ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2016) **14**(3); 200–210
doi:10.1017/S1479262115000246

Diversity and genetic structure of cassava landraces and their wild relatives (*Manihot* spp.) in Colombia revealed by simple sequence repeats

E. Tovar[1]*, J. L. Bocanegra[1], C. Villafañe[2], L. Fory[3], A. Velásquez[3], G. Gallego[3] and R. Moreno[1]

[1]*The Alexander von Humboldt Biological Resources Research Institute, Bogotá, Colombia,*
[2]*Ministry of the Environment and Sustainable Development, Bogotá, Colombia and*
[3]*International Center for Tropical Agriculture (CIAT), Palmira, Colombia*

Received 15 January 2015; Revised 5 May 2015; Accepted 15 May 2015 – First published online 24 August 2015

Abstract

Understanding the genetic composition and population structure of plant species at a molecular level is essential for the development of adequate strategies aimed at enhancing the conservation and use of their genetic resources. In addition, such knowledge can help to plan ahead for a scenario under which wild and cultivated species come into contact with their genetically modified (GM) counterpart(s). Using ten simple sequence repeat markers, we genotyped 409 samples pertaining to the species in the *Manihot* genus known to occur in Colombia, i.e. cassava (*Manihot esculenta)* and its wild relatives *Manihot brachyloba*, *Manihot carthaginensis* and *Manihot tristis*. High genetic variation was observed in all the species ($H_E = 0.212–0.603$), with cassava showing highest diversity. Most of the genetic variation was found within species populations. Our results suggest that outcrossing events among populations occur much more frequently in *M. tristis* and *M. esculenta*, and particularly so in the latter, where the exchange of varieties among local farmers plays a key role in maintaining and introducing new genetic diversity. The occurrence of gene flow within and among populations of *Manihot* species in Colombia becomes relevant in a biosafety context, where gene flow from GM cassava, if introduced to the country, might have detrimental effects on the structure and dynamics of populations of wild species. The baseline information on the genetic diversity and structure of the four Colombian species that we have presented here provides a first and indispensable step towards the development of targeted interventions necessary to preserve their genetic resources.

Keywords: cassava; crop wild relatives; genetic diversity; GMO; *Manihot*; simple sequence repeat

Introduction

The *Manihot* genus includes 98 species distributed through the Neotropic (Rogers and Appan, 1973). In Colombia, the wild species *Manihot brachyloba* Müll.Arg., *Manihot carthaginensis* (Jacq.) Müll.Arg., *Manihot tristis* Müll.Arg. are known to occur alongside cultivated cassava (*Manihot esculenta* Crantz). Cassava is an allogamous species native to South America (Allem, 1994; Olsen and Schaal, 1999). It ranks fourth among the principal carbohydrate sources in the tropic, and is the sixth most important crop for the human diet, providing food to approximately one billion people from 105 countries. The wide distribution

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

and importance of this crop relates to its resilience to abiotic stress factors and the low requirements for its production, allowing its cultivation in low-fertility soils (Food and Agriculture Organization (FAO), 2008; Montagnac *et al*., 2009).

The combination of cassava's adaptive traits, its importance in terms of food security of rural populations around the world, and its promising industrial applications, have attracted attention to the generation of genetically modified varieties with beneficial traits, from high β-carotene content to the quality of the starch (Ihemere *et al*., 2006; Sayre *et al*., 2011). While research in Colombia on the development of genetically modified (GM) cassava is ongoing, no variety has been released in the country to date. However, if GM cassava were to be released in Colombia there is high probability of introgression of genetically modified organisms (GMO) genes into its wild relatives and traditional cultivars through gene flow, owing to the crop's predominant allogamous nature with pollination mediated by insects (Kawano, 1980), high outcrossing rates from 60 to 100% (Kawano, 1980; Meireles da Silva R *et al*., 2003), and the possibility of producing interspecific hybrids (Nassar, 2003). To allow the development of a proper risk management strategy, the establishment of a genetic baseline of the wild relatives of cassava in Colombia is indispensable. The characterization of species at a molecular level is an essential step of programmes aimed at developing conservation and use strategies of biological resources. Just as artificial selection has played a key role in the transformation of wild plants into their cultivated forms, similarly the introduction of GM varieties, could through gene flow and introgression of alien genes, influence the evolutionary and population dynamics of their wild relatives and traditional cultivated forms. Determining where and how (*ex situ* or *in situ*) to conserve is crucial to maintain the genetic variation of the species. In this study, ten simple sequence repeat (SSR) markers were used to assess the genetic diversity and population structure of the Colombian species *M. brachyloba*, *M. carthaginensis*, *M. tristis* and *M. esculenta*.

Materials and methods

Plant material

A total of 409 samples were collected over a period of three years (2009–2011) for the four taxa, taking into consideration previous reports on the distribution of *Manihot* species in Colombia, and ecological niche modelling (potential distribution) based on those reports. The geographical coverage included 12 of the 32 departments of Colombia (Table 1).

For the study, 180 controls provided by the Genetic Resources Unit of the International Center for Tropical Agriculture (CIAT) were used, corresponding to: 31 of *M. esculenta* ssp. *flabellifolia* and 32 of *M. tristis* from Brazil; 18 of *M. carthaginensis* of Colombia; and 99 of *M. esculenta* from 13 countries of Latin America, Africa

Table 1. Location and number of samples of *Manihot* species collected in Colombia

Species	Department	Predominant natural region	Number of samples
Manihot brachyloba	Antioquia	Andean	40
	Caldas	Andean	24
	Casanare	Orinoquia	10
	Choco	Pacific	23
	Cundinamarca	Andean	18
	Meta	Orinoquia	26
	Vichada	Orinoquia	16
	Total		157
Manihot carthaginensis	Antioquia	Andean	38
	La Guajira	Caribbean	15
	Magdalena	Caribbean	30
	Total		83
Manihot esculenta	Amazonas	Amazon	18
	Antioquia	Andean	16
	Caldas	Andean	2
	Casanare	Orinoquia	37
	Meta	Orinoquia	3
	Vaupes	Amazon	52
	Total		128
Manihot tristis	Vichada	Orinoquia	41
	Total		41
	Grand total		409

https://doi.org/10.1017/S1479262115000246 Published online by Cambridge University Press

and Asia. The high number of controls was used to confirm the taxonomic status of the collected samples and to identify possible hybrids.

