Genetic diversity of cassava (*Manihot esculenta* Crantz) landraces and cultivars from southern, eastern and central Africa

R. S. Kawuki[1,2,3], L. Herselman[2], M. T. Labuschagne[2], I. Nzuki[3,4],
I. Ralimanana[5], M. Bidiaka[6], M. C. Kanyange[7], G. Gashaka[7],
E. Masumba[8], G. Mkamilo[8], J. Gethi[9], B. Wanjala[9], A. Zacarias[10],
F. Madabula[10] and M. E. Ferguson[3]*

[1]*National Crops Resources Research Institute (NaCRRI), PO Box 7084, Kampala, Uganda,*
[2]*Department of Plant Sciences, University of the Free State, PO Box 339, Bloemfontein
9300, South Africa,* [3]*International Institute of Tropical Agriculture (IITA), PO Box 30709,
Nairobi, Kenya,* [4]*Bioscience eastern and central Africa, c/o International Livestock
Research Institute, PO Box 30709, Nairobi, Kenya,* [5]*FOFIFA/DRA, PO Box 1444,
Antananarivo, Madagascar,* [6]*Institut National pour l'Etude et la Recherche Agronomique
(INERA), Kinshasa, Democratic Republic of Congo,* [7]*Institut des Sciences Agronomiques
du Rwanda (ISAR), BP 138, Butare, Rwanda,* [8]*Root and Tuber Research Program,
Naliendele Agricultural Research Institute, PO Box 509, Mtwara, Tanzania,* [9]*Kenya
Agricultural Research Institute (KARI), Katumani, PO Box 340, Machakos, Kenya and*
[10]*Agricultural Research Institute (IIAM), PO Box 1922, Maputo, Mozambique*

Abstract

Studies to quantify genetic variation in cassava germplasm, available within the national breeding programmes in Africa, have been limited. Here, we report on the nature and extent of genetic variation that exists within 1401 cassava varieties from seven countries: Tanzania (270 genotypes); Uganda (268); Kenya (234); Rwanda (184); Democratic Republic of Congo (DRC; 177); Madagascar (186); Mozambique (82). The vast majority of these genotypes do not exist within a formal germplasm conservation initiative and were derived from farmers' fields and National Agricultural Research Systems breeding programmes. Genotypes were assayed using 26 simple sequence repeat markers. Moderate genetic variation was observed with evidence of a genetic bottleneck in the region. Some differentiation was observed among countries in both cultivars and landraces. Euclidean distance revealed the pivotal position of Tanzanian landraces in the region, and STRUCTURE analysis revealed subtle and fairly complex relationships among cultivars and among landraces and cultivars analysed together. This is likely to reflect original germplasm introductions, gene flow including farmer exchanges, disease pandemics, past breeding programmes and the introduction of cultivars from the International Institute of Tropical Agriculture – Nigeria. Information generated from this study will be useful to justify and guide a regional cassava genetic resource conservation strategy, to identify gaps in cassava diversity in the region and to guide breeding strategies.

Keywords: Africa; cassava; genetic differentiation; genetic diversity; *Manihot esculenta* Crantz; simple sequence repeats

---

* Corresponding author. E-mail: m.ferguson@cgiar.org

Introduction

Cassava (*Manihot esculenta* Crantz) is one of the most important staple crops in sub-Saharan Africa, where it is particularly popular with low-income, women-led households (Nweke *et al.*, 2001). Its popularity is derived from its inherent resilience to abiotic stresses, low input requirements, the diversity of uses to which its starchy roots are amenable and its flexible harvest date which allows it to be 'banked' in the soil as a food reserve (Nweke *et al.*, 2001). In sub-Saharan Africa, cassava roots are primarily used as food in either fresh or processed form. In addition, cassava leaves are an important vegetable and animal feed in some communities (Lutaladio and Ezumah, 1981). Some of the major cassava-producing countries in Africa include: Nigeria; Democratic Republic of Congo (DRC); Ghana; Tanzania; Mozambique; Uganda; Madagascar. To date, the vast majority of cassava varieties from southern, eastern and central (SEC) Africa are not genetically conserved in a formal conservation programme, but they do exist in farmers' fields and National Agricultural Research Systems (NARS) breeding programmes. Cassava germplasm in this region comprises both local farmer varieties (landraces) and cultivars that are derived from formal breeding either by the International Institute of Tropical Agriculture (IITA) or by the NARS. The effect of formal breeding on the genetic diversity of cassava is not known.

The genetic diversity of cassava in SEC Africa has most probably been influenced by early introductions, and also by major pandemics of cassava mosaic disease (CMD) and the current pandemic of cassava brown streak disease (CBSD) (Cours *et al.*, 1997; Otim-Nape *et al.*, 1997; Legg *et al.*, 2011). In addition, farmers retain their planting material from year to year and make selections according to their own preferences. Traditionally, farmers also exchange germplasm with friends and relatives in other regions. It is therefore a rational hypothesis that genetic diversity of cassava in the region has been moulded into patterns of diversity.

Genetic diversity of cassava in Africa has been assessed using a range of molecular markers (Fregene *et al.*, 2000; Fregene *et al.*, 2003; Zacarias *et al.*, 2004; Balyejusa Kizito *et al.*, 2005; Benesi, 2005). These assessments have resulted in varying outputs, involving different analysis methods and applications. Other key cassava-producing countries in the region, notably DRC, Madagascar and Rwanda, have not undertaken any significant assessment of cassava diversity. Detailed knowledge on genetic structure and variability of cassava germplasm is critically important for setting effective national and regional conservation priorities, for identifying 'gaps' in diversity and/or for defining breeding strategies. The objective of this study was to understand the nature,

extent, distribution and hierarchical organization of genetic variation that exists within the NARS breeding programmes of seven countries in SEC Africa: Tanzania; Uganda; Kenya; Rwanda; DRC; Madagascar; Mozambique. Genetic diversity was assessed using simple sequence repeat (SSR) markers that are frequent in eukaryotic genomes (Tautz and Renz, 1984), show high levels of allelic diversity and are relatively stable and co-dominant (Morgante and Olivieri, 1993; Toro *et al.*, 2009).

