Genet Resour Crop Evol
DOI 10.1007/s10722-014-0175-5

RESEARCH ARTICLE

Characterization of the genetic diversity of Uganda's sweet potato (*Ipomoea batatas*) germplasm using microsatellites markers

Barbara M. Zawedde · Marc Ghislain · Eric Magembe · Geovani B. Amaro · Rebecca Grumet · Jim Hancock

Received: 7 April 2014 / Accepted: 1 September 2014

Abstract Knowledge about the genetic diversity and structure of crop cultivars can help make better conservation decisions, and guide crop improvement efforts. Diversity analysis using microsatellite markers was performed to assess the level of genetic diversity in sweet potato in Uganda, and evaluate the genetic relationship between the Uganda's germplasm and some genotypes obtained from Kenya, Tanzania, Ghana, Brazil and Peru. A total of 260 sweet potato cultivars were characterized using 93 microsatellite loci. The Ugandan collection showed a large number of distinct landraces, and very low (3 %) levels of genetic diversity between genotypes obtained from the different agro-ecological zones. There was low (6 %) levels of genetic diversity observed between the East African genotypes; however unique alleles were present in collections from the various sources. Pairwise comparisons of genetic differentiation indicated that Uganda's germplasm was significantly different ($P < 0.001$) from cultivars from Tanzania, Ghana, Brazil and Peru. The presence of unique alleles in populations from various Uganda's agro-ecological zones and other global regions, as well as the regional diversity patterns, suggest that efforts should be made to further collect and characterize the germplasm in more depth.

Keywords Characterization · Crop breeding · *Ipomoea batatas* · Molecular markers · SSR

B. M. Zawedde · R. Grumet · J. Hancock (✉)
Graduate Program in Plant Breeding, Genetics and Biotechnology, Michigan State University, Plant and Soil Science Building, 1066 Bogue Street, East Lansing, MI 48824, USA
e-mail: hancock@msu.edu

B. M. Zawedde
e-mail: zawedde@msu.edu

Present Address:
B. M. Zawedde
Uganda Biosciences Information Center (UBIC), National Crop Resources Research Institute, 27 km Kampala - Zirobwe Road, Namulonge, Kampala, Uganda

M. Ghislain · E. Magembe
CIP Sub-Saharan Africa, International Potato Center, P.O. Box 25171, Nairobi, Kenya

G. B. Amaro
Embrapa Vegetable Crops, P.O. Box 218, Brasília, DF CEP 70359-970, Brazil

Introduction

A germplasm collection of crop cultivars with varying environmental adaptive capacity can be both a source of genes for future crop improvement, as well as a critical resource for farmers. The highest levels of genetic diversity for the majority of the important global food crops is in the South, where crop centers of origins are commonly found, and centers of diversity emerged due to prolonged periods of farmer selection (FAO 2008).

Published online: 17 September 2014

Sweet potato, *Ipomoea batatas* (L.) Lam., is the fifth most important food crop in terms of weight harvested in Eastern Africa (FAO 2012). Sweet potato was introduced to the East African borders from South America by Portuguese explorers during the 16th century (Zhang et al. 2004). The oldest remains of sweet potato have been found in the caves of the Chilca Canyon in Peru and dated as 8,000 years old (Lebot 2010). However, based on morphological relationships among related species, the center of origin appears to be between the Yucatan Peninsula in Mexico and the Orinoco River in Venezuela (Austin 1977). It is also in that region that the wild species of the section Batatas, considered as putative ancestors and wild relatives of the cultivated sweet potato, are found (Andersson and de Vicente 2010). Evaluations of genetic diversity patterns among germplasm from different parts of the world have resulted in the suggestion that China, Southeast Asia, New Guinea and East Africa are secondary centers of diversity (Yen 1982; Austin 1983).

Uganda has the highest production per capita in Sub-Saharan Africa and numerous, diverse sweet potato landraces are grown. In 2005, the national sweet potato program collected over 1,300 landraces, which were characterized using morphological methodologies to determine the level of genetic diversity and 946 of these were found to be morphologically distinct genotypes (Yada et al. 2010a). This high level of diversity can be attributed primarily to the allogamous and hexaploidy nature of sweet potato (Lebot 2010), as well as variations in farmers' preferences (Veasey et al. 2008). The method of propagation by vine cuttings contributes also indirectly by maintaining cultivar diversity.

Knowledge about the genetic diversity and structure of existing crop cultivars can aid in making better conservation decisions, and help direct breeding programs. Characterization of crop diversity can be achieved through morphological and molecular tools. Morphological characterization is an important first step in assessment of diversity; however there are major limitations in relying only on morphological characterization including low levels of polymorphism, low repeatability, late expression for certain traits; phenotypic plasticity and parallel evolution (Karuri et al. 2010; Yada et al. 2010b). A number of molecular markers including random amplified polymorphic DNAs (RAPDs), restriction fragment length polymorphism (RFLPs), amplified fragment length polymorphism (AFLP), microsatellites or simple sequence repeats (SSRs), single nucleotide polymorphisms (SNPs) have been developed and used to complement morphological characterization. Selection of any particular DNA marker in a crop depends largely on the objective of the research, available resources and technical skills (Otoo et al. 2009).

In sweet potato research, a number of molecular markers have been used to study the crop's genetic diversity including RAPDs (Connolly et al. 1994; Gichuki et al. 2003; He et al. 2006), DNA amplification fingerprinting (DAF) (He et al. 1995), AFLPs (Zhang et al. 2004, Elameen et al. 2008), inter simple sequence repeats (ISSRs) (Hu et al. 2003), selective amplification of microsatellite polymorphic loci (SAMPL) (Tseng et al. 2002), and SSRs markers (Gichuru et al. 2006; Veasey et al. 2008) and microsatellites or SSRs (Jarret and Bowen 1994; Buteler et al. 1999; Hu et al. 2004; Yada et al. 2010b; Tumwegamire et al. 2011).

Sweet potato is a hexaploid ($2n = 6x = 90$) crop believed to have originated from natural hybridization between several wild species (Lebot 2010). There are three hypotheses for generation of hexaploid sweet potato: autopolyploidy (Kobayashi 1983; Shiotani and Kawase 1987, 1989); allo-autopolyploidy (Schafleitner et al. 2010) and allopolyploidy (Austin 1977; Nishiyama 1971; Srisuwan et al. 2006; Gao et al. 2011). The hexaploid nature of sweet potato and the complexity of its genomic makeup create a challenge when using molecular markers for germplasm characterization (Lebot 2010). However, SSR markers have proved to be useful in unveiling genetic relationships that are closely similar to morphological characterization in hexaploid Guinea yam *Dioscorea rotundata* Poir. (Mignouna et al. 2003) and detection of polymorphism in sweet potato (Veasey et al. 2008). Small sets of SSR markers, as low as 10 primer pairs, have been found to efficiently trace the movement of specific alleles between populations (Lebot 2010), and distinguish distinct individuals within a population (Yada et al. 2010b).

