African Crop Science Journal, Vol. 12. No. 3, pp. 171-187, 2004
Printed in Uganda. All rights reserved
ISSN 1021-9730/2004 $ 4.00

ASSESSMENT OF GENETIC DIVERSITY OF LOCAL VARIETIES OF CASSAVA IN TANZANIA USING MOLECULAR MARKERS

F. HERZBERG, N.M. MAHUNGU[1], J. MIGNOUNA[2] and A. KULLAYA[3]
University of Göttingen, Germany (Present address: PhytoPharm Consulting Berlin, Germany)
[1]Southern Africa Root Crops Research Network (IITA/SARRNET) Dar-es-Salaam, Tanzania
[2]International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria
[3]ARI-Micocheni, Dar-es-Salaam, Tanzania

ABSTRACT

Twenty four arbitrarily selected cassava (*Manihot esculenta* Crantz) landraces, supplied by the germplasm collections in Kibaha and Ukiriguru research stations of Tanzania, were analysed by two different molecular marker techniques. Genetic distances on the basis of RAPD (random amplified polymorphic DNA) revealed separate clustering of almost all coast region-derived varieties. Inter-sequence-tagged repeat (ISTR) analysis, the second method, reproduced the geographical segregation, though the effect was less pronounced. Two genotypes were found to be very similar with either method, indicating a possible double accession. There is high overall genetic variability of the cassava germplasm. However, the variability is non- homogenous; rather it represents varieties of the same geographical origin grouped together. These findings are useful in the rationalisation of sampling and management of germplasm collections.

Key Words: DNA, genetics, *Manihot esculenta*

RÉSUMÉ

Vingt quatre variétés locales de manioc (*Manihot esculenta* Crantz) étaient arbitrairement sélectionnées dans la collection de germe plasme de la station de recherché d'Ikiriguru-Kibahade en Tanzanie. Elles étaient analysées par deux techniques de molécules marqueuses différentes. La distance génétique sur base de RAPD (Amplification Polymorphique de l'ADN) révéla une séparation en classe de presque toutes les variétés en provenance de la région côtière. La première méthode concernant les analyses des répétitions des inter-sequences marquées, la seconde méthode, réproduisa la ségrégation géographique, même si l'effet était moins prononcé. Deux génotypes étaient trouvés similaires par les deux methods, indiquant une possibilité d'une double accession. Il y a une variation génétique élévée de germeplasme de manioc. Cependant, la variabilité est non-homogène; mais représente les variétés de la même origine géographique groupées ensemble. Ces résultants sont très importants dans la rationalization de l'échantillonnage et gestion des collections des germeplasmes.

Mots Clés: AND, génétiques, *Manihot esculenta*

INTRODUCTION

Cassava (*Manihot esculenta* Crantz) ranks sixth among the world food crops within Sub-Saharan Africa and is predominant in many regions of Tanzania with a total production of 6.2 million tonnes in 1998 (FAOSTAT, 1999).

Over 70% of cassava is commonly found in four major zones in Tanzania. These include the southern zone (Mtwara and Lindi regions),

accounting for 32% of the total production; the eastern zone (Tanga, Coast, Dar-es-Salaam and Morogoro regions), accounts for 18%; the lake zone (Bukoba, Mwanza, Mara regions and some parts of Shinyanga), accouts for 13%; and the southern highlands (Ruvuma and Rukwa region, some parts of Iringa and areas along Lake Malawi), account for 8.7%. The Western Zone (Kigoma and along Lake Tanganyika, Tabora region, some parts of Dodoma region) accounts for the rest.

As in all African countries, farmers grow more than one variety at a time and turnovers of varieties are high (Nweke *et al.*, 1994). A survey (Eminger, 1994), conducted in the coast region, revealed that forty-four different 'sweet" and eight 'bitter' varieties were grown by small farmers in the area, and local preferences for certain cultivars varied considerably. In addition, farmers are eager to exchange planting material and to some extent, make use of the open-pollinated seedlings on their farms as observed in Mtwara region.

The International Institute of Tropical Agriculture (IITA) has developed improved clones, which are high yielding, of high quality, and resistant to pests and diseases. The Tanzania National Root Crops Program, in concert with the Southern Africa Root Crops Research Netowrk (SARRNET), IITA executed project, carries out the development and distribution of improved lines in the country. Breeding and selection trials using botanical seeds and clonal lines from local and introduced material aim at improved productivity and root characteristics, high level of resistance to pests and diseases, and adaptability to environmental stresses in various ecological zones. Overall, 294 accessions are maintained including 102 local clones, 151 IITA introduces clones and 41 local breeding lines (Jeremiah *et al.*, 1994). Participatory approaches with reference to variety selection for cassava and identification of research targets are being used in Tanzania (Thro *et al.*, 1996; Kapinga *et al.*, 1997; Mahungu, 1998).

Distinct variants of cassava mosaic virus occur in Tanzania. The African cassava mosaic virus (ACMV) is restricted to Western Tanzania. Another type, the East African Mosaic Virus (EACMV), was originally thought to be restricted to the coast region. However, recent findings, using a discriminating ELISA, suggest that EACMV also occurs in the Lake Zone in Western Tanzania (Swanson and Harrison, 1994; Ogbe *et al.*, 1997). In the late 90's, the pandemic of CMD observed in the Lake zone confirmed the occurrence of the Uganda variant (*UgV*) of EACMD in the country.

On the other hand, exclusive reliability on few improved cultivars would necessarily narrow the genetic base leading to genetic erosion. Such a case has been reported with two clones, which had been successfully adopted in West and Central Africa (TMS 30572 and TMS 4(2)1425). These were derived from a single source of ACMD resistance (Mignouna and Dixon, 1997). Hence, in addition to "*in vitro* conservation" facilities like germplasm collections, "on-farm conservation" by the small farmer plays a crucial role in maintaining a broad genetic variability of cassava (Brush, 1995).

The estimation of the diversity of plant genetic resources has become much more simple and reliable since the advent of the molecular marker technology. In contrast to morphological or biochemical marker techniques, DNA-based methods are independent of environmental factors and give rise to a high number of polymorphic loci (Karp *et al.*, 1997). This holds particularly true for DNA finger printing or "DNA-profiting" methods based on the polymerase chain reaction (PCR).

Two distinct finger printing methods were used in the present study, namely; the Random Amplified Polymorphic DNA (RAPD) technique, which is a modification of the basic PCR protocol (Williams *et al.*, 1990; Inverse sequence-tagged repeat (ISTR)) which exploits the ubiquitous dispersal of so-called copia-like sequences in the eucaryotic genome (Rohde, 1996).

The objective of the present study was to put the description of cassava germplasm in Tanzania on a more solid basis by employing PCR-based molecular market techniques that are presently established at the Mikocheni Agricultural Research Institute.

MATERIALS AND METHODS

Plant material. Random samples of local cassava varieties were obtained from the germplasm collections at Ukiringuru and Kibaha in Tanzania.