Materials and methods

Cassava germplasm

The germplasm included in this study consisted of cultivars developed through formal breeding and landraces (farmer varieties). Some of this germplasm was available within the NARS breeding programmes but the majority of the NARS undertook specific collections from farmers' fields to supplement the germplasm base for this study. It provides a reasonable representation of cassava germplasm in each country. Germplasm from each country was grown in a phenotyping trial (Kawuki *et al.*, 2011) from which young fresh leaf samples were collected on dry ice: Tanzania (270 genotypes); Uganda (268); Kenya (234); Rwanda (184); DRC (177); Madagascar (186); Mozambique (82). Of the 1443 plants sampled originally, 1401 cassava genotypes (848 landraces and 553 cultivars) were successfully assayed and used in the analysis. Genotype name, country of origin and classification into cultivars and landraces are given in Supplementary Material 1 (available online only at http://journals.cambridge.org) and summarized in Table 1. Total genomic DNA was extracted using a modified protocol described by Dellaporta *et al.* (1983). DNA quality was assessed on 0.8% agarose gels, quantified using the NanoDrop ND-1000 and then diluted to 50 ng/μl.

Microsatellite genotyping and allele calls

Cassava genotypes were analysed for polymorphism using 26 genomic SSR markers (SSRY5, SSRY9, SSRY12, SSRY19, SSRY21, SSRY38, SSRY51, SSRY52, SSRY59, SSRY63, SSRY64, SSRY69, SSRY100, SSRY102, SSRY110, SSRY135, SSRY147, SSRY148, SSRY151, SSRY155, SSRY161, SSRY169, SSRY171, SSRY181 and SSRY182 (Mba *et al.*, 2001) and NS911 (A. Zarate, personal communication)). Twenty-four of these SSRs had previously been used in a global diversity assessment of more than 2000 cassava genotypes and found to be highly informative and provided a similar overall population structure to that provided by 30 SSR loci (Hurtado *et al.*, 2008; M. Ferguson, personal

Table 1. Allelic richness and gene diversity of cassava germplasm used in this study

Country-landrace/ cultivar	No. of genotypes	Allelic richness[a]	Mean effective no. of alleles (A_e)	H_o[b]	Gene diversity[c] (H_e)	Shannon's index
Tan	270	128.4	2.89	0.61	0.60	1.15
Uga	268	121.8	2.63	0.54	0.56	1.08
Ken	234	125.4	3.05	0.59	0.61	1.19
Rwa	184	123.9	2.79	0.55	0.58	1.13
DRC	177	139.5	3.06	0.58	0.63	1.25
Mad	186	114.6	2.77	0.61	0.59	1.10
Moz	82	124.2	2.69	0.58	0.56	1.09
Tan-land	248	89.58	2.91	0.60	0.60	1.15
Uga-land	205	81.94	2.50	0.53	0.53	1.02
Ken-land	37	82.24	2.69	0.59	0.58	1.06
Rwa-land	47	88.36	2.74	0.59	0.59	1.11
DRC-land	104	93.51	2.85	0.58	0.61	1.19
Mad-land	134	84.47	2.72	0.60	0.58	1.08
Moz-land	73	85.29	2.61	0.58	0.55	1.05
Tan-cult	22	82.52	2.52	0.58	0.55	1.00
Uga-cult	63	87.84	2.66	0.56	0.57	1.09
Ken-cult	197	92.35	2.98	0.58	0.61	1.17
Rwa-cult	137	85.71	2.63	0.53	0.56	1.07
DRC-cult	73	97.53	3.06	0.57	0.63	1.24
Mad-cult	52	86.95	2.73	0.61	0.60	1.10
Moz-cult	9	85.68	2.37	0.57	0.55	0.95

Tan, Tanzania; Uga, Uganda; Ken, Kenya; Rwa, Rwanda; DRC, Democratic Republic of Congo; Mad, Madagascar; Moz, Mozambique; land, landrace; cult, cultivar.
[a] Allelic richness was computed using the rarefaction method as described by El Mousadik and Petit (1996). [b] H_o is the proportion of heterozygous individuals in a population. [c] Nei's unbiased estimator of gene diversity.

communication). These loci were furthermore selected based on single-locus amplification, high degree of polymorphism and reproducibility.

Amplifications with SSR primers were performed in 10 µl reactions containing 50 ng DNA, 1 pmol of each primer, 1 × *Taq* polymerase buffer, 2 mM MgCl$_2$, 0.2 mM dNTPs and 0.375 U *Taq* polymerase (New England Biolabs, Inc., Ipswich, MA). The PCR profile was 95°C for 2 min, followed by 30 cycles at 95°C for 30 s, 55–57°C for 1 min and 72°C for 1 min and a final extension at 72°C for 30 min. Different-sized amplification products at different loci and variable fluorescent labels (NED, 6-FAM, PET and VIC; MWG-Biotech) on the forward primer allowed multiplexing of amplicons from the same individual. Seven co-loading sets were optimized and used for the entire genotyping. For each co-loading set, 1–2 µl, depending on amplification efficiency, of the different amplicons were mixed and briefly vortexed. Aliquots of 1 µl of the mixture were added to 9 µl of a master mix containing 1 ml HiDi formamide and 12 µl GeneScan 500 LIZ size standard.

Amplicons were denatured (95°C, 3 min) and subjected to capillary electrophoresis using the ABI 3730 DNA sequencer (Applied Biosystems) and allele calls made using GeneMapper® software version 3.7 (Applied Biosystems® by Life Technologies, Carlsbad, CA). All genotyping was done at the Biosciences eastern and central

Africa (BecA) Hub in Nairobi, Kenya. The DNA of sample TMS 30 572 was used as a control. All genotypic data are provided in Table S1 (available online only at http://journals.cambridge.org).

Statistical analysis

Initially, genetic diversity statistics were calculated according to country of origin and according to 14 groups representing country-landrace/cultivar. These statistics included count of alleles per locus, allele frequency per locus, effective number of alleles (A_e) (Kimura and Crow, 1964), observed heterozygosity (H_o), expected heterozygosity (Nei's gene diversity; H_e) and Shannon's information index (Lewontin, 1972) calculated using POPGENE v1.32. H_e is defined by Nei (1973) as the probability that two alleles chosen at random from a population are different, and was computed using the formula $H_e = 1 - \Sigma p_{il}^2$, where p_{il} is the frequency of the *i*th allele at the *l*th locus. Allelic richness (number of alleles segregating in the population), compensating for differences in sample size, was calculated by adapting the rarefaction index as suggested by El Mousadik and Petit (1996). Allele frequency and the number of rare alleles, defined as having a frequency <0.05, were also computed on a per-country

basis using FSTAT version 2.9.3.2 (Goudet, 2001). Quantification of random mating and genetic differentiation was examined using F-statistics (Weir, 1990) in PowerMarker version 3.25 (Liu and Muse, 2005).

A non-model-based approach, principal component analysis (PCA), was applied to all genotypes according to country of origin, and landraces and cultivars separately, according to country of origin, to elucidate genetic relationships. For this analysis, any clone that had more than two missing SSRs ($\sim 10\%$) was omitted, leaving 1004 clones when landraces and cultivars were considered together, 616 when only landraces were analysed and 388 when cultivars were considered alone. The remaining missing data were replaced with the average value, and analysis conducted using Prcomp function in R (R Development Core Team, 2010).