In Uganda, two studies have previously used SSR markers to analyze the genetic diversity of the sweet potato germplasm in Uganda. One study focused on assessed the genetic relationship among 192 selected superior landraces (high yielding, sweet potato virus disease or *Alternaria* blight disease resistance, or high

dry matter content) and concluded that there were 190 distinct landraces (Yada et al. 2010b). The second study used 75 cultivars from the same collection to assess the genetic relationship between East African orange-fleshed and the white-fleshed landraces, and how they related to cultivars from China, USA, Papua New Guinea and Peru (Tumwegamire et al. 2011). They found that the orange-fleshed types which were first developed in Kenya were distinct from those of non-African origin, and that orange-fleshed and white-fleshed landraces from East Africa are closely related. Since 1995, the national breeding programme has released twenty cultivars (Mwanga et al. 2011). Most of these cultivars (Sowola and NASPOT 1 to 11) were selected from bulked seed from a polycross between 18 and 24 parents.

The purpose of this study was to use SSR markers to determine the structure of genetic diversity in Ugandan's sweet potato, to assess genetic relationship between landraces and the released cultivars in Uganda, as well as to relate it to germplasm from other parts of the world. This information can be used to make recommendations that will improve and possibly enhance efficient low-cost conservation of sweet potato diversity.

Materials and methods

Plant material

Using the collection database maintained by the Sweet potato Program in the National Crop Resources Research Institute at Namulonge, a total of 168 Ugandan genotypes were selected based on three categories: response to weevil infestation (Muyinza et al. 2012); agro-ecological zone (Yada et al. 2010a); and varietal classification (Mwanga et al. 2011). Landraces were selected from the five major sweet potato growing agro-ecological zones of Uganda: northern region; eastern region; central region; western region; and southern region. To determine their genetic relationship with genotypes from other parts of the world, the Ugandan genotypes were compared with genotypes from other sweet potato growing African countries, Brazil and Peru. Improved cultivars were selected from bulked seed of three different polycrosses, each containing between 18 and 24 parents (Mwanga et al. 2003, 2009, 2011). Planting material of selected cultivars was grown in a screen house for 2 months and young leaves were put immediately on ice and stored at −80 °C until DNA extraction.

DNA extraction and simple sequence repeat (SSR) amplification

Genomic DNA was isolated from 200 mg of frozen leaf tissue using a modified CTAB method (Doyle and Doyle, 1990), according to (Yada et al. 2010b). A total of 308 DNA samples were quantified, diluted to 20 ng/µl and prepared for polymerase chain reaction (PCR) amplification. 31-labeled SSR primer pairs used in previous sweet potato studies (Karuri et al. 2010; Yada et al. 2010b; Tumwegamire et al. 2011) for DNA amplification were tested andonly 19 SSR markers (Table 1) were selected and used in this study. Other SSR markers were rejected due to non-specific amplification or for being monomorphic. A total reaction volume of 10 µl containing 1 µl of the 10X PCR buffer (10 mM Tris–HCl pH 9.0, 30 mM KCl, 1.5 mM $MgCl_2$), 0.1 µl of lyophilized 5 U/µl of Taq DNA polymerase, 0.2 µl of 10 mM dNTPs, 0.5 µl of 10 µM of the primers, 2.5 µl of 20 ng/µl DNA template and 5.7 µl of double distilled water, was used for PCR. The PCR program was set as follows: 94 °C for 3 min for initial denaturation, followed by 35 cycles each consisting of denaturation at 94 °C for 30 s, annealing at the indicated temperature for each primer pair (Table 1) and polymerization at 72 °C for 30 s, and then final extension at 72 °C for 20 min. Amplified PCR products were prepared for each DNA sample and used for fragment analysis using an ABI 3730 capillary sequencer (Applied Biosystems).

Allele scoring and data analysis

Genemapper v3.7 software (Applied Biosystems) was used for peak detection and fragment size estimation. AlleloBin software (Prasanth et al. 1997) was used to correct any errors in the scored alleles due to slippage of DNA polymerase during PCR (Schlotterer and Tautz 1992). Analyses were performed using two methods of data coding that were previously employed to describe diversity in polyploids (Esselink et al. 2004; Jørgensen et al. 2008; Kloda et al. 2008; García-Verdugo et al. 2009; Sampson and Byrne 2012). Firstly, the multilocus data were transformed into binary arrays of the presence/absence of an allele for each individual, using ALS-Binary software (Prasanth and Chandra 1997).

Table 1 Characteristics of SSR markers used to evaluate East African sweet potato cultivars: name, labeled dye, motifs, annealing temperature and reference

Name	Dye	Motif	Ta[a]	Reference
IB-S07	6-FAM	(tgtc)7	60	Benavides (unp.)[b]
IB-R03	PET	(gcg)5	58	Benavides (unp.)
IB-R16	VIC	(gata)4	59	Benavides (unp.)
IB-R13	NED	(ttc)6	58	Benavides (unp.)
J10A	PET	(aag)6	TD (57–62)	Solis et al. (unp.)[c]
J175	VIC	(aatc)4	TD (57–62)	Solis et al. (unp.)
IBCIP-9	6-FAM	(cca)2ac(acc)6	TD (50–60)	Yañez (2002)
IBCIP-13	NED	(acc)3 + (cgg)2 + (tgc)3 + (gtc)2	TD (57–62)	Yañez (2002)
IB-R19	PET	(cag)5b	TD (57–62)	Benavides (unp.)
BU691984	6-FAM	(tgg)6	TD (57–62)	Hu et al. (2004)
JB1809	VIC	(cct)6(ccg)6	TD (50–60)	Solis et al. (unp.)
IBSSR09	NED	(gaa)5(gag)3	TD (57–62)	Hu et al. (2004)
IB-R08	PET	(t3a)4	TD (50–60)	Benavides (unp.)
BU690524	VIC	(cag)5	TD (57–62)	Hu et al., 2004
BU690708	6-FAM	(ccg)5	TD (57–62)	Hu et al. (2004)
1B-S18	6-FAM	(tagc)4	TD (57–62)	Benavides (unp.)
IBCIP-1	6-FAM	(acc)7a	TD (57–62)	Yañez (2002)
IBSSR07	PET	(ct)7a(tc)4	TD (57–62)	Hu et al. (2004)
IB-R12	NED	(cag)5a	TD (57–62)	Benavides (unp.)

[a] Ta- annealing temperature; [b] unpublished data developed from 2002 to 2003 at CIP, [c] unpublished data developed from 2005 to 2006 at CIP

Secondly, the MAC-PR (microsatellite DNA allele counting-peak ratios) method (Esselink et al. 2004) was employed to obtain allele-dosage of 16 markers for each of the polyploidy individuals. The MAC-PR method makes use of the quantitative values for peak areas provided by the GeneMapper software. For each locus, all alleles were analysed in pairwise combinations in order to determine their copy numbers in the individual samples. This was accomplished by calculating ratios between peak areas for two alleles in all samples where these two alleles occurred together. Access to individuals with six different alleles or three different alleles provided a baseline for calculating the single- and double copy dosage, respectively.

