Molecular differentiation and diversity of cassava (*Manihot esculenta*) taken from 162 locations across Puerto Rico and assessed with microsatellite markers

Maria Montero-Rojas, Ana María Correa and Dimuth Siritunga*

Department of Biology, University of Puerto Rico, P.O. Box 9000, Mayaguez, PR 00680, USA

Received: 20 January 2011; Returned for revision: 8 March 2011; Accepted: 19 March 2011; Published: 24 March 2011

Citation details: Montero-Rojas M, Correa AM, 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

Abstract

**Background and aims**

Knowledge of germplasm diversity and its genetic characterization is an invaluable asset in crop improvement as well as conservation strategies. Although an important crop and present in the Caribbean since the 15th century, the origins of today’s cassava in the Caribbean are poorly understood. Genetic variability and long growing cycles provide a unique set of challenges when researching cassava, and it is in part due to these challenges that little knowledge regarding the genetic diversity of cassava in Puerto Rico exists.

**Methodology**

In order to evaluate the genetic diversity of cassava in Puerto Rico, 162 samples with unknown genetic background were collected from different townships of the island and were analysed by 33 single sequence repeat markers. For comparative purposes, 23 accessions of the Puerto Rican cassava germplasm collection were also evaluated.

**Principal results**

Our results show that the genetic diversity ($H_E$) of unknown cassava samples (0.7174) is slightly higher than in the current Puerto Rican cassava collection (0.6996). The observed proportion of heterozygotes ($H_O$) was higher in the Puerto Rican cassava collection compared with the unknown samples. An unweighted pair-group method with arithmetic averaging analysis showed that most of the samples clustered with the accessions of the Puerto Rican cassava collection. However, there was one cluster of only unknown cassava samples and no accessions of the Puerto Rican collection. Members of this cluster most likely share alleles which are not represented by the accessions in the collection.

**Conclusion**

This broad study shows the presence of high levels of genetic diversity in the unknown samples. Traditional practices, especially intercropping and incorporation of volunteer seedlings, have led to the contribution of recombinant genotypes to cultivated stocks. This study revealed a high potential for local Puerto Rican cassava to be a useful part of future genetic improvement programmes.

**Introduction**

The incorporation of volunteer seedlings, especially of predominantly vegetatively propagated crops, by traditional farmers has become an important mechanism for increasing genetic variability and a potential avenue for avoiding genetic erosion (Pujol et al. 2005).

* Corresponding author’s e-mail address: dimuth.siritunga@upr.edu

Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
During the common farming practice of ‘slash and burn’, volunteer seedlings generated by sexual production can remain in the farmed land and then be cultivated along with a known accession during the following growing season. Based on their morphological characteristics, farmers readily incorporate the new seedlings with similar known accessions even though they are genetically different (Elias et al. 2000). Such traditional practice contrasts with the modern agricultural approach of cultivating large areas with genetically uniform accessions. The older tradition is important for maintaining or even increasing genetic diversity (Altieri and Merrick 1987; Faroldo et al. 2000; Sambatti et al. 2000; Cabral et al. 2002; Resende et al. 2004).

Cocultivation of different accessions in the same traditionally managed field has been documented for root and tuber crops such as sweet potato, _Ipomoea batatas_ L. (Veasey et al. 2007); yam, _ Dioscorea _ spp. (Malapa et al. 2005); and cassava, _ Manihot esculenta _ Crantz (Mühlen et al. 2000; Sambatti et al. 2001; Peroni et al. 2007). In cassava, such traditional farming practices have been shown to maintain genetic variability (Elias et al. 2000; Sambatti et al. 2001) by means of recombination, gene flow and somatic mutations (Elias et al. 2001). The accumulation of fixed somatic mutations in cassava transmitted through vegetative propagation is an important factor attributed to the intra-varietal polymorphisms found today (Elias et al. 2001; Sardos et al. 2008).

With the genus _Manihot _ containing nearly 100 wild species, researchers have suggested that cassava was domesticated independently more than once and is a complespecies with several different wild relatives contributing to its genetic make-up (Rogers 1963; Rogers and Appan 1973; Jennings 1995). But Olsen and Schaal (1999) and Olsen (2004), using sequence variation of single-copy nuclear gene glyceraldehyde 3-phosphate dehydrogenase and molecular markers, respectively, showed that cassava was domesticated from a single wild progenitor, _M. esculenta_ subsp. _flabelifolia_, which shares a strong morphological similarity to modern cultivated cassava. Although research on cassava ancestors has revealed high levels of genetic diversity (Elias et al. 2001; Fregene et al. 2003), the same high levels of genetic diversity (measured as levels of heterozygosity) are not expected at the field level due to the clonal nature of cassava propagation (Pujol et al. 2005).

Evaluating genetic diversity and further characterizing cassava populations is an invaluable asset in the genetic improvement strategies of cassava. It can also lead to the discovery of new accessions thanks to traditional farming practices still being continued today. Since the maintenance of accessions is carried out in the field by farmers themselves, correct assessment of the genetic diversity of those farmer-held accessions can be invaluable in a crop’s conservation. Correct assessment can lead to the identification of new combinations with maximum genetic variability, which can then be used for further selection and introgression of desirable genes from diverse germplasm into the available genetic base (Smith 1984; Cox et al. 1986; Mohammadi and Prasanna 2003).

