Genome Mapping and Molecular Breeding in Cassava

Y .Lokko[1], E. Okogbenin[1], C.Mba[2], A. Dixon[1], A. Raji[1] and M. Fregene[3]

1. International Institute of Tropical Agriculture (IITA), Oyo Road PMB 5320, Ibadan, Nigeria. www.iita.org

2. Plant Breeding Unit, Joint FAO/IAEA Agriculture and Biotechnology Laboratory, International Atomic Energy Agency Wagramer Strasse 5, P.O. Box 100, A-1400 Vienna AUSTRIA www.iaea.org

3. Centro Internacional de Agricultura Tropical, Apartado Aéreo 6713, Cali Columbia, www.ciat.cgiar.org

Formatted: Spanish (Spain-Modern Sort)

I. Introduction

Brief History of Crop

Cassava (*Manihot esculenta* Crantz) belongs to the family Euphorbiaceae, which are characterized by lactiferous vessels composed of secretory cells. Its relatives in the euphorbiaceae family include several commercially important plants, such as rubber trees (*Hevea brasiliensis*), castor oil plants (*Ricinus comunis*) and ornamental plants (*Euphorbia* spp.). Cassava it's believed to have originated by hybridization between two wild *Manihot* species, followed by vegetative reproduction of the hybrid.

The center of origin of cassava was first reported to be Central America including Colombia, Venezuela, Guatemala and Southern Mexico due to the large number of varieties present there (Sauer (1952; Roger 1965). This was later referred to as the "minor" center of origin, and Brazil as the "major" center of origin (Ekanayake et al 1997). Furthermore, *Manihot* species found in the Central America region are only distantly related to cassava (Fregene et. al. 1994;Schaal et. al. 1994; Roa et. al. 1997). Schaal and Oslen, (1999), investigated the crop's domestication based on haplotypes of the single-copy nuclear gene glyceraldehyde 3-phosphate dehydrogenase (*G3pdh*), in cassava and its relatives, which confirmed that cassava was likely domesticated from wild *M. esculenta*, mainly from populations of subspecies *flabellifolia* along the southern border of the Amazon basin.

Cassava grows in tropical and subtropics and was introduced into Africa and Asia by Portuguese travelers in the 15th century (Jennings and Hershey, 1985; Allem, 1994).

Botanical Descriptors

Cassava, *Manihot esculenta* Crantz, is a perennial woody shrub with edible tuberous root. It is also called yuca, (Spanish), manioc (French), and mandioca (Portugues). The height of a mature cassava plant usually ranges from 1 to 2 m, although some cultivars reach 4 m. Branching height can be as low as 20 cm, while some varieties never branch and, as a result, never flower.

Cassava leaves are simple, with 3 and 9 leaf lobes (usually odd numbers) arranged spirally around the stem. It is monoecious: male and female flowers are located on the same plant. Some varieties flower frequently and regularly, while others flower rarely or not at all. Flower production is important for breeding. Environmental factors, such as photoperiod and temperature, influence flowering. Male flowers develop near the tip, while female flowers develop closer to the base of the inflorescence (Ekanayake et al 1997).

Cassava is generally propagated with stem cuttings, thereby maintaining a genotype. Under natural conditions as well as in plant breeding, propagation by seed is common and farmers in African are known to occasional use of spontaneous seedlings for subsequent planting (Silvestre and Arraudeaus, 1988).

Cassava plant propagated from stem cuttings produces adventitious roots at the base of the cuttings within a week, while Cassava plants propagated from seeds first develop a tap root system. The adventitious roots develop into a fibrous root system which within 30-60 days increase in diameter and become tuberous roots (also called 'storage roots'). These swell with time due to starch accumulation (Ekanayake et al 1997).

Cassava generally has a diploid genome (2n=36). However, some authors have described it as a segmental allotetraploid with basic chromosome number x=9. Jos and Nair, (1979) however, conducted studies on the meiotic behaviour of several cassava genotypes and observed regular 18 bivalent formation of the chromosomes typical of its diploid (2n=2x=36 chromosomes). Furthermore, the breeding behaviour of allotetraploids is similar to that of diploid (Wricke and Weber, 1986). Although most cassava genotypes studied are diploid, spontaneous polyploid such as triploids (3n) and tetraploids (4n) of some genotypes have been reported (IITA 1980; Hahn *et al*., 1980). Triploid and tetraploid plants differ from diploid plants in plant vigor and leaf shape and size. Triploid plants usually grow and yield better than tetraploid and diploid plants. The nucleic acid content of diploid cassava is 1.67 picogram per nucleus that is 772 mega base pairs in the haploid genome (Awoleye *et al*., 1994). The heterozygosity levels of different cassava accession have been further confirmed from diversity studies using isozymes (Lefevere and Charrier, 1993), AFLP and SSRs (Fregene et al 2000; Fregene et al 2003; Sanchez, et al 1999).

Economic Importance

Cassava accounts for approximately one-third of the total staple food production and provides over 50% of the energy for more than 200 million people in sub-Saharan Africa (IITA 1992). In Central Africa alone, cassava is estimated to provide over 1000 Kcal per day to 30 million people (Cock 1985). The storage roots form the basic carbohydrate component of the diet and the leaves, which contain appreciable amounts of vitamins, minerals and proteins are consumed as a preferred green vegetable in many parts of Africa, providing protein, mineral and vitamins (Hahn 1989; Lancaster and Brooks, 1983).

The crop is grown almost exclusively as food in 39 African countries stretching through a wide belt from the Island of Madagascar in the southeast to Senegal and the

Cape Verde Islands in the north-west (Hahn and Keyser 1985). Production is in areas where the rainfall exceeds 600 mm over a period of at least 2 to 3 months and altitudes range from sea level to 1800m. It is particularly important in those areas where food supply is constantly threatened by environmental constraints such as drought and pest outbreaks because of its ability to grow under conditions considered as sub optimal for the majority of food crops. It can be harvested anytime from 6 to 24 months after planting, and can be left in the ground as a food reserve for household food security in times of famine, drought and war (FAO, 2000; Best and Henry 1992; Cock, 1985).

Global cassava production projected for 2004 is expected to be around 192 million tonnes as was recorded in 2003, with Africa contributing 103 million (FAO, 2004).