The binary data was used to determine the polymorphic information content (PIC), which is the measure of the usefulness of each marker in distinguishing between individuals that was formulated by Weir (1996) as:

$$\mathrm{PIC}i = 1 - \sum_{j=1}^{n} Pij^2$$

where PIC_i is the PIC of a marker i; P_{ij} is the frequency of the jth pattern for marker i, and the summation extends over n patterns. The standard measures of genetic diversity for each population included number of polymorphic loci (P), percentage of polymorphic loci (%P) and Nei's (1973) gene diversity (D), estimated from binary data using GenAlex 6.4 (Peakall and Smouse 2006).

Analysis of molecular variance (AMOVA) was also performed using GenAlEx version 6.4 to estimate the total variance and distribution of diversity within and between populations. Wright's F-Statistic (F_{ST}, fixation index) was also computed, using GenAlEx software, to estimate the amount of genetic variance that can be explained by population structure (Holsinger and Bruce 2009).

$$\text{Fixation index, } F_{ST} = \frac{H_T - H_I}{H_T}$$

where H_I is the mean observed heterozygosity per individual within subpopulations and H_T is the expected heterozygosity in a random mating total population. F_{ST} can range from 0.0 (no differentiation) to 1.0 (complete differentiation, that is, subpopulations fixed for different alleles).

The phylogenetic relationship among populations was assessed using DARwin version 5 (Perrier and Jacquemoud-Collet 2006). Similarity matrices were constructed from the binary data with Jaccard's

coefficients (Jaccard 1908). Jaccard's coefficient = $N_{ab}/(N_a + N_b)$, where N_{ab} is the number of alleles shared by two individuals a and b, N_a is total number of alleles in sample a, and N_b is total number of alleles in sample b. Genetic distances between populations were obtained by computing the usual Euclidian distance matrix based on haplotype frequencies. From this matrix, a dendrogram was constructed using the neighbor joining method (NJ) from Saitou and Nei (1987). The significance of each node was evaluated by bootstrapping data over a locus for 5,000 replications of the original matrix. We examined hierarchical genetic variation between individuals using the un-weighted pair group method analysis (UPGMA), as suggested by Sneath and Sokal (1973).

Clustering patterns of individuals and populations were examined using STRUCTURE version 2.3.3 (Pritchard et al. 2000), which is reported to have the capability to generate population structuring (Pritchard et al. 2009). Using the allele dosage (MAC-PR) data for each individual, individuals were assigned probabilistically to genetic clusters (*K*). The STRUCTURE program was run using no prior assumptions of population structure with an admixture ancestry model and the recommended methods for recessive alleles, and allele frequencies correlated. The analysis was used to determine whether biologically relevant clusters could be determined among the plants sampled, and establish the proportion of an individual's genome (*Q*) that originated from each cluster. For all analyses, the Markov chain Monte Carlo (MCMC) parameters were set to a burn-in period of 50,000 with 50,000 iterations. The optimum *K*, indicating the number of true clusters in the data, was determined from 20 replicate runs for each value of K (K set to 10) using the method described by Evanno et al. (2005) and the *ad hoc* Quantity *Delta K*, based on the rate of change in the log probability of the data between successive *K* values. Parameters of the method of Evanno et al. (2005) were calculated using the program Structure Harvester version 0.6.92 (Earl and vonHoldt 2012). Similarity among different runs was calculated by the method of Jakobsson and Rosenberg (2007) as used in their computer program CLUMPP 1.1.2. This method calculates a similarity coefficient h', which allows the assessment of the similarity of individual runs of the program STRUCTURE. The optimal alignment of 20 replicates of *K* values was determined using the computer program CLUMPP 1.1.2 (Jakobsson and Rosenberg 2007) and clusters were visualized using the program DISTRUCT 1.1 (Rosenberg 2004).

Table 2 Observed base pair (bp) range, number of alleles and polymorphic information content (PIC) for the SSR markers used to characterize sweet potato genotypes from Uganda, Kenya and Tanzania

Name	bp range	No. of alleles	PIC
IB-S07	173–177	5	0.22
IB-R03	244–258	5	0.60
IB-R16	194–212	5	0.66
IB-R13	226–296	9	0.77
J10A	175–229	8	0.82
J175	112–144	6	0.81
IBCIP-9	176–194	4	0.80
IBCIP-13	196–374	7	0.52
IB-R19	191–208	4	0.63
BU691984	252–267	3	0.67
JB1809	196–264	4	0.19
IBSSR09	196–208	5	0.31
IB-R08	205–216	4	0.64
BU690524	272–293	8	0.80
BU690708	235–259	7	0.81
1B-S18	211–235	4	0.76
IBCIP-1	135–189	8	0.89
IBSSR07	158–178	4	0.70
IB-R12	303–341	7	0.79

Results

SSR markers amplification

A total of 107 alleles were scored for the 19 SSR markers (Table 2). The number of alleles per locus ranged from 3 to 9. Three markers had very low PIC; IB-S07 (0.22), JB1809 (0.19) and IBSSR09 (0.31) thus were excluded from further analyses.

Determining relatedness between cultivars in Uganda

A total of 10 newly improved cultivars released by the national program were compared with 158 Ugandan landraces. The unweighted neighbor joining (NJ) algorithm cluster analysis generated numerous clusters (Fig. 1). Improved cultivars were scattered into

