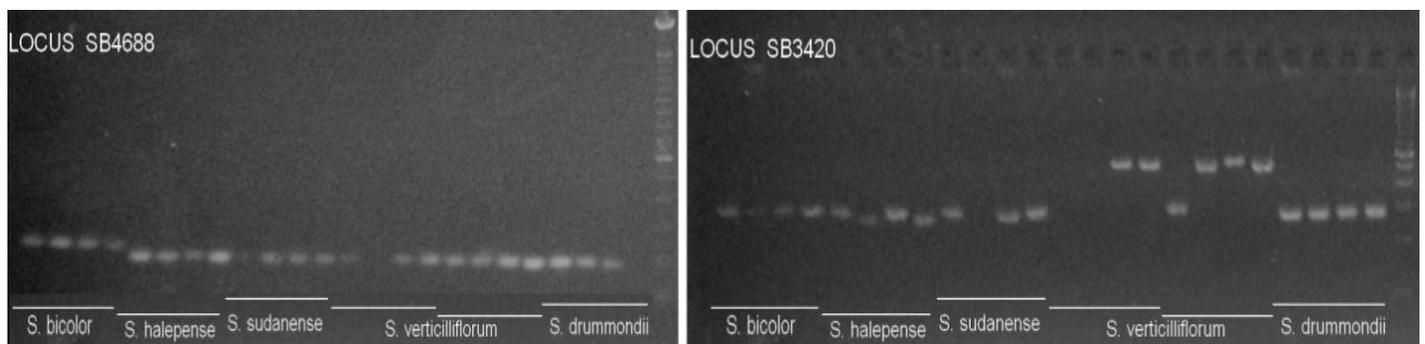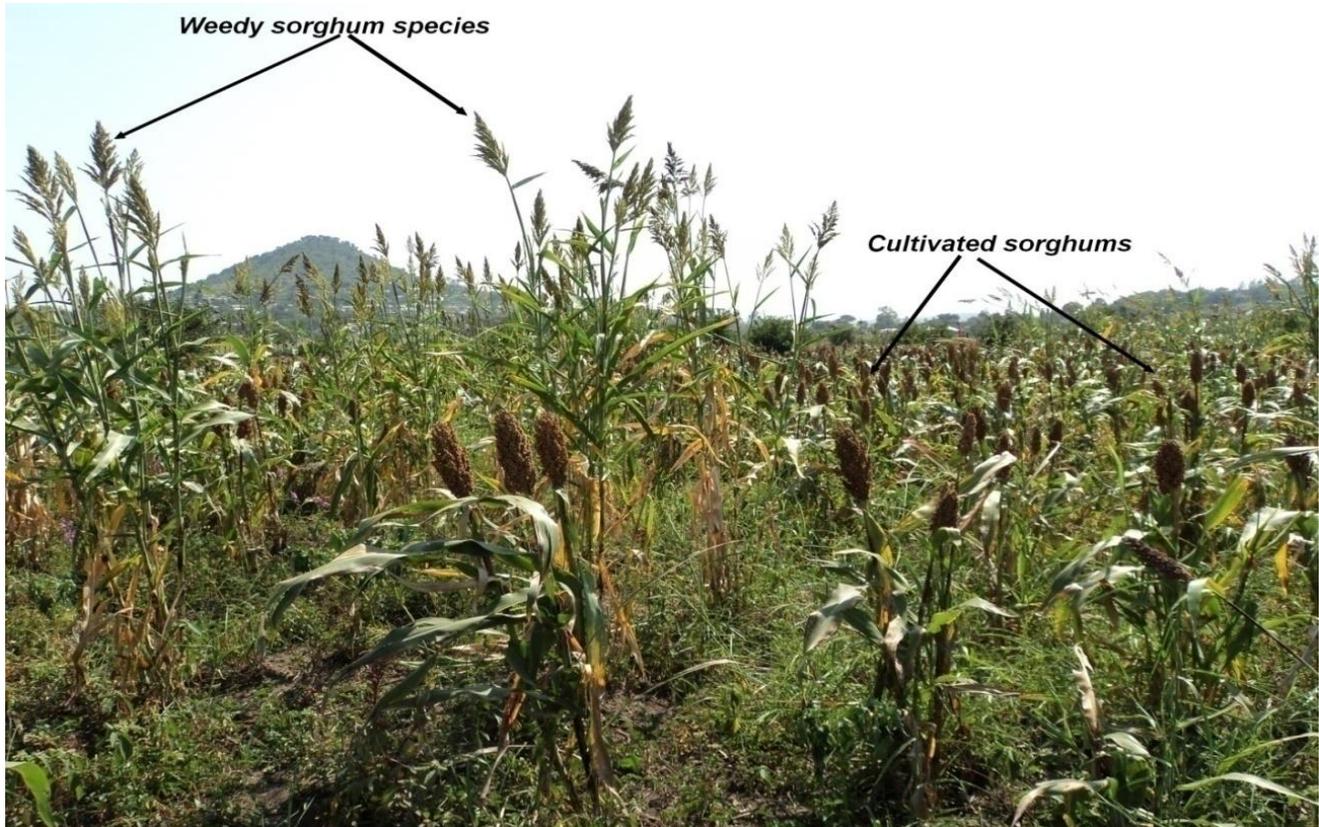Locus SB1764 showed high (above 0.7) frequencies of alleles A and C in $SYLM_3$, $SYLM_4$ and $BULM_4$ (Table 5). Alleles A and C of SB1764 were the most frequent in the population (Table 5). Allele C had medium (0.4 to 0.7) frequencies in all AEZs except in $HBLM_4$, $HBLM_1$, $SYLM_4$, $BULM_1$ and $BULM_2$ (Table 5). Alleles B, D and E of locus SB1764 had low (below 0.4) frequencies in all AEZS (Table 5). Locus SB3420 alleles had medium (0.4 to 0.7) to low (below 0.4) in all populations except in $SYLM_1$ where allele E had a frequency of 0.7. Allele B had medium frequencies (0.4 to 0.7) in 8 AEZ populations except $HBLM_3$, $HBLM_4$, $SYLM_1$, $SYLM_2$ and $BULM_4$. Medium frequencies were also observed on allele D and allele E (Table 5). Locus SB44688 showed high frequencies on allele C in $BULM_4$ and allele E in $SYLM_1$, and $SYLM_3$. Medium frequencies on this locus were seen on alleles B, C, E, F and G. Low frequencies were observed on alleles B, C, F and G. Allele E was the most frequent in the population (Table 5).

Genetic diversity within and among wild sorghum populations from AEZS in Western Kenya

Total heterozygosity was high in AEZs from Busia (0.712),

followed by those from Homabay (0.665) and Siaya (0.564). Observed heterozygosity (H_O) was high (0.762) in $BULM_1$ (1200-1440 masl) wild sorghum groups. However, H_O was low towards the lower AEZs LM_2, LM_3, and LM_4 in Busia, similar trend was seen in Homabay AEZs with high H_O values of 0.62 in $HBLM_1$ and $HBLM_2$ and low values in $HBLM_3$ (0.444) and $HBLM_4$ (0.567) (Table 6). In Siaya all the zones had H_O of less than 0.4 (Table 6).

