

Comparison of simple sequence repeat (SSR) and diversity array technology (DArT) markers for assessing genetic diversity in cassava (*Manihot esculenta* Crantz)

P. Hurtado¹, K. M. Olsen³, C. Buitrago¹, C. Ospina¹, J. Marin¹, M. Duque¹, C. de Vicente², P. Wongtiem^{4,5}, P. Wenzel⁵, A. Killian⁵, M. Adeleke⁶ and M. Fregene^{1*}

¹Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia, ²Bioversity International, Cali, Colombia, ³Biology Department, Washington University, St. Louis, MO, USA, ⁴Field Crop Research Station, Rayong, Thailand, ⁵Center for the Application of Molecular Biology to Agriculture (CAMBIA), Canberra, Australia and ⁶Department of Biology, Rivers State University of Science and Technology (RUST), Port Harcourt, Nigeria

Received 13 September 2007; Accepted 24 January 2008; First published online 22 August 2008

Abstract

Several molecular marker systems have been developed for assessing genetic diversity in crop germplasm collections. A trade-off often exists between the number of loci that can feasibly be sampled by a marker system and the amount of information provided by each locus. We compared the usefulness of two marker systems for revealing genetic diversity and population structure in cassava (*Manihot esculenta* Crantz): simple sequence repeats (SSRs) and diversity array technology (DArT) markers. DArTs survey many more loci per reaction than do SSRs; however, as bi-allelic, dominant markers, DArTs provide less polymorphism information per locus. Genetic differentiation was assessed in a randomly selected set of 436 cassava accessions, consisting of 155 African and 281 Latin American accessions. A genome-wide set of 36 SSR markers and a DArT array of approximately 1000 polymorphic clones were used to assess genetic diversity and differentiation. Cluster analyses were performed using principal coordinate analysis (PCoA). Results were compared with *a priori* expectations of genetic differentiation based on previous genetic analyses. Analyses of the two datasets generated broadly similar clustering patterns. However, SSRs revealed greater differentiation than DArTs, and more effectively recovered patterns of genetic differentiation observed in previous analyses (differentiation between Latin American and African accessions, and some geographical differentiation within each of these groups). These results suggest that SSR markers, while low throughput in comparison with DArTs, are relatively better at detecting genetic differentiation in cassava germplasm collections. Nonetheless, DArTs will likely prove useful in 'orphan crop' species, where alternative molecular markers have not been developed.

Keywords: cassava; DArT marker; genetic differentiation; genetic diversity; SSR markers

Introduction

The starchy root of cassava (*Manihot esculenta* Crantz) feeds over 500 million of the world's poorest people. This crop is the primary source of carbohydrates in

*Corresponding author. E-mail: M.Fregene@cgiar.org

sub-Saharan Africa, and it is a major food and industrial crop in Latin America and south-east Asia (FAO, 2001). Cassava shows a remarkable degree of phenotypic variability and a wide range of adaptability to agriculturally marginal environments. It is routinely cultivated in conditions ranging from the humid tropics of the Amazon basin to arid regions such as the north-east of Brazil and northern West Africa; it also tolerates a wide range of altitudes, soil fertility and pH.

Understanding the genetic diversity present in cassava is the first step in harnessing its phenotypic variability for crop improvement efforts. The Generation Challenge Programme (GCP; www.generationcp.org), of the Consultative Group on International Agricultural Research, seeks to use advances in molecular biology to harness natural crop genetic variation for the creation of a new generation of hardy crops for small farmers. By examining genetic diversity in a large sample of cassava germplasm, the GCP proposes to identify a subset of germplasm to be used in breeding programs and in association mapping.

The largest cassava germplasm collections are held at two international agriculture research centres, namely the International Center for Tropical Agriculture (CIAT, Colombia), which holds over 5500 accessions from 23 countries worldwide; and the International Institute for Tropical Agriculture (IITA, Nigeria), which holds >2000 African cassava accessions.