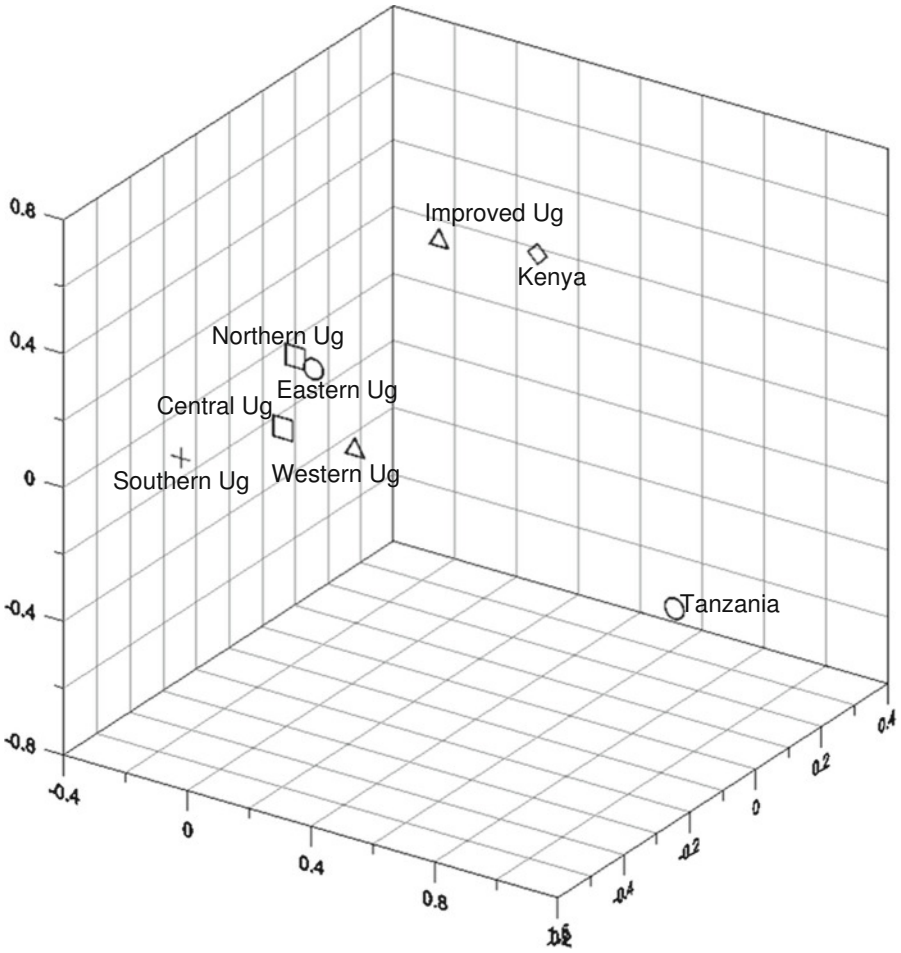

Fig. 1 Non-metric multidimensional scaling representation of the principal coordinates' analysis (PCoA) comparing genotypes from different Uganda's agro-ecological zones and other East African countries. The percentage of variation explained by the three axes is 89.1 %

many of these clusters together with landraces. Noteworthy (6/10) improved cultivars were grouped together with a Kenyan cultivar Kakamega, which was purposely included in this analysis because it is a known maternal parent for many of these improved cultivars.

Genetic relationship between genotypes from Uganda's agro-ecological zones and cultivars collected from other East African countries

Analysis of Molecular Variance (AMOVA) indicated that only 6 % of the genetic variation was explained by differences among the sources (Table 3). Analysis of the sixteen microsatellites yielded a total of 93 presumptive loci in the 228-sweet potato genotypes from the eight predefined populations (Table 4). An average of 70 polymorphic loci was observed in each population. The level of genetic diversity varied among the different populations. Most regions in Uganda had populations with few unique alleles (1–4), except the south-western region, which had none. Tanzanian cultivars also had few unique alleles (4), but had the highest level of heterozygosity (*D*). Overall the level of heterozygosity (*D*) for the collected samples was low. The significant difference between the cultivars from Tanzania and the populations from Uganda and Kenya is clearly shown in the genetic distance matrix (Fig. 2).

Genetic relationship between Ugandan genotypes and cultivars collected from elsewhere

Analysis of Molecular Variance (AMOVA) indicated that only 24 % of the genetic variation was explained by differences among the countries (Table 5). Pairwise comparisons of genetic differentiation among countries indicated that Uganda's germplasm was

Table 3 AMOVA for genetic diversity within and among the populations of sweet potato from different Uganda's agro-ecological zones and other East African countries (Kenya and Tanzania) as well as Uganda's improved cultivars

Source	*df*	SS	MS	Est. Var.	%	FST	*P* value
Among regions	7	275.24	34.41	0.81	6	0.056	0.010
Within regions	220	3,206.70	13.65	13.65	94		
Total	227	3,481.93		14.45	100		

Table 4 Number of individuals, number of unique alleles, and genetic diversity parameters[a] for sweet potato populations from various Uganda's agro-ecological zones and other East African countries (Kenya, Tanzania), as well as Uganda's improved cultivars

Regions	Individuals	RA	*P*	%P	*D*
Northern Uganda	23	2	69	75.3	0.433 ± 0.032
Eastern Uganda	66	1	78	84.4	0.433 ± 0.029
Central Uganda	34	1	73	79.2	0.442 ± 0.031
Western Uganda	43	4	79	85.7	0.449 ± 0.027
South-western Uganda	14	0	62	67.5	0.431 ± 0.032
Improved cultivars	11	0	57	62.3	0.347 ± 0.028
Kenya	15	2	62	67.5	0.457 ± 0.028
Tanzania	22	4	78	84.4	0.514 ± 0.025
Mean	–	–	70	75.8	0.438 ± 0.010
Total	228	13	93	–	–

[a] Where P is total number of polymorphic loci per region, %P is percentage of polymorphic loci, and D is Nei's (1973) gene diversity estimated with computer program GenAlex 6.4 (Peakall and Smouse 2006)

significantly different ($P < 0.001$) from genotypes from Brazil, Peru and Ghana (Table 6). The Jaccard's similarity coefficients ranged from 0.0 to 0.95 with a mean of 0.56. More than 70 % of the pair-wise similarity coefficients were between 0.50 and 0.63.

The dendrogram generated by DARwin software revealed three clusters (Fig. 2). To efficiently visualize the results, the dendrogram was pruned from the complete tree to show clustering only between genotypes that had bootstrap values greater than 60. This pruned tree showed similar broad clustering patterns as the complete tree (data not shown). East African germplasm and cultivars from USA were found in Cluster A, the majority of cultivars from Brazil and Peru were in cluster B while most Ghanaian cultivars were found in Cluster C.

The Bayesian model of STRUCTURE (Pritchard et al. 2000) assigned the individuals to two major genetic clusters, as the highest Delta K was observed at K = 2. All individuals appeared to have a component of both clusters in their genome; however, the Ugandan and Kenyan cultivars had a very high proportion of their genome originating from cluster K1, Brazil, Ghana and Peru had a very high proportion of their genome originating from cluster K2, while Tanzanian cultivars were composed of a mixture of the two clusters (Fig. 3).

Discussion

The number of alleles per primer pair observed in this work is close to that obtained by Yada et al. (2010b) using the same SSR markers. However, our number of alleles varies somewhat from those reported by Tumwegamire et al. (2011) on similar sweet potato germplasm using the same markers. Higher number of alleles was observed for some markers in this work and this is likely due to a larger genotype sample size and a higher resolution of DNA fragment. Yada et al. (2010b) assessed 192 samples using the ABI system similar to what was used in this work, while Tumwegamire et al. (2011) screened 75 samples with the LiCOR system.

A total of 92 out of 106 markers were highly polymorphic, which confirms the high discriminating power of the SSR markers (Hwang et al. 2002; Gichuru et al. 2006, Veasey et al. 2008, Yada et al. 2010a, b; Tumwegamire et al. 2011). Hwang et al.