DNA extraction and SSR amplification

DNA extraction from leaf tissue was carried out using the DNeasy® Plant Mini Kit system from QIAGEN (Germantown, Maryland, USA). Total DNA quality was verified through visualization by electrophoresis in 0.8% agarose gels stained with SYBR® Safe, Life Technologies (Carlsbad, California, USA). DNA concentration was determined by quantification using the BioPhotometer Eppendorf® (Hamburg, Germany) spectrophotometer.

The samples were genetically characterized by means of the following ten SSRs developed and studied by Mba *et al.* (2001) and Raji *et al.* (2009a): SSRY164, SSRY161, SSRY59, SSRY21, SSRY175, SSRY5, SSRY20, C283Y, MEESR15, and MEESR60. Amplification of the SSRs was carried out using a PTC-100TM thermal cycler (Programmable Thermal Controller; Bio-Rad Laboratories, Inc. (Hercules, California, USA)). The PCR profile included a first cycle of denaturation at 94°C for 5 min, followed by 35 cycles at 94°C for 30 s, annealing T° (specific for each pair of primers) for 25 s, and 72°C for 30 s, and a final extension cycle at 72°C for 10 min. The reaction mixture, with a final volume of 13 µl, contained 20 ng of DNA of the sample, 1X buffer (Tris–HCl 100 mM, KCl 500 mM), 1.5 mM of $MgCl_2$, 0.18 mM of each deoxynucleoside triphosphate (dNTP), 0.24 µM of each primer, and 0.5 U of Taq polymerase. The PCR products were visualized in 6% acrylamide gels, 7.5 M of urea, stained with silver nitrate (Bassam *et al.*, 1991). Allele sizes were visually determined through comparison with the 10 bp DNA ladder (Invitrogen™ Corporation, Carlsbad, CA, USA).

Statistical analysis

Principal Coordinate Analyses (PCoA) based on the Nei genetic distance matrices were undertaken with the *cmdscale* function available in version 1.3 of the *gstudio* package (Dyer, 2014) for R software for statistical computing version 3.1 (R Core Team, 2014). Hierarchical cluster analysis was carried out with the *hclust* function using the unweighted pair group method with arithmetic mean. We used the Kelley–Gardner–Sutcliffe penalty function (KGS) (Kelley *et al.*, 1996) for hierarchical trees implemented in the *kgs* function of the *maptree* package (White and Gramacy, 2012) to suggest the optimal number of clusters or populations in the data set.

The FSTAT program version 2.9.3 (Goudet, 2001) was used to calculate genetic diversity parameters, average number of alleles per locus (A), observed heterozygosity (H_O), expected heterozygosity or average genetic diversity (H_E), total heterozygosity (H_T), coefficient of differentiation (G_{ST} and F_{ST}) and inbreeding coefficient (F_{IS}).

Analysis of the distribution of variation at different levels within the data set of each species was calculated with the ARLEQUIN program version 3.5 (Excoffier and Lischer, 2010) using the analysis of molecular variance (AMOVA) with 1000 permutations. Likewise, correlations between genetic (F_{ST}) and geographical distance (calculated from the geographical coordinates) matrices of the populations of the species were assessed using ARLEQUIN by means of Mantel tests with 10,000 permutations.

We calculated an indirect estimate of gene flow based on the equation $N_m = (1 - G_{ST})/4G_{ST}$ (Slatkin and Barton, 1989), where N_m is the estimated number of migrants per generation.

Polymorphism information content (PIC) values (Botstein *et al.*, 1980), as a measure of the informativeness of a genetic marker, were calculated for each locus according to the following formula:

$$\mathrm{PIC} = 1 - \sum_{i=1}^{n} P_i^2 - 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} P_i^2 P_j^2,$$

where P_i = allele frequency of the marker and n = number of different alleles.