Several genetic marker systems have been used for the assessment of genetic diversity of cassava germplasm. They include allozymes (Ocampo *et al.*, 1992), random amplified polymorphic DNA (RAPDs Bonierbale *et al.*, 1995), amplified fragment length polymorphisms (AFLPs; Second *et al.*, 1997), simple sequence repeats (SSRs; Chavarriaga-Aguirre *et al.*, 1998; Fregene *et al.*, 2003) and diversity array technology (DArT; Xia *et al.*, 2005). These molecular markers typically show a trade-off between the number of loci that can be efficiently sampled and the amount of information provided per locus.

At one end of the spectrum, SSR genotyping is a fairly low-throughput and lab-intensive approach. This constrains the number of SSR loci that can realistically be used in genotyping a large germplasm collection. On the other hand, SSRs often show high levels of allelic diversity per locus, and as co-dominant markers they can be used to infer genotype information. At the other end of the spectrum, DArT is a new genotyping method that offers the highest throughput genotyping available to date. DArT uses a microarray platform to score hundreds of polymorphic markers across the genome in a single assay (Jaccoud *et al.*, 2001; Wenzl *et al.*, 2004); however, DArT loci are scored as binary characters and must be treated as dominant markers and this limits the genetic information provided by a given locus.

The first report of DArTs for assessing diversity in cassava (Xia *et al.*, 2005) has opened up the possibility of applying this technology for large-scale germplasm characterization in this crop. However, the potential benefits of this newest high-throughput marker system can extend only as far as its ability to accurately detect patterns of genetic diversity and differentiation in the samples analysed. In this study, we compared the ability of DArTs and SSRs to reveal genetic differentiation in a sample of global cassava genetic resources. A previous genetic analysis, using a large number of SSR markers (67 loci) in a group of 283 cassava accessions, revealed genetic differentiation between African and Neo-tropical accessions, with identifiable structure within each of these groups (Fregene *et al.*, 2003). Genetic differentiation between African and Latin American landraces has also been detected using AFLPs (Fregene *et al.*, 2000). In the present study, we compared the ability of SSRs and DArTs to reveal these previously detected patterns of genetic differentiation in a larger sample of accessions.

Materials and methods

Plant materials

A random set of 436 accessions from over 3000 genotypes maintained at CIAT and IITA was selected for this study. These accessions represent the diversity of agronomic traits present in the crop (Table 1), and include 155 from the IITA collection and 281 from CIAT's cassava germplasm bank. The former are composed principally of west African landraces and elite IITA varieties, while the latter comprise a wide diversity of south and central American landraces (Supplementary Table 1, available online only at <http://journals.cambridge.org>).

Table 1. Description of selection criteria for assembly of 3000 cassava accessions included in Generation Challenge Programme

1	CIAT and IITA core collections
2	CIAT and IITA elite clones
3	Officially released varieties from CIAT, IITA and CNPMF
4	Drought tolerance
5	Tolerance to all major cassava pests and diseases worldwide
6	Good culinary quality
7	Tolerance to acid soils and low phosphorus soils
8	High beta-carotene varieties (yellow varieties)
9	Low and high levels of amylose and amylopectin
10	Low and high levels of iron, zinc and HCN in the root
11	High level of protein in the root
12	Good adaptation to different tropical and sub-tropical agro-ecologies of south and central America, south-east Asia and sub-Saharan Africa
13	Dwarf accessions

Molecular marker analysis

Leaf tissue was collected for DNA extraction from tissue culture plantlets or plants maintained in the screen house or field. Fresh leaves were crushed in extraction buffer or dried at 50°C for 48 h before grinding with sand. In both cases, DNA extractions were carried out using the Dellaporta *et al.* (1983) extraction method. DNA extraction and determination of quantity and quality of DNA samples were carried out at CIAT, where SSR analyses were also conducted. DNA aliquots were sent to the Center for the Application of Molecular Biology to Agriculture (CAMBIA), Canberra, Australia, for DArT analysis.