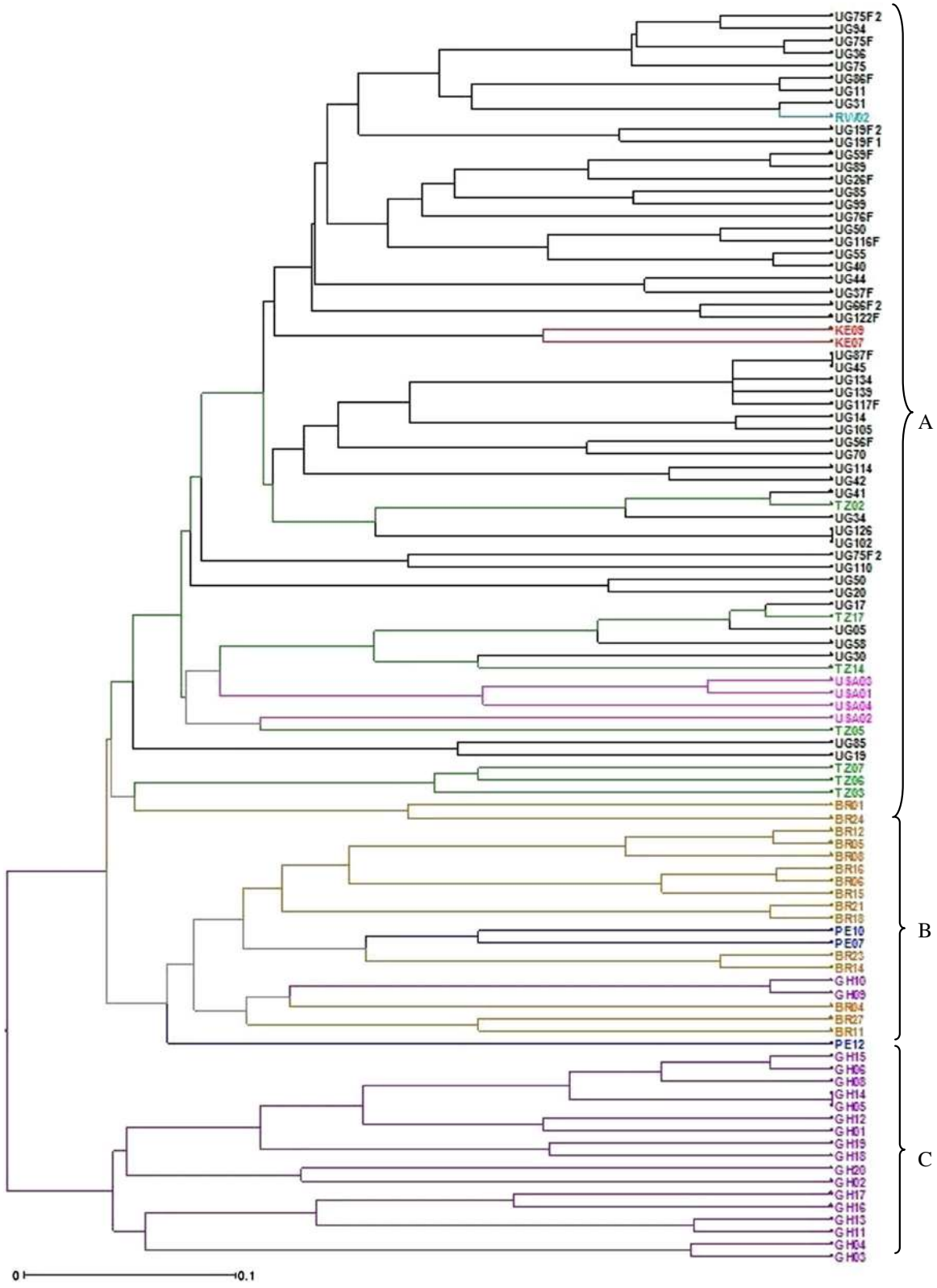
UG75F2
UG94
UG75F
UG36
UG75
UG86F
UG11
UG31
RWA02
UG19F2
UG19F1
UG59F
UG89
UG26F
UG85
UG99
UG76F
UG50
UG116F
UG55
UG40
UG44
UG37F
UG66F2
UG122F
KE09
KE07
UG87F
UG45
UG134
UG139
UG117F
UG14
UG105
UG56F
UG70
UG114
UG42
UG41
TZ02
UG34
UG126
UG102
UG75F2
UG110
UG50
UG20
UG17
TZ17
UG05
UG58
UG30
TZ14
USA03
USA01
USA04
USA02
TZ05
UG85
UG19
TZ07
TZ06
TZ03
BR01
BR24
BR12
BR05
BR08
BR16
BR06
BR15
BR21
BR18
PE10
PE07
BR23
BR14
GH10
GH09
BR04
BR27
BR11
PE12
GH15
GH06
GH08
GH14
GH05
GH12
GH01
GH19
GH18
GH20
GH02
GH17
GH16
GH13
GH11
GH04
GH03
A
B
C
0
0.1

Fig. 2 A dendrogram of the unweighted pair group method analysis (UPGMA) cluster analysis on the basis of Jaccard's simple sequence repeat (SSR) showing genetic similarities among 102 cultivars. BR, GH, KE, PE, RW, TZ, UG stands for Brazil, Ghana, Kenya, Peru, Rwanda, Tanzania, and Uganda, respectively. Other cultivars were obtained from USA. The dendrogram was pruned from the complete tree to show clustering between genotypes that had bootstrap values greater than 60. This pruned tree shows similar broad clustering as observed for the complete tree (data not shown)

(2002) suggested that the high level of polymorphism of molecular markers in sweet potato was due to its large genome size and high levels of heterozygosity. Such high levels of heterozygosity would be expected if the sweet potato is an allopolyploid as many suggest (Austin 1977; Nishiyama 1971; Srisuwan et al. 2006; Gao et al. 2011) and high levels of heterozygosity would be maintained by the common practice of vegetative propagation of sweet potato. In addition, the mating system of sweet potato, which is outcrossing due to frequent self-incompatibility, would support high levels of heterozygosity.

There was a low mean genetic similarity coefficient of 0.56 between the Ugandan cultivars and cultivars from other African countries, similar to that observed by Tumwegamire et al. (2011). This suggests that there is a large amount of sweet potato genetic diversity in Uganda that needs to be preserved. Very few Ugandan landraces had high genetic similarity coefficients (<0.05) with other cultivars, suggesting that there is high level of genetic diversity being maintained by the farmers. This observation was supported by the AMOVA results, which showed over that 94 % of the variation was found within populations. Similar results were observed in previous studies assessing East African germplasm (Gichuru et al. 2006; Tumwegamire et al. 2011). The high level of genetic diversity within a given gene pool is influenced by a number of factors such gene flow through intra-specific introgression, and movement of plant materials between localities during farmer plant material exchange and levels of farmer selection (Gichuru et al. 2006; Veasey et al. 2008). Landrace exchange can explain our observation that a number of

Table 5 AMOVA for genetic diversity within and among sweet potato populations from Uganda, other African countries (Ghana, Kenya, Tanzania, Others[a]), and American countries (Brazil, Peru, USA). [a] Others stand for other African countries including Rwanda and Mozambique

Source	*df*	SS	MS	Est. Var.	%	FST	*P* value
Among countries	7	328.09	46.87	1.95	24	0.244	0.001
Within countries	252	1,513.79	6.03	6.03	76		
Total	259	1,841.88		7.98	100		