All coast region varieties were from the latter, whereas Ukiringuru supplied material of broader origin, including all Lake region cultivars (Fig. 1). Leaf materials were checked CMD contamination. For diversity studies, leaves of young shoots from potted stem cuttings were preferred. For reference and long-term storage, shoots from the sprouted cuttings were also maintained in vitro, according to the IITA protocol for the set up of note cultures (Ng and Ng, 1997).

CMV screening. For screening for the presence of viral DNA, a rapid small-scale extraction method was employed (*Dellaporta et al.*, 1983) which used only milligramme amounts of leaf material. Detection of contaminated viral DNA was performed by PCR using primers, which amplify a DNA fragment of the core protein-coding region of both, ACMV and EACMV. A standard PCR protocol of 35 cycles was adopted (annealing temperature 53°C).

DNA extraction and analysis. Genomic DNA was extracted in the presence of CTAB following the protocol of Doyle and Doyle (1990), starting from 5-10 g of fresh leaves. The extracts were treated with RNAse A (50 Fg/ml) and stored as isopropanol precipitates at -20°C. Working dilutions were stored as aqueous solutions for not more than two weeks. The integrity of DNA and the efficiency of RNA depletion were loaded on 1% agarose gels electrophoresis. Usually, 2 Fl of a 1:100 dilution were loaded on 1% agarose gels. The nucleic acid amount and purity of the sample was estimated by measuring the absorbance at 260 mm (OD_{260}) and by calculating the $OD_{260/280}$ ratio, respectively.

Random Amplified Polymorphic DNA (RAPD). The RAPD PCR was carried out according to Williams *et al.* (1990). A hundred ng of genomic DNA (according to the OD_{260}) were used as template for random-primed amplification with commercial 10-mer oligonucleotides (Operon Technologies, Alameda, USA). The composition of the reaction mix and cycling parameters were determined empirically. Optimum conditions found for a 25 Fl reaction were as follows: 100 mM Tris-HCl, pH 8.3, 50 mM KCL, 2 mM MgCL2, 0.2 mM of each dATP, dCTP, dGTP, and dTTP, 1 unit of Taq Polymerase, 5 pMol 10-mer primer, and 25 ng of DNA template.

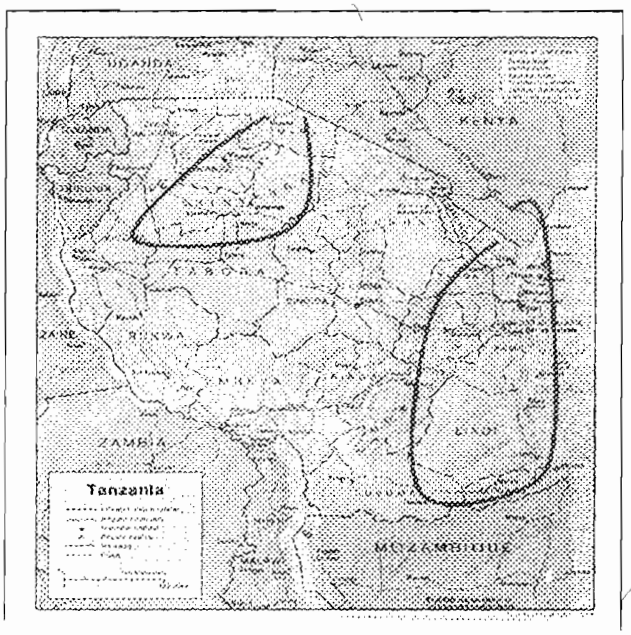


Figure 1. Map of Tanzania showing areas where cassava materials were obtained (circled)

Random-primed amplification using a "Genius" thermocycler (Techne) was achieved by 45 cycles of the sequence 94°C/1 minutes, 35°C/1 minutes, 72°C/2 minutes, preceded by a denaturation step and followed by a final extension step (72°C/10 minutes). Products were analysed by agarose gel electrophoresis using the Sub-cell Model 96 (Bio-RAD, Richmond, USA) providing a separation distance of about 25 cm. 1.5% agarose gels were run in 0.5 x TBE at 2.5 V/cm. Gels were cast with ethidium bromide (0.5 Fg/ml) but were usually post-stained to improve visualisation of the bands. Photographs were taken under short-wave UV light on Polaroid film.

Inter-sequence-tagged repeat (ISTR). The protocol described by Rohde (1996) was adopted. Forward (F) and backward (B) primers were radioactively labelled using T4 polynucleotide kinase and [γ-^{33}P] ATP. PCR reactions were performed in a final volume of 20 µl containing 20 ng of DNA, 1 pMol of each primer, 2.5 mM $MgCl_2$, 0.2 mM dNTP's, 1 x PCR buffer, and 1U of Taq DNA polymerase. Reaction took place in a 'UNO' Thermoblock (Biometra) following the programme: 1.95°C/3 minutes, 2.95°C/30 s, 3.45°C/30s, 4.72°C/2 minutes, 5.72°C/ 0 minutes. Steps 2 to 4 were looped and repeated for 40 cycles. Amplicons were separated on 4% polyacrylamide sequencing gels, run in 1 x TBE (pH 8.8) at 29 W for about 2 h. 1.5 µl aliquots were loaded per slot after addition of sequencing loading buffer and heating to 93°C for 5 minutes. The gel was exposed to XAR film (Kodak, USA) at room temperature for 24 h.

Data analysis. Polymorphism data were obtained by scoring shared bands between pairs of genotypes. The binary code designated the presence (1) or absence (0) of a band at a specific location. Missing values (e.g. resulting from a failed PCR reaction or faint, i.e. uncertain bands) were designated by 2. Genetic distances (GD = 1 - JC) were computed and arranged in a dissimilarity matrix on the basis of Jaccard's coefficient (JC) (*Jaccard*, 1908) using the software DDAT, 2.1 (E. Ritter, Centro de Investigacion y Mejora Agraria, CIMA, Spain). Cluster analysis based on the Unweighted Pair Group Method with Arithmetic Averages (UPGMA) and a phenetic dendrogram was computed with NTSYS-pc software, V. 1.80 (Applied Biostatistics, Inc.; Rohlfs, 1989). Additionally, a three-dimensional scatter plot was generated on the basis of a Principal Component Analysis (PCA).

RESULTS

Prior to the analysis of genetic diversity, preliminary experiments had been performed aiming at providing clean plant material and to adjust the experimental parameters of the RAPD-PCR. These experiments were necessary because of the inherent difficulties of the RAPD technique. The use of short arbitrary oligomers as primers in combination with low stringency conditions (35°C annealing temperature) give rise to false-priming resulting in unreliable fingerprints easily. In addition, the universality of the primers and the sensitivity of the PCR could lead to amplification of any contaminating DNA, e.g. by viral infection.

Therefore, in a first attempt, the leaf material from different varieties, supplied by the germplasm collections in Kibaha and Ukiriguru, was screened against contamination by cassava mosaic disease (CMD). For this purpose, a PCR was established which made use of 'universal' primers. These primers anneal to a conservative stretch coding for the core protein of any CMV subtype, amplifying a 680 bp fragment in PCR. Figure 2 shows typical result of this screening. A specific 680 bp fragment, comigrating with positive control DNA, was amplified in some cases, indicating contamination. Samples, which had been proved positive, were excluded from further studies.

