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 × 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 *Xho*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 $8 \times 10^5$ 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 carrried 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 homoeologous 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 $\geqslant 2.0$. Markers that could not be placed with LOD $\geqslant 2.0$ 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

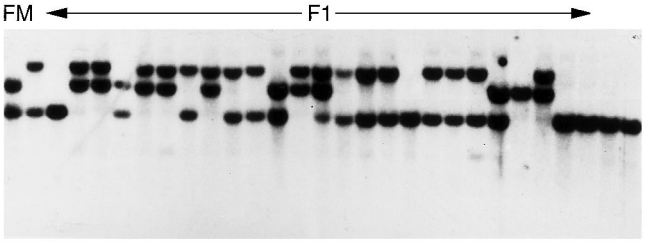



**Fig. 1** Southern hybridization showing segregation of heterozygous unique alleles from both parents (allelic bridges). *F* Female parent, *M* male parent

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 *Hin*dIII (Fig. 2). The majority of sequences detected

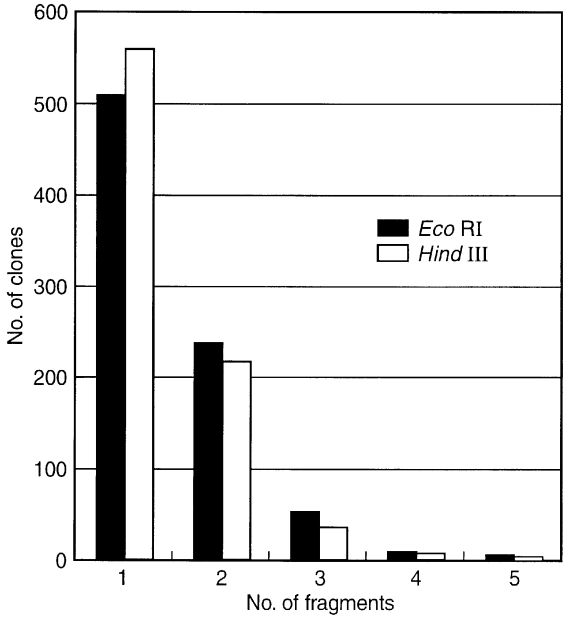



**Fig. 2** Number of DNA low-copy sequences detecting 1, 2, 3, 4, and 5 fragments with *Eco*RI nad *Hin*dIII restriction enzymes

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.

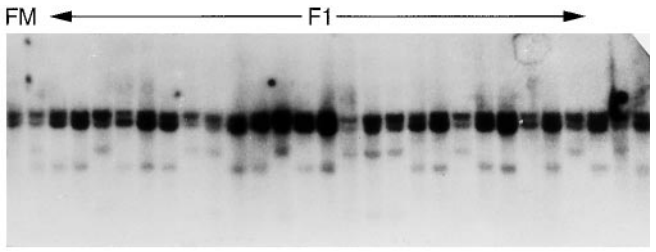



**Fig. 3** Southern hybridization of clone GY92 with *Hin*dIII digests of parental line and mapping population. *F* Female parent, *M* male parent. Two faint fragments can be observed to be present and polymorphic in the male but absent in the female parent