Genetic relationships among landraces from different countries, cultivars from different countries, and landraces and cultivars from different countries were also calculated using the simple-matching Euclidean distance (DE) followed by cluster analysis using the weighted neighbour-joining algorithm in PowerMarker version 3.25 (Liu and Muse, 2005). Relationships among the groups were visualized as a dendrogram using Power-Marker version 3.25 (Liu and Muse, 2005).

In addition, a model-based approach was also applied to determine genetic structure. The number of populations (K) was determined with the software STRUCTURE using the admixture model (Pritchard et al., 2000). This Bayesian-based model assumes that each individual inherited some portion of its ancestry from each one of the K populations. Initially, the length of the burn-in period and Markov Chain Monte Carlo (MCMC) replications were set at 10,000 for $K = 1-17$ (number of putative populations in the model); this was repeated ten times. The ΔK statistic, which is based on the rate of change in the log probability of the data between successive K-values, was then used to detect the true number of K populations in the dataset (Evanno et al., 2005). The determined population K was rerun at a burn-in period and MCMC replications of 100,000 to verify its consistency. A membership coefficient of $Q > 0.9$ was used throughout.

Results

Allelic richness and gene diversity

Of the 1443 cassava cultivars and landraces genotyped for 26 SSR loci, 42 had more than 10% missing data and were excluded from this analysis. All genotypic data are provided in Table S1 (available online only at http://journals.cambridge.org). A total of 192 alleles

were detected, giving an average of 7.38 alleles per locus. All loci were polymorphic in all countries. Generally, rare alleles ($< 5\%$ frequency) were more prevalent in the studied germplasm than alleles of higher frequency (see Table S2 for actual allele frequencies and Table S3 for a summary, available online only at http://journals.cambridge.org). Alleles with frequencies > 0.3 were few. When only rare alleles ($< 5\%$ frequency) were analysed (Table S3, available online only at http://journals.cambridge.org), it was generally observed that they were more frequent in germplasm from the DRC, although landraces from Tanzania had the largest proportion of rare alleles (35.2%). The results further indicated that no rare alleles were observed in cultivars from Mozambique and that two loci (SSRY19 and SSRY110) were associated with higher numbers of rare alleles.

Gene diversity and allelic richness statistics among cassava germplasm from the seven countries and country-landrace/cultivar are presented in Table 1. Landraces harboured 186 alleles, whereas cultivars had 169. Allelic richness was very similar in pooled cultivars (88.37), as it was in pooled landraces (86.48) as was the mean A_e (cultivars (3.131) and landraces (3.146)), Nei's (1973) gene diversity (H_e) (cultivars (0.623) and landraces (0.633)) and Shannon's information index (cultivars (1.272) and landraces (1.261)). The overall mean H_o was 0.57.

Landraces from all countries had similar levels of diversity in terms of allelic richness, mean A_e, H_e and Shannon's information index, with the DRC having marginally more diversity for all parameters, followed by Tanzania. The lowest diversity for all parameters was observed in Ugandan landraces. In terms of the cultivars, cassava from the DRC once again had the highest values for allelic richness, A_e, H_e and Shannon's information index, followed by those from Kenya. The lowest values were observed in the Tanzanian and Mozambican cultivars.

Genetic variance distribution and genetic differentiation

Wright's F-statistics on all 1401 genotypes grouped into 14 populations (countries-landraces/cultivars) indicated a very slight deficit of heterozygotes within populations with $F_{IS} = 0.0084$. The overall F_{IT} value (0.1113) also indicated a slight deficit of heterozygotes in the total population. The F_{ST} value between all 14 populations (countries-landraces/cultivars) was 0.1038, indicating moderate differentiation. When seven landrace populations were considered, representing the seven countries, F_{IS} was slightly negative (-0.0053), basically indicating
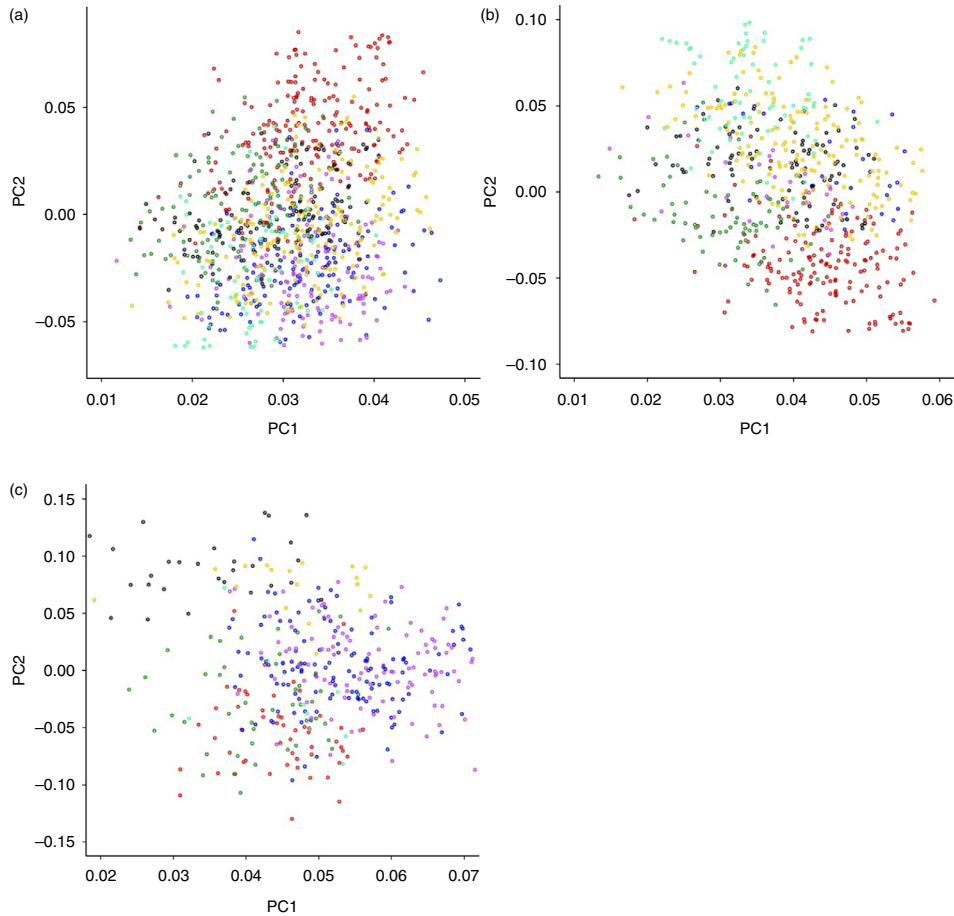
Fig. 1. (colour online) Principal component (PC) analysis shows some differentiation among cassava: (a) cultivars and land-races pooled, (b) landraces and (c) cultivars from the seven countries: Uganda (red); DRC (dark green); Rwanda (pink); Kenya (blue); Madagascar (black); Mozambique (light green); Tanzania (gold).