Results

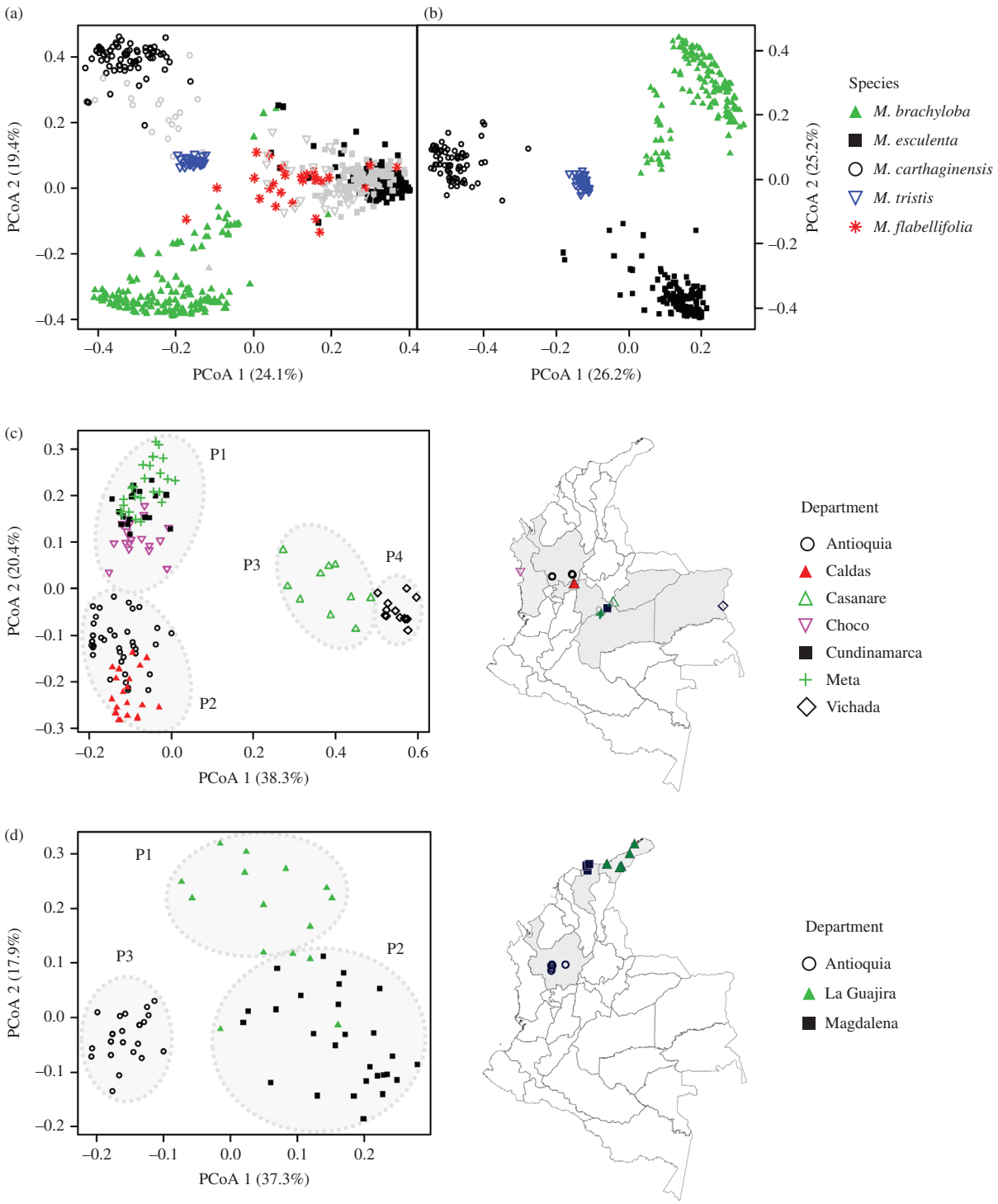
The ten SSRs used in this study, across all 409 samples, registered a very high (0.816) average value of PIC, ranging from 0.629 (MEESR60) to 0.903 (SSRY161). Lower values were obtained at individual species level (Supplementary Table S1, available online). *M. esculenta* showed the highest average PIC value (0.622), with values of individual loci ranging between 0.221 (C283Y) and 0.868 in (SSRY175). The PIC values for *M. brachyloba* ranged between 0.090 (MEESR15) and 0.911 (SSRY5), with an average of 0.569. *M. carthaginensis* yielded an average value of 0.420 and a range between 0.060 (MEESR15) and 0.785 (SSRY21). The lowest average PIC value of 0.233 was obtained for *M. tristis*, which relates to the low number of alleles per locus observed in the species: in four cases only one allele was found, resulting in a PIC value of 0. Nevertheless, the highest value was 0.806 for locus SSRY5 with 12 alleles.

Genotyping of all 409 individuals with the ten SSRs yielded a total of 195 different alleles, with an average of 19.5 alleles per locus. The lowest number of alleles (10) was observed for locus MEESR60, and the highest number (31) for locus SSRY161. At species level (Supplementary Table S1, available online), highest allele richness was observed for *M. esculenta* (107 alleles;

https://doi.org/10.1017/S1479262115000246 Published online by Cambridge University Press

range from 6 (C283Y) to 18 (SSRY175)). Followed by *M. brachyloba* (95 alleles; range from 2 (SSRY21) to 22 (SSRY5)) and *M. carthaginensis* (67 alleles; range from 2 (MEESR15) to 13 (SSRY161)). Lowest alleles richness was found for *M. tristis* with a total of 34 alleles, ranging from 1 (SSRY59, SSRY175, SSRY20, C283Y) to 12 (SSRY5).

To explore the relationships among the different *Manihot* species considered in this study, we first carried out a PCoA based on all individuals, including the control samples. The first axis (PCoA 1) of Fig. 1(a), explaining 24.1% of the total variation in data, separates the different species relatively well. However, there is a clear overlap between the



https://doi.org/10.1017/S1479262115000246 Published online by Cambridge University Press