## 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

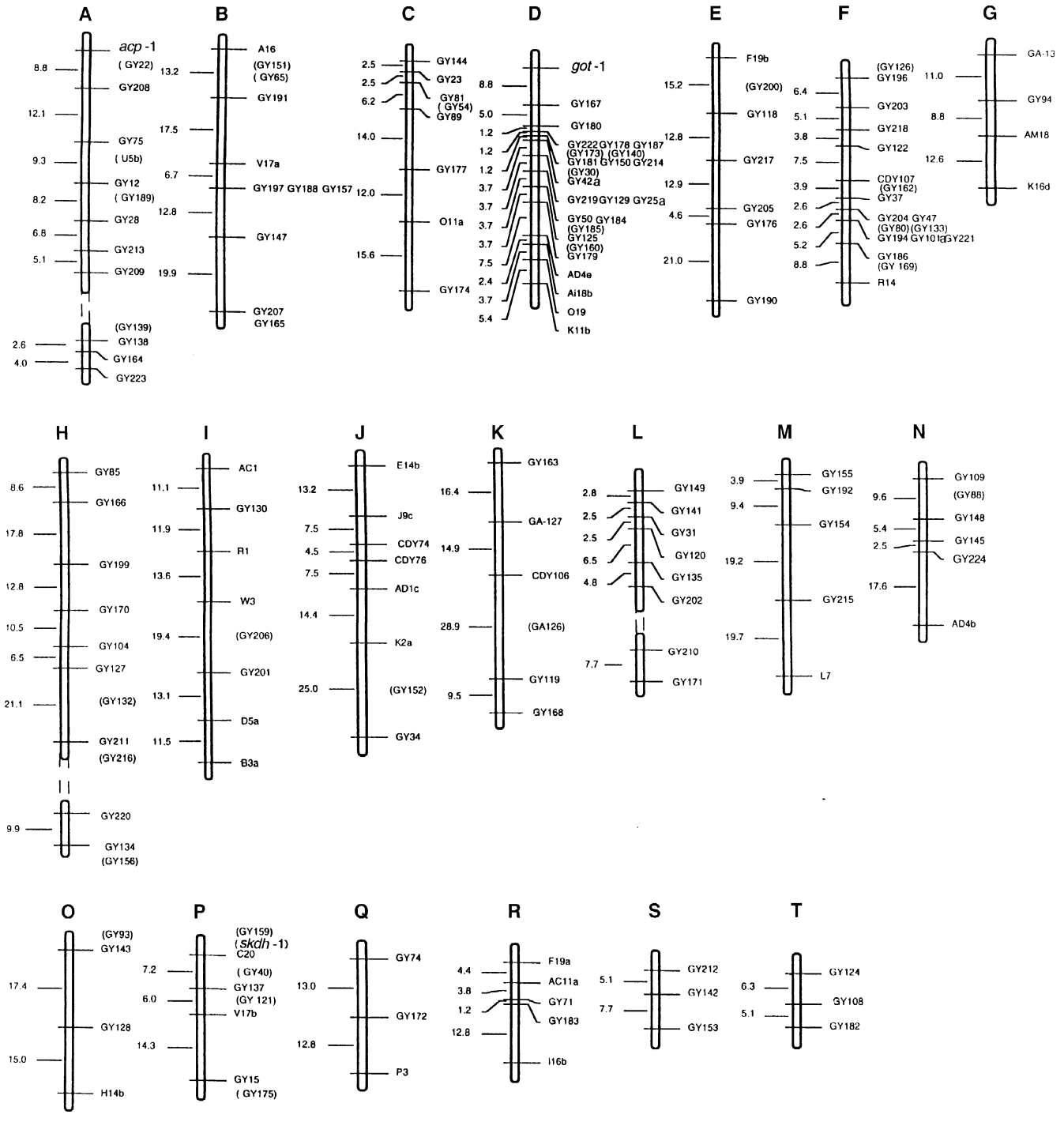
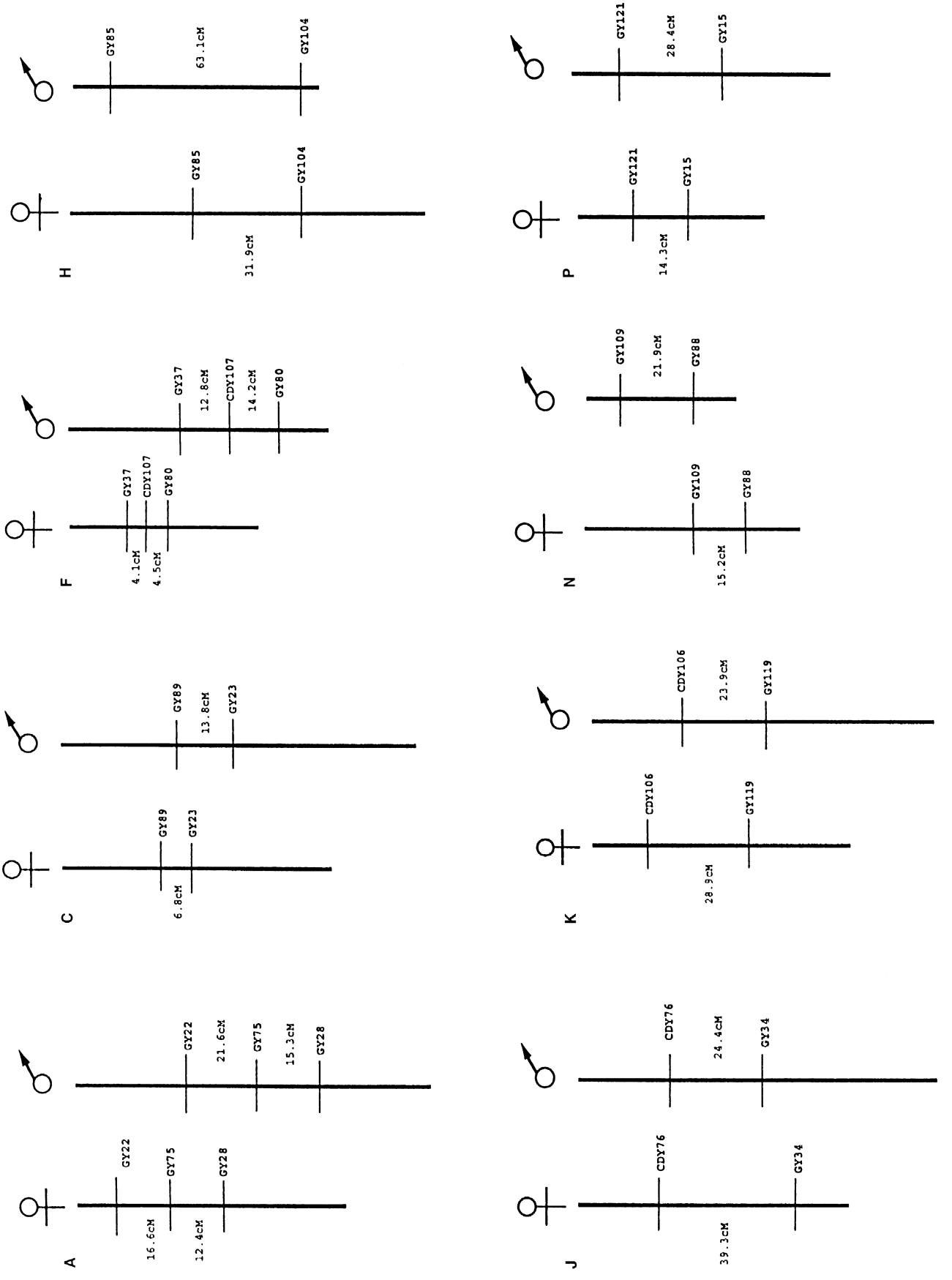