More than 800 SSR markers are available for cassava (Mba *et al.*, 2001; Fregene *et al.*, 1997; Fregene *et al.*, 2002 unpublished data) and 200 have been placed on a genetic map of cassava (Okogbenin *et al.*, 2006). A subset of 36 markers, 2 from each of the 18 haploid chromosomes of cassava, was selected for assessing genetic diversity. All 36 loci have previously been used in analysis of cassava diversity and show high polymorphism information content (PIC; CIAT, 2002, 2003, 2004).

PCR assays of SSR markers used 30 ng of genomic DNA in 15 µl reactions containing 0.1 mM dNTP, 1 × PCR buffer, 2.5 mM MgCl₂, 0.2 µM each primer (Supplementary Table 1, available online only at <http://journals.cambridge.org>) and 1 U *Taq* DNA polymerase. PCR amplification was performed using following thermal profile: 2 min at 95°C, 30 cycles of 30 s at 94°C, 55/45°C at 1 min and 60 s at 72°C and a final step of 5 min at 72°C. PCR products were denatured and electrophoresed on 4% polyacrylamide gels and visualized by silver staining. The raw SSR data were scored as allele sizes and allele numbers and used for statistical analysis. A strictly bi-allelic model of inheritance was assumed; SSR markers showing three or more alleles were eliminated.

DArT markers were obtained from a cassava DArT array of approximately 1000 polymorphic clones constructed at CAMBIA (Xia *et al.*, 2005). This DArT chip was developed using cassava accessions, a wild progenitor (*M. esculenta* ssp. *flabellifolia*) and three more distantly related wild species (Xia *et al.*, 2005). About 100 ng of total genomic DNA from the 436 accessions was labelled with Fluorescent Cy3-dUTP or Cy5-dUTP (Pharmacia, USA) and used to probe the DArT arrays according to standard methods (Jaccoud *et al.*, 2001). Slides were scanned using the Affymetrix 418 Scanner adjusting the PMT voltage as required.

Spot signal intensities were analysed by Scanlyse v. 2.44 (Stanford University) as well as GenPix Pro v. 3 (Axon Instruments) and GMS Pathways (Affymetrix v. Beta). Outputs of image analysis programs were further analysed using a program, Mathcard v. 8, developed for DArT markers (Jaccoud *et al.*, 2001). To ensure statistical significance, the median of log-transformed signal ratios of three replicates on each slide was used for comparison. A matrix with binary scores for accessions analysed (66.5–94.8% data/genotype) at polymorphic spots was used as the input file for genetic diversity analysis.

Statistical analysis

The PIC for each marker in both marker systems (SSR and DArT) was determined using Botstein *et al.*'s (1980) formula, in which

$$\text{PIC} = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^n \sum_{j=i+1}^n 2p_i^2 p_j^2$$

where p_i is the frequency of the i th allele and n is the number of alleles.