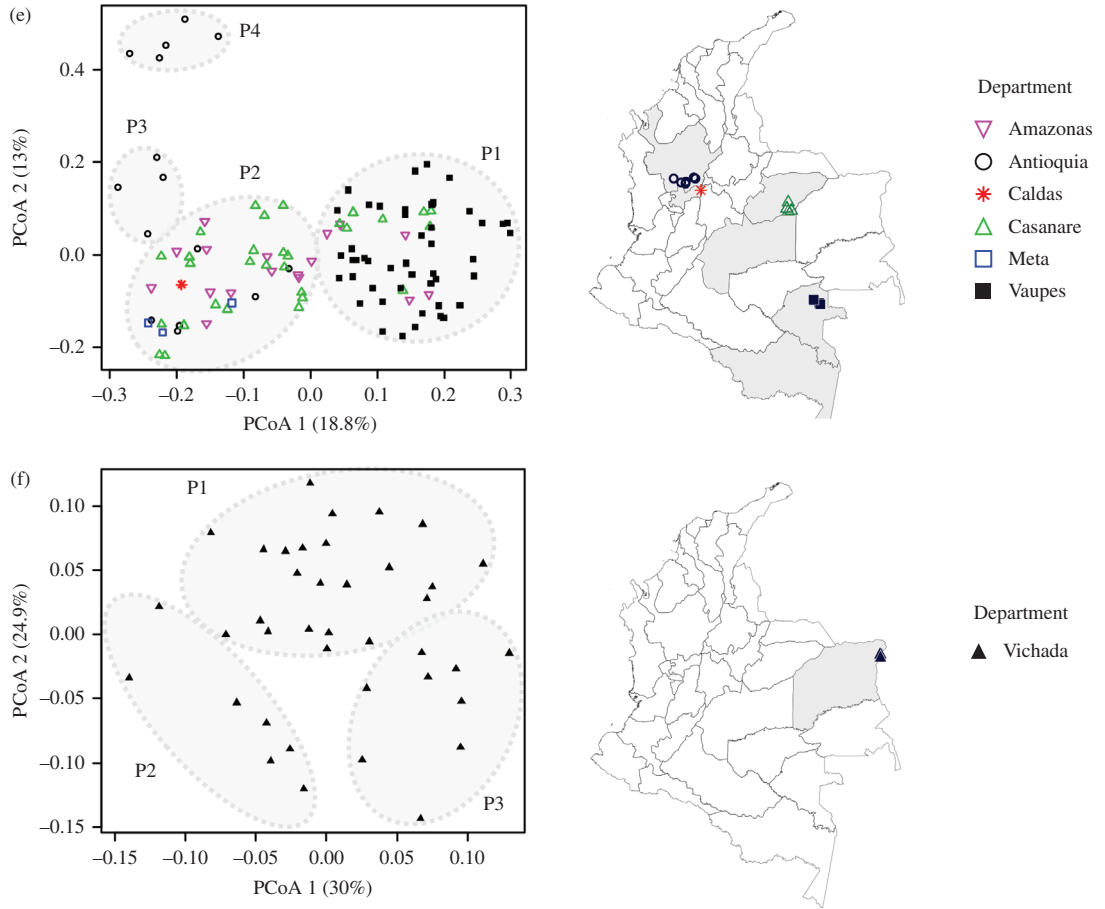


Fig. 1. Principal Coordinate Analysis (PCoA) and sample distribution of the *Manihot* species studied. (a) All species + controls. Controls of the species studied in grey colour, (b) *Manihot* species of Colombia, (c) *Manihot brachyloba*, (d) *Manihot carthaginensis*, (e) *Manihot esculenta* and (f) *Manihot tristis*. Shaded areas in the graph indicate populations suggested by the KGS analysis.

projection of individuals pertaining to *M. esculenta* ssp. *flabellifolia* (controls), *M. esculenta* (samples and controls) and *M tristis* samples from Brazil (controls). Also the separation of cultivated and wild species was relatively well marked, with a tendency for cultivated species to be projected on the right and wild on the left hand side of the graph. The wild species studied are separated along the second axis (PCoA 2), which explains 19.4% of the total genetic variation. In a PCoA scatterplot without control samples (Fig. 1(b)), the samples are clearly divided into four discrete groups, corresponding to the species studied.

At the species level, the KGS analysis suggested the separation of the *M. brachyloba* individuals in four clusters (Fig. 1(c)): the first one composed of samples from the departments of Choco, Cundinamarca and Meta (P1); the second one of samples from Antioquia and Caldas (P2); the third one of individuals from Casanare (P3); and the fourth one of specimens from Vichada department (P4). The first PCoA axis, explaining 38.3% of the total genetic variation, differentiates the ensemble of populations P1 and P2 from both populations P3 and P4. The differentiation of populations P1 and P2 is apparent along the second axis, which explained 20.4% of the total genetic variation.

For *M. carthaginensis*, the KGS analysis identified three clusters (Fig. 1(d)): a first group composed of samples from the department of Guajira (P1), a second one of individuals from Guajira and Magdalena (P2), and a third one of samples from Antioquia (P3). Populations P1 and P2 are separated from population P3 along the first PCoA axis, which explained 37.3% of the total variation. Separation of populations P1 and P2 is evident along the second axis, which explains 17.9% of the total variation.

The *M. esculenta* were grouped into four clusters (Fig. 1(e)) but unlike the two previous species, clear separation of the individuals according to their department of origin was not observed. Population P1 included individuals from Amazonas, Casanare and Vaupes; P2 individuals from Antioquia, Amazonas, Caldas, Casanare and Meta; and P3 and P4, samples from Antioquia.

https://doi.org/10.1017/S1479262115000246 Published online by Cambridge University Press

The first PCoA axis explains 18.8% of total genetic variation and differentiates population P1 from the remaining ones, while the latter ones are separated along the second axis, explaining 13% of total variation.

Three populations were identified in the *M. tristis* samples, all originating from the department of Vichada (Fig. 1(f)). Total variation explained by the first and second PCoA axes was 30 and 24.9%, respectively.

Of all species studied, *M. esculenta* showed the highest average values of observed ($H_O = 0.564$), expected ($H_E = 0.603$) and total heterozygosity ($H_T = 0.792$). The different populations identified in this species showed no major differences in H_E scores, which ranged from 0.554 (P1) to 0.693 (P3). Accordingly, high genetic diversity levels were detected for all the loci, ranging between 0.375 (C283Y) and 0.789 (SSRY164). For *M. brachyloba* average H_O, H_E and H_T were 0.255, 0.406 and 0.708, respectively (Supplementary Table S1, available online). Within populations, average H_E ranged between 0.261 (P4) and 0.493 (P3); and within loci, it varied between 0.105 (MEESR15) and 0.574 (SSRY5). In *M. carthaginensis*, average H_O, H_E and H_T values were 0.244, 0.355 and 0.511. Population P1 was least diverse ($H_E = 0.215$) and population P3 was most diverse ($H_E = 0.463$). H_E per locus ranged between 0.663 (C283Y) and 0.056 (MEESR15). Compared with the other species studied, *M. tristis* evidenced the lowest values of diversity parameters, with average values for H_O, H_E and H_T of 0.191, 0.212 and 0.258, respectively. These low scores are mainly because four of the ten SSR markers were monoallelic, and because two additional markers detected variability only in one of the three populations. Hence, only four markers revealed heterozygosity in all the populations established. The highest H_E value was found for SSRY5 (0.783). P3 was the most diverse population ($H_E = 0.252$), while P2 was the least diverse (0.172).