Table 6 Pair-wise F_{ST}^{a} for populations from Uganda, other African countries (Ghana, Kenya, Tanzania, Others[b]), and American countries (Brazil, Peru, USA)[c]

Regions	Others	Brazil	Ghana	Kenya	Peru	Tanzania	Uganda	USA
Others	0.000							
Brazil	*0.187*	0.000						
Ghana	*0.258*	*0.347*	0.000					
Kenya	0.006	*0.294*	*0.314*	0.000				
Peru	*0.181*	*0.127*	*0.221*	*0.308*	0.000			
Tanzania	0.010	*0.197*	*0.246*	*0.190*	*0.166*	0.000		
Uganda	0.000	*0.285*	*0.387*	0.068	*0.299*	*0.100*	0.000	
USA	0.029	*0.225*	*0.257*	*0.210*	0.116	0.033	*0.133*	0.000

[a] F_{ST} is the mean reduction in observed heterozygosity of a subpopulation (relative to the total population) due to genetic drift among subpopulations. 0.0 means no differentiation while 1.0 means complete differentiation

[b] Others stand for cultivars obtained from other African countries including Rwanda and Mozambique

[c] Italicised values indicate significant difference between the populations ($P < 0.001$)

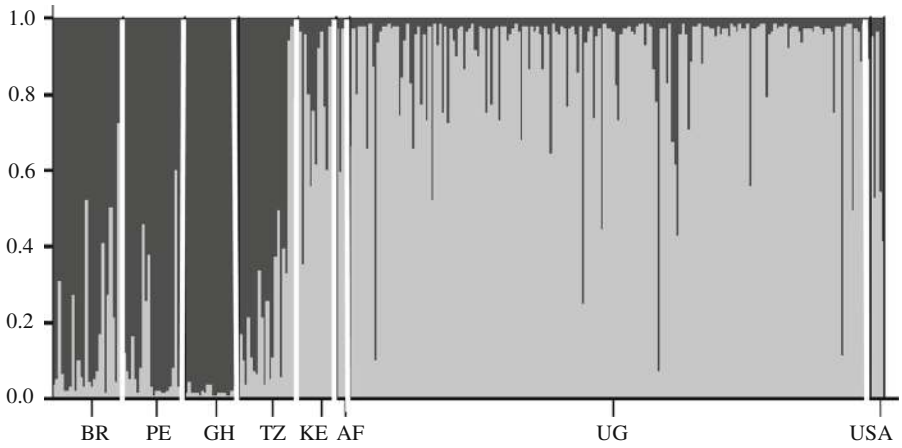

Fig. 3 Probabilities of membership for each individual in clusters K1 and K2 as determined by the Bayesian clustering method implemented in STRUCTURE version 2.3.3 (Pritchard et al. 2000). BR, PE, GH, TZ, KE, AF, and UG stands for Brazil, Peru, Ghana, Tanzania, Kenya, other African countries (Mozambique and Rwanda) and Uganda, respectively. Other cultivars were obtained from USA. Each individual is represented as a vertical bar and the shading correspond to its membership probabilities in clusters K1 (*light*) and K2 (*dark*)

Ugandan landraces had high genetic similarity with cultivars from Kenya, Rwanda and Tanzania.

While the majority of genetic diversity was spread evenly across Uganda and East Africa, there were distinct regional clusters of genotypes. Pairwise comparisons of genetic differentiation among regions indicated that cultivars from Tanzania were significantly different ($P < 0.001$) from all the Ugandan populations and the Kenyan one; Uganda's germplasm was significantly different ($P < 0.001$) from cultivars from Brazil, Peru and Ghana. This suggests that the same genes were being assorted in different arrays in the various regions.

Even though sweet potato was likely introduced into Africa from Brazil, we observed substantial differentiation between the Brazilian cultivars and all of the African populations. This could be explained by the low representation of the Brazilian germplasm, which in this study consisted of only 22 genotypes. Another explanation for the higher degree of differentiation among populations is that the sweet potato gene pool evolved to local conditions and human preferences as it was spread from America throughout Africa. This would explain why the majority of the American cultivars were found in one cluster, and most of the East African germplasm were found in another cluster. However, some of the cultivars from Tanzania clustered more closely with the American group. This may be because the Tanzanian collection contained ancient genotypes that have been cultivated in isolation on some of Tanzania's islands. The Ghanaian germplasm was more closely related to the American cultivars than with the East African genotypes, although they still formed a unique cluster. The cultivars from the USA had a higher genetic similarity with the East African germplasm compared with other American genotypes, probably because these genotypes were introduced in Uganda about 15 years ago and have been used as paternal parents in the polycrosses (Mwanga et al. 2003). It is likely that hybridization and introgression has occurred between these genotypes and Ugandan germplasm.

The majority of the improved cultivars in Uganda were found in one cluster probably because of by the source of parents used in breeding the improved cultivars. These improved cultivars were generated from a polycross of a single cultivar crossed with 18–24 parents including introductions from other parts of the world (Mwanga et al. 2003, 2009, 2011). For example NASPOT 7-11 were generated from a polycross of 24 parents with 'Kakamega' (a Kenyan landrace) as the maternal parent. The cross, included 'Zapallo' (from Peru), 'Beauregard' and 'Jewel' (from USA), and three introductions from the International Institute of Tropical Agriculture (IITA) based in Nigeria. Hence, it is quite remarkable to note the level of genetic diversity present in this relatively small and interbred sample of improved cultivars. This

finding is of particular significance for genetic diversity conservation because farmers are likely to gradually shift from their poor performing cultivars to improved ones. The three sub-clusters identified, which were composed of only landraces also suggests that there is still a large genetic diversity in Uganda that has not been tapped by breeding programs.

Conclusions

Overall, the sweet potato has high levels of genetic diversity. However, the presence of unique alleles in populations from various Uganda's agro-ecological zones and other global regions, as well as the regional diversity patterns, indicates the value of collecting and characterizing the germplasm in more depth. The use of microsatellite marker data can be particularly useful to make better choices of what needs to be preserved in order to increase genetic diversity and representation of landraces across Africa. These genotypes need to be incorporated in the collections at the national gene bank and managed to ensure their long-term conservation. Finally, the origin of sweet potato germplasm in East Africa doesn't appear to be strictly of a single Brazilian origin but rather successive introduction from several sources.

Acknowledgments We are very grateful to the Norman E. Borlaug Leadership Enhancement in Agriculture Program (LEAP) for funding this research. Our sincere gratitude goes to Dr. Joseph Nduguru and Luambano Nessie at Mikocheni Agricultural Research Institute, Tanzania for providing us with the samples from Tanzania. We are also grateful to Francis Osingada and Jimmy Akono at the Biosciences Facility of the National Crop Resources Research Institute, Uganda, Bramwel Wanjala of Biosciences eastern and central Africa (*BecA*) Hub and Maggie Mwathi of CIP-Office at the International Livestock Research Institute in Nairobi, Kenya, for the technical assistance provided to conduct the research.