Classical breeding Objectives

Breeding for increased yield, multiple pest and disease resistance, desirable agronomic and consumer preference traits such as early vigor in plant growth (for high foliage yield for leaf vegetable), appropriate plant architecture, and early bulking of storage roots, combined with high dry matter content, low cyanide content, and other favorable traits, e.g., easy peeling, have been the main breeding objectives. Recently breeding for improved micronutrient content is being emphasized. Broad-based populations for different agroecologies are developed through recurrent selection (intra and interspecific population improvement techniques) and backcrossing as well as multiple crossing schemes to achieve. Superior individuals are selected as parents, segregating families are generated by multiple crossing among these elite clones complementing one another for various agronomic, consumer quality and major pest resistance traits. The breeding values of the parents are evaluated through progeny testing in seedling nurseries. Based on evaluations, selected parental clones or half sib progenies are hybridized for further improvement in a recurrent selection scheme. Backcrossing has also been a useful procedure for the transfer of resistance into elite populations by providing resistant lines quickly to prevent the severe infestation of relevant pests.

Classical breeding Achievements

Achievement in cassava breeding in Africa is in the development of a range of elite genotypes , such as TMS 30572 and TMS 4(2)1425that combine high stable yields, agronomy and consumer quality with acceptable levels of resistance to CMD and CBB which are widely cultivated In Africa (Hahn et al 1989). The adoption of some of these multiple pest and disease resistance genotypes in Uganda for instance,

curtailed the devastating effect of the CMD on cassava production in that country (Otim Nape et al., 1994).

The introduction of Latin American gemplasm into the breeding programs in Africa has resulted in significant broadening of the genetic base of cassava in Africa.

Classical breeding has contributed to improved dry matter and starch content as well as reducing the cyanogenic potential in cassava. (could you fill in with some eg from CIAT). An IFAD-financed research project executed by the National Center for Research on Cassava and Tropical Fruits (CNMPF, the Brazilian acronym), Cruz das Almas, CIAT and IITA beginning 1991 led to the collection and evaluation at representative semi-arid sites of cassava germplasm in Northeast Brazil. Genotypes with high efficiency for extraction of limited soil water, low levels of water loss through transpiration, and resistance to severe mites attack were selected and recombined by genetic crosses to form the basis of a breeding population for the semi-arid agroecosystem (CIAT 1996). Subsequently, a farmer participatory plant breeding (PPB) scheme was employed to evaluate recombinant progenies and select several varieties adapted to the semi-arid region of Northeastern Brazil (Fukuda and Saad 2001). Improved cassava varieties developed in the project had 25-100% yield increases over traditional cultivars in 41 communities of the states of Bahia and Pernambuco of Northeast Brazil (CIAT and CNMPF 1999).

Recombinant seeds from the above breeding population for semi-arid ecosystems were also introduced to Sub Saharan Africa (hereafter Africa) through IITA and evaluated in the drier areas of Northern Nigeria. Evaluations revealed the potential of the germplasm to increase productivity in the dry and hot conditions of Northern Nigeria (IITA 1999). However, the presence of a major biotic constraint the cassava mosaic disease (CMD) in Africa, a disease that is not found in Latin America, limited immediate use of the germplasm but required introgression of CMD resistance into the Latin American germplasm. Resistance to CMD was introduced by backcrosses and thousands of recombinant seeds were produced and distributed to participating countries in West and Central Africa. Evaluation under typical farmer conditions and selection has lead to the selection of 25, 32, 10, 28 and 122 improved genotypes that combine CMD resistance and adaptation to water stress for the semi-arid areas of Burkina Faso, Chad, Ghana, Niger and Nigeria respectively (IITA 2000a). The improved germplasm has extended considerably the range of cultivation of cassava beyond its traditional area in the humid and sub-humid tropics into the semi-arid zone of West and Central Africa by more than 100,000 hectares between 1989 and 1999 (IITA 2000b).

Limitations of classical breeding and utility of molecular mapping

Most traits studied in cassava are polygenic (Bryne, 1987; Rajendran, 1989; Amma and Sheela 1995; Hahn *et al*. 1989). Variation in polygenic traits is attributed to quantitative trait loci, QTL's. Quantitave traits in plants and animals are studied using a variety of genetic models and designs incluing the analysis of mating designs in segregating populations, estimate effective factors using biometrical estimating techniques or molecular markers or biometrical methods can be used to estimate effective factors for quantitative traits (Lynch and Walsh, 1998; Zeng, *et al*., 1990). Most of the biometrical estimating techniques do not estimate the exact number of genes but rather the number

of effective factors. Furthermore the detection of multiple genes for a trait using segregation analysis alone is not efficient because of the differences due to genotype by environment interaction. Molecular markers, which are not affected by environmental conditions and are insensitive to gene interactions, are suitable for such studies.

The two main strategies used to identify molecular markers associated with traits of interest are genetic linkage mapping and bulk segregant analysis (BSA) (Tanksley *et al.*, 1989; Michelmore *et al.*, 1991; Giovannoni *et al.*, 1991).

Molecular markers, which include biochemical (isozymes and storage proteins) and DNA markers, exist in every genotype. They occur in large numbers and their expression is independent of phenotypic value and as such are powerful tools of genetic research (Soller and Beckmann, 1986; Beckmann and Soller, 1983). They allow geneticist and plant breeders to locate and follow the numerous interacting genes that determine a complex trait as well as tagging those controlled by single genes. Molecular markers offer a spectacular improvement in the efficiency and sophistication of plant breeding. They are i) developmentally stable, ii) detectable in all tissues, iii) unaffected by environmental conditions, and iv) virtually insensitive to epistatic or pleiotropic effects. They provide a choice of codominant markers which can be identified in heterozygotes or dominant marker which are identified as present or absent subclasses (Botstein *et al*., 1980; Williams *et al*., 1990). Molecular markers can provide. It is now generally accepted that markers represent the most significant advance in breeding technology that has occurred in the last few decades and are currently the most important application of molecular biology to plant breeding.

Molecular markers have made an immense contribution to cassava breeding and genetics, in the assessment of genetic diversity, taxonomical studies, understanding the phylogenetic relationships in the genus, confirmation of ploidy and in the development of genetic maps (Fregene *et al*., 1997; Fregene 2001; Lokko et al 2004).