With respect to genetic structure, for *M. brachyloba* (Table 2) the results of an AMOVA indicated that most of the genetic variation (62.8%) occurred within the populations identified with KGS analysis, and 37.2% among them. The high F_{ST} value (0.372) (Supplementary Table S1, available online) suggested large genetic variation between the populations considered. Similarly, a high and positive F_{IS} value was observed (0.437). The value of the estimated migrants (N_m) was low (0.337), which could be related to isolation by distance, as confirmed by the strong and significant correlation between genetic and geographical distance ($r = 0.71$, $P = 0.04$) calculated by a Mantel test. Also for *M. carthaginensis*, most of the genetic variation was found within the populations (65.0%; Table 2), but still more than one third (35.0%) of the variation was attributed to the differentiation among populations. A high and positive F_{IS} value (0.358) (Supplementary Table S1, available online) and a low N_m (0.567) were also observed. While the Mantel test indicated a very strong correlation between genetic and geographic distance ($r = 0.99$), it was not significant ($P = 0.16$); hence, there was no isolation effect derived from distance. With respect to *M. esculenta*, only a minor of genetic variation (13.8%; Table 2) was observed among populations, and the large majority within populations (86.3%). The average F_{IS} for this species was low (0.049) (Supplementary Table S1, available online) and the N_m was higher (1.059), compared with the previous species. A Mantel test showed a very weak and non-significant correlation ($r = -0.02$, $P = 0.36$) between the genetic and geographical distances of the populations. For *M. tristis*, moderate genetic differentiation among populations was noticed ($F_{ST} = 0.178$) (Supplementary Table S1, available online). Most of the genetic variation (82.23%; Table 2) was found within populations. As for *M. esculenta*, F_{IS} observed in *M. tristis* was low (0.085) (Supplementary Table S1, available online) while the

Table 2. AMOVA results for individuals of *Manihot* species based on ten SSR

Species	Partitioning	degrees of freedom (df)	Sum of squares	Variance components	Percentage of variation
Manihot brachyloba	Among populations	3	257.084	1.240	37.18
	Within populations	310	645.455	2.094	62.82
	Total	313	902.539	3.334	
Manihot carthaginensis	Among populations	2	89.863	0.852	34.99
	Within populations	163	256.312	1.583	65.01
	Total	165	346.175	2.435	
Manihot esculenta	Among populations	3	75.443	0.464	13.75
	Within populations	252	732.608	2.909	86.25
	Total	255	808.051	3.373	
Manihot tristis	Among populations	2	13.265	0.227	17.77
	Within populations	79	82.870	1.052	82.23
	Total	81	96.135	1.280	

https://doi.org/10.1017/S1479262115000246 Published online by Cambridge University Press

N_m was the highest (1.703) among all the species, which could be associated with the geographical closeness of the populations.

Discussion

Overall, the markers used in this study evidenced high PIC values, demonstrating their ability to discriminate among species and populations at intraspecific level. In the specific case of *M. esculenta*, the PIC values obtained here (0.221 (C283Y)–0.868 (SSRY175), $\bar{X} = 0.622$) are within the range of previous studies. In a study of cassava germplasm, Kawuki *et al.* (2009) reported PIC values within a range of 0.358–0.759 with an average of 0.571. Similarly, Turyagyenda *et al.* (2012) found an average PIC value of 0.611 for cassava landraces from Uganda.

Relationship of control samples and Manihot *species of Colombia*

The PCoA carried out with the four species analysed, along with the controls (Fig. 1(a)), showed a trend of wild species samples from Colombia to form discrete groups, while the cultivated one was grouped with control samples corresponding to the wild species *M. esculenta* ssp *flabellifollia* and *M. tristis* from Brazil. A clearer separation of all the four Colombian *Manihot* species was observed in the PCoA without the controls (Fig. 1(b)), confirming their status as separate species. Similar observations of wild *Manihot* species showing a tendency to separate from cultivated varieties have been made in previous studies (Roa *et al.*, 1997; Elias *et al.*, 2000, 2004). This pattern is consistent with the hypothesis of a limited, or even unique, domestication event in a restricted area, followed by the dispersion of cultivated phenotypes (Elias *et al.*, 2000). Phylogenetic studies conducted by Schaal *et al.* (1997), Olsen and Schaal, 1999, 2001) and Olsen (2004) have confirmed Allem's (1994) hypothesis that suggests *M. esculenta* ssp. *flabellifolia* occurring in southern Amazonia, as the wild progenitor of cultivated cassava. The close relationship observed between the species *M. esculenta* and *M. esculenta* ssp. *flabellifolia* in this study supports this hypothesis.

The clustering of *M. tristis* control samples from Brazil with those of *M. esculenta* and *M. esculenta* ssp. *flabellifolia* suggests that the *M. tristis* specimens from Colombia and Brazil may be different enough to be classified as two different species, or at least as two distinct sub-species. On the other hand, according to Allem (1994) and Allem *et al.* (2001), the *M. tristis* species from the Brazilian region is considered a synonym of *M. esculenta* ssp. *flabellifolia* placing it in the *M. esculenta* complex, in agreement with what was observed in the PCoA (Fig. 1(a)).

Genetic diversity and gene flow

We found high levels of intraspecific genetic diversity for all the species studied. While the lowest scores of all genetic diversity parameters were observed for *M. tristis* ($H_E = 0.212$, $H_T = 0.258$, $A = 3.4$), these may still be considered high when taking into account the number of samples studied, which was the lowest of all species here considered. The considerably high H_E value (0.603) obtained for *M. esculenta*, the highest of all species analysed, is in agreement with previous findings by Kawuki *et al.* (2009) who studied germplasm from Asia, Africa and America and found a H_E value of 0.615 for samples from the Americas. Another study on germplasm from different countries in Africa yielded a similar H_E value of 0.630 (Raji *et al.*, 2009b). Similarly, Turyagyenda *et al.* (2012) reported an average H_E value of 0.661 in landrace varieties from Uganda. The average number of alleles per locus in cassava obtained in this study (10.7) was higher than that of previous studies. Montero-Rojas *et al.* (2011) reported 7.15 alleles per locus in a study on cassava from Puerto Rico. Fregene *et al.* (2003) found alleles number equivalent to 6.0 and 5.2 for landraces of Colombia and Brazil, respectively. Likewise, Turyagyenda *et al.* (2012) observed an average number of 5.9 alleles per locus in landrace varieties from Uganda. The foregoing evidences a high allelic richness in the loci sampled within the Colombian populations.