Optmising of RAPD-PCR protocol included DNA template concentration, $MgCl_2$, dNTP's concentrations, Taq-Polymerase activity, annealing temperature, type of thermocycler and selection of RAPD primers.

Low concentrations of magnesium ions (which act as cofactors for the activity of the Taq polymerase) favour the amplification of shorter fragments. Raising $MgCl_2$ results in fragments of increasing size (Fig. 3A). The amount of template is critical as is shown in Figure 2B. Too concentrated as well as too diluted DNA results in poor or no patterns. The reason for the former is probably due to accumulation of inhibitors in the extract (e.g. polysaccharides), which have to be

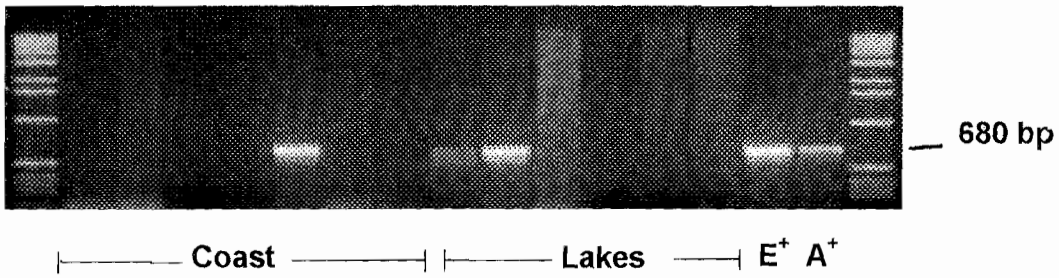
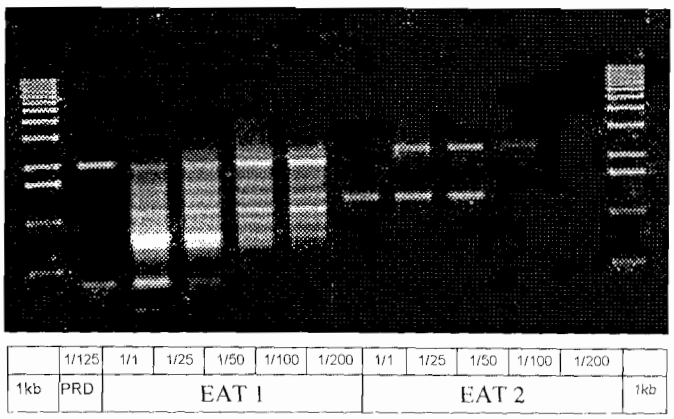


Figure 2. DNA screening for viral contamination of cassava varieties originating from different geographical regions. PCR was performed with the universal UV Ba, CP and ACMV D1, which result in a 680 bp fragment of both, ACMV and EACMV DNA. NC = negative control; E^+ = positive control for EACMV (No. 54); A+ = positive control for ACMV (No. 27). The 1 kb ladder, serving as size standard, was loaded in the outer slots. Reaction products were analysed on 1% agarose gels, visualised by ethidium bromide staining under UV light and photographed on polaroid film.



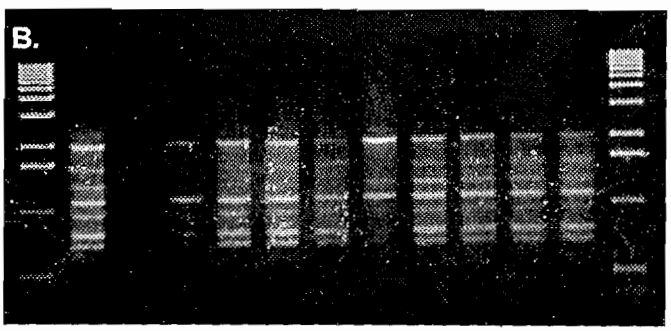


Figure 3. Adjustment of RAPD-PCR by variation of critical parameters. A. Amplification of a 1:125 dilution of coconut genomic DNA (PRD) in the presence of 1-5 mM $MgCl_2$. Two different Operon primers were tested. B. Dilutors of DNA extracts ranging from 1:1 to 1:200 from two different coconut varieties (EAT 1, EAT 2) were used as templates. PRD DNA from experiment A was co-amplified. 4 mM $MgCl_2$ and primer OPA-02 was used throughout. Note the reproducible band pattern of the two independent experiments.

diluted to become inactive. Figure 3 also demonstrates the reproductively of patterns from one experiment to another under optimised conditions and the variability in the number of bands yielded by different primers. Therefore, following adjustment of all other parameters, informative primers had to be identified as shown below.

1kb 1 2 3 4 5 1 2 3 4 5 1kb
OPA-02 OPA-05

1/125 1/1 1/25 1/50 1/100 1/200 1/1 1/25 1/50 1/100 1/200
1kb PRD EAT 1 EAT 2 1kb

Two other factors were found to profoundly affect the outcome of a RAPD experiment, namely, storage conditions of the genomic DNA and the type of tehrmocycler in combination with extreme climatic conditions. It had been observed that after a series of experiments, the number of blanks within the set of samples gradually increased, although all other parameters were carefully adjusted.

Successful amplification turned out to be highly dependent on the storage conditions of the DNA. Since it is generally not recommended to store genomic DNA frozen, working dilutions were prepared and stored at 4°C for several weeks. However, storage for more than four weeks led to massive degradation (data not shown). To overcome this problem, DNA was kept for a maximum of two weeks at 4°C, and thereafter, replaced by a stock, which was maintained as isopropanol precipitates at - 20°C.

A serious drawback was encountered by the thermocycler, which had been used for the bulk of optimisation experiments. The instrument, a Biometra UNO equipped with a 96-well thermoblock, exhibited a delayed temperature adjustment and pronounced sensitivity towards increased ambient temperatures. As a result, the annealing temperature interval, the most critical step in RAPD-PCR, was drastically reduced. This problem was addressed only after new instruments for comparison became available. Re-evaluation of experimental parameters using a new thermocycler (Techne Genius) revealed that most of the reagents could be reduced considerably. Most notably, Taq polymerase could be reduced by 50% (1 instead of 2 units per sample). Correspondingly, optimum magnesium concentration was decreased from 4 to 2 mM.

Selection of suitable primers was done by screening 50 arbitrary oligonucleotides derived from the different commercial sets (Operon Technologies Inc., Alameda). The Operon primers were tested on a subset of five cassava genotypes. A snapshot of the screening by testing primers of the OPB-series, is given in Figure 4. Interestingly, some primers did not amplify any bands (e.g., OPB-13), yet some gave poor patterns only (e.g., OPB - 14). For the genetic diversity analysis, primers that fulfilled such criteria as multitudes of bands, high degree of polymorphic bands, and few or no failures (blanks) were chosen.