**Fig. 4** Molecular genetic map of cassava based on segregation of RFLP (*CDY* cDNA, *GY* genomic), microsatellite (*GA*), isozyme (*acp*, *skdh*, and *got*), and RAPD markers (prefix *A-AP* through *Z*, Operon primer nomenclature) in gametes of the female parent (TMS30575), in a $F_1$ cross (with CM2177-2 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
GY80
4.1cM
4.5cM
GY37
12.8cM
CDY107
14.2cM
GY80
H
GY85
GY104
31.9cM
GY85
63.1cM
GY104
J
CDY76
GY34
39.3cM
CDY76
24.4cM
GY34
K
CDY106
GY119
28.9cM
CDY106
23.9cM
GY119
N
GY109
GY88
15.2cM
GY109
21.9cM
GY88
P
GY121
GY15
14.3cM
GY121
28.4cM
GY15

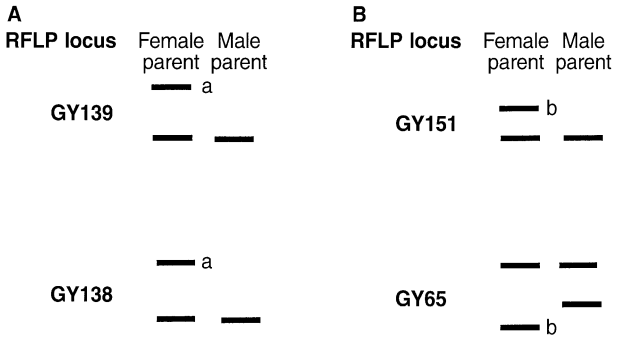



**Fig. 6A, B** Examples of the category of markers considered in identifying linkage in repulsion, with one parent heterozygous (**A**) or both parents heterozygous (**B**). In both examples, unique alleles (*a* and *b*) from the female parent segregated in repulsion in the $F_1$ progeny

## 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 culitvars 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

Lagercrantz 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