random mating, F_{ST} was 0.1040 and F_{IT} was 0.0992, indicating moderate differentiation among populations. When seven cultivars were considered, representing the seven countries, F_{IS} was slightly higher, but still very low (0.0298), F_{ST} was 0.0845 and F_{IT} was 0.1118, indicating some deviation from random mating and moderate differentiation among populations. The first two principal components (PCs) of PCA of 1004 genotypes accounted for 52 and 7% of the variation, respectively (Fig. 1(a)). This revealed a tendency for the germplasm from Uganda and DRC to be somewhat differentiated from the germplasm from other countries.

Table 2. Frequency-based Euclidean distance among the 848 landraces from the seven countries within southern, eastern and central Africa

	DRC	Ken	Mad	Moz	Rwa	Tan	Uga
DRC	0	0.378	0.368	0.377	0.300	0.313	0.326
Ken	0.378	0	0.252	0.343	0.263	0.195	0.290
Mad	0.368	0.252	0	0.370	0.305	0.220	0.309
Moz	0.377	0.343	0.370	0	0.391	0.254	0.408
Rwa	0.300	0.263	0.305	0.391	0	0.250	0.331
Tan	0.313	0.195	0.220	0.254	0.250	0	0.290
Uga	0.326	0.290	0.309	0.408	0.331	0.290	0
Mean	0.343	0.287	0.304	0.357	0.307	0.254	0.326

DRC, Democratic Republic of Congo; Ken, Kenya; Mad, Madagascar; Moz, Mozambique; Rwa, Rwanda; Tan, Tanzania; Uga, Uganda.

Table 3. Frequency-based Euclidean distance among the 553 cultivars from the seven countries within southern, eastern and central Africa

	DRC	Ken	Mad	Moz	Rwa	Tan	Uga
DRC	0	0.242	0.378	0.347	0.235	0.382	0.223
Ken	0.242	0	0.355	0.292	0.159	0.313	0.267
Mad	0.378	0.355	0	0.480	0.376	0.328	0.394
Moz	0.347	0.292	0.480	0	0.322	0.400	0.360
Rwa	0.235	0.159	0.376	0.322	0	0.354	0.273
Tan	0.382	0.313	0.328	0.400	0.354	0	0.413
Uga	0.223	0.267	0.394	0.360	0.273	0.413	0
Mean	0.301	0.271	0.385	0.366	0.286	0.365	0.321

DRC, Democratic Republic of Congo; Ken, Kenya; Mad, Madagascar; Moz, Mozambique; Rwa, Rwanda; Tan, Tanzania; Uga, Uganda.

Genetic relationships among landraces

The first two PCs accounted for 56 and 10% of the genetic variation in landraces from the seven countries (Fig. 1(b)). The greatest separation appeared to be in the germplasm from Uganda and DRC, to that from Mozambique. DEs among landraces on a country-of-origin basis are provided in Table 2. This shows that landraces from Tanzania are the most closely related to landraces from all other countries except for the DRC which is most closely related to Rwanda, followed by Tanzania. This relationship is not clearly depicted in a dendrogram. Landraces from Tanzania had the lowest mean genetic distance (0.218). Landraces from Mozambique and DRC are overall most distantly related to other countries with mean DEs of 0.306 and 0.294, respectively. Landraces from Mozambique were most distantly related to those from Uganda (0.408), followed by those from Rwanda (0.391). Landraces from the DRC were most distantly related to those from Kenya (0.378) and Mozambique (0.377).
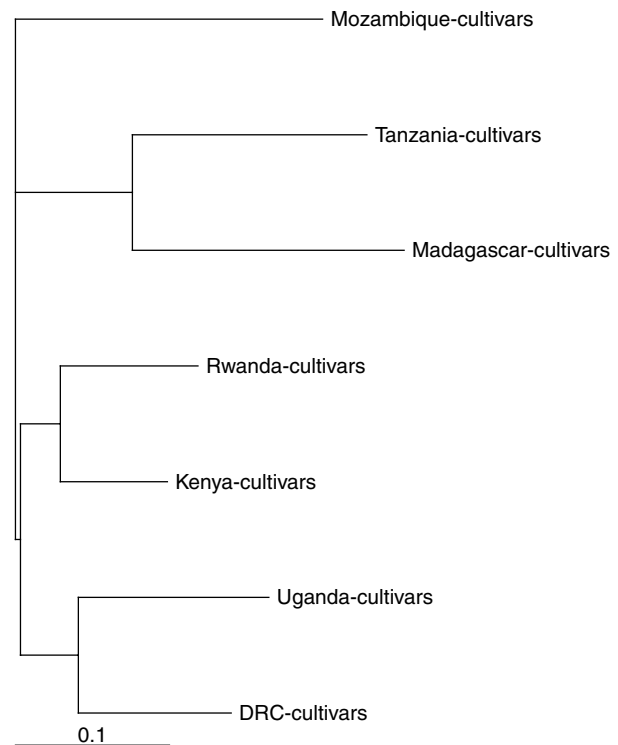
Genetic relationships among cultivars

Less differentiation was found among cultivars from different countries with the first two PCs accounting for 59 and 6% of the genetic variation (Fig. 1(c)). DE measures among cultivars from different countries are provided in Table 3, and are represented in a dendrogram (Fig. 2). Cultivars from Kenya, Uganda, DRC and Rwanda appear relatively similar, whereas those from Madagascar, Tanzania and Mozambique are more distinct. Madagascar had the highest mean DE (0.330), followed by Mozambique (0.314) and Tanzania (0.313). These were represented by 52, 9 and 21 cultivars each. Cultivars from Madagascar were most similar to those from Tanzania (0.328), and those from Tanzania were most similar to those from Kenya (0.313), followed by Madagascar. Cultivars from Mozambique were most

similar to those from Kenya (0.292), followed by Rwanda (0.322).