A number of molecular methods such as restriction fragment length polymorphisms, random amplified polymorphic DNAs, amplified fragment length polymorphisms and, most recently, single-nucleotide polymorphisms and single sequence repeat (SSR) markers (also known as microsatellites) have been used to study cassava diversity (Fregene et al. 2003; Kizito et al. 2005). Single sequence repeat molecular markers are a very effective way of assessing genetic diversity since they can be easily adapted for the classification and identification of many organisms and are particularly useful in studying the variation in allelic frequency of unlinked loci. Single sequence repeats exhibit high levels of polymorphisms, are somatically stable, are inherited in a co-dominant Mendelian manner and are conducive to automation (Morgante and Olivieri 1993; Fregene et al. 2003). Using SSRs, several studies have shown a high level of heterozygosity in various populations of cassava (Chavarriaga-Aguirre et al. 1998; Olsen and Schaal 2001; Fregene et al. 2003; Kizito et al. 2005; Rocha et al. 2008). Rocha et al. (2008) analysed the genetic diversity of cassava relatives in traditional Costa Rican farm settings and found that new plants from germinated seed (e.g. products of sexual reproduction) co-exist with their progenitors or wild relatives. These seedlings add higher levels of heterozygosity and genetic diversity within these populations.

Although cassava cultivation in the Caribbean can be traced back to the Taino Indians of the 15th century (Rouse 1992), to our knowledge no concerted effort has been made to assess the genetic diversity of this crop in the region. Two experimental research stations of the University of Puerto Rico at Corozal and Isabela, Puerto Rico, are responsible for maintaining an official cassava germplasm collection. The collection currently consists of 23 accessions that have either been introduced from the International Center for Tropical Agriculture (CIAT) (from its acronym in Spanish, Cali, Colombia) or have been included from local selections of unknown genetic backgrounds. Seven of the accessions (CM523, CM3064, CM3311, CM3380, CM4448, SG804 and SM494) have been introduced to Puerto Rico by the USDA—Tropical Agriculture Research Station (TARS) in 1994 from the International Center for Tropical Agriculture (CIAT) in...
Colombia. Yield data obtained from these seven accessions showed that CM3311 and CM3380 had the highest yield when grown in Western Puerto Rico (Goenaga et al. 2002). The rest of the present collection consists of accessions that have been added and maintained at the Agriculture Experimental Stations of Puerto Rico for ~26 years. These accessions of unknown genetic background are ‘Abuelo’, ‘Bravo’, ‘Chilena’, ‘Cubana’, ‘Forastera’, ‘Jamaica_18’, ‘Llanera’, ‘PI12900’, ‘PI12902’, ‘PI12903’, ‘Seda’, ‘Senon’, ‘Serrallés’, ‘Tremesiana’, ‘Trinidad14-56’ and ‘Valencia’.

In this study, 33 SSR markers were used to assess the genetic diversity of the Puerto Rican cassava germplasm. In addition, the same set of 33 markers was utilized in estimating the genetic diversity of samples of unknown genetic background collected from farmers’ fields throughout the island. These 33 SSR markers were selected primarily for being unlinked, for a good coverage of the cassava genome as well as for their high polymorphic content (Fregene et al. 1997; Chavarriaga et al. 1998; Mba et al. 2001; Hurtado et al. 2008). Furthermore, these markers are a subset of 36 markers which represent the 18 haploid chromosomes of cassava (2 per chromosome) (Hurtado et al. 2008).

The objective of this study was to define genetic diversity within the current Puerto Rican cassava collection as well as to measure the role played by farming practice in maintaining and/or adding to cassava genetic diversity.

**Materials and methods**

**Plant material**

Leaf samples for the 23 cassava germplasm accessions were collected from the Corozal Agriculture Experimental Station of the University of Puerto Rico (latitude 18°20’N, longitude 66°31’W, altitude 185 m), located in the highland agricultural zone of Puerto Rico. Cassava leaf samples from an additional 162 plants of unknown genetic background were collected from townships around the island by students as part of a laboratory module implemented in the Department of Biology Genetics Laboratory at the University of Puerto Rico Mayagüez Campus. The majority of the samples were collected from the Northwest and Southwest part of the island [see Additional information]. This is the main area of cassava cultivation in Puerto Rico due to the presence of well-drained soil in the area, which is ideal for cassava growth.

**DNA extraction**

DNA extraction was performed according to the Della-porta et al. (1983) protocol with some modifications. Approximately 0.5 g of young leaf tissue was ground with sterile sea sand in 500 μL of extraction buffer (0.1 M Tris-HCl pH 8.0, 0.05 M ethylenediaminetetraacetic acid (EDTA), 0.5 M NaCl, 1 % polyvinylpyrrolidone, 1.6 % sodium dodecyl sulphate (SDS)). This was followed by the addition of 50 μL of 20 % SDS, and after mixing by inversion the tubes were incubated for 8 min at 65 °C. Next 250 mL of potassium acetate (at −20 °C) were added, followed by incubation for 5 min at −20 °C. The samples were centrifuged for 3 min and 500 μL of isopropanol (at −20 °C) were added to the supernatant in a new tube. The mixture was incubated for 5 min at −20 °C, and after centrifugation the supernatant was discarded and the pellet air-dried. The pellet was washed once with 700 μL of 70 % ethanol at −20 °C and resuspended in 150 μL of 10:1 TE (10 mM Tris:1 mM EDTA) buffer. After adding 2.25 μL of 10 mg mL−1 RNase and incubating for 5 min at 65 °C, the DNA was stored at −20 °C until further use. The DNA was quantified using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific Inc., Wilmington, DE, USA). Samples were diluted to 10 ng μL−1 concentrations with deionized distilled water for polymerase chain reaction amplification.