II. Construction of genetic maps:

Brief history of mapping efforts

Genetic mapping uses genetic techniques such as crossbreeding and pedigree analysis to construct genomic maps. Genetic linkage maps provide a direct method for selecting genes via their easily detectable markers, a guide for sequencing experiments by showing the positions of genes and other sequence features. Linkage maps are based on recombination frequencies and the distance between points on a genetic map is a reflection of the recombination frequencies between the points.

The first genetic maps constructed in an organism was in the fruit fly Drosophilia which used genes which specified distinct phenotypes such as eye colour and wing shape as markers (Sturtevant, 1931). In plants genetic maps of some crops species such as maize and tomato were also among the earliest genetic linkage maps to be constructed (Emerson et al 1935: MacAurther, 1934). These were based on morphorlogical characters such as dwarfism, albinism and alter leaf morphology.

The first DNA marker based genetic linkage maps in plants were Brassica were published in the late 1998's to the early 1990 (Table1). These were mainly based on RFLP markers, which led to the rapid construction of linkage maps in many plant species and the placement of genes that control qualitative traits on these maps. In cassava RFLP markers were also the first set of DNA markers to be used in mapping.

Table1 below

First generation maps constructed in cassava

Angel *et al*. (1993), initiated work on genome mapping in cassava with an investigation on polymorphism in a range of cassava accessions and a wild relative with different random genomic clones and restriction enzymes. They concluded that a combined use of RFLP and RAPDs markers would lead to the construction of detailed map of cassava and Gomez *et al* (1996) subjected 328 RAPD markers to linkage analysis in cassava. Following this, the first cassava linkage map to be developed was published by Fregene et al, (1997).

The map was based on an F_1 population of two geographic divergent parents. The female parent TMS I30572, with resistance to CMD, which was derived through introgression from M. glazovii while, the South American male, parent CM 2177-2 (ICA-Cebucan) with no resistance to CMD. A hundred of 150 individuals were available in the cross however a subset of 90 individuals were used to develop the first map.

In a heterozygous species, the segregating F_1 is used (Lui, 1998; Pillay and Kenny, 1996; Fregene *et al*., 1997). They scored single-dose restriction fragments (SDRFs), for the linkage analysis and developed two separate linkage maps based on male and female parents. 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 and are suitable for linkage analysis in an F_1 population when there is the presence of a number of unique segregating polymorphisms (heterozygosity) and normal meiosis in either or both parents in mapping polyploid genomes (Wu *et al.*, 1992; Williams, 1998).

One hundred and fifty-eight RFLP, 30 RAPD, 3 microsatellite, and 4 isozyme single-dose markers, donated by the female parent of mapping population, were tested for linkage using the MAPMAKER computer package (Lander et al. 1987). 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 with LOD value ≥2.0 were assigned to the framework map. Markers that could not be placed with LOD≥2.0 were added to the map in the most likely interval between framework markers. One hundred and thirty-two RFLP, the 30 RAPD and 3 microsatellite, and 3 isozyme loci defined 20 linkage groups spanning 931.6 cM, with an average marker density of 1 marker every 8 cM (Fig. 1); 26 RFLP markers and 1 isozyme marker remained unlinked. The linkage groups were 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≥2.0).

They estimated the length of the cassava genome based on the female map to be 1, 610 cM, covering about 60% of the cassava genome. This estimate was based on 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 (Hulbert et al. 1988).

The male derived map consisted of 107 RFLP, 50 RAPD, 1 microsatellite, and 1 isozyme single-dose markers, in 24 linkage groups with a total distance of 1, 220 cM. The genome length was estimated 2, 010 cM. Intervals were observed to be larger in the male-derived map than in the female-derived map and a paired *t*-test of showed significantly (*P*(0.01) greater distances in the male-derived map, suggesting a reduced recombination rate in gametes of the female parent. The mean interval length between adjacent allelic bridges (markers common to both parents) in the female-derived map was 38% less than in the male derived map.

Second generation maps constructed in cassava

In an attempt to increase the marker density on the map the next strategy was to develop simple-sequence repeat (SSR) markers, markers that are polymerase chain

reaction (PCR)-based and highly polymorphic and map them on to the exiting map and new maps to be developed. Mba *et al*., 2001, developed and characterised 172 SSR markers to saturate the existing linkage map. The markers were screening in 150 progenies of the mapping population TMS I30572 X CM 2177-2. Thirty six markers were placed on the map and were evenly spread over the linkages groups. A higher number of "allelic bridges" which are required for rigorous marker-assisted quantitative genetic analysis in F1 progeny from non-inbred parents were obtained using the SSR markers (30%, as against 10% obtained with RFLPs). Currently, 80 RAPD markers 239 RFLP markers and six EST's have been placed on the cassava genetic map (Fregene *et al*., 2001).

Map Coverage Now?????

The total number of RFLP, RAPD, and SSR markers on the cassava map is 830 markers, SSR markers make up roughly 25% of markers on the map. Genetic mapping in a full-sib cross from non-inbred parents involves linkage analysis of markers segregating independently in the female and male gametes and leads to 2 independent maps (Fig???). Analogous groups can be identified amongst the female- and male-derived linkage groups via markers that are unique in both maps, also known as allelic bridges". For the first time, 18 analogous linkage groups that may represent the 18 haploid chromosomes of cassava have been identified. Previous mapping efforts with only RFLP markers failed to resolve this problem due to insufficient allelic bridge markers 27% of all markers compared to 30% with SSR markers. Efforts are currently ongoing to develop a single map using the JOIN MAP (Stam 1993) program. Nevertheless, the linkage map is not saturated as 22 and 40 markers remain unlinked as well as the presence of 2 non-analogous linkage groups, one from each parent.

With over 1000 cassava SSR primers, developed (Chavariagga et. al. 1998; Mba et. al. 2001; CIAT 2003; Lopez and Verdier 2004, unpublished data), and the suitability of AFLP marker technology to non-radioactive detection, efforts were made to develop linkage maps of cassava from diverse genetic backgrounds based on simple, PCR based techniques. Three genetic linkage mapping studies were initiated at IITA specifically to map the genes conferring resistance to cassava mosaic disease (CMD), develop molecular markers for CMD resistance in African cassava germplasm and to

Deleted: 300

assist the incorporation and enhancement of resistance to this disease in African, Latin American and Asian cassava gene pools (IITA, XXX; Akano et al 1998). The F_1 populations were generated from crosses between the CMD resistant breeders line TMS I30572 and a Nigerian landrace TME117 (local name Isunikankiyan); crosses between the breeders line TMSI30555 with moderate susceptibility to CMD and CMD resistant landraces TME3 (2^{nd} Agric) and TME7 (Oko Iyawo) (Akano et al 1998; Lokko, et al., 2001;).