Genetic relationships among cultivars and landraces

The relationship between cultivars and landraces from different countries is best revealed through STRUCTURE analysis. The results from STRUCTURE analysis indicated that the 1401 individuals genotyped largely belonged to

Fig. 2. Genetic relationships among cultivars from different countries in southern, eastern and central Africa based on simple-matching Euclidean distance and the weighted neighbour-joining algorithm using PowerMarker version 3.25.
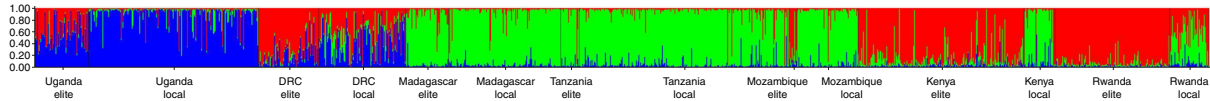
Fig. 3. (colour online) Assignment of population structure using the admixture model of the 1401 individuals to three subpopulations (subpopulation 1 = red; subpopulation 2 = green; subpopulation 3 = blue): 1 = Ugandan cultivars; 2 = Ugandan landraces; 3 = DRC cultivars; 4 = DRC landraces; 5 = Malagasy cultivars; 6 = Malagasy landraces; 7 = Tanzanian cultivars; 8 = Tanzanian landraces; 9 = Mozambican cultivars; 10 = Mozambican landraces; 11 = Kenyan cultivars; 12 = Kenyan landraces; 13 = Rwandan cultivars; 14 = Rwandan landraces.

three subpopulations, i.e. $K = 3$ (assignment of individuals to the three populations is presented in Fig. 3). ΔK values were $K = 2\,(24.982)$, $K = 3\,(33.601)$ and $K = 4\,(8.636)$. Cultivars from DRC, Mozambique, Kenya and Rwanda had over 70% of their individuals assigned to subpopulation 1. Nearly 40% of Ugandan cultivars were assigned to this subpopulation as well. Landraces from Madagascar, Tanzania, Mozambique and Kenya had over 70% of their individuals assigned to subpopulation 2. The majority of cultivars from Tanzania and Madagascar were also included in this group. A relatively similar number of individuals from the landraces from Rwanda were assigned between subpopulations 1 and 2. Over 80% of Ugandan landraces were assigned to subpopulation 3, together with 56.9% of Ugandan cultivars, while the landraces from the DRC were assigned between subpopulations 1 and 3.

Discussion

A major objective of this research was to study the extent of genetic diversity in cassava landraces and cultivars immediately available within the cassava breeding programmes of Tanzania, Uganda, Kenya, Rwanda, DRC, Madagascar and Mozambique, and to ascertain any patterns of variation. It was hypothesized that the independent introductions of cassava into the East African region (Jones, 1959), contrasting farmers' selection needs across the region, disease pandemics and germplasm movement, could have contributed to shape cassava into distinct groups or subpopulations. For this purpose, selectively neutral microsatellites that are suited for understanding population structure or history were used (Toro *et al.*, 2009). It is important to note that genetic models used for analysis are based on several assumptions, including random mating and sexual reproduction (Weir, 1990). Data presented in this research were generated from representative samples of landraces (farmer varieties) and cultivars (derived from formal breeding). Cassava is a clonally propagated, outcrossing crop that is propagated through stem cuttings. Flowering is variable and the number of pods and seeds limited. Silva *et al.* (2003) performed a quantitative analysis of mating system parameters of ethno varieties of cassava

using progeny arrays assayed for eight allozyme markers. They found the multilocus outcrossing rate (t_m) to be 0.915 ± 0.04, with a range from 0.69 to 1.00, revealing high levels of outcrossing with variable levels of selfing.

Cassava was one of the first crops to arrive in Africa from the east coast of Brazil. After several years of debate on its origin (Allem, 1994), molecular markers have provided strong evidence that cassava was probably domesticated from a single wild *Manihot* species (*M. esculenta* spp. *flabellifolia* Pohl) and that the crop originated from the southern Amazon basin (Olsen, 2004). It was from this region that cassava was exported to other parts of the world. The crop was first introduced to West Africa some time during the 1700 s (Jones, 1959; Carter *et al.*, 1992). Thereafter, it was quickly adopted and rapidly spread within the West African region. Cassava was subsequently introduced to the continent via the East African coastline in the 1750 s, after the French introduced cassava from Brazil to Mauritius (Jones, 1959). Thereafter, the crop was introduced to Madagascar in the early 19th century and to mainland Africa where it spread to various countries in the eastern, central and southern African region (Jones, 1959; Langlands, 1966). Cassava also reached Lake Tanganyika in Tanzania from West Africa via Congolese farmers and farmer-to-farmer spread (Carter *et al.*, 1992). Further spread was reinforced by administrators in the region who encouraged farmers to grow cassava and increased cultivation in the late 19th and 20th centuries. The natural spread of the crop was facilitated by its vegetative propagation, hardiness and viability of cuttings (Masumba, 2006). The introduction of cassava into East and West Africa thus represents a 'genetic bottleneck' whereby Africa received a portion of the genetic diversity present in the crop's centre of origin. In addition, cassava in Africa was isolated from the great diversity available in Brazil and from the majority of its wild relatives.

Genetic diversity

Moderate to low levels of genetic variation were observed in all countries. Although the number of alleles per locus was relatively high (see Table S2, available

online only at http://journals.cambridge.org), the majority of alleles occurred at a very low frequency (<5%), and therefore do not contribute much to heterozygosity. This is reflected in low values of Shannon's information index (between 0.95 and 1.25) and A_e (between 2.37 and 3.06). Much higher values have been noted in Guyana where landraces grown at one location had the same level of genetic diversity (Shannon's information index = 4.293) as a sample from the core collection at the International Centre for Tropical Agriculture (CIAT) (Shannon's information index = 4.289) (Elias et al., 2000). This indicates a genetic bottleneck in both landraces and cultivars in SEC Africa. If a stepwise mutation model is assumed for the SSRs used in this study, then several alleles appear to be missing or at low frequency (see Table S2, available online only at http://journals.cambridge.org), providing further evidence of a genetic bottleneck.

Marginal differences were observed in allelic richness among the seven countries, with the highest number of segregating alleles being registered in the DRC (96) and lowest in Uganda (85). It appears that this difference largely stems from the number of rare alleles. This indicates that cassava from the DRC is slightly more genetically diverse than cassava from the other six surveyed countries. It is possible that, due to its location, the DRC may have received more cassava germplasm from West Africa, as well as from SEC Africa, contributing to the gene pool. New gene combinations are generated through unmanaged sexual reproduction between different cultivars within cassava stands in farmers' fields, but this does not increase SSR diversity.

Allelic richness, A_e, H_e and Shannon's information index were very similar in landraces and cultivars. Vellvé (1993) reported that formal breeding significantly reduced genetic diversity in European agriculture. Other studies have strongly demonstrated increased wheat genetic variability with formal breeding (Maccaferri et al., 2003). To date, it appears that there is limited genetic variability in both landraces and cultivars in SEC Africa, and that there is scope for increasing diversity through the introduction of germplasm from the centre of origin and diversity of cassava in Brazil.