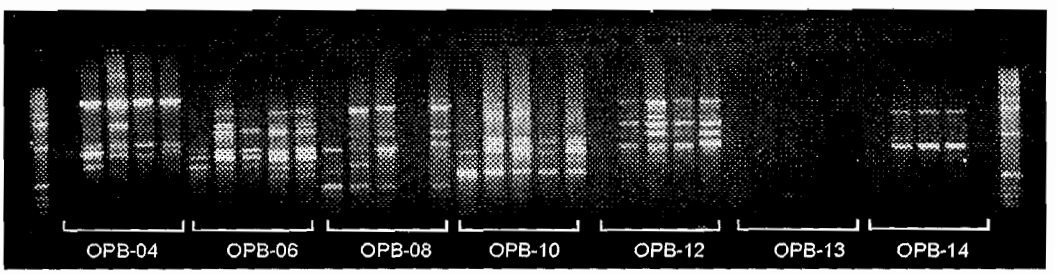


Figure 4. A typical screening result obtained by a subset of five cassava varieties probed with different Operon primers of the OPB series.

In the present example, primer OPB - 06 met all three criteria and was, therefore, included in the study. The final RAPD analysis comprised of 24 cassava landraces, 12 coast region varieties obtained from the germplasm collection in Kibaha, and 12 others supplied by the Ukiriguru collection. The majority of the latter had their origin in the regions of Lake Victoria, Lake Tanganyika as well as Tabora. Two varieties, however, were originally collected from Mtwara in Southern Tanzania. Table 1 presents the names and provenance of the varieties, together with the available passport data.

Extraction of genomic DNA in the presence of CTAB yielded between 800-1600 Fg of DNA from up to 7.5 of green leaves, as determined spectrophotmetrically. The OD ratio 260 nm/280 nm was between 1.7 - 1.9, indicating pure nuclei acids. Eight Operon primers gave a variety of polymorphic bands and were considered in the RAPD study. The fragment yields of the primers are depicted in Table 2. The eight primers together generated 103 scorable loci (i.e., band levels or fragments); 10-18 fragments per primer were produced, (mean: 12.9). Of these 103 fragments, 93 (90.3%) were found to be polymorphic, i.e. there was at least one unmatched at a specific locus within the set of 24 genotypes. The most informative primers were OPB-12 and OPC-01, both of which generated 22 different patterns, i.e., one single primer was able to discriminate between almost all varieties. Combining the data for each single primer resulted in 22 patterns, i.e. the two primers together had the same discriminatory power as OPB-12 or OPC-01 alone.

Three fingerprints resulting from primers OPB-06, OPB-12 and OPC-01 are presented in Figure 5. These demonstrate the high variability of the patterns among the 24 individuals and also between the primers. However, total failure of amplification was also observed sporadically (black lanes). Single repeated reactions proved that these blanks resulted from incomplete reaction mixes or degraded templates and not because of absent priming sites.

A dissimilarity matrix based on Jaccard's coefficient (JC) was computed by the pair-wise comparison of all 24 genotypes. Genetic distances (GD = I - JC) ranged from close similarity by GD = 0.128 (pair K 18/K 24) to a maximum divergence of 0.712 (K 1/K23). Mean distance was calculated by 0.457.

The genetic distance data were clustered using the UPGMA algorithm. To compute the clusters, the dissimilarity matrix had to be transformed into a cophenatic matrix. The goodness-of-fit between the two matrices was calculated, revealing a correlation of 78% significance. The resulting phenetic dendgrogam revealed three major clusters at the 60% similarity level (Fig. 6). Most interestingly, two clusters (I, II) contained only coast region genotypes (plus "Mwasunga", K22), whereas the third cluster (III) comprised the bulk of all other varieties. Few genotypes did not segregate to either cluster. Varieties K 3 ("Kilokote") and K II ("Agrikacha") of the coast region as well as K21 ("Mapangano") and K 23 ("Dalama") were found to be the most unrelated varieties with respect to all other cultivars. On the other hand, two varieties, K 18 ("Zolo") and K 24 ("Mganda") revealed more than 90% similarity, suggesting a double accession.

The above findings are supported by the principal component analysis (Fig. 7). Though PCA is based on a different algorithm using Euclidian distances, the grouping of most of the coast region varieties in the resulting three-dimensional scatter plot was striking (again "contaminated" only by K22). The separation from the other varieties was already evident along the first principal coordinate, which alone explained 48.6% of the total variance. K11 and K23 segregated as expected and only K 3 remained intermediate.

In conclusion, the analysis of genetic distances based on RAPD data clearly reconstructed the provenance of the genotypes, suggesting the existence of geographical varieties of cassava in Tanzania.

In co-operation with the Max-Planck-Institute for Breeding Research in Cologne, Germany, the same 24 cassava varieties were analysed using a different approach to estimate genetic distances, based on the variability of genomic DNA caused by insertion of ubiquitous copia-like repetitive elements. Primer combinations flanking those replicons were designed, which gave rise to high-resolving finger prints after applying the PCR reactions to sequencing gels. An example of the resolution power of this technique is given in

TABLE 1. Key to accessions and available passport data. Genotypes were number K1 - K24 to simplify analysis. K 1-12 were obtained from the Kibaha collection, K 13-24 from Ukiriguru. CMD = cassava mosaic disease; CBSD = Cassava Brown Streak Disease (resistance judged on the basis of aerial symptoms only); CGM = Cassava green mite. S = sweet; B = bitter, 1 = intermediate; MAP = months after planting. "others" refers to a blend a predominant petiole colours (green, light green, reddish), n/a present. The feature "stem colour" was excluded as it is not representative for the stage when plants were collected (ca. 5 MAP)

Key	Kibaha collection						
	Variety (local name)	Origin	Branching	Morphological description 1. Petiole 2. Mature leaf 3. Young leaf 4. Leaf shape	Pest and disease resistance 1. CMD 2. CBSD 2. CBSD 3. CGM	Root taste	Maturing (MAP)
K1	Kibaha	Pangani	High	1 2 1 horizontal arrangement	1 + 2 + 3 -	S	>9
K2	Gezaulole	Tanga	No (sometimes high)	3 2 2 2	1 + 2 +/- 3 +/-	S	Late 12
K3	Kilokote	Kibaha	Low	2 2 2 2	1 + 2 +/- 3 +/-	S	8
K4	Kititu	Kirogwe	Low	3 2 2 2	1 - 2 + 3 -	S	Early 6
K5	Kiroba	Rufiji	Low	1 2 2 2	1 + 2 - 3 -	S	Early 6
K6	Mzungu	Bagamoyo	Low	3 2 1 2	1 - 2 + 3 +	S	Early 7
K7	Nsufi-3	Korogwe	High	2 2 2 2	1 - 2 + 3 -	S	8
K8	Msufi-3	Korogwe	High	2 4 3 2	1 - 2 + 3 -	S	8
K9	Athumani-1	Kilosa	Low	3 2 1 2	1 +/- 2 +/- 3 +/-	S	8

TABLE 1. Contd.