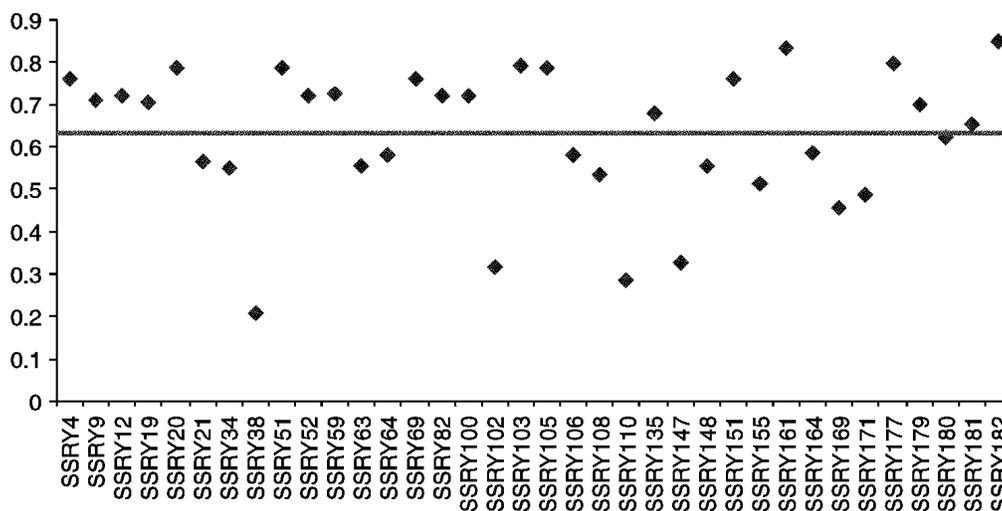


Fig. 1. Polymorphic information content (PIC) calculated for each of the 36 simple sequence repeat markers included in the study. The horizontal line indicates the PIC average in this marker set (0.63).

Similarity matrices generated from SSR and DArT data employing Jaccard's similarity index were used as input files for principal coordinate analysis (PCoA). Coordinates of each point in a Cartesian space were used as input file in a UPGMA cluster analysis based on Euclidian distances in NTSYS v. 2.1 (Rohlf, 1998). To define the number of clusters, a cophenetic correlation and the incremental contribution of each new subdivision were used. Once the clusters were identified, average similarities among and between the clusters were calculated.

Results

For the SSRs, all 36 loci were polymorphic among the 436 accessions; a total of 200 alleles were obtained. Between 3 and 12 alleles per locus were observed (average of six alleles per locus). A total of DArT 251 loci (~25%) were polymorphic. The PIC of each marker is shown in Figs 1 and 2. The average PIC for the SSRs was 0.63 with a standard deviation of 0.16; the lowest value was observed for SSRY38 (0.2) and the highest for SSRY182 (0.85). Among the DArT markers, the average PIC value was 0.33 (standard deviation = 0.05, range = 0.09–0.37). Thus, SSRs were expected to more effectively detect genetic differentiation than DArT.

Genetic differentiation by continent

Previous genetic analyses, using AFLPs and SSRs, have indicated differentiation between African and Latin American cassava accessions (Fregene *et al.*, 2000,

2003). PCoAs in the present SSR and DArT datasets indicate that this pattern of genetic differentiation was apparent in the former (Fig. 3A) but not in the latter, where accessions from the two continents are largely intermingled (Fig. 3B). For the DArT dataset, quantification of genetic identities between African and Latin American accessions indicates higher genetic similarity (0.25) and thus low differentiation between these two groups, whereas the similarity value for the SSR dataset is lower (0.23), indicating a more clear differentiation between these two groups (Table 2). In the DArT dataset, the similarity among accessions from the same continent was higher (Table 2) indicating a low genetic differentiation between them.

Genetic differentiation within continents

A cophenetic correlation (r^2) was used to define the number of clusters in DArT and SSR datasets. Six DArT and five SSR groups explained 80% of the genetic variance ($r^2 = 0.8$). The composition of each cluster was assessed based on the origin of the accessions. Both SSRs and DArTs indicate that, among the Latin American accessions, samples from Guatemala are clearly differentiated from other samples from other countries (Fig. 4A,B). Genetic similarity estimates confirm that the Guatemalan accessions show a high degree of similarity to one another (0.67 for SSRs and DArTs) and a low similarity to other groups (Table 3A,B). This same genetic differentiation of Guatemalan accessions was also detected in a previous SSR analysis (Fregene *et al.*, 2003). Within the African accessions, a group of primarily Nigerian accessions