Regarding the population structure of *M. brachyloba* and *M. carthaginensis* observed in the PCoA (Fig. 1(c and d)), samples tended to group according to their geographical origin and their relative proximity. The congruence between the geographical distribution of populations and their genetic differentiation is generally related to constrained gene flow between them (Schaal *et al.*, 1998). *M. brachyloba* displayed a highly significant pattern of isolation by distance, meaning that the reduced or almost inexistent gene flow between populations at larger distances may have led to the stochastic differentiation among them due to a genetic drift (Ellstrand and Elam, 1993). The high F_{ST} value, along with the low gene flow (N_m), observed in *M. brachyloba*, is consistent with the possibility that genetic drift has been the main modelling force of the current genetic structure of the species. In the case of *M. carthaginensis*, although the strong correlation ($r = 0.99$) between the geographical and genetic distances of the populations was not significant, similar

https://doi.org/10.1017/S1479262115000246 Published online by Cambridge University Press

F_{ST} and N_m values were found as for *M. brachyloba*, which could suggest that the process of genetic differentiation by distance of *M. carthaginensis* populations is still ongoing. Despite the fact that F_{IS} values were high and positive for the previous two species, indicating high levels of inbreeding within the populations identified by the KGS analysis, it is important to take into account that many of the individuals within these groups were separated by long distances, in some cases hundreds of kilometres. However, if there were no gene flow between individuals of *M. brachyloba* and *M. carthaginensis* it probably would have an impact on the heterozygosity of these; nevertheless, they were high for both species. Our results would suggest the occurrence of gene flow among individuals of previous species but at a much smaller scale than the proposed for the analysis, occurring mainly between neighbouring individuals, thus implying the predominance of an open reproductive system in these crop wild relatives.

Originating from more compact, geographically closer populations, the *M. tristis* samples seemed to be well connected by gene flow both among and within populations, as suggested by the high number of migrants ($N_m = 1.703$) and the low level of inbreeding ($F_{IS} = 0.085$) observed, respectively. To assess the effect of geographic distance of individuals within populations on F_{IS} estimator, we calculated anew the inbreeding coefficient for *M. brachyloba* and *M. carthaginensis*, this time establishing populations based on their geographical origin and closeness. Results revealed a decrease in the F_{IS} value, obtaining 0.244 and 0.265, respectively. These results suggest that the relationships among individuals and populations of species, mediated by gene flow, may show different patterns depending on the spatial scale they are assessed.

In *M. esculenta*, the heterozygotic nature of the species and its elevated levels of gene flow among and within the populations were confirmed by the F_{ST}, F_{IS}, and N_m values, consistent with observations in the PCoA scatter plot (Fig. 1(e)), where projections of samples from different regions of the country tended to cluster together, without forming apparent sub-populations. The existence of elevated gene flow in cassava has been previously demonstrated in genetic diversity studies conducted in different countries around the world, (Fregene *et al.*, 2003; Siqueira *et al.*, 2009; Montero-Rojas *et al.*, 2011; Kawuki *et al.*, 2013). The high variability and gene flow in cassava could be associated mainly with two factors. First, there is the allogamous nature of the species. Even though cassava varieties are essentially vegetatively propagated, the production of voluntary seedlings generated from sexual reproduction is frequently incorporated by farmers in their crop gene pool, thus favouring the recombination and production of new genotypes which are subsequently subjected to natural and artificial selection pressures. This traditional practice helps to maintain genetic diversity and counters genetic erosion (Elias *et al.*, 2001; Pujol *et al.*, 2005). Second, there is the exchange of cassava varieties among local farmers, which can play a key role in promoting and preserving the genetic variability of cassava populations. This factor has been shown to have a strong impact on the diversity of landraces, as has been reported in previous studies with different ethnic groups (Salick *et al.*, 1997; Elias *et al.*, 2000; Sambatti *et al.*, 2001). In addition, the great interest of local farmers to acquire new landraces and the fact that they seldom discard unproductive varieties with the argument that these may have a higher productivity in other environments, leads to a high varietal diversity on their farms (Elias *et al.*, 2000). Clearly, the exchange of varieties and the incorporation of volunteer seedlings strongly contribute to maintain and augmenting genetic variability among and within crops on smallholder farmers' fields.

GM cassava and the conservation of landraces and wild relatives

The potential impact of transgene introgression on the genetic structure and population dynamics of wild relatives, if GM cassava cultivated were to be permitted in Colombia, would be determined essentially by time, since the effect and the level of fixation of genes in populations are related to the adaptive advantage they confer to the species. According to the results of our study, the gene flow between genetically improved cassava varieties and wild relatives is a feasible event. Although little scientific information exists about the exchange of genes among species in the *Manihot* genus, many plant breeders have shown that cassava can be crossed with several of its wild congeners, thus overcoming reproductive barriers (Rogers and Appan, 1973; Byrne, 1984; Asiedu *et al.*, 1992; Wanyera *et al.*, 1992; Blair *et al.*, 2007). In addition, based on molecular evidence, Duputié *et al.* (2007) confirmed the existence in nature of fertile hybrids between cassava and its wild relative *M. esculenta* ssp. *flabellifolia* at several locations in French Guiana. In the case of *M. carthaginensis*, a frequently used species in cassava breeding (Chavarriaga *et al.*, 2004), natural hybrid populations of both species have been found in Africa, based on morphological and isozyme marker studies (Wanyera *et al.*, 1992; Lefevre and Charrier, 1993). The uncertainty associated with the possibility of gene flow from GM cassava to its wild relatives and the potential effect thereof, calls for the development of strategies to preserve the current gene

https://doi.org/10.1017/S1479262115000246 Published online by Cambridge University Press

pool of wild *Manihot* species in Colombia if GM cassava were to be released. To achieve this establishing the genetic diversity and population structure of wild species, which we have attempted to do in this study, is an indispensable step. Our results indicate the existence of high genetic diversity in wild *Manihot* species in Colombia, which are significant taking into consideration the small size of the populations. Aside from potential future threats of pollution with transgenes, also the current population sizes of the wild species make them more sensitive to losing genetic variation, e.g. through the effects of the random genetic drift and natural selection (Milligan *et al.*, 1994). Thus, actions should be undertaken in the short-term in order to preserve these species' genetic resources.