**Single sequence repeat marker amplification**

Table 1 shows the sequences of the primers, the modifications made to the forward primer adding the M13 sequence at the 5’ end (Oetting et al. 1995; Zho et al. 2002), and the allele size range observed for each primer set. Three of the original 36 SSR markers were discarded due to incompatibility with the diallelic model of inheritance adopted. Polymerase chain reaction with a final volume of 12.5 μL was performed as follows: 0.125 μL of 0.1 μM for both forward and reverse primers, 0.5 μL of 0.5 pmol μL−1 M13 tailed SSR primer (LI-COR Biosciences, Lincoln, NE, USA), 2.0 μL of 10 ng μL−1 template DNA, 1.0 μL of 2 mM MgCl2, 0.25 μL of 0.2 mM dNTPs and 1 U Taq polymerase (Applied Biological Materials, Richmond, BC, Canada) and 2.5 μL of 1× reaction buffer.

Amplification reaction conditions were 95 °C for 5 min followed by 34 cycles of 94 °C for 30 s, 45 or 55 °C for 45 s and 72 °C for 1 min, with a final extension of 5 min at 72 °C. The amplified product was visualized on 6.5 % denaturing polyacrylamide gels on a LI-COR 4300 automated DNA sequencer. The bands were scored as present or absent. The molecular weight of each band was assessed by running a 50–350 base pair molecular size ladder (LI-COR) in each gel. Scoring of the bands was performed using the Saga GT software (LI-COR).
Table 1 Properties of the 33 SSR loci used in the study of genetic diversity of cassava in Puerto Rico and the respective primer pairs utilized for the amplification of those loci. All primers shown are from the 5′–3′ direction and the 5′ end of each forward primer was modified with the M13 tail sequence (5′-CACGACGTTGTAAACGAC-3′).

| Locus   | Repeat motif         | Forward primer (F)/reverse primer (R)                  | Annealing temp. (°C) | Allele size range |
|---------|----------------------|--------------------------------------------------------|----------------------|-------------------|
| SSRY4   | GA(16)TA GA(3)       | F—ATAGAGCCAGAAGTGCAGGCG R—CTAACGCACACGACCTGCTTGCA   | 55                   | 278–320           |
| SSRY9   | GT(15)               | F—ACAATTCATCATGAGTCATCAACT R—CCGATTGTTTTTGTCTCTCT   | 55                   | 267–293           |
| SSRY12  | CA(19)               | F—AACTGCTAAACCATTCTATCTTTG R—GCCACACACGTTTGCTACAT   | 55                   | 264–284           |
| SSRY19  | CT(8)GA(18)          | F—TGTAGGGACATCCAGAAATATCA R—CTCTCTGGAAAGGATGTAAGA | 55                   | 203–237           |
| SSRY20  | GT(14)               | F—CATTTGACTTCTCTACCAATTAATGAAT R—TGATGGAAAGGTTATGTCCTTT | 55 | 146–188 |
| SSRY21  | GA(26)               | F—CTGCCACAATATGGAATGG R—CAACAATGGACTAAGCAAGCA      | 55                   | 180–214           |
| SSRY34  | GGC(5)GGT GGC GGT(2) | F—TTCCAGACTGTTTCCACCAT R—ATTGCAGGATTATGATGCTCG     | 55                   | 288–306           |
| SSRY59  | CA(20)               | F—GCAATGACGTAACCATTTTGTGT R—CTCCAATCATCTCATGCTTTGGCA | 55 | 146–202 |
| SSRY63  | GA(16)               | F—TCGAATTGCTCAATTCTGGA R—AGACACACTTTTGTGCTGCA      | 55                   | 303–315           |
| SSRY64  | CT(13)CG CT(6)       | F—CGACAGTGTATGATGATGATTCAGC R—GCAAGGTGGGCTACGAGAC | 55                   | 192–220           |
| SSRY69  | CT(18)ATT AT(2) CTTCTTT CT(TT(2)CCTCTTT    | F—CGATCTGTACGTAATACCAAG R—CACCACGTGGTCGCCAATTATTA | 55                   | 213–267           |
| SSRY82  | GA(24)               | F—TGAGACATTCTCTGCAGAGTCTTTCA R—CGTGAGGAAATTTTTTGTGC | 55 | 166–208 |
| SSRY100 | CT(17)TT CT(7) CCCTT | F—ATCTGAGGTGCACTATTGTC R—TGTGAGTCTATTCCCTTGAATGGC | 55                   | 209–273           |
| SSRY102 | (GT)11               | F—TGGTCTGCTTCTCCTCTTTAATGC R—TGAGCAGAAGATTGAAACCA | 55                   | 198–200           |
| SSRY103 | GA(22)               | F—TGAGAACGAAACTCTGTGACAGC R—CAGCAAGACCATCACCAGTGT | 55                   | 274–308           |
| SSRY106 | CT(24)               | F—GGAGATGTCTGTGCAAAAGA R—CAGCAAGACCATCACCAGTGT    | 55                   | 268–302           |
| SSRY120 | CA(7)AA GA(8)GGA     | F—TCACGTTAATTGTAAGCTGTCCG R—GCAGGTGTCAAATATGCGAT | 55                   | 150–174           |
| SSRY135 | (CT)16               | F—CCAGAAATGAAATGCATCG R—AACATGTGCACAGTATTG        | 45                   | 183–277           |
| SSRY147 |                     | F—GTACATCCCCCAACAGGCG R—ACAGAAGACCATCACCAGTGT      | 45                   | 118–136           |