Key	Variety (local name)	Origin	Branching	Morphological description	Pest and disease resistance	Root taste	Maturing (MAP)
				1. Petiole 2. Mature leaf 3. Young leaf 4. Leaf shape	1. CMD 2. CBSD 2. CBSD 3. CGM		
K10	Ex-Msangazi	Muheza	Medium	2. 2 3	deed purple (apical)	1 + 2 +/- 3 +/-	S 8
K11	Agirikacha	Bagamoyo	Low	3 2 2 2	1 - 2 - 3 -	S	Early 6
K12	Mkukumkuku	Bagamoyo	High	1 2 1 2	1 +/- 2 - 3 +/-	S	>12
				Ukiriguru collection			
K13	Mapolu	Tabora	Low	1. Others 2. light green 3. bluish 4. broad	1 - 2 (n/p) 3 +	S	>12
K14	Rushura	Bukoba	High	1. Red 2. dark green 3. reddish 4. broad	1 + 2 (n/p) 3 +	S	>8
K15	Yuda	Mtwara	Low	1. Light green 2. dark green 3. reddish 4. broad	1 - 2 (n/p) 3 +	S	>8
K16	Binti Athumani	Tabora	Low	1. Others 2. green 3. reddish 4. narrow	1 + 2 (n/p) 3 +	S	>8
K17	Kabumba	Shinyanga	Low	1. Red 2. green 3. reddish blue 4. broad	1 +/- 2 3 +	S (n/p)	>8
K18	Zolo	n.d.	High	1. red 2. green 3. reddish 4. narrow	1 +/- 2 (n/p) 3 +	S	>12

MAP = Months after planting

TABLE 1. Contd.

Key	Variety (local name)	Origin	Branching	Morphological description	Pest and disease resistance	Root taste	Maturing (MAP)
				1. Petiole 2. Mature leaf 3. Young leaf 4. Leaf shape	1. CMD 2. CBSD 2. CBSD 3. CGM		
K19	Kajanga	Musoma	Low	1. red 2. green 3. dark green 4. narrow	1 +/- 2 (n/p) 3 +/-	S	8
K20	Rangi Mbili	Kwimba	High	1. Others 2. green 3. light green 4. narrow	1 + 2 (n/p) 3 +	S	>9-12
K21	Mapangano	Mtwara	High	1. red 2. green 3. light green 4. narrow	1 +/- 2 (n/p) 3 +	S	>12
K22	Mwasunga	Tabora	High	1. Red 2. green 3. light green 4. broad	1 + 2 (n/p) 3 +	S	>8
K23	Dalama	Kigoma	Low	1. Others 2. green 3. reddish blue 4. narrow	1 + 2 (n/p) 3 +	1	>12
K24	Mganda	Musoma	Low	1. Red 2. dark green 3. reddish blue 4. broad	1 + 2 (n/p) 3 +	S	>8

MAP = Months after planting

Figure 8, showing the result of ISTR-PCR using the primer combination F3/B2A.

The investigation was performed with seven different primer combinations, four of which resulted in scorable banding patterns. A total of 144 fragments was obtained, 131 or 91% of which were polymorphic. On the average, almost three times as much fragments were produced by each ISTR primer combination compared to the RAPD primers (32.75) versus 12.9). Each single primer combination was sufficient to discriminate between almost all 24 varieties (minimum 23 different patterns). Furthermore, specific fragments were amplified in high numbers by two of the two primer combinations (F3/B2A and F3/B5, making them particular useful tools market assisted selection).

Maximum genetic distance of all pair-wise comparisons was found for K 12/K 24 (GD = 0.64), whereas K 18/K 24 showed closest similarity (GD = 0.19). Average distance was calculated as GD = 0.49, i.e. overall intraspecific homology was 50%.

UPGMA clustering resulted in the dendrogram displayed in Figure 9. Two major clusters could be defined close to the 55% similarity level. The first cluster (1) was composed of the same 7 Coast Region varieties plus K 22 as the RAPD cluster 1. The second cluster (II) comprised 9 genotypes, including only one Coast region variety (K 2).

Thus, the hypothesis of geographical segregation is confirmed by ISTR analysis. The subcluster K 17/K 18/K 24, with K18 and K24 sharing closest similarity of all varieties, was also reproduced by the ISTR analysis.

On the other hand, substantial differences between the two data sets were observed. For example, cluster II included several other genotypes compared to the major cluster of the RAPD dendrogram. Moreover, K 23 segregated in the major cluster, whereas K12, which grouped with K7 and K10 in the RAPD dendrogram, was the most distant genotype here. In general, the ISTR data appeared to be more random, i.e. less varieties were grouped into clusters. This assumption was confirmed by the PCA-plot, which exhibited a more randomly scattered distribution of the 24 varieties compared to the corresponding plot computed on the basis of the RAPD data. The discrepancy could have been due to a relatively low numbers of loci included in the study to compute genetic distances between 24 closely related genotypes.

DISCUSSION

RAPD analysis yielded 93 polymorphic fragments or loci, ISTR analysis 131. The dendrograms resulting from either analysis showed significant similarities but also some divergence.

One cluster was completely reproduced. It contained five coast region varieties which has been collected originally in Pagani, Korogwe, Rufiji, Bagamoyo and Kilosa. The interspersed variety K 22 ('Mwasunga") from Tabora suggests a secondary introduction from the coast.

A second major cluster obtained with both methods excluded coast region varieties (with the exception of one variety from Tanga). In this case, however, the composition of the cluster by Ukiriguru-derived genotypes varied.

One subcluster of the RAPD dendrogram comprising the Ukiriguru-derived genotypes K 17 ("Kabumba"), K 18 ("Zolo") and K 24 ("Maganda") was also found in the ISTR dendrogram. Moreover, K 18 and K 24 exhibited the closet similarity to all other varieties (>90% identify). This raises the question whether the accessions are identical. They differ in some morphological passport data (leaf shape and colour) and some ISTR loci would imply distinct genotypes (e.g. compare lanes 18 and 24 in figure 8). In any case, the close relationship suggests one origin. Unfortunately the provenance for "Zolo" remained unclear but it is deduced from the genetic data that it was originally collected in the Musoma area.

Several genotypes did not segregate to any cluster or their relationship varied depending on the applied analysis. To put the data on a more solid basis, more polymorphic loci have to be added. A recent study estimating the genetic

TABLE 2. Fragments analysis of the RAPD primers included in genetic distance analysis of 24 cassava genotypes

Operon Primer	Sequence (5' >3')	Total no. of fragments	Polymorphic fragments[†]	Unique fragments S	Patterns
OPC-01	TTCGAGCCAG	15	14	1(1)	22
OPB-06	TGCTCTGCCC	10	7	0	13
OPB-08	GTCCACACGG	10	10	1	16
OPE-19	ACGGCGTATG	15	15	0(1)	20
OPC-02	GTGAGGCGTC	10	7	1	13
OPC-05	GATGACCGCC	14	13	0(2)	21
OPC-06	GAACGGACTC	11	9	0(1)	17
OPB-12	CCTTGACGCA	18	18	0(1)	22
	total	103	93(90.3%)		
	Mean	12.9			

*single bands plus missing values were scored as monomorphic
*values in brackets include specific bands plus missing values
*fragment differences (numbers of unmatches between each pair of genotypes summed over all fragments)