Partitioning of variation

A high level of mean H_o ($H_o = 0.57$) was expected due to the outbreeding nature of cassava, as was an F_{IS} close to zero, indicating random mating. This is consistent with reports of some degree of inbreeding in cassava which is likely to occur spontaneously in farmers' fields with progeny being selected from volunteer seedlings. In addition, if outcrossing occurs between two clonally propagated plants, it would appear as inbreeding. Farmers tend to grow a number of plants of the same clonally propagated genotype in close proximity; this may tend to overestimate levels of true inbreeding. In addition, it appears that the cassava inflorescence architecture and seed dispersal mechanisms limit long-distance gene flow, but favour mating among plants in close proximity, which would also tend to reduce heterozygosity. Besides inbreeding, high frequencies of null alleles (Brookfield, 1996) have been reported to cause heterozygosity deficits within populations. The estimate on the frequency of null alleles was computed using the method of Brookfield (1996), which assumes that null homozygotes are present in the sample. The estimated frequency of null alleles was low (0.04), suggesting that inbreeding (or outcrossing among identical plants) could account for random mating. Elsewhere, heterozygosity deficits have been reported for both cultivated and wild relatives of cassava (Olsen and Schaal, 2001).

Genetic differentiation

In general, if F_{ST} is 0.05–0.15, then differentiation in allele frequencies between populations is considered moderate. Here, F_{ST} among all 14 populations when landraces and cultivars were considered was 0.1038, indicating moderate differentiation. Traditionally, it is customary for farmers to exchange germplasm frequently, and often across substantial distances, including across international borders. This would tend to reduce genetic differentiation among countries.

Slightly greater differentiation was observed, in terms of deviation from random mating, in landraces ($F_{ST} = 0.1040$), compared with cultivars ($F_{ST} = 0.0845$). This is likely to reflect the common origin from IITA-Nigeria of many of the cultivars now present in a number of countries, tending to reduce population differentiation. The levels of differentiation observed here are slightly higher than those observed by others. Balyejusa Kizito et al. (2005), though working with fewer cassava genotypes (ranging between 19 and 195), observed no appreciable gene differentiation among cassava genotypes from Uganda (East Africa) and Ghana (West Africa) ($F_{ST} = 0.04$) and between cassava genotypes from Ghana and Tanzania (East Africa) ($F_{ST} = 0.05$). Peroni et al. (2007) analysed sweet (76 entries) and bitter (60 entries) cassava and observed higher genetic diversity within populations (higher within sweet cassava), and moderate genetic differentiation ($R_{ST} = 0.057$) among the two populations. Several other studies using neutral microsatellites observed no appreciable gene differentiation within different cassava populations (Asante and Offei, 2003; Lokko et al., 2006; Siqueira et al., 2009).

Genetic relationships of landraces from different countries

PCA showed a tendency for landraces from a specific country to group together (Fig. 1(b)) and some loose structure among the countries. DE revealed that landraces from Tanzania have a pivotal position in the region, being most closely related to germplasm from all other countries apart from that from the DRC which were most closely related to landraces from Rwanda, followed by Tanzania (Table 2). Tanzania is centrally located in the region, having borders with Rwanda, Uganda, Kenya and Mozambique. This is likely to reflect traditional practices of exchanging cassava germplasm, even across international borders. The relative distance of Mozambique and DRC from other germplasm in the region could reflect their geographical peripheral location in the region, with DRC probably receiving a relatively greater proportion of germplasm from West Africa. Germplasm from West Africa has been shown on the basis of 1190 single nucleotide polymorphism markers and a small number of genotypes to be somewhat genetically different from that from SEC Africa (Ferguson *et al.*, 2012). The close relationship of landraces from the DRC to those of Rwanda indicate substantial exchange between these countries. It is interesting to note that landraces from Madagascar were not more distinct from the rest of the region, due to its relative geographical isolation; however, Madagascar was struck by a severe epidemic of CMD in 1934–1936. This led to the almost total elimination of local varieties (Cours *et al.*, 1997). Many of the 'landraces' may be relatively recent introductions from mainland Africa. A similar epidemic struck Uganda in the 1990s where the severity of symptoms commonly led to the almost complete elimination of the most vulnerable varieties (Legg *et al.*, 2011). This epidemic spread rapidly to the Great Lakes region and beyond acquiring the status of 'pandemic' (Otim-Nape *et al.*, 1997). Many areas of SEC Africa are currently experiencing another pandemic, of CBSD (Legg *et al.*, 2011). This is likely to affect the nature and frequency of particular landraces in the region. The lack of conservation of these landraces in any kind of germplasm repository in the region is of particular concern in the face of disease pandemics.

Genetic relationships of cultivars from different countries

The observed close relationship among cultivars from Kenya, Uganda, Rwanda and DRC from PCA and DE can be explained by the fact that many of these improved varieties grown in these countries have been acquired through the IITA through the 'International Nurseries' initiative. This germplasm often has a West African genetic background and/or a Latin American genetic background if CIAT parental lines were used in hybridization schemes at the IITA.

The isolation of cultivars from Madagascar, and their slight proximity to cultivars from Tanzania, may reflect efforts to combat, particularly, CMD in these countries. As a result of the CMD epidemic in Madagascar, a broad diversity of genotypes were screened for resistance to the disease at Alaotra Agricultural Research Station, Madagascar from 1935. Adequate resistance to CMD was found in Javanese varieties and intra-specific crosses (Cours, 1951; Cours-Darne, 1968). In addition, inter-specific crosses with ceara rubber (*Manihot glaziovii*) and other species were conducted to obtain increased resistance. From the 1930s, a similar but unconnected programme at Amani, Tanzania, conducted extensive hybridization between cassava and its wild relatives (Nichols, 1947). Examples of these hybridization schemes included: (1) utilization of *Manihot dichotoma* Ule, *M. glaziovii* Muell.-Arg. and *Manihot catingae* Ule for sources of disease resistance (Hillocks and Jennings, 2003) and (2) *Manihot melanobasis* Muell.-Arg. and *Manihot saxicola* Lang. for disease resistance and protein enhancement (Jennings, 1959). These latter two species are now considered to be forms of *M. esculenta* subsp. *flabellifolia*. High levels of resistance to CMD were obtained from *M. glaziovii* × *M. esculenta* crosses (Jennings, 1994). This cassava improvement programme continued during the 1940s and 1950s, supported by the British Government and under the auspices of the East African Agriculture and Forestry Research Organization. In the 1950s, a number of selections derived from crosses were distributed to research stations across East Africa (Nichols, 1947), and some of this material has recently been retrieved for use in current disease resistance breeding programmes. These initiatives generated new genetic variability in the region, and could explain the relative isolation of Malagasy and Tanzanian cultivars from those from Uganda, Rwanda and DRC. The relative isolation of Mozambique was somewhat surprising; however, this may be due to its peripheral location in the region, and the fact that it was only represented by nine cultivars.