Continued
Analysis of SSR marker data

Gene diversity parameters were estimated utilizing the 33 SSR marker data, chosen for their diallelic nature. The software package GEN-SURVEY (Vekemans and Lefevre 1997) was used for generating the following statistics: percentage of polymorphic loci, mean number of alleles per locus, average observed heterozygosity ($H_O$), average expected heterozygosity ($H_E$) and average expected heterozygosity corrected for small samples ($H_{EC}$) (Nei, 1978). The inbreeding coefficient $F_{IS}$ was calculated as $H_{EC} - H_O / H_{EC}$. For all loci and samples, the total heterozygosity ($H_T = \text{total heterozygosity in the entire data set}$) was partitioned into within-population diversity ($H_S = \text{heterozygosity within populations averaged over the entire data set}$) and between-population diversity ($D_{ST} = \text{heterozygosity between populations averaged over the entire data set}$) estimates, where $H_T = H_S + D_{ST}$. The proportion of among-accession differentiation ($G_{ST}$) was estimated according to Nei (1978). Standard deviations for the above parameters were

### Table 1 Continued

| Locus | Repeat motif | Forward primer (F)/reverse primer (R) | Annealing temp. (°C) | Allele size range |
|-------|--------------|---------------------------------------|----------------------|-------------------|
| SSRY148 | R—AGAGCCTGTTGGGCGCAAAGGC | F—GGCTTCATGCGGAAAAACCC | 45 | 128–138 |
| SSRY151 | R—CAATGCTTTTACCGGAAAGGCC | F—AGTGGAAATAAGCCATGTGAT | 45 | 194–234 |
| SSRY155 | R—CCATAATTTGTGCGTGGTTT | F—CGTGTGAAATGGTGGAAGACCA | 55 | 163–175 |
| SSRY161 | R—ACTCCACCTCCCGATGCTGC | F—AAGGAACACTTCTCCTAGAATCA | 55 | 188–256 |
| SSRY164 | R—CCAGCTGTAGTGAGTGGACG | F—TCAACAAGAATTACCAGAAGTGG | 45 | 156–204 |
| SSRY175 | R—TGAGATTTCGAATATTCTCATTTTTC | F—TGACTAGCAGACCGTTCA | 55 | 100–156 |
| SSRY177 | R—GCTAACAGTCCTACTCAACTTAA | F—ACCAAAACATAGCAAGCAG | 45 | 244–286 |
| SSRY179 | R—CAGGCTCAGGTGAATTAAAGG | F—CACCAATTCACAGAAATTCAC | 55 | 195–253 |
| SSRY180 | R—GCGAAAGTAAAGCCTCCTGAACTTAA | F—CCTTGGCAGAGATGAATTAGAG | 55 | 131–145 |
| SSRY181 | R—GGGGCTCTAAGTGATGACAAATTA | F—GGTGATCTGGTGACGAGGG | 55 | 192–216 |
| SSRY182 | R—CAATCGAAACCGACGAATA | F—GAATTCTTGTGGTGGTGGTG | 55 | 238–258 |
| NS189 | R—TTGTCTTACAACTTCTGGGACG | F—TGGGCTGGTGTGATGCCCTTA | 55 | 106–124 |
| NS376 | R—CATGAGTTTAAAAATATCATCCG | F—TCAAGACCCCTGCTTTGGTGT | 55 | 213–233 |
| NS911 | R—GGACTATCAAGGCGCAAAAG | F—TGTTGTGAGCGTGAGTCCA | 55 | 135–149 |
estimated over loci by jackknifing (200 replications), and bootstrapping (1000 bootstraps) over loci was used to obtain 95 % confidence intervals for F-statistics. Pairwise genetic distances of populations were calculated with the software GEN-SURVEY using the following statistics: original Nei’s distance (Nei 1972) and corrected Nei’s distance (Nei 1978) for small sample sizes. Genetic differentiation between pairs was estimated over all loci using F-statistics (Wright 1965) and pairwise calculations of $F_{ST}$ (Weir and Cockerham 1984) by the software GEN-SURVEY. Genetic similarity among accessions was estimated by principal coordinate analysis (PCoA) and clustering analysis based on unweighted pair-group method with arithmetic averaging (UPGMA) using the program NTSYSpc version 2.2 by selecting the association coefficient of Jaccard to construct a similarity matrix of the cassava individuals. The similarity matrix was subjected to PCoA to evaluate the relationships among cassava accessions. Further cluster analyses were performed to measure the degree of relatedness between accessions.

### Results

**Number of alleles and genetic diversity**

The 33 loci revealed a total of 293 alleles varying from 2 to 14 alleles per locus [see Additional information]. The 162 samples presented an average number of alleles per locus and per polymorphic locus of 7.15 ± 1.03. Average $H_T$ across all loci and all accessions was high at 0.6705 ± 0.0226. When comparing the two populations, the cassava collection had a slightly higher $H_T$ and the unknown cassava samples had a slightly higher $H_S$ (Table 2). The $H_T$, measured as total heterozygosity, was 0.7357 ± 0.1193 (range 0.3499–0.8583), and was high in the total population (the unknown samples plus the Puerto Rican cassava germplasm) for the 33 SSR markers. The average $H_S$ within population was $0.7085 ± 0.12$ (Table 3) and the inbreeding coefficient ($F_{IS}$), corrected for small samples, was close to zero, implying that neither inbreeding nor outbreeding/genetic drift was occurring in accessions within populations (averaged across the 33 loci, $F_{IS} = 0.0663 ± 0.0364$) (Table 2). Differentiation between populations had an $F_{ST}$ of 0.0299. The average $D_{st}$ between populations was 0.0272 ± 0.0356. The coefficient of gene differentiation revealed a low genetic differentiation between populations ($G_{ST} = 0.0367 ± 0.0478$). However, unique alleles were detected in some of the unknown cassava samples being evaluated. Some of these alleles were considered ‘rare’: e.g. alleles 5 and 6 for SSR181 (frequency of 0.003), allele 5 for SSR164 (frequency of 0.003) and allele 4 for SSR161 (frequency of 0.006).