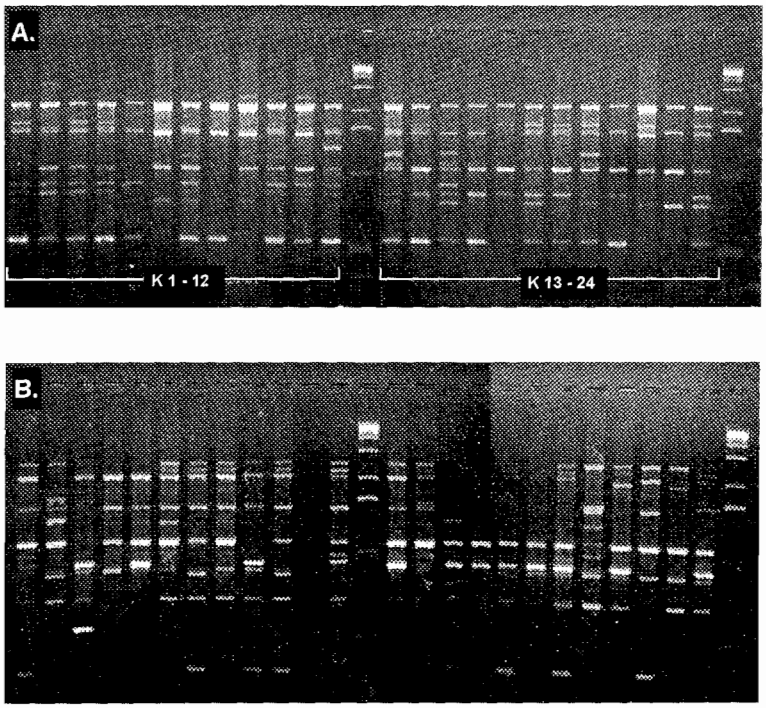


Figure 5. RAPD profiles from 24 cassava landraces generated by PCR amplification with the arbitrary primers OPC-01 (A) and OPB - 12 (B). Notice the polymorphic pattern obtained within the set of genotypes, but also notice putative monomorphic bands and the blank lane in B, where amplification failed. Slot loading is identical in both experiments. Samples were electrophoresed in 1.4% agarose.

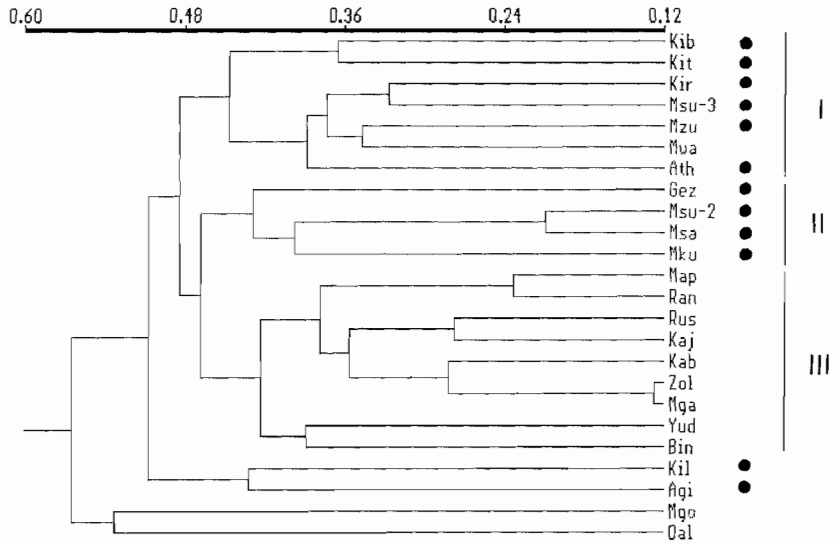


Figure 6. Phenetic dendrogam from 103 RAPD's based on UPGMA resulting from pair-wise comparisons employing Jaccard's coefficients (JC). Scale on top represents genetic distance (GD = 1-JC). Three clusters at the 60% homology level were defined. Notice that the Coast Region varieties (marked by black dots) segregate between cluster 1 and II, respectively.

variability among theobroma caccao accessions (Lecertau *et al.*, 1997) calculated a minimum of 506 RAPD fragments (corresponding 228 primers) required to reach a coefficient of variation of only 10% for genetic distances analysing a set of 155 genotypes.

A comparison between RAPD and ISTR lacks predecessors in the literature. RAPD analyses have been adopted excessively for genetic diversity studies of various tropical crops (Shah *et al.*, 1994; Yu and Nguyen, 1994; Asemota *et al.*, 1996) and comparison with other methods like RFLP has proved the value of this method (Ngoran *et al.*, 1994; Thorman *et al.*, 1994), whereas ISTR has so gar only been applied sporadically.

Few other RAPD studies have investigated the genetic variation of cassava germplasm. Laminski *et al.*, (1997) characterised South African elite varieties on the basis of 68 RAPD loci. The cluster analysis did not differentiate between local and international cultivars, which could have been due to the small number of polymorphic loci. Marmey *et al.* (1994) analysed 19 African cassava cultivars using eight RAPD primers, which gave rise to 85 loci. Although their plant material was of broader origin than used in this study, genetic variability was found to be homogeneously distributed without significant cluster formation. One explanation could be related to the different coefficient used for the estimation of genetic distance. They preferred simple matching (Sokal and Michener, 1958), which, in contrast to Jaccard's coefficient, takes 0/0 matches into account. This algorithm may lead to false positive matches with pair-wise comparisons, which results in less pronounced variability, especially when the number of loci is not very large. This assumption is confirmed by the high overall similarity among their tested genotypes (about 70%). Mignouna and Dixon (1997) investigated the genetic relationships among 35 cassava landraces with varying levels of ACMD resistance collected from various countries of West Africa. They applied 9 primers, which gave 74 fragments, 54 of which were found to be polymorphic. They could identify six clusters at the 60% similarity level, which showed a good correlation with the provenance of the genotypes. However, the objective of the analysis was to identify cassava landraces with high resistance to ACMD, which were genetically distinct to established lines. RAPD has also been used to assess the interspecific relationships of the genus Manihot aiming in tracing wild relatives (Schaal *et al.*, 1997).

Compared to the above data, the finding of the present study were unique, in so far as geographical races have been identified within a sample of relatively narrow distribution, i.e., within the

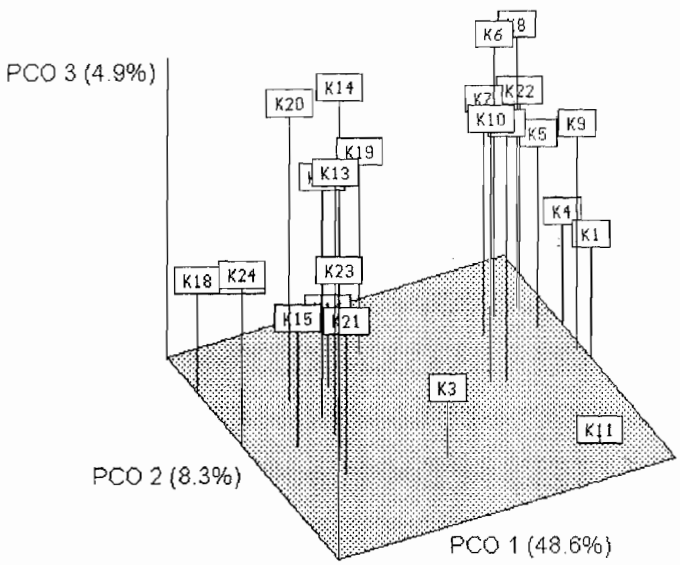


Figure 7. PCA ordination based on 103 RAPDs of 24 cassava genotypes. (Note: the first principal component (PCA1) accounts for 48.6% of variance among individuals).