One hundred and thirty two genotypes each of populations TMS I30572 X TME117 and TMS I30555 X TME3 were used for linkage map construction. A total of 125 markers from 85 SSR primers were generated in for were subjected to linkage analysis in the populations TMS I30572 X TME117. Sixty-two markers mapped in 19 linkage groups (LGs), which covered 816.6 cM of the genome.

In the population TMS I30555 X TME3, 66 markers from 47 SSR primers and 85 AFLP markers were subjected to linkage analysis also mapped in 19 LGs (Fig 3). Thirty-three out of the 85 AFLP's markers and 37 out of the 66 SSR's, mapped in 19 LGs using a LOD of 3.0 and recombination fraction of 0.17 to declare linkage. The map covered 788 cM. Generally in population C, the SSR and AFLP markers clustered in specific groups, nine groups were made up of ALFP markers only, two groups had both AFLP and SSR markers and the remaining eight were made up of SSR markers only.

Sixty-nine F_1 s of TMS I30555 x TME7 were screened with 15 SSR primers and 11 AFLP markers and the data subjected to linkage analysis with Join Map (Van Ooijen et al 2001). Thirty-six <aaxab> and 41 <abxaa> markers were scored from AFLP markers, while five <aaxab>, three <abxaa> and five <abxac> markers were scored for the SSR analysis giving a total of 90 markers. The chi square analysis revealed that 34 (37.78%) <aaxab> markers , 37 (41.11%) markers and five (5.55%) <abxac> markers were significant for the1:1 segregation ratio while 10 (11.11%) <abxaa> markers and 4 (4.44%) <aaxab> markers were significant for 3:1 segregation ratio. Forty-five out of the 90 markers mapped in 15 linkage groups with a total length of 812.32 cM of the genome.

Despite the low saturation of the maps, the loci were randomly distributed over linkage groups and the information from it and the map itself can be utilised in cassava genetics. For instances, the markers could also be applied in studies with similar cassava populations. AFLP primer pair E-ACT/M-CAG which contributed markers in four linkage groups in the TMS I30555 X TME7 map, could be used in experiments such as in diversity studies, which require

coverage of most of the genome. The maps could also be used in linkage map pooling with the TMS I30572 X CM 2177-2 map.

[Representative maps in figures??]

III Gene Mapping

The cassava genetic linkage map developed at CIAT currently has five known genes including two CMD resistance genes *CMD1* and *CMD2* placed on it and a number of quantitative trait loci (QTL) associated with some linkage groups (Fregene *et al*., 2001). To date two CMD resistance genes *CMD1* and *CMD2* have been placed on the map (Akano; *et al.*, 2002; Fregene, 2001).

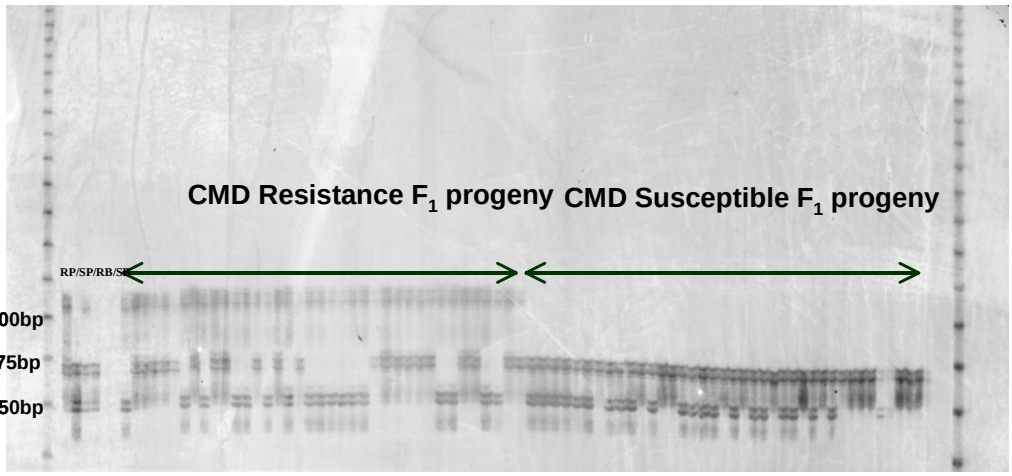
Bulk segregant analysis (BSA) was used to screen several SSR and AFLP markers in the mapping populations TMS I30555 X TME3 and TMS I30555 x TME7 and to map genes for resistance to CMD (Akano et al 2001; Lokko et al 2001). The populations were characterised in multiple environments and the CMD status of each genotype used to classify them into highly resistant, resistant susceptible and highly susceptible individuals. In TMS I30555 X TME3, the CMD resistant and CMD susceptible bulks were made up of 40 individuals. In TMS I30555 x TME7 10 individuals made up either DNA bulk. In the two studies, the SSR marker SSRY28 was tightly linked to CMD resistance (Fig XX).

Akano *et al*., (2002) identified a dominant cassava resistance gene *CMD2* on linkage group R of the male cassava framework map 8 cM from the marker SSRY28.

Marker-trait association detected by regression analysis of TMS I30555 x TME7 showed that the marker, which accounted for 57.41% of total phenotypic variation for resistance. The analysis further showed that another SSR marker, SSRY106 and AFLP marker E-ACC/M-CTC-225, accounted for 35.59% and 22.5% of the total phenotypic variation for resistance respectively. Correlation coefficients for SSRY28-180 and E-ACC/M-CTC-225 with CMD responses were significant and negative while correlation between SSRY106-270 and the CMD responses were significant and positive. This suggests that the CMD resistance genes associated with E-ACC/M-CTC-225 and SRY28-180 are different from genes associated with SSRY106-270. Interval mapping analysis further revealed the large QTL regions between SSRY28-180 and E-ACC/M-CTC-225 covering about 50% of the length of the linkage group and 3% of the length of genome covered, which explained most of the phenotypic variation.