### Table 2  Analysis of within-population variation of Puerto Rican germplasm accessions and unknown farm-collected samples

| Population              | $H_O$  | $H_E$  | $H_EC$ | $F_{IS}$ |
|-------------------------|--------|--------|--------|----------|
| Unknown cassava         | 0.6545 | 0.7174 | 0.7198 | 0.0921   |
| Cassava collection      | 0.6865 | 0.6996 | 0.7153 | 0.0406   |
| Mean                    | 0.6705 | 0.7085 | 0.7175 | 0.0663   |
| Standard deviation      | 0.0226 | 0.0126 | 0.0032 | 0.0364   |

$H_O$: mean observed heterozygosity.

$H_E$: mean expected heterozygosity.

$H_EC$: mean expected heterozygosity corrected for small samples.

$F_{IS}$: average inbreeding coefficient corrected for small samples.

### Table 3  Average genetic diversity index ($H_T$) and average genetic diversity within populations ($H_S$) across all cassava accessions (unknown farm-collected samples plus Puerto Rican germplasm) for 33 SSR loci

| Locus      | $H_T^a$ | $H_S^b$ | $H_T^a$ | $H_S^b$ |
|------------|---------|---------|---------|---------|
| SSR4       | 0.8044  | 0.7894  | SSR151  | 0.7435  | 0.7273  |
| SSR9       | 0.7757  | 0.765   | SSR155  | 0.6827  | 0.64    |
| SSR12      | 0.7527  | 0.7413  | SSR161  | 0.7044  | 0.6996  |
| SSR19      | 0.8244  | 0.8101  | SSR164  | 0.8454  | 0.7589  |
| SSR20      | 0.8106  | 0.7838  | SSR175  | 0.7561  | 0.7414  |
| SSR21      | 0.8506  | 0.8242  | SSR177  | 0.812   | 0.8055  |
| SSR34      | 0.7188  | 0.7121  | SSR179  | 0.6777  | 0.6685  |
| SSR59      | 0.8321  | 0.8188  | SSR180  | 0.7336  | 0.6995  |
| SSR63      | 0.6874  | 0.5302  | SSR181  | 0.7049  | 0.6835  |
| SSR64      | 0.7884  | 0.7823  | SSR182  | 0.7836  | 0.7685  |
| SSR69      | 0.8583  | 0.8429  | NS189   | 0.3499  | 0.3476  |
| SSR82      | 0.8454  | 0.8361  | NS376   | 0.8172  | 0.8129  |
| SSR100     | 0.4868  | 0.4809  | NS911   | 0.5925  | 0.563   |
| SSR102     | 0.8116  | 0.8016  |         |         |         |
| SSR106     | 0.8315  | 0.8214  | Mean    | 0.7357  | 0.7085  |
| SSR120     | 0.6837  | 0.6494  | STDc    | 0.1193  | 0.1202  |
| SSR135     | 0.7889  | 0.6925  | 95 % CI | 0.6889  | 0.6632  |
| SSR147     | 0.4627  | 0.4463  | 95 % CI | 0.7746  | 0.7461  |
| SSR148     | 0.6375  | 0.6271  |         |         |         |

$a$: Genetic diversity index.

$b$: Genetic diversity within populations.

$c$: Standard deviation (STD) with 95 % confidence interval.

### Cluster analysis

Using Euclidean distance and the UPGMA clustering method, a dendrogram was constructed showing the
genetic relationships of accessions (Fig. 1). Three major groups were defined through cluster analysis (Cluster 
#1, #2 and #3). Similarly, PCoA, which represents the relationship between individual cassava accessions based on the genetic similarity matrix calculated using the Jaccard association coefficient, also showed three main clusters (Fig. 2). Most of the unknown samples were grouped in Cluster #2 as well as 21 out of 23 accessions from the cassava germplasm collection. Interestingly, Cluster #1 did not contain any representative accessions from the cassava germplasm collection. Conversely, Cluster #3 included two accessions, ‘Valencia’ and ‘Serralle’s’, from the cassava germplasm collection and four unknown cassava samples (A-27, A-35, A-58 and A-62) which share >92 % similarity with these two accessions. Likewise, in Cluster #2 several of the unknown samples collected on the island showed high genetic similarity to known accessions from the collection. These were four unknown samples (A-4, A-6, A-7 and C-2) with ‘Jamaica_18’, one unknown sample (B-24) with ‘Trinidad’, one unknown sample (E-5) with ‘Chilena’, one unknown sample (D-6) with ‘SM494’, one unknown sample (B-21) with ‘Abuelo’, one unknown sample (E-16) with ‘Brava’ and ‘Seda’, and two unknown samples (A-52 and E-20) with ‘CM3380’ (Fig. 1). In addition, in two instances, unknown samples (B-43/G-9 and A-53/B-50) shared identical genotypes to each other (Fig. 1), implying
that these samples could be duplicates of the same accessions.

