Theor Appl Genet (2003) 107:1083–1093
DOI 10.1007/s00122-003-1348-3

M. A. Fregene · M. Suarez · J. Mkumbira · H. Kulembeka · E. Ndedya · A. Kulaya · S. Mitchel · U. Gullberg · H. Rosling · A. G. O. Dixon · R. Dean · S. Kresovich

Simple sequence repeat marker diversity in cassava landraces: genetic diversity and differentiation in an asexually propagated crop

Received: 27 January 2003 / Accepted: 23 May 2003 / Published online: 11 July 2003

Abstract Cassava (*Manihot esculenta*) is an allogamous, vegetatively propagated, Neotropical crop that is also widely grown in tropical Africa and Southeast Asia. To elucidate genetic diversity and differentiation in the crop's primary and secondary centers of diversity, and the forces shaping them, SSR marker variation was assessed at 67 loci in 283 accessions of cassava landraces from Africa (Tanzania and Nigeria) and the Neotropics (Brazil, Colombia, Peru, Venezuela, Guatemala, Mexico and Argentina). Average gene diversity (i.e., genetic diversity) was high in all countries, with an average heterozygosity of 0.5358 ± 0.1184. Although the highest was found in Brazilian and Colombian accessions, genetic diversity in Neotropical and African materials is comparable. Despite the low level of differentiation [F_{st}(theta) = 0.091 ± 0.005] found among country samples, sufficient genetic distance (1-proportion of shared alleles) existed between individual genotypes to separate African from Neotropical accessions and to reveal a more pronounced substructure in the African landraces. Forces shaping differences in allele frequency at SSR loci and possibly counterbalancing successive founder effects involve probably spontaneous recombination, as assessed by parent-offspring relationships, and farmer-selection for adaptation.

Keywords Cassava · SSR markers · Genetic diversity · Genetic differentiation · Parent-offspring analysis

Communicated by H.C. Becker

M. A. Fregene (✉) · M. Suarez
Centro Internacional de Agricultura Tropical (CIAT),
Cali, Colombia
e-mail: M.Fregene@cgiar.org

J. Mkumbira · U. Gullberg
Swedish University of Agricultural Sciences (SLU),
Uppsala, Sweden

H. Kulembeka
Agricultural Research Institute (ARI), Ukiguru, Mwanza, Tanzania

E. Ndedya
Rural Integrated Project Support (RIPS) Programme,
Mtwara, Tanzania

A. Kulaya
Agricultural Research Institute (ARI), Mikocheni,
Dar es Salaam, Tanzania

H. Rosling
Division of International Health,
Department of Public Health Sciences, Karolinska Institute,
Stockholm, Sweden

S. Mitchel · A. G. O. Dixon
International Institute of Tropical Agriculture (IITA),
Ibadan, Nigeria

S. Kresovich
Institute for Genomic Diversity (IGD), Cornell University,
Ithaca, NY, USA

R. Dean
Plant Genetic Research Consortium,
University of Georgia at Griffin, USA

Introduction

The starchy root crop cassava (*Manihot esculenta* Crantz) is grown in most tropical countries of the world, ranking fourth in importance as a dietary staple and sixth on an overall basis (Cock 1985). In sub-Saharan Africa, average consumption exceeds 300 kg per person per year in some countries, and the region is the world's largest producer of cassava with a production of 90 million metric tons in 1999 (FAO 2000). Cassava was introduced into Africa, arriving at the western and eastern coasts by Portuguese slave ships from Brazil, during the 1500s until the 1800s (Jones 1969). It then spread along trade routes into the interior of the continent. Because of its excellent adaptability to erratic rainfall and low-fertility soils, it became a major dietary staple, a famine-reserve crop and a source of cash. In the New World, it had been established within the last 10,000 years and, even today, continues to be extensively cultivated from Meso-America to Argentina, with Brazil, Colombia and Paraguay being the principal

producers (Rogers 1963; Pearsall 1992; Sauer 1993). Cassava was not introduced into Asia until the 18th century but, by the 19th century it was firmly established in South and Southeast Asia (Cock 1985).

Cassava probably originated in wild *M. esculenta* populations growing along the southern rim of the Amazon Basin in Brazil. Olsen and Schaal (1999) demonstrated that cassava haplotypes of the gene glyceraldehyde-3-phosphate dehydrogenase (*G3pdh*) are present in these natural populations of wild *M. esculenta* subspecies, but absent in populations from other regions of Brazil. The savannas of Colombia, Venezuela, Guatemala and southern Mexico had earlier been proposed as likely places of origin due to the large number of landraces present there (Sauer 1952; Rogers 1965). But the *Manihot* species found in these locations are only distantly related to cassava (Fregene et al. 1994; Schaal et al. 1994; Roa et al. 1997).

Given the "out of Brazil" theory for the spread of cassava to the rest of the world, many founder events should have occurred, with the concomitant effect of reduced diversity and increased genetic differentiation. Furthermore, the primary method of propagation – vegetative, or asexual, propagation – would be expected to lead to further reduction in genetic diversity over time because of the accumulation of systemic pathogens and the spread of a few, vigorous, well-adapted landraces with a capacity to produce many stakes for planting. Nonetheless, levels of molecular marker diversity observed for cassava landraces from Africa and several Neotropical countries are comparable with those from Brazil (Beeching et al. 1993; Fregene et al. 2000).

However, the traditional farming system of slash and burn, followed by 3 to 15 years of fallow, as practiced by small farmers, by far the largest producers of cassava in the Neotropics and Africa, and the allogamous nature of cassava produce a large pool of volunteer seedlings. Natural and human selection acts on these to produce new varieties that maintain a high level of genetic diversity (Doyle et al. 2001). An elevated level of inter- and intravarietal diversity is also known to exist in farm fields; for example, the diversity found in a single field of a Makushi Amerindian community in southern Guyana was shown to equal that of the core-of-core collection of 38 accessions representing the world's cassava collection held at the International Center for Tropical Agriculture (CIAT, its Spanish acronym), Cali, Colombia (Elias et al. 2000).

Variation in allele frequency at many unlinked loci is the preferred way to assess genetic diversity and differentiation, and to estimate the strengths of the various forces shaping them. A study of unlinked loci, particularly the simple sequence repeat (SSR) marker loci, from cassava landraces across the three continents would shed more light on the dynamics of genetic diversity and differentiation, the latter that may represent heterotic pools. High levels of genetic differentiation, potentially representing heterotic pools, have been described for maize (Shull 1952; Tomes 1998) and robusta coffee (Leroy et al. 1993).

Simple sequence repeat (SSR) markers are particularly attractive to study because they are abundant in plant and animal genomes, they have high levels of polymorphism, and are adaptable to automation (Donini et al. 1998). In cassava, SSR markers have been used to search for duplicates in the CIAT core collection (Chavarriaga-Aguirre et al. 1999) and to analyze variation in natural populations of putative progenitors of cassava (Olsen and Schaal 2001). We describe our SSR marker assessment, at 67 marker loci, of cassava landraces from the primary and African centers of diversity. We also examine the genetic diversity and differentiation derived from the SSR data and the forces shaping differences in allele frequencies. This study of the organization of genetic diversity is expected to benefit cassava germplasm conservation and enhancement, and contribute to the elucidation of forces that shape genetic differentiation in an asexually propagated allogamous crop.

Materials and methods

Plant materials

Cassava landraces were collected in June 1999, from the districts of Mtwara, Newala, Masasi, Tandahimba and Nachingwea in southeastern Tanzania, situated between 10° and 12°S and 30° and 40°E. This region is important in terms of the view of genetic diversity, because the coastal town of Mtwara was a key port in the transatlantic slave trade (Jones 1969). Cassava has remained an important staple crop in the coastal area and remote surrounding plateau since its arrival in the early 18th century. The forces affecting genetic differentiation in southeastern Tanzania over the last 300 years under representative conditions (in small farmers' fields) can be assessed here. Eleven villages were visited, they lie between latitude 811.9 and 854.7, longitude 465.2 and 627.5, and altitude 90 and 600 m above sea level.

Woody stakes, 20 to 30-cm long, were obtained from all genotypes, a total of 96, and planted in 20-l pots in the greenhouse at the ARI at Kibaha. Another 23 and 56 landraces were included from the cassava germplasm banks at ARI – Naliendele, a subregional collection from southeastern Tanzania, and at ARI – Kibaha, the national collection, respectively. A summary of the plant materials, their source, morphological, and taste characteristics of the genotypes collected can be viewed at http://www.ciat.cgiar.org/molcas/appendix1.jsp?cod_est=3. A number of genotypes collected from farmers fields had poor germination and were eliminated; only 84 of the 96 genotypes from the collection, were included in subsequent analysis.

The world germplasm collection held at CIAT has 5,724 cassava landraces from 24 Neotropical and Asian countries. The largest numbers of accessions are from Colombia (2,003 accessions) and Brazil (1,340 accessions). A core collection of about 634 genotypes representing the entire collection in terms of country, agroecology and isozyme diversity had earlier been made (Hershey et al. 1992). This collection was recently evaluated with four SSR markers (Chavariagga-Aguirre et al. 1999). Based on this study, we sampled 76 landraces from Brazil, Colombia, Peru, Venezuela, Guatemala, Mexico and Argentina by selecting 12% of accessions representing the diversity in each country to give a manageable number for SSR analysis. Fifteen of the best parents in terms of general combining ability (GCA), with average performance when crossed with other individuals, were included from the CIAT breeding program to compare their genetic diversity with that of the landraces.

Also included were 19 Nigerian landraces from the international collection at the International Institute of Tropical Agriculture

(IITA) and ten improved genotypes from the Institute's cassava breeding program. Four of the ten genotypes came from the 1950s breeding efforts at the Moor Plantation Experiment Station, Ibadan, Nigeria; one genotype, 58,308, has been used extensively as the donor parent for resistance to the cassava mosaic disease (CMD) in the IITA program.

In all, 283 accessions were used, 84 coming from Tanzanian farms, 79 from local germplasm banks in Tanzania, and 120 from holdings at CIAT and IITA. Fourteen groups or samples, based on the country of origin, were created to study genetic diversity and differentiation within each country and among countries (see Table 1).

Simple sequence repeat marker analysis

DNA isolation was carried out according to Dellaporta et al. (1983), using 3–5 g of fresh young leaf tissue obtained from 3 to 4-week-old plants. Between 500 μg and 1 mg of high quality DNA was obtained from each extraction and quantified by fluorometry. A subset of 92 SSR markers, with broad coverage of the genome, was selected from the 186 SSR markers developed at CIAT (Chavariaga-Aguirre et al. 1998; Mba et al. 2001), and organized into 23 quadriplexes (multiplexes of four markers). The quadriplexes were designed by searching for sets of four markers, from the pool of 92 SSR markers, whose primers do not form heteroduplexes at the 3′ end, using the computer package DNAMAN (Lynon Biosoft Inc, USA). From the initial 23 quadriplexes, 12 quadriplexes, 8 triplexes (multiplexes of three SSR markers) and 5 duplexes (multiplexes of two SSR markers), 82 markers, were found to adequately amplify DNA from two test cassava genotypes in PCR multiplex reactions.

A new total of 82 SSR markers were chosen for the PCR multiplexing of the entire Tanzanian collection, and the CIAT genotypes. The primers were multiplexed on all the 283 genotypes, using fluorescently labeled primer pairs. The PCR product was denatured and electrophoresced on 4% polyacrylamide gels, using an automated DNA sequencer ABI model 377 (Perkin Elmer Inc.). The raw gel data were extracted, using the GENESCAN package of ABI PRISM and GENOTYPER software (Perkin Elmer Inc.). The extracted data were exported as allele sizes to Microsoft Excel (Microsoft Inc.) for further formatting as input files for statistical analysis. A strictly diallelic model of inheritance was adhered to; in other words, markers having three or more alleles were eliminated. Fifteen SSR markers yielded complex patterns, i.e. more than two alleles per genotype, and were eliminated from the final analysis.

Statistical analysis of SSR data

Gene diversity and genetic differentiation analyses were carried out with data from 67 SSR markers, chosen for their diallelic nature, clear patterns and broad coverage of the cassava genome. Table 2 reveals the map position (where available) of the SSR markers. Genetic diversity within and among accessions was estimated, using the software package GEN-SURVEY (Vekemans and Lefebvre 1997), with the following statistics: percentage of polymorphic loci, mean number of alleles per polymorphic locus, average observed heterozygosity (H_o), and average gene diversity (H_e) (Nei 1978). For all loci and for all accessions, total heterozygosity (H_t) and the proportion of among-accession differentiation (G_{st}) were estimated according to Nei (1978). Standard deviations for the above parameters were estimated over loci and sampled by jackknifing (Quenoille 1956; Efron 1982) using 200 replications. Given the small evolutionary divergence times for the accessions, the infinite allele model (IAM) (Kimura and Crow 1964) was assumed for all calculations.

Genetic differentiation was quantified by the F-statistics estimator (F_{st}, theta) (Wright 1951), as described by Weir and Cockerham (1984), using FSTAT 2.9 (Goudet 1995), and by G_{st} (Nei 1978). F_{st} gives a similar estimate of genetic differentiation as G_{st}, but G_{st} takes into account variation in sample sizes, as is necessary in this study. F_{st} values were estimated per allele, per locus and overall. FSTAT performs bootstrapping (Efron 1982) over loci and, given the large number of unlinked SSR loci employed in this study, provides rigorous testing of hypotheses of genetic differentiation. Confidence intervals were calculated per locus over samples, and over loci by jackknifing, using 200 replications, and by bootstrapping, 1,000 bootstraps, over loci. Pairwise values of F_{st} between samples (landrace groups) were also estimated. The pairwise matrix was analyzed by cluster analysis, using the unweighted pair-group method with arithmetic averaging of NTSYS-PC (Rohlf 1993).

Pairwise genetic distances between individual accessions were calculated from the raw allele size-data based on the 1-proportion of shared alleles (PSA) (Bowcock et al. 1994), using the computer program "microsat", developed by Eric Minch (http://hpgl.stanford.edu/projects/microsat/microsat.html). Distances based on 1-PSA give a more exact representation of genetic relationships when using closely related genotypes and SSR markers, compared with other distance estimates (Takezaki and Nei 1996; Bertin et al. 2001). The distance matrix generated from "microsat" was subjected to principal component analysis (PCA), using the program JMP (SAS Institute 1995) to deduce multivariate relationships among the cassava genotypes. The first and second principal components were then used to draw a scatter plot.

To assess the extent to which spontaneous recombination played a part in the evolution of genetic diversity, parent-offspring relationships were sought in the SSR data obtained from the African samples, using the methods described by Thompson (1975) and Meagher (1986). The computer program CERVUS (Marshall et al. 1998) was used for the analysis. Assuming a Hardy-Weinberg equilibrium, CERVUS simulates a maternal and paternal genotype from allele frequencies observed in the study population, and derives an offspring-genotype by Mendelian sampling of the parental alleles. The simulation also alters the genotypic data to reflect the existence of unsampled males, missing loci, and incorrectly typed loci according to the values of the simulation parameters. Next, each candidate parent is considered in turn as the alleged father, and the log of odds (LOD) ratio scores are calculated for all males for whom genetic data exist. Once all males have been considered, the most-likely and second most-likely males are identified, and the delta score (difference in LOD scores) calculated. The delta value is recorded, together with the status of the most-likely male (i.e., whether or not it is the true father). The final stage of the simulation is to find critical values of delta so that the significance of delta values found in paternity inference in the study population can be tested.

A total of 67 SSR markers were employed in this study to provide a rigorous estimate of genetic diversity and differentiation. This number of markers quickly becomes unrealistic when very large numbers of genotypes are to be analyzed, particularly where resources are limited. An analysis to evaluate the number of markers that will provide sufficient information on allele diversity, i.e. an addition of more markers would not provide an increase in allele diversity, was conducted using the parameter polymorphism-information content (PIC) = $1 - \Sigma Pi^2$, where Pi is the allele frequency of the ith allele. PIC values were calculated for 1, 2, 5, 10, 15, 20, 30, 40 and 50 randomly chosen SSR markers for the whole data set of 283 genotypes. The PIC value was then plotted against the number of SSR markers. The PIC score is expected to show asymptotic behavior when enough markers have been added.

Results

Allelic richness of SSR loci

Genetic variation at 67 unlinked SSR loci was assessed in 283 accessions of cassava landraces grouped into 14 samples according to their country of origin and, in the case of Tanzania, local origin within the country. The number of alleles observed at each locus in the data set

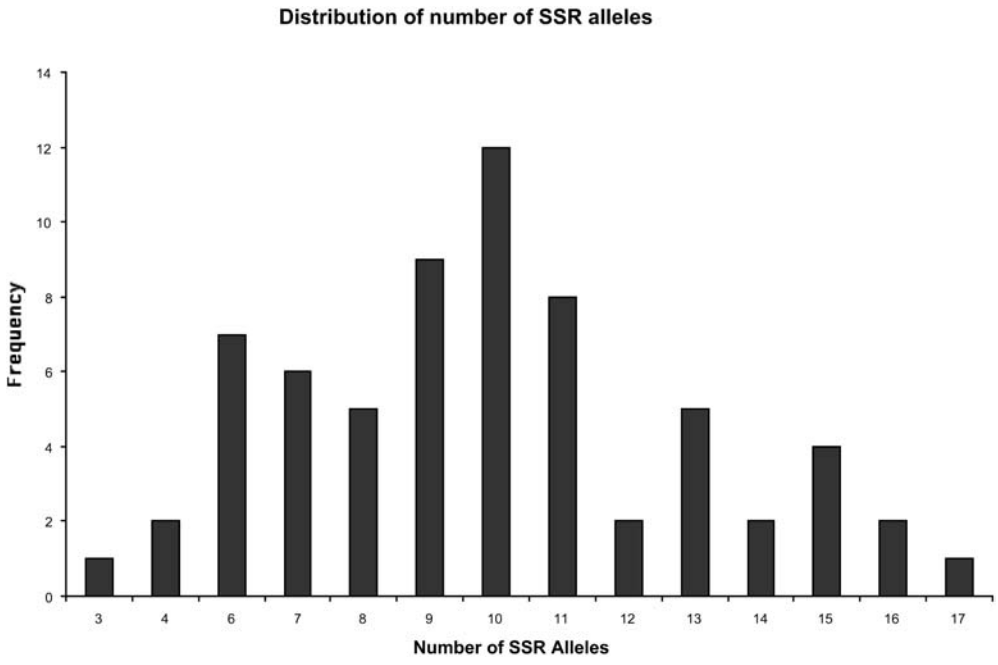

Fig. 1 Number of simple sequence repeat (SSR) alleles per locus and their frequency in cassava landraces from southeastern Tanzania, Nigeria, and some selected Latin American countries

ranged widely from 3 to 17 alleles per locus. The distribution of the number of alleles at each locus over all 67 loci is shown in Fig. 1. No unique alleles with a frequency of more than 25% were found within country samples, except for Guatemala and Nigeria. Accessions from Guatemala had unique alleles for five markers, with frequency ranging between 25% and 100%, and from Nigeria for one marker at 50%.

Genetic diversity

Genetic diversity parameters were calculated from the SSR data within and between country samples (Table 1). Very high genetic polymorphism was found for all the tested SSR loci when using the criterion that the frequency of the most common allele is less than 0.98. On an average, there were 94% ± 1.06% polymorphic loci across all accessions and countries. Brazil and Colombia had the largest number of polymorphic loci, at 100% and 98.5%, respectively. A particularly high number of alleles was found per locus, at an average of 4.03 ± 0.0378, with Colombian accessions having the highest at an average of 6.0 alleles per locus, followed by accessions from the collection made in southeastern Tanzanian (5.3) and Brazil (5.2). The probability that two randomly selected alleles in a given accession are different, estimated by H_e, was, on average, more than half (0.538 ± 0.118), with the highest H_e being found for Colombian accessions, followed by Brazilian accessions. The average proportion of observed heterozygous individuals (H_o) was high (0.5136 ± 0.0495), which confirms cassava's outcrossing and highly heterozygous nature.

Genetic diversity estimates H_e and H_o of cassava landraces from the Neotropics were not significantly different from those from Africa (data not shown). Sampling and sample sizes obviously affected the results presented here but in general these reveal a comparable level of polymorphism in Africa to that found in the primary center of diversity. The overall heterozygosity (H_t) in all accessions was high at 0.6499 ± 0.1595, but only 10% (G_{st} = 0.1075 ± 0.0565) of this was due to differentiation among samples from cassava-producing countries as diverse as Tanzania and Guatemala. Almost two thirds of the diversity was due to within-country variation (mean H_s = 0.5812 ± 0.1470).

Genetic differentiation

The F_{st} (theta) estimators of genetic differentiation, averaged over all loci, were 0.091 ± 0.005 as estimated by jackknifing, and 0.079 ± 0.107 as calculated by bootstrapping, with a confidence interval of 99% (Table 2). These results again reveals a low level of differentiation between country samples as was observed with estimates of G_{st}. Pairwise calculations of F_{st} (theta) over all loci between pairs of country landrace groups provide a picture of germplasm exchange between countries and of which Neotropical countries contributed most germplasm to Africa (Table 3). F_{st} values ranged widely, from 0.2696 between accessions from Nigeria and Guatemala, to a minimum of 0.0076 between accessions collected in Mtwara, southeastern Tanzania, and the cassava germplasm bank at ARI – Naliendele, which serves the same region. The very low genetic differentiation between the last two groups reveals that germplasm at the ARI – Naliendele gene bank is representative of that grown by farmers.

Table 1 Genetic diversity within groups of cassava landraces classified according to the country of origin. Standard deviations (SD) were estimated by jackknifing over loci (200 replications). H_t, H_s, D_{st} and G_{st}[a] are given over loci and over groups (country collections)

Group[b]	Sample size	No. of loci	No. of pol.[c] loci	Percent of pol.[c] loci	Mean no. alleles/ locus	Mean no. alleles/ pol.[c] locus	H_o[d]	H_e[e]	H_{e_p}[f]
Argentina	3	67	57	85.1	2.6	2.9	0.5174	0.4635	0.5672
Brazil	20	67	67	100.0	5.2	5.2	0.5311	0.6129	0.6285
Colombia	32	67	66	98.5	6.0	6.0	0.5012	0.6177	0.6277
GCA	15	67	65	97.0	4.5	4.6	0.5244	0.5754	0.5952
Guatemala	4	67	57	85.1	2.4	2.7	0.4925	0.3960	0.4554
Mexico	5	67	64	95.5	3.6	3.7	0.4915	0.5600	0.6251
Peru	7	66	62	93.9	3.7	3.9	0.4892	0.5596	0.6067
Venezuela	5	66	64	97.0	3.5	3.6	0.4297	0.5692	0.6340
Tanzania-Mtwara	84	67	65	97.0	5.3	5.4	0.5430	0.5580	0.5616
Tanzania-Naliendele	23	67	64	95.5	4.5	4.7	0.5448	0.5545	0.5667
Tanzania-Kibaha	56	67	64	95.5	5.1	5.3	0.5274	0.5334	0.5382
Nigeria	19	66	62	93.9	3.9	4.0	0.5965	0.5296	0.5440
IITA	6	67	61	91.0	3.2	3.4	0.4915	0.4866	0.5340
Moor Plantation	4	67	63	94.0	2.7	2.8	0.5100	0.4852	0.5596
Mean		66.79	62.93	94.23	4.03	4.17	0.5136	0.5358	0.5745
SD		0.43	2.97	4.45	1.11	1.06	0.0378	0.0602	0.0495
	H_t	H_s	D_{st}	G_{st}					
Mean	0.6499	0.5812	0.0687	0.1075					
SD	0.1595	0.1470	0.0318	0.0565					
95% CI	0.6100	0.5463	0.0621	0.0953					
99% CI	0.6871	0.6136	0.0758	0.1195					

[a] H_t = total heterozygosity in the entire data set; H_s = heterozygosity within the country averaged over the entire data set; D_{st} = average gene diversity between populations; G_{st} = coefficient of gene differentiation
[b] GCA = best parents in terms of general combining ability in the CIAT collection; IITA = International Institute of Tropical Agriculture; CI = confidence interval
[c] pol. = polymorphic
[d] H_o = average observed heterozygosity within country
[e] H_e = average expected heterozygosity within country
[f] H_{e_p} = average expected heterozygosity within country corrected for small sample sizes (Nei 1978)

In contrast the high differentiation between Guatemalan and Nigerian landraces suggests limited germplasm exchange between these two countries. Unique alleles were found in accessions from both countries. Pairwise comparison between Guatemalan landraces and those from other countries revealed moderate (with Colombian landraces, 0.1122) to high differentiation (with African germplasm, >0.25). The UPGMA of the pairwise F_{st} estimates resulted in a dendrogram of the landraces that divides the African from the Neotropical accessions (Fig. 2). It also confirms the fact that the most commonly used parents in breeding at CIAT are mostly Colombian landraces.

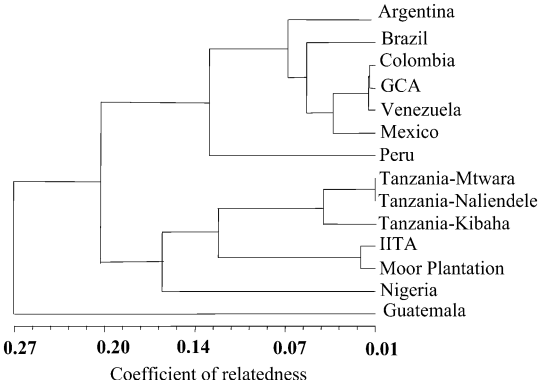

Fig. 2 Unweighted pair-group method with arithmetic averaging (UPGMA) dendogram of the pairwise fixation index (F_{st}) between cassava landraces, grouped by country and by source

Genetic relationships among accessions

To examine genetic relationships among accessions, genetic distances between all pairs of individual accessions were calculated by the 1-proportion of shared alleles (1-PSA) statistic (Bowcock et al. 1994). The genetic distance estimator 1-PSA has been found to be the most reliable for SSR data, assuming the infinite allele model (IAM) (Kimura and Crow 1964) and close relationships (Bertin et al. 2001). Nei's standard distance D (Nei 1978) is not appropriate for SSR data, as many accessions do not share a single allele. To present a graphical representation of genetic relationships between accessions, a principal coordinate analysis (PCA) was performed on the genetic distance matrix, and a scatter plot of the first two principal components (PC) made (Fig. 3).

Table 2 Map location of the 67 SSR loci used in this study, proportion of among accession differentiation (Gst), and genetic differentiation according to F statistics estimator (Fst), estimated - over all country samples for all loci. Mean JK values were obtained by jackknifing over samples (200 replications) and loci, with Means BS by bootstrapping over loci (1000 bootstraps)

Locus	Genetic Map location[a]	Gst	F_{st}	Locus	Genetic Map location[a]	Gst	F_{st}
SSRY169	D	0.102	0.056 ± 0.013	SSRY82	K	0.145	0.080 ± 0.017
SSRY5	J	0.131	0.046 ± 0.030	SSRY52	H	0.126	0.031 ± 0.019
SSRY183	C	0.546	0.535 ± 0.114	SSRY65	nd	0.217	0.060 ± 0.030
SSRY50	B	0.147	0.085 ± 0.030	SSRY119	nd	0.179	0.071 ± 0.036
SSRY34	UMA	0.154	0.086 ± 0.031	SSRY101	J	0.186	0.102 ± 0.035
SSRY144	nd	0.117	0.063 ± 0.028	SSRY94	nd	0.138	0.079 ± 0.040
SSRY28	UMB	0.197	0.073 ± 0.019	SSRY80	nd	0.137	0.100 ± 0.053
SSRY61	nd	0.185	0.101 ± 0.024	SSRY87	nd	0.100	0.048 ± 0.022
SSRY27	nd	0.206	0.080 ± 0.046	SSRY153	nd	0.150	0.046 ± 0.019
SSRY89	nd	0.141	0.064 ± 0.027	SSRY106	G	0.127	0.084 ± 0.037
SSRY164	H	0.150	0.059 ± 0.028	SSRY49	K	0.179	0.090 ± 0.040
SSRY9	D	0.116	0.055 ± 0.010	SSRY26	UMB	0.132	0.097 ± 0.021
SSRY114	nd	0.148	0.041 ± 0.028	SSRY105	U	0.146	0.103 ± 0.062
SSRY182	UMA	0.168	0.099 ± 0.028	SSRY177	U	0.143	0.113 ± 0.027
SSRY38	G	0.141	0.073 ± 0.035	SSRY170	U	0.227	0.159 ± 0.091
SSRY21	B	0.155	0.100 ± 0.029	SSRY180	E	0.200	0.141 ± 0.062
SSRY69	nd	0.111	0.038 ± 0.015	SSRY142	nd	0.154	0.059 ± 0.030
SSRY29	nd	0.115	0.090 ± 0.031	SSRY161	E	0.209	0.119 ± 0.026
SSRY35	nd	0.437	0.462 ± 0.165	SSRY62	nd	0.163	0.079 ± 0.029
SSRY59	M	0.160	0.074 ± 0.036	SSRY148	nd	0.146	0.047 ± 0.019
SSRY19	V	0.144	0.046 ± 0.020	SSRY155	nd	0.152	0.095 ± 0.030
SSRY47	J	0.315	0.123 ± 0.041	SSRY151	nd	0.159	0.060 ± 0.031
SSRY171	C	0.180	0.078 ± 0.030	SSRY154	nd	0.253	0.151 ± 0.053
SSRY123	nd	0.206	0.063 ± 0.029				
SSRY64	nd	0.169	0.039 ± 0.025		Mean JK	0.175 ± 0.72	0.091 ± 0.005
SSRY185	X	0.279	0.178 ± 0.072		Mean BS (CI 95%)[a]	0.16	0.082 ± 0.103
SSRY75	nd	0.136	0.086 ± 0.037		Mean BS (CI 99%)[a]	0.196	0.079 ± 0.107
SSRY175	K	0.138	0.089 ± 0.016				
SSRY181	nd	0.150	0.074 ± 0.026				
SSRY110	X	0.134	0.024 ± 0.021				
SSRY93	nd	0.177	0.074 ± 0.031				
SSRY70	nd	0.162	0.052 ± 0.018				
SSRY68	P	0.114	0.039 ± 0.021				
SSRY20	X	0.155	0.066 ± 0.026				
SSRY108	D	0.097	0.075 ± 0.038				
SSRY135	G	0.202	0.145 ± 0.063				
SSRY85	K	0.323	0.202 ± 0.101				
SSRY79	nd	0.139	0.084 ± 0.040				
SSRY141	nd	0.235	0.127 ± 0.078				
SSRY32	C	0.177	0.081 ± 0.057				
SSRY100	L	0.140	0.057 ± 0.023				
SSRY12	H	0.162	0.043 ± 0.026				
SSRY51	N	0.209	0.076 ± 0.045				
SSRY145	N	0.207	0.073 ± 0.027				

[a] genetic map location = reference, Mba et al. 2001
[b] nd: no linkage data
[c] CI = confidence interval

Principal components 1 and 2 accounted for 36% and 28% of the total variance, respectively. The PCA roughly splits the accessions into African and Neotropical on the PC1 axis, as was observed for the UPGMA of F_{st}, with a considerable overlap. On the PC2 axis, however, the PCA divides the Tanzanian accessions into two groups. An attempt was made to search for an underlying basis for the two PC2 groups, using available data such as taste (bitter or non-bitter, according to the presence of cyanogenic glucosides) and source (villages, research centers), but the two clusters did not correspond to the above (data not shown). All villages had accessions from both clusters, and although more bitter accessions were found in the lower group, both groups had a mix of bitter and non-bitter. An interesting observation from both clusters of the Tanzanian accessions is that there were no accessions in the lower cluster from Tanzania in the ARI-Kibaha national collection. This collection was assembled largely from other parts of the country (H. Kulembeka 1999, personal communication). Accessions from ARI-Naliendele are, however, represented in the lower group, suggesting that this group comprises germplasm unique to southeastern Tanzania and which might have been selected by farmers particularly for adaptation to local conditions.

Table 3 Pairwise estimator of F_{st} (theta) between pairs of country groupings of cassava landraces[a]

Population	Argentina	Brazil	Colombia	GCA	Guatemala	Mexico	Peru	Venezuela	Tanzania-Mtwara	Tanzania-Naliendele	Tanzania-Kibaha	Nigeria	IITA	Moor
Argentina	0	0.0632	0.0708	0.0680	0.2364	0.0429	0.1279	0.0283	0.1453	0.1270	0.1654	0.2067	0.1447	0.1631
Brazil	0.0632	0	0.0572	0.0442	0.1454	0.0397	0.1119	0.0297	0.0791	0.0787	0.1150	0.1076	0.1296	0.1054
Colombia	0.0708	0.0572	0	0.0114	0.1122	0.0379	0.0782	0.0123	0.0922	0.0827	0.1290	0.1290	0.1237	0.1137
GCA	0.0680	0.0442	0.0114	0	0.1227	0.0283	0.0967	0.0119	0.0809	0.0682	0.1077	0.1388	0.1275	0.1117
Guatemala	0.2364	0.1454	0.1122	0.1227	0	0.1468	0.1914	0.1205	0.1682	0.1638	0.2103	0.2696	0.2689	0.2506
Mexico	0.0429	0.0397	0.0379	0.0283	0.1468	0	0.0853	0.0488	0.0984	0.0820	0.1185	0.1405	0.1170	0.1118
Peru	0.1279	0.1119	0.0782	0.0967	0.1914	0.0853	0	0.0538	0.1417	0.1212	0.1529	0.1877	0.1687	0.1612
Venezuela	0.0283	0.0297	0.0123	0.0119	0.1205	0.059	0.0538	0	0.0567	0.0476	0.0839	0.1147	0.0795	0.0593
Tanzania-Mtwara	0.1453	0.0791	0.0922	0.0809	0.1682	0.0984	0.1417	0.0567	0	0.0076	0.0452	0.1402	0.1180	0.1081
Tanzania-Naliendele	0.1270	0.0787	0.0827	0.0682	0.1638	0.0820	0.1212	0.0476	0.0076	0	0.0097	0.1358	0.0888	0.0919
Tanzania-Kibaha	0.1654	0.1150	0.1290	0.1077	0.2103	0.1185	0.1529	0.0839	0.0452	0.0097	0	0.1625	0.1121	0.1215
Nigeria	0.2067	0.1076	0.1290	0.1388	0.2696	0.1405	0.1877	0.1147	0.1402	0.1358	0.1625	0	0.1605	0.1142
IITA	0.1447	0.1296	0.1237	0.1275	0.2689	0.1170	0.1687	0.0795	0.1180	0.0888	0.1121	0.1605	0	0.0182
Moor	0.1631	0.1054	0.1137	0.1117	0.2506	0.1118	0.1612	0.0593	0.1081	0.0919	0.1215	0.1142	0.0182	0