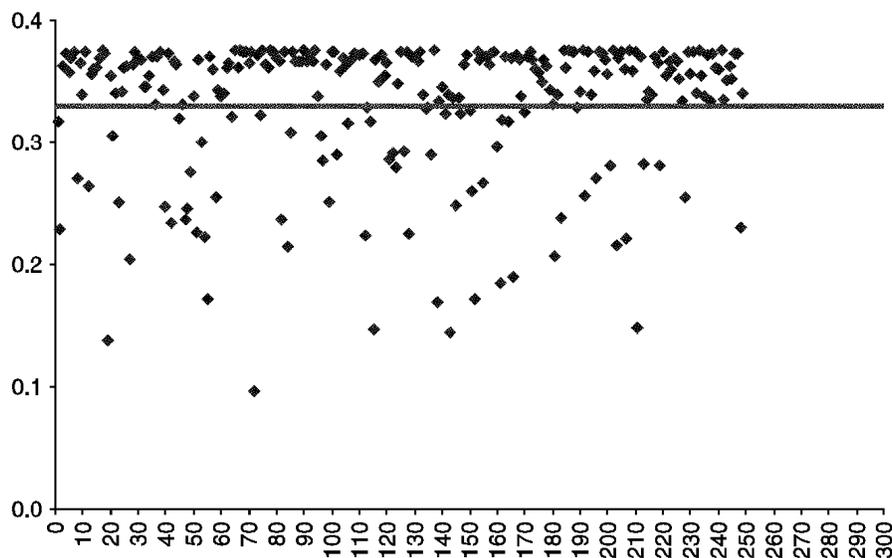


Fig. 2. Polymorphic information content (PIC) calculated for each of the 251 diversity array technology markers included in the study. The horizontal line indicates the PIC average in this marker set (0.33).