Genetic relationships of cultivars and landraces from different countries

When the robust Bayesian model-based analysis was done on all the 1401 cassava individuals, they were assigned into three subpopulations. Most cultivars from DRC, Mozambique, Kenya and Rwanda, and nearly 40%

of cultivars from Uganda were assigned to the same subpopulation, together with a proportion of landraces from DRC and Rwanda. This subpopulation is likely to be characterized by a higher frequency of alleles from West Africa. This was hypothesized earlier to explain the relatively high genetic distance of landraces from the DRC to other landraces in the region. Cultivars derived from IITA-Nigeria are also more likely to have a West African genetic background with more of its characteristic allelic composition and frequency.

The second subpopulation consisted of cultivars and landraces of more East African origin. This included landraces from Madagascar, Tanzania, Mozambique, Kenya and a proportion from Rwanda. This subpopulation also included cultivars from those two countries, Tanzania and Madagascar, which had early breeding programmes, including inter-specific breeding programmes (Jennings, 1994), and disseminated cassava varieties to neighbouring countries (Nichols, 1947). This indicates a relatively close relationship and introgression between cultivars and landraces from these countries, and less influence of West African-derived germplasm from IITA-Nigeria. This is not necessarily the case in Kenya and Mozambique, where the landraces and cultivars appear more distinct, indicating a lack of introgression between landraces and cultivars, possibly due to the maintenance of integrity of introductions from IITA-Nigeria.

It is interesting to note that STRUCTURE analysis of cultivars and landraces together revealed the relative difference between Ugandan cultivars and landraces. This is likely to reflect the breeding programme in Uganda where IITA-Nigeria varieties are frequently introgressed with local landraces. This, together with its intermediate position between West African-influenced DRC and East Africa, could explain why germplasm from this country is largely allocated to a separate subpopulation.

Conclusions

In conclusion, the results from this research highlight that cassava landraces and improved cultivars available within the NARS of Tanzania, Kenya, Uganda, DRC, Madagascar, Mozambique and Rwanda contain moderate genetic variability resulting from a dominance of a few alleles with the majority of alleles having low frequency. This indicates a genetic bottleneck in the region and the need to increase diversity through the importation of germplasm from South America if substantial genetic gain is to be made in breeding. The majority of variation exists within individuals as opposed to among countries. This reflects the heterozygous nature of cassava. PCA, genetic distance and STRUCTURE analysis revealed subtle and rather complex patterns of genetic diversity

of cassava germplasm of SEC Africa. This appears to be shaped by gene flow from West Africa into the landraces from the DRC; introduction and preservation of IITA-improved varieties in the breeding programmes of DRC, Rwanda, Kenya and Mozambique; independent introductions of cassava into East Africa and its subsequent dissemination throughout the region which may have contributed to the pivotal role of Tanzanian landraces in the region. This and the close relationship of cultivars and landraces in Tanzania and Madagascar are likely to reflect early independent breeding programmes in these countries, which involved germplasm selections and intra- and inter-specific crosses with landraces. This is the first large assessment of genetic variation in the SEC African region, and the first study of variation in the NARS of DRC, Madagascar and Rwanda. Information generated from this study has immediate application to justify and guide a regional cassava genetic resource conservation strategy, to identify gaps in cassava diversity in the region in relation to global diversity, to contribute information to a global cassava reference set and to guide parental selection during breeding.

Acknowledgements

References

Allem AC (1994) The origin of *Manihot esculenta* Crantz (Euphobiaceae). *Genetic Resources and Crop Evolution* 41: 133–150.

Asante IK and Offei SK (2003) RAPD-based genetic diversity study of fifty cassava (*Manihot esculenta* Crantz) genotypes. *Euphytica* 131: 113–119.

Benesi IRM. (2005) Characterisation of Malawian cassava germplasm for diversity, starch extraction and its native and modified properties. PhD Thesis, University of the Free State, Bloemfontein, South Africa, pp. 199.

Balyejusa Kizito E, Bua A, Fregene M, Egwang T, Gullberg U and Westerbergh A (2005) The effect of cassava mosaic disease on the genetic diversity of cassava in Uganda. *Euphytica* 146: 45–54.

Brookfield JFY (1996) A simple new method for estimating null frequency from heterozygous deficiency. *Molecular Ecology* 5: 453–455.

Carter SE, Fresco LO, Jones PG and Fairbairn JN (1992) *An Atlas of Cassava in Africa: Historical, Agroecological and Demographic Aspects of Crop Distribution*. Cali: CIAT.

Cours G (1951) Le manioc à Madagascar. Mémoires de l'Institut Scientifique de Madagascar. *Série B Biologie Végétale* 3: 203–400.

Cours G, Fargette D, Otim-Nape GW and Thresh JM (1997) The epidemic of cassava mosaic virus disease in Madagascar in the 1930s–1940s: lessons for the current situation in Uganda. *Tropical Science* 37: 238–248.

Cours-Darne G. (1968) Improving cassava in Africa. The Abidjan Conference: Agricultural Research Priorities for Economic Development in Africa, vol. 2, pp. 330–339.

Dellaporta SJ, Wood J and Hicks JB (1983) A plant DNA mini preparation. Version II. *Plant Molecular Biology Reporter* 1: 19–21.

El Mousadik A and Petit RJ (1996) Chloroplast DNA phylogeography of the argan tree of Morocco. *Molecular Ecology* 5: 547–555.

Elias M, Panaud O and Robert T (2000) Assessment of genetic variability in a traditional cassava (*Manihot esculenta* Crantz) farming system using AFLP markers. *Heredity* 85: 219–230.

Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620.

Fregene MA, Suarez M, Mkumbira J, Kulembeka H, Ndedya E, Kulaya A, Mitchel S, Gullberg U, Rosling H, Dixon AGO, Dean R and Kresovich A (2003) Simple sequence repeat marker diversity in cassava landraces: genetic diversity and differentiation in an asexually propagated crop. *Theoretical and Applied Genetics* 107: 1083–1093.

Ferguson ME, Hearn SJ, Close TJ, Wanamaker S, Moskal WA, Town CD, de Young F, Marri PR, Rabbi IY and de Villiers EP (2012) Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. *Theoretical and Applied Genetics* 124: 685–695.

Fregene M, Bernal A, Duque M, Dixon A and Tohme J (2000) AFLP analysis of African cassava (*Manihot esculenta* Crantz) germplasm resistant to the cassava mosaic disease (CMD). *Theoretical and Applied Genetics* 100: 678–685.