In general, the internal branches of the dendrogram were short while the external branches were long, indicating that within-group variability was higher than that between groups. This is in agreement with our results shown in Table 3, where within-population diversity ($H_S = 0.7085 \pm 0.1202$) is higher than $D_{ST}$ ($0.0272 \pm 0.0356$).

**Discussion**

Cassava, a preferential outcrosser, is propagated by clonal cuttings. It has been shown to be highly heterozygous and the clonal propagation preserves the highly heterozygous genotypes that show hybrid vigour (Balloux et al. 2003; Fregene et al. 2003). Single sequence repeat markers, with >800 available for cassava, have been successfully used in assessing the genetic diversity of different cassava populations (Fregene et al. 1997; Chavarriaga et al. 1998; Mba et al. 2001; Peroni et al. 2007; Hurtado et al. 2008). In this research, Puerto Rico cassava accessions showed high genetic diversity with an average of 7.15 alleles per locus (ranging between 2 and 14 alleles per locus) when analysed by 33 SSR markers. Even though the range of alleles per locus was in agreement with similar studies in cassava (Faraldo et al. 2000; Mühlen et al. 2000; Elias et al. 2001; Fregene et al. 2003; Mkumbira et al. 2003), our average alleles per locus of 7.15 was higher than that reported for other recent studies. For example, when analysing 283 accessions from various countries with 67 SSR loci (the subset of which was used in this study), Fregene et al. (2003) found an average of 5.02 alleles per locus, ranging between 3 and 17 alleles per locus.
A high $H_e$ and $H_o$ across all loci and all accessions were found in our study. The total genetic diversity found in cassava accessions in Puerto Rico was high at $0.7357 \pm 0.1193$. In similar studies, where research involved cassava germplasm, $H_o$ was high ranging from 0.591, 0.538 and 0.654 for work from Fregene et al. (2003), Lokko et al. (2006) and Peroni et al. (2007), respectively. These high $H_o$ values for cassava support both the cassava outcrossing breeding system, with multi-locus outcrossing rates estimated at 91.5% (Silva et al. 2003), as well as its natural abilities to preserve heterozygosity due to its vegetative mode of reproduction. The protogynous nature of cassava flowers, where the male flowers open 7–8 days after female flowers, plays a role in promoting outcrossing in cassava as well (Purseglove 1968). The $H_o$ was high, but comparable between the Puerto Rican cassava germplasm accessions and that of the unknown samples at 0.6865 and 0.6545, respectively. This is probably due to the selection imposed by the local farmer in the field during clonal propagation. Pujol et al. (2005) demonstrated that a positive correlation between plant size and heterozygosity could be found and that this observation was attributed, in part, to the in-field elimination of less vigorous plants by farmers.

Fregene et al. (2003), in their study of 283 cassava accessions from different countries, attributed the high genetic diversity found to agricultural practices of ‘slash and burn’ by Amerindian farmers. Owing to the preferential outcrossing nature of cassava, a large number of volunteer seedlings that survive the ‘slash and burn’ practice can germinate in the field. Natural and artificial selection then acts on these seedlings, leading to new accessions of cassava in the field. The incidence of new accessions in farming systems had long been described in cassava (Altieri and Merrick 1987; Elias et al. 2000, 2001; Pujol et al. 2005; Rocha et al. 2008; Siqueira et al. 2009). Another farming practice that has been documented to increase genetic diversity is the exchange of planting material.
between farmers (Elías et al. 2000; Pujol et al. 2005; Sardos et al. 2008). A combination of these farming practices is thought to be acting upon the cassava populations in Puerto Rico.

Owing to the large number of unknown samples from field sites assessed in this study, a proper comparative evaluation could be performed between the population consisting of the unknown samples and the cassava germplasm population, both of which show high genetic diversity inherited by their ancestors (Olsen and Schaal 2001). When comparing the $H_0$, a higher heterozygosity was observed within the collection than in the unknown samples. An explanation for this finding could be that all unknown samples originated from seedlings of the known germplasm accessions. Approximately 71% of the genetic variability was found to be concentrated within populations ($H_S = 0.7085$), and there was very low genetic differentiation between populations ($G_{ST} = 0.0367$). Siqueira et al. (2009), who conducted a study with cassava landraces selected from different regions in Brazil, also found that most of the genetic variability was concentrated within ethnovariety groups ($H_S = 0.552$) and low genetic differentiation between groups ($G_{ST} = 0.131$).

Genetic distances (Fig. 2) revealed the relative similarities between samples (E-1/F-1, A-9/A-10, B-40/B-43/G-9, A-25/B-51, B-44/B-45, B-30/A-29, C-17/C-29, F-3/H-14, A-6/H-8, F-2/A-37) from different townships in Puerto Rico [see Additional information]. This close relationship could be a result of recent exchange of material between local farmers. Interestingly, four pairs of accessions from the cassava collection, ‘Forastera’/‘PI12900’, ‘Cubana’/‘PI12903’, ‘Serralle’s’/‘Valencia’ and ‘Seda’/‘Brava’, share $\geq 96\%$ genetic similarity. These pairs of accessions could be misnamed and in actuality the same genotype. Such misnaming could

![Fig. 2 Principal coordinate analysis of the microsatellite amplified pattern of the unknown cassava samples and the Puerto Rican cassava collection showing the level of relatedness and diversity among the populations. Three major clusters are shown by circles.](https://academic.oup.com/aobpla/article-abstract/doi/10.1093/aobpla/plr010/148209)
happen in the hands of the farmer due to morphological similarities or perceived differences.