References

Andersson MS, de Vicente MC (2010) Gene flow between crops and their wild relatives. Johns Hopkins University Press, Baltimore, p 564

Austin DF (1977) Hybrid polyploids in *Ipomoea* section of *Batatas*. J Hered 68:259–260

Austin DF (1983) Variability in sweet potato in America. Proc. Amer. Soc. Hort. Sci. 27(pt. B):15–26

Buteler MI, Jarret RL, La Bonte DR (1999) Sequence characterization of microsatellites in diploid and polyploid *Ipomoea*. Theor Appl Genet 99:123–132. doi:10.1007/s001220051216

Connolly AG, Godwin ID, Cooper M, Delacy IH (1994) Interpretation of randomamplified polymorphic DNA marker data for fingerprinting sweet potato [*Ipomoea batatas* (L.) Lam.] genotypes. Theor Appl Genet 88:332–336. doi:10.1007/BF00223641

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15. http://ci.nii.ac.jp/naid/10003365693/

Earl DA, von Holdt MB (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4(2):359–361. doi:10.1007/s12686-011-9548-7

Elameen A, Fjellheim S, Larsen A, Rognli AO, Sundheim L, Msolla S, Masumba E, Mtunda K, Klemsdal SS (2008) Analysis of genetic diversity in a sweet potato (*Ipomoea batatas* L.) germplasm collection from Tanzania as revealed by AFLP. Genet Resour Crop Evol 55:397–408. doi:10.1007/s10722-007-9247-0

Esselink GD, Nybom H, Vosman B (2004) Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting-peak ratios) method. Theor Appl Genet 109:402–408. doi:10.1007/s00122-004-1645-5

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

FAO (2008) The state of food insecurity in the world 2003. Food and Agriculture Organization, Rome. http://www.fao.org/3/a-i0291e.pdf. Accessed 16 May 2012

FAO (2012) FAO crop production statistics. Food and Agriculture Organization. http://faostat.fao.org/. Accessed 03 March 2013

Gao M, Ashu GM, Lakeisha S, Akwe WA, Njiti V, Barnes S (2011) Wx intron variations support an allohexaploid origin of the sweet potato [*Ipomoea batatas* (L.) Lam.]. Euphytica 177:111–133. doi:10.1007/s10681-010-0275-z

García-Verdugo C, Fay MF, Granado-Yela C, Rubio De Casas R, Balaguer L, Besnard G, Vargas P (2009) Genetic diversity and differentiation processes in the ploidy series of *Olea europaea* L.: a multiscale approach from subspecies to insular populations. Mol Ecol 18:454–467. doi:10.1111/j.1365-294X.2008.04027.x

Gichuki ST, Berenyi M, Zhang D, Herman M, Schmidt J, Glossl J, Burgh K (2003) Genetic diversity in sweet potato [*Ipomoea batatas* (L.) Lam.] in relationship to geographic sources as assessed with RAPD markers. Genet Resour Crop Evol 50:429–437. doi:10.1023/A:1023998522845

Gichuru V, Aritua V, Lubega GW, Edema R, Adipala E, Rubaihayo PR (2006) A preliminary analysis of diversity among East African sweet potato landraces using morphological and simple sequence repeats (SSR) markers. Acta Hort (ISHS) 703:159–164. http://www.actahort.org/books/703/703_18.htm

He G, Prakash CS, Jarret RL (1995) Analysis of genetic diversity in a sweet potato (*Ipomoea batatas*) germplasm collection using DNA amplification fingerprinting. Genome 38(5):938–945. doi:10.1139/g95-123

He X, Liu Q, Ishiki K, Zhai H, Wang Y (2006) Genetic diversity and genetic relationships among Chinese sweet potato

landraces revealed by RAPD and AFLP markers. Breed Sci 56:201–207. doi:10.1270/jsbbs.56.201

Holsinger KE, Bruce SW (2009) Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat Rev Genet 10(9):639–650. doi:10.1038/nrg2611

Hu J, Nakatani M, Lalusin AG, Kuranouch T, Fujimura T (2003) Genetic analysis of sweet potato and its wild relatives using inter-simple sequence repeat (ISSRs). Breed Sci 53:297–304. doi:10.1270/jsbbs.53.297

Hu J, Nakatani M, Mizuno K, Fujimura T (2004) Development and characterization of microsatellite markers in sweet potato. Breed Sci 54:177–188. doi:10.1270/jsbbs.54.177

Hwang SY, Tseng YT, Lo HF (2002) Application of simple sequence repeats in determining the genetic relationships of cultivars used in sweet potato polycross breeding in Taiwan. Sci Hort 93:215–224. doi:10.1016/S0304-4238(01)00343-0

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44:223–270 (In French)

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23(14):1801–1806. doi:10.1093/bioinformatics/btm233

Jarret RL, Bowen N (1994) Simple sequence repeats (SSRs) for sweet potato characterization. Plant Genet Resour Newsl 100:9–11

Jørgensen MH, Carlsen T, Skrede I, Elven R (2008) Microsatellites resolve the taxonomy of the polyploidy *Cardamine digitata* aggregate (Brassicaceae). Taxon 57:882–892. http://www.jstor.org/stable/27756715

Karuri HW, Ateka EM, Amata R, Nyende AB, Muigai AWT, Mwasame E, Gichuki ST (2010) Evaluating diversity among Kenyan sweet potato genotypes using morphological and SSR markers. Int J Agric Biol. 12:33–38. http://www.fspublishers.org/published_papers/31561_..pdf

Kloda JM, Dean PDG, Maddren C, MacDonald DW, Mayes S (2008) Using principal component analysis to compare genetic diversity across polyploidy levels within plant complexes: an example from British Restharrows (*Ononis spinosa* and *Ononis repens*). J Hered 100:253–260. doi:10.1038/sj.hdy.6801044

Kobayashi M (1983) The *Ipomoea trifida* complex closely related to sweet potato. In: Shideler SF, Rincon H (eds.) Proceedings of the 6th symposium of the International Society of Tropical Root Crops, Lima, Peru, 21–26 February, 1983, International Potato Center, Lima, Peru, pp 561–568

Lebot V (2010) Sweet potato, chapter 3. In: Bradshaw JE (ed) Root and tuber crops. Handbook of Plant Breeding 7. Springer Science & Business Media, Berlin, pp 97–125

Mignouna HD, Abang MM, Fagbemi SA (2003) A comparative assessment of molecular marker assays (AFLP, RAPD and SSR) for white yam (*Dioscorea rotundata* Poir.) germplasm characterization. Ann Appl Biol 142:269–276. doi:10.1111/j.1744-7348.2003.tb00250.x

Muyinza H, Talwana HL, Mwanga ROM, Stevenson PC (2012) Sweet potato weevil (Cylas spp.) resistance in African sweet potato germplasm. Int J Pest Manag 58(1):78–81. doi:10.1080/09670874.2012.655701