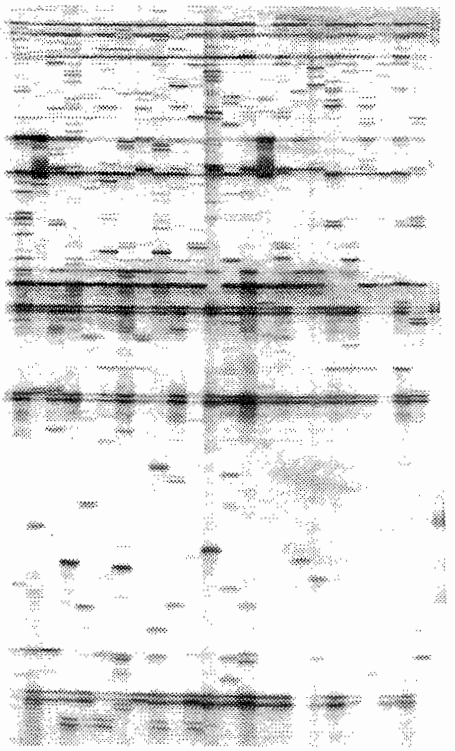


Figure 8. Autoradiograph of fragments generated by ISTR-PCR using primers F3 and B2A and resolved in 4% sequencing PAGE gels. 1.5 Fl of each PCR reaction had been loaded from left to right for cultivars K1 - K24.

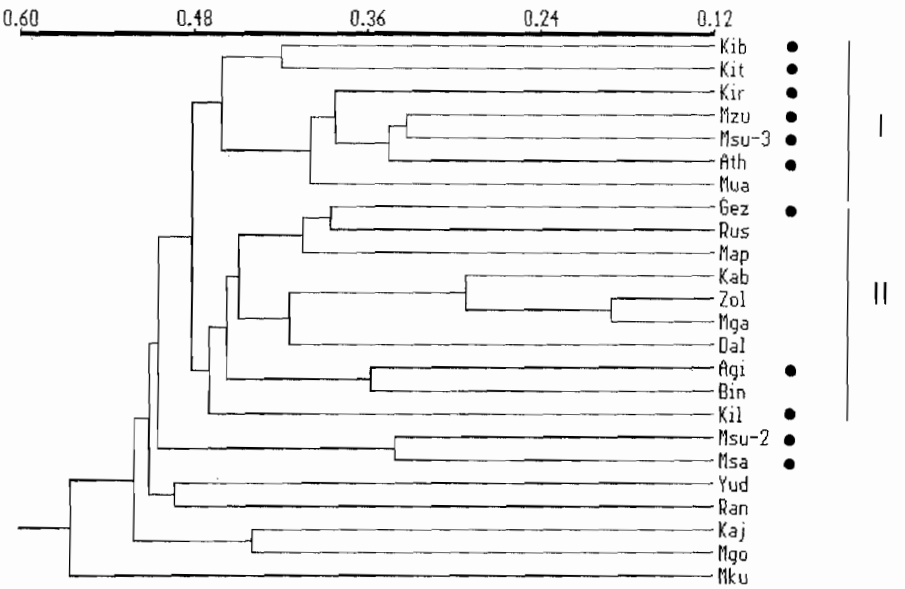


Figure 9. Phenetic dendrogram of ISTR analysis computed from a dissimilarity matrix of genetic distance based on Jaccard's Coefficient and UPGMA. Two clusters at the 55% homology level can be defined.

Tanzania borders. The key for the occurrence of geographical races in Tanzania is most likely a historical one. As mentioned earlier, Tanzania experienced a dual introduction of cassava in the past. For long periods, a mixing of the two secondary genepools was precluded by the central highland barrier, favouring a divergent evolution. Furthermore, it can be speculated that in the first place it was the coast region which benefited from the colonial promotion of improved lines distributed by the Amani station, thereby narrowing the genetic basis of that region. It is very likely that these historical circumstances are still influencing the pattern of current geographical distribution. The reason for this is linked to the mode of propagation of cassava. Traditionally, cassava is propagated vegetatively *via* stem cuttings, resulting in genetic clones. In contrast with seeds, the stem cuttings are more bulky and less durable, precluding greater dispersal.

Several other RAPD studies have also characterised geographical races. For example by the aid of RAPD markets, East and West European geographical races of the weed *Silene latifolia* were unambiguously identified (Vekkellop *et al.*, 1996). Ashburner *et al.* (1997) have defined geographical distinct populations of coconut in the Southern Pacific. *Johns et al.* (1997) have classified common bean landraces in Chile on the basis of RAPD data. The accessions they analysed, segregated between two major clusters, which corresponded to the Andean and the Mesoamerican genepool, respectively. This segregation was not found by using morphological descriptors. Likewise, geographical variation patterns encompassing almost all parts of Sub-Saharan Africa were described for the medicinal plant Phytolacca dodecandra by a RAPD study (Demeke and Adams, 1994).

Taken together, the data of this study suggest a high overall genetic variance of cassava landraces in Tanzania by the fact that at least 23 different genotypes of a random sample of 24 accessions were identified. On the other hand, this variability is not homogenous, as clusters of similar geographical origin and can be defined.

These findings may have direct consequences for sampling strategies because they imply that random but country-wide collection is superior to regional sampling. Furthermore, re-evaluation of the accessories in existing collections on the basis of genetic variance may help to identify gaps as well as doublets. Thus, linguisitic polymorphism may pretend on overestimated variability, since one genotype is given the same common name and vice versa.

To put sampling strategies and the management of germplasm collections on a rationale basis is best achieved by the establishment of molecular market technology. PCR-based methods are particular and useful tools not only to characterise the genetic diversity but also to develop marker-assited breeding strategies.

ACKNOWLEDGEMENTS

This work has been supported by a fellowship of the German Academic Exchange Service (DAAD). The authors are indebted to Prof. H. Becker from the Institute of Plant Production and Plant Breeding at the University of Gottingen for valuable supervisory role to the main author. To Mrs. Esther Masumba for technical support, and D. Becker, Maxplanck-Institute for Breeding Research, Germany, for help in generating the ISTR data. Dr. S. Winter of German Collection of Microorganisms and Cell Cultures, DSMZ, Brunswick, Germany for supplying the CMD primers.

REFERENCES

Asemota, H.N., Ramser, J., Lopez-Peralta, Weising, K., and Kahl, G. 1996. *Euphytica* 92:341-351.

Ashburner, G.R., Thompson, W.K., and Halloran, G.M. 1997. RAPD analysis of South Pacific Coconut palm populations. *Crop Science* 37:992-997.

Brush, S. 1995. In situ conservation of landraces in centres of crop diversity. *Crop Science* 35:346-354.

Carter, S.E., Fresco, L.O., Jones,. P.G. and Fairbairn, J.N. 1992. An atlas of cassava in Africa. Historical, agroecological and demographic aspects of crop production. Centro Internacional de Agricultural Tropical, Cali, Colombia.

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

Demeke, T. and Adams, R.P. 1994. The use of RAPDs to determine germplasm collection strategies in the African species, Phytolacca dodecandra. In: *conservation of plant genes II: Utilization of ancient and modern DNA.* Adams, R.P., Miller, J.S., Golenberg, E.M., and Adams, J.E (Eds.), Missouri Botanical Garden Press, St. Louis Chapter II.

Doyle, J.J. and Doyle, J.L. 1990. Isolation of plant DNA from fresh tissue. *FOCUS* 12:13 - 15.

Eminger, P. 1994. Intercropping of cassava and sweet potato in the coconut based farming systems of Dar-es-Salaam, Pwani & Tanga regions. Internal Report of the National Coconut Development Programme.

FAOSTAT. 1999. Statistical database. Food and Agricultural Organisation. Rome, Italy.

Jacard, P. 1908. Nouvelles recherches sur la distribution floralle. *Bull Soc. Vaud. Sci. Nat* 44:223-270.

Jeremiah, S.C., Ndunguru, M.R. and Kibibi, H. 1994. Tanzania root crops research. In: Root crops for food security in Africa. Akoroda, M.O (Ed.). ISTRC-AB, CTA, IITA. pp. 425-426.

Johns, M.A., Skroch, P.W., Nienhuis, J., Hinrichsen, P., Bascur, G. and Munoz-Schick, C. 1997. Genepool classification of common bean landraces from Chile based on RAPD and morphological data. *Crop Science* 37:605-613.

Jones, M.A. Skroch, P.W., Nienhuis, J., Hinrichnsen, P., Bascur, G. and Munoz-Schick, C. 1997. Genepool classification of common bean landraces from Chile based on RAPD and morphological data. *Crop Science* 37:605-613.

Jones, W.O. 1959. Manioc in Africa. Stanford University Press, Stanford, CA, USA. .

Kapinga R., de Steenhuisen-Piters, B., Kajiru, S. Shwagara, D., Rugutu, C., and Mahungu, N. 1997. Selection of cassava varieties by farmers in the Lake Zone of Tanzania. *African Journal of Root and Tuber Crops* 2:248-253.

Karp, A., Kresovich, S., Bhat, K.V., Ayad, W.G. and Hodgkin, T. 1997. Molecular tools in plant genetic resources conservation: a guide to the technologies. IPGRI technical bulletins, International Plant Genetic Resources Institute, Rome, Italy.

Laminski, S., Robinson, E.R., and Gray, V.M. 1997. Application of molecular markets to describe South African elite cassava cultivars. *African Journal of Root Tuber Crops* 2:132-134.

Lecerteau, E., Robert, T., Petiard, V. and Crouzillant, D. 1997. Evaluation of the extent of genetic variability among Theobrma cacao accessions using RAPD and RFLP markers.

Mahungu, N.M. 1998.Cassava germplasm enhancement in Southern Africa. In: *Food Security and Crop Diversification in SADC Countries: The Role of Cassava and Sweetpotato.* Akoroda, M.O. and Teri, J.M. (Eds.), pp. 100 - 109. Proceeding Workshop, Lusaka 17 - 19 August 1998.

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

Mignouna, H.D. and Dixon, G.O. 1997. Genetic relationship among cassava with varying levels of resistance to African mosaic disease using RAPD markers. *African Journal of Root Tuber Crops* 2:28-32.

Morgan, W.T.W. 1973. East Africa. Longman, London, U.K.

N'goran, J.A.K., Laurent, B., Resterucci, A.M. and Lanaud, C. 1994. Comparative genetic diversity studies of Theobroma cacoa L. using RFLP and RAPD markers. *Heredity* 73:589-597.

Ng, S.Y.C. and Ng, N.Q. 1997. Cassava in vitro germplasm management at the International Institute of Tropical Agriculture. *African Journal of Root Tuber Crops* 2:232-234.

Nweke, F.I., Dixon, A.G.O., Asiedu, R. and Folayan, S.A. 1994. Attributes of cassava varieties by farmers in sub-Saharan Africa. In: *Root Crops for Food Security in Africa.* Akoroda, M.O. (Ed.), pp. 65-92. ISTRC - AB, CTA, IITA.

Ogbe, F.O., Legg, J., Raya, M.D., Muimba-Kankalongo, A.M. Theu, M.P., Kaitisha, G., Pkiri, N.A. and Chalwe, A. 1997. Diagnostic

survey of cassava mosaic viruses in Tanzania, Malawi and Zambia. *Roots* 4:12-17.

Rohde, W. 1996. Inverse sequence-tagged repeat (ISTR) analysis, a novel and universal PCR-based technique for genome analysis in the plant and animal kingdom. *Journal of Genetics and Breeding* 50:249-261.

Rohlfs, W. 1989. NTSYS-pc Numerical Taxonomy and Multivariate Analysis System. Exeter Publications, Setauket, N.Y.)

Schaal, B., Carvalho, L.J.C.B., Prinzie, T., Olsen, K., Hernandez, M., Cabral, G. and Moller, D. 1997. Phylogenetic relationships and genetic diversity in Manihot species. *African Journal of Root Tuber Crops* 2:147-149.

Shah, H., Rashid, O., Simons, A.J. and Dunsdon, A. 1994. The utility of RAPD markers for the determination of genetic variation in oil palm *(Elaeis guinensis).*

Sokal, R. and Michener, C.D. 1958. A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* 38:1409-1438.

Swanson, M.M. and Harrison, B.D. 1994. Properties, relationships and distribution of cassava mosaic geminiviruses. *Tropical Science* 34:15-25.

Thormann, C.E., Ferreira M.E., Camargo, L.E.A., Tivang, J.G. and Osborn, T.C. 1994. Comparison of RFLP and RAPD markers to estimating genetic relationship within and among cruciferous species. *Theoretical Application Genetics* 88:973-980.

Thro, A.M., Lynam, J.K. Msabaha, M., Kulembeka, H., Shengero, W., Kapande, A., Mlingi, N., Hemed, L., Digges, P. and Cropley, J. 1996. From village concerns to national priorities: using the results of a case study with cassava farmers in the Tanzanian Lake Zone. In: Komen, J., Cohen, J.I., and Ofir, Z. (Eds,). Turning priorities into feasible programs: Proceedings of a seminar on planning, priorities and policies for agricultural biotechnology. South Africa, 23-28 April, 1995. The Hague/Pretoria: Intermediary Biotechnology Service/Foundation for Research Development.

Vellekoop, P., Butjer, J.B., Maas, J.W. and van Brederode, J. 1996. Can the spread of agriculture in Europe be followed by tracing the spread of the weed Silene latifolia. A RAPD study. *Theoretical Application Genetics* 92:1085-1090.

Williams, J.G.K., Kubelik, A.R., Ivak, K.J., Rafalski, J.A. and Tingy, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 6531-6535.

Yu, L. X and Nguyen, H.T. 1994. Genetic variation detected with RAPD marker among upland and lowland rice cultivars (*Oryza sativa* L.). *Theoretical Application Genetics* 87:668-672.