Goudet J. (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available at http://www.unil.ch/izea/softwares/fstat.html

Hillocks RJ and Jennings DL (2003) Cassava brown streak disease: a review of present knowledge and research needs. *International Journal of Pest Management* 49: 225–234.

Hurtado P, Ospina C, Marin J, Buitrago C, Castelblanco W, Correa A, Alfonso P, Barrera E, Gutierrez I, Fregene M, Hearne S, Ferguson M, Alves A, Fortes-Ferreira C, de Vicente C (2008) Assessment of the diversity in global cassava genetic resources based on simple sequence repeat (SSR) markers. In: Fauquet CM (ed). *Cassava: Meeting the challenges of the new millenium. Proceedings of the First Scientific Meeting of the Global Cassava Partnership*, 21–25 July 2008, Ghent, Belgium.

Jennings DL (1959) *Manihot melanobasis* Muell. Arg. – a useful parent for cassava breeding. *Euphytica* 8: 157–162.

Jennings DL (1994) Breeding for resistance to African cassava mosaic geminiviruses in East Africa. *Tropical Science* 34: 110–122.

Jones WO (1959) *Manioc in Africa*. Stanford, CA: Standard University Press, pp. 315.

Kawuki RS, Ferguson M, Labuschagne M, Herselman L, Orone J, Ralimanana I, Bidiaka M, Lukombo S, Kanyange MC, Gashaka G, Mkamilo G, Gethi J and Obiero H (2011) Variation in qualitative and quantitative traits of cassava germplasm from selected national breeding programmes in sub-Saharan Africa. *Field Crops Research* 122: 151–156.

Kimura M and Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics* 49: 725–738.

Langlands BW (1966) Cassava in Uganda 1860–1920. *Uganda Journal* 30: 211–218.

Legg JP, Jeremiah SC, Obiero HM, Maruthi MN, Ndyetabula I, Okao-Okuja G, Bouwmeester H, Bigirimana S, Tata-Hangy W, Gashaka G, Mkamilo G, Alicai T and Kumar LP (2011) Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Research* 159: 161–170.

Lewontin RC (1972) The apportionment of human diversity. *Evolutionary Biology* 6: 381–398.

Liu K and Muse SV (2005) PowerMarker. Integrated analysis environment for genetic marker data. *Bioinformatics* 21: 2128–2129.

Lokko Y, Dixon A, Offei S, Danquah E and Fregene M (2006) Assessment of genetic diversity among African cassava *Manihot esculenta* Crantz accessions resistant to the cassava mosaic virus disease using SSR markers. *Genetic Resources and Crop Evolution* 53: 1441–1453.

Lutaladio NB and Ezumah HC (1981) Cassava leaf harvesting in Zaire. In: Tenry ER, Oduri KA and Cavaness F (eds) *Tropical Root Crops: Research Strategies for the 1980's*. Ohawa: IDRC, pp. 134–136.

Maccaferri M, Sanguineti MC, Donini P and Tuberosa R (2003) Microsatellite analysis reveals a progressive widening of the genetic basis in the elite durum wheat germplasm. *Theoretical and Applied Genetics* 107: 783–797.

Masumba E.A. (2006) Genetic diversity and field performance of cassava (*Manihot esculenta*) landraces commonly grown in eastern, southern and lake zones, Tanzania. A Dissertation submitted in partial fulfilment of the requirements for the degree of Master of Science in Agriculture of Sokoine University of Agriculture, Morogoro, Tanzania.

Mba REC, Stephenson P, Edwards K, Melzer S, Nkumbira J, Gullberg U, Apel K, Gale M, Tohme J and Fregene M (2001) Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. *Theoretical and Applied Genetics* 102: 21–31.

Morgante M and Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. *Plant Journal* 3: 175–182.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences USA* 70: 3321–3323.

Nichols RFW (1947) Breeding cassava for virus resistance. *East African Agricultural Journal* 12: 184–194.

Nweke F, Spencer D and Lynam J (2001) *The Cassava Transformation: Africa's Best Kept Secret*. East Lansing, MI: Michigan State University Press.

Olsen KM (2004) SNPs, SSRs and inferences on cassava origin. *Plant Molecular Biology* 56: 517–526.

Olsen KM and Schaal BA (2001) Microsatellites variation in cassava (*Manihot esculenta*, Euphorbiaceae) and its wild relatives: further evidence for a southern Amazonian origin of domestication. *American Journal of Botany* 88: 131–142.

Otim-Nape GW, Bua A, Thresh JM, Baguma Y, Ogwal S, Semakula GN, Acola G, Byabakama B and Martin A (1997) *Cassava mosaic virus disease in Uganda: the current pandemic and approaches to control*. Chatham: Natural Resources Institute.

Peroni N, Kageyama PY and Begossi A (2007) Molecular differentiation, diversity, and folk classification of "sweet" and "bitter" cassava (*Manihot esculenta*) in Caicara and Caboclo management systems (Brazil). *Genetic Resources and Crop Evolution* 54: 1333–1349.

Pritchard JK, Stephens M and Donnelly P (2000) Inferences of population structure from multilocus genotypic data. *Genetics* 155: 945–959.

R Development Core Team (2010) R: A language and environment for statistical computing. Available at http://www.Rproject.org (verified 3 November 2010).

Silva R, da Bandel G and Martins PS (2003) Mating system in an experimental garden composed of cassava (*Manihot esculenta* Crantz) ethnovarieties. *Euphytica* 134: 127–135.

Siqueira MVBM, Queiroz-Silva JR, Bressan EA, Borges A, Pereira JCK, Pinto JG and Veasey AE (2009) Genetic characterisation of cassava (*Manihot esculenta*) landraces in Brazil assessed with simple sequence repeats. *Genetics and Molecular Biology* 32: 104–110.

Tautz D and Renz M (1984) Simple sequences are ubiquitous components of the eukaryotic genomes. *Nucleic Acids Research* 12: 4127–4138.

Toro AA, Fernàndez J and Caballero A (2009) Molecular characterization of breeds and its use in conservation. *Livestock Science* 120: 174–195.

Vellvé R (1993) The decline of diversity in European agriculture. *Ecologist* 23: 64–69.

Weir BS (1990) *Genetic Data Analysis: Methods for Discrete Population Genetic Data*. Sunderland, MA: Sinauer Associates, Inc.

Zacarias AM, Botha AM, Labuschagne MT and Benesi IRM (2004) Characterisation and genetic distance analysis of cassava (*Manihot eculenta* Crantz) germplasm from Mozambique using RAPD fingerprinting. *Euphytica* 138: 49–53.