VARIATION IN CASSAVA (*MANIHOT ESCULENTA* CRANTZ) BASED ON SINGLE NUCLEOTIDE POLYMORPHISMS, SIMPLE SEQUENCE REPEATS AND PHENOTYPIC TRAITS

by

ROBERT SEZI KAWUKI

Thesis submitted in fulfilment of the requirements for the degree Philosophiae Doctor in the Department of Plant Sciences: Plant Breeding, in the Faculty of Natural and Agricultural Sciences at the University of the Free State

UNIVERSITY OF THE FREE STATE
BLOEMFONTEIN
SOUTH AFRICA

Promoter: Prof. Liezel Herselman
Co-promoters: Prof. Maryke T. Labuschagne
Dr. Morag Ferguson

November 2009

DECLARATION

"I declare that the thesis hereby submitted by me for the degree Philosophiae Doctor in Agriculture at the University of the Free State is my own independent work and has not previously been submitted by me at another University/faculty

I furthermore cede copyright of the thesis in favour of the University of the Free State"

…… ……..

Robert Sezi Kawuki Date

DEDICATION

The work is dedicated to my father Theophilus Mukasa Kaddu; my mother Lydia Naggalabuzi Kaddu; sisters Sylivia, Judith, Joan, Justine, Flavia, Hellen and Stella and brothers John and Ronald. They have provided me with immense support throughout my education.

ACKNOWLEDGEMENTS

The work presented in this thesis involved the interaction and contribution from several individuals and institutions. The ones listed herein are just a few of the many contributors.

Foremost, I thank the Almighty God for guiding and keeping me going throughout the execution of this research and more importantly during times when I was near to the ground.

The Biosciences of Eastern and Southern Africa (BecA) under the stewardship of Prof. Bruno Kubata provided me with the PhD fellowship. To the entire BecA secretariat based at Nairobi, I sincerely thank you for the support rendered.

I thank the Director General of the National Agricultural Research Organisation (NARO) Uganda, Dr. Dennis Kyetere; the former and current Director of Research, National Crops Resources Research Institute (NaCRRI), Drs. Fina Opio and James Thomas Ogwang and the Head of Cassava Programme, Dr. Anton Bua, all for fully granting me the study leave and support. I also thank the scientists of the Cassava Programme notably Dr. Yona Baguma, Dr. Chris Omongo, Dr. Titus Alicai, Mr. Stephen Kashub, Mr. Anthony Pariyo and Ms. Teddy Amuge who supported me throughout the PhD study.

Special appreciation goes to my promoters: Dr. Morag Ferguson, Prof. Maryke Labuschagne, Prof. Liezel Herselman and Dr. Dong-Jin Kim. This team guided me with intelligence, sensitivity and passion for the subject. Morag conceived the idea and sourced funds to support most of the research activities. Maryke provided funds for the starch analysis. To all, I owe them a debt I cannot hope to pay.

Colleagues and friends of Laboratory 4 at the International Livestock Research Institute (ILRI), Nairobi, Kenya: Evans Mutegi, Heneriko Kulembeka, Eric Magembe, Inosters Nzuki, Rosemary Mutegi, Titus Kathurima, Martina Kyalo, Elizabeth Njuguna, Veronica Ogugo, Maggie Mwathi, Mercy Kitavi, Mercy Mbogori and Michael Kimani, are all appreciated for the assistance and patience during my early and shaky days in the lab! I enjoyed interacting and working with you all. Dr. Henry Ojulong, I thank you for reading through this work and the wonderful suggestions that you proposed. I sincerely thank Alice Muchiri, the administrator at IITA-Nairobi. She was an expert hand-holder; she assisted with all laboratory, equipment and travel logistics.