Cluster #1 of our study consisted only of samples from cassava of unknown nature found in the hands of the farmers while none from the cassava germplasm collection were present. This suggests that the genetic information present in the germplasm collection is not representative of the genetic diversity that actually exists in cassava grown in Puerto Rico. Attempts are now being made to collect propagatable nodal samples from Cluster #1 in order to incorporate them into the Puerto Rican cassava collection being maintained in vitro as well as in the field. The overall genetic diversity found in the unknown cassava samples in Puerto Rico is in part attributable to the high genetic diversity of the genus *Manihot* and more specifically to the ancestors of *M. esculenta*. Previous studies have established that the closest ancestor to the domesticated cassava, *M. esculenta* subspecies *flabellifolia*, has a high level of heterozygosity and that the heterozygosity in modern-day cassava is a direct derivative of that (Olsen and Schaal 2001).

### Conclusions and forward look

Our results highlight the need to redefine the biological unit incorporated into strategies for conserving local cassava diversity. These strategies should be based not only on the major agronomic groups or even on the variety, but also on the ecological and human factors that contribute to the differentiation of the genetic stocks. Traditional practices, especially intercropping and incorporation of volunteer seedlings, promote the contribution of recombinant genotypes to the cultivated stocks, therefore allowing selection and adaptation to continue in crops such as cassava, which are mainly propagated vegetatively. Even though the extent of the influence of traditional practices on the genetic diversity of local accessions must be evaluated more precisely in future work, we conclude that abandoning these practices would lead to long-term reduction of genetic variability in local cassava accessions. Conservation strategies should therefore aim to maintain such traditional farming methods. The levels of genetic diversity found in our study reveal high potential for local Puerto Rican cassava to be part of a genetic improvement programme in the future.

### Additional information

The following additional information is available in the online version of this article –

File 1. Figure: Collection of samples from cassava of unknown genetic background from farmers in different areas of Puerto Rico.

File 2. Figure: The frequency and number of alleles per SSR marker in the Puerto Rican cassava collection and in unknown cassava samples.

File 3. Table: Accessions sharing close relationships with each other measured as >95 % similarity coefficient.

### Sources of funding

This work was supported by a grant from the Course, Curriculum and Laboratory Improvement Program of the US National Science Foundation (#0736727).

### Contributions by the authors

All the authors contributed equally in the preparation of this manuscript.

### Acknowledgements

We thank Agenol González, Pedro Márquez and Gladys Toro of the University of Puerto Rico Mayaguez for their contributions in the development of the study. We thank the undergraduate students of the Genetics (Bio3300) course during spring and fall 2009 semesters for providing leaf samples from different townships of Puerto Rico used in this study. We also thank Dr Brian Irish (USDA-TARS, Mayaguez, PR) and Dr Jim Beaver (University of Puerto Rico Mayaguez, PR) for their critical review of the manuscript.

### Conflict of interest statement

None declared.

### References

Altieri M, Merrick L. 1987. In situ conservation of crop genetic resources through maintenance of traditional farming systems. *Economic Botany* 41: 86–96.

Balloux F, Lehmann L, de Meëus T. 2003. The population genetics of clonal and partially clonal diploids. *Genetics* 164: 1635–1644.

Cabral B, Souza J, Ando A, Vasey E, Cardoso ER. 2002. Isoenzymatic variability of cassava accessions from different regions in Brazil. *Scientia Agricola* 59: 521–527.

Chavarriaga-Aguirre P, Maya M, Bonierbale M, Kresovich S, Fregene M, Tohme J, Kochert G. 1998. Microsatellites in cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. *Theoretical and Applied Genetics* 97: 493–501.

Cox T, Murphy J, Rodgers D. 1986. Changes in genetic diversity in the red winter wheat regions of the United States. *Proceedings of the National Academy of Science (USA)* 83: 5583–5586.
Manihot esculenta

Fregene M, Goenaga R, Mba R, Fregene M, Malapa R, Faraldo M, 2011 plr010 doi:10.1093/aobpla/plr010, available online at www.aobplants.oxfordjournals.org © The Authors 2011

Dellaporta S, Wood J, Hicks J. 1983. A plant DNA minipreparation: version II. Plant Molecular Biology Reporter 1: 19–21.

Elías M, Panouédé O, Robert T. 2000. Assessment of genetic variability in a traditional cassava (Manihot esculenta Crantz) farming system, using AFLP markers. Hereditas 85: 219–230.

Elías M, Penet L, Vindry P, Mckey D, Panaud O, 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.

Faraldo M, Silva M, Ando A, Martins P. 2000. Variabilidade genética de etnovariedades de mandioca em regiões geográficas do Brasil. Scientia Agricola 57: 499–505.

Fregene M, Angel F, Gómez R, Rodríguez F, Chavarriaga P, Roca W, Tohme J, Bonierbale M. 1997. A molecular genetic map for cassava (Manihot esculenta Crantz). Theoretical and Applied Genetics 95: 431–441.

Fregene M, Suárez M, Mkwumba J, Kulembeha K, Ndedyo E, Kulya A, Mitchel S, Gulberg U, Rosling H, Dixon A, Dean R, Kresovich S. 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.

Goenaga R, Rivera-Amador E, Chardon U. 2002. Yield performance of introduced cassava clones in an ultisol in Puerto Rico. Journal of Agriculture of the University of Puerto Rico 86: 27–33.

Hurtado P, Olsen K, Buitrago C, Osprina C, Marin J, Duque M, de Vicente C, Wongtiem P, Wenzel P, Killian A, Adeleke M, Fregene M. 2008. Comparison of simple sequence repeat (SSR) and diversity array technology (DArT) markers for assessing genetic diversity in cassava (Manihot esculenta Crantz). Plant Genetic Resources: Characterization and Utilization 6: 208–214.