Finally, the practice of exchange of cassava landraces among local farmers, combined with the allogamous nature of the species, could itself constitute an effective mechanism to facilitate gene flow, not only within populations but also among populations separated by large distances. This may be an efficient manner to maintain the variability within the species. However, within a biosafety context this practice may also bear risks since it could be the ideal mean for the dispersion of transgenes. In addition, the absence of a genetic barrier between GM cassava and traditional landraces, along with the wide distribution of the latter, are factors that may increase the risk of contact between wild and landrace populations contaminated with transgenes.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1479262115000246

Acknowledgements

The authors wish to thank the Ministry of the Environment and Sustainable Development, the International Center for Tropical Agriculture (CIAT), and the Alexander von Humboldt Biological Resources Research Institute, for their cooperation and support to the Biosafety Line in Colombia. The authors express sincere gratitude to Evert Thomas for reviewing and improving substantially the text, and for his useful comments and suggestions. We are also grateful to Francisco Castro for his great cooperation during the expeditions carried out. A grateful acknowledgement is given to Myriam Cristina Duque and Juan Guillermo Perez for their valuable cooperation during the development of the study. And we thank all those who, in one way or another, contributed to this study.

Conflict of interest

The authors declare that they have no conflict of interest.

References

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

Allem AC, Mendes RA, Salomão AN and Burle ML (2001) The primary gene pool of cassava (*Manihot esculenta* Crantz subspecies *esculenta*, Euphorbiaceae). *Euphytica* 120: 127–132.

Asiedu R, Hahn SK, Bai KV and Dixon AGO (1992) Introgression of genes from wild relatives into cassava. In: Akoroda MO and Arene OB (eds) *Proceedings of the 4th Triennial Symposium of the International Society for Tropical Root Crops – Africa Branch*. Nigeria: ISTRC-AB/IDRC/CTA/IITA, pp. 89–91.

Bassam B, Caetano G and Gresshoff P (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. *Analytical Biochemistry* 190: 80–83.

Blair MW, Fregene MA, Beebe SE and Ceballos H (2007) Marker assisted selection in common beans and cassava. In: Guimaraes E (ed) *Marker-Assisted Selection. Current Status and Future Perspectives in Crops, Livestock, Forestry and Fish*. Italy: Food and Agriculture Organization (FAO), pp. 81–116.

Botstein D, White RL, Skolnick M and Davis RW (1980) Construction of a genetic linkage map in the man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32: 314–331.

Byrne D (1984) Breeding cassava. *Plant Breeding Reviews* 2: 73–134.

Chavarriaga P, Prieto S, Herrera CJ, López D, Bellotti A and Tohme J (2004) Screening transgenics unveils apparent resistance to hornworm (*E. ello*) in the non-transgenic, African cassava clone 60444. In: Alves A and Tohme J (eds) *Adding Value to a Small-farmer Crop. Proceedings of the 6th International Scientific Meeting of the Cassava Biotechnology Network*. Cali: CIAT, p. 4.

Duputié A, David P, Debain C and McKey D (2007) Natural hybridization between a clonally propagated crop, cassava (*Manihot esculenta* ssp. *esculenta*) and a wild relative in French Guiana. *Molecular Ecology* 16: 3025–3038.

Dyer RJ (2014) *gstudio*: Analyses and functions related to the spatial analysis of genetic marker data. R package version 1.3. http://CRAN.R-project.org/package=gstudio (accessed July 2014).

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.

Elias M, Penet L, Vindry P, McKey D, Panaud O and Robert T (2001) Unmanaged sexual reproduction and the dynamics of genetic diversity of a vegetatively propagated crop plant, cassava (*Manihot esculenta* Crantz), in a traditional farming system. *Molecular Ecology* 10: 1895–1907.

Elias M, Santos Mühlen G, McKey D, Roa AC and Tohme J (2004) Genetic diversity of traditional south american landraces of cassava (*Manihot esculenta* Crantz): an analysis using microsatellites. *Economic Botany* 58: 242–256.

Ellstrand NC and Elam DR (1993) Population genetic consequences of small population size: implications for plant

conservation. *Annual Review of Ecology, Evolution, and Systematics* 24: 217–242.

Excoffier L and Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.

Food and Agriculture Organization (FAO) (2008) Cassava for food and energy security. http://www.fao.org/news room/en/news/2008/1000899/index.html (accessed 14 August 2013).

Fregene M, Suarez M, Mkumbira J, Kulembeka H, Ndedya E, Kulaya A, Mitchel S, Gullberg U, Rosling H, Dixon A and Kresovich S (2003) Simple sequence repeat (SSR) diversity of cassava (*Manihot esculenta* Crantz) landraces: genetic diversity and differentiation in a predominantly asexually propagated crop. *Theoretical and Applied Genetics* 107: 1083–1093.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet (1995).

Ihemere U, Arias-Garzon D, Lawrence S and Sayre RT (2006) Genetic modification of cassava for enhanced starch production. *Plant Biotechnology Journal* 4: 453–465.

Kawano K (1980) Cassava. Chapter 13. In: Fehr WR and Hadley HH (eds) *Hybridization of Crop Plants*. Madison: American Society of Agronomy, pp. 225–233.