At the University of the Free State (UFS), Mrs. Sadie Geldenhuys was another expert hand-holder; she swiftly helped with all the university logistics and accommodation arrangements during my stay in "The City of Roses", Bloemfontein (Mangaung). I also extend a special word of thanks to fellow students at UFS with whom I interacted: Davies Mweta, Oscar Olago, Godwin Amenorpe, Elizabeth Parks, Scott Sydenham, Onoufrios Philippou and Rouxléne van der Merwe. I also thank Sarah Chalo for the assistance in the Biochemistry Laboratory.

Colleagues from the NARS of Kenya (Mr. Bramwel Wanja, Mr. Hannington Obiero and Dr. James Gethi), Tanzania (Mr. Simon Jeremiah and Dr. Geoffrey Nkamilo), Rwanda (Mrs. Claire Kanyange and Mr. Gervais Gashaka), DRC (Mr. Mpansu Bidiaka and Mr. Singi Lukombo), Mozambique (Dr. Anabela Zacarius and Mr. Frederico Madabula), Madagascar (Ms. Isabelle Ralimanana) and Uganda (Mr. Orone Joseph, Mr. Charles Majara, Mr. Robert Oba, Mr. Francis Osiganda, Mr. Jimmy Akano and Mr. Jacob Omara). I thank you all together with your respective teams for willing to work with me and the passion you displayed while conducting the research. Keep up the spirit please!

Lastly, but not least, I thank members of my family who supported me throughout my studies. To all, may the Almighty continue to bless you abundantly.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	Adenine
AFLP	Amplified fragment length polymorphism
AKR	*Aldo/keto reductase*
AL	Colour of apical leaf
AMOVA	Analysis of molecular variation
AmyA	*Alpha-amylase gene*
ANOVA	Analysis of variance
APX3	*Ascorbate peroxidise*
BAC	Bacterial artificial chromosome
BC	Before Christ
BecA	Biosciences Eastern and Central Africa
BC_2	Second backcross
Bgla	Beta glucosidase
BH	Branching habit
Bp	Base pair
C	Cytosine
CAT2	*Catalase*
CBSD	Cassava brown streak disease
CEB	Colour of end branches
CI	Allergenic-related protein Pt2L4
CIAT	International Centre for Tropical Agriculture
CLS	Shape of central leaf
CLV	Colour of leaf vein
CMD	Cassava mosaic disease
COEEP	Chloroplast oxygen-evolving enhancer protein
CPI	*Cysteine protease inhibitor*
CRC	Colour of root cortex
CRP	Colour of root pulp
CS	Colour of stem exterior

CSC	Colour of stem cortex
CSE	Colour of stem epidermis
CV	Coefficient of variation
CYP79D2	*N-hydroxylating cytochrome P450 gene*
DArT	Diversity array technology
DGGE	Denaturing gradient gel electrophoresis
D_E	Euclidean distance
DMC	Dry matter content
DNA	Deoxyribose nucleic acid
DNASP	DNA sequence polymorphism
dNTPs	Deoxynucleotide triphosphates
d_R	Rogers's genetic distance
DRC	Democratic Republic of Congo
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
EP	Cortex ease of peeling
ERC	External colour of root
ERF	*Ethylene response factor*
eSNP	Electronic single nucleotide polymorphism
EST	Expressed sequence tags
f	Coefficient of co-ancestry
F_1	First filial generation
FAO	Food and Agriculture Organisation
F3H	Flavanone-3 hydroxylase
F_{IS}	Correlation of genes within individuals within populations
F_{IT}	Correlation of genes within individuals over all populations
FS	Prominence of foliar scars
F_{ST}	Correlation of genes of different individuals in the same population
G	Guanine
G3pdh	*Glyceraldehyde 3-phosphate dehydrogenase*
GBSSII	*Granule bound starch synthase II precursor*
GCA	General combining ability