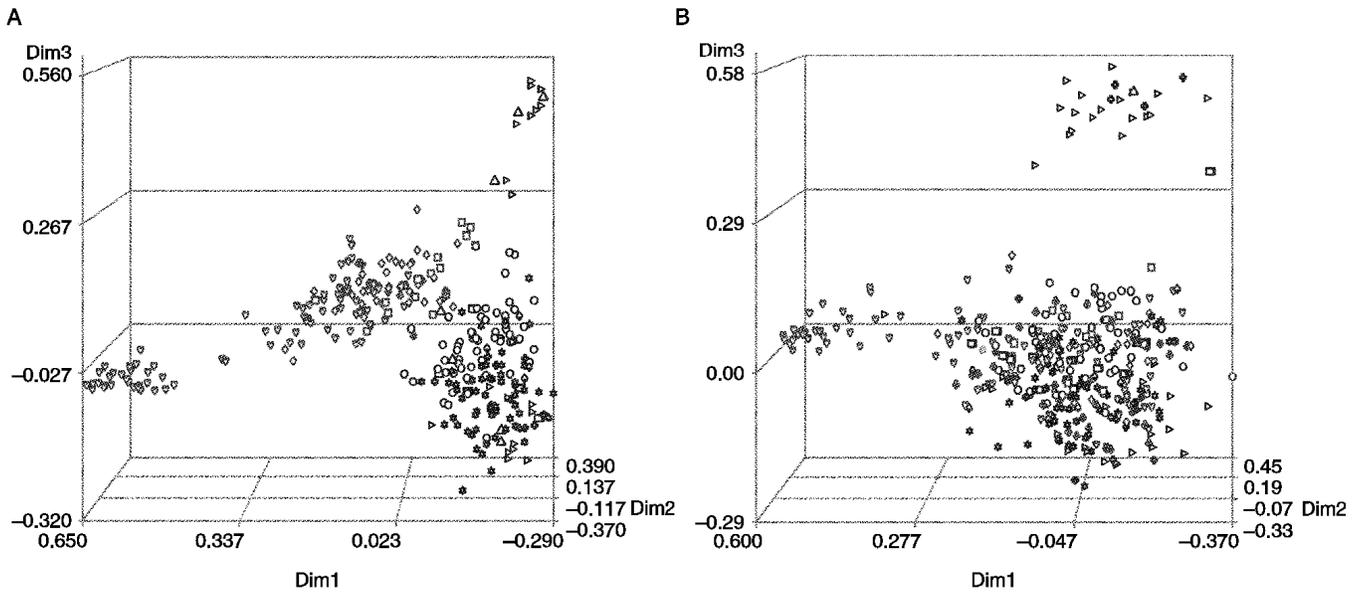


Fig. 3. Principal coordinate analysis of (A) simple sequence repeat and (B) diversity array technology markers representing genetic differentiation between continents. African accessions are represented in grey and Latin American accessions in black. The shapes represent the countries as follow: triangle, Costa Rica and Puerto Rico; flag, Guatemala and Cuba; square, Cape Verde, Gambia and Sierra Leone; diamond, Cameroun; heart, Nigeria; circle, Brazil; club, Ghana, Benin and Togo; cross, Mexico and Panama; start, Colombia; cube, Tanzania; spade, Peru, Ecuador, Venezuela and Argentina.

(labelled B in Fig. 4) was clearly differentiated from the others. A high degree of relatedness obtains within the Nigerian cluster (genetic similarities were 0.78 (SSRs) and 0.75 (DArTs); Table 3A,B). The SSRs, but not the DArTs also revealed a second distinguishable set of primarily Nigerian accessions (group E; Fig. 4; Table 3).

Discussion

Genetic resolution of DArTs and SSRs

Our analysis of 251 polymorphic DArTs in comparison with 36 SSRs indicates that the SSRs are more effective at revealing the patterns of genetic differentiation. The SSRs more clearly recover the population structure inferred in earlier genetic analyses (Fregene *et al.*, 2000, 2003), including differentiation between African and Latin American cassava accessions (Fig. 3A), as well as

Table 2. Genetic identities (similarities) between and among African and Latin American accessions in the simple sequence repeat (SSR) and diversity array technology (DArT) datasets

	SSR		DArT	
	Africa	America	Africa	America
Africa	0.08	0.23	0.13	0.25
America	0.23	0.09	0.25	0.19

differentiation of Guatemalan accessions from other Latin American and differentiation of some Nigerian accessions from other African samples (Fig. 4A). In addition, the SSR analyses suggest that there may be additional differentiation among the African accessions (e.g. group E in Fig. 4A) that is not identified in the DArT analysis.

Taken at face value, these results suggest that there is a greater pay-off in using a low number of SSRs than a large number of DArTs. At least one caveat should be considered, however. Our primary criterion for 'correct' detection of genetic differentiation is whether or not the cluster analyses recover patterns of differentiation observed in earlier genetic analyses. Thirty of the 36 SSR loci in the present study were used in the analysis by Fregene *et al.*, 2000. Thus, if there were some inherent bias in the type of polymorphism information revealed by these SSRs, we may have biased ourselves in favour of detecting the same patterns with the present SSR analysis. It is conceivable that lower genetic differentiation revealed by the DArT analysis (Fig. 3B) is actually a more accurate representation of genetic relationships in the sample set. Nonetheless, the fact that the African–Latin American differentiation is detectable with both SSRs (Fregene *et al.*, 2003) and AFLPs (Fregene *et al.*, 2000) strongly suggests that the patterns of differentiation observed in the present SSR analyses are real.

Although the DArT analysis included 1000 loci, only about one-quarter were polymorphic. This likely reflects the selection of taxa used in designing the cassava DArT