[a] GCA = best parents in terms of general combining ability in CIAT collection; IITA = International Institute of Tropical Agriculture; Moor = Moor Plantation, Ibadan

Estimating spontaneous recombination in farm fields in Southeastern Tanzania

To estimate if spontaneous sexual recombination had contributed significantly to the genetic structure of cassava in Tanzania, parent-offspring relationships were investigated within accessions collected there. A total of 51 significant parent-offspring relationships were found among the 84 collected accessions, using a delta threshold level of 1.0. Results of the parent-offspring analyses can be viewed at the URL http://www.ciat.cgiar.org/molcas/appendix2.jsp?cod_est=3

Analysis of parent-offspring relationships is confounded by the presence of closely related offspring among the accessions collected. The delta statistic (Marshall et al. 1998), calculated by CERVUS, compares LOD scores of the two best putative parents to reduce the confounding effects of full- or half-sibs. Results successfully identified a known parent of TMS 30,572, itself an improved line from IITA that was included as an internal control. Genotype 58,308 from the Moor Plantation breeding program of the 1950s had served as a parental source of resistance to cassava mosaic disease (CMD) for TMS 30,572.

Determination of an optimum number of SSR markers for an assessment of genetic diversity in cassava

A plot of polymorphism information content (PIC) against the number of SSR marker loci yielded an asymptotic curve (Fig. 4). The plot reveals that little or no additional increase in PIC is obtainable with more than 30 SSR markers. This analysis shows that 67 SSR loci is more than enough to assess genetic diversity in cassava land races and that a minimum of 30 unlinked markers is required to obtain the maximum amount of information on allelic diversity.

Discussion

The large number (67) of unlinked SSR loci employed in this study enabled a rigorous estimation, not previously carried out, of genetic differentiation and diversity in cassava landraces from the primary and African centers of diversity. The reliability of estimates for genetic variation, such as H_e, H_o, F_{st}, and genetic distances, depends more on the number of loci than on the number of individuals sampled (Baverstock and Moritz 1996). Estimates of genetic differentiation ranged widely from locus to locus, underscoring the danger of assessing SSR diversity with a small set of SSR markers. For example, the genetic diversity of maize as a subset of the diversity found in its teosinte progenitors varies from 25% to 75%, depending on what location of the genome the diversity analysis was based on (Eyre-Walker et al. 1998). PIC values also increased considerably with an increase in marker loci below a total of 30 markers.

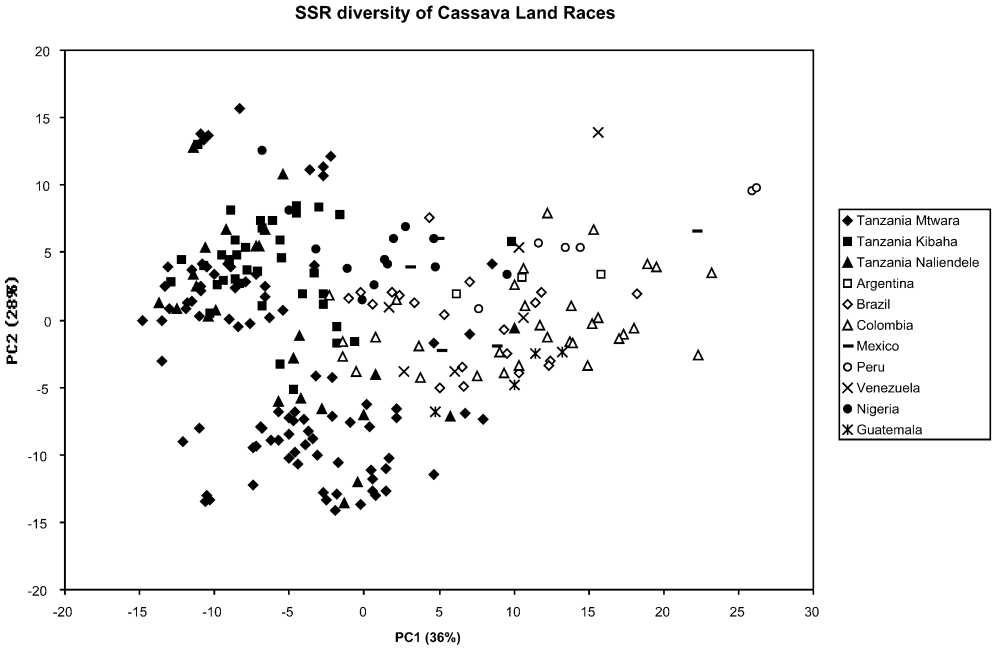

Fig. 3 Scatter plot of simple sequence repeat (SSR) marker diversity in cassava accessions was derived from plotting the first two principal components (PC1 and PC2), which were computed from the genetic distance matrix (1-proportion of shared alleles)

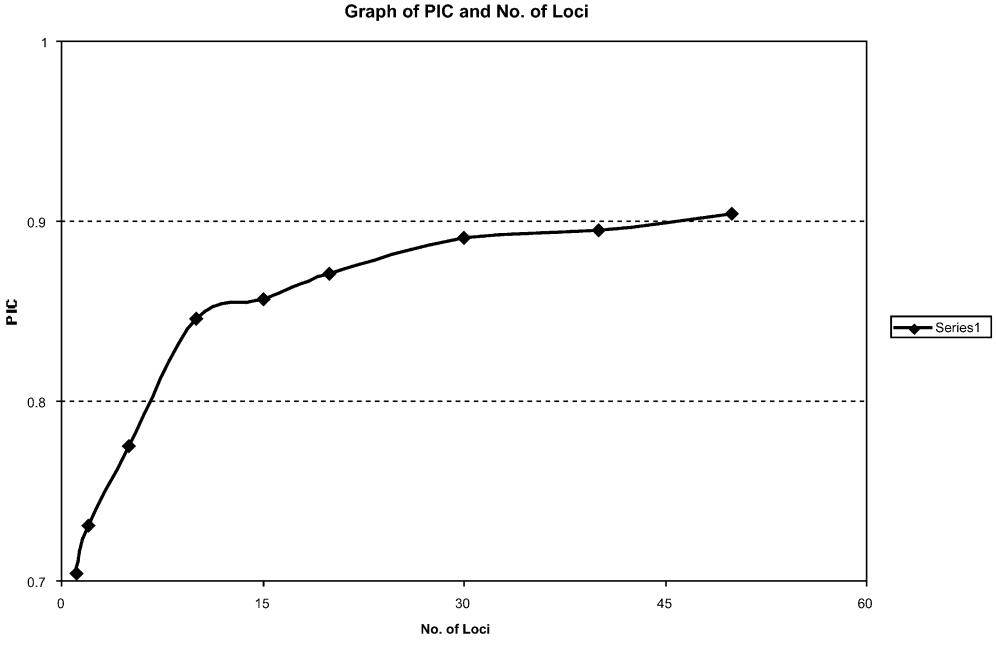

Fig. 4 Plot of polymorphism information content (PIC) against the number of loci

Cassava was probably domesticated from populations of *M. esculenta* ssp. *flabellifolia* along the southern rim of the Amazon Basin within the last 10,000 years ago (Olsen and Schaal 2001). Its expansion into other regions of Neotropics, Africa and Asia could have led to a founder effect of reduced diversity and an increase in genetic differentiation. The high genetic diversity of cassava landraces in all countries is therefore unexpected. However, the agricultural practices of slash and burn by Amerindian farmers and the allogamous nature of cassava have been demonstrated to produce a large pool of volunteer seedlings that natural and human selection can act on to produce new varieties (Doyle et al. 2001).

These observations agree with this study's findings, in which half of the genotypes collected in 11 villages of southeastern Tanzania had putative parent-offspring relationships with other accessions from the same area but not with those from other regions within the country or outside it. Furthermore, this study revealed that spontaneous recombination and farmer selection from the volunteer seedlings of new varieties is also occurring in Africa. The overall effect of spontaneous recombination in the farmers' field observed in this study is the

maintenance of high levels of genetic diversity. The production of new varieties not only maintains a high level of genetic diversity but also serves as insurance against crop failure due to biotic and abiotic stresses. Subsistence cassava farmers in Africa and Latin America may have more than 17 distinct varieties in a single field with a high level of inter- and intra-varietal diversity (Elias et al. 2000; M.A. Fregene 1999, personal observation).

The differences in allele frequencies seen among landraces in this study are probably due to genetic drift effects subsequent to mutation, since "divergence" among crop accessions occurred within the last 10,000 years (Sauer 1952; Rogers 1965). The average genetic differentiation (0.091 ± 0.005; Table 3) found among the country accessions conforms with the above assumptions. The high level of genetic differentiation between landraces from Guatemala and Africa [Nigeria and the ARI – Kibaha, Tanzanian national collection, pairwise F_{st} values of 0.2696 and 0.2103 (Table 3), respectively] is therefore unusual.

Also unexpected is the higher than expected average genetic differentiation between the Guatemalan and Neotropical samples (0.156), compared with the overall average genetic differentiation within the Neotropics (0.063). The small size of the Guatemalan sample could bias the estimate of F_{st}, but the percentage of unique alleles (8%) found for Guatemalan varieties makes bias from the small-sample-size unlikely. The principal argument to explain this peculiarity is that Guatemalan landraces represent an independent domestication event or have introgressed with different wild relatives; the lowlands of Guatemala have been suggested as a likely domestication site for cassava (Rogers 1963, 1965). Even though *Manihot* species present in Guatemala and Meso-America were shown as distantly related to cassava (Fregene et al. 1994; Roa et al. 1997), Rogers' explanation cannot be completely ruled out because progenitor species could have become extinct.

The highly differentiated landraces from Guatemala and Africa may represent heterotic pools, like those for maize (Shull 1952). A principal reason for this study was to assess genetic diversity in cassava landraces as a first step to delineating heterotic pools for a more systematic improvement of combining ability via recurrent reciprocal selection (Keeratinijakal and Lamkey 1993). The heterotic patterns found in maize populations at the turn of the century is the basis of a very successful maize hybrid industry and has raised maize yields by 500% since 1928 (Shull 1952; Tomes 1998). A high level of genetic differentiation, as revealed by molecular markers, was later found between these populations (Melchinger et al. 1990). Ongoing activities for cassava include SSR analysis of a larger sample of landraces and diallel crosses of representative landraces from Nigeria and Guatemala.

Other principal findings of the study are the separation between Neotropical and African landraces and the more pronounced substructure in the African accessions. These results agree with a previous AFLP marker study of 29 African and 11 Neotropical landraces that placed African and Neotropical landraces in two distinct clusters with a substructure for the African accessions (Fregene et al. 2000). The forces that have shaped the genetic structure of landraces in the primary and African gene pools obviously include sexual recombination and selection for adaptation to biotic and abiotic stresses, and for end uses.

The most important constraint of cassava production in Africa, the cassava mosaic disease (CMD) does not occur in the Neotropics, and has undoubtedly served as selection pressure towards resistant genotypes and therefore new variability. An older form of resistance to CMD was introgressed from wild *Manihot glaziovii*, and a more recent source of resistance that confers high levels of resistance on cassava was found in Nigerian land races that were obviously mutants (Jennings 1976; Akano et al. 2002). These land races and their progeny would quickly become endemic in the collection of any region that has suffered a CMD epidemic.

A unique allele of marker SSRY183 was found in 50% of accessions from Nigeria and, coincidentally, these accessions have extreme levels of resistance to CMD. This association is consistent with a bottleneck imposed by adaptation to the disease. On the other hand, reduction in genetic diversity that should have resulted from disease epidemics in Africa, such as CMD, is not evident in this study. This suggests a mechanism that rapidly compensates for the loss of diversity over a short period of time. More insight to this is expected from a study recently initiated to assess diversity of cassava in Uganda after the CMD epidemic that swept through the country in the late eighties and early nineties (Gullberg 2003, personal communication). That study seeks to assess the immediate and long-term effect of such epidemics to genetic diversity.

Two findings of this study are of direct application to cassava breeding. First, the unique and broad diversity of cassava land races found in southeastern Tanzania reveals an invaluable germplasm resource for cassava improvement targeted to the region. The unique diversity suggests that the germplasm might have genes, in high frequencies, for adaptation to the area, while the high genetic diversity implies a high amount of additive genetic variance, upon which progress in plant breeding depends, is present. Secondly, the high level of differentiation between land races from Guatemala and Nigeria may represent a heterotic pool and provide an opportunity for the systematic exploitation of hybrid vigor in cassava.

Acknowledgements We thank Eric Minch, and Xavier Vekemans (Université Libre de Bruxelles, Brussels, Belgium) for their assistance with the computer packages "microsat" and GENSURVEY, and for their helpful suggestions. We acknowledge the useful comments from Joe Tohme and Paul Chavariagga-Aguirre (CIAT), Luigi Guarino (IPGRI), Ken Olsen (North Carolina State University), Ryohei Terauchi (Iwate Biotechnology Research Center, Kitakami, Japan) and two anonymous reviewers. We also thank Matthew Raya of the Agricultural Research Institute at Naliendele, Tanzania, for his assistance with field collection. This work was supported in part by the International Program for the

Chemical Sciences, Uppsala University, Sweden, and we are grateful to its director Malin Akerbolm for her abiding faith in our application of chemistry to unravel genetic diversity issues.

References

Akano A, Barrera E, Mba C, Dixon AGO, Fregene MA (2002) Genetic mapping of a dominant gene conferring resistance to the cassava mosaic disease (CMD). Theor Appl Genet 105:521–525

Baverstock PR, Moritz C (1996) Project design. In: Hillis DM, Moritz C, Mable BK (eds) Molecular systematics. Sinauer Associates, Sunderland, Massachusetts, pp 17–27

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

Bertin P, Gregoire D, Massart S, de Froidmont D (2001) Genetic diversity among European cultivated spelt revealed by microsatellites. Theor Appl Genet 102:148–156

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Mich E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457

Chavarriaga-Aguirre P, Maya MM, Bonierbale MW, Kresovich S, Fregene MA, Tohme J, Kochert G (1998) Microsatellites in cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. Theor Appl Genet 97:493–501

Chavarriaga-Aguirre P, Maya MM, Tohme J, Duque MC, Iglesias C, Bonierbale MW, Kresovich S, Kochert G (1999) Using microsatellites, isozymes and AFLPs to evaluate genetic diversity and redundancy in the cassava core collection and to assess the usefulness of DNA-based markers to maintain germplasm collections. Mol Breed 5:263–273

Cock JH (1985) Cassava: new potential for a neglected crop. Westview, London

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

Donini P, Stephenson P, Bryan GJ, Koebner RMD (1998) The potential of microsatellite markers for high throughput genetic diversity assessment in wheat and barley. Gene Resour Crop Evol 45:415–421

Doyle M, Pujol B, Elias M (2001) Ecology and genetics of populations of cassava managed by Amazonian Amerindians: or, how to maximize the benefits and minimize the costs of outcrossing in a vegetatively propagated crop. In: Fauquet C, Taylor NN (eds) Proc 5th Int Scientific Meeting of the Cassava Biotechnology Network, Danforth Plant Science, St. Louis, Missouri (in press)

Efron B (1982) The jackknife, the bootstrap and other resampling plans. NSF-CBMS Regional Conference Series in Applied Mathematics, Monograph 38. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pennsylvania

Elias M, Panaud O, McKey DB, Robert T (2000) Traditional cultivation of cassava among Amerindians: consequences on genetic diversity assessed with AFLP markers. In: Carvalho LJCB, Thro AM, Vilarinhos A (eds) *Cassava* biotechnology; IV Int Sci Meeting of the Cassava Biotechnology Network. Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), Brasília, DF, pp 111–117

Eyre-Walker A, Gaut RL, Hilton H, Feldman D, Gaut B (1998) Investigation of the bottlenecks leading to the domestication of maize. Proc Natl Acad Sci USA 95:4441–4446

FAO (Food and Agriculture Organization of the United Nations) (2000) Food outlook, Rome

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

Fregene MA, Bernal A, Duque MC, Dixon AGO, Tohme J (2000) AFLP analysis of African cassava (*Manihot esculenta* Crantz) germplasm resistant to the cassava mosaic disease (CMD). Theor Appl Genet 100:678–685

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate *F*-statistics. J Hered 86:485–486

Hershey C, Iglesias C, Iwanaga M, Tohme J (1992) Definition of a core collection for cassava. In: Report of the First Meeting of the International Network for Cassava Genetic Resources, CIAT, Cali, Colombia, 18–23 August 1992. International Crop Network Series No. 10. International Plant Genetic Resources Institute (IPGRI), Rome, pp 145–156

Jennings DL (1976) Breeding for resistance to African cassava mosaic. In: African cassava mosaic, report of an interdisciplinary workshop held at Muguga, Kenya. International Development Research Centre (IDRC), Ottawa, Canada, pp 39–44

Jones WO (1969) Manioc in Africa. Stanford University Press, Stanford, California

Keeratinijakal V, Lamkey KR (1993) Responses to a reciprocal recurrent selection in BSSS and BSCBI maize populations. Crop Sci 33:73–77

Kimura M, Crow JT (1964) The number of alleles that can be maintained in a finite population. Genetics 49:725–738

Leroy TC, Momtagnon C, Charrier A, Eskes AB (1993) Reciprocal recurrent selection applied to *Coffea canephora* Pierre. I. Characterization and evaluation of breeding populations and value of intergroup hybrids. Euphytica 67:113–125

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639–655

Mba REC, Stephenson P, Edwards K, Melzer S, Mkumbira J, Gullberg U, Apel K, Gale M, Tohme J, Fregene MA (2001) Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. Theor Appl Genet 101:21–31

Meagher TR (1986) Analysis of paternity within a natural population of *Chamaelirium luteum*. I. Identification of most-likely male parents. Am Natl 128:199–215

Melchinger AE, Lee M, Lamkey KR, Woodman WL (1990) Genetic diversity for restriction fragment length polymorphisms: relation to estimated genetic effects in maize inbreds. Crop Sci 30:1033–1040

Nei M (1978) Estimation of average heterozygosity and genetic distances from a small number of individuals. Genetics 89:583–590

Olsen K, Schaal B (1999) Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. Proc Natl Acad Sci USA 96:5586–5591

Olsen K, Schaal B (2001) Microsatellite variation in cassava *Manihot esculenta*, Euphorbiaceae and its wild relatives: evidence for a southern Amazonian origin of domestication. Am J Bot 88:131–142

Pearsall DM (1992) Agriculture in the low land tropics of the Americas. In: Cowan CW, Watson PJ, Benco N (eds) The origins of agriculture: an international perspective. Smithsonian Institute, Washington, DC, pp 173–205

Quenoille M (1956) Notes on bias in estimation. Biometrika 43:253–260

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

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

Rogers DJ (1965) Some botanical and ethnological considerations of *Manihot esculenta*. Econ Bot 194:769–777

Rohlf FJ (1993) NTSYS-PC numerical taxonomy and multivariate analysis system. Version 1.8 Exeter Publ., Setauket, New York

SAS Institute, Inc (1995) JMP (version 3.1). Cary, North Carolina

Sauer CO (1952) Agricultural origins and dispersals. American Geographical Society, New York, USA

Sauer JD (1993) Historical geography of crop plants: a select roster. (CRC Press, Boca Raton, Florida)

Schaal BA, Olson PD, Prinzie TP, Carvalho LCB, Tonukari J, Hayworth DA (1994) Phylogenetic analysis of the genus *Manihot* based on molecular markers. In: Roca WM, Thro AM (eds) Proc 1st Scientific Meeting of the Cassava Biotechnology Network, Cartagena, Colombia, 25–28 August 1992 Working Document No. 123. Centro Internacional de Agricultura Tropical (CIAT), Cali, Columbia, pp 106–122

Shull GF (1952) Beginnings of the heterosis concept. In: Gowen JW Heterosis. Iowa Sate College Press, Ames, Iowa, pp 14–48

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics 139:457–462

Thompson EA (1975) The estimation of pairwise relationships. Ann Am Naturalist 128:199–215

Tomes D (1998) Heterosis: performance stability, adaptability to changing technology and the foundation of agriculture as a business. In: Lamkey K, Staub JE (eds) Concepts and breeding of heterosis in crop plants. CSSA Special Publication Number 25. Crop Science Society of America, Madison, Wisconsin

Vekemans X, Lefebvre C (1997) On the evolution of heavy metal tolerant populations in *Armeria maritima*: evidence from allozyme variation and reproductive barriers. J Evol Biol 10:175–191

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370

Wright S (1951) The genetic structure of populations. Ann Eugen 15:323–354