Jennings DL. 1995. Manihot esculenta (Euphorbiaceae). In: Smartt J, Simmonds NW, eds. Evolution of crop plants. New York: Wiley, 128–132.

Kizito E, Bua A, Fregene M, Egwang T, Gulberg U, Westerberg A. 2005. The effect of cassava mosaic disease on the genetic diversity of cassava in Uganda. Euphytica 146: 45–54.

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

Malapa R, Arnau G, Noyer JL, Lebot V. 2005. Genetic diversity of the greater yam ( Dioscorea alata L.) and relatedness to D. nummularia Lam. and D. transversa Br. as revealed with AFLP markers. Genetic Resources and Crop Evolution 52: 919–929.

Mba R, Stephenson P, Edwards K, Melzer S, Mkwumba J, Gulberg U, Apel K, Gale M, Tohme J, 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.

Mkumba J, Chiwona-Karltun L, Langercrantz U, Mahungu N, Saka J, Mhone A, Bokanga M, Brimer L, Gulberg U, Rosling H. 2003. Classification of cassava into ‘bitter’ and ‘cool’ in Malawi: from farmers’ perception to characterization by molecular markers. Euphytica 132: 7–12.

Mohammadi S, Prasanna B. 2003. Review & interpretation. Analysis of genetic diversity in crop plants. Crop Science 43: 1235–1248.

Morgante M, Olivieri A. 1993. PCR-amplified microsatellites as markers in plant genetics. The Plant Journal 3: 175–182.

Mühlen G, Martins P, Ando A. 2000. Variabilidade genética de etnovariedades de mandioca, avaliada por marcadores de DNA. Scientia Agricola 57: 319–328.

Nei M. 1972. Genetic distance between populations. American Naturalist 106: 283–292.

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

Oetting W, Lee H, Flanders D, Wiesner G, Sellers T, King R. 1995. Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13-tailed primers. Genomics 30: 450–458.

Olsen K. 2004. SNPs, SSRs and inferences on cassava’s origin. Plant Molecular Biology 56: 517–526.

Olsen K, Schaal B. 1999. Evidence on the origin of cassava: phylogeography of Manihot esculenta. Proceedings of the National Academy of Sciences of the USA 96: 5586–5591.

Olsen K, Schaal B. 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.

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

Pujol B, Gigot G, Laurent G, Pinheiro-Kluppel M, Elias M, McKey HM, McKey D. 2005. Germination ecology of cassava (Manihot esculenta Crantz, Euphorbiaceae) in traditional agroecosystems: seed and seedling biology of a vegetatively propagated domesticated plant. Economic Botany 56: 366–376.

Purseglove JW. 1968. Tropical crops: dicotyledons. London: Longman.

Resende A, Filho P, Machado M. 2004. Esterase polymorphism marking cultivars of Manihot esculenta, Crantz. Brazilian Archives of Biology Technology 47: 347–353.

Rocha O, Zaldimar M, Castro L, Castro E, Barrantes R. 2008. Microsatellite variation of cassava (Manihot esculenta Crantz) in home gardens of Chibchan Amerindians from Costa Rica. Conservation Genetics 9: 107–118.

Rogers DJ. 1963. Studies of Manihot esculenta Crantz and related species. Bulletin of the Torrey Botanical Club 90: 43–54.

Rogers DJ, Appan SG. 1973. Manihot and maniholioides (euphorbiaceae): a computer assisted study New York, NY: Hafner.

Rouse I. 1992. The Tainos: rise and decline of the people who greeted Columbus. New Haven, CT: Yale University Press, 211 p.

Sambatti J, Martins P, Ando A. 2000. Distribuição da diversidade isoenzimática e morfológica da mandioca na agricultura autóctone de Ubatuba. Scientia Agricola 57: 75–80.

Sambatti J, Martins P, Ando A. 2001. Folk taxonomy and evolutionary dynamics of cassava: a case study in Ubatuba, Brazil. Economic Botany 55: 93–105.
Sardos J, Mackey E, Duval MF, Malapa R, Noyer JL, Lebot V. 2008. Evolution of cassava (Manihot esculenta Crantz) after recent introduction into a South Pacific Island system: the contribution of sex to the diversification of a clonally propagated crop. Genome 51: 912–921.

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

Siqueira M, Jurema R, Queiroz-Silva E, Bressan A, Borges K, Pereira J, Pinto J, Veasey E. 2009. Genetic characterization of cassava (Manihot esculenta) landraces in Brazil assessed with simple sequence repeats. Genetics and Molecular Biology 32: 104–110.

Smith J. 1984. Genetic variability within U.S. hybrid maize: multivariate analysis of isozyme data. Crop Science 24: 1041–1046.

Veasey E, Silva J, Rosa M, Borges A, Bressan E, Peroni N. 2007. Phenology and morphological diversity of sweet potato (Ipomoea batatas) landraces of the Vale do Ribeira. Scientia Agricola 64: 416–427.

Vekemans X, Lefevre C. 1997. On the evolution of heavy metal tolerant populations in Armenia maritima: evidence from allozyme variation and reproductive barriers. Journal of Evolutionary Biology 10: 175–191.

Weir B, Cockerham C. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358–1370.

Wright S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19: 395–420.

Zho Y, Bui T, Auckland L, Williams C. 2002. Direct fluorescent primers are superior to M13-tailed primers for Pinus taeda microsatellites. BioTechniques 32: 46–52.