Mwanga ROM, Odongo B, Ocitti p'obwoya C, Gibson RW, Smith N, Carey E (2003) Release of six ("NASPOT 1" to "NASPOT 6") sweet potato cultivars in Uganda. HortScience 38:475–476. http://hortsci.ashspublications.org/content/38/3/475.full.pdf

Mwanga ROM, Odongo B, Niringiye CN, Alajo A, Kigozi B, Makumbi R et al (2009) "NASPOT 7", "NASPOT 8", "NASPOT 9 O", "NASPOT 10 O", and "Dimbuka-Bukulula" sweet potato. HortScience 44:828–832. http://hortsci.ashspublications.org/content/44/3/828.full.pdf

Mwanga ROM, Niringiye CN, Alajo A, Kigozi B, Namakula J, Mpembe I et al (2011) "NAPOT 11", a sweet potato cultivar bred by a participatory plant-breeding approach in Uganda. HortScience 46:317–321. http://hortsci.ashspublications.org/content/46/2/317.full

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Nishiyama I (1971) The origin of the sweet potato plant. In: Barrau J (ed) Plants and the migrations of Pacific peoples. Press Honolulu, Bishop Mus, pp 119–128

Otoo E, Akromah R, Kolesnikova-Allen M, Asiedu R (2009) Delineation of pona complex of yam in Ghana using SSR markers. Int J Genet Mol Biol 1:6–16 (Article number 77EE4812732)

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295. doi:10.1111/j.1471-8286.2005.01155.x

Perrier X, Jacquemoud-Collet JP (2006) DARwin software. http://darwin.cirad.fr/darwin. Accessed 26 Aug 2012

Prasanth VP, Chandra S (1997) ALS-Binary: A C Program for converting allele size in microsatellite markers to 0–1 (Binary Data). Biometrics. ICRISAT. International Crops Research Institute for the Semi-Arid Tropics, India. http://www.icrisat.org/bt-software-d-als.htm. Accessed 05 July 2012

Prasanth VP, Chandra S, Jayashree B, Hoisington D (1997) AlleloBi—a program for allele binning of microsatellite markers based on the alogirithm of Idury and Cardon (1997). ICRISAT. International Crops Research Institute for the Semi-Arid Tropics, India. http://www.icrisat.org/bt-software-d-allelobin.htm. Accessed 05 July 2012

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure from multilocus genotype data. Genetics 155:945–959. http://www.genetics.org/content/155/2/945.full.pdf

Pritchard JK, Wen X, Falush D (2009) Documentation for structure software. Version 2.3. http://pritch.bsd.uchicago.edu/structure.html. Accessed 18 Oct 2012

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138. https://web.stanford.edu/group/rosenberglab/software/distructManual.pdf

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425. http://mbe.oxfordjournals.org/content/4/4/406.full.pdf+html

Sampson JF, Byrne M (2012) Genetic diversity and multiple origins of polyploid *Atriplex nummularia* Lindl. (Chenopodiaceae). Biol J Linn Soc 105:218–230. doi:10.1111/j.1095-8312.2011.01787.x

Schafleitner R, Tincopa LR, Palomino O, Rossel G, Robles RF, Alagon R, Rivera C, Quispe C, Rojas L, Pacheco JA (2010) A sweet potato gene index established by de novo assembly of pyrosequencing and Sanger sequences and mining for gene-based microsatellite markers. BMC Genome 11:604. doi:10.1186/1471-2164-11-604

Schlotterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. Nucleic Acid Res 20:211–215 (PMCID: PMC310356)

Shiotani I, Kawase T (1987) Synthetic hexaploids derived from wild species related in *Ipomoea trifida*. Jpn J Breed 39:57–66

Shiotani I, Kawase T (1989) Genomic structure of the sweet potato and hexaploids in *Ipomoea trifida* (H.B.K) Don. Jpn J Breed 39:57–66. http://hdl.handle.net/10076/8643

Sneath PHA, Sokal RR (1973) Numerical Taxonomy. Freeman, San Francisco, p 573

Srisuwan S, Sihachakr D, Siljak-Yakovlev S (2006) The origin and evolution of sweet potato (*Ipomoea batatas* Lam.) and its wild relatives through the cytogenetic approaches. Plant Sci 171:424–433. doi:10.1016/j.plantsci.2006.05.007

Tseng YT, Lo HF, Hwang SY (2002) Genotyping and assessment of genetic relationships in elite polycross breeding cultivars of sweet potato in Taiwan based on SAMPL polymorphisms. Bot Bull Acad Sin 43:99–105. http://ejournal.sinica.edu.tw/bbas/content/2002/2/bot432-02.html

Tumwegamire S, Kapinga R, Rubaihayo PR, LaBonte Do R, Grüneberg WJ, Burgos G, zum Felde T, Carpio R, Pawelzik E, Mwanga ROM (2011) Evaluation of dry matter, protein, starch, sucrose, β-carotene, iron, zinc, calcium, and magnesium in east african sweet potato [*Ipomoea batatas* (L.) Lam.] Germplasm. HortScience 46:348–357. http://hortsci.ashspublications.org/content/46/3/348.full

Veasey EA, Borges A, Rosa MS, Queiroz-Silva JR, Bressan EDA, Peroni N (2008) Genetic diversity in Brazilian sweetpotato (Ipomoea batatas (L.) Lam., Solanales, Convolvulaceae) landraces assessed with microsatellite markers. Genet Mol Biol 31:725–733. doi:10.1590/S1415-47572008000400020

Weir BS (1996) Genetic data analysis II. Methods for discrete population genetic data. Sinauer Associates, Inc., Sunderland, p 445

Yada B, Tukamuhabwa P, Villordon A, Alajo A, Mwanga ROM (2010a) An online database of sweet potato germplasm collection in Uganda. Hortscience 45(1):153–153. http://hortsci.ashspublications.org/content/45/1/153.full

Yada B, Tukamuhabwa P, Wanjala B, Kim D-J, Skilton RA, Alajo A, Mwanga ROM (2010b) Characterization of Ugandan sweet potato germplasm using fluorescent labeled simple sequence repeat markers. Hortscience 45(2):225–230. http://hortsci.ashspublications.org/content/45/2/225.full

Yañez AVO (2002) Aislamiento y caracterización de marcadores moleculares microsatelites a partir de la construcción de librerias genomicas enriquecidas de camote (*Ipomoen batatas* (L.) Lam.). Universidad Nacional Mayor de San Marcos. Facultad de Ciencias Biológicas. EAP, Lima 108

Yen DE (1982). Sweet potato in historical perspective. In: Villareal RL, Griggs TD (eds) Proceedings of the first international symposium, AVRDC, Publ. No. 82–172, pp 17–30

Zhang DP, Rossel G, Kriegner A, Hijmans R (2004) AFLP assessment of diversity in sweet potato from Latin America and the Pacific region: its implications on the dispersal of the crop. Genet Resour Crop Evol 51:115–120. doi:10.1023/B:GRES.0000020853.04508.a0