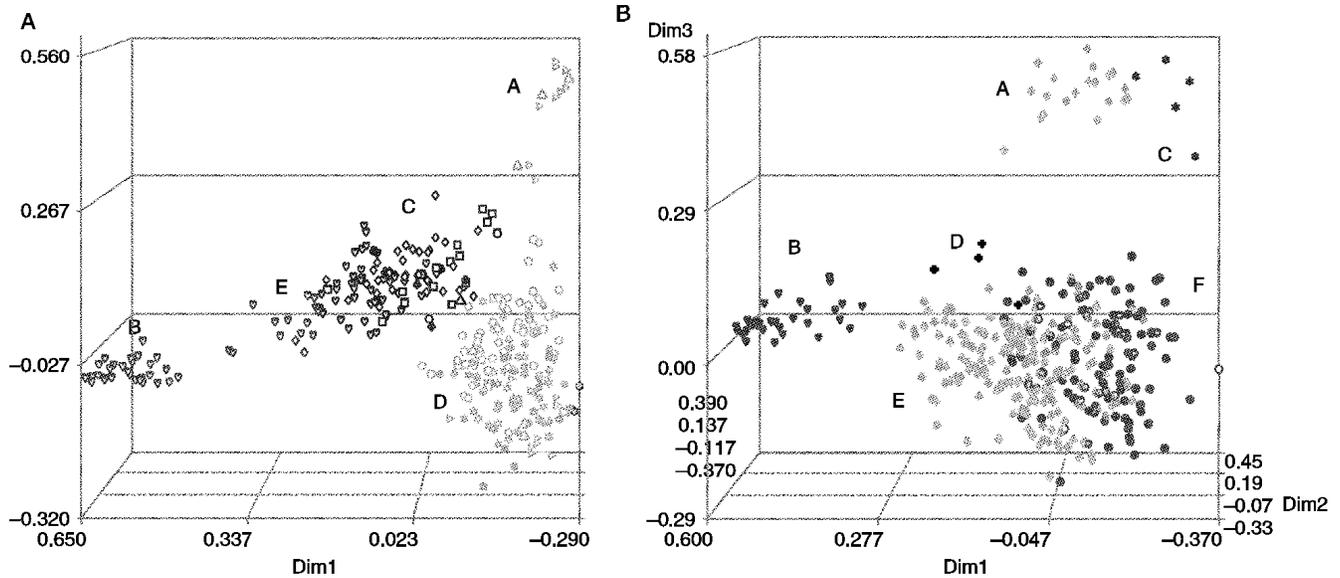


Fig. 4. Principal coordinate analysis of (A) simple sequence repeat and (B) diversity array technology markers showing the groups explaining 80% of the genetic variance in the sample. In (A) and (B), the letters represent the groups and shape, the countries. In (A), the letters represent the groups [A, group A (80% from Guatemala); B, group B (100% from Nigeria); C, group C (37% from Cameroun); D, group D (86% from Brazil); E, group E (86% from Nigeria)] and the shapes correspond to country of origin (triangle, Costa Rica and Puerto Rico; flag, Guatemala and Cuba; square, Cape Verde, Gambia and Sierra Leone; diamond, Cameroun; heart, Nigeria; circle, Brazil; club, Ghana, Benin and Togo; cross, Mexico and Panama; start, Colombia; cube, Tanzania; spade, Peru, Ecuador, Venezuela and Argentina). In (B), the letters and shapes represent the groups [(A) diamond, group A (80% from Guatemala); (B) heart, group B (93% from Nigeria); (C) start, group C (40% from Guatemala); (D) cross, group D (100% from Nigeria); (E) club, group E (23% from Brazil); (F) ball, group F (12% from Ghana)].

array. The chip was designed to reflect diversity present not only in cassava but also its wild ancestor (*M. esculenta* ssp. *flabellifolia*) and three *Manihot* species that are distantly related to the crop. Cassava's wild relatives contain much greater genetic diversity than the crop (e.g. Roa *et al.*, 1997; Olsen and Schaal, 1999; Olsen, 2004). Designing a DArT array solely from cassava germplasm would be likely to yield an array with more power to detect genetic variation within the crop.

