Saturation of the Genetic Map of Cassava with PCR-Based Markers and the Use of the Genetic Map in the Improvement of Cassava

(RF 96010 #20)

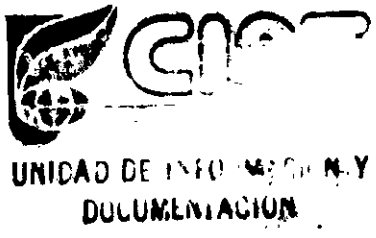
Progress Report

January 1997 – September 1998

and

Proposed Two Year Extension

Submitted to: The Rockefeller Foundation



By: Centro Internacional de Agricultura Tropical (CIAT)

October 1998

Contents

Progress Report to the Rockefeller Foundation on:

Saturation of the Genetic Map of Cassava with PCR-based Markers and the Use of the Genetic Map in the Improvement of Cassava

CIAT Project Component (RF96010 #20): "Saturation of the Genetic Map of Cassava with PCR-based Markers and the Use of the Genetic Map in the Improvement of Cassava"

Period of Report: January 1997- September 1998

Staff Involved at CIAT

Full Time:

Dr. M. Fregene, PDF, Nigeria (with support from RF96010 #20).

Janneth Gutierrez (B.Sc) Research Assistant, Colombia (with support from RF96010 #20).

Student Involved at CIAT:

Emmanuel Okogbenin, Ph.D. student, Nigeria (with support from RF)
Maria Christina Suarez, Undergraduate project student, Colombia (with support from RF96010 #20).

Part Time (CIAT Contribution to the Project):

Dr. J. Tohme, Geneticist, Biotechnology Research Unit, with CIAT core.

CIAT Units Involved in the Project

Biotechnology Research Unit

Summary

1. During the two-year period of the Rockefeller Foundation grant RF96010 #20, from January 1997 to September 1998, the following major outputs were achieved;

1.1 Generation of simple sequence repeat (SSR) enriched libraries from cassava; the identification of 1,220 putative di-, tri-, and tetra-nucleotide repeats after screening with the appropriate oligonucleotides; Isolation of Plasmid DNA from >500 of the putative SSRs in readiness for sequencing; Forward and reverse sequencing of >100 positive clones revealed simple and compound SSR with repeats ranging from 6-30. Primers have been designed for 60 of the new SSR markers.

1.2 An important output has been the ability of ongoing activities, specifically the development of a SSR-based genetic map of cassava, to attract support from non-traditional donors. CIAT has sought support from, and has been requested to submit a proposal to the International Science program, Uppsala, Sweden, to: apply SSR markers in the quantitative assessment of cassava genetic diversity in Africa, and to compare it to the total available in South America and recently identified wild progenitors, on a genome wide basis. The study, the first of its kind in cassava, will provide crucial information as to the existing genetic diversity of cassava in Africa and how to systematically broaden it for improving cassava yields given the irreducible uncertainties facing small holder farming systems.

1.3 Forty ESTs have been developed for cassava from >250 cDNA/AFLP polymorphisms identified between the parents of the F1 mapping population. They consist of mostly house keeping genes. Six of the ESTs have been mapped to the existing genetic map. ESTs generated as cDNA/AFLP fragments show a much higher level of polymorphism, 55%, than that found with random cDNAs, 20%.

1.4 The project has initiated a multilocational QTL trial to identify quantitative trait loci (QTL) controlling earliness, dry matter content, post-harvest deterioration (PHD), starch quality/content, culinary quality, and important morphological traits using the F1 mapping population. The QTL experiment, meant to run for two seasons, has yielded data for earliness for the first season; other traits will be measured at the end of the year. Heritability scores for earliness was 60 and 64% in the 2 locations, and two regions of the same linkage group was found to control >80% of genotypic variance in one location.

1.5 Construction of a Cassava (*Manihot esculenta* Crantz) Bacterial Artificial Chromosome (BAC) Library in collaboration with the Clemson University genome Center (Dr Rod Wing): Towards cloning of disease and pest resistance Genes. The

cassava BAC library has 55,296 clones, of average size 80kb and has a 5X coverage of the cassava genome.

1.6 The cassava mapping project is providing training for an African scientist in QTL mapping of traits of agronomic importance and their application in cassava breeding.

1.7 Three manuscript have been prepared, based on the outputs described in this progress report; one has been accepted in a refereed journal, another submitted as a book chapter, and a third is in preparation. Two posters, one describing the BAC library construction, has been presented in scientific meetings, and a third, reporting recent QTL mapping work, is in preparation for presentation at Plant Genome VI conference taking place in San Diego early 1999.

2. While the mapping project has generated several useful outputs, the biological nature of cassava, long growth cycle/ vegetative propagation, and the limited genetic knowledge and resources available, considerably lengthens the progress for developing markers for cassava breeding. A request for a two year extension, to complete work started in this project, is being submitted with this report to the Rockefeller Foundation for consideration. Principal activities enumerated in the proposal includes:

2.1 Genetic mapping of a few hundred SSR markers: towards a PCR-based map of cassava.

2.3 BAC contig mapping of regions of the cassava genome bearing disease resistance genes to the cassava bacterial blight (CBB), and the African cassava mosaic disease (ACMD).

2.2 Repetition of the QTL trial, to obtain data for the second year, and verification of putative QTLs by extending the study to a backcross population, developed by crossing 5 F1s from the mapping population to their female parent.

2.4 Sequencing and mapping of the ESTs.

2.5 Technology transfer of molecular marker technology for the assessment of genetic diversity to African national programs.

Technical Report

Saturation of the Genetic Map of Cassava with PCR-Based Markers and the Use of the Genetic Map in the Improvement of Cassava

Project RF96010 #20 (January 1997- September 1998)

Introduction

With support from the Rockefeller foundation, the this project is an effort to take results obtained from genetic mapping of the cassava genome and to apply them to overcoming constraints of cassava breeding. Considering the low technology status of most cassava breeding programs, the use of RFLP markers, currently the majority on the molecular genetic map of cassava, in marker-assisted studies of genetic diversity, genetic analysis of complex traits and marker-assisted improvement of cassava was questioned by collaborators in the National Agricultural Research services (NARs). RFLP markers, though highly informative, and the marker of choice for the heterozygous genome of cassava, are beyond the capacity of all centers in Africa and Latin America and most centers in Asia, and will be a major limitation of technology transfer.

Recent advances in the identification and development of simple sequence repeat (SSR) markers, combined with the PCR-based, co-dominant nature, speed of assay, and high levels of allelic diversity assay, make SSRs the second generation markers of choice for the saturation of the cassava genetic map and for transfer of marker technology to collaborators in the NARs. The major emphasis of the first 18months of this project was to bring together the resources, technology, and personnel required for the generation of a few hundred SSR markers for the cassava map. Modest financial support was received from the Swiss Development Cooperation (SDC) and the Swedish International Development Agency (SIDA). A cassava breeder from the National Root and Tuber Program in Nigeria was invited to participate in the entire process of SSR marker identification, development and mapping. An immediate application of the SSR markers to assessing cassava genetic diversity is also discussed.

A useful marker for genetic analysis of complex traits are expressed sequence tags (ESTs). ESTs are particulary useful markers for the candidate loci approach to mapping polygenic traits, to organizing YACs, BACs, or other large-insert DNA clones, into physical maps, and for identification of DNA regulatory sequences responsible for gene expression. ESTs markers have been widely used in the human, mouse and several plant genome, and the EST data base division of Gene Bank contains over 1.2 million sequences. ESTs are PCR-based and their development for cassava genome mapping is a component of this project.

A principal strategy in the application of markers to crop genetic improvement is the development of molecular tags, preferably PCR-based markers, for traits of agronomic interest. Their development requires scoring molecular markers, inherited in a simple Mendelian fashion, and the trait in segregating populations of simple pedigrees. Linkage disequilibrum between the markers and loci affecting the trait provides molecular tags that can be used to select parents with high breeding values for several complex traits simultaneously, and to select early for traits expressed late in the crop growth cycle. One of the major objectives of the project was the design and execution of an experimental trial to tag genes controlling traits of agronomic design in cassava.

Another dimension of applying the map to cassava breeding is by improving the efficiency of moving useful genes, especially disease and pest resistance genes, around cassava genepools. The rapid advance in cloning of plant disease resistance genes, together with reproducible transformation protocols in cassava makes genetic transformation with homologous disease and pest resistance genes, a faster and more efficient way of moving around useful disease resistance genes amongst cassava genepools. Cloning genes known only by their phenotypes and position relative to molecular markers on a genetic map requires libraries of large DNA fragments that can be ordered into contigs, that span the genomic region carrying the gene(s) of interest. The project prioritized generation of BAC libraries for cassava considering that disease resistance genes, controlling resistance to several strains of the cassava bacterial blight

(CBB), have been mapped, using the genetic map of cassava. Similar experiments are ongoing to identify genes controlling resistance to the cassava mosaic disease (CMD).

Report

1. Development of SSR Markers

A DNA library from variety TMS30572, was enriched for SSR markers using the following di-, tri, and tetra- nucleotides: $(CT)_8$, $(GT)_8$, $(CAA)_6$, $(CAG)_6$, $(AAT)_6$, $(ACG)_6$, $(GATA)_4$, and $(CAGA)_4$ and cloned into pUC18 plasmid, according to Edwards et al. (1996). The library was transformed into DH10B competent cells and total of 18,000 colonies from the enriched library was screened for SSR markers by hybridization with $(CT)_{20}$, $(GT)_{20}$, $(CAA)_{14}$, $(CAG)_{14}$, $(AAT)_{14}$, $(ACG)_{14}$, $(GATA)_{10}$, and $(CAGA)_4$ oligonucleotides and by anchored primer PCR (Stephenson 1998, unpublished data). A second enriched library from variety CMC40 enriched for $(CT)_8$, $(GT)_8$, developed in collaboration with Keith Edwards, Bristol University, UK, and cloned in pJV1, to increase the percent enrichment. About 2,100 clones were picked from the library and screened with the oligos $(CT)_{20}$, and $(GT)_{20}$ by dot blot screening. Plasmid DNA was isolated from putative clones by the QIAprep plasmid isolation kit, and 1-3ul of plasmid preparation was sequenced on an ABI377 automated sequencer, using the Universal and reverse M13 primers. Forward and reverse sequence strands were aligned using, either the GCG software, or the Basic Alignment Search Tool (BLAST) version 2.0 site of the National Center for Biotechnology Information found at http://www.ncbi.nlm.nih.gov/gorf/wblast2.cgi. Primer design was by the web based software Primer 3.0 found at http://waldo.wi.mit.edu/cgi-bin/primer/primer3.cgi.

A total of 120 putative clones were identified in the first library and sequenced. Of this, 64 clones contained unique SSR markers; primers, 20-mers long, could be designed for 60 clones from regions flanking the di-or tri—nucleotide repeats. Other clones had the SSR too close to the end of the fragment, or had very short, <4 di- or tri-nucleotide repeats. Results reveal a low efficiency of enrichment <1% enrichment. More than 1100 putative SSR markers were identified from the second library, corresponding to an

enrichment of 50%. Plasmid DNA was isolated from 550 clones and 100 of these clones have so far been sequenced, all clones contained the corresponding di-nucleotide repeat, with some 10% having the microsatellite too close to one extreme. Primers are being designed that will yield three different PCR product sizes or bins, to facilitate multiplexing with the ABI377 automated sequencer (Perkin Elmer) and fluorescent primers. Sequencing of putative SSR clones from the second enrichment is continuing; the target by the end of the year is 500 sequenced clones with their corresponding primer pair. Genetic mapping of the SSR markers is expected to begin at the beginning of next year.

2. Application of SSR Markers

The irreducible uncertainties of crop production in Sub-Saharan Africa occasions an austere description for a major staple crop. Cassava's hardiness, low labor and resource requirements fits the harsh realities of most farming systems in the sub continent, demonstrated by its increasing importance as a staple and food security crop of the region (Jiggins et. al. 1996). A major concern of cassava production in Africa today is steadily decreasing or stagnating yields due to pests, diseases, soil fertility, arising from shorter cycles of shifting cultivation and cropping of marginal soils, a lack of labor and resource inputs, drought and other climatic factors. It has been suggested that broadening local germplasm with disease, pest, root quality and nutrient-use efficiency genes, while endeavoring to preserve most of farmer-preferred traits might be a solution to cassava improvement in Africa (Gullberg 1998 pers. comm.). This is a task that requires maximizing genetic variation to achieve adaptable combination of alleles. But cassava is of South American origin. It is not known what proportion of genetic diversity of cassava present in South America has been transferred to Africa. . The recent discovery that cassava has only 25% of genetic variation found in its wild progenitors (Olson et. al. 1998 in preparation), compared to 75% in maize (Eyre-Walker et. al. 1998) has raised prospects for the broadening the genetic base of the crop.

A cassava molecular diversity network to evaluate genetic diversity of cassava in Africa has been proposed and support sought from the International Science Foundation in

Sweden. The Network, consists of a geneticist, molecular geneticist, population geneticist, plant breeder, and germplasm scientists from CIAT, the Washington University, St. Louis, USA, the Swedish Agricultural University, Uppsala, Sweden, IITA and National Programs in Africa. The Network intends to collect cassava varieties, GIS, and agronomic information from selected African countries, and evaluate the accessions with a sub-set of SSR markers spread over the entire genome at CIAT. Ongoing and concluded cassava genetic diversity studies of South American accessions, with SSR markers, will be expanded to obtain a representative estimate of diversity in the crop's center of diversity. Further collections of wild progenitors of cassava, recently identified in the Southern rim of the Amazonian region will also be evaluated. The study will yield a quantitative estimate of allelic richness of African cassava varieties versus the total available in South America and wild progenitors of cassava. The genetic diversity study is expected to begin in 1999.

3. ESTS

Total RNA was isolated from young leaves and roots of the two genotypes of the parents of the cassava mapping population TMS30572 y CM2177-2 at 4 weeks after planting. Messenger RNA was isolated from total RNA by Dyna beads and cDNA was synthesized from the mRNA templates as described by Bachem et al. (1996). Using the modified AFLP method for DNA templates (Bachem et al, 1996), a preamplification PCR without selective bases primers and a subsequent amplification with two selective nucleotides primers, using all the possible 256 combinations, was done on the cDNA library. The PCR reaction was electrophoresed on a 5% sequencing gel. Polymorphic AFLP bands corresponding to TDFs differentially expressed in the parents were excised from the dried gels, eluted and reamplified.

The eluted cDNA fragment was cloned into pGEMT (Promega Inc) and transformed into *E.Coli* strain, DH5 alpha. Plasmid DNA was isolated from overnight miniprep cultures and TDFs were PCR labelled for use as probes in an RFLP screen of parental filters. Filters contained genomic DNA from the parental genotypes digested with five enzymes: EcoRI, HindIII, HaeII, AseI, TaqI. TDFs found polymorphic in the parental survey were

hybridized to progeny filters prepared with the corresponding enzyme. Raw marker data was combined with previous segregation data from the framework map of cassava and linkage analysis was by the MAPMAKER program. New markers are added to the existing framework map with a LOD thresh hold score of 3.0. TDFs were also sequenced, regardless of if they have been mapped or not using the automated ABI 377 sequencer in the Biotech Research Unit, CIAT. The Basic Alignment Search Tool (BLAST) version 2.0, of the National Center for Biotechnology Information found at http://www.ncbi.nlm.nih.gov/gorf/wblast2.cgi was used to search for homology to known genes published sequences of Gene Bank.

More than 500 TDF polymorphisms were obtained from the AFLP screening using 256 2-nucleotide selective primer combinations. All TDFs were eluted from the gel and PCR amplification preformed on a 200. A sub-set of forty eight TDFs were randomly chosen, and cloned. Forty four TDFs were successfully cloned and selected for mapping; the parents of the mapping population were screened with the 44 TDFs using 5 restriction enzymes viz: EcoRI, HindIII, Hae III, Ase I, and Taq I. Level of RFLP polymorphism in the parents was 55, and six of the polymorphic TDFs have so far been screened in the progeny and mapped to the existing frame work of cassava, four to the female-derived map and two to the male-derived map. The others have to be confirmed in the hybridization with the progeny. Forty four of the TDFs were also sequenced and homology to known genes in revealed mostly housekeeping genes. They include Chlorophyll binding proteins, in *Lycopersicum esculentum* and *Brassica napus,* a transcription factor from *Nicotiana tabaccum* PRL-2 genes from *Arabidopsis thaliana* , and an AFLP fragment from *Dioscorea tocoro.* These sequences, in the process of submission to the EST data base of the GenBank.

Complementary DNA derived from cDNA/AFLPs show much higher RFLP polymorphism compared to random full length cDNAs found during genetic mapping of cassava, 55% versus 20%. The technique is therefore a powerful way to map many more cDNAs, using the highly informative RFLP technique, towards a transcript map of cassava. Preliminary QTL mapping with mapped TDFs reveals the power of mapped

cDNAs to uncover genes involved in complex trait, a chlorophyll binding protein mapped close to a QTL for dry matter content and explained 10% of phenotypic variance. However, the short length of the TDFs, 200-400bp considerably reduces sensitivity of Southern hybridization resulting in rather weak signals. Other marker techniques, such as single nucleotide polymorphisms (SNP) might provide a way around the RFLP bottle neck encountered in mapping TDFs, primers are being designed from TDFs sequences to implement the SNP mapping approach. Results thus far have shown that it is possible to develop a transcript map of cassava that might simplify genetic analysis of agronomic traits of interest. More TDFs will be cloned, sequenced and mapped and their putative identities determined

4. Genetic Analysis of Agronomic Traits

The F1 mapping population, from which the cassava map was developed, was used as the QTL mapping population; the F_1 cross has been genotyped with over 200 RFLP, SSR, isozyme and RAPD markers, including known starch biosynthesis genes, cyanogenesis genes, and expressed sequence tags (ESTs). In 1997 the 150 genotypes of F_1 cross was multiplied in preparation for the multi-locational, replicated trial. Thirty plants from each genotype were established in the field, April 1997, and stakes, or planting materials harvested at 9 months. The QTL trial was a triple partially balanced lattice design with 20 plant per genotype, three replications, in two sites, CIAT, Palmira and CIAT, Quilichao. The Quilichao experiment was planted on the 15th of January 1998, while the CIAT plot was established on the 29th of January, both plots were irrigated prior to planting. A total of 144 genotypes were planted in both locations.

To test for early genotypes, storage roots from three plants from the border row of genotype plots, were harvested at 7 months, or early August, and dry matter content of the roots determined. Leaf morphology was also be measured at 6 months. Dry matter, starch content/quality, post-harvest deterioration, culinary quality and important plant morphological characteristics such as plant height, height of first branching, stem color etc.harvested at maturity, will be scored at harvest in November '98. Heritability of the different traits will be calculated from environmental, genotypic and total variance

derived from the ANOVA of phenotypic data. Single point marker analysis, by regression, and t-test with conditioning will be carried out on the mean of phenotypic scores from the three replication by the QTL analysis software package Q-gene and PGRI.

Storage roots was harvested from three border plants from each plot at 7 months. Fresh roots from all three plants were weighed in air and water, and the specific gravity of roots determined, which was later used to calculate dry matter. Phenotype values were averaged over replication for each site and analyzed by single point analysis by regression, with markers from the female-derived map being the independent variable, using the Q-gene package. Table 1 shows significant QTLs ($P<0.01$) found only on chromosome D for weight of storage roots at 7 months for the data from the Quilichao trial. Two regions on Chromosome D explains more than 60% of phenotypic variance. Heritability estimates, calculated from the ANOVA, using SAS (SAS Institute) was 60% for Palmira, and 64% for Quilichao. Genotypic variance explained by markers from the two regions was approximately 60%. Inter-locus interaction, or epistasis, is being analysed to uncover other sources of genotypic variance. No significant QTLs were found with the CIAT, Palmira data. The experiment will be repeated the following year and in controlled trials in the green house to confirm the identified QTLs. The leaf morphology data is currently being analyzed. The earliness trait will be re-analyzed in controlled green house experiments with 40 genotypes from the extremes of the trait to measure components of earliness such as, increase in girth of secondary roots, time of first starch accumulation, or differentiation of cambia, rate of starch accumulation, etc, and thus increase the power of the experiment to detect earliness genes.

5. BAC Library Construction

Megabase size DNA was isolated from the cassava variety TMS30001 and embedded in agarose plugs as described by Zhang *et al.* (1995). TMS30001, developed at the International Institute of Tropical Agriculture (IITA), shows extreme resistance to the cassava mosaic disease (CMD) and resistance to the cassava bacterial blight (CBB). Large genomic DNA, in one-third of an agarose plug, was partially digested with Hind

III, 1.5U for 20 min at 37°C, and DNA fragments of sizes 100-300kb size selected by pulse field gel electrophoresis (CHEF MAPPER, Bio-Rad Corp). Size selected DNA was ligated into the Hind III cloning site of pBeloBAC11 in a vector:insert ratio of 10:1, using 14U of ligase, in a final volume of 100µl. Twenty microliters of DH10B competent cells (GIBCO BRL) were transformed, with 2µl of the ligation reaction, by electroporation, and white colonies picked for DNA insert sizing. Colonies were grown for 14hr in LB+ 30mg/ml Chlorampenicol and plasmid DNA isolated by the Autogen automatic plasmid isolation robot (Kurabo Inc.). Plasmid DNA was digested with 10U of *Not* I to liberate inserts and separated on a 1% agarose gel by pulse field gel electrophoresis. The rest of the ligation was transformed, plated out and picked with the Q-bot robot (Genetix PLC).

Twenty colonies carrying recombinant plasmids were randomly picked and inserts were sized. DNA insert sizes ranged from 40 to 110kb, with an average of 80kb. Using the Q-bot robot (Genetix PLC), 55296 clones were picked and transferred to 384-well plates containing culture media suitable for storage of bacteria colonies at -80°C. A total of 144 plates were required to store the library. The BAC library was also gridded onto 3 high density filters, each containing 18,432 clones per filter, for rapid screening of the library. The cassava genome has a size of 760 million base pairs. The cassava BAC library therefore has a 5X coverage of the genome and a 95% probability of finding any desired DNA clone. Screening the BAC library with cassava genomic DNA clones has so far revealed a satisfactory representation, up to 30 hits. The BAC library was constructed with the objective of map-based cloning of disease and pest resistance genes in cassava. A good commencement point of cloning resistance genes will be already identified markers tightly linked to CBB resistance genes. Fine mapping of these genomic regions has also been initiated, to be followed by assembling a contig of BAC clones traversing the region.

The cassava BAC library has 55,296 clones, of average size 80kb available for gene cloning. The library is publicly available through CIAT. The next stage of this project will be to use the BAC library to construct contigs of regions already identified in cassava carrying resistance genes to CBB. A USAID University/International Agricultural

Research Centers Linkage (UNIARCL) Program proposal was submitted along with Dr Rod Wing of the Genomics Institute, Clemson University to clone resistance genes to the African cassava mosaic disease and to the white fly disease. The proposal was approved and work on contig mapping will begin shortly.

Future Work Planned

- Completion of the Sequencing of putative SSR markers, primer design and genetic mapping of the SSR markers
- Transfer of the SSR marker technology to national programs through the activities of the Cassava Molecular diversity Network.
- Assessment of genetic diversity of cassava in Africa versus total diversity prsent in the crop's center if diversity, and wild progenitors: towards a systematic approach of broadening genetic diversity for cassava improvement.
- Completion of the first year field evaluation, of the F1 mapping population for dry matter, starch content/quality, post-harvest deterioration, culinary quality and important plant morphological characteristics such as plant height, height of first branching, stem color etc. harvested at maturity. Second year evaluation.
- BAC contig mapping of regions of the cassava genome carrying disease resistance genes.
- Genetic mapping of all 44 sequenced ESTs and sequencing of the remaining TDFs identified.

Other Outputs of the Project

Important spin-offs from this project has been:

a) SSR markers develop earlier in this project are being used by collaborators in Sweden for farmer-participatory breeding schemes, to identify farmer varieties and elucidate selection methods practiced by farmers (Nkumbira et. al. in preparation).

b) The cassava molecular diversity network, intends to take SSR technology out of the laboratory and apply it to the evaluation of genetic variation of cassava in Africa.

c) Disease resistance genes, controlling resistance to all know strains of the cassava bacterial blight (CBB) have been localized on the existing map of cassava. This work has been conducted at CIAT under a ORSTOM-CIAT collaboration.

d) Training of an African cassava breeder, and an African Ph.D. student in new tools for breeding of cassava

e) Increased donor support and attention, in particular a USAID University/International Agricultural Research Centers Linkage (UNIARCL) Program grant to CIAT and Clemson University for gene cloning in cassava

f) The Cassava genome data base has been completed and will be made publicly available through the USDA www site of the plant genome initiative.

Publications from the Project

Thro A.M.T., Roca W., and Fregene M. (1997). Genetic Biotechnologies and cassava-based development of marginal rural areas. In: Dennis E.S., Hohn B., Hohn Th, Meins F., Schell J., Verma D.P.S. (eds). Plant Gene Research Series. Springer Verlag, Berlin.

Chavarriaga-Aguirre P., Bonierbale M., Fregene M., Kresovich S., Tohme J., and Kochert G. (1998). Microsatellites in cassava (*Manihot esculenta* Crantz): characterization, inheritance and variability. Theor Appl Genet (in press).

M. Foregone, F. Angel, R. Gomez, F. Rodriguez, M. Suarez, J. Gutierrez, P. Chavarriaga, M. Bonierbale, W. Roca, V. Jorge, V. Verdier, and J. Tohme. 1998 Molecular Genetic Mapping in cassava. Red Bio, Havana Cuba June 1998

Fregene M. A., Gallego G., Lopez J., Wing R., Roca W., Tohme J. 1998. Construction of a Cassava (*Manihot esculenta* Crantz) Bacterial Artificial Chromosome (BAC) Library: Towards Cloning of Disease and Pest Resistance Genes. Red Bio, Havana Cuba June 1998.

Suarez M. , Bernal A., Gutierrez J., Tohme J., Fregene M. 1998. Development of expressed sequence tags for genetic mapping in cassava. (In preparation).

Fregene, M.A., Okogbenin , E., Duque M., Tohme, J., and Roca, W.M. (1999). Mapping of root quality traits and earlines in cassava (*Manihot Esculenta* Crantz). Poster for presentation at Plant Genome V, January 17-21 1999, San Diego California.

References

Bachem CWB, Van der Hoeven R.S., de Bruijn S. M., Vreugdenhill D., Zabeau M., and Visser R. G. F. 1996. Visualization of differential gene expression using a novel method of RNA fingerprinitng based on AFLP: Analysis of gene expression during potato tuber development. The Plant Journal 11996) 9(5), 745-753.

Edwards KJ, Barker JHA, Daly A, Jones C., and Karp A 1996. Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques vol 20, No. 5:758-760

Eyre-WalkerA., Gaut R. L., Hilton H., Feldman D., and Gaut B. 1998. Investigation of the bottlenecks leading to the domestication of maize.

Jiggins J., Reinjntjes C., Lightfoot C. 1996. Mobilizing Science and Technology to get Agriculture moving in Africa: A response to Borlaug and Dorswell. Development Policy Review March 01 1996, v14, n1, 89.

Olson K., Allem C., Schaal B. 1998 Phylogeography of *Manihot esculenta* based on a single-copy nuclear gene: Implications for the origins of cassava (in preparation).

Zhang H., Zhao X., Ding X., Paterson A., Wing R. 1995. Preparation of megabase-size DNA from plant nuclei. The Plant Journal 7(1): 175-184.

Table 1. Markers of linkage group D of the genetic map of cassava showing significant linkage to a QTL effect for storage root dry matter weight at 7 months. QTL analysis used only data from the Quilichao trial.

Marker	Chrom	F	RSq	P	AA	SE	N	Aa	SE
Ai18b	NgD	15.72	0.1733	0.0002	521.72	32.46	39	695.5	29.35
GY219	NgD	14.45	0.1598	0.0003	530.49	33.07	41	696.76	27.86
GY42	NgD	14.43	0.1632	0.0003	530.8	33.25	41	701.23	28.97
rGY167	NgD	14.37	0.1608	0.0003	536.93	31.7	44	706.36	29.55
rGY180	NgD	14.24	0.1578	0.0003	532.95	32.6	42	698.5	28.23
GY181	NgD	14.01	0.1556	0.0004	535.42	31.93	43	700.2	29
GY50	NgD	13.96	0.1569	0.0004	533.93	32.57	42	700.29	29.14
rGOT-2	NgD	13.81	0.1521	0.0004	537.47	30.49	47	704.25	30.81
GY125	NgD	13.8	0.1554	0.0004	536.26	32.11	43	702.24	29.55
GY222	NgD	12.18	0.1447	0.0008	532.27	33.4	41	692.3	29.82
GY179	NgD	10.51	0.1229	0.0018	539.05	34.14	40	685.76	29.16
O19	NgD	8.97	0.1068	0.0037	554.42	30.85	40	684.22	30.26
AD4e	NgD	8.89	0.1086	0.0039	550.79	31.97	38	682.32	30.34
K11b	NgD	7.52	0.097	0.0077	536.29	35.21	38	669.35	32.88

Theor Appl Genet (1997) 95:431–441

M. Fregene · F. Angel · R. Gomez · F. Rodriguez
P. Chavarriaga · W. Roca · J. Tohme
M. Bonierbale

A molecular genetic map of cassava (*Manihot esculenta* Crantz)

Received: 5 July 1996 / Accepted: 22 November 1996

Abstract A genetic linkage map of cassava has been constructed with 132 RFLPs, 30 RAPDs, 3 microsatellites, and 3 isoenzyme markers segregating from the heterozygous female parent of an intraspecific cross. The F_1 cross was made between 'TMS 30572' and 'CM 2177-2', elite cassava cultivars from Nigeria and Colombia, respectively. The map consists of 20 linkage groups spanning 931.6 cM or an estimated 60% of the cassava genome. Average marker density is 1 per 7.9 cM. Since the mapping population is an F_1 cross between heterozygous parents, with unique alleles segregating from either parent, a second map was constructed from the segregation of 107 RFLPs, 50 RAPDs, 1 microsatellite, and 1 isoenzyme marker from the male parent. Comparison of intervals in the male- and female-derived maps, bounded by markers heterozygous in both parents, revealed significantly less meiotic recombination in the gametes of the female than in the male parent. Six pairs of duplicated loci were detected by low-copy genomic and cDNA sequences used as probes. Efforts are underway to saturate the cassava map with additional markers, to join the male- and female-derived maps, and to elucidate genome organization in cassava.

Key words Cassava · Molecular markers · Genetic mapping · Polyploidy

Communicated by J. W. Snape

M. Fregene (✉) · F. Angel · F. Rodriguez
W. Roca · J. Tohme
Biotechnology Research Unit, CIAT, Cali, Colombia

M. Bonierbale
Cassava Program, CIAT, Cali, Colombia

R. Gomez
Centro Internacional de Fisica (CIF), Bogota, Colombia

P. Chavarriaga
Department of Botany, University of Georgia at Athens, Athens, GA, USA

Introduction

The genetics of cassava (*Manihot esculenta* Crantz) are the least understood of any of the major staple crops (including rice, maize, wheat, and potatoes) that feed mankind. This discrepancy is due to the heterozygous nature of the crop, a lack of classical markers, its long growing cycle of 9–18 months, its low seed yield per pollination (a maximum of three seeds per cross), and the limited funding for research on this important crop. Cassava is a strongly outcrossing monoecious species (mediated by protogyny) and suffers from inbreeding depression, making it difficult to develop appropriate stocks for classical genetic studies. Although the crop is considered to be a segmental allopolyploid (Magoon et al. 1969) or an allopolyploid (Umannah and Hartman 1973), nothing is known about the diploid ancestors of cassava's 36 somatic chromosomes, nor has a classical genetic map been developed. Biochemical marker sudies have recently provided new insights into the cassava genome. Analysis of ten isoenzyme loci in cassava revealed predominantly disomic inheritance (Roca et al. 1992; Levefre 1993). Comparison of intralocus and interlocus heterodimeric bands from isoenzyme analysis of diploid tissue (leaves) and a haploid mixture (pollen grains) made it possible to differentiate between heterozygosity (which is characteristic of cassava) and duplicated loci, providing evidence of locus duplication in at least two of the isoenzyme loci studied (Levefre and Charrier 1993).

Native to the New World tropics and a member of family Euphorbiaceae, cassava and some 90 other species make up the genus *Manihot*. Cassava is the only widely cultivated member, being valued for its starchy tuberous roots, which provides food or an export commodity in nearly all tropical countries of the world. Evidence of the allopolyploid origin of cassava ($2n = 36$) relies heavily on its possession of two sets of dissimilar nucleolar organizing regions, on the

repetition of chromosome types (Magoon et al. 1969; Umannah and Hartman 1973), and on the basic chromosome numbers of other genera in the Euphorbiaceae, which range from 6 to 11 (Perry 1943). No evidence of tetrasomic inheritance or of wild *Manihot* relatives with chromosome numbers of $2n = 18$ has been found, which supports the allopolyploid theory in cassava. On the contrary, chromosome pairing in interspecific hybrids between cassava and 10 *Manihot* species, all possessing $2n = 36$ chromosomes, shows comparatively normal meiosis, with 18 bivalents at prophase and occasionally 2–4 univalents or one tetravalent, suggesting a high percentage of genome similarity between cassava and its wild relatives (Hahn et al. 1990; Bai et al. 1992). Normal chromosome pairing at meiosis occurs in hybrids of *Manihot* species that are morphologically very different and belong to separate primary phylogenetic lineages according to cpDNA studies (Fregene et al. 1994). This combined with the lack of a strong interspecies boundaries in the genus has led to the theory that a polyploidization event occurred before species differentiation (Lefevre and Charrier 1993; Bryne 1984).

Flow cytometry measurements of nuclear DNA in cassava have revealed a diploid DNA content of 1.67 pg per cell nucleus (Awoleye et al. 1994). This value corresponds to 772 mega-base pairs in the haploid genome and puts cassava's genome size at the lower end of the range for higher plants (Bennet et al. 1992). The relatively small size of this genome favours the development of a saturated genetic map and molecular tags, which may contribute to an understanding of the inheritance of many important quantitative traits despite the heterozygous nature of cassava.

The literature describes genetic approaches to mapping polyploid genomes with molecular markers (Ritter et al. 1991; Wu et al. 1992; Al Janabi et al. 1993; Da Silva et al. 1994). These approaches attempt to simplify the determination of allelism by analyzing a special class of markers known as single-dose restriction fragments (SDRFs) (Wu et al. 1992). SDRFs are DNA markers that are present in one parent and absent in the other and segregate in a 1:1 ratio (absence:presence) in the progeny. They represent the segregation equivalent of an allele at a heterozygous locus in a diploid or an allopolyploid genome or a simplex allele in an autopolyploid. Linkage analysis using SDRFs in an F_1 population requires the presence of a number of unique segregating polymorphisms (heterozygosity) and normal meiosis in either or both parents. It results in two separate linkage maps based on male and female sources of markers.

Our objectives were to construct a molecular genetic linkage map of the cassava genome and elucidate genome organization using an intraspecific *Manihot esculenta* F_1 cross designed to segregate for several important traits. As a first step, several restriction enzyme-by-genomic probe combinations were assessed in a set of cultivars from different cassava growing regions (Angel et al. 1993). Restriction fragment length polymorphism (RFLP) between the most genetically dissimilar pair was around 40%, indicating that intraspecific crosses are suitable for mapping in cassava. A molecular genetic map of cassava has the promise of helping to identify molecular markers linked to traits of interest and to apply these in cassava improvement. Marker-enhanced selection aids the identification of recombinant individuals by permitting a more exact assessment of genotypes bearing introgressed genes of interest, with minimum flanking DNA, than is possible through phenotype or progeny testing alone (Young and Tanksley 1989). In addition, markers are a powerful tool for studying the genetics of complex traits.

Materials and methods

Plant materials

The cassava mapping population is comprised of 90 F_1 plants from an intraspecific cross between 'TMS 30572' (the female parent), an elite cassava cultivar developed at the International Institute of Tropical Agriculture (IITA), Nigeria, and 'CM 2177-2' (the male parent), a successful cassava cultivar resulting from breeding at the Centro Internacional de Agricultura Tropical (CIAT) in Colombia. The 'TMS 30572' × 'CM 2177-2' cross (CM 7857) was chosen from a preliminary evaluation of three intraspecific crosses using RFLP and random amplified polymorphic DNA (RAPD) markers to determine the percentage of single-dose markers (Gomez et al. 1995). Apart from being heterozygous, due to the fairly large number of diverse cassava accessions in their pedigrees, the parental accessions have the following advantages:

1) 'TMS 30572' is highly tolerant to the African cassava mosaic disease (ACMD).
2) 'CM 2177-2' shows high photosynthetic rates, good cooking quality, and tolerance to the cassava mealy bug.
3) both are tolerant to cassava bacterial blight (CBB).

Resistance to ACMD is thought to have been introgressed from *M. glaziovii* into lines leading to 'TMS 30572'. While a subset of 90 individuals was used to develop the first framework map, a total of 150 individuals from the same cross is available for higher resolution and quantitative trait mapping. Clonal propagation from stem cuttings in the greenhouse and the field provided leaf tissue for molecular analysis and stocks for the production of roots for isoenzyme analysis.

DNA isolation and restriction enzyme analysis

Genomic DNA was isolated from fresh, young tissue of cassava leaves according to Dellaporta et al. (1983). The total DNA obtained was dissolved gently overnight at 4° C and quantified by fluorimetry (TKO 100 Hoefer). Large restriction enzyme digestions to produce uniform Southern blots were performed with *Eco*RI, *Eco*RV, *Hae*III, *Hin*dIII, and *Pst*I according to the manufacturer's instructions (Amersham PLC). Ten micrograms of digested cassava genomic DNA was loaded onto 0.9% agarose gels and electrophoresed in TBE buffer for 16–18 h. Separated DNA fragments were transferred by the alkaline method of Southern blotting to Hybond N$^+$ nylon membranes (Amersham), as recommended by the manufacturer.

Parental survey filters contained DNA of the two parents digested with the five above-mentioned restriction enzymes. Progeny filters containing restricted DNA from the 90 F_1 plants, including DNA from both parents in the first 2 lanes, were replicated, four to eight times per restriction enzyme.

Genomic DNA, cDNA libraries, and Southern hybridization

Angel et al. (1993) have described the construction of nuclear genomic libraries, totaling 200 clones, from cassava DNA sequences generated with *Hin*dIII, *Xba*I, *Eco*RI, and *Pst*I. Two other *Pst*I genomic libraries of about 1,500 clones each were also constructed from cassava nuclear DNA according to standard procedures (Sambrook et al. 1989; Vayda et al. 1991). The copy number of DNA sequences was estimated by Southern hybridization of total cassava DNA, labelled with Horseradish peroxidase, according to the "ECL direct" protocol of Amersham PLC, with dot blots of whole plasmid preparations. Genomic clones that produced a strong signal after two 0.5× SSC, 0.4% SDS washes at 55 C for 10 min, detection, and autoradiography were judged to be repetitive sequences and were left out of the parental survey.

For the cDNA libraries, total RNA was extracted from leaf tissue using the method of Bothwell et al. (1990); polyadenylated RNA was purified away from the total RNA preparation using Dyna beads oligo (dT)25 (Dynal inc), and a cDNA library was constructed from messenger RNA in lambda ZAP II vector, following the protocol of the suppliers (Stratagene). Approximately 17,500 recombinant clones were obtained in the primary library. A total of about 200 cDNA clones were isolated; they ranged in size between 0.6 and 1.3 kb, as determined by polymerase chain reaction (PCR) amplification of inserts using T3 and T7 primers (Stratagene). For both cDNA and genomic clones, probes were prepared for Southern hybridization by PCR amplification using the appropriate primers. T3 and T7 primers were used for cDNA clones in lambda ZAP II and genomic clones in pBluescript, and M13 forward and reverse primers (New England Biolabs) for clones in PUC 18. A PCR thermal profile of 1 min at 94 C, 34 cycles of 30 s at 94 C, 1 min at 55 C, and 2 min at 72 C, with a final extension time of 10 min at 72 C in a Perkin Elmer-Cetus thermo-cycler, was used in insert amplification.

Southern hybridization of parental surveys of the cDNA and genomic libraries was carried out using the ECL direct nucleic acid hybridization and detection kit of Amersham PLC. A batch of 30 survey filters (6 × 4 cm) was routinely hybridized in small plastic boxes (6 × 6 cm) with 10 ml of ECL direct hybridization buffer and 100–200 ng of labelled probe for 16–18 h; this was followed by two medium stringency washes of 0.5× SSC, 0.4% SDS for 10 min at 55 C, then chemiluminescent detection, and autoradiography for 1–2 h. Filters were used more than 20 times. This provided a fast and relatively inexpensive way to screen large numbers of RFLP clones; about 150 were handled in a 5-day week. Southern hybridization of progeny filters, with a greater emphasis on clarity of autoradiograms to permit unambiguous scoring of RFLP data, was done with $\alpha[^{32}P]$dNTP-labelled probes. Purified PCR products of DNA inserts (using Sephadex G-50 columns) from both genomic and cDNA libraries were labelled with $\alpha[^{32}P]$dATP by the random primer method of Feinberg and Volgestein (1983) to a specific activity of about $1–8.0 \times 10$ cpm and employed as probes in Southern hybridization using the Church and Gilbert (1984) hybridization buffer at 65 C. Prehybridization and hybridization were for 4 and 18 h, respectively. Post-hybridization washes were at 65 C with 2× SSC, 0.1% SDS for 30 min and 0.5× SSC, 0.1% SDS for 20 min, followed by autoradiography for 2–4 days at −80 C with one or two intensifying screens. Filters were reprobed. The previous probe was stripped off by incubating the filters in 0.5% SDS solution (initial temperature 100 C) at 65 C for 30 min, up to 12 times without any loss of the hybridization signal.

RAPD analysis

A slightly modified version of the protocol described by Yu and Pauls (1992) was used to obtain RAPD data (Gomez et al. 1995). A total of 740 10-nucleotide oligos of random sequence (Operon Technologies, Alameda, Calif.) served as primers for the amplification of parental DNA and a subset of the progeny in order to detect polymorphisms. Amplification reactions were carried out in a 12.5-µl volume containing 25 ng of cassava genomic DNA, 10 m*M* TRIS-HCl (pH 9.0), 50 m*M* KCl, 0.01% Triton X-100, 2.5 m*M* $MgCl_2$, 0.2 m*M* of each dNTP (New England Biolabs), 0.8 µ*M*, of primer and 1 U *Taq* DNA polymerase (Perkin Elmer-Cetus Corporation).

Microsatellite analysis

Cassava nuclear DNA was digested to completion with *Eco*RI and *Xba*I restriction enzymes (New England Biolabs) according to the manufacturer's instructions; 0.3- to 0.8-kb fragments were isolated from a 0.5% low-melting-point agarose gel using the phenol chloroform purification procedure (Sambrook et al. 1989). Purified fragments were cloned in lambda ZAP II (Stratagene), and 8×10^5 recombinants clones were obtained, as determined from the IPTG, X-gal screening procedure. The library was then screened with (GA)15, (GT)15, (AT)15, (TTAG)8, and (TCT)10 oligonucleotides. Positive clones were sequenced on an automated sequencer (Applied Biosystems) using T3 and T7 primers. Primers were designed for microsatellites containing no less than ten perfect or imperfect repeats using the PRIMER 0.5 software. PCR reactions to search for microsatellite polymorphism between the parents and to score microsatellite loci in the progeny were carried out in a 100-µl volume containing 0.1–0.7 ng/µl genomic DNA, 0.2 µ*M* of each primer in 10 m*M* TRIS-HCl, 50 m*M* KCl, 1.5 m*M* $MgCl_2$, 0.01% gelatin, 200 µ*M* each of dNTPs, and 2.5 U of Ampli-*Taq* polymerase (Perkin Elmer-Cetus). One denaturation cycle was performed at 94 C for 5 min prior to 30 cycles of denaturation at 94 C for 1 min, annealing at 45 C or 56 C for 1 min, extension at 72 C 1 min, and a final extension at 72 C for 5 min. PCR-amplified products were visualized on 5% metaphor agarose gels stained with EtBr, or labelled directly by incorporation of $\alpha[^{32}P]$dATP during PCR, and run on 6% polyacrylamide gels.

Isoenzyme analysis

Nine isoenzyme systems (i.e., alpha and beta-esterase, acid phosphatase (acp), diaphorase, peroxidase, glutamate oxaloacetate transaminase (got), shikimate dehydrogenase (skdh), and malate dehydrogenase) were examined in the parents of the mapping population. Beta-esterase yielded a single-dose fragment segregating from the male gametes, while shikimate dehydrogenase, diaphorase, glutamate oxaloacetate, and acid phosphatase had unique alleles segregating from the female gametes. Protocols for the isoenzyme assay have been described elsewhere (Ocampo et al. 1992).

Data analysis and map construction

Monogenic segregation ratios of present, absent classes for each marker were tested for goodness of fit to the expectation of 1 : 1 to identify single-dose markers by chi-square analysis. Two separate mapping data sets were obtained from the segregation of single-dose markers in the F_1 mapping population. One data set was from single-dose markers segregating in the gametes of the female parent, and the other was from single-dose markers segregating in the gametes of the male parent. Markers with segregation ratios

significantly different from the expected ratio of 1 : 1 were tested for other possible ratios – such as 3 : 1, which is expected for the segregation of double-dose markers on homocologous chromosomes of an allo- or autopolyploid (Wu et al. 1992). Double-dose markers represent the duplex (double simplex) condition at a heterozygous locus. In order to compensate for the random assignment of "1" or "0" to alternate alleles at a locus and to detect linkage in repulsion, we duplicated the data matrices, and markers of the second half were recoded by inverting the scores before linkage analysis. This resulted in a mirror image of each linkage group, which was later discarded.

The unduplicated mapping data sets consisted of 195 markers for the female and 203 for the male parent. The test for linkages was done using the computer package MAPMAKER 2.0 running on a Macintosh Centris 650 and Mapmaker 3.0 Unix version on a SPARC workstation (Lander et al. 1987). A LOD score of 4.0 and recombination fraction of 0.30 served as the threshold for declaring linkage. Map units (in centiMorgans, cM) were derived using the Kosambi function (Kosambi 1944). Maximum likelihood orders of markers were verified by the "ripple" function, and markers were said to belong to the framework map if the LOD value, as calculated by the "ripple" command, was ≥ 2.0. Markers that could not be placed with LOD ≥ 2.0 were added to the map in the most likely interval between framework markers.

Once linkage groups were drawn, they were checked for markers linked in repulsion to distinguish between random chromosome assortment, as in autopolyploids, and preferential pairing, as in diploids or allopolyploids. Only pairs of adjacent loci with one shared allele and one parent-specific allele, for which the 'presence' class had not been assigned at random, were considered adequate for this comparison. Pairs of loci for which 'presence' was linked with 'absence' of segregating alleles were counted as being linked in repulsion.

Recombination rates in the gametes of the male and female parents were compared by a t-test ($P < 0.01$) of 10 map intervals bounded by markers for which both parents were heterozygous and had one allele in common, termed allelic bridges (Ritter et al. 1990).

Estimation of genome size

A simple and useful method of estimating genome length, G, from linkage data of organisms that undergo normal meiosis has been described (Hulbert et al. 1988). The method estimates G based on the probability that a randomly chosen pair of loci will lie within xcM of each other is approximately $2x/G$; where x is assumed to be small compared to the mean genetic length of the chromosome. G is mathematically determined from linkage data by solving the equation:

$$G = MX/K$$

Where M = number of informative meioses, X = an interval in cM at some minimum LOD score, K = actual number of pairs of markers observed that border the interval x or less.

Results

Library characterization and parental survey

About 2,700 clones, or 90% of the total from the two *Pst*I libraries of 1,500 clones each, were judged to be low-copy sequences based on dot blot hybridization of whole recombinant plasmid with total cassava genomic DNA as probe. One hundred low-copy clones (50%) were obtained from 200 genomic clones derived from four libraries generated with non-methylation-sensitive enzymes (*Hin*dIII, *Xba*I, *Bam*HI, and *Eco*RI). About 900 low-copy sequences from two *Pst*I libraries, 100 genomic clones from the four smaller libraries, and 75 cDNA clones have so far been screened for detecting polymorphism between the parents of the mapping population with five restriction enzymes, *Eco*RI, *Eco*RV, *Hae*III, *Hin*dIII, and *Pst*I. Of the low-copy genomic clones 41% and 37% detected unique segregating alleles in the gametes of the male and female parents, respectively, with at least one restriction enzyme. Of the cDNA clones, 20% and 26% revealed similar markers in gametes of the male and female parents, respectively, with at least one restriction enzyme. The percentage of RAPD markers with a unique allele in the male and female parents has been described elsewhere (Gomez et al. 1995). Twelve microsatellite loci, ranging in length from 12 to 21 di- or tri-nucleotide repeats were screened for their ability to detect polymorphisms between the parents; 10 microsatellites or 83%, were heterozygous in the female parent and 7, or 58%, in the male parent. Nine isoenzyme loci detected four unique alleles segregating in the female (roughly 50%) but only one (13%) in those of the male parent. Percentage polymorphism of RAPD markers, with unique alleles, in the male and female parents, and percentage polymorphism of similar RFLP markers, for the same cross, showed no significant difference. Table 1 gives a breakdown of polymorphism found in both parents for the 1075 clones screened. Greater levels of polymorphism were found with *Eco*RI and *Hin*dIII; *Hae*III was the least successful (Table 1).

Table 1 Percentage polymorphism (unique allele) found with respect to male and female parent with RFLP, RAPD, microsatellite, and isoenzyme loci

RFLP	% polymorphism detected		Average number of fragments detected by probe	
	Female	Male	Female	Male
*Eco*RI	10.0	12.0	1.5	1.5
*Eco*RV	7.2	7.7	1.5	1.5
*Hae*III	5.2	4.0	1.3	1.3
*Hin*dIII	8.9	1.1	1.4	1.4
*Pst*I	5.7	6.1	1.2	1.2
Total RFLP	37	40.8	1.4	1.4
RAPD	40.0	38.0	–	–
Microsatellite	83.0	58.0	–	–
Isoenzyme	50	12.5	–	

Segregation analysis

Three hundred and seventy-two genomic and cDNA clones were scored in the F_1 mapping population, yielding segregation data for 158 single-dose RFLPs for the

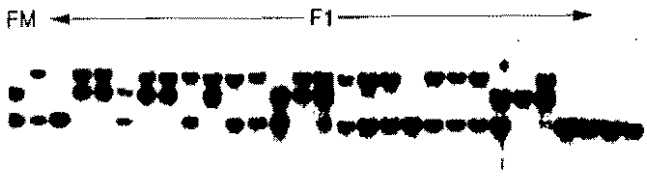


Fig. 1 Southern hybridization showing segregation of heterozygous unique alleles from both parents (allelic bridges). *F* Female parent, *M* male parent

female and 145 for the male. Five additional polymorphic RFLP markers segregated in the mapping population with a ratio of 3:1 (3 from the female and 2 from the male; these were not included in the linkage analysis). Eighteen markers, polymorphic in the parents, with more than one restriction fragment, did not segregate in the F_1 mapping population (10 from the female and 8 from the male). Another 95 of the polymorphic markers were either pseudo F_2 markers, monomorphic between the parents but heterozygous and segregating in the F_1 population (12 markers, excluded from linkage analysis), or were difficult to score after Southern hybridization and are being reanalysed. About a quarter of the polymorphic RAPD markers segregating as single-dose markers were chosen for linkage analysis based on several factors, including consistency of banding pattern after two or three reamplifications, clarity of gels, and number of amplified fragments, with fewer fragments being more acceptable. All microsatellite and isozyme markers polymorphic in the male or female parent segregated as single-dose markers and were scored in the F_1 mapping population.

Thirty genomic clones detected a unique segregating fragment in each parent and a common allele in both parents and were mapped to similar positions on the male/female derived linkage group. Such allelic bridges (Ritter et al. 1991) are crucial for identifying the analogous linkage groups in the male- and female-derived maps, as they detect the same locus on both parental chromosomes, except when they represent duplicated sequences. Figure 1 shows an example of segregation of a marker heterozygous in both parents with a shared allele, or allelic bridge, used to reconcile linkage groups drawn on the independent segregation of markers in male and female gametes.

Sequence duplication in the cassava genome

We assessed the number of fragments detected by 1075 single- and low-copy DNA sequences with the two most polymorphic restriction enzymes, *Eco*RI and *Hind*III (Fig. 2). The majority of sequences detected

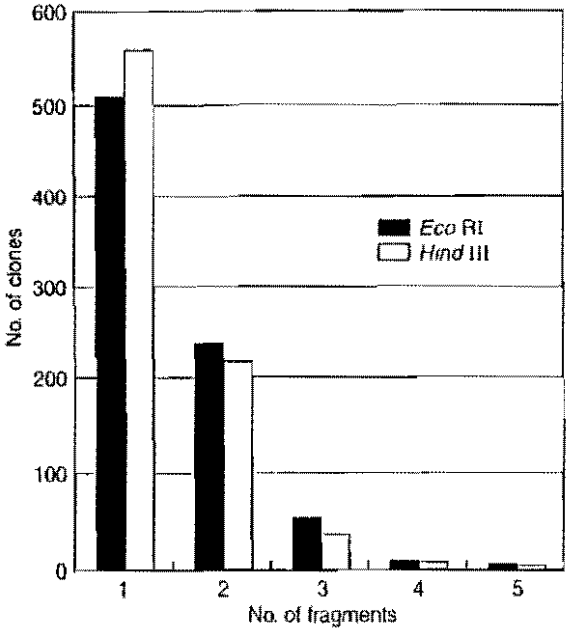


Fig. 2 Number of DNA low-copy sequences detecting 1, 2, 3, 4, and 5 fragments with *Eco*RI nad *Hind*III restriction enzymes

only one or two fragments, which was to be expected for unique loci in the homozygous or heterozygous state. About 100 single- and low-copy sequences detected more than two fragments. This was expected for unique loci in an allo- or autopolyploid with at least one of the two homologous groups having two alleles at duplicated loci or was due to the presence of an internal site for the restriction enzyme used in at least one of the alleles at a marker locus. Duplication was confirmed, by linkage analysis, for six of the sequences presenting more than 2 alleles. These markers detected nonallelic segregating fragments in the male or female gametes and were mapped to different linkage groups. Three of the duplicated loci had 1 locus, each segregating in the gametes of the male and the female parent, and were mapped to linkage groups identified as nonanalogous in the male- and female-derived framework maps. The duplicated loci, GY25a and GY42a on linkage group D and GY 101a on linkage group F, are shown on the female-derived map described below. Three other duplicated loci segregated in the gametes of the male parent (not shown). The rest of the RFLP loci having three or more fragments consisted of: loci having 1 mapped locus and additional monomorphic fragments; loci with two nonallelic segregating fragments (unmapped); or loci with no heterozygous fragments. Efforts continue to map secondary (or primary) loci, as shown in Fig. 3, using additional restriction enzymes. The addition of microsatellites to the map may also assist here, as preliminary evaluation indicates that they detect higher levels of allelic diversity/heterozygosity than genomic or cDNA clones.

Fig. 3 Southern hybridization of clone GY92 with *Hind*III digests of parental line and mapping population. *F* Female parent, *M* male parent. Two faint fragments can be observed to be present and polymorphic in the male but absent in the female parent

The genetic linkage map

One hundred and fifty-eight RFLP, 30 RAPD, 3 microsatellite, and 4 isoenzyme single-dose markers, segregating in the gametes of the female parent of our F_1 mapping population, were tested for linkage using the MAPMAKER computer package. One hundred and thirty-two RFLP, 30 RAPD, 3 microsatellite, and 3 isoenzyme loci were found to define 20 linkage groups spanning 931.6 cM, with an average marker density of 1 marker every 8 cM (Fig. 4); 26 RFLP markers and 1 isoenzyme marker remained unlinked. Linkage groups are named alphabetically until they can be correlated to earlier named chromosome karyotypes (Magoon et al. 1969). The most densely populated linkage group (D) spanned 51.2 cM, with 26 markers, while the least populated group (I), also the longest group, had 8 markers spanning 80.6 cM. This wide range of marker density indicated differing degrees of saturation of linkage groups with markers.

In Fig. 4, 139 (84%) attached to linkage groups by horizontal bars make up the LOD 2.0 framework map of cassava, the remaining 29 markers (in parenthesis) are placed in the most likely intervals between framework markers (LOD $\geqslant$ 2.0). The existence of 27 unlinked markers suggested that some regions of the cassava genome have not yet been mapped. Based on the segregation data reported here, we estimated the length of the cassava genome to be 1, 610 cM. With a total map length of 932 cM, this implied that the present framework map covers roughly 60% of the cassava genome. Regions with distorted segregation ratios could not be identified in the cassava genetic map, since only those polymorphic restriction fragments segregating in a 1 : 1 ratio (SDRF) were included in the linkage analysis. The non-inclusion of markers with distorted monogenic ratios may have excluded certain genomic regions from any map based on these criteria.

Comparison of the male and female genetic maps

One hundred and seven RFLP, 50 RAPD, 1 microsatellite, and 1 isoenzyme single-dose markers, segregating from the gametes of the male parent scored in the F_1 mapping populations, defined 24 linkage groups (not shown) with a total distance of 1, 220 cM. A similar estimate of genome size, conducted as in Hulbert et al. (1988), based on mapping data from the male gametes suggested that the length of the cassava genome is 2, 010 cM. The most salient difference observed between the male- and female-derived framework maps is the greater genetic distances on the male-derived map between markers common to both parents (allelic bridges); see Fig. 5. Intervals were observed to be larger in the male-derived map than in the female-derived map in eight instances, and in only two instances, did the female-derived map display larger genetic distances (linkage groups J and K). A paired *t*-test of the ten intervals showed significantly ($P < 0.01$) greater distances in the male-derived map, suggesting a reduced recombination rate in gametes of the female parent. This observation was reinforced by the greater overall length of the male-derived famework map and its larger estimate of genome size compared to the female. The mean interval length between adjacent allelic bridges in the female-derived map was 38% less than in the male-derived map. Though distances between adjacent allelic bridges may not be representative of all parts of the cassava genome, the similarity between means of interval differences and the overall difference in length of the two maps provided a reasonable basis for comparing recombination frequency between the male- and female-derived parents. We are currently working with the computer package JOIN MAP 2.0 (Stam et al. 1993) to develop a consensus map based on male- and female-derived framework maps.

Chromosome pairing in cassava

To test the hypothesis of preferential pairing against the alternative of random chromosome assortment, we examined the framework map for markers linked in the repulsion phase compared to those linked in the coupling phase. The expected ratio of single-dose markers linked in repulsion to markers linked in coupling is 1 : 1 for disomic inheritance, while no markers are expected to be linked in repulsion for autosomic inheritance (Wu et al. 1992). A total of 40 RFLP markers (30%) were found to be linked in repulsion. Only cases with one allele in common between the parents and the alternate allele linked in repulsion were considered in this calculation (Fig. 6). Linkage groups B and H presented a 1 : 1 ratio of 6 markers linked in repulsion to 6 markers linked in coupling, as expected for disomy. For all other linkage groups there appeared to be some degree of random pairing, based on the number of markers linked in the repulsion phase. Linkage groups A, C, D, I, J, M, N, and T have 2 markers linked in the repulsion phase as against 14, 8, 26, 8, 8, 5, 6, and 3 markers in the groups, respectively. Groups E, K, and P have

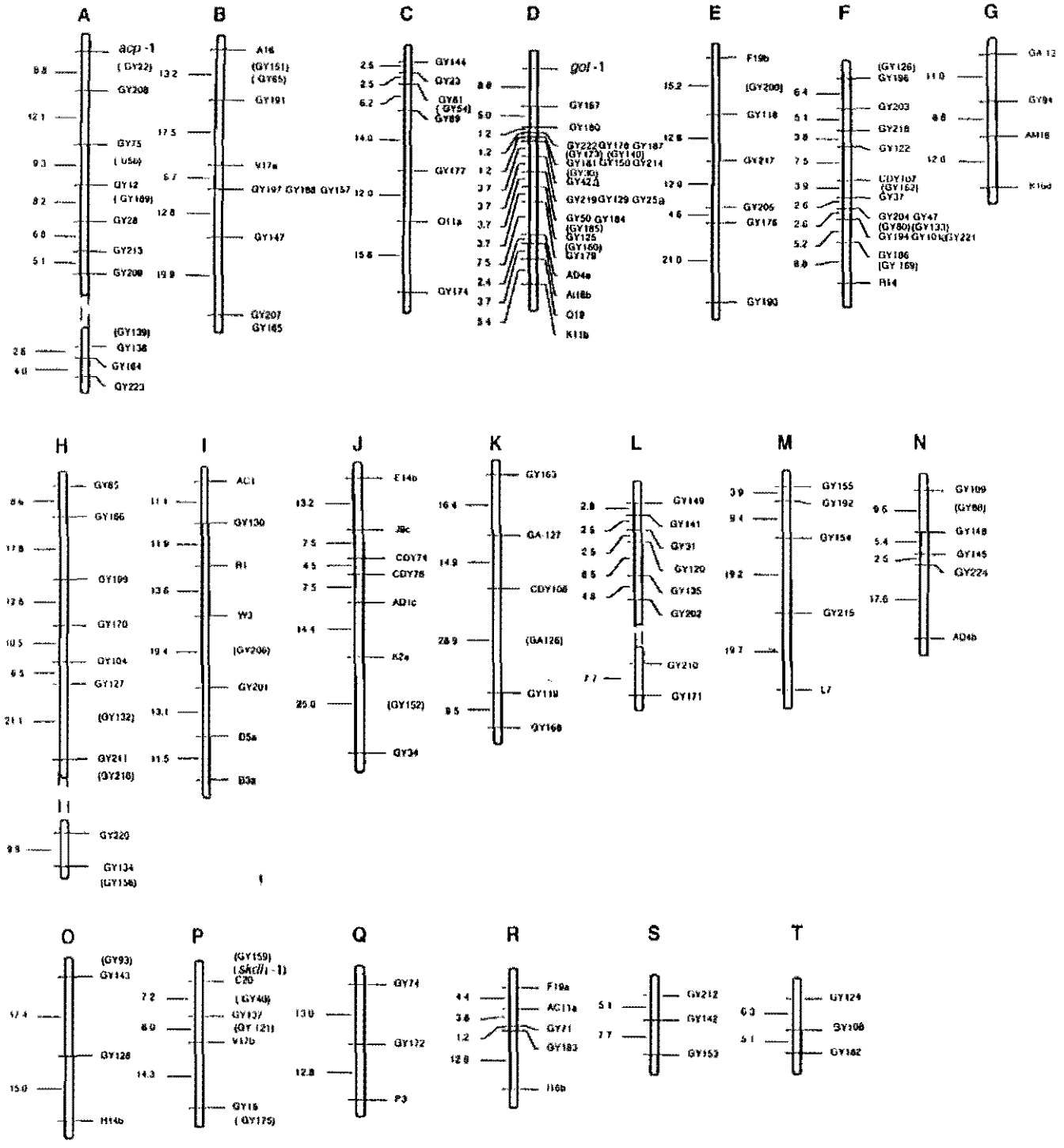
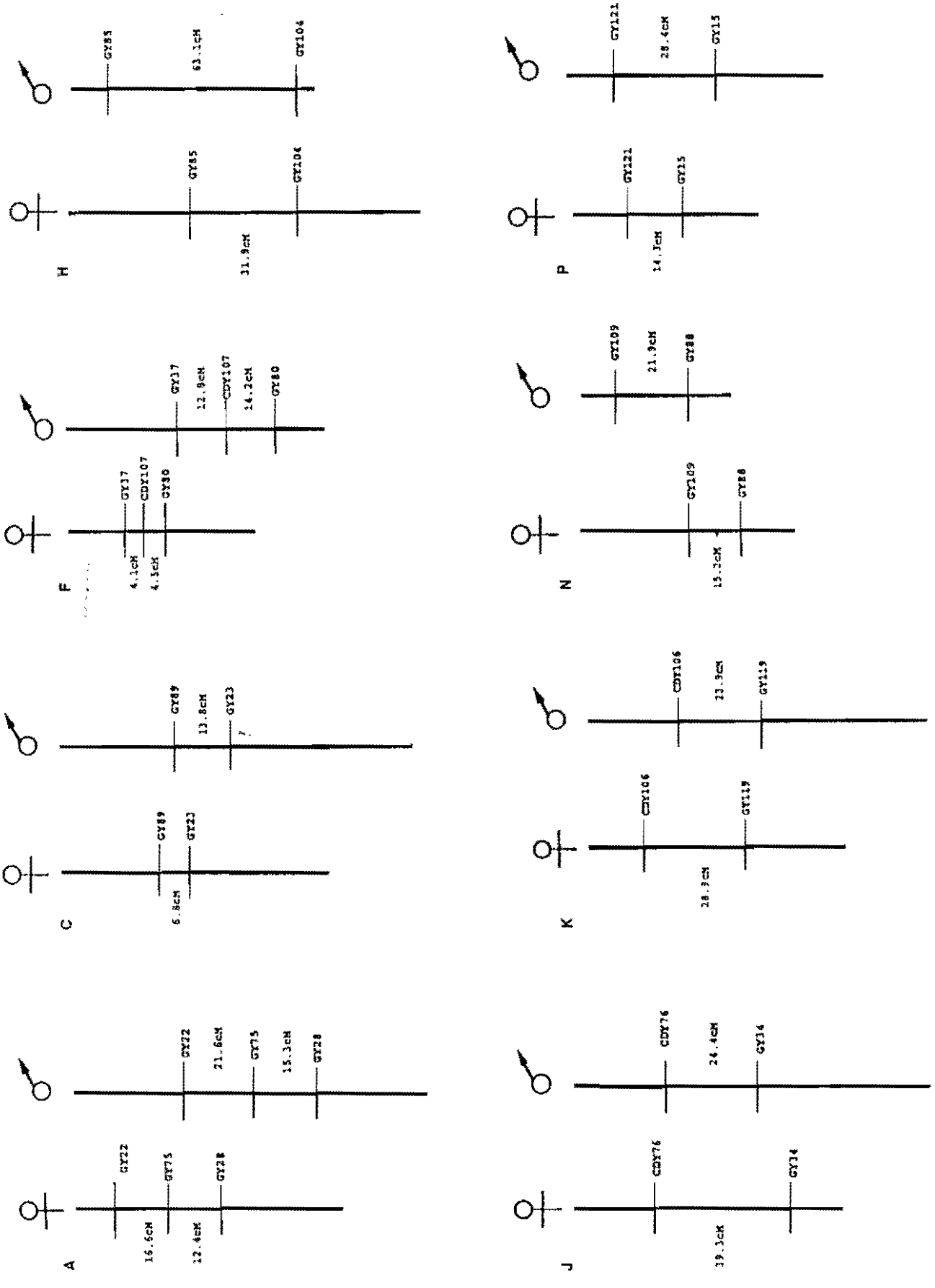


Fig. 4 Molecular genetic map of cassava based on segregation of RFLP (*CDY* cDNA, *GY* genomic), microsatellite (*GA*), isozyme (*acp*, *skdh*, and *got*), and RAPD markers (prefix *A-AP* through Z, Operon primer nomenclature) in gametes of the female parent (TMS30575), in a F_1 cross (with CM2177-2 as male parent) of 90 individuals. Markers with the suffix "*a*" represent duplicated loci. Markers *adjacent* to *horizontal lines* belong to the framework (LOD > 2.0) map; those *following* on the *same line* cosegregate, and remaining markers (*in parenthesis*) are placed in the most probable interval. Map distances, shown on the *left*, are indicated in Kosambi map units

4 markers linked in repulsion as against 7, 6 and 9 markers in the linkage groups, respectively. Group L has 6 of its 8 markers linked in the repulsion phase. No markers were found linked in repulsion phase in linkage groups F, G, O, Q, R, and S, which agrees with the expected behaviour of random assortment under autopolyploidy. The sizes of linkage groups G, O, Q, and R, however, suggested that a sizeable part of the chromosomes remains to be mapped.


A
GY22
GY75
GY28
16.6cM
12.4cM
GY22
21.6cM
GY75
15.3cM
GY28
C
GY89
GY23
6.8cM
GY89
13.8cM
GY23
F
GY37
CDY107
GY30
4.1cM
4.3cM
GY37
12.9cM
CDY107
14.2cM
GY30
H
GY85
GY104
31.9cM
GY85
63.1cM
GY104
J
CDY76
GY34
39.3cM
CDY76
24.4cM
GY34
K
CDY106
GY119
28.3cM
CDY106
23.3cM
GY119
N
GY109
GY88
15.2cM
GY109
21.9cM
GY88
P
GY121
GY15
14.1cM
GY121
28.4cM
GY15

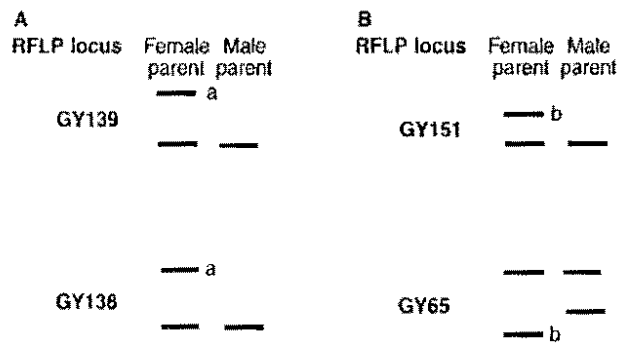


Fig. 6A, B Examples of the category of markers considered in identifying linkage in repulsion, with one parent heterozygous (**A**) or both parents heterozygous (**B**). In both examples, unique alleles (*a* and *b*) from the female parent segregated in repulsion in the F_1 progeny

Discussion

The cassava genetic linkage map reported here spans 931.6 cM, has 20 linkage groups, and is estimated to cover 60% of the cassava genome. If a polyploid origin for the cassava genome is assumed, the expected number of linkage groups in cassava would depend on chromosome assortment. Eighteen groups are expected where preferential pairing is exclusive, up to 36 where there is mostly random pairing, and a number between 18 and 36 where there is a mixture of preferential and random pairing. Thirty percent of all RFLP markers on the cassava genetic map were found to be linked in the repulsion phase. This figure is less than the 50% expected for linkage of single-dose markers in allopolyploids or diploids (Wu et al. 1992). On the other hand, it differs markedly from the complete absence of linkages in the repulsion phase expected for autopolyploids. The finding that fewer than 50% of the markers were linked in the repulsion phase and of entire groups not presenting markers linked in repulsion implies that a significant amount of random pairing occurs in cassava. The karyology of the 18 haploid chromosomes of cassava reveals six identical pairs and three different pairs of homologous chromosomes (Magoon et al. 1969). Assuming random assortment between homoeologous chromosomes of the six pairs of identical chromosomes, a reduction of 66.7% would be expected of all markers linked in repulsion (16.7% as against 50%). The percentage of markers linked in the repulsion phase in the genetic map of cassava reported here, 30%, is significantly higher than this. It is not clear at this stage if the higher number of markers found in repulsion than expected from the karyology of cassava (Magoon et al. 1969) supports the predominance of disomic inheritance as suggested by pairing behaviour (Bai 1992) and inheritance of isoenzyme loci (Hussain et al. 1987; Roca et al. 1992; Lefevre and Charrier 1993).

More confounding is the fact that only 6 duplicated loci were detected with the segregation of single-dose markers derived from 36 cDNA, and over 200 genomic sequences. In well-known allopolyploids such as maize, wheat, and cotton, blocks of duplicated loci have been clearly identified by RFLP mapping (Helentjaris et al. 1988; Devos et al. 1993; Reinisch et al. 1994). Our results, in contrast, have revealed only a few randomly distributed duplicated loci, less than 5% of the total number of markers, a number corresponding roughly to that reported in many diploids (Causse et al. 1994; Tanksley et al. 1995). There is no doubt that the cassava genome contains duplicated regions, but it remains to be determined whether the duplicated loci represent vestiges of an ancient allopolyploid or random genomic duplication events in a diploid or diploidized genome. We are currently placing more cDNA markers on the cassava map in an attempt to resolve the enigma of genome duplication in cassava, a putative allopolyploid. cDNA sequences are known to be relatively more conserved than genomic clones (Helentjaris et al. 1988), and are expected to be very valuable additions to the cassava map.

The clusterings of markers on some groups, compared to others, separated by intervals larger than 15 cM, suggests that recombination is not occurring uniformly across the cassava genome. Nonrandomness of recombination frequency has been reported for several other crops (Bonierbale et al. 1988; Causse et al. 1994; Lagercrantz and Lydiate 1995).

Allelic bridges (markers that are heterozygous in the gametes of both the male and female parents and share a common allele) have been useful in reconciling the separate maps drawn on segregation from gametes of the male and female parents. Specifically, these markers help compare rates of recombination in the gametes of the male and female parents, a task of biological interest and possible practical use. Genetic distances between pairs of analogous markers were found to be larger in the male-derived map than in the female in eight out of ten intervals bordered by allelic bridges, suggesting a reduced recombination in gametes segregating from the female parent. Earlier reports in *Drosophila* (Baker et al. 1976), humans (Bowden et al. 1989), and tomatoes (de Vincente and Tanksley 1991) found a significant genome-wide reduction in recombination in gametes segregating from the male parent. Enhanced male recombination frequencies were, however, found to be associated with terminal chromosomal regions in *Brassica*, while enhanced female

Fig. 5 Genetic maps of single-dose markers segregating in the female and male gametes. The intervals shown are those bounded by markers heterozygous in both parents, and distances are as calculated by MAPMAKER's "pairwise" comand. Alphabetical names of linkage groups are according to the female map

recombinations were adjacent to putative centromere positions (Lagercrantz and Lydiate 1994). We cannot determine at this point whether the reduced rates of recombination observed in the gametes of the female parent of this cross are due to its genetically diverse ancestry ('TMS 30572' is understood to contain introgressed genes from wild *Manihot* species), or to potential genetic sex differences in cassava. It is known, however, that increased recombinations, such as those found in the male parent, can serve as a tool to minimize linkage drag during backcrossing schemes (Young and Tanksley 1989).

To our knowledge, the cassava genetic map is the first of its kind to be constructed for the family Euphorbiaceae. It could prove useful for comparative mapping in other economically important genera of the family, such as *Hevea* (rubber), or the markers could be applied to minor related species, for which molecular cloning may not be justified. Furthermore, the map provides an important tool for cassava improvement, based on the association of molecular markers with useful genes. Cassava is a staple throughout the tropics, and important agronomic and culinary trait as well as adaptation to pests and diseases vary in cassava cultivars from one ecological zone to another. Although intra- and even interspecies crosses are easy to make in cassava, the introgression of traits from one cassava genepool to another is complicated by difficulties in understanding the inheritance of complex traits in highly heterozygous genetic backgrounds. The map can serve as a tool for efficiently transferring traits between gene pools. Future perspectives for the genetic map of cassava includes the addition of more molecular markers, preferably microsatellite markers and cDNA clones, and joining the male/female-derived maps. Microsatellite markers showed a higher level of polymorphism than any other marker in the intraspecific cross used in generating the cassava genetic map, while cDNAs will be most useful in defining the genome structure.

Acknowledgements We would like to thank Dr. C. Iglesias for generating the mapping population, Maria Mercedes Maya and Matt "Mateo" Sayre for excellent technical support, Cesear Ocampo for the isoenzyme analysis, and Nathan Russell for editing. Thanks also to Dr. S. Bebe for helpful comments. This work was supported by grant RF 91060 No. 2 of the Rockefeller Foundation's biotechnology programme. MF was supported by a post-doc fellowship, also from the Rockefeller foundation and RG by COLCIENCIAS project No. 228-07-003-92.

References

Al Janabi SM, Honeycutt RJ, McClelland M, Sobral BWS (1993) A genetic map of *Saccharum spontaneum* L. "SES 208". Genetics 134:1249–1260

Angel F, Arias D, Tohme J, Iglesias C, Roca W (1993) Towards the construction of a molecular map of cassava (*Manihot esculenta* Crantz): comparison of restriction enzymes and probe source in detecting RFLPs. J Biotechnol 31:103–113

Awoleye F, Van Duren M, Dolezel J, Novak FJ (1994) Nuclear DNA content and in vitro induced somatic polyploidization (*Manihot esculenta* Crantz) cassava breeding. Euphytica 76: 195–202

Bai KV (1992) Cytogenetics of *Manihot* species and interspecific hybrids. In: Roca W, Thro AM (eds) Proc 1st Int Sci Meet Cassava Biotechnol Network. CIAT, Cali, Colombia, pp 51–55

Baker BS, Carpenter ATC, Esposito RE, Sandler L (1976) The genetic control of meiosis. Annu Rev Genet 10:53–134

Bennet MD, Smith JB, Heslop-Harrison JS (1992) Nuclear DNA amounts in Angiosperms. Proc R Soc London Ser B 216: 179–199

Bonierbale MW, Plaisted RL, Tanksley SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095–1103

Bothwell AL, Yancopoulis GD, Alt FW (1990) Methods for cloning and analysis of eukaryotic genes. Jones and Bartlett Publ, Boston, Mass., pp 8–19

Bowden DW, Gravius TC, Green P, Falls K, Wuster-Hill D, Noll W, Muller-Kahle H, Donis-Keller H (1989) A genetic linkage map of 32 loci on human chromosome 10. Genomics 5:718–726

Bryne D (1984) Breeding cassava. Plant Breed Rev 2:73–134

Causse MA, Fulton TM, Cho YG, Ahn SN, Chunwongse J, Wu K, Xiao J, Yu Z, Ronald PC, Harrington SE, Second G, McCouch SR, Tanksley SD (1994) Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138:1251–1274

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA 81:1991–1995

Da Silva JAG, Sorrels ME, Burnquist WL, Tanksley SD (1993) RFLP linkage map and genome analysis of *Saccharum spontaneum*. Genome 36:782–791

Dellaporta SL, Wood J, Hicks JR (1983) A plant DNA minipreparation: version II. Plant Mol Biol Rep 1:19–21

de Vincente MC, Tanksley SD (1991) Genome-wide reduction in recombination of backcross progeny derived from male versus female gametes in an interspecific cross of tomato. Theor Appl Genet 83:173–178

Devos KM, Millan T, Gale MD (1993) Comparative RFLP maps of the homeologous chromosomes of wheat, rye, and barley. Theor Appl Genet 85:784–792

Feinberg AP, Vogelstein B (1981) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137:266–267

Fregene MA, Vargas J, Ikea J, Angel F, Tohme J, Asiedu RA, Akoroda MO, Roca WM (1994) Variability of chloroplast DNA and nuclear ribosomal DNA in cassava (*Manihot esculenta* Crantz) and its wild relatives. Theor Appl Genet 89:719–727

Gomez R, Angel F, Bonierbale M, Rodriguez F, Tohme J, Roca W (1994) Selection of heterozygous parents and single dose markers for genetic mapping in cassava. In: Roca W, Thro AM (eds) Proc 2nd CBN Meeting. CIAT, Cali, Colombia, pp 113–124

Hahn S, Bai KV, Asiedu RA (1990) Tetraploids, triploids, and 2n pollen from diploid interspecies crosses with Cassava. Theor Appl Genet 79:433–439

Helentjaris T, Weber D, Wright S (1988) Identification of the genome locations of duplicate nucleotide sequences in maize by anlysis of restriction fragments length polymorphisms. Genetics 118:353–363

Hulbert SH, Ilott TW, Legg EJ, Lincoln SE, Lander ES, Michelmore RW (1988) Genetic analysis of the fungus, *Bremia lactucae*, using restriction fragments length polymorphisms. Genetics 120: 947–958

Hussain A, Ramirez H, Roca W, Bushuk W (1987) Identification of *cassava* (*Manihot esculenta* Crantz) by electrophoretic patterns of esterase isozymes. Seed Sci Technol 15:19–22

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175

Lagererantz UFL, Lydiate DJ (1995) RFLP mapping in *Brassica nigra* indicates differing recombination rates in male and female meioses. Genome 38:255–264

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Lefevre F, Charrier A (1993) Heredity of seventeen isozyme loci in cassava (*Manihot esculenta* Crantz). Euphytica 66:171–178

Magoon ML, Krishnan R, Bai KV (1969) Morphology of the pachytene chromosomes and meiosis in *Manihot esculenta* crantz. Cytol Tokyo 34:612–626

Ocampo C, Hershey C, Iglesias C, Iwanaga M (1992) Esterase isozyme fingerprinting of the cassava germplasm collection held at CIAT. In: Roca W, Thro AM (eds) Proc 1st Int Sci Meet Cassava Biotechnol Network. CIAT, Cali, Colombia, pp 81–89

Perry BA (1943) Chromosome number and phylogenetic relationship in the Euphobiacea. Am J Bot Club 30:527–543

Reinisch AJ, Dong J, Brubaker CL, Stelly DM, Wendel JF, Paterson AH (1994) A detailed RFLP map of cotton *Gossypium hirsutum* × *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829–847

Ritter E, Debener T, Barone A, Salamini F, Gebhardt C (1991) RFLP mapping on potato chromosomes of two genes controlling extreme resistance to potato virus X (PVX). Mol Gen Genet 227:81–85

Roca WM, Henry G, Angel F, Sarria R (1992) Biotechnology research applied to cassava improvement at the International Center of Tropical Agriculture(CIAT). AgBiotechnol News Inf 4:303N–308N

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory Press, New York

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JOIN MAP. Plant J 3:739–744

Tanksley SD, Ganal MW, Prince PJ, deVincente MC, Bonierbale MW, Broun P, Fulton MW, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1995) High density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141–1160

Umanah EE, Hartmann RW (1973) Chromosome numbers and karyotypes of some *Manihot* species. Am Soc Hortic Sci 98:272–274

Vayda ME, Freeling M (1986) Insertion of the *Mu1* transposable element into the first intron of maize *Adh1* interferes with transcription elongation but does not distrupt chromatin structure. Plant Mol Biol 6:441–454

Wu KK, Burnquist W, Sorrels ME, Tew TL, Moore PH, Tanksley SD (1992) The detection and estimation of linkage in polyploids using single dose restriction fragments. Theor Appl Genet 81:471–476

Young ND, Tanksley SD (1989) RFLP analysis of the size of chromosomal segments retained around the *Tm-2* locus of tomato during backcross breeding. Theor Appl Genet 77:353–354

Yu K, Pauls KP (1992) Optimization of the PCR program for RAPD analysis. Nucelic Acids Res 20:2606

Theoretical and Applied Genetics (in press).

Microsatellites in Cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability

[1]P. Chavarriaga-Aguirre, [2]M.M. Maya, [1,5]M.W. Bonierbale, [3]S. Kresovich, [4]M.A. Fregene, [4]J. Tohme and [1]G. Kochert.

(0) Author for correspondance

(1) P. Chavarriaga - G. Kochert: 2502 Plant sciences, Department of Botany, University of Georgia, Athens GA 30605, USA

(2) M.M. Maya - M.W. Bonierbale: Cassava Program, CIAT, A.A. 6713, Cali, Colombia

(3) S. Kresovich: USDA-ARS, Plant Genetic Resources Conservation Unit, University of Georgia, Griffin, GA, 30223-1797, USA

(4) M.A. Fregene - J. Tohme: Biotechnology Research Unit, CIAT, A.A. 6713, Cali, Colombia

(5) Present address: Merideth Bonierbale, International Potato Center (CIP), Apartado 1558, Lima 12, Peru

Abstract

Fourteen microsatellites containing GA-repeats were isolated and characterized in cassava (*Manihot esculenta* Crantz, Euphorbiaceae). Microsatellite heterozygosity (*h*) was estimated in 48 accessions using (^{32}P)-end-labeled primers and in more than 500 accessions using fluorescence-based genotyping. Heterozygosity values ranged from 0.00 to 0.88 and the number of alleles detected varied from 1 to 15. The reproducibility of allele sizing was also assessed using fluorescence-based genotyping. The average inter-gel size difference was 1.03 nucleotides. Chi square tests (X^2) were performed to analyse segregation distortion and linkage between alleles segregating from either or both parents in an F_1 mapping population. Most microsatellite loci segregated in the expected 1:1, 1:2:1 or 1:1:1:1 ratio. Linkage was detected between loci segregating from either parent, and segregation distortion from the male parent was detected for locus GA-131. Approximately 80% of microsatellites detected one or two alleles per accession, suggesting a low degree of microsatellite loci duplication, an unexpected finding for a putative allopolyploid, highly heterozygous species. The high *h* values of most microsatellites, their amplification in other *Manihot* taxa and their suitability for high-throughput, fluorescence-based genotyping, make microsatellites the marker of choice for germplasm characterization and saturation of the cassava map.

Keywords: cassava, microsatellites, fluorescence-based genotyping, heterozygosity, linkage.

Introduction.

Cassava (*Manihot esculenta* Crantz, Euphorbiaceae) is a root crop that originated in the new world tropics. Although the exact place of cassava domestication is still a matter of debate (Roa et al. 1997; Schaal et al. 1995; Fregene et al. 1994; Bertram 1993), it was taken from South America and spread throughout the Old World tropics by Portuguese sailors in the 16th century (Rogers 1963). Since then, cassava has gradually become an economically important crop in developing countries. In Africa, the starchy roots of cassava constitute a staple food. In Thailand, Indonesia and Brazil, cassava is an important export commodity (CIAT 1993; FAO 1992). Today cassava ranks fourth among the major sources of carbohydrates in the tropics, where it is an important food for over 500 million people.

DNA-based molecular markers such as RAPDs, nuclear RFLPs and microsatellites were used to develop the cassava molecular genetic map (Fregene et al. 1997). Similarly, RAPDs, AFLPs and rDNA, cpDNA and cDNA RFLPs have been used to assess the genetic variability of small sets of cassava germplasm and to establish relationships between cassava and its wild relatives (Bertram 1993; Beeching et al. 1993; Marmey et al. 1994; Fregene et al. 1994; Schaal et al. 1995; Roa et al. 1997). Assessing the genetic variability of larger germplasm collections, such as the cassava core collection that contains over 600 accessions (Hershey et al. 1994), would require highly polymorphic markers as well as high-throughput genotyping systems (Ziegle et al. 1992). Such systems have been used for plant germplasm characterization (Mitchell et al. 1997) and for large-scale genotyping of humans (Ghosh et al. 1997; for review see Hall et al. 1996). In the case of cassava, germplasm characterization will contribute to the establishment of relationships between accessions of the cultivated genepool and, consequently, facilitate the development of cassava cultivars to satisfy market standards and to respond to diverse biotic and abiotic challenges. One type of molecular marker that may be suitable for cassava germplasm characterization is microsatellites.

Microsatellites are short stretches of tandemly repeated, 1-5 nucleotides sequences, such as $(GA)_n$. They are ubiquitously present in eukaryotic genomes and are highly polymorphic (Tautz 1989; Beckman and Soller 1990, Weber 1990). Conservation of the microsatellite flanking sequences allows the designing of primers for PCR amplification. Microsatellites, like RFLPs, are considered codominant markers. However, insertions or base substitutions in priming sites may convert them to dominant markers. Their high polymorphism makes microsatellites suitable markers for population genetics, development of linkage maps and phylogenetic studies (Broun and Tanksley 1996; Zhao and Kochert 1993; Saghai Maroof et al. 1994). They have been useful for germplasm characterization of grapes, soybean, Brassica and wheat (Thomas et al. 1994; Rongwen et al. 1995; Kresovich et al.1995; Roder et al. 1995). The objective of this work was to clone cassava microsatellites suitable for mapping and for characterization of the genetic variability of the cassava core collection. Fourteen cassava microsatellite loci containing GA-repeats are described, along with heterozygosity estimates, mode of inheritance and linkage analysis of 11 loci and relevant aspects of microsatellite allele sizing using fluorescence-based, semi-automated genotyping systems.

Materials and methods.

Plant material and DNA extraction.

A segregating cassava F_1 population of 83 individuals (cross CM7857), produced by crossing TMS30572 and CM2177-2 cultivars, was received *in vitro*, or as DNA samples, from CIAT (Cali, Colombia). *In vitro* material was transferred to the greenhouse after a subculture period of six weeks. Leaf tissue was collected for DNA extraction from the population mentioned above and from 522 accessions of the cassava core collection grown in the field at CIAT. For passport data of all accessions consult SINGER database at the web site http://nocl.cgiar.org/seartype.htm.

Genomic DNA was isolated from fully-expanded young leaves of greenhouse-grown plants by a modification of the Kochert et al. (1991) procedure. Basically, 6-8 g of fresh tissue were powdered in liquid nitrogen and the DNA extracted by adding 5 ml of lysis buffer, avoiding the isolation of a a crude nuclear preparation. All DNA solutions were treated with 1 µl of 10 µg/ml RNAse for 1 hour at 37°C. DNA extractions for field-grown plants were carried out according to Dellaporta et al. (1983), with the modifications reported by González et al. (1995).

Microsatellite cloning and sequencing.

Approximately 100 µg of DNA from cassava accession MCol22 was double-digested with *Eco*RI and *Xho*I, according to instructions provided by the manufacturer (New England Biolabs). The restricted DNA was run

on 0.5% low melting point agarose gels until good separation of fragments between 0.3 and 0.8 kb was achieved. DNA fragments within this size range were isolated from the gel, purified using a phenol-chloroform procedure (Sambrook et al. 1989), and cloned into the vector λ ZAP II (Stratagene). The total number of recombinant clones obtained was $8x10^5$, as assessed with the IPTG X-Gal screening procedure. Plaque lifts (10^4 phages/plate) were screened with $(GA)_{15}$, $(GT)_{15}$, $(AT)_{15}$, $(TTAG)_8$, $(TCT)_{10}$, $(GTAG)_7$ and $(GAGG)_7$ (^{32}P)-end-labeled synthetic oligonucleotides. Hybridizations were conducted overnight at 42-50°C in 6XSSC, 0.5% SDS, 1 mM EDTA pH 8.0, 0.1% powdered milk and 10 mM monobasic/dibasic potassium phosphate. One to three filter washes were performed at 42-50°C for 15 min each in 6XSSC and 0.1% SDS. Positive clones were sequenced after at least one more round of screening. DNA sequencing was performed on an automated sequencer (Perkin Elmer/Applied Biosystems model 373A) using T3 and/or T7 primers.

Primer Design , PCR conditions and electrophoresis.

Primers were designed for microsatellites containing at least 10 perfect or imperfect repeats, using PRIMER version 0.5 software (Whitehead Institute for Biomedical Research) and/or by visual inspection of the sequences. Nonradioactive PCR reactions (100 µl) contained 0.1-0.6 µg genomic DNA, 0.2 µM of each primer in 10 mM TRIS-HCl, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, 200 mM of each dNTP and 0.5-2.5 Units of Amplitaq DNA polymerase (Perkin Elmer/Cetus). Microsatellites were amplified with a Perkin-Elmer (Cetus) 480 or a PTC-100 Hot Bonnet (MJ-Research) thermal cycler. An initial denaturation cycle was done at 94°C for 2 min, followed by 30 cycles of amplification by denaturing at 94°C for 1 min, annealing at 56 or 45°C for 1 min and extension at 72°C for 2 min. The final step was a single extension cycle at 72°C for 5 min. Nonradioactive PCR products were run on 5% metaphor agarose gels following recommendations from the supplier (FMC Bioproducts).

Radioactive PCR reactions (10 µl) followed the protocol described above with 0.1-0.2 µg of genomic DNA/reaction and (^{32}P)-end-labeled primers. PCR products were run on 6% polyacrylamide sequencing gels for a minimum of three hours at 80W. Since all microsatellites isolated contained dinucleotide GA-repeats (see results), microsatellite allele sizing was visually estimated based on the sequence-determined size of the original clone from cassava accession Mcol22, and assuming a two base-pair difference between the most intense band and stutter bands.

Semi-automated genotyping using fluorescent primers.

Primers labeled with fluorescent dyes were commercially synthesized by Perkin Elmer/Applied Biosystems or Research Genetics (Foster City, CA; Huntsville, AL). To confirm the accuracy of semi-automated allele sizing, four primer sets (GA-21, GA-126, GA-134, GA-136) were fluorescently labeled to perform quadriplex PCR amplifications. Data from ten independent gel runs were collected from PCR amplified products of cassava accession MCol22. The range, mean and standard deviation of inter-gel allele sizes were determined for each locus. Allele sizes were then compared to those obtained by sequencing the original microsatellite clones from the same accession.

Multiplex PCR amplifications were carried out in 25 µl volumes containing 25 ng of target DNA, 5 pmol of each primer, 1X Perkin Elmer PCR buffer II, 1.5 mM $MgCl_2$, 0.25 mM dNTPs and 0.5 U AmpliTaq-Gold polymerase (Perkin Elmer). Temperature cycling was done on a GeneAmp PCR System 9600 (Perkin Elmer) using 1 s ramp times. The amplification profile for all PCR reactions using fluorescently labeled primers consisted of one polymerase activation cycle at 94°C for 10 min; one DNA denaturation cycle at 95°C for 4 min; 25 amplification cycles at 95°C for 1 min, 55°C for 2 min and 72°C for 2 min; and a final elongation cycle at 72°C for 10 min.

After PCR amplification, 1.5 µl of PCR product was mixed with 0.5 µl GeneScan 500 internal lane standard labeled with TAMRA (Perkin Elmer/Applied Biosystems) and 2.0 µl of 50% formamide. The mixture was incubated at 95°C for 5 min, and 1-2 µl were loaded on 6% denaturing (7 M urea) acrylamide:bisacrylamide (19:1) gels (12 cm well-to-read). The samples were separated by electrophoresis in 1X TBE at 29 W for a minimum of 3 h on an automatic DNA sequencer (Perkin Elmer/Applied Biosystems

model 373A with GeneScan 672 software version 1.2.1). Fragment sizes were calculated using the "local Southern" algorithm (Elder and Southern 1987).

Heterozygosity.
To estimate informativeness of the 14 microsatellites in cassava, heterozygosity (h) values were calculated from single gel runs of PCR-amplified products from accessions of the cassava core collection. If (^{32}P)-end-labeled primers were used to PCR-amplify microsatellites, 48 accessions were analyzed for h determination. With fluorescently-labeled primers, 522 accessions were analyzed to estimate h (see Table 3). Heterozygosity was calculated according to the formula: $h = 1-\Sigma(p_i)^2$ (Nei 1978; see also Rongwen et al. 1995 for application), where h represents the probability that two alleles from the same locus would be different if chosen at random in the population (Li and Graur 1991), and (p_i) is the freqoency of the i^{th} allele at a locus. h gives an indication of the polymorphism of a marker in a given population.

Segregation and linkage analysis.
Chi square tests (α=0.01, with n-1 degrees of freedom, where n = number of phenotypic classes) were performed to ascertain the significance of segregation distortion and linkage of eleven microsatellite loci. The tests were performed using 83 F_1 individuals of the cassava mapping population. For segregation analysis, two kinds of segregation ratios were expected: backcross-like- or F_2-like ratio. In the former, microsatellite alleles segregate from only one of the parents in a 1:1 ratio, while in the latter both parents are heterozygous for the microsatellite and segregation results in a 1:2:1 or 1:1:1:1 ratio. Linkage analysis was performed using the gamete segregation ratios for pairs of microsatellite loci for each parent, in the case of backross-like expected ratios, and for both parents in the case of F_2-like expected ratios.

Results and Discussion.

Microsatellite characterization and primer design.
A size-fractionated cassava DNA library was screened with seven different oligonucleotide probes to detect microsatellites. Sixty positive clones were purified and sequenced to confirm the expected repeats. Thirty-two clones had the expected repeats, while the rest contained unexpected or no repeats. The majority of the microsatellites detected contained GA-repeats, which were variable in length and composition, some having up to 19 perfect repeats. Microsatellites containing GA-repeats are common in many plants (Lagercrantz et al. 1993; Wang et al. 1994; Depeiges et al. 1995; Terauchi 1994), which may be the case for cassava as well. The isolation of other types of microsatellites may require the screening of libraries enriched for specific repeats.

Primers were designed for 22 microsatellites, although only those primers successfully amplifying the expected products are reported in Table 1. All primers were used to amplify cassava DNA at an annealing temperature of 56°C, unless otherwise indicated.

Semi-automated genotyping.
The results obtained for inter-gel allele size comparisons are summarized in Table 2. Only inter-gel allele size differences are reported since they are usually higher than intra-gel size differences according to Mitchell et al. (1997) and Smith (1995). Mitchell et al. (1997) report that the maximum inter-gel allele size difference should not exceed 0.5 nucleotides, since alleles can be classified incorrectly if sizes are rounded off to the next integer. This problem is accentuated when DNA fragment sizes vary by one nucleotide, due to real size differences or to terminal transfer activity of *Taq* DNA polymerase. The literature reports inter-gel size differences of 0.7 to 1.3 nt (see Mitchell et al. 1997 and references therein), which are similar to the ranges reported here, 0.42 to 1.61 nt, except for one allele of locus GA-21, where successive determinations varied up to 2.17 nt. This discrepancy may be due to terminal transfer activity of *Taq* DNA polymerase, and gel or electrophoresis detection conditions that make sizing of alleles less precise on 373A sequencers (Ghosh et al. 1997). Deletion and insertion events that occur in the clone once the cloning vector is introduced into *E.coli* for replication may also explain large size differences between the sequenced clone and the PCR-amplified product (Freund et al. 1989. In: Mitchell et al. 1997). Dinucleotide repeats are difficult to score due to overlapping alleles. However, the use of external adjustment and binning algorithms, like those reported by Ghosh et al. (1997), may be used to reduce inaccuracies in allele sizing.

Cassava germplasm characterization is often limited by budget constrains. Thus, unequivocal identification of each accession is essential to avoid duplicate samples, which increases germplasm maintenance costs.

The set of microsatellites described here, combined with appropriate allele sizing algorithms and multiplex, semi-automated genotyping systems, could be used to characterize cassava accessions from the core collection or from other collections. This would expedite the process of selecting representative sets of germplasm for conservation, characterization or breeding.

Heterozygosity.
To determine how useful microsatellites would be for fingerprinting or characterization of cassava germplasm and for mapping purposes, heterozygosity (h) estimates were made using (^{32}P)-end-labeled- or fluorescently-labeled primers. The number of alleles detected for microsatellite loci ranged from 1 to 15 (see example in Figure 1), and h varied from 0.00 to 0.88 (Table 3). An example of low h value was microsatellite GA-123, for which only one fluorescently labeled PCR product of equal size was detected in 3 out of 39 accessions analyzed, the other 36 accessions did not show detectable PCR products (data not shown), probably due to mutations in priming sites. However, more than 75% of the microsatellites analyzed had h values above 0.5, which makes them useful markers for cassava germplasm characterization and mapping. A study of the genetic diversity of cultivated and wild *Manihot* germplasm using microsatellites GA-12, GA-21 and GAGG-5, revealed between 8-18 alleles for each locus (Roa 1997). Roa (1997) detected 18 alleles for locus GA-12 alone, ten more than the eight reported here, probably due to the inclusion of seven *Manihot* taxa. Nevertheless, the number of alleles observed in the Roa's study is comparable with the ones reported here and suggest microsatellite priming site conservation among cassava and its wild relatives.

Assessment of microsatellite heterozygosity was done using small or large sets of cassava germplasm, though the number of alleles was naturally higher in larger sample sets (Table 3). A 27% increase in number of alleles was observed for microsatellite GA-134 when the number of accessions screened increased from 38 to more than 500 individuals. However, Table 3 shows that h values for microsatellites GA-21, GA-126 and GA-134 remained around 0.6, 0.8 and 0.5 respectively, regardless of sample size. This indicated that the small core subsamples fairly represented the variability existing in the larger core collection.

The set of cassava microsatellites reported here thus have heterozygosity values that may be useful for germplasm characterization of cultivated cassava and its wild related species. This will undoubtedly contribute to the understanding of genetic relationships among cassava accessions, wild taxa or both.

Microsatellite segregation and linkage analysis
Chi-square tests were performed on segregation ratios of 11 microsatellite loci that segregated in the mapping population from the female, the male or both parents. Three microsatellites, GA-13, GA-57 and GA-123 were not polymorphic in the parents, yet the former two were polymorphic in other cassava accessions (data not shown). The results of the Chi-square tests are summarized in Table 4. All microsatellite loci, except GA-131, fit expected 1:1, 1:2:1 or 1:1:1:1 disomic segregation ratios for one or two loci, a finding previously reported for isozyme and RFLP loci (Roca et al. 1992; Sarria et al. 1993; Lefevre and Charrier 1993; Fregene et al. 1997) and of significance to cassava breeding programs. Microsatellite GA-131 did not fit an expected 1:1:1:1 ratio, and exhibited an excess of alleles derived from the female parent (Table 4), although it has been mapped to linkage group G of the male- and female-derived cassava maps (Fregene et al. 1997; Table 5).

The cassava genetic linkage maps reported by Fregene et al. (1997) were developed based on segregation of unique alleles in the gametes of both parents, which resulted in one linkage map for each parent. A total of five microsatellite loci, GA-12, GA-126, GA-127, GA-131 and GAGG-5 were assigned to at least three linkage groups in both maps (Fregene et al. 1997, Table 5). Here we tested for linkage between pairs of all segregating microsatellite loci using Chi-square tests performed on the segregation ratios of gametes from both parents (Table 5). As reported by Fregene et al. (1997), linkage between GA-126 and GA-127 was also detected here, although the observed Chi-square value (11.33) was slightly lower than the critical value (11.34). These two loci have been assigned to group K of the female-derived cassava map at approximately 30cM from each other. In our study, no linkage was detected for any other pair of microsatellite loci in the female parent. However, in the male parent, microsatellite GA-131 appeared to be linked to five other loci, including GA-12 and GA-127, with observed Chi-square values larger than the critical value. Yet, these three microsatellite loci have been mapped to three different linkage groups of the male-derived cassava map (Table 5). We assumed that the segregation distortion observed for this microsatellite on the male-derived gametes may explain the discrepancy between our results and those of Fregene et al. (1997).

Mapping microsatellites in cassava helped to identify common linkage groups in the female- and male-derived maps for microsatellites GA-12, GA-127 and GA-131 (Table 5; Fregene et al. 1997). Microsatellite GA-12 revealed heterozygosity in both parents, although it has been mapped from the male parent only (segregation data was not available for the female parent), falling into a linkage group that has not been reconciled with the female-derived map. Success in mapping microsatellites like GA-12, with unique alleles segregating from both parents, would accomplish the recognition of common linkage groups from among those based on female- or male-derived segregation data.

Duplicated loci.

Barring priming site mutations, PCR amplification of highly polymorphic microsatellites should detect all alleles from duplicated loci in highly heterozygous, true auto- or allotetraploid species. However, in the case of cassava, 78% (11/14) of microsatellites detected only one or two alleles in most accessions, and segregation analysis of 11 microsatellites showed only one locus or two linked loci (Table 4). Microsatellite GA-161 was an exception detecting up to five alleles per genotype (Figure 1), though segregation analysis uncovered two linked loci. These results were unexpected considering that cassava is an outcrossing, very heterozygous crop of presumable allotetraploid origin (Magoon et al. 1969; Umanah and Hartman 1973). Although the knowledge on the evolution of microsatellite loci in plant genomes is limited, these findings may suggest a low degree of microsatellite loci duplication in cassava, which agrees with the low number of duplicated RFLP loci (less than 5%) found by Fregene et al. (1997). Duplicated microsatellite and RFLP loci are present in the cassava genome, though their origin as random genomic duplications or polyploidization events remains unknown.

Acknowledgments This research was supported by funds from the USAID and the Rockefeller Foundation, for which the authors are most grateful. We are grateful for excellent technical assistance provided by the late Christopher A. Jester and by Dr. Sharon E. Mitchell, both from USDA-ARS, Plant Genetic Resources Conservation Unit of the University of Georgia, Griffin, GA. We also thank Perkin Elmer/Applied Biosystems for providing fluorescently-labelled primers.

Table 1.

Forward and reverse primer sequences for 14 cassava microsatellites. (@) Annealing temperature at 45°C.

Microsatellite name	5' to 3' primer sequences
GAGG-5	TAATGTCATCGTCGGCTTCG GCTGATAGCACAGAACACAG
GA-12	GATTCCTCTAGCAGTTAAGC CGATGATGCTCTTCGGAGGG
GA-13	TTCCCTCGCTAGAACTTGTC CTATTTGACCGTCTTCGCCG
GA-16	GTACATCAACACCAACGGGC AGAGCGGTGGGGCGAAGAGC
GA-21	GGCTTCATCATGGAAAAACC CAATGCTTTACGGAAGAGCC
GA-57	AGCAGAGCATTTACAGCAAGG TGTGGAGTTAAAGGTGTGAATG
GA-123	CAATGCAGGTGAAGTGAATACC AGGGTGCTCTTCAGAGAAAGG
GA-126	AGTGGAAATAAGCCATGTGATG CCCATAATTGATGCCAGGTT
@GA-127	CTCTAGCTATGGATTAGATCT GTAGCTTCGAGTCGTGGGAGA
GA-131	TTCCAGAAAGACTTCCGTTCA CTCAACTACTGCACTGCACTC
GA-134	ACAATGTCCCAATTGGAGGA ACCATGGATAGAGCTCACCG
GA-136	CGTTGATAAAGTGGAAAGAGCA ACTCCACTCCCGATGCTCGC
GA-140	TTCAAGGAAGCCTTCAGCTC GAGCCACATCTACTCGACACC
GA-161	TGTTCTTGATCTTCTGCTGCA TGATTGTGGACGTGGGTAGA

Table 2.
Sizing of four fluorescently-labeled microsatellite loci amplified in a quadriplex PCR reaction from cassava cultivar MCol22 and run on 12 cm denaturing gels using a 373A DNA sequencer (Perkin Elmer/Applied Biosystems). Sizing is relative to the GeneScan 500 internal standard (TAMRA) using the "local Southern" algorithm. Inter-gel allele size differences are summarized for ten independent gel runs. @ sequence of original clone from accession Mcol22. (nt) nucleotides.

Locus	Size range (nt)	Inter-gel size differences (nt)	Mean size (nt)	Standard deviation (nt)	@ Size in nt of the sequenced clone
GA-21					
allele #1	106.23-107.79	1.56	107.19	0.4162	114
allele #2	114.02-115.63	1.61	114.98	0.4764	
allele #3	119.65-121.82	2.17	120.79	0.6329	
GA-126					
allele #1	182.28-183.25	0.97	182.68	0.3001	182
allele #2	188.13-188.76	0.63	188.49	0.2094	
allele #3	189.78-190.51	0.73	190.26	0.2467	
GA-134					
allele #1	307.82-308.56	0.74	308.22	0.3343	318
allele #2	317.65-318.89	1.24	318.52	0.3946	
GA-136					
allele #1	151.18-151.95	0.77	151.69	0.2296	158
allele #2	155.42-155.84	0.42	155.66	0.1535	
allele #3	159.12-159.68	0.56	159.50	0.1994	
		Mean = 1.0363		Mean = 0.3267	

Table 3. Heterozygosity estimates for 14 cassava microsatellites using (^{33}P)-end-labeled and/or fluorescently labeled primers for the detection of PCR amplification products. A maximum of two alleles per accession was the general observation, except for loci GA-161 and GA-127, for which up to five alleles per accession were detected. No null alleles were considered for *h* estimations, which may result in underestimation of *h*. (a) Microsatellites amplified in quadriplex reactions. (*) 39 accessions screened but only three showed a PCR product (see text). (@) Alleles detected using fluorescently labeled primers, all others using (^{33}P)-labeled primers.

Microsatellite	# of accessions scored for *h*	# of alleles detected	Size Range (nt)	Heterozygosity (*h*)
GA-12	39	6	131-157	0.54
GA-13	42	2	137-139	0.07
GA-16	41	6	89-129	0.63
GA-21	42	6	104-126	0.61
GA-21&	522	8@	107-121	0.65
GA-57	28	6@	153-183	0.67
GA-123	3*	1@	175	0.00
GA-126	34	8	178-214	0.85
GA-126&	521	12@	178-220	0.79
GA-127	34	10	203-239	0.74
GA-131	45	12	95-119	0.88
GA-134	38	4@	309-337	0.46
GA-134&	512	15@	308-338	0.55
GA-136&	522	9@	145-161	0.67
GA-140	35	5	154-164	0.76
GA-161	39	12	64-140	0.85
GAGG-5	28	6@	109-127	0.46

Table 4.
Goodness-of-fit X^2 test of expected and observed disomic segregation ratios for 11 microsatellite loci in the F_1 cassava mapping population CM7857. Segregation analysis was done using (^{33}P)-labeled primers and sequencing gels (*), or by running nonradioactive PCR products on 5% Metaphor Agarose gels stained with Ethidium Bromide (#). The null hypothesis that there were not significant differences between observed and expected segregation ratios was tested at α = 0.01 and n-1 degrees of freedom, where n = number of phenotypic classes (if n=2, 3 or 4, the critical regions are = 6.63, 9.21 or 11.34 respectively). (@) GA-127 and GA-134 had two linked loci in each parent but only one locus is shown here. GA-13, GA-57 and GA-123 did not segregate in the mapping population. (++) Segregating, weakly-amplified allele from male parent for locus GA-161a.

Microsatellite	Parental genotypes TMS30572 (left) and CM2177-2	Genotypic classes: expected (top) and observed ratios	Total individuals scored	Calculated X^2 and significance at a = 0.01
GAGG-5#	— / — —	— / — — 1 : 1 33: 43	75	1.08 NS
GA-12*	— — — / —	— — — / — — 1 : 1 : 1 : 1 20: 30: 19: 14	83	6.49 NS
GA-16*#	— — — —	— — — — 1 : 2 : 1 21: 43: 19	83	0.20 NS
GA-21#	— — —	— — — — 1 : 1 34: 43	77	1.05 NS
GA-126*	— — —	— — — — 1 : 1 35: 35	70	0.00 NS
GA-127*@	— — — —	— — — — — — — — 1 : 1 : 1 : 1 17: 20: 17: 14	68	1.06 NS
GA-131*#	— — — — —	— — — — — — — — 1 : 1 : 1 : 1 22: 8 : 11: 31	72	18.55 Significant
GA-134@	— — —	— — — 1 : 1 41: 38	79	0.11 NS

Table 4. (continued)

Microsatellite	Parental genotypes TMS30572 (left) and CM2177-2	Genotypic classes: expected (top) and observed ratios	Total individuals scored	Calculated X^2 and significance at $\alpha = 0.01$
GA-136*	— — — —	— — — — 1 : 1 36: 43	79	0.62 NS
GA-140*	— — — —	— — — — 1 : 2 : 1 21: 39: 16	76	0.71 NS
GA-161a*	— —	— — — ++ 1 : 1 31: 48	79	4.01 NS
GA-161b*	— — — —	— — — — 1 : 2 : 1 14: 40: 25	79	3.21 NS

Table 5.

Goodness-of-fit X^2 test to estimate the significance of linkage between pairs of microsatellite loci that segregate from th female (top table), the male (middle table), or both parents (lower table). Critical values for $X^2_{0.01}$ with 2, 3 or 5 degrees o freedom are 9.21, 11.34 and 15.08 respectively. Bold numbers indicate values too close to, or larger than the critical value: Also shown are the linkage groups to which loci have been assigned (source: Fregene et al. (1997); this report). (nd) no determined yet

Microsatellite	GA-12	GA-126	GA-127	GA-131	GA-136	GA-161b	Linkage group
GAGG-5	1.74	1.62	4.44	2.57	3.94	6.61	Q
GA-12		0.35	2.06	0.66	0.87	4.08	nd
GA-126			**11.33**	4.20	0.80	5.56	K
GA-127				0.57	3.02	6.15	K
GA-131					1.79	4.81	G
GA-136						7.73	nd

Microsatellite	GA-21	GA-127	GA-131	GA-134	GA-161a	GA-161b	Linkage group
GA-12	4.284	7.623	**15.94**	3.98	8.94	6.51	unnamed
GA-21		0.632	**11.74**	0.25	5.79	5.59	nd
GA-127			**14.63**	0.70	5.19	5.95	K
GA-131				**13.9**	**13.69**	13.32	G
GA-134					2.86	5.84	nd
GA-161a						**49.05**	nd
GA-161b							nd

Microsatellite	GA-140	GA-161b	linkage group
GA-16	2.662	3.451	nd
GA-140		0.794	nd

References

Beckman JS, Soller M (1990) Toward a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. Biotechnology 8: 930-932

Beeching JR, Marmey P, Gavalda MC, Noirot M, Haysom HR, Hughes MA, Charrier A (1993) An assessment of genetic diversity within a collection of cassava (*Manihot esculenta* Crantz) germplasm using molecular markers. Ann Bot 72:515-520

Bertram RB (1993) Application of molecular techniques to genetic resources of cassava (*Manihot esculenta* Crantz - Euphorbiaceae): interspecific evolutionary relationships and intraspecific characterization. Ph.D. thesis, University of Maryland. 465 p.

Broun P, Tanksley SD (1996) Characterization and genetic mapping of simple repeat sequences in the tomato genome. Mol Gen Genet (1996) 250:39-40

CIAT (1993) Cassava : The latest facts about an ancient crop. CIAT communication unit. AA 6713, Cali, Colombia

Dellaporta SL Wood J, Hicks JR (1983) A plant DNA minipreparation: version II. Plant Mol Biol Rep 1:19-21

Depeiges A, Goubely C, Lenior A, Cocherel S, Picard G, Raynal M, Grellet F, Delseny M (1995) Identification of the most represented repeated motifs in *Arabidopsis thaliana* microsatellite loci. Theor Appl Genet 91:160-168

Elder JK, Southern EM (1987). Computer-aided analysis of one-dimensional restriction fragment gels. In: Bishop MJ, Rawlings CJ (eds) Nucelic acid and protein sequence analysis - A practical approach. IRL Press, Oxford, UK, pp165-172

FAO (1992) Agrostat database 1992

Fregene M, Vargas J, Ikea J, Angel F, Tohme J, Asiedu RA, Akoroda MO, Roca WM (1994). Variability of chloroplast DNA and nuclear ribosomal DNA in cassava (*Manihot esculenta* Crantz). Theor Appl Genet 89:719-727

Fregene M, Angel F, Gomez R, Rodriguez F, Chavarriaga P, Roca WM, Tohme J, Bonierbale MW (1997) A Molecular genetic map for cassava (*Manihot esculenta* Crantz). Theor Appl Genet 95:431441

Freund A, Bichara M, Fuchs RPP (1989) Z DNA-forming sequences are spontaneous deletions hot spots. Proc Natl Acad Sci. USA 86:7465-7496

Ghosh S, Karanjawala ZE, Hauser ER, Ally D, Knapp JI, Rayman JB, Musick A, Tannenbaum J, Te C, Shapiro S, Eldridge W, Musick T, Martin C, Smith JR, Carpten JD, Brownstein MJ, Powell JI, Whiten R, Chines P, Nylund SJ, Magnuson VL, Boehnke M, Collins FS, FUSION Study Group (1997) Methods for precise sizing, automated binning of alleles and reduction of error rates in large-scale genotyping using fluorescently labeled dinucleotide markers. Genome Res 7:165-178

González DO, Palacios N, Tohme J (comps., eds.) (1995) Protocolos para Marcadores Moleculares. Unidad de Investigación en Biotecnología. Centro Internacional de Agricultura Tropical (CIAT) Cali-Colombia. 78p. [Publicación CIAT No. 258]

Hall JM, LeDuc CA, Watson AR, Roter AH (1996). An approach to high-throughput genotyping. Genome Res 6:781-790

Hershey C, Iglesias C, Iwanaga M, Tohme J (1994) Definition of a core collection for cassava. In: International network for cassava genetic resources. Report of the first meeting of the international network for cassava genetic resources. CIAT, Cali, Colombia, 18-23 August 1992. International crop network series No 10. International plant genetic resources, Rome, Italy. p 145-156

Kochert G, Halward T, Branch DW and Simpson CE (1991) RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. Theor Appl Genet 81:565-570

Kresovich S, Szewc-McFadden AK, Bliek SM, McFerson JR (1995) Abundance and characterization of simple sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. (rapeseed). Theor Appl Genet 91:206-211

Lagercrantz U, Ellengren H, Anderson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acid Res 21:1111-1115

Lefevre F and Charrier A (1993) Heredity of seventeen isozyme loci in cassava (*Manihot esculenta* Crantz) germplasm. Euphytica 66:171-178

Li WH, Graur D (1991) Fundamentals of molecular evolution. Sinauer Assoc Inc Sunderland, Massachusetts, USA, pp35-36

Magoon ML, Krishnan R, Bai KV (1969) Morphology of the pachytene chromosomes and meiosis in *Manihot esculenta* crantz. Cytologia 34:612-626

Marmey P, Beeching JR, Hamon S, Charrier A (1994) Evaluation of cassava (*Manihot esculenta* Crantz) germplasm collections using RAPD markers. Euphytica 74:203-209

Mitchell SE, Kresovich S, Jester CA, Hernandez J, Szewc-McFadden AK (1997) Application of Multiplex and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. Crop Sci. 37:617-624

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-90

Roa AC (1997) Estimación de la diversidad genética en *Manihot* spp. mediante morfología y marcadores moleculares. Tesis de Maestría. Universidad del Valle, Facultad de Ciencias, Programa Magister en Ciencias-Biología. Cali, Valle, Colombia. 220p.

Roa AC, Maya MM, Duque MC, Tohme J, Allem AC, Bonierbale MW (1997) AFLP analysis of relationships among cassava and other *Manihot* species. Theor Appl Genet 95:741-750

Roca WM, Henry G, Angel F, Sarria R (1992) Biotechnology research applied to cassava improvement at the International Center of Tropical Agriculture (CIAT). Ag Biotechnol News Inf 4:303N-308N

Roder MS, Plaschke J, König SU, Börner A, Sorrells ME, Tanksley SD, Ganal MW (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 246:327-333

Rogers DJ (1963) Studies on *M.esculenta* Crantz cassava and related species. Bull Torrey Bot Club 9:43-54

Rongwen J, Akkaya MS, Bhagwat AA, Lavi U, Cregan PB (1995) The use of microsatellite DNA markers for soybean genotype identification. Theor Appl Genet 90:43-48

Saghai Maroof MA, Biyashev RM, Yang GP, Zhang Q, Allard RW (1994) Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations, and population dynamics. Proc Nat Acad Sci USA 91:5466-5470

Sambrook J, Fritsch EF and Maniatis T (1989) Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory. Press, Cold Spring Harbor, NY

Sarria R, Ocampo C, Ramirez H, Hershey C, Roca WM (1993) Genetics of Esterase and Glutamate Oxaloacetate Transaminase isozymes in cassava (*Manihot esculenta* Crantz). In: Roca WM and Thro AM 1993. Proceedings of the first international scientific meeting of the Cassava Biotechnology Network, Cartagena, Colombia, 25-28 August 1992. -- Cali, Colombia: Centro Internacional de Agricultura Tropical, 1993. 75 p. [Working document, no. 123]

Schaal B, Olson P, Prinzie T, Carvalho LJCB, Tonukari NJ, Hayworth D (1995) Phylogenetic analysis of the genus *Manihot* based on molecular markers. In: The cassava biotechnology network: Proceedings of the second international scientific meeting, Bogor, Indonesia, 22-26 August 1994. -- Cali, Colombia: Centro Internacional de Agricultura Tropical, 1995. 62 p. 2v. -- [Working document, no. 150]

Smith RN (1995) Accurate size comparison of short tandem repeat alleles amplified by PCR BioTechniques 18(1):122-128

Tautz D (1989) Hipervariability of simple sequences as a general source of polymorphic DNA markers. Nucl Acid Res 17:6463-6471

Terauchi R (1994) A polymorphic microsatellite marker from the tropical tree *Dryobalanops lanceolata* (Dipterocarpaceae). Jpn J Genet 69:567-576

Thomas MR, Cain P, Scott NS (1994) DNA typing of grapevines: A universal methodology and database for describing cultivars and evaluating genetic relatedness. Plant Mol Bio 25:939-949

Umannah EE, Hartman RW (1973) Chromosome numbers and karyotypes of some *Manihot* species. J Am Soc Hort Sci 98:272-274

Wang Z, Weber JL, Zhong G, Tanksley SD (1994) Survey of plant short tandem repeats. Theor Appl Genet 88:1-6

Weber JL (1990) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Genomics 7:524-530

Zhao X, Kochert G (1993) Phylogenetic distribution and genetic mapping of a $(GGC)_n$ microsatellite from rice (*Oryza sativa* L.) Plant Mol Bio 21:607-614

Ziegle JS, Su Y, Corcoran KP, Nie L, Mayrand PE, Hoff LB, McBride LJ, Kronick MN, Diehl SR (1992) Application of automated DNA sizing technology for genotyping microsatellite loci. Genomics 14:1026-1031

Inheritance of random amplified polymorphic DNA markers in cassava (*Manihot esculenta* Crantz)

Rocio Gomez, Fernando Angel, Merideth W. Bonierbale, Fernando Rodriguez, Joseph Tohme, and William M. Roca

Abstract: The informativeness and inheritance of randomly amplified polymorphic DNA (RAPD) markers were investigated in an intraspecific F_1 progeny derived from two heterozygous parents. The analysis confirmed the utility of RAPD markers for comparing candidate parents for the development of a molecular genetic map, and provided numerous markers for linkage analysis in a crop with a very limited history of classical or molecular genetic studies. Six potential parental lines (themselves F_1 hybrid clones) showed between 1.82 and 0.62 segregating bands per primer in three hybrid families. Forty-three percent (309) of 722 primers produced polymorphic products in the most informative of these three crosses, revealing 328 single-dose (SD) markers segregating 1:1 for presence/absence in a progeny of 90 individuals. A second class of informative markers were those present in both parents but segregating in the progeny. Fifty-seven or 67% of the monomorphic but segregating markers exhibited the 3:1 ratio expected for SD dominant markers in a cross between heterozygotes. Linkage groups were constructed from the segregation of SD RAPD markers originating in the female (TMS 30572) and the male (CM2177-2) parent.

Key words: RAPDs, molecular markers, genetic segregation, *Manihot*, single-dose markers.

Résumé : Le caractère informatif et l'hérédité de marqueurs d'ADN polymorphe amplifié au hasard (RAPD) ont été étudiés chez la progéniture F_1 issue d'un croisement intraspécifique impliquant deux parents hétérozygotes. L'analyse a confirmé l'utilité des marqueurs RAPD en vue de la comparaison de parents potentiels en vue de l'élaboration d'une carte génétique. De plus, de nombreux marqueurs ont été obtenus pour des fins d'analyse de liaison génétique chez cette espèce qui a fait l'objet de très peu d'études en génétique classique et moléculaire. Six lignées parentales potentielles (elles-mêmes des clones hybrides F_1) ont montré entre 1.82 et 0.62 bandes en ségrégation par amorce chez trois familles hybrides. Quarante-trois pour cent (309) de 722 amorces ont amplifié des produits polymorphes chez le plus informatif des trois croisements. Ces amorces ont permis de révéler 328 marqueurs présents en simple dose, c'est-à-dire qui montraient une ségrégation 1:1 (absence/présence) chez une progéniture comprenant 90 individus. Une seconde classe de marqueurs informatifs étaient présents chez les deux parents mais étaient en ségrégation chez la progéniture. Cinquante-sept, soit 67%, des marqueurs monomorphes en ségrégation montraient le rapport 3:1 attendu de marqueurs dominants à simple dose suite au croisement de deux hétérozygotes. Des linkats ont pu être établis à partir de la ségrégation des marqueurs RAPD à simple dose présents chez les parents femelle (TMS 30572) et mâle (CM2177-2).

Mots clés : RAPD, marqueurs moléculaires, ségrégation génétique, *Manihot*, marqueurs à simple dose.
[Traduit par la Rédaction]

Cassava, *Manihot esculenta* Crantz is an important species in tropical agriculture. In the tropical regions where it is cultivated, cassava is the fourth most important source of calories, after rice, maize, and sugarcane. It is estimated that about 70 million people obtain more than 500 Calories (1 Cal. = 4.1855 kJ) a day from cassava, particularly in Africa (Cock 1985). Considered as a diploid species of allopolyploid origin (Umanah and Hartman 1973), cassava is genetically the least understood among the major staple crops that feed mankind. Only a handful of morphological mutants (Hershey and Ocampo 1989) and several isozyme markers have been developed for cassava genetics. Analysis of 10 isozyme loci has recently revealed predominantly disomic inheritance (Sarria et al. 1993; Lefevre and Charrier 1993). In the course of developing a molecular genetic map of cassava using DNA markers, we have investigated the utility of random amplified polymorphic DNAs or RAPDs (Williams et al. 1990). RAPD markers have been used to construct genetic maps (Tulsieram et al. 1992; Al-Janabi et al. 1993; Rowland and Levi 1994), to detect genetic variation within plant species (Williams et al. 1990), to evaluate levels of gene flow between species (Arnold et al. 1991), and to obtain markers linked to resistance genes (Martin et al. 1991; Michelmore et al. 1991).

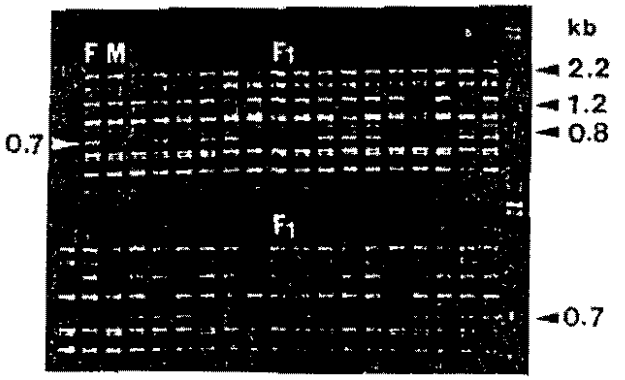
Corresponding Editor: G. Jenkins.
Received January 28, 1996. Accepted July 4, 1996.

R. Gomez, F. Angel, M.W. Bonierbale, F. Rodriguez, J. Tohme, and W.M. Roca. Biotechnology Research Unit, Centro Internacional de Agricultura Tropical (CIAT), Apartado Aereo 6713, Cali, Colombia.

With RAPD markers, polymorphisms are usually detected according to the presence or absence of amplified fragments. Wu et al. (1992) described a specific class of molecular markers, called single-dose (SD) markers, that are present in one (heterozygous) parent, absent in the other, and segregate 1:1 in the progeny. SD markers have been used to develop genetic maps of potato (Bonierbale et al. 1988), sugarcane (Al-Janabi et al. 1993), and *Eucalyptus* (Grattapaglia and Sederoff 1994) based on segregation from heterozygous parents. A single primer usually amplifies multiple fragments, and each polymorphic fragment can be treated as a separate SD marker if it segregates 1:1 for presence/absence in the progeny. Since SD markers may segregate simultaneously and independently from each heterozygous parent, the two sources of segregation are considered separately in linkage analyses. RAPDs present in the heterozygous state in both parents are also informative, with the expected presence/absence ratio of 3:1 in the progeny.

Three different intraspecific crosses, CM7857 (TMS 30572 × CM2177-2), CM8382 (TMS 30572 × CM3299-4), and CM8224 (THAI-1 × CM3372-4), were examined as candidates for genetic mapping. Total genomic DNA was extracted from greenhouse- or field-grown leaf tissue of the parents and 90, 20, and 20 F_1 individuals, respectively, using the method described by Dellaporta et al. (1983). RAPD primers were screened for revealing polymorphisms and segregation, using the parents and a subset of 15 progeny, before the whole F_1 population was analyzed. Target DNA sequences were amplified by PCR on polyvinylchloride microtest plates (Falcon) placed in a 96-well MJ Research Programmable Thermal Controller (PTC-100). A single decanucleotide (Operon Technologies Inc., Alameda, Calif.), containing 60–70% G–C, served as a random primer for each reaction. Each 12.5-μL amplification reaction contained 25 ng of plant genomic DNA, 10 mM Tris–HCl (pH 9.0), 50 mM KCl, 0.01% Triton X-100, 2.5 mM $MgCl_2$, 0.2 mM of each dNTP (New England Biolabs), 0.8 μM of primer, and 1 U of Taq DNA polymerase (Perkin Elmer). The mixture was overlayed with mineral oil. With a genome size of 1.43 pg (C. Martinez, unpublished data; Arumuganathan and Earle 1991), haploid genome content, the amount of genomic DNA used in an RAPD reaction corresponded to 17 000 haploid genome equivalents. Amplification was performed as described by Yu and Pauls (1992) with slight modifications (94°C for 5 min, followed by 34 cycles at 94°C for 5 s, 36°C for 30 s, 72°C for 1 min, and ending with 5 min at 72°C). Seven hundred and twenty-two primers of arbitrary sequences were screened. The PCR products were assayed by electrophoresis in 1.4% agarose gels run with TBE buffer (89 mM Tris-borate plus 2 mM EDTA, pH 8.3). Markers were scored directly on the ethidium bromide stained gel under uv light and confirmed using black and white Polaroid film. Segregating bands were classified according to their origin in the pistillate or staminate parent, or both. χ^2 tests were conducted to assess the goodness of fit of segregation in the progeny to the expected 1:1 or 3:1 ratios, and separate data sets were developed for the 1:1 segregation of SD markers from each parent. Nomenclature for individual RAPD markers describes the primer and relative RAPD marker size from largest to smallest, i.e., K16a, K16b, etc. To determine the linkage relationships among RAPD markers we ran MAPMAKER (Lander et al. 1987) using a minimum LOD score of 5.00 and a maximum θ value (recombination fraction) of 0.25. The proportion of markers in coupling versus repulsion phase was compared as discussed in Wu et al. (1992). This was achieved by copying the markers and inverting the scores for each of the polymorphisms (called recoding in the MAPMAKER version 2.0 manual), and then running MAPMAKER again (LOD = 5.0, θ = 0.25, two-point analysis) looking for linkages between the first markers and the newly created second set of markers. Markers in repulsion phase should appear on the same linkage group when this is done, but each linkage group is presented twice and mirror images have to be discarded. Centimorgan (cM) values were calculated using the Kosambi function (Kosambi 1944).

Fig. 1. Genetic inheritance and segregation of a 0.7-kb RAPD marker in the F_1 population. RAPD marker is present in the female parent (F), absent in the male parent (M), and segregates 1:1 in the progeny. Right lane contains lambda DNA digested with *Pst*I.

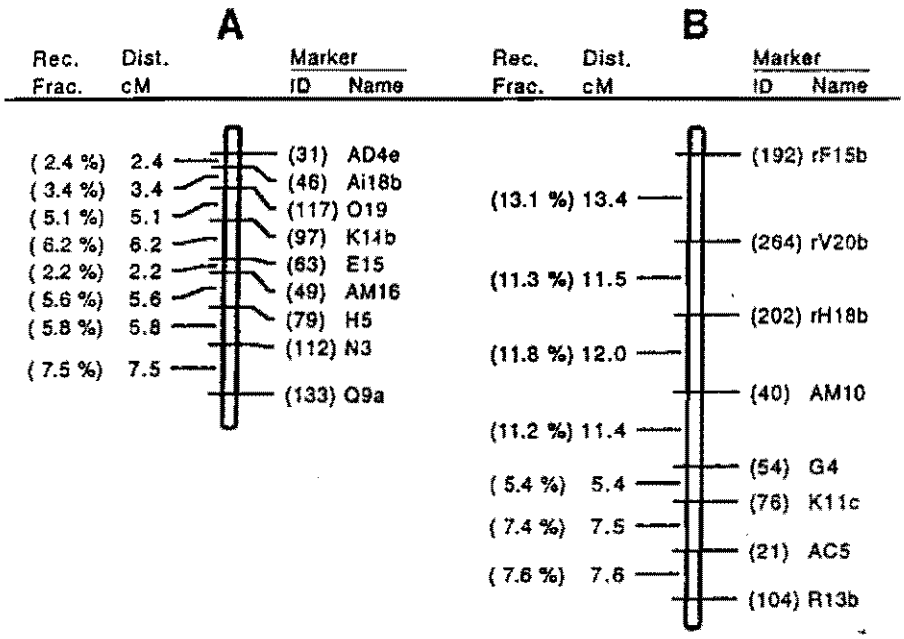


Sixty-six random primers were used to amplify genomic DNA from the parents and 15 individuals of each of the three candidate crosses. The proportion of primers that revealed polymorphisms was highest in CM7857 (Gomez et al. 1995). The total numbers of amplified and segregating bands observed with 66 primers in the progenies of the three sexual crosses are

	Amplified bands:		Segregating bands:	
Cross	total	per primer (median)	total	per primer (median)
CM7857	405	6.1	120	1.82
CM8382	389	5.9	68	1.03
CM8224	355	5.4	41	0.62

A χ^2 for independence of segregation and family indicated a significant difference among families ($p < 0.005$) in the presence of segregating bands. As CM7857 presented more segregating bands over all, as well as more segregating bands per primer, it was considered the best candidate of the

Fig. 2. Examples of linkage groups of RAPDs segregating in the female (A) and male (B). Single-dose RAPD markers segregate 1:1 in F_1 progeny and are ordered within groups with a confidence of LOD >2.5. Rec. frac., recombination fraction; Dist., distance.



three for genetic mapping. Furthermore, this cross is expected to present segregation for a number of agronomically important traits that could eventually be associated with DNA polymorphisms. Further studies were carried out using CM7857. A total of 722 PCR primers of arbitrary sequences, 10 nucleotides in length, were used singly in PCR assays to screen for DNA polymorphisms among 90 progeny of the F_1 population. Primer screening was efficiently carried out using both parents and a sample of 15 F_1 individuals. With this format, parental origin of the markers, as well as their allelic state (homozygous or heterozygous), was directly inferred from the presence of the fragment in one parent, the absence in the other, and segregation (presence/absence) in the F_1 progeny sample. An RAPD marker was counted only if the phenotypic classes were consistent and clearly distinguishable. In instances, where more than two phenotypic classes were evidenced by band intensity differences, the band was counted only if one of the classes was a virtual absence of the marker, and the intensity variants were summed in a single "presence" class.

Of the 722 arbitrary primers screened in CM7857, 127 (18%) did not yield any amplified product, 254 (35%) did not detect polymorphism or segregation, 309 (43%) revealed at least one polymorphism, and 32 (4%) revealed monophormic bands (present in both parents) that segregated in the progeny. The 309 primers showing at least one polymorphism generated a total of 554 segregating RAPD markers. Of these markers, 328 fit a 1:1 ratio ($p > 0.05$). The markers segregating 1:1 were generated by 246 primers, yielding an average of 1.33 RAPD markers per primer, similar to the number found in peach (Chaparro et al. 1994) and lower than reports of 1.93 in *Picea abies* Karst. (Binelli and Bucci 1994), 1.8 in *Stylosanthes*, (Kazan et al. 1993), and 1.88 in sugarcane (Al-Janabi et al. 1993). Amplified fragment size ranged from 3.2 to 0.1 kilobases (Fig. 1). Segregation of the 328 SD markers was scored in 90 F_1 individuals of CM7857. The number of SD markers inherited from each parent was 186 from TMS 30572 (pistillate) and 142 from CM2177-2 (staminate). Eighteen polymorphic markers segregated as double dose (DD) markers, 10 from TMS 30572 and 8 from CM2177-2. Only 85 markers were present in both parents and segregated in the progeny. Of these markers, 57 fit a 3:1 ratio, as expected for SD markers, and 12 segregated as DD markers. The number of observed and expected SD markers to multiple dose markers was 0.59, which fits the expectation for disomic inheritance. This observation supports earlier reports made by Lefevre and Charrier (1993) based on isozyme analyses of F_1 populations from intra- and inter-specific crosses. Under the described conditions, individual RAPD phenotypes were highly reproducible for an individual among DNA replicates. Samples lacking DNA giving amplification products, which have been reported in some cases (Torres et al. 1993), were not present in our experiments. Furthermore, we never observed a fragment in the F_1 that was not observed in at least one of its parental clones. In summary, 385 SD RAPD markers were considered in this study, 328 polymoprhic in the parents and 57 present in both parents. Furthermore, 30 markers fit the expectation for DD markers. Twenty-five percent of RAPD markers showed distorted segregation. Distorted segregation may be the result of linkages between markers and the genes operating in prezygotic and postzygotic phases of reproduction (Zamir and Tadmor 1986). Although we did not examine the reason(s) for such segregation ratios, other

reasons may include preferential chromosome elimination (Tanksley 1984), preferential fertilization, and selective elimination of particular zygotes (Koening and Gepts 1989).

The molecular nature of RAPD markers and their Mendelian inheritance make them valuable tools for genetic analyses in cassava. Two linkage groups, based on segregation of these markers in the female and male, are shown in Fig. 2. The groups were constructed using the MAPMAKER program with a LOD score of 5.0 and a recombination fraction of 0.25. The first linkage group drawn for TMS 30572 (pistillate parent) is 38.2 cM in length and is composed of 9 markers with a maximum map distance of 7.5 cM. No markers in repulsion are present in this linkage group. The sample linkage group of CM2177-2 (staminate parent) is composed of 8 markers, five of them linked in coupling and three in repulsion (rF15b, rV20b, and rH18b), indicating that they represent dominant markers from homologous chromosomes. For allopolyploids, the proportion of linkages that can be detected in the coupling phase is expected to be the same as that in the repulsion phase (Wu et al. 1992). Thirty-eight percent of SD RAPD markers were found to be linked in repulsion phase (results not shown). In an integrated linkage map of cassava constructed from RFLPs, microsatellites, isozymes, and RAPD markers, 30% of markers were found to be linked in repulsion, suggesting preferential pairing among some chromosomes and random pairing among others (M. Fregene, F. Angel, R. Gomez, F. Rodriguez, P. Chavarriaga, M.W. Bonierbale, J. Tohme, and W.M. Roca, in preparation).

The RAPDs described here represent the largest collection of genetic markers that have been subjected to segregation analysis in cassava to date. Data from these markers will be pooled with data from RFLPs, using cloned cDNA, genomic DNA probes; microsatellites; and isozymes to create a saturated molecular linkage map of cassava, which in theory, should have a minimum of bias for coding and noncoding chromosomal regions. RAPD markers are less informative than codominant markers, such as RFLPs or microsatellites; nevertheless, we feel that RAPDs are potent molecular markers not only for the construction of linkage maps, but also for assessing the validity of controlled crosses. The mapping population used in this study revealed a wide range of variability for resistance to bacteriosis (CBB), root perishability, and physiological parameters of photosynthesis. The cassava map can serve to identify important loci controlling these traits by their cosegregation with mapped molecular markers, and the resulting genetic information can be applied to the improving of selection schemes. In summary, in this work we have established a methodology of RAPD analysis for cassava and have demonstrated the use of these markers for genetic analyses on an intraspecific cross.

Acknowledgments

This research was supported by the Rockefeller Foundation's Biotechnology Program, RF 91060 No. 2. R.G. is a graduate student from the Universidad Javeriana, sponsored by COLCIENCIAS, Project No. 2228-07-003-92 and by Centro Internacional de Fisica (CIF).

References

Al-Janabi, S.M., Honeycutt, R.J., McClelland, M., and Sobral, B.W.S. 1993. A genetic linkage map of *Saccharum spontaneum* L. "SES 208". Genetics, **134**: 1249–1260.

Arnold, M.L., Buckner, C.M., and Robinson, J.J. 1991. Pollen mediated introgression and hybrid speciation in Louisiana irises. Proc. Natl. Acad. Sci. U.S.A. **88**: 1398–1402.

Arumuganathan, K., and Earle, E.D. 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. **9**: 208–218.

Binelli, G., and Bucci, G. 1994. A genetic linkage map of *Picea abies* Karst, based on RAPD markers, as a tool in population genetics. Theor. Appl. Genet. **88**: 283–288.

Bonierbale, M.W., Plaisted, R.L., and Tanksley, S.D. 1988. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics, **120**: 1095–1103.

Chaparro, J.X., Werner, D.J., O'Malley, D., and Sederoff, R.R. 1994. Targeted mapping and linkage analysis of morphological isozyme, and RAPD markers in peach. Theor. Appl. Genet. **87**: 805–815.

Cock, J.H. 1985. Cassava: new potential for a neglected crop. Boulder and London, Western Press.

Dellaporta, S.L., Wood, J., and Hicks, J.B. 1983. A plant DNA minipreparation: version II. Plant. Mol. Biol. Rep. **1**: 19–21.

Gomez, R., Angel, F., Bonierbale, M.W., Rodriguez, F., Tohme, J., and Roca, W.M. 1995. Selecting heterozygous parents and single-dose markers for genetic mapping in cassava. *In* CBN II: Proceedings of the Second International Scientific Meeting of the Cassava Biotechnology Network, held at Bogor, Indonesia, 22–26 August 1994. *Edited by* A.M. Thro and W.M. Roca. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia. pp. 113–124.

Grattapaglia, D., and Sederoff, R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics, **137**: 1121–1137.

Hershey, C.H., and Ocampo, C.H. 1989. New marker genes found in cassava. Cassava Newsletter No. 13. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia. pp. 1–5.

Kazan, K., Manners, J.M., and Cameron, D.F. 1993. Inheritance of random amplified polymorphic DNA markers in an interspecific cross in the genus *Stylosanthes*. Genome, **36**: 50–56.

Koening, R., and Gepts, P. 1989. Segregation and linkage of genes for seed proteins, isozymes, and morphological traits in common bean (*Phaseolus vulgaris*). J. Hered. **80**: 455–459.

Kosambi, D. 1944. The estimation of map distances from recombinant values. Annu. Eugen. **12**: 172–175.

Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E., and Newbburg, L. 1987. Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics, **1**: 174–181.

Lefevre, F., and Charrier, A. 1993. Heredity of seventeen isozyme loci in cassava (*Manihot esculenta* Crantz). Euphytica, **66**: 171–178.

Martin, G.B., Williams, J.G.K., and Tanksley, S.D. 1991. Rapid identification of markers linked to a *Pseudomonas* resistance in tomato by using random primers and near-isogenic lines. Proc. Natl. Acad. Sci. U.S.A. **88**: 2336–2340.

Michelmore, R.W., Paran, I., and Kesseli, R.V. 1991. Identification of markers linked to disease-resistance genes by bulked segregating analysis. A rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Natl. Acad. Sci. U.S.A. **88**: 1212–1216.

Rowland, L.J., and Levi, A. 1994. RAPD-based genetic linkage map of blueberry derived from a cross between diploid species (*Vaccinium darrowi* and *V. elliottii*). Theor. Appl. Genet. **87**: 863–868.

Sarria, R., Ocampo, C., Ramirez, H., Hershey, C.H., and Roca, W.M. 1993. Genetics of esterase and glutamate oxaloacetate transaminase isozymes in cassava (*Manihot esculenta* Crantz). *In* CBN I: Proceedings of the First Scientific Meeting of the Cassava Biotechnology Network, held at Cartagena, Colombia, 20–24 August 1992. *Edited by* A.M. Thro and W.M. Roca. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia. pp. 75–80.

Tanksley, S.D. 1984. Linkage relationships and chromosomal location of enzyme coding genes in pepper, *Capsicum annuum*. Chromosoma, **89**: 352–360.

Torres, A.M., Weeden, N.F., and Martin, A. 1993. Linkage among isozyme, RFLP and RAPD markers in *Vicia faba*. Theor. Appl. Genet. **85**: 937–945.

Tulsieram, L.K., Glaubitz, J.C., Kiss, G., and Carlson, J.E. 1992. Single tree genetic linkage mapping in conifers using haploid DNA from megagametophytes. Bio/Technology, **10**: 686–690.

Umanah, E.E., and Hartmann, R.W. 1973. Chromosome numbers and karyotypes of some *Manihot* species. J. Am. Soc. Hort. Sci. **98**: 272–274.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids. Res. **18**: 6531–6535.

Wu, K.K., Burnquist, W., Sorrells, M.E., Tew, T.L., Moore, P.H., and Tanksley, S.D. 1992. The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor. Appl. Genet. **83**: 294–300.

Yu, K.F., and Pauls, K.P. 1992. Optimization of the PCr program for RAPD analysis. Nucleic. Acids. Res. **20**: 2606.

Zamir, D., and Tadmor, Y. 1986. Unequal segregation of nuclear genes in plants. Bot. Gaz. **147**: 355–358.

reasons may include preferential chromosome elimination (Tanksley 1984), preferential fertilization, and selective elimination of particular zygotes (Koening and Gepts 1989).

The molecular nature of RAPD markers and their Mendelian inheritance make them valuable tools for genetic analyses in cassava. Two linkage groups, based on segregation of these markers in the female and male, are shown in Fig. 2. The groups were constructed using the MAPMAKER program with a LOD score of 5.0 and a recombination fraction of 0.25. The first linkage group drawn for TMS 30572 (pistillate parent) is 38.2 cM in length and is composed of 9 markers with a maximum map distance of 7.5 cM. No markers in repulsion are present in this linkage group. The sample linkage group of CM2177-2 (staminate parent) is composed of 8 markers, five of them linked in coupling and three in repulsion (rF15b, rV20b, and rH18b), indicating that they represent dominant markers from homologous chromosomes. For allopolyploids, the proportion of linkages that can be detected in the coupling phase is expected to be the same as that in the repulsion phase (Wu et al. 1992). Thirty-eight percent of SD RAPD markers were found to be linked in repulsion phase (results not shown). In an integrated linkage map of cassava constructed from RFLPs, microsatellites, isozymes, and RAPD markers, 30% of markers were found to be linked in repulsion, suggesting preferential pairing among some chromosomes and random pairing among others (M. Fregene, F. Angel, R. Gomez, F. Rodriguez, P. Chavarriaga, M.W. Bonierbale, J. Tohme, and W.M. Roca, in preparation).

The RAPDs described here represent the largest collection of genetic markers that have been subjected to segregation analysis in cassava to date. Data from these markers will be pooled with data from RFLPs, using cloned cDNA, genomic DNA probes; microsatellites; and isozymes to create a saturated molecular linkage map of cassava, which in theory, should have a minimum of bias for coding and noncoding chromosomal regions. RAPD markers are less informative than codominant markers, such as RFLPs or microsatellites; nevertheless, we feel that RAPDs are potent molecular markers not only for the construction of linkage maps, but also for assessing the validity of controlled crosses. The mapping population used in this study revealed a wide range of variability for resistance to bacteriosis (CBB), root perishability, and physiological parameters of photosynthesis. The cassava map can serve to ··ify important loci controlling these traits by their ·egation with mapped molecular markers, and the genetic information can be applied to the improv- ection schemes. In summary, in this work we lished a methodology of RAPD analysis for cas- nd have demonstrated the use of these markers for etic analyses on an intraspecific cross.

Acknowledgments

This research was supported by the Rockefeller Foundation's Biotechnology Program, RF 91060 No. 2. R.G. is a graduate student from the Universidad Javeriana, sponsored by COLCIENCIAS, Project No. 2228-07-003-92 and by Centro Internacional de Fisica (CIF).

References

Al-Janabi, S.M., Honeycutt, R.J., McClelland, M., and Sobral, B.W.S. 1993. A genetic linkage map of *Saccharum spontaneum* L. "SES 208". Genetics, **134**: 1249–1260.

Arnold, M.L., Buckner, C.M., and Robinson, J.J. 1991. Pollen mediated introgression and hybrid speciation in Louisiana irises. Proc. Natl. Acad. Sci. U.S.A. **88**: 1398–1402.

Arumuganathan, K., and Earle, E.D. 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. **9**: 208–218.

Binelli, G., and Bucci, G. 1994. A genetic linkage map of *Picea abies* Karst, based on RAPD markers, as a tool in population genetics. Theor. Appl. Genet. **88**: 283–288.

Bonierbale, M.W., Plaisted, R.L., and Tanksley, S.D. 1988. RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics, **120**: 1095–1103.

Chaparro, J.X., Werner, D.J., O'Malley, D., and Sederoff, R.R. 1994. Targeted mapping and linkage analysis of morphological isozyme, and RAPD markers in peach. Theor. Appl. Genet. **87**: 805–815.

Cock, J.H. 1985. Cassava: new potential for a neglected crop. Boulder and London, Western Press.

Dellaporta, S.L., Wood, J., and Hicks, J.B. 1983. A plant DNA minipreparation: version II. Plant. Mol. Biol. Rep. **1**: 19–21.

Gomez, R., Angel, F., Bonierbale, M.W., Rodriguez, F., Tohme, J., and Roca, W.M. 1995. Selecting heterozygous parents and single-dose markers for genetic mapping in cassava. *In* CBN II: Proceedings of the Second International Scientific Meeting of the Cassava Biotechnology Network, held at Bogor, Indonesia, 22–26 August 1994. *Edited by* A.M. Thro and W.M. Roca. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia. pp. 113–124.

Grattapaglia, D., and Sederoff, R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics, **137**: 1121–1137.

Hershey, C.H., and Ocampo, C.H. 1989. New marker genes found in cassava. Cassava Newsletter No. 13. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia. pp. 1–5.

Kazan, K., Manners, J.M., and Cameron, D.F. 1993. Inheritance of random amplified polymorphic DNA markers in an interspecific cross in the genus *Stylosanthes*. Genome, **36**: 50–56.

Koening, R., and Gepts, P. 1989. Segregation and linkage of genes for seed proteins, isozymes, and morphological traits in common bean (*Phaseolus vulgaris*). J. Hered. **80**: 455–459.

Kosambi, D. 1944. The estimation of map distances from recombinant values. Annu. Eugen. **12**: 172–175.

Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E., and Newbburg, L. 1987. Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics, **1**: 174–181.

Lefevre, F., and Charrier, A. 1993. Heredity of seventeen isozyme loci in cassava (*Manihot esculenta* Crantz). Euphytica, **66**: 171–178.

Martin, G.B., Williams, J.G.K., and Tanksley, S.D. 1991. Rapid identification of markers linked to a *Pseudomonas* resistance in tomato by using random primers and near-isogenic lines. Proc. Natl. Acad. Sci. U.S.A. **88**: 2336–2340.

Michelmore, R.W., Paran, I., and Kesseli, R.V. 1991. Identification of markers linked to disease-resistance genes by bulked segregating analysis. A rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Natl. Acad. Sci. U.S.A. **88**: 1212–1216.

Annex 1

Proposed Two Year Extension of the Project:

Saturation of the Genetic Map of Cassava with PCR-Based Markers and the Use of the Genetic Map in the Improvement of Cassava

Project Goal:

The two year extension intends to consolidate ongoing work on the development of SSR and EST markers, genetic analysis of agronomic traits, and the application of large DNA insert size libraries to gene cloning of disease resistance genes, towards enhancing the undisputed position of cassava as a food security crop and engine of economic development in Sub-Saharan Africa.

Project Objectives:

1. Second year multilocational trial, with the F_1 genetic map population, to evaluate tentative quantitative trait loci (QTL) controlling earliness, dry matter content, post-harvest deterioration (PHD), starch quality/content, culinary quality, and other important morphological traits, identified in the first year. Marker fidelity studies with a BC_1 cross developed from the F_1 map population. Marker-assisted population development, by selection of parents with greatest breeding values for the above traits.
2. Genetic mapping of a few hundred SSR markers, already in the process of development, on to the molecular genetic map of cassava.
3. Genetic mapping of all sequenced ESTs and sequencing of the remaining transcript derived fragments (TDF) identified.
4. Fine mapping and BAC contig mapping of regions of the cassava genome carrying disease resistance genes.
5. Transfer of SSR marker technology to national programs, through the activities of the Cassava Molecular diversity Network, for the evaluation and structuring of cassava genetic diversity in Sub-Saharan Africa compared to the total available in cassava and wild progenitors.

Introduction

The first two years of this project has yielded important outputs, described in the attached progress report, a two year extension is required to bring the activities begun in 1996 to a conclusion. Furthermore one of the principal objectives of the extension, specifically development and genetic mapping of SSR markers, is key to the success of a proposed study to assess genetic diversity of cassava in Africa with molecular markers, a study which is expected to generate information, critical, and complementary to other efforts towards a systematic improvement of the crop in Africa.

Project Justification:

The irreducible uncertainties of cassava farming systems in Sub-Saharan Africa, due to labor and agricultural input scarcities, pests, diseases, soil fertility problems, arising from shorter cycles of shifting cultivation and cropping of marginal soils, drought and other climatic factors has meant that production is highly vulnerable. In Africa more than 80% of cassava is produced in poverty hot-spot areas, where as much as 70% of the rural population lives under absolute poverty and practice low resource input agriculture (Fresco 1993). It has been suggested that broadening local germplasm with disease, pest, root quality and nutrient-use efficiency genes, while endeavoring to preserve most of farmer-preferred traits might be a more realistic paradigm of cassava improvement for the region (Gullberg 1998 pers. comm.). This is a task that requires introducing specific traits, while maintaining existing ones, in an efficient manner, and maximizing genetic variation to achieve combination of favorable alleles for widely varying agro-ecological and production niches

A molecular marker-assisted approach is required to upgrade populations, with specific genes, while maintaining genes controlling desired traits at a high frequency, in a cost-effective and efficient manner (Tanksley et. al. 1989). During the two year extension of these project, ongoing experiments to map quantitative trait loci (QTL) controlling traits of agronomic interest in the F1 map population will be concluded and extended to a BC1 cross derived from it to examine genetic background effects. Markers linked to identified

QTLs will employed to chose parents with greatest breeding values for traits analyzed simultaneously.

Should marker-aided selection of parents be found to be more efficient compared to phenotypic selection, a unique opportunity for doing recurrent selection, the most powerful breeding method for cassava, in cycles of one year becomes feasible. Cassava is vegetatively propagated, current methods of recurrent selection requires: one year to establish seedlings and obtain clones, another year to multiply clones for replicated multi-locational trials, two years for multi-locational trials and a last year for making crosses to complete the cycle. Markers linked to QTLs known to contribute most to the traits, cuts out the need for clonal evaluation in multi-locational trials, making a one year cycle feasible. Compared to the current 6 years required for one cycle, one year of marker-assisted recurrent selection is immense savings in time and resources, and a big boost to cassava breeding.

Several ESTs, which represent genes, will be added to the map. Together with genes of known function already on the map, including cloned starch genes, (Munyinkwa et al., 1994), cyanogenic biosynthesis genes (Anderson et al. unpublished data, 1998), and a few mapped ESTs, these highly informative markers are expected to increase the probability of tagging QTLs controlling traits of agronomic importance.

Cassava is of South American origin, and it is not known what proportion of genetic diversity of cassava present in South America and wild progenitors has been transferred to Africa. What is required is a genome-wide quantitative assessment of genetic divsersity of cassava in Africa versus that in South America. Molecular markers have been recognized as unbiased tools in the characterization of genetic diversity (Kresovich et al. 1995; Roder et al. 1995; Saghai Maroof et al 1994). The molecular markers of preference in cassava are simple sequence repeats (SSR) due to their genome-wide availability, ease of use, and co-dominant nature. SSR markers have been used in genetic diversity and linkage studies in cassava (Chavarriaga et al. 1998a; 1998b; Mkumbira et al. 1998, in preparation; Muller-Dos Santos et al. 1998, in preparation;

Fregene et al. 1997). One of the principal objectives of this project is to saturate the genome of cassava with SSR markers, which will be available for studies of genetic diversity of cassava in Africa.

Disease and pest epidemics are a major source of production instability in major cassava growing regions, with the attendant loss of valuable germplasm. They are combated, mostly, by resistance breeding in cassava, but cassava's long growth cycle, outcrossing and vegetatively propagated nature considerably slows down resistance breeding, and restricts the powerful option of introgression by marker-assisted backcrossing. The rapid advance in cloning of plant disease resistance genes, together with reproducible transformation protocols in cassava makes genetic transformation with homologous disease and pest resistance genes, a faster and more efficient way of moving around useful disease resistance genes amongst cassava genepools.

Cloning genes known only by their phenotypes and position relative to molecular markers on a genetic map requires libraries of large DNA fragments that can be ordered into contigs, that span the genomic region carrying the gene(s) of interest. The cassava mapping project has identified regions of the genome carrying resistance genes to the cassava bacterial blight (CBB) (Jorge et al., in preparation), and similar studies are underway for the African cassava mosaic disease (ACMD), and the white fly disease. A BAC library has been generated for cassava from a CBB, and ACMD resistant variety (Fregene et. al. 1998). An objective of this two year extension is the fine mapping and BAC contig mapping of identified regions bearing resistance gene.

Project Description:

QTL Mapping of Agronomic traits

The F1 mapping population, from which the cassava map was developed, is also the QTL mapping population; the F_1 cross has been genotyped with over 200 markers, including known starch biosynthesis genes, cyanogenesis genes, and expressed sequence tags. The field experiment for the second year will be the same as the first year evaluation, a triple partially balanced lattice design with 20 plant per genotype/plot per replication, with

three replications, in two sites, CIAT, Palmira and CIAT, Quilichao. Adjusted means for the traits will be used for the QTL analysis by single regression using the Q-gene and PGRI package. Heritability of the different traits will be calculated from environmental, genotypic and total variance derived from the ANOVA of results. Inter-allelic interactions, or epistasis, will be estimated by a two-way ANOVA. Earliness measurements will be taken at 7 months on three plants per plot, and also in controlled green house experiments for components of earlines such as, initiation of dry matter accumulation, rate of dry matter accumulation and completion of bulking as described in the original proposal.

Marker inheritance and fidelity studies

Comportment and inheritance of markers linked to tentative QTLs associated with post-harvest deterioration, earliness, starch content and disease resistance would next be examined in a different genetic backgounds to propose a model inheritance of key QTLs involved in genetic control of the agronomic traits. An ideal population for marker fidelity studies is a BC1 half-sib population developed from crossing five individuals of the F_1 mapping population to the female parent in a reciprocal fashion.

Development of improved populations

Once the principal QTLs, stable over environments, years, and genetic backgrounds have been identified, they will be employed in marker assisted development of populations that contain higher levels of starch, post-harvest deterioration, earliness, CBB disease resistance, through selection of parents, with greatest breeding value for the listed traits. The strategy includes choosing the best individuals of the F_1 cross, for all the traits combined, as parents of a breeding population, using the computer software, PGRI (Liu, 1998), which incorporates breeding and marker data. Results from marker-aided selection will be compared to phenotypic selection. Marker-assisted selection of parents will also be extended to other elite breeding lines and particularly outstanding land races, for different agro-ecologies, based on results obtained with the F_1 cross.

Genetic Mapping of SSR and EST markers

More than 1200 putative SSR markers have been identified from two libraries enriched for SSR markers (Fregene et al. unpublished data). They are being sequenced, and primer pairs are being designed to yield three groups, corresponding to 100-150bp, 150-200bp, and 200-250 bp PCR product sizes, to facilitate multiplexing of the primer pairs. SSR markers found to be polymorphic between the parents of the cassava map population will be scored in the 150 progeny by florescent primer multiplexing, 2-3 primer pairs, on an ABI377 automated sequencer located at the Biotechnology Research Unit (CIAT). Data will be automatically collated and linkage analysis will be by MAPMAKER 3.0 (Lander et. al. 1987). More than 250 transcript derived fragments (TDFs) have been found to be polymorphic between the parents of the map cross. They were obtained by applying the AFLP techniques to cDNA libaries obtained from the two parents, and represent potential ESTs. About 40 TDFs have been sequenced and are ready for primer design. The rest will be sequenced and primer pairs also generated. Mapping of the TDFs will be by the single strand conformation polymorphisms (SSCP) as described by Slaybaugh et. al. (1997).

Fine Mapping and Contig Mapping of disease resistance genes

Constructing a contig of BAC clones across a region, bearing resistance genes, requires an identification and fine mapping of such regions. Genetic mapping of resistance to the cassava bacterial blight (CBB) has already identified regions of the genome carrying resistance genes to 5 strains of the pathogen. Gene tagging of ACMD and white fly resistance genes has also been initiated and is expected to yield results shortly. The most efficient way of fine mapping in cassava is by employing a variant of the bulk segregant analysis and AFLP markers. The method takes advantage of the unique ability of the AFLP technique to sample many loci spread all over the genome using large number of primer combinations, and the ability of finding additional linked markers in any region by screening bulks of genotypic classes of markers adjoining that region. A backcross population (BC1), generated by crossing 5 of the F_1s from the mapping population to TMS30572, the resistant parent, for ACMD and CBB, is available, in addition to the F_1 mapping population, for genetic mapping and fine mapping. The half-sib BC1

population has a total of 270 genotypes and is already being genotyped for molecular markers from the frame-work map of cassava in preparation for gene tagging.

Technology transfer of markers to African NARs

Transfer of marker technology, as it relates to cassava breeding, is a key component of this project. It will be conducted through long term, Ph.D. fellowships, and short term trainings, visits of collaborators from NARs to CIAT. An African Ph.D. student, Mr Emmanuel Okogbenin joined the mapping team at CIAT recently, with support from the Rockefeller foundation. Mr Okogbenin is a young breeder and is working on the QTL mapping. A series of short term visits, by collaborators from NARs in Africa, has been planned under the auspices of the proposed cassava molecular diversity network, to be based at CIAT. Training on the use of SSR markers in cassava germplasm evaluation and use will be the principal activity of the short term visits. The international nature of cassava germplasm and its usage makes complementary collaborative efforts indispensable to achieving a useful characterization of genetic diversity of cassava. The proposed network aims to assess genetic variation of cassava in Africa compared to the total available variation in cassava and wild progenitors from South America, and to determine underlying genetic factors of successful cultivars and how to complement them with favorable alleles from other cultivars and wild progenitors. The Network will carry out surveys in sub-Saharan Africa, with National program collaborators, initially in Uganda and Nigeria, starting 1999, to collect cassava varieties grown by farmers, anthropological and agro-ecological information. DNA will be isolated from these samples and brought over to the CIAT for genotyping with genome-wide SSR markers.

Expected Outputs:

- Identification of QTLs and genes controlling earliness, dry matter content, post-harvest deterioration (PHD), starch quality/content, culinary quality, and other important morphological traits.
- Tools for the development of cassava populations with enhanced keeping quality, starch content/quality, earliness, and CBB disease resistance.
- Saturation of the genetic map of cassava with a few hundred SSR and EST markers.

- Isolation of candidate BAC clones bearing disease and pest resistance genes.
- Capacity building of African National Program scientist.

Capacity of Executing Institution

CIAT's Biotechnology Research Unit:

Facilities and trained personnel for PCR-based marker development, QTL analysis and availability of an existing genetic map of cassava. CIAT also has expertise, and resources in databases for storing and analysing molecular marker data.

Time Frame:

1999

- Mapping of SSR, and EST markers in the F_1 mapping populations.
- Second year phenotypic evaluation of QTL mapping populations in two different
- Crosses between best F_1 individuals for population development.
- Rapid propagation of BC1 QTL mapping populations.
- Training visit by Ugandan and Nigerian national scientist(s), under the auspices of the proposed cassava molecular diversity network.

2000

- Mapping of available PCR-based markers in the F_1 mapping populations (contd).
- Phenotypic evaluation of BC1 QTL mapping populations in three different sites for the second year.
- Establishment of populations from F_1 full sib cross.
- QTL analysis.
- Training visit by Zairean and Tanzanian national scientist(s), under the auspices of the proposed cassava molecular diversity network.

Budget (US dollars)

CENTRO INTERNACIONAL DE AGRICULTURA TROPICAL - CIAT

Project: Saturation of the Genetic Map of Cassava with PCR-Based Markers and the Use of the Genetic Map in the Improvement of Cassava

PROPOSED BUDGET TO THE ROCKEFELLER FOUNDATION FOR TWO YEARS EXTENSION

	(in U.S. dollars)		
Item	**1999**	**2000**	**Total**
Personnel			
Associate Scientist	55,000	55,000	110, 000
Technician	12,000	12,000	24, 000
Supplies and services	15,000	15,000	30,000
Indirect costs (6%)	5,000	5,000	10,000
T O T A L	**85,000**	**85,000**	**170,000**

References:

Chavarriaga-Aguirre P., Bonierbale M., Fregene M., Kresovich S., Tohme J., and Kochert G. (1998). Microsatellites in cassava (*Manihot esculenta* Crantz): characterization, inheritance and variability. Theor Appl Genet (in press).

Chavarriaga-Aguirre P., Maya M., Tohme J., Duque M., Iglesias C., Bonierbale M., Kresovich S., Kochert G. 1998b. 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. (in preparation).

Fregene, M.A., F. Angel, R. Gómez, F. Rodríguez, W. Roca, J. Tohme, and M. Bonierbale (1997). A molecular genetic map of cassava (*Manihot esculenta* Crantz). Theor. and Appl. Genet. TAG 95 (3) 431-441.

Fregene M. A., Gallego G., Lopez J., Wing R., Roca W., Tohme J. 1998. Construction of a Cassava (*Manihot esculenta* Crantz) Bacterial Artificial Chromosome (BAC) Library: Towards Cloning of Disease and Pest Resistance Genes. Red Bio, Havana Cuba June 1998.

Fresco L.O. 1993. The dynamics of cassava in Africa: An outline of research issues. COSCA working paper No. 9 Collaborative study of cassava in Africa. International Institute of Tropical Agriculture, Ibadan .

Mkumbira J., Lagercrantz U., Rosling H., Gullberg U. 1998. SSR evaluation of bitter and sweet cassava varieties from Malawi (in preparation)

Muller-Dos Santos G., Roa C., Tohme J. 1 998. RAPD, SSR, and AFLP assessment of Brazilian cassava land races (in preparation).

Muyinkwa T.R.I., Chipangura, Salehuzzaman S.N.I.M., Jacobsen E., and, Visser R. F. G. 1996. Cloning and characterization cassava genes involved in starch biosynthesis. African Journal of Root and Tuber crops, Vol 1, special edition.

Kresovich S., Szewc-McFadden A. K. Bliek S. M., McFerson J. R. 1995. Abundance and characterization of simple sequence repeats (SSRs) isolated from a size-fractionated genomic library of Brassica napus L. 9rapeseed). Theor Appl Genet 91:206-211.

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER : An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181

Liu B. 1995. PGRI. In: Statistical Genomics: Linkage, Mapping and QTL analysis. 1998. CRC Press, Boca Raton, FL 33431-9868.

Roder M. S., Plaschke J., Konig S. U., Borner A., Sorrells M. E., Tanksley S. D., Ganal M. W. 1995. Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 246: 327-333

Tanksley, S.D., N.D. Young, A.H. Paterson & M.W. Bonierdale. 1989. RFLP mapping in plant breeding: New tools for an old science. A review. Biotech. 7:257-264.

Saghai Maroof M.A., R.M. Biyashev, G.P. Yang, Q. Zhang, R.W. Allard 1994. Extraordinary polymorphic microsatellite DNA in barley: Species diversity,

chromosomal locations, and population dynamics. Proc. Natl Acad Sci. (USA) 91:5446-5470

Slaybaugh, M.B., Huestis G.M., Leonard J., Holloway J., Rosato C., Hongtrakul, V. Martini N., Toffeler R., Vortz M., Shell J., and Knapp S. J. 1997. Sequence-based genetic markers for genes and gene families: single strand conformation polymorphisims for the fatty acid synthesis genes of *Cuphea*. Theor Appl Genet 94: 400-408.