Development of a Diversity Array Technology (DArT) Chip for Cassava

K. Andrzej[1], P. Wenzel[1], C. de Vicente[2], E. Barrera[3], A. M. Correa[3] and M. Fregene[3]

1. CAMBIA (www.cambia.org) GPO Box 3200, Canberra, ACT 2601 Australia.
2. IPGRI, Ofice of the Americas, CIAT, Cali, Colombia
3. Centro Internacional de Agricultura Tropical, CIAT, Cali, Colombia.

INTRODUCTION

Genetic resources, mostly held by small farmers, represent a critical resource for the future productivity and stability of production of the crop. How to evaluate and use in a systematic manner the vast amount of variability present in cassava is still a challenge to most cassava breeding programs. Numerous genotyping methods including RFLPs, SSRs, RAPDs, AFLPs, and other techniques, have contributed to our understanding of genome organization and genetic variation. Nevertheless, these techniques present some disadvantages related to their dependence on gel electrophoresis; some of these methods require pre-identification of a polymorphism or a potential site for it, before analysis of other individuals is possible. Furthermore, all methods based in size separation of multiple DNA fragments suffer from difficulties in precisely correlating bands on gels with allelic variants [1]. Genotyping micro-array technologies offer the highest throughput available up to date. One of them call Diversity Array Technology, DArT ™ (CAMBIA), is sequence-independent (low-input) and allows the fingerprint of an individual's genome based on a high number of polymorphic sites spread over the genome. These screening procedures should allow testing of thousands of individual samples in a speedy manner. We describe here a proof of concept on using the DArT tool as a cost-effective way for measuring and characterizing genetic diversity of cassava germplasm.

MATERIALS AND METHODS

The project was initiated in early October 2002, with the shipment of cassava DNA samples from CIAT to CAMBIA. Plant materials used for the generation of the DArT chip was chosen to represent a broad as possible diversity of the cultivar, a few genotype of its wild progenitors and 2 wild species were included to capture a large number of polymorphic fragments. They include 14 accessions from Brazil, 14 from Colombia, 4 from Guatemala, 2 each from Nigeria, Cuba, and Ecuador, Peru and Thailand respectively. Others include, one accession each from Argentina, Bolivia, Costa Rica, Fiji islands, Indonesia, Mexico, Panama, Venezuela, and USA. Six and 2 improved varieties were included from CIAT and IITA respectively. The wild species accessions were 29 of *Manihot esculenta* sub spp *flabellifolia*, 7 of *M. carthaginensis* and 1 of *M. walkerae*. DNA isolation was according to Dellaporta et al. (1983) followed by two washes of phenol/chloroform.

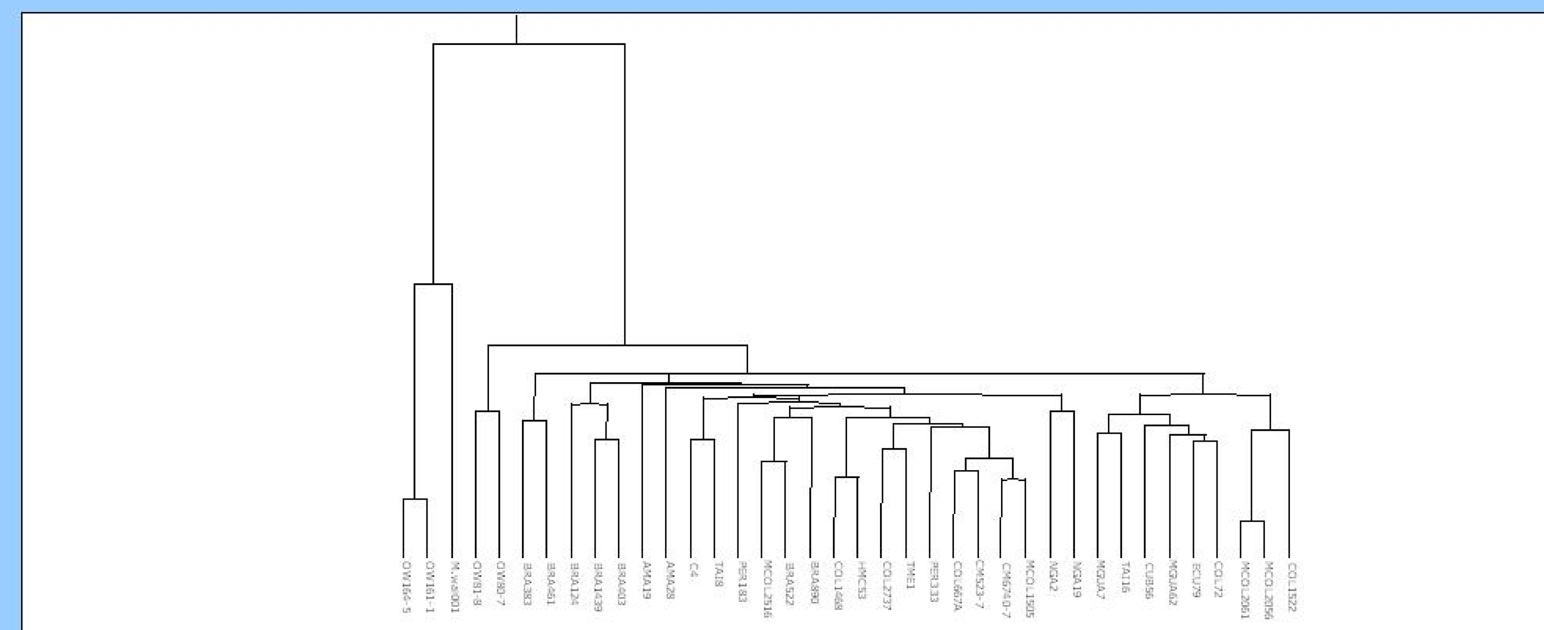
A critical step in DArT is the complexity reduction step. Work at CAMBIA with several other plant genomes has shown that digestion with PstI restriction enzyme (RE) in combination with more frequently cutting RE is an efficient method to reduce complexity. To generate representations, a preliminary experiment was conducted to determine the best enzyme combinations. A mixture of genomic DNA from twenty cassava genotypes was digested with PstI, then ligated to adapters and further digested with several frequently cutting RE (BstNI, ApoI, TaqI and BanII), followed by amplification with an adapter-specific primer. The cassava genomic PstI fragments lacking the recognition site for the frequent cutting RE (BstNI, ApoI and TaqI) were individualized by transformation into E. coli, amplified from bacterial colonies and micro-arrayed. From each of the libraries 760 clones were arrayed. Genomic representations prepared in the same way (RE digestion/ligation followed by amplification) from each of the twenty genotypes separately, were labelled with Cy3-dUTP and hybridized together with Cy5-dUTP-labelled reference DNA to these microarrays. Slide preparation, hybridizations, washes and scanning are as described by Andrzej et al. (2002). Images generated by the scanner were used to extract quantitative fluorescence signal data for each array feature using a software. Same software was used to binarize the data (score as 0/1) for all slides. Binary scoring table was used to prepare the Hamming distance matrix and to obtain a dendogram. Library expansion to obtain more polymorphic clones was carried out using the enzyme combinations PstI/BstNI, and PstI/TaqI, and 80 DNA samples from CIAT. About 3000 clones were evaluated for polymorphism from the 2 libraries. The arrays were printed and DArT analysis carried out as described earlier.

RESULTADOS Y DISCUSION

Among the three libraries PstI/BstNI, PstI/ApoI, and PstI/TaqI tested in the preliminary experimnt, PstI/BstNI gave the largest number of polymorphic clones (132), followed by *TaqI* (112) and ApoI (69). In total, 313 candidate polymorphic clones were obtained in the initial experiment to determine the best enzyme combination. DArT analysis based on 296 polymorphic clones without a single missing data point was used to generate a binary matrix and obtain a dendrogram, based on Hammings distance, representing genetic relationship between the 20 samples (Fig. 1).

Figure1. Genetic relationship among the cassava accession used to identify polymorphic clones in the PstI/BstNI, PstI/TaqI and PstI/ApoI array. The dendrogram was created using the distance table based on 296 polymorphisms and the UPGMA clustering algorith

The libray expansion with 80 clones yielded 440 polymorphic clones (14.3%), for the PstI/TaqI array, consistent with the results obtained with the initial, smaller array (14.6% polymorphic clones). A dendrogram based upon the analysis of the 80 cassava samples with the polymorphic clones is presented in Figure 2.



e 2. **Figure 2.** Genetic relationship among the cassava accession used to identify polymorphic clones in the PstI/TaqI array. The dendrogram was created using the distance table based on 440 polymorphisms and the UPGMA clustering algorithm.

In the PstI/BstNI array 554 polymorphic clones (18.0%) were identified, also consistent with the polymorphism frequency in the smaller PstI/BstNI array in the first project phase (17.2%). The dendrogram resulting from this array is shown in Figure 3.

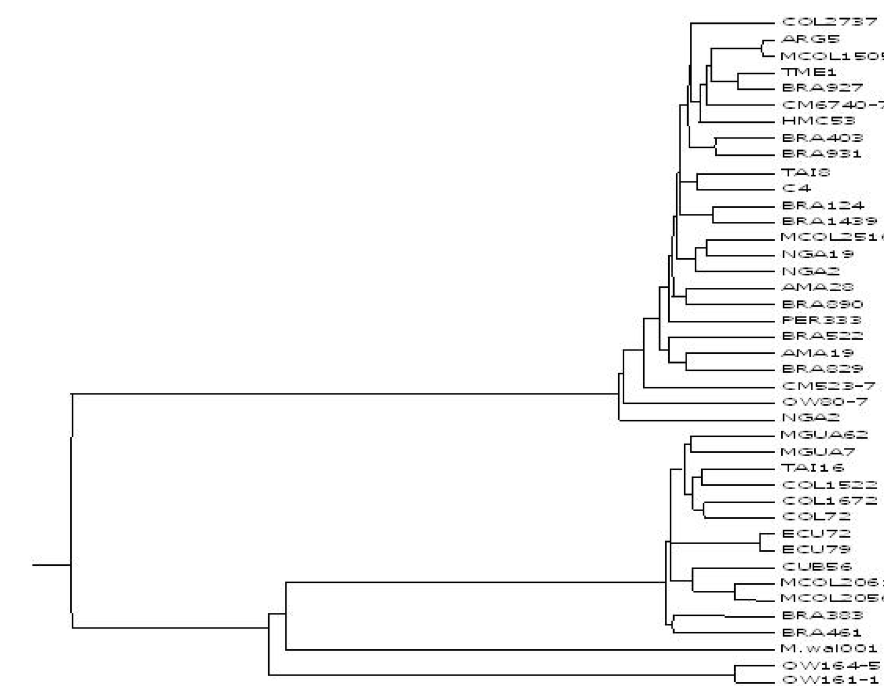


Figure 3. Genetic relationships among the cassava accessions used to identify polymorphic clones (markers) in the PstI/BstNI array. The dendrogram was created using the distance table based on 554 polymorphisms and the UPGMA clustering algorithm. Topology and branch lengths of this dendrogram are biased due to the presence of significant numbers of polymorphic clones derived from repetitive sequences.

There is a difference between the two dendrograms obtained with the PstI/BstNI and the PstI/TaqI array, respectively. A thorough inspection of the data suggests that the BstNI array contains a higher proportion of clones derived from repetitive sequences than PstI/TaqI array. Typing using repetitive sequences introduces a bias in genetic diversity analysis due to over-representation. The PstI/TaqI array does not show a high proportion of clones with repeated sequences and can be used as a routine genotyping tool for genetic diversity analysis. Work is ongoing to analyze a larger sub set of accessions from the CIAT core collection with at least 800 polymorphic clones

CONCLUSION AND PERSPECTIVES

This feasibility project has resulted in the design and validation of a reliable complexity reduction method that is uncovering several hundred polymorphic DArT markers in cassava germplasm. Because DArT markers can be scored in parallel in a single analysis, a high throughput, cost-effective whole genome genotyping is now available for determining high-density genome profiles of cassava. As this study has demonstrated, the availability of such a method has the potential to improve dramatically the recognition, conservation and exploitation of cassava genetic diversity. This platform is now available for research on genetic diversity and genetic mapping studies in cassava. With current improvements of DArT's throughput and its cost reduction it is feasible to molecularly characterize very large germplasm collections within a reasonable period of time and cost.

REFERENCES

Jaccoud, D., Peng, K., Feinstein, D., Kilian, A. (2001). Diversity Arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Research. 29: 4e25.

Kawano K. (2003). Thirty Years of Cassava Breeding: Biological, and Social Factors for Success. Crop Science 43:1325-1335.

Kawano K. (2001). The role of improved cassava cultivars in generating income for better farm management. In:Howeler R.H., and Tan S.L. (eds). Cassava`s Potential in the 21st Century: Present Situation and Future Research and Development Neds. Proceedings of the Sixth Regional Workshop, held in Ho Chi Minh City, Vietnam. Feb 21-25, 2000. CIAT-Bangkok, Thailand. Pp5-15.

ACKNOWLEDGEMENTS

Funds for this work was provided by IPGRI to CDV. We wish to thank Janneth Gutierrez for technical assistance