Later, the group at CIAT identified NS158 to be tightly linked to CMD resistance.

Representative gel pictures??



IV QTLs detected

QTLs associated with dry matter yield and dry matter percentage have been linked to linkage group D of the cassava framework map (Fregene *et al*, 2001).

V Marker Asisted Breeding

Tools of genomics particularly molecular markers permit an immediate increase of efficiency in cassava breeding pending when inbred lines will be developed. Molecular markers are particularly useful for:

1. Using a large of progenitors, especially when no information on combining ability is available, in breeding schemes and the subsequent elimination of a large part of the resulting progeny very early in the breeding cycle.
2. Increasing the number of replications and environments for testing breeding lines (thus achieving a more accurate performance of the genotypes), by working with reduced progeny sizes.
3. Increasing heritability by the elimination of the confounding effect of the environment, this is particularly true in cassava where the first stage of selection is conducted on a single plant per genotype or low/absence of pathogen/pest pressure in initial sites of selection (nurseries), or a trait that has a large environmental influence.

Formatted: Bullets and Numbering

4. The pyramiding of different sources of resistance against a single pathogen or the combination of resistance to several pests and diseases and other traits without resort to time consuming field evaluations, also known as 'express breeding'.

The success of conventional cassava breeding is inversely proportional to the number of traits it attempts to improve. Combining many genes controlling quantitative traits from diverse sources into a single variety is a long-term, high-risk venture that requires every available tool for success. MAS can be used to achieve these goals more efficiently. For example, CMD-resistant donor parents can be crossed to other parents with excellent resistance to CGM, and markers used to select recombinants that combine resistance to CMD and CGM in a single generation, without the need for field trials. Resulting selections can then be crossed to other genotypes that carry , for example, high beta-carotene content to produce multi-trait hybrids, again without need for field evaluations. The best of these selections are then crossed to elite progenitors of the appropriate gene pool, to capture genes for yield and adaptation, and the resulting hybrids are selected with markers to eliminate those progenies that do not have resistance to CMD, CGM and high beta-carotene content, leaving a smaller number of progeny to be thoroughly evaluated in the regular breeding scheme.

Molecular marker-assisted selection for breeding resistance to CMD has been successfully implemented for introducing resistance into elite gene pools at CIAT (Fregene et al. 2004;CIAT, 2003,) and also to introgress resistance to CGM and CMD in Local cassava varieties in Tanzania (Kullaya et al. 2004). With the successful development of markers for resistance to CMD and CGM, efforts have turned to development of markers for another major source of loss to cassava production -- postharvest physiological deterioration (PPD). Dramatically delayed PPD was found in *Manihot walkerae*, a wild relative of cassava found in Mexico and USA (state of Texas) (Fregene and Mba 2004). An accession of *M. walkerae* (MWal 001) was crossed extensively to elite cassava varieties. A single successful genotype was found with delayed PPD. The storage roots of the hybrid remained intact a month after harvest. Backcrosses of this hybrid to elite progenitors of the CIAT cassava gene pools and selfed (S_1) populations were made for genetic mapping of the delayed PPD traits. Genetic

Formatted

mapping of the delayed PPD genes is progressing and following identification of genes involved in the regulation of PPD, MAS will be used to combine these genes with progenitors that already have combined CMD and CGM genes. CIAT and a number of partners have initiated a project to genetically fortify cassava with the inherent ability to produce higher levels of beta-carotene. This is one way of combating the deficiency of this key micronutrient in areas where cassava is a major staple. The experimental approach to increasing cassava beta-carotene content includes conventional breeding and genetic transformation. The discovery of a wide segregation pattern of root color in two S_1 families from the Colombian land race MCol 72 (cross code AM 273) and the Thai variety MTAI 8 (AM 320), was the basis for molecular genetic analysis of beta-carotene content in cassava (CIAT 2003). Three SSR markers were found to be associated with beta-carotene content (Fregene and Mba 2004). One of these markers explained 30% of phenotypic variation for beta-carotene content in the population used for this study. The homozygous state of certain alleles of these markers translates into higher beta-carotene content, suggesting that breeding for beta-carotene can benefit from molecular markers to assist in combining these favorable alleles in breeding populations. Few key traits in cassava hold greater potential for increasing cost-effectiveness via molecular marker assisted selection, than root dry matter content (DMC). This trait is usually measured at the end of the growth cycle. A number of genetic and environmental effects influence DMC. It is usually highest before the onset of the rains, but drops after the rains begin as the plant mobilizes starch from the roots for re-growth of leaves (Byrne 1984). Defoliation from pest and disease attacks can lower DMC. Breeding programs have been quite successful in improving DMC, especially for industrial markets.

The entry point for developing markers associated with DMC was three diallel experiments carried out from 2000 to 2002 (Jaramillo et al 2004). Diallels, in this case made up of 90 families, are an ideal method to identify genes controlling DMC that are useful in many genetic backgrounds. Estimates of general and specific combining ability (SCA and GCA, respectively) for many traits of agronomic interest were calculated, with an emphasis on DMC. Based on GCA estimates, parents were selected to generate larger sized

progenies for DMC mapping. Sizes of families in the original diallel experiment were about 30 progenies, a rather small size for genetic mapping. Parallel to the development of mapping populations was the search for markers associated with DMC using two F_1 families – GM 312 and GM 313 selected from the diallel experiment having parents with high GCA for DMC.

VI Map based cloning

High resolution mapping around the cassava genome region bearing a dominant CMD resistance gene, *CMD2*, has lead to the identification of 2 markers, RME1 and NS158 that flank the gene at 1 and 2 cM respectively (Moreno 2004, unpublished data). This year, positional cloning of *CMD2* made progress with the construction of a bacterial artificial chromosome (BAC) library and BAC contigs around *CMD2*. The BAC library construction was carried out at the Clemson University Genome Institute with the participation of CIAT via a graduate student, Ms Isabel Moreno. The BAC library was brought back to CIAT and BAC plate, column and plate pools were created. PCR amplification, using primers of the SSR marker NS158 and the SCAR marker RME1, of the 'BAC Pools' were employed to identify BAC clones containing the markers closest to *CMD2* for BAC contig construction. The BAC contigs will be the start-off point for chromosome walking to the gene of interest.

Plant material for BAC library construction was the African local variety TME3 that has resistance to CMD mediated by the dominant gene *CMD2*. Construction of the library was as described earlier by Tomkins et al. (1999a; 1999b; 2004). To estimate the distribution and average size of the clones, a total of 370 clones from the TME3 library were picked at random and grown overnight in 3ml of liquid LB medium + 12.5µg /µl Chlorampenicol. Plasmid DNA was isolated, digested with *Not* I restriction enzymes and inserts separated from the vector by pulsed-field electrophoresis.

Contig construction was by PCR amplification of 'BAC pools', namely 'plate pools' (PP), 'column pools' (CP) and 'row pools' (RP). Briefly, all 192 384-well plates were duplicated using a 384-pin replicator and allowed to grow in LB/ Cloramfenicol (12.5 ug/ml) medium at 37^{o} C over night. For BAC plate pools, all the bacteria culture in a 384-well plate was combined into an omnitray and 200ul of this transferred into a single well in a 96-well plate to yield 2 'BAC pool' plates. Simultaneously, every 10 plates of the library were inoculated into a single 384-well plate using a 384-pin replicator to give 20 384-well plates. Each row of each of the 20 plates was inoculated, using a sterile tooth pick, into a single well containing LB/ Cloramfenicol (12.5 ug/ml) medium to form 'Row plates (RP)' 5 RPs of 96-well plates in all (Figure 12.24). The same was done for each column of the 20 384-well plates

combined into a single well to form 4 96-well 'Column plates (CP)' (Figure 12.25).

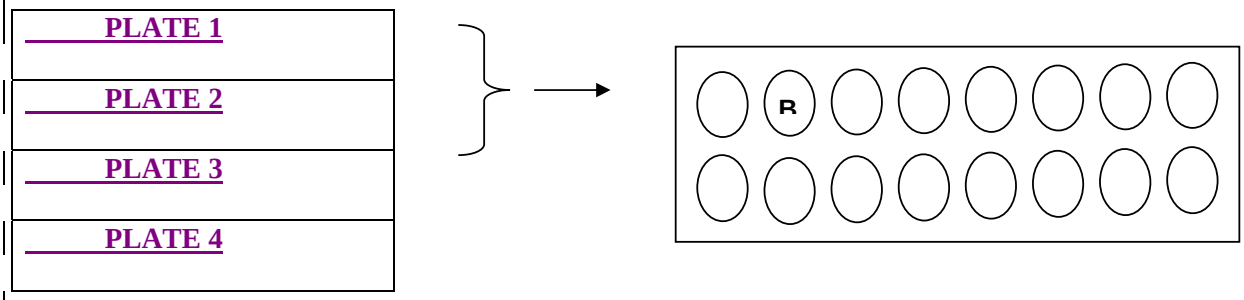
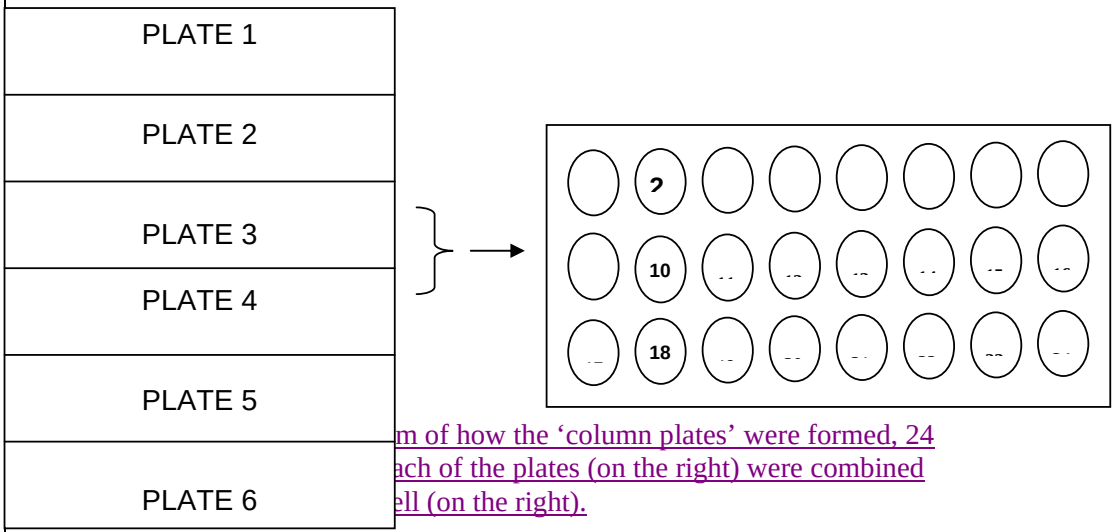


Figure 12.24. Schematic diagram of how the 'Row plates' were formed, sixteen rows from each of the plates (on the right) were combined into a single well (on the right).



m of how the 'column plates' were formed, 24
ach of the plates (on the right) were combined
ell (on the right).

The 4 column and 5 row plates were incubated at 37°C overnight. A total of 11 plates PP, RP and CP were obtained. For PCR amplification, 5ul of the bacteria culture was transferred using a multi-pipette to a clean 96 well plate and the bacteria pelleted at 4500rpm for 10 minutes in a Sorvall centrifuge. The supernatant was discarded and the pellet re-suspended in 5ul of sterile water and used as template for PCR amplification. PCR amplification condition for the SSR marker was 2 mM of $MgCl_2$, 0.2 mM of DNTPs, 0.2 uM of each primer, 1 U of taq-polimerasa, in a final volume of 50ul. Thermal cycle profile was an initial denaturation step at 95ºC 2 min, 30 cycles 94º 30s, 55ºC 1 min, 72ºC 1 min, and a final extension step of 72ºC for 5 min. For the SCAR marker, $MgCl_2$ and DNTP concentrations were increased to 2.5 mM and 0.4 mM respectively. PCR cycling conditions were: 95ºC 2 min, 30 cycles: 94ºC 30 seg, 52ºC 1 min, 72ºC 1 min; and a final extension of 72ºC for 5 min. PCR products were visualized in a 1.5% agarose gel stained with ethidium bromide.

The TME3 BAC library is made up of 73 728 clones in 192 384-well plates. Insert size ranged from 20kb to 130kb with an average insert size of 100kb (Figure 12.26). The BAC library has a 10X coverage of the cassava genome.