Kawuki RS, Ferguson M, Labuschagne M, Herselman L and Kim DJ (2009) Identification, characterization and application of single nucleotide polymorphisms for diversity assessment in cassava (*Manihot esculenta* Crantz). *Molecular Breeding* 23: 669–684.

Kawuki RS, Herselman L, Labuschagne MT, Nzuki I, Ralimanana I, Bidiaka M, Kanyange MC, Gashaka G, Masumba E, Mkamilo G, Gethi J, Wanjala B, Zacarias A, Madabula F and Ferguson ME (2013) Genetic diversity of cassava (*Manihot esculenta* Crantz) landraces and cultivars from southern, eastern and central Africa. *Plant Genetic Resources: Characterization and Utilization* 11: 170–181.

Kelley LA, Gardner SP and Sutcliffe MJ (1996) An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally-related subfamilies. *Protein Engineering* 9: 1063–1065.

Lefevre FA and Charrier S (1993) Isozyme diversity within African *Manihot* germplasm. *Euphytica* 66: 73–80.

Mba REC, Stephenson P, Edwards K, Melzer S, Mkumbira 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: 22–31.

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

Milligan BG, Leebens-Mack J and Strand AE (1994) Conservation genetics: beyond the maintenance of marker diversity. *Molecular Ecology* 12: 844–855.

Montagnac JA, Davis CR and Tanumihardio SA (2009) Nutritional value of cassava for use as a staple food and recent advances for improvement. *Comprehensive Reviews in Food Science and Food Safety* 8: 181–194.

Montero-Rojas M, Correa AM and Siritunga D (2011) Molecular differentiation and diversity of cassava (*Manihot esculenta*) taken from 162 locations across Puerto Rico and assessed with microsatellite markers. *AoB Plants* 2011: plr010. doi:10.1093/aobpla/plr010.

Nassar N (2003) Gene flow between cassava (*Manihot esculenta* Crantz), and wild relatives. *Genetics and Molecular Research* 2: 334–347.

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

Olsen KM and Schaal BA (1999) Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. *Proceedings of the National Academy of Sciences of the United States of America* 96: 5586–5591.

Olsen KM and Schaal BA (2001) Microsatellite 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.

Pujol B, David P and McKey D (2005) Microevolution in agricultural environments: how a traditional Amerindian farming practice favours heterozygosity in cassava (*Manihot esculenta* Crantz. *Euphorbiaceae). Ecology Letters* 8: 138–147.

R Core Team (2014) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org (accessed July 2014).

Raji AAJ, Anderson JV, Kolade OA and Ugwu CD (2009*a*) Gene-based microsatellites for cassava (*Manihot esculenta* Crantz): prevalence, polymorphisms, and cross-taxa utility. *BMC Plant Biology* 9: 118.

Raji AAJ, Fawole I, Gedil M and Dixon AGO (2009*b*) Genetic differentiation analysis of African cassava (*Manihot esculenta*) landraces and elite germplasm using amplified fragment length polymorphism and simple sequence repeat markers. *Annals of Applied Biology* 155: 187–199.

Roa AC, Maya MM, Duque MC, Tohme J, Allem AC and Bonierbale MW (1997) AFLP analysis of relationships among cassava and other *Manihot* species. *Theoretical and Applied Genetics* 95: 741–750.

Rogers DJ and Appan SG (1973) Manihot *and* Manihotoides *(Euphorbiaceae). A Computer Assisted Study*. Flora Neotropica, Monograph No. 13. New York: Hafner Press.

Salick J, Cellinese N and Knapp S (1997) Indigenous diversity of cassava: generation, maintenance, use and loss among the Amuesha, Peruvian upper Amazon. *Economic Botany* 51: 6–19.

Sambatti JBM, Martins PS and Ando A (2001) Folk taxonomy and evolutionary dynamics of cassava: a case study in Ubatuba. *Brazil. Economic Botany* 55: 93–105.

Sayre R, Beeching JR, Cahoon EB, Egesi C, Fauquet C, Fellman J, Fregene M, Gruissem W, Mallowa S, Manary M, Maziya-Dixon B, Mbanaso A, Schachtman DP, Siritunga D, Taylor N, Vanderschuren H and Zhang P (2011) The biocassava plus program: biofortification of cassava for sub-saharan Africa. *Annual Review of Plant Biology* 62: 251–272.

Schaal B, Carvalho LJCB, Prinzie T, Olsen K, Hernandez M, Cabral G and Moeller D (1997) Phylogenetic relationships and genetic diversity in *Manihot* species. *African Journal of Root and Tuber Crops* 2: 147–149.

Schaal BA, Hayworth DA, Olsen KM, Rauscher JT and Smith WA (1998) Phylogeographic studies in plants: problems and prospects. *Molecular Ecology* 7: 465–474.

Siqueira MVBM, Queiroz-Silva JR, Bressan EA, Borges A, Pereira KJC, Pinto JG and Veasey EA (2009) Genetic characterization of cassava (*Manihot esculenta*) landraces in Brazil

assessed with simple sequence repeats. *Genetics and Molecular Biology* 32: 104–110.

Slatkin M and Barton NH (1989) A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43: 1349–1368.

Turyagyenda LF, Kizito EB, Ferguson ME, Baguma Y, Harvey JW, Gibson P, Wanjala BW and Osiru DSO (2012) Genetic diversity among farmer-preferred cassava landraces in Uganda. *African Crop Science Journal* 20: 15–30.

Wanyera NMW, Hahn SK and Aken'ova ME (1992) Introgression of ceara rubber (*Manihot glaziovii* Muell-Arg) into cassava (*M. esculenta* Crantz): a morphological and electrophoretic evidence. In: Akoroda MO (ed) *Proceedings of Root Crops for Food Security in Africa*. Ibadan: Africa Branch, pp. 125–130.

White D and Gramacy RB (2012) *maptree*: Mapping, pruning, and graphing tree models. R package version 1.4-7. http://CRAN.R-project.org/package=maptree. (accessed July 2014).