Table 3. Similarity within and between the groups defined by principal coordinate analysis of (A) simple sequence repeat and (B) diversity array technology markers

A						
Group	C	B	A	D	E	
C	0.36					
B	0.35	0.78				
A	0.25	0.25	0.67			
D	0.26	0.25	0.28	0.28		
E	0.35	0.50	0.24	0.25	0.43	
B						
Group	B	E	F	D	C	A
B	0.75					
E	0.26	0.32				
F	0.28	0.27	0.28			
D	0.19	0.22	0.23	0.26		
C	0.24	0.27	0.26	0.23	0.49	
A	0.19	0.27	0.23	0.24	0.22	0.67

Because loci assessed by DArT markers can be significantly increased beyond the number employed here with minimal additional costs, it may well be possible that the level of genetic resolution obtained with DArTs could be fairly easily made comparable to that of SSRs. More conclusive evidence will have to await analysis using denser DArT arrays and a larger sample set. While the initial investment for development of both marker systems (labour and consumables) is approximately similar, the cost per assay is far less for DArTs than SSRs, making DArTs an attractive marker alternative. This will be especially true for 'orphan crop' species, where existing marker systems have not yet been developed.

References

Bonierbale MW, Maya MM, Claros JL and Iglesias C (1995) Application of molecular markers to describing the genetic structure of cassava gene pools. In: *The Cassava Biotechnology Network: Proceedings of the Second International Scientific Meeting of the Cassava Biotechnology Network*. Working document no 50. Cali, Colombia: Centro Internacional de Agricultura Tropical, 2v.

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

- Chavarriga-Aguirre P, Maya MM, Bonierbale MW, Kresovich S, Fregene MA, Tohme J and Kochert G (1998) Microsatellites in cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. *Theoretical and Applied Genetics* 97: 493–501.
- Chavarriga-Aguirre P, Maya MM, Tohme J, Duque MC, Iglesias C, Bonierbale MW, Kresovich S and Kochert G (1999) Using microsatellites, isozymes and AFLPs to evaluate genetic diversity and redundancy in the cassava core collection and to assess the usefulness of DNA-based markers to maintain germplasm collections. *Molecular Breeding* 5: 263–273.
- Dellaporta SL, Woods J and Hicks JR (1983) A plant DNA mini-preparation: version II. *Plant Molecular Biology Reports* 1: 19–21.
- FAO (2001) *FAOSTAT Statistical Databases*. Rome, Italy: Food and Agriculture Organization.
- Fregene M, Angel F, Gómez R, Rodríguez F, Bonierbale M, Iglesias C, Tohme J and Roca W (1997) A molecular genetic map of cassava (*Manihot esculenta* Crantz). *Theoretical and Applied Genetics* 95: 431–441.
- Fregene MA, 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.
- 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.
- International Center for Tropical Agriculture (2002) Annual report, Cali, Colombia.
- International Center for Tropical Agriculture (2003) Annual report, Cali, Colombia.
- International Center for Tropical Agriculture (2004) Annual report, Cali, Colombia.
- Jaccoud D, Peng K, Feinstein D and Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research* 29: e25.
- 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: 21–31.
- Ocampo C, Hershey C, Iglesias C and Iwanaga M (1992) Esterase isozyme fingerprinting of the cassava germplasm collection held at CIAT. In: Roca W and Thro AM (eds) *Proceedings of the First International Scientific Meeting of the Cassava Biotechnology Network*. Cali, Colombia: CIAT, pp. 81–89.
- Okogbenin E, Marin J and Fregene M (2006) An SSR-based molecular genetic map of cassava. *Euphytica* 147: 433–440.
- 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.
- 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.
- Rohlf FJ (1998) *NTSYS-pc Numerical Taxonomy and Multivariate Analysis System Ver21*. New York: Exeter Software.
- Second G, Allem A, Emperaire L, Ingram C, Colombo C, Mendes R and Carvalho L (1997) AFLP based *Manihot* and cassava numerical taxonomy and genetic structure analysis in progress: implications for dynamic conservation and genetic mapping. *African Journal of Root and Tuber Crops* 2: 140–147.
- Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A and Kilian A (2004) Diversity array technology (DART) for whole-genome profiling of barley. *Proceedings of the National Academy of Sciences of the United States of America* 101: 9915–9920.
- Xia L, Peng K, Yang S, Wenzl P, De Vicente MC, Fregene M and Killian A (2005) DART for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. *Theoretical and Applied Genetics* 110: 1092–1098.