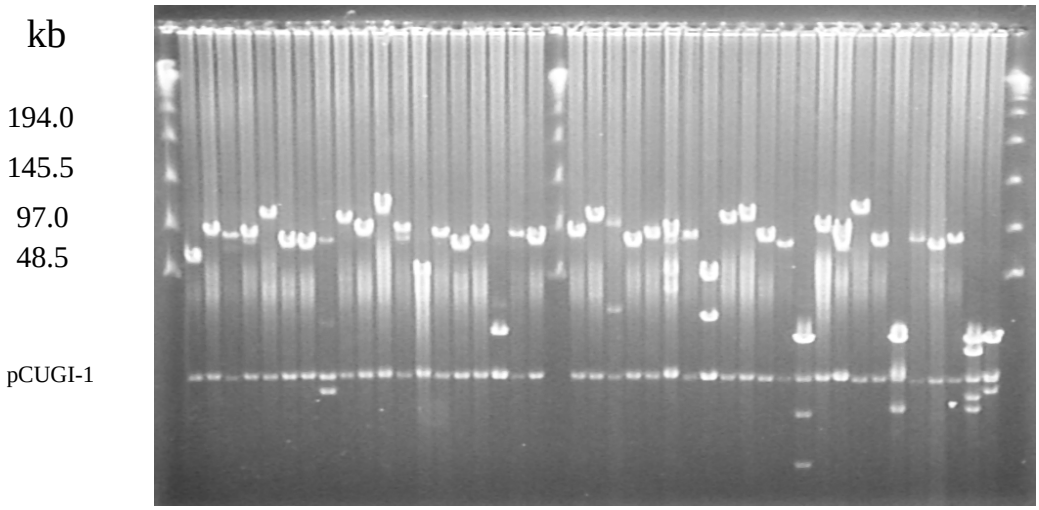


Figure 12.26. Random BAC clones from the TME3 library digested with *Not 1* and run in a 1% agarose gel in 0.5X TBE at 14°C at 6V/cm with a switch time of 5s-15s pulse time for 14 hours.

Results of screening the 'BAC pools' with NS158 yielded 2 positive clones while screening with RME1 yielded 34 positive clones. NS158 is a single copy SSR marker while RME1 was developed from a multiple copy RAPD marker. The clones were digested with 20U of HindIII overnight and run for 24 hours on 1.2% agarose gel to obtain a BAC clone fingerprint (Figure ???.)

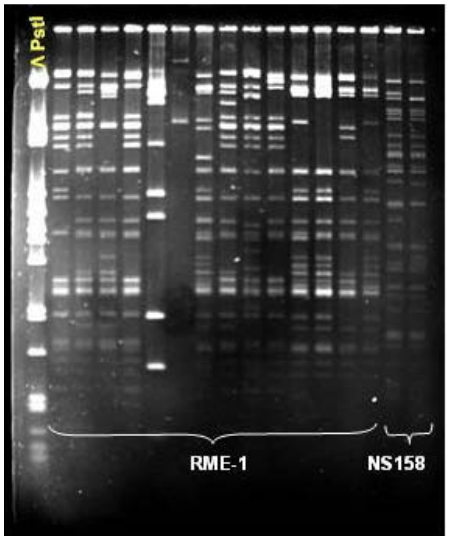


Figure ???. Fingerprinting of positive BAC clones.

The positive clones are listed below (Fig.8.28)

"PP" ("Pool Plate")	"RP"("Row Plate")	"CP"(Column Plate")
90	N	18
189	M	19

"PP" ("Pool Plate")	"RP"("Row Plate")	"CP"(Column Plate")
12	N	14
17	M	21
34	L	16
47	O	3
51	I	1
52	A	23
54		
85	J	23
130	D	18
136	J	8
139	G	22
155	K	24
173	F	1

Formatted: Bullets and Numbering

Figure ???. List of positive clones from the 'BAC pool' screening using the SSR marker NS158 and the SCAR marker RME1.

VII Advanced works (*Kole recently included this section.*)

Third Generation maps; physical, RH

Yvonne we don't have physical maps yet, we have some small BAC contigs nothing close to physical maps

Tools developed: ESTs Microarrays

EST in cassava were first reported by Suárez, et al (2000) based on cDNA–AFLP analysis of the mRNA from the parents of the genetic mapping population, and obtained over 500 transcript-derived fragments (TDFs). ESTs were identified in about half of the TDFs and six were assigned to the both the male and female derived maps. Currently, further development of SSR markers involving untranslated regions of cassava ESTs for SSR repeats are underway (Mba *et al*., 2001).

A collaborative project between IITA and the USDA-Agricultural Research Service, Biosciences Research Laboratory at Fargo, North Dakota USA was initiated to generate ESTs from cassava, and other related Euphorbs such as leafy spurge, in order to prepare gene catalogues that are requisite for development of DNA arrays. A hybridisation test microarray containing clones from a leafy spurge EST-database successfully identified gene expression changes associated with cytokinin imbalances in poinsettia samples. Analysis of gene sequences from cassava and leafy spurge databases indicated that many genes retain sufficient conservation (sequence similarity) to cross-hybridise (Anderson et al., 2001). Using DNA macroarrays developed using clones from a leafy spurge EST-database (Anderson and Horvath, 2001), we further demonstrated that at least 35% of the leafy spurge clones showed greater than 2X above background hybridisation with dehydration-stressed cassava target sample (Anderson et al., 2004). Based on this clones from the leafy spurge EST-database are being used as probes to monitor gene expression profiles in cassava tissues challenged with various environmental stresses as well as other Euphorbiaceae family members. Genes involved in the G1/S and G2/M cell cycle progression in cassava have also been identified (Anderson et al. unpublished).

To data 20,000 ESTs have been generated for clones isolated from the two cassava libraries. Preliminary analysis of a subset of these revealed that 73.3% of the sequences had coding potential, 3741 sequences were single unique genes and1451 were unigene clusters with an average cluster size of 3.53. A wide range of gene functions could be assigned and microsatellite and SNP sites, which could be used in marker development, were identified (Lokko et al unpublished).

Details on structural and function genomics

VII. Future Work

A comparative assessment of gaps vis a vis other crops/ models

The three CMD linkage maps require further saturation with markers for a more extensive coverage of the genome. Addition SRR markers, high throughput markers such as EST and SNPs, may be employed in these population SNP could also facilitate mapping CMD resistance and other disease resistance genes of interest due to the high frequency of SNPs in the genome, there is high possibility of them being closely associated with resistance (Lohmann *et al*., 2000). SNP's also offer the opportunity to uncover allelic variation directly within expressed sequences of candidate genes to develop halplotypes based on gamete phase disequilibrium for analysis of quantitative traits.

With the recent development in EST in cassava, sequence tagged sites (STS), which are a short DNA sequence generally, between 100-500 bp that is easily recognisable and occurs only once in a chromosome or genome studied (Brown, 1999) can be developed to further saturate the maps and facilitate fine mapping.

With the availability and enhance of the computer package JOIN MAP future work on the cassava framework map TMS I30572 X CM 2177-2 would be to develop consensus map, based on male and female-derived framework map. Following this, a composite map of cassava, based on all exiting maps could then be used in planning experiments, to construct a genome database, to compare QTL identities in different genetic backgrounds, and for comparative mapping with other species. Composite maps have been reported in *Arabidopsis thaliana* and in Maize (Beavis and Grant, 1991; Hauge *et al*., 1993). Comparative mapping is a strategy that uses information obtained from the study of a species with a simple genome to make inferences about the map position and function of genes in species with a complex genome (Brown, 1999). Comparative mapping has been used successful in human and mouse genome mapping to increase the efficiency of mapping (Liu, 1998). It has been suggested that comparative mapping has definite advantages in mapping plant genomes. For instance, genes in wheat with a genome size of 17,000 Mb would be mapped from information obtained from a small member of the *Gramineae,* rice, with a genome size of 400 Mb (Gale and Devos, 1998; Liu 1998; Brown 1999).

For effective efficient map based gene cloning and associating candidate genes with important biological or agronomic traits there is a need for a physical map and an established correlation between the genetic maps and the physical map. Efforts at fine mapping specific regions of the cassava framework map at CIAT, followed by contig mapping with bacterial artificial chromosome (BAC) libraries is underway (Fregene *et al.,* 2001).

Table 1 First generation linkage maps of various plant species compared with the first cassava linkage map.

Crop	Marker	Population Type	Reference
Maize	RFLP	F_2	Helentjars et al 1987
Lettuce	RFLP, Isozymes	Intraspecific cross	Landry et al 1987
Potato	RFLP	Interspecific cross	Bonierbale et al 1988
Tomato	Random cDNA		Tanskley et al 1988
Pepper	RFLP, Isozymes	Interspecific cross	Tanskley et al 1988
Rice	RFLP	F_2	McCouch et al 1988
Arabidopsis	RFLP	F_2/F_3	Chang et al 1988
B. oleracea	RFLP	F_2	Slocum et al 1990
B. rapa	RFLP	F_2	Song et al 1991
B. napus	RFLP	F_2	Landry et al 1991
Cassava	RFLP, RAPD, Isozymes, SSR	Segregating F_1	Fregene et al., 1997

Martin's references (please check for constituency with text)

CIAT (1996) Development of Cassava Germplasm for the Drier Tropics and Subtroical Agroecosystems of Africa, Asia and Latin America. A Report by CIAT and CNMPF/EMBRAPA to the International Fund for Agricultural Development (IFAD). CIAT, Cali, Colombia. 30pp

CIAT (1999) Pogramme for the Enhancement of Income-Generation Potential through the Development and Transfer of Improved Cassava Genetic Material and Adoption of Post-Harvest Technologies. A Report by CIAT and CNMPF/EMBRAPA to the International Fund for Agricultural Development (IFAD). CIAT, Cali, Colombia. 40pp

CIAT, (2001). Annual Report Project SB2, Assessing and Utilizing Agrobiodiversity through Biotechnology, CIAT, Cali, Colombia, pp 239-241.

CIAT 2003a. Annual Report IP3. Improved cassava for the developing world. Pp8-1 to 8-3

DeVries, J., and Toenniessen, G. (2001). Securing the Harvest. Biotechnology, Breeding and Seed Systems for African Crops. CABI Publishers, New York, USA. 208pp.

FAO, (2000) www.fao.org. Internet Database.

(Fregene and Mba 2004). Marker-Assisted Selection (MAS) In: Hershey C. (ed) Cassava Breeding. FAO, Rome Italy (in press)

Fukuda, W.M., and Saad, N. (2001). Participatory Research in Cassava Breeding with Farmers in Northeastern Brazil. Working Document No. 99 CNPMF, Cruz das Almas-Bahia, Brazil, 42pp.

IITA (1999) Annual report Project 14. International Institute of Tropical Agriculture, Ibadan, Nigeria

IITA (2000b) Enhancement of Income-Generation Potential through the Development and Transfer of Improved Cassava Genetic Material and Adoption of Post-Harvest Technologies in Africa. A report by IITA to the International Fund for Agricultural Development (IFAD). IITA, Ibadann, Nigeria. 200pp

IITA (2000b). Impact: The Contribution of IITA-imporved cassava to food security in sub-Saharan Africa: an impact study. IITA, Ibadan, Nigeria. 10pp

Formatted

Kawano, K. (2001). The role of improved cassava cultivars in generating income for better farm management. In: Howeler R.H., and Tan S.L. (eds). Cassava's Potential in the 21st Century: Present Situation and Future Research and Development Needs. Proceedings of the Sixth Regional Workshop, held in Ho Chi Minh city, Vietnam. Feb 21-25, 2000, CIAT, Bangkok, Thailand, pp 5-15.

Nweke, F., Spencer, D., Lynam, J. (2001) The Cassava Transformation: Africa's Best-Kept Secret. Michigan State University Press, East Lansing, USA. 273 pp.

Kullaya A.; K. Mtunda; H. Kulembeka; M. Ferguson; J. Marin; C. Ospina; E. Barrera; A. Jarvis; N. Morante; H. Ceballos; Tohme J; M. Fregene (2004) Molecular marker-assisted and farmer participatory improvement of cassava germplasm for farmer/market preferred traits in Tanzania In: Alves and Tohme (eds). Adding Value to a Small-Farmer Crop: Proceedings of the Sixth International Scientific Meeting of the Cassava Biotechnology Network. 8-14 March 2004, CIAT, Cali Colombia. Book of Abstracts Pp70