High expected heterozygosity (H_E) values were observed in Lower midlands in Busia. Population from $BULM_2$ had H_E of 0.715 while $BULM_1$, $BULM_3$ and $BULM_4$ had H_E values of between 0.590 to 0.690. Expected heterozygosity in HBLMs ranged between 0.531 ($HBLM_1$) to 0.683 ($HBLM_2$). Expected heterozygosity was low in AEZs from Siaya (0.453 to 0.557). Expected heterozygosity was greater than the Inter-population gene diversity (D_{ST}) and the proportion of inter-population gene diversity (G_{ST}) in all wild sorghum populations from different AEZS. However the D_{ST} and G_{ST} of $HBLM_1$ (0.134 and 0.202), $SYLM_1$ (0.111 and 0.197) and $BULM_3$ (0.116 and 0.163) had higher values than other AEZS (Table 6).

The degree of inbreeding and therefore heterozygote deficiency (F_{IS}) was high in $HBLM_1$, $HBLM_2$, $HBLM_3$ and $HBLM_4$. $HBUM_1$ had an F_{IS} index of 0.596 which was almost similar to that observed in $SYLM_1$ (0.582) (Table

Table 5. Allele frequency for loci SB1764, SB3420 and SB4688 in the wild sorghum populations obtained from three sorghum growing counties around Lake Victoria in Western Kenya.

Population	HBLM$_1$	HBLM$_2$	HBLM$_3$	HBLM$_4$	HBUM$_1$	SYLM1	SYLM2	SYLM3	SYLM4	BULM$_1$	BULM$_2$	BULM$_3$	BULM$_4$
Locus/Allele													
SB1764/A	0.5625	0.2500	0.1250	0.2250	0.0909	0.1667	0.1818	0.1364	0.7105	0.5357	0.4286	0.2000	0.0455
SB1764/B	0.3125	0.2500	0.2500	0.3250	0.1364	0.0000	0.2727	0.0000	0.2368	0.2500	0.0000	0.1333	0.0455
SB1764/C	0.1250	0.5000	0.6250	0.3500	0.5455	0.6667	0.5000	0.7273	0.0526	0.1071	0.3571	0.5333	0.7273
SB1764/D	0.0000	0.0000	0.0000	0.1000	0.2273	0.0833	0.0455	0.1364	0.0000	0.0000	0.1071	0.1333	0.0909
SB1764/E	0.0000	0.0000	0.0000	0.0000	0.0000	0.0833	0.0000	0.0000	0.0000	0.1071	0.1071	0.0000	0.0909
SB3420/A	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0455	0.0000	0.0526	0.1071	0.1429	0.2000	0.2273
SB3420/B	0.5625	0.4375	0.1667	0.2250	0.5455	0.2917	0.3636	0.4545	0.4474	0.4286	0.5000	0.4333	0.3182
SB3420/D	0.4375	0.3125	0.4167	0.1250	0.0909	0.0000	0.0000	0.0455	0.0000	0.1429	0.2500	0.2333	0.1818
SB3420/E	0.0000	0.2500	0.2917	0.6250	0.3636	0.7083	0.5909	0.3636	0.5000	0.3214	0.1071	0.1333	0.2273
SB3420/F	0.0000	0.0000	0.1250	0.0250	0.0000	0.0000	0.0000	0.1364	0.0000	0.0000	0.2857	0.0000	0.0455
SB4688/B	0.6875	0.4375	0.1667	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
SB4688/C	0.3125	0.2500	0.3333	0.4000	0.4091	0.2500	0.3636	0.2727	0.4211	0.2500	0.2143	0.7667	0.2273
SB4688/E	0.0000	0.3125	0.5000	0.6000	0.5909	0.7500	0.6364	0.7273	0.5263	0.2857	0.2500	0.1333	0.3636
SB4688/F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4286	0.2500	0.0667	0.0909
SB4688/G	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0526	0.0357	0.0000	0.0333	0.3182

HB, Homabay: SY, Siaya; BU, Busia counties. LM, lower midlands; UM, upper midlands; HBLM$_1$, Homabay lower midlands 1.

Table 6. Molecular diversity indices of wild sorghum populations obtained from three sorghum growing counties around lake Victoria in Western Kenya.

Region	Population	Number of individuals	H$_O$	Sd	H$_E$ (H$_S$)	Sd	H$_T$	Sd	F$_{IS}$	D$_{ST}$	G$_{ST}$
Homabay	HBLM$_1$	8	0.625	0.250	0.531	0.075	0.665	0.031	0.19318	0.134	0.202
Homabay	HBLM$_2$	8	0.625	0.125	0.683	0.014	0.665	0.031	0.09091	0.019	0.028
Homabay	HBLM$_3$	12	0.444	0.210	0.640	0.087	0.665	0.031	0.31518	0.025	0.037
Homabay	HBLM$_4$	20	0.567	0.104	0.593	0.123	0.665	0.031	0.04508	0.072	0.108
Homabay	HBUM$_1$	11	0.242	0.189	0.583	0.074	0.665	0.031	0.59596	0.082	0.123
Siaya	SYLM$_1$	12	0.194	0.127	0.453	0.075	0.564	0.098	0.58152	0.111	0.197
Siaya	SYLM$_2$	11	0.364	0.091	0.557	0.100	0.564	0.098	0.35829	0.007	0.013
Siaya	SYLM$_3$	11	0.333	0.189	0.514	0.138	0.564	0.098	0.36232	0.050	0.089
Siaya	SYLM$_4$	19	0.456	0.030	0.523	0.065	0.564	0.098	0.13011	0.042	0.074
Busia	BULM$_1$	14	0.762	0.149	0.684	0.030	0.712	0.010	0.11828	0.028	0.039
Busia	BULM$_2$	14	0.548	0.419	0.715	0.052	0.712	0.010	0.24112	0.003	0.004
Busia	BULM$_3$	15	0.511	0.139	0.596	0.171	0.712	0.010	0.14702	0.116	0.163

Table 6. Contd.

Region	Population	Number of individuals	HO	Sd	HE (HS)	Sd	HT	Sd	FIS	DST	GST
Busia	BULM$_4$	11	0.606	0.292	0.670	0.174	0.712	0.010	0.09910	0.042	0.060
	Total	166*	0.483	0.178	0.595	0.090	0.648	0.045		0.056	0.087

H_O= Observed heterozygosity; sd=standard deviation; H_E (H_S); expected heterozygosity (intra-population gene diversity); H_T; total gene diversity; F_{IS}= heterozygote deficiency/ degree of inbreeding; D_{ST}= inter-population gene diversity (H_T-H_S); G_{ST}= proportion of inter-population gene diversity (D_{ST}/H_T).

Table 7. Expected heterozygosity (H_E) of wild sorghum populations obtained from three sorghum growing counties around lake Victoria in Western Kenya.

Locus#	HBLM$_1$	HBLM$_2$	HBLM$_3$	HBLM$_4$	HBUM$_1$	SYLM$_1$	SYLM$_2$	SYLM$_3$	SYLM$_4$	BULM$_1$	BULM$_2$	BULM$_3$	BULM$_4$	Mean	S.D.	Tot. Het
SB1764	0.608	0.667	0.554	0.729	0.654	0.536	0.671	0.455	0.448	0.651	0.690	0.662	0.472	0.600	0.096	0.608
SB3420	0.525	0.692	0.728	0.556	0.589	0.431	0.515	0.671	0.562	0.706	0.680	0.724	0.797	0.629	0.106	0.525
SB4688	0.458	0.692	0.638	0.492	0.506	0.391	0.485	0.416	0.558	0.696	0.775	0.402	0.740	0.558	0.135	0.458
Mean	0.531	0.683	0.640	0.593	0.583	0.453	0.557	0.514	0.523	0.684	0.715	0.596	0.670	0.595	0.079	0.531
S.D.	0.075	0.014	0.087	0.123	0.074	0.075	0.100	0.138	0.065	0.030	0.052	0.171	0.174	0.090	0.049	0.075

HB, Homabay: SY, Siaya; BU, Busia counties. LM, lower midlands; UM, upper midlands; HBLM$_1$, Homabay lower midlands 1.

6). These F_{IS} values imply low to moderate inbreeding the two populations. Low F_{IS} values were observed in HBLM$_4$ (0.045), SYLM$_4$ (0.130) and BULM$_4$ (0.099) (Table 6), these populations had higher levels of outcrossing.

Expected heterozygosity on each locus varied across populations in the AEZs. Locus SB1764 gave high H_E values in HBLM$_4$ (0.729) (Table 7). Midway H_E values of between 0.69 and 0.455 were observed in the rest of the populations. Locus SB3420 showed higher diversity within HBLM$_3$, BULM$_1$, BULM$_3$ and BULM$_4$ (>0.7). The least H_E value on the locus was observed in SYLM$_1$ (0.431) (Table 7) suggesting low intra-population gene diversity. Locus SB4688 had higher intra-population gene diversity in BULM$_2$ and BULM$_4$ with H_E values of 0.775 and 0.740 respectively. The least H_E value on the locus was observed in the SYLM$_1$ population just as was the

case with locus SB3420 (Table 7).

Analysis of molecular variance (AMOVA) as a weighted average over the three loci SB1764, SB3420, SB4688 showed that 3.2% of the total variation was explained by among group variations, 10.5% by among populations within groups' differences and 16.3% by among individuals within populations. Large variations of 69.97% were explained by differences within individuals on the three loci (Table 8).

Population specific indices in wild sorghum populations from AEZS in Western Kenya

Wild sorghum from BULM$_2$ showed high degree of inbreeding (F_{IS} of 0.9) on locus SB1764, while loci SB3420 had high inbreeding in HBUM$_1$ (0.7) and SYLM$_1$ (0.81). Loci SB4688 had high inbreeding in

HBUM$_1$ (0.83) (Table 9). Midway heterozygosity deficiency (F_{IS}) values were observed on locus Sb1764 in HBLM$_1$ (0.4) HBLM$_3$ (0.56) SYLM$_3$ (0.41) and BULM$_4$ (0.43) (Table 9). Similar values were observed on locus SB3420 in SYLM$_2$ (0.48) and BULM$_3$ (0.46) Similar inbreeding degrees were seen on locus SB4688 in SYLM$_1$ (0.59) and SYLM$_3$ (0.57). Low values of less than 0.4 were observed on locus SB1764 in AEZs in Homabay, Siaya and Busia counties. However negative F_{IS} valves were more present on loci SB3420 and SB4688 in sorghum species obtained from HBLM$_1$, HBLM$_2$, HBLM$_4$, BULM$_2$, BULM$_3$ and BULM$_4$ (Table 9). This shows differential heterozygosity deficiency coefficients on the three loci in wild sorghum population with higher affinity for fixation on loci SB1764. On a whole population basis of wild sorghums from Western Kenya, the degree of inbreeding (F_{IS}) values were generally

Table 8. AMOVA showing percentage of variation on each of the three loci assayed.

Source of variation	Locus SB1764 Percentage of variation	Locus SB3420 Percentage of variation	Locus SB4688 Percentage of variation	3 loci Percentage of variation
Among groups	-3.14523	5.81122	7.01745	3.20
Among populations within groups	15.81786	4.81671	10.98644	10.52
Among individuals within populations	28.66450	8.02441	12.18045	16.30
Within individuals	58.66288	81.34767	69.81566	69.97
	100%	100%	100%	100%

Table 9. Population specific F_{IS} (heterozygote deficiency) indices per polymorphic locus (absolute values) in wild sorghums from AEZs in Western Kenya

LOCUS	$HBLM_1$	$HBLM_2$	$HBLM_3$	$HBLM_4$	$HBUM_1$	$SYLM_1$	$SYLM_2$	$SYLM_3$	$SYLM_4$	$BULM_1$	$BULM_2$	$BULM_3$	$BULM_4$
SB1764	0.400	0.263	0.560	0.389	0.315	0.389	0.333	0.412	-0.059	0.013	0.900	-0.007	0.434
SB3420	-0.750	-0.091	0.088	-0.173	0.701	0.814	0.483	0.195	0.161	-0.331	-0.273	0.456	-0.029
SB4688	-0.400	0.103	0.357	-0.226	0.828	0.585	0.259	0.574	0.250	-0.028	0.081	-0.167	0.018
Mean	0.193	0.091	0.315	0.045	0.596	0.582	0.358	0.362	0.130	0.118	0.241	0.147	0.099

HB - Homabay: SY - Siaya and BU - Busia counties. LM -Lower Midlands and UM-Upper Midlands. $HBLM_1$ -Homabay Lower Midlands 1.

Table 10. Whole population degree of inbreeding (F_{IS}), deviation from Hardy Weinberg (HW) equilibrium (F_{IT}), degree of population differentiation (F_{ST}) and estimated gene flow, based on loci SB1764, SB3420 and SB4688.

Locus	F statistics			
	F_{IS}	F_{IT}	F_{ST}	Nm*
SB1764	0.3015	0.4147	0.1620	1.2935
SB3420	0.0336	0.1521	0.1226	1.7887
SB4688	0.1196	0.3022	0.2074	0.9555
Mean	0.1514	0.2901	0.1634	1.2800

*Nm = Gene flow estimated from Fst = 0.25(1 - Fst)/Fst.

low on all loci giving a population mean of 0.1514. Locus SB1764 showed F_{IS} index of 0.3015, locus SB3420 had 0.0336 while locus SB4688 had an index of 0.1196 on all collections from Homabay, Siaya and Busia AEZs (Table 10). The whole population leaned towards the Hardy- Weinberg (HW) equilibrium with F_{IT} index (deviation from Hardy Weinberg (HW) equilibrium (=1-H/H_E) of (0.2901). Attainment of the HW equilibrium was observed more on Locus SB3420 with F_{IT} of 0.1521. Locus SB4688 had an index of 0.3022, while deviation from the HW equilibrium was most visible on locus SB1764 with an F_{IT} index of 0.4147 (Table 10).

Inter-population differences were most observed

Table 11. χ^2 Test of HW equilibrium within wild sorghum populations in AEZs in Homabay, Siaya and Busia counties by estimation of expected genotypic frequencies under random mating (Levene, 1949).

Locus	HBLM$_1$ χ^2	P	HBLM$_2$ χ^2	P	HBLM$_3$ χ^2	P	HBLM$_4$ χ^2	P	HBUM$_1$ χ^2	P
SB1764	15.250	0.002	15.571	0.001	10.489	0.015	43.500	0.000	8.149	0.227
SB3420	4.083	0.043	3.357	0.340	12.250	0.057	3.617	0.728	25.580	0.000
SB4688	1.273	0.259	7.714	0.052	4.049	0.256	1.012	0.314	8.205	0.004
	SYLM$_1$ χ^2	P	SYLM$_2$ χ^2	P	SYLM$_3$ χ^2	P	SYLM$_4$ χ^2	P	BULM$_1$ χ^2	P
SB1764	46.650	0.000	9.380	0.153	8.658	0.034	38.872	0.000	20.114	0.003
SB3420	8.874	0.003	3.173	0.366	8.767	0.187	3.108	0.375	9.313	0.157
SB4688	4.606	0.032	0.769	0.380	4.050	0.044	4.263	0.234	6.400	0.380
	BULM$_2$ χ^2	P	BULM$_3$ χ^2	P	BULM$_4$ χ^2	P				
SB1764	36.727	0.000	18.833	0.004	42.200	0.000				
SB3420	5.646	0.464	21.202	0.002	15.400	0.118				
SB4688	16.857	0.010	1.162	0.979	10.250	0.115				

on locus SB4688 with an F_{ST} index of 0.2074. Locus SB1764 had an F_{ST} index of (0.1620) while locus SB3420 had an index of 0.1226). All three loci gave a degree of population differentiation of 0.1634 (Table 10). Gene flow (Nm) estimated from the degree of population differentiation (F_{ST}) had a whole population mean of 1.28. Gene flow was most observed on locus SB3420, moderate on locus SB1764 (1.294) and least on locus SB4688 (0.96) (Table 10).

Population equilibrium within wild sorghum populations in agro-ecological zones in in Western Kenya

Homogeneity tests of gene frequencies in wild sorghums across counties and AEZs within counties using χ^2 tests demonstrated that the wild sorghum populations were within the HW equilibrium (Table 11). This equilibrium was true for all loci (SB1764, SB3420 and SB4688). χ^2 computation for wild sorghum from lower midlands in Homabay, Siaya and Busia showed significant adherence for the HW equilibrium. Wild sorghums from HBLM$_4$, SYLM$_2$, SYLM$_4$, BULM$_1$ and BULM$_4$ did not adhere to the HW equilibrium on more than two loci, as shown by the high non-significant P values (Table 11).

Population bottlenecks in wild sorghum populations in the different Agro-ecological zones in Homabay, Siaya and Busia counties

There was no evidence to suggest existence of population bottleneck on wild sorghums growing in AEZs in Homabay, Siaya and Busia. The Wilcoxon test for heterozygosity deficiency under the infinite allele model (IAM) had a probability of 1.00, under the two–phase model (TPM) P was 1.00 and under the step wise mutation model (SMM) P was 1.00 showing no significant pattern. Heterozygosity excess was not significantly different under the IAM, TPM and SMM model (p=0.0625). A similar situation was seen when both heterozygosity excess and deficiency was tested (p=0.125) (Table 12a).

The sign test showed that the expected number of loci with heterozygosity excess under IAM was 1.68 under TPM was 1.74 and under SMM was 1.82. No loci were observed with heterozygosity deficiency but none had significant heterozygosity excess under all models (Table 12a). Despite lack of significant heterozygosity excess or deficiency

Table 12a. Sign and Wilcoxon tests for heterozygosity excess and deficiency under the IAM, SMM and TPM models.

Model	Sign test				Wilcoxon test		
	Expect number of loci with heterozygosity excess	loci with heterozygosity deficiency	loci with heterozygosity excess	p	Probability heterozygosity deficiency	Probability heterozygosity excess	Probability heterozygosity excess and deficiency
I.A.M	1.68	0	3	0.17553	1.00000	0.06250	0.12500
T.P.M	1.74	0	3	0.19409	1.00000	0.06250	0.12500
S.M.M	1.82	0	3	0.22393	1.00000	0.06250	0.12500

I.A.M, infinite allele model; T.P.M, two-phase model; S.M.M, stepwise mutation model.

Table 12b. Comparison between (Ho) Heterozygosity (He) Heterozygosity and the expected values under the IAM, SMM and TPM models.

Locus	Heterozygosity		Population bottlenecks					
	H_O	H_E	I.A.M. H_{EQ}	Probability	S.M.M. H_{EQ}	Probability	T.P.M. H_{EQ}	Probability
SB1764	0.3750	0.6083	0.432	0.0430	0.648	0.3460	0.543	0.0960
SB3420	0.8750	0.5250	0.428	0.0450	0.647	0.3880	0.538	0.1090
SB4688	0.6250	0.4583	0.434	0.1070	0.641	0.4810	0.540	0.2020
Mean	0.6250	0.5306						

H_O, Observed heterozygosity; H_E, expected heterozygosity; H_{EQ}, heterozygosity under mutation-drift equilibrium.

pattern in the models applied to explain bottlenecks in the populations there was variation in the calculated heterozygosity values. This was true when compared to the heterozygosity under mutation - drift equilibrium (H_{EQ}) values in IAM, TPM and SMM models (Table 12b). On locus SB1764 H_E was 0.6083 but H_{EQ} was computed at 0.432, 0.648 and 0.543 under IAM, SMM and TPM model (Table 12b). Locus SB3420 had H_E of 0.5250 but under IAM, SMM and TPM H_{EQ} values of 0.428, 0.647 had 0.538 respectively (Table 12b). Locus SB4688 had H_E of 0.4583 and H_{EQ} 0.434, 0.641 and 0.540 values under IAM, SMM and TPM respectively.

Genetic distance and genetic identity among wild sorghum populations from agro-ecological zones in Homabay, Siaya and Busia counties

Wild sorghum population within AEZS in three Western Kenya counties (Homabay, Siaya and Busia) showed significant levels of genetic identity and distance (Table 13). Population $HBLM_1$ was less identical to other population. It was most distant to $SYLM_1$ (1.3148) $SYLM_3$ (1.028), $BULM_4$ (0.9838), $SYLM_2$ (0.9017), $HBLM_4$ (0.8464) and $HBUM_1$ (0.791) (Table 13). Moderate genetic distances were observed between $HBLM_1$ and

$HBLM_3$ (0.5727), $SYLM_4$ (0.5228) and Busia AEZs. Similar midway genetic distance values were obtained between $SYLM_1$ and $BULM_1$ (0.4729), $BULM_2$ (0.4357) and $BULM_3$ (0.4563). Population from $SYLM_4$ and $BULM_1$ showed moderate to high genetic distances 0.6316 and 0.5314 respectively. Low genetic distances were observed between population from HBUM and $SYLM_1$ (0.0771), $SYLM_2$ (0.0325) and $SYLM_3$ (0.0173). Similar low values were observed between populations $HBLM_4$ and $HBUM_1$ (0.0906); $HBLM_4$ and $SYLM_2$ (0.0039): $HBLM_2$ and $HBLM_3$ (0.0701) (Table 13).

Population from Homabay, Siaya and Busia

Table 13. Nei's unbiased measures of genetic identity and genetic distance among wild sorghum populations from AEZs in Homabay, Siaya and Busia counties

Pop ID	HBLM$_1$	HBLM$_2$	HBLM$_3$	HBLM$_4$	HBUM$_1$	SYLM$_1$	SYLM$_2$	SYLM$_3$	SYLM$_4$	BULM$_1$	BULM$_2$	BULM$_3$	BULM$_4$
HBLM$_1$	****	0.8564	0.5640	0.4290	0.4634	0.2685	0.4059	0.3577	0.5928	0.6462	0.6509	0.5988	0.3739
HBLM$_2$	0.1550	****	0.9323	0.7948	0.8505	0.7496	0.8221	0.8091	0.7024	0.7142	0.7777	0.7827	0.7883
HBLM$_3$	0.5727	0.0701	****	0.8777	0.8752	0.8414	0.8771	0.8839	0.6307	0.6163	0.6915	0.7818	0.8622
HBLM$_4$	0.8464	0.2296	0.1304	****	0.9134	0.9278	0.9962	0.8674	0.8571	0.7462	0.6436	0.7120	0.7349
HBUM$_1$	0.7691	0.1619	0.1332	0.0906	****	0.9258	0.9680	0.9828	0.7552	0.6881	0.7706	0.8338	0.8707
SYLM$_1$	1.3148	0.2882	0.1727	0.0750	0.0771	****	0.9752	0.9637	0.7305	0.6232	0.6468	0.6336	0.8285
SYLM$_2$	0.9017	0.1959	0.1312	0.0039	0.0325	0.0251	****	0.9437	0.8367	0.7329	0.6848	0.7393	0.8156
SYLM$_3$	1.0280	0.2118	0.1234	0.1422	0.0173	0.0369	0.0580	****	0.6907	0.6143	0.7361	0.7307	0.8957
SYLM$_4$	0.5228	0.3532	0.4610	0.1542	0.2808	0.3140	0.1783	0.3701	****	0.8871	0.7398	0.6458	0.5317
BULM$_1$	0.4367	0.3366	0.4840	0.2927	0.3738	0.4729	0.3108	0.4873	0.1198	****	0.8908	0.6697	0.5878
BULM$_2$	0.4295	0.2514	0.3688	0.4408	0.2605	0.4357	0.3786	0.3064	0.3014	0.1156	****	0.7911	0.8361
BULM$_3$	0.5128	0.2451	0.2462	0.3397	0.1817	0.4563	0.3021	0.3138	0.4373	0.4009	0.2343	****	0.8020
BULM$_4$	0.9838	0.2378	0.1482	0.3080	0.1384	0.1881	0.2038	0.1102	0.6316	0.5314	0.1790	0.2206	****

Nei's genetic identity (above diagonal) and genetic distance (below diagonal).

made significant clusters in dendogram analysis. Moderate to high bootstrap values (30 to 83) were observed on county branches (Figure 4). Populations from Siaya clustered away from Busia and Homabay, however, three sub-clusters were observed within the major cluster (Figure 4). Populations from Busia clustered away from the population from Homabay. Less minor clusters were observed in Busia than those that were seen in Homabay. Some populations from Siaya were found to cluster with those from Homabay. A similar situation was observed with populations from Busia clustering with those from Siaya and Homabay (Figure 4). Populations sampled from AEZs in the three counties showed a trend on the distribution of wild sorghum genotypes in given ecological zonation (Figure 5). Wild sorghums from LM$_4$ zones made clusters with wild sorghums from other counties. SYLM$_4$ clustered with populations from lower midlands from Busia. BULM$_4$ clustered with SYLMs while HBLM$_4$ clustered with SYLMs. Wild sorghums from each of the counties clustered together in most instances irrespective of their AEZ of origin and showed no significant pattern on the distribution (Figure 5).

DISCUSSION

Crop alleles were observed in the wild sorghum populations obtained from sorghum growing Western Kenya counties around Lake Victoria. Crop allele frequencies varied among the loci studied and in the AEZs from where the material was sampled. Low (<0.4), moderate (0.4-0.7) and high (0.7) crop allele frequencies were observed in this study. The presence of crop alleles at varying frequencies in wild forms could be directly attributed to the wild sorghum sympatric growth patterns and cultivation of crop sorghum. Localized sympatry was observed in all regions of growth, where wild sorghums were seen growing in crop stands as weeds, on boarder rows, hedges and road side reserves. Tillering in sorghums (especially wild genotypes) extended their flowering period and increased the chances of interspecific hybridization. The climatic conditions in different geographic locations and the rainfall pattern encourage synchrony in germination and flowering and possibly interspecific hydridization. Previous studies in sorghum have shown the existence of crop alleles in populations of weed sorghums (Morrell et al., 2005). Proximity of wild sorghums to crop sorghums has been shown to result in both crop to weed and weed to crop interspecific hybridization (Warwick et al., 2009; Arriola and Ellstrand, 1996; Sahoo and Schmidt, 2010). The persistence of crop alleles in wild populations would be attributed to their weedy nature in given environments (Paterson et al., 1995).

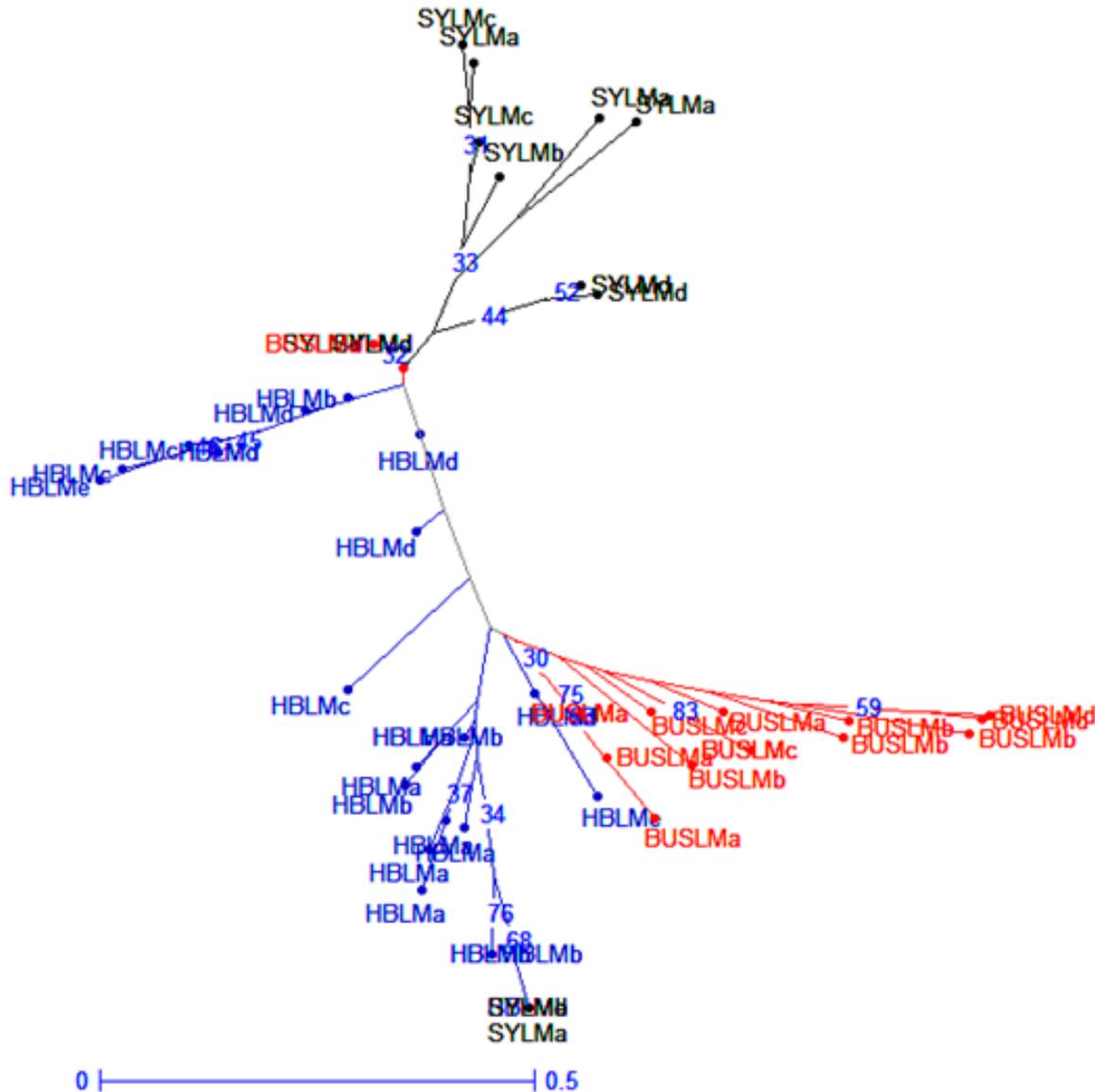Wild sorghum population had moderate to high

Figure 4. Dendogram showing clusters of populations from Homabay, Siaya and Busia counties of Western Kenya.

diversity on the SSR loci assayed from crop sorghum producing AEZs in Western Kenya. H_E values of between 0.453 in LM_1 to 0.715 in LM_2 were obtained in the wild populations. No distinct patterns were observed on the magnitude of diversity of wild sorghums based on AEZs. There was minimal differences in the magnitude of diversity among the counties (Busia H_E = 0.59 - 0.71; Homabay H_E = 0.58-0.68, Siaya H_E = 0.45-0.59). Weed control and seed dissemination agricultural practices by local human populations may have influenced the diversity of the weeds. Such practices may have had low input on the diversity of wild sorghums in Busia and most in Siaya. High diversity in wild sorghums was also observed in other regions of Kenya H_E = 0.69 (Mutegi et al., 2011), in Africa H_E = 0.59 (Casa et al., 2005) in

Cameroon (Barnaud et al., 2007) and in Mali and Guinea (Sagnard et al., 2011).

Wild sorghum populations exhibit intraspecific and interspecific hybridization. Intraspecific hybridization results in low heterozygosity values in populations while interspecific hybridization explains gene flow within wild populations and thus the evolution and diversity of biotypes. Several intermediate types may also be observed in populations, which may play the role of "bridge species". However, interspecific hybridization seems to increase variability of traits per species as a result of disruptive selection in wild sorghums (Doggett and Majisu, 1968).

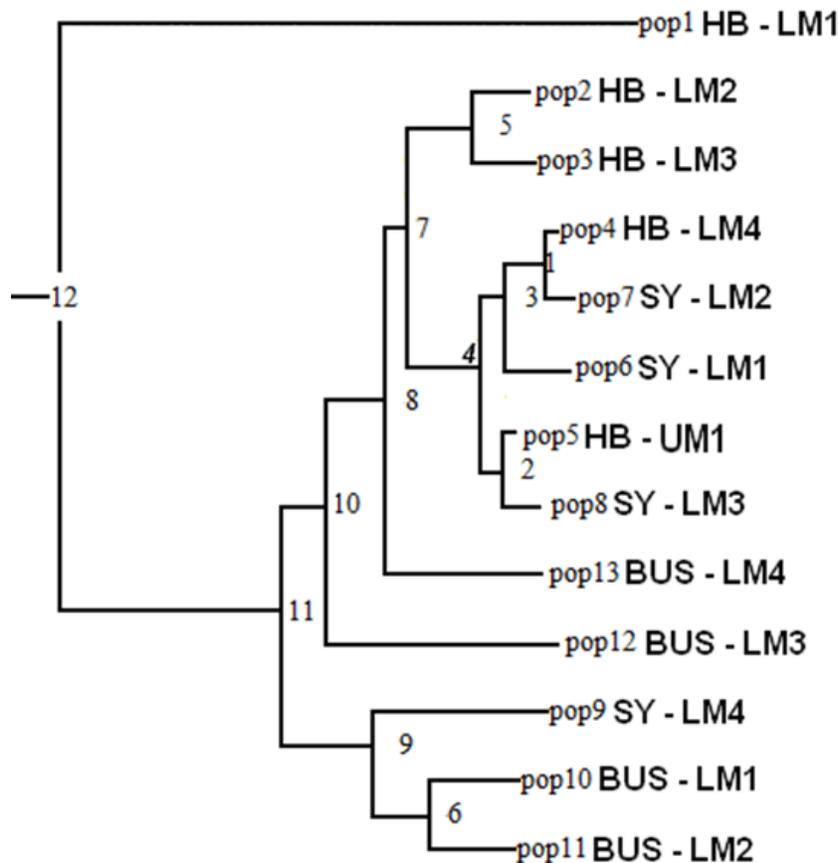Wild sorghums in Western Kenya were maintained at HW equilibrium due to low inbreeding and high

Figure 5. Cluster of populations from study AEZs. Dendrogram Based Nei's (1978) Genetic distance: Method = UPGMA, Modified from neighbor procedure of PHYLIP Version 3.5, in popgene.

heterozygosity. The degree of inbreeding (F_{IS}) was low in most AEZs except in UM_I from Homabay and LM_1 from Siaya. These two regions also had fairly low heterozygosity (H_E). F_{IS} and the degree of genetic differentiation of the population (F_{ST}) have substantial impact on the HW equilibrium at whole population level. Low F_{IS}, F_{IT} and F_{ST} values were observed at a whole population basis, indicating that the population was at HW equilibrium. However, populations from Homabay UM_1 and Siaya LM_1 had moderate F_{IS} values of 0.59 and 0.58 respectively and did not conform to the HW equilibrium. The deviation was probably due to farmer selection practices during weeding. Farmers might have allowed some species of wild sorghums to grow to maturity in or around sorghum fields. This may have a similar effect to inbreeding in wild sorghum population. However, on a whole population basis, inbreeding was low and heterozygosity was high. Previous results show that populations of crop and wild sorghum were heterozygous with low F_{IS} in other parts of Kenya (Mutegi et al., 2011) and in other parts of Africa (Barnaud et al., 2007, Sagnard et al., 2011). Allele frequencies within populations of wild sorghums were high on the loci

assayed. This indicates the presence of gene flow among wild sorghums. Intra-population diversity ($H_S=H_E$) was moderate to high in these wild sorghum populations signifying existence of pollen or seed mediated outbreeding and geneflow among the wild sorghums. Thus, there is potential for proliferation and maintenance of exotic crop genes in these wild populations. In addition, low F_{IS} values were observed in the populations showing low inbreeding. Partitioning of variances shows that variation within individuals in populations was higher than variations among populations, confirming diminished inbreeding within individuals in populations. This also indicates that pollen and seed mediated gene flow was important in the wild sorghum populations. Furthermore, the intra-population diversity (H_S) was larger than the inter-population diversity (D_{ST}) in all populations, thus allelic differences between populations was not huge. The flow of genes from population to population could be accomplished by both pollen flow and seed distribution. This is important due to the movement towards adopting transgenics in cropping system in the near future. If this happened there would be a risk of unintentional escape of transgenes into wild populations thereby boosting their

adaptive advantage and therefore their wild character (Warwick et al., 2009).

The wild sorghum populations did not have significant patterns to explain recent population bottlenecks. Heterozygosity under mutation drift equilibrium (H_{EQ}) varied under IAM, TPM and SMM models. The H_{EQ} estimated by the mutation models show lower values probably due to selective weeding in sorghum fields and the presence of sorghum wild volunteers. Bird preference for the larger seeded weedy sorghums may have had some impact on reducing the estimated H_{EQ}. Growth of small-scale agriculture in the regions around Lake Victoria may have led to heavy weeding of some wild sorghum species to extinction and therefore contributed to the lower H_{EQ} values estimated in the mutation models. Previous studies in sorghum growing regions of Ethiopia did not show recent bottlenecks in sorghum populations (Adugna and Bekele, 2015).

Phylogenetic analysis of wild sorghums showed that counties and AEZs clustered in different ways. The dendograms from the counties had consistent clusters in contrast to those from the various AEZs that did not show a consistent pattern. For instance in Figure 4, genetic material from one county clustered away from material from other counties. However, in Figure 5, genetic material from specific AEZs did not cluster together, indicating that wild sorghums were not clearly segregated along AEZ boundaries in Western Kenya. The presence of consistent geographic county clusters and lack of consistent AEZs clusters indicate higher influence of human population activities rather than climatic conditions on the distribution and diversity of wild sorghums. This could be attributed to practices such as seed contamination with specific weed seed during harvest and also due to selective weeding of the non crop sorghums. Furthermore, selective maintenance of certain non-crop sorghums on hedges and sharing of contaminated seed with specific wild sorghum seeds could be important in the distribution and diversity of wild sorghums. These practices enhance the populations of given wild sorghums within certain geographic locations. Selective weeding and seed distribution practices seem to differ from county to county, culminating in differences in diversity of weed types present. The influence of strong selection by man has been shown to influence sorghum phylogeographic position in Africa (de Alencar Figueiredo et al., 2008). The diversification trend of the wild types found in Western Kenya seem to follow the diversification pattern of crop sorghum which is strongly linked to geographic biotype / race classifications (Deu et al., 1995; 2006, Dje et al., 2000; Barnaud et al., 2007).

Conclusions

Crop alleles were observed in wild, sorghum populations. In addition wild sorghum population had moderate to high diversity on SSR loci assayed. H_E values of between 0.453 in LM_1 to 0.715 in LM_2 were obtained in the wild populations. The wild populations had low inbreeding, low genetic differentiation and low to moderate deviation from HW equilibrium. Intra-populations diversity (H_S) was larger than inter-population diversity (D_{ST}) in all wild populations. This shows pollen and seed mediated gene flow is important in the wild sorghum populations. Analysis of H_{EQ} values under IAM, TPM and SMM mutation models suggest absence of recent bottlenecks in the wild populations. Human influence maybe more important than climatic conditions in explaining the distribution and diversity of wild sorghums populations. Future concerns over the persistence of robust crop alleles including transgenes in wild sorghums will need to be addressed. There is need to expand the capacity for testing and evaluation of the presence and effect of crop genes and in wild sorghums growing around crop sorghum production regions.

Conflict of interests

The authors did not declare any conflict of interest.

REFERENCES

Adugna A, Bekele E (2015). Assessment of recent bottlenecks and estimation of effective population size in the Ethiopian wild sorghum using SSR allele diversity and mutation models. Plant Genet. Resour. DOI:10.1017/S1479262115000015.

Adugna A, Snow AA, Sweeney PM, Bekele E, Mutegi E (2013) Population genetic structure of in situ wild *Sorghum bicolor* in its Ethiopian center of origin based on SSR markers. Genet. Resour. Crop Evol. 60: 1313-1328.

Arriola PE, Ellstrand NC (1996). Crop-to-weed gene flow in the genus sorghum (Poaceae): Spontaneous interspecific hybridization between johnsongrass, *Sorghum halepense*, and crop sorghum, *S. bicolor.* Am. J. Bot. 83:1153–1160.

Barnaud A, Deu M, Garine E, McKey D, Joly HI (2007). Local genetic diversity of sorghum in a village in northern Cameroon: structure and dynamics of landraces. Theor. Appl. Genet. 114(2):237-48.

Barnaud A, Trigueros G, Garine E, McKey D, Joly HI (2008). High outcrossing rates in fields with mixed sorghum landraces: How are landraces maintained? Heredity 101: 445-452.

Casa HM, 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(1):23-30.

Cornuet JM, Luikart G (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001-2014.

de Alencar Figueiredo LF, Calatayud C, Dupuits C, Billot C, Rami JF,

Brunel D, Perrier X, Courtois B, Deu M, Glaszmann JC (2008). Phylogeographic evidence of crop neodiversity in sorghum. Genetics 179(2): 997–1008.

Delobel C, Moens W, Querci M, Mazzara M, Cordeil S, Van den Eede G (2007). Maize seeds sampling and DNA extraction; report on the validation of a DNA extraction method from maize seeds and grains. European commission, Directorate General- joint research centre institute for health and consumer protection (IHCP). Biotechnology and GMOs unit. Website www.gmo-crl.jrc.ec.europa.eu/summaries/3272_DNAExtrReport.pdf Date: 10-1-2012.

Deu M, Hamo P, Chantereau J, Dufour P, D'Hont A, Lanaud C (1995). Mitochondrial DNA diversity in wild and cultivated sorghum. Genome 38:635–645.

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994). Mutational processes of simple sequence repeat loci in human populations. Proc. Natl. Acad. Sci. USA 91:3166–3170.

Doebley JF (2004). The genetics of maize evolution. Annu. Rev. Genet. 38:37-59.

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

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

Doyle JF, Doyle JL (1990). Isolation of plant DNA from fresh tissue. Focus 12:13-15.

Ejeta G, Grenier C (2005). Sorghum and its weedy hybrids, In: Crop ferality and Volunteerism, edited by Jonathan Gressel, CRC Taylor and Francis, London.

Ewens WJ (1972). The sampling theory of selectively neutral alleles. Theor. Popul. Biol. 3:87-112.

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

Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491.

Gepts P, Papa R (2003). Possible effects of (trans) gene flow from crops on the genetic diversity from landraces and wild relatives. Environ. Biosafety Res. 2:89-103.

Goodman SJ (1998). Patterns of extensive genetic differentiation and variation among European harbor seals (*Phoca vitulina* vitulina) revealed using microsatellite DNA polymorphisms. Mol. Biol. Evol. 15(2):104-118.

Hartl DL, Clark AG (1989). Principles of Population Genetics. Sinauer Associates, Sunderland, MA.

Hokanson KE, Ellstrand NC, Ouedraogo JT, Olweny PA, Schaal BA, Raybould AF (2010). Biofortified sorghum in Africa: using problem formulation to inform risk assessment. Nat. Biotechnol. 28(9):900-903.

James C (2010). Global Status of Commercialized Biotech/GM Crops: 2010. ISAAA Brief No. 42. ISAAA: Ithaca, NY.

Levene H (1949). On a matching problem arising in genetics. Ann. Math. Stat. 20:91-94.

Lewontin RC (1964). The interaction of selection and linkage. I. General considerations: Heterotic models. Genetics 49:49-67.

Lewontin RC (1972). The apportionment of human diversity. Evolut. Biol. 6:381-398.

Lewontin RC, Kojima K (1960). The evolutionary dynamics of complex polymorphisms. Evolution 14:458-472.

Mohammadi SA, Prasanna BM (2003). Review interpretation analysis of genetic diversity in crop plants salient statistical tools. Crop Sci. 43:1235-1248.

Mondini L, Noorani A, Pagnotta MA (2009). Assessing Plant Genetic Diversity by Molecular Tools. Diversity 1(1):19-35.

Morrell PL, Williams-coplin D, Lattu AL, Bowers EJ, Chandler JM, Paterson AH (2005). Crop-to-weed introression has impacted allelic composition of Johnson grass populations with and without recent exposure to cultivated sorghum. Mol. Ecol. 14:2143-2154.

Muraya MM, Sagnard F, Parzies HK (2010). Investigation of recent population bottlenecks in Kenyan wild sorghum populations (Sorghum bicolor (L.) Moench ssp. verticilliflorum (Steud.) De Wet)

based on microsatellite diversity and genetic disequilibria. Genet. Resour. Crop Evol. 57(7):995-1005.

Mutegi 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 (2009). 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(2):243-253.

Mutegi E, Sagnard F, Semagn K, Deu M, Muraya M, Kanyenji B, de Villiers S, Kiambi D, Herselman L, Labuschagne M (2011). Genetic structure and relationships within and between cultivated and wild sorghum (*Sorghum bicolor* (L.) Moench) in Kenya as revealed by microsatellite markers. Theor. Appl. Genet. 122(5):989-1004.

Nei M (1973). Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. U.S.A. 70:3321-3323.

Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590.

Paterson AH, Schertz KF, Lin YR, Liu SC, Chang YL (1995). The weediness of wild plants: Molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum halepense* (L.) Pers. Proc. Natl. Acad. Sci. U.S.A. 92:6127-6131.

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

Piñeyro-Nelson A, van Heerwaarden J, Perales H, Serratos J, Rangel A, Hufford, M, Gepts P, Garay-Arroyo A, Rivera-Bustamante R, Alvarez-Buylla E (2009). Transgenes in Mexican maize: molecular evidence and methodological considerations for GMO detection in landrace populations. Mol. Ecol. 18:750-761.

Rueffler C, Van Dooren TJM, Leimar O, Abrams PA (2006). Disruptive selection and then what? Trends Ecol. Evol. 21(5):238-45A.

Sagnard F, Deu M, Dembélé D, Leblois R, Touré L, Diakité M, Calatayud C, Vaksmann M, Bouchet S, Mallé Y, Togola S, Traoré PCS (2011). Genetic diversity, structure, gene flow and evolutionary relationships within the *Sorghum bicolor* wild-weedy-crop complex in a western African region. Theor. Appl. Genet. 123(7):1231-1246.

Sahoo L, Schmidt J (2010). Growth and fitness components of wild × cultivated *Sorghum bicolor* (Poaceae) hybrids in Nebraska. Am. J. Bot. 97(10):1610-1617.

Saitou N, Nei M (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406-425.

Schlotterer C (2000). Evolutionary dynamics of microsatellite DNA. Chromosoma 109: 365-371.

Sherwin, WB, Jabot F, Rush R, Rossetto M (2006). Measurement of biological information with applications from genes to landscapes. Mol. Ecol. 15(10): 2857–2869

Song ZP, Lu BR, Zhu YG, Chen JK (2002). Pollen competition between cultivated and wild rice species (*Oryza sativa* and *O. rufipogon*). New Phytol. 153: 289–296.

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

Warwick SI, Beckie HJ, Hall LM (2009). Gene Flow, Invasiveness, and Ecological Impact of Genetically Modified Crops. The Year in Evolutionary Biology 2009: Ann. N.Y. Acad. Sci. 1168:72-99.

Weir BS (1990). Genetic data analysis. 1st ed. Sinauer Associates, Sunderland, MA.

Yang RC, Yeh FC (1993). Multilocus structure in Pinus contoria Dougl. Theor. Appl. Genet. 87:568–576.

Yeh FC, Boyle TJB (1997). Population genetic analysis of codominant and dominant markers and quantitative traits. Belg. J. Bot. 129:157.

Yonemaru J, Ando T, Mizubayashi T, Kasuga S, Matsumoto T, Yano M (2009). Development of Genome-wide Simple Sequence Repeat Markers Using Whole-genome Shotgun Sequences of Sorghum (*Sorghum bicolor* (L.) Moench). DNA Res. 16(3):187-193.