HA	Heteroduplex analysis
h^2	Heritability
HCN	Hydrogen cyanide content
HCN	Hydrogen cyanide levels
H_d	Haplotype based gene diversity
H_e	Gene diversity or Expected heterozygosity
HI	Harvest index
Hnl	ά-Hydroxynitrile
H_o	Observed heterozygosity
HPSEC	High performance size exclusion chromotography
IAC	Instituto Agronômico de Campinas
HRGP	Hydroxyproline-rich glycoprotein
IAM	Infinite allele model
ID	Inbreeding depression
IFPRI	International Food Policy Research Institute
IITA	International Institute of Tropical Agriculture
Indels	Insertion and deletions
IPGRI	International Plant Genetic Resources Institute
K	Number of putative populations
Kb	kilobase
LB	Levels of branching
LC	Leaf colour
LD	Linkage disequilibrium
LL	Number of leaf lobes
LM	Lobe margins
LR	Leaf retention
LS	Length of stipule
LSD	Least significant difference
MAP	Months after planting
MAS	Marker-assisted selection
MCMC	Markov Chain Monte Carlo

MDS	Multidimensional scaling
MeEF1	*Manihot esculenta elongation factor 1-alpha gene*
$MgCl_2$	Magnesium Chloride
MOLCAS	Cassava Molecular Diversity Network
NaCRRI	National Crops Resources Research Institute
NARS	National Agricultural Research Systems
NCBI	National Centre for Biotechnology Information
NCSS	Number cruncher statistical system
Nd	Nucleotide diversity
Nm	Number of migrants coming into a population per generation
OP	Orientation of petiole
P	Pubescence on apical leaf
PAL2	*Phenylalanine ammonia-lyase 2 gene*
PC	Principal component
PC	Petiole colour
PCA	Principal component analysis
PCR	Polymerase chain reaction
PIC	Polymorphic information content
pLIN	*Linamarase gene*
PPD	Post-harvest physiological detoriationdeterioration
PS	Shape of plant
QTL	Quantitative trait loci
RAPD	Random amplified polymorphic DNA
RC	Root constrictions
REML	Restricted maximum likelihood
RFLP	Restriction fragment length polymorphisms
RP	Extent of root peduncle
RS	Root shape
RUBISCO	Ribulose 1,5-bisphosphate carboxylase small chain precursor
S	Segregating sites
S_0	Non-inbred progenitor

S_1	Selfing of first generation
SG	Growth habit of stem
Sig_a	Among sample variance component
Sig_b	Between individual within sample variance component
Sig_w	Within individual variance component
SM	Stipule margin
SMM	Stepwise mutation model
SNP	Single nucleotide polymorphism
SOD	*Copper/zinc superoxide dismutase*
SSA	Sub-Saharan Africa
SSCP	Single strand conformation polymorphism
Ssp	Sub species
SSR	Simple sequence repeat
T	Thymine
Taq	*Thermus aquaticus*
TK	*Pto-like serine/threonine kinase*
TRE	Texture of root epidermis
Wa	Weight in air
Ww	Weight in water
Wx	Waxy gene

SI UNITS

µl	Microlitre
^{o}C	Degree centigrade
cM	CentiMorgan
Cm	Centimetre
G	Gram
H	Hour
Ha	Hectare
Kg	Kilogram
L	Litre
M	Metre
M	Molar
Mg	Milligram
Min	Minute
Ml	Millilitre
Mm	Millimetre
mM	Millimolar
ng	Nanogram
nm	nanometre
pmole	Picomole
Sec	Second
U	Unit

CHAPTER 1

GENERAL INTRODUCTION

As part of national, regional and international efforts to reduce poverty and food security problems prevalent in sub-Saharan Africa (SSA), increased crop productivity is being advocated. A strong justification for this approach stems from the benefits of the "Green Revolution" that were driven by science-led initiatives, which resulted in a dramatic increase in crop yields in the Asian and meso-American countries during the 1960s and 1970s. These interventions consequently resulted in substantial reduction in famine, increased calorie intake and stimulation of rural economic growth. Wheat and rice were specifically targeted for improvement, as they were commonly grown in the developing world (IFPRI, 2002). In SSA, cassava (*Manihot esculenta* Crantz) is a popular and widely grown crop (Nweke et al., 2002), suggesting that interventions to increase its productivity and/or utilisation will make a significant contribution towards the improvement of the quality of life of rural communities that primarily depend on it.

Studies conducted in cereals between 1960 and 1990 indicated that over 50% of the increase in crop production resulted from improved crop cultivars and that technologies generated through plant breeding played an important role in providing food, feed and fibre to the ever increasing human population (Frey, 1992). Compared to this, interventions to increase cassava productivity in southern, eastern and central Africa in order to mitigate hunger and poverty have not been very successful. Key limitations towards attainment of this goal are the array of constraints presented by several biotic (insect pests and diseases) and abiotic (drought, post-harvest deterioration and hydrogen cyanide) stresses. These stresses cut across regions and pose a huge challenge towards increased cassava productivity and utilisation (Hahn et al., 1979; IITA, 1990; Nweke et al., 2002). National policies and funding to support cassava research have also been limited.

Clearly, to achieve the goal of increased income and sustained food security in the SSA region, efforts must be made to address key production constraints through modern crop improvement programmes, especially via the adoption of new breeding methods and techniques (Ceballos et al., 2004; Ojulong et al., 2008).

In the last decades, considerable research has been devoted towards addressing key cassava production constraints in Africa, but with varying levels of success (Hahn et al., 1980; IITA, 1990; Mahungu et al., 1994; Alicai et al., 2007; Hershey, 2010). A case in point is the devastating cassava mosaic disease (CMD) that is endemic in Africa. Although considerable progress was made to deploy CMD resistance genes from both wild relatives (Jennings, 1957; 1994) and domesticated varieties (Akano et al., 2002; Hershey, 2010), CMD incidences are still high in most cassava growing communities in eastern and southern Africa, primarily due to cultivation of CMD susceptible varieties (Hershey, 2010).

Indeed, both formal and informal interactions with farmers indicate that farmers are reverting back to CMD susceptible local varieties chiefly because most of the deployed CMD resistant varieties have inferior culinary root qualities. Chang et al. (1979) observed that promising sources of genes will continue to be sourced from local varieties, as they have adaptive gene complexes that they have accumulated through long-term selection. Local varieties have, over generations, been selected for farmer-preferred characteristics and if new varieties are to be adopted they must address farmer preferences.

The systematic characterisation of farmer varieties facilitates their enhanced utilisation through: 1) knowledge of the range of existing variation, 2) knowledge of geographical distribution of adaptive gene complexes and 3) identification of individuals with preferred characteristics. Some breeding programmes are incorporating local varieties into their hybridisation programmes to identify, among the resulting progeny, improved clones with agronomic traits that match those of the local varieties (Hershey, 2010).

An equally challenging objective in cassava breeding is the improvement of starch quality, as this will spur the industrial utilisation of cassava. Efforts to change starch characteristics have been limited. The enthusiasm that was placed on the genetic improvement of other key agronomic cassava traits (Kawano, 2003) can also be reciprocated in starch quality traits. Work on starch quality has began (Raemakers et al., 2005; Ceballos et al., 2007) and needs to be further accelerated to increase the competitiveness of cassava in the booming starch industry. Ceballos et al. (2004) noted that development of cassava varieties with high quality starch will provide an alternative to the expensive carbohydrate sources from the temperate regions.

Genetic progress requires that: 1) genetic variability for the trait of interest exists, 2) the trait is heritable and 3) the trait can be accurately selected for using phenotypic data, and/or molecular markers, where possible (Hershey, 2010). Evidently, the ability to develop cassava varieties that meet specifications of growers, processors and consumers will largely depend on the availability of sufficient and useful genetic diversity. However, to be of benefit, this genetic variation should be useful in attaining higher yields under favourable productions systems and/or extending cultivation to unfavourable regions characterised by low input technologies. Maunder (1992) observed that the maximum potential to be attained via breeding rests on choice of germplasm and that the actual breeding approach will determine how much of the potential can be realised. Selection and optimal utilisation of germplasm will require that it is well characterised and/or evaluated with efficient techniques. If this is not done, the crop's potential will never be realised.

Smale (1998) documented that in traditional, resource poor farming communities located in marginal, variable environments, the crop populations that endure are those that meet production and consumption standards. These populations possess the genetic variability to respond to continual changes in farmers' needs and growing environments. Cassava farmers have for several decades cultivated the crop and have accumulated knowledge that needs to be incorporated in genetic diversity studies.

Classic examples of the advantages of evaluating and/or characterising cassava germplasm are illustrated by the identification of 1) CMD resistance genes in local cassava varieties (Akano et al., 2002), 2) waxy cassava through expression of a naturally occurring mutation through inbreeding (Ceballos et al., 2007) and 3) sugary cassava (Carvalho et al., 2004). Evaluation of large germplasm sets has furthermore provided valuable information as demonstrated by recently conducted cassava studies (Chávez et al., 2005; Sánchez et al., 2009).

Genetic diversity in most domesticated crops has been shaped by pre-domestication evolutionary forces involving wild species progenitors and/or by post-domestication influences, particularly by human and natural selection (Vavilov, 1951). Since the introduction of cassava to the east African coast in the eighteenth century (Jones, 1959), cassava varieties have evolved and adapted in response to selection for human preferences and adaptation to local environments. These evolutionary forces can result in changes in one of a few genes or whole genomes (Buckler and Thornsberry, 2002). Whether there are individual gene or whole genome changes, the resulting genetic variation presents itself in various forms (ecological adaptation traits, agronomic and consumer related traits, morphological traits, chromosomal morphology and behaviour and biosynthetic pathways), which can be exploited for cassava improvement (Hershey, 2010).

Tools available for dissecting diversity include both molecular and morphological markers. The earliest markers for cassava were morphological and were identified on leaves, stems and roots. These markers are generally under genetic control with little or no environmental influence (Hershey and Ocampo, 1989). Stable morphological traits are recommended for studying genetic diversity (Gulick et al., 1983). However, for genetic gains in plant breeding, it is important to understand variation in key quantitative traits (those of moderate to high heritability) that are relevant to cassava improvement and/or commercialisation. Examples of such phenotypic quantitative traits include root dry matter content, harvest index, leaf retention and root cortex thickness.

Simple sequence repeats (SSRs), which are densely interspersed in eukaryotic genomes (Tautz and Renz, 1984) are a classic example of molecular markers that have proved to be effective in studying genetic variation both within and between populations (Matsuoka et al., 2002; Maccaferri et al., 2003; Toro et al., 2009). Specifically, SSRs can be used to infer allele frequencies distribution, allelic richness and population structure in cassava germplasm available within the different national cassava breeding programmes in the region. Another class of molecular markers, single nucleotide polymorphism (SNPs), are one of the new generation markers reported to be the most frequent form of naturally occurring genetic variation in populations (Kruglyak, 1997). SNPs may hence provide sufficient variation to discriminate between closely related individuals.

Cassava is widely grown in SSA by the rural poor communities. Thus, efforts aimed at increasing cassava productivity will positively impact on these communities that primarily depend on cassava. Because of the benefits associated with cassava inbreeding, particularly the identification of novel starches as witnessed by the waxy S_1 cassava clone (Ceballos et al., 2007), it is important that these new breeding approaches (like cassava inbreeding) be initiated by national cassava breeding programmes whose mandate is to serve cassava communities. Further, it is important to examine the diversity of cassava germplasm (local and improved genotypes) available within the national cassava breeding programmes. Evaluation and/or characterisation of this germplasm should give consideration for some farmer preferred traits.

Taken together, all this information helps elucidate whether or not the genetic variation has a hierarchical organisation. For instance, is the cassava germplasm differentiated among the national breeding programmes? Are local varieties genetically differentiated from improved cassava varieties? This information could be important in designing germplasm conservation schemes and/or in defining breeding objectives.

Therefore this thesis presents results from four research objectives. The first objective presents analysis of cassava germplasm from six countries (Uganda, Kenya, Tanzania, Rwanda, Democratic Republic of Congo (DRC) and Madagascar) using 29 qualitative traits. Germplasm from these countries was further quantified for root dry matter content, harvest index, leaf retention and root cortex thickness. The second objective presents analysis of allele frequency distribution in cassava germplasm from Uganda, Kenya, Tanzania, Rwanda, Democratic Republic of Congo, Mozambique and Madagascar using 26 highly polymorphic SSR markers. The third objective characterises and examines the utility of SNPs in cassava and compared them to commonly used SSRs. And finally, the fourth objective quantifies variation in root dry matter content, harvest index and amylose content in S_1 cassava inbreds.

References

Akano, A.O., A.G.O. Dixon, C. Mba, E. Barrera, and M. Fregene, 2002. Genetic mapping of a dominant gene conferring resistance to the cassava mosaic disease (CMD). Theoretical and Applied Genetics 105:521-525.

Alicai, T., C.A. Omongo, M.N. Maruthi, R.J. Hillocks, Y. Baguma, R. Kawuki, A. Bua, G.W. Otim-Nape, and J. Colvin, 2007. Re-emergence of cassava brown streak disease in Uganda. Plant Disease 91:24-29.

Buckler, E.S., and J.M. Thornsberry, 2002. Plant molecular diversity and applications to genomics. Current Opinion in Plant Biology 5:107-111.

Carvalho, L.J.C.B., C.R.B. de Souza, J.C.M. Cascardo, C.B. Junior, and L. Campos, 2004. Identification and characterisation of a novel cassava (*Manihot esculenta* Crantz) clone with high free sugar content and novel starch. Plant Molecular Biology 56:643-659.

Ceballos, H., C.C. Iglesias, J.C. Pérez, and A.G.O. Dixon, 2004. Cassava breeding: opportunities and challenges. Plant Molecular Biology 56:503-515.

Ceballos, H., T. Sánchez, N. Morante, M. Fregene, D. Dufour, A.M. Smith, K. Denyer, J.C. Perez, F. Calle and C. Mestres, 2007. Discovery of an amylose-free starch mutant in cassava (*Manihot esculenta* Crantz). Journal of Agricultural and Food Chemistry 55: 7469-7476.

Chang, T.T., W.L. Brown, J.G. Boonman, J. Sneep, and H. Lamberts, 1979. Crop genetics resources. In: Sneep, J., A.J.T. Hendriksen, and O. Holbek (Eds.), Plant breeding perspective: Centennial publication of Koninklijk Kweekbedrijf en Zaadhandel D.J. van der Have 1879-1979, pp. 83-103. Centre for Agricultural publishing and documentation, Wageningen, Netherlands.

Chávez, A.L., T. Sánchez, G. Jaramillo, J.M., Bedoya, J. Echeverri, E.A. Bolaños, H. Ceballos, and C.A. Iglesias, 2005. Variation of quality traits in cassava roots evaluated in landraces and improved clones. Euphytica 143:125-133.

Frey, K.J., 1992. Plant breeding perspectives for the 1990s. In: Stalker, H.T., and J.P. Murphy (Eds.), Plant breeding in the 1990s, Proceedings of the symposium on plant breeding in the 1990s held at North Carolina State University Raleigh, NC, pp. 1-13. CAB International Wallingford Oxon OX10 8DE, UK.

Gulick, P., C. Hershey, and J. Esquinas-Alcazar, 1983. Genetic resources of cassava and wild relatives. Rome, International Board for Plant Genetic Resources Rome: IBPGR, pp. 56.

Hahn, S.K., E.R. Terry, and K. Leuschner, 1980. Breeding cassava for resistance to cassava mosaic disease. Euphytica 29:677-683.

Hahn, S.K., E.R. Terry, K. Leuschner, I.O. Akobunda, C. Okali, and R. Lal, 1979. Cassava improvement in Africa. Field Crops Research 2:193-226.

Hershey, C., 2010. Cassava breeding: theory and practice. A publication by Food and Agriculture Organisation (FAO), in press.

Hershey, C., and C. Ocampo, 1989. New marker genes found in cassava. Cassava Newsletter 13:1-5.

IFPRI, 2002. Green Revolution: curse or blessing? International Food Policy Research Institute (IFPRI), pp. 4. http://www.ifpri.org/pubs/ib/ib11.pdf. Accessed April 3, 2008.

IITA, 1990. Cassava in tropical Africa. A Reference Manual. International Institute of Tropical Agriculture, Ibadan, Nigeria, pp. 176.

Jennings, D.L., 1957. Further studies in breeding cassava for virus resistance. East African Agricultural Journal 22:213-219.

Jennings, D.L., 1994. Breeding for resistance to African cassava mosaic geminiviruses in East Africa. Tropical Science 34:110-122.

Jones, W.O., 1959. *Manioc* in Africa. Standard University Press, Stanford, California, pp. 315.

Kawano, K., 2003. Thirty years of cassava breeding for productivity-biological and social factors for success. Crop Science 43:1325-1335.

Kruglyak, L., 1997. The use of a genetic map of biallelic markers in linkage studies. Natural Genetics 17:21-24.

Maccaferri, M., M.C. Sanguineti, P. Donini, and R. Tuberosa, 2003. Microsatellite analysis reveals a progressive widening of the genetic basis in the elite durum wheat germplasm. Theoretical and Applied Genetics 107:783-797.

Mahungu, N.M., A.G.O. Dixon, and J. Mkumbira, 1994. Breeding cassava for multiple pest resistance in Africa. African Crop Science Journal 2:539-552.

Matsuoka, Y., S.E. Mitchell, S. Kresovich, M. Goodman, and J. Doebley, 2002. Microstaellites in Zea-variability, patterns of mutation, and use for evolutionary studies. Theoretical and Applied Genetics 104:436-450.

Maunder, A.B., 1992. Identification of useful germplasm for practical plant breeding programmes. In: Stalker, H.T., and J.P. Murphy (Eds.), Plant Breeding in the 1990s. Proceedings of the symposium on plant breeding in the 1990s held at North Carolina State University Raleigh, NC, pp. 147-169. CABI International Wallington Oxon OX108DE, UK.

Nweke, F., D. Spencer, and J. Lynam, 2002. The cassava transformation: Africa's best kept secret. Michigan State University Press, East Lansing, USA, pp. 273.

Ojulong, H., M.T. Labuschagne, M. Fregene, and L. Herselman, 2008. A cassava clonal evaluation trial based on a new cassava breeding scheme. Euphytica 160:119-129.

Raemakers, K., M. Schreuder, L. Suurs, H. Furrer-Verhorst, J.P. Vincken, N. de Vetten, E. Jacobsen, and R.G.F. Visser, 2005. Improved cassava starch by antisense inhibition of granule-bound starch synthase I. Molecular Breeding 16:163-172.

Sánchez, T., E. Salcedo, H. Ceballos, D. Dufour, G. Mafla, N. Morante, F. Calle, J.C. Pèrez, D. Debouck, G. Jaramillo, and I.X. Moreno, 2009. Screening of starch quality traits in cassava (*Manihot esculenta* Crantz). Starch 61:12-19.

Smale, M., 1998. Varietal diversity in bread wheat. In : Evenson, R.E. (Ed.), Agricultural values of plant genetic resources. Food and Agriculture Organization of the United Nations (FAO), Center for International Studies on Economic Growth - Tor Vergata University, and CAB International, Wallingford, UK.

Tautz, D., and M. Renz, 1984. Simple sequences are ubiquitous components of the eukaryotic genomes. Nucleic Acids Research 12:4127-4138.

Toro, A.A., J. Fernàndez, and A. Caballero, 2009. Molecular characterization of breeds and its use in conservation. Livestock Science 120:174-195.

Vavilov, N.I., 1951. Phytogeographic basis of plant breeding. The origin, variation, immunity and breeding of cultivated plants. Chronicle of Botany 13:1-366.

CHAPTER 2

LITERATURE REVIEW

2.1 The cassava plant: its domestication and genetic uniqueness

Cassava, also known in different parts of the world as yuca, manioc and mandioca, is a widely distributed crop in the tropics. Cassava belongs to the family Euphorbiaceae and is the agriculturally most productive plant among the 98 species of the genus *Manihot* (Rogers and Appan, 1973). After several years of debate on its origin (Allem, 1994), molecular markers have provided strong evidence that cassava was likely domesticated from a single wild *Manihot* species (*M. esculenta* spp. *flabellifolia* Pohl) and that the crop originated from the southern Amazon basin (Olsen, 2004). That study involved analysis of 212 individuals collected from wild *Manihot* populations and 20 cassava varieties representative of the crop's diversity, using both SNPs and SSRs. Earlier studies had however pointed to cassava being a hybrid, with a wild species as a progenitor (Allem, 1999; Olsen and Schaal, 1999) and that cassava is a recently diverged crop (Olsen and Schaal, 2001).

Cassava cytogenetics has been another contradicatory aspect of the crop. Cytogenetic studies have provided conflicting results with some indicating that it is a segmental allotetraploid (Magoon et al., 1969), while others indicated that it is an allopolyploid (Umannah and Hartman, 1973). Other studies have however indicated that it is a diploid species (Chavarriaga-Aguirre et al., 1998) with 2n = 36 (Nasser, 2005). This contradiction in cassava cytogenetics needs to be resolved. It is likely that accurate answers to this discrepancy will be provided by further molecular studies.

It is estimated that cassava domestication began 5000-7000 years BC in the Amazon region (Gibbons, 1990). It was from this region that cassava was exported to other parts of the world. Cassava was one of the first crops exported to Africa from the east coast of Brazil. In Africa, the crop was first introduced to west Africa some time during the 1700s (Jones, 1959; Carter et al., 1992). Thereafter it was quickly adopted and rapidly spread within the west African region (Hershey, 2010). Subsequent introduction of cassava to Africa was via the east African coastline in the 1750s, when the French introduced cassava from Brazil to Mauritius (Jones, 1959). Thereafter the crop was introduced to Madagascar and then inland from where it spread to various countries in the eastern, central and southern African region (Jones, 1959; Langlands, 1966).

It is therefore rational to suggest that Africa received a portion of the genetic diversity present in the crop's centre of origin and that this was achieved through two major routes, namely west and east Africa. It should also be noted that during the early breeding activities in the east African region during the 1930s, extensive hybridisation between cassava and its wild relatives occurred. These initiatives too, generated new and broadened genetic variability in the region. Examples of these hybridisation schemes included: 1) utilisation of two species *M. melanobasis* Mueller von Argau. and *M. saxicola* Lang., for protein enhancement (Jennings, 1959; 1963) and 2) *M. dichotoma* Ule and *M. glaziovii* Mueller von Argau., for resistance gene sources. However, it was only *M. glaziovii* that contributed valuable genes (Storey and Nichols, 1938; Hahn et al., 1980). Today, cassava is a major crop in most west, central and east African countries. From Africa, cassava was introduced to Asia, where it is currently a major crop, particularly in Indonesia, Thailand, India, China, Philippines and Vietnam (Hershey, 2010).

For several thousand years, farmers have been altering the genetic makeup of crops they grow, a process which has considerably changed domesticated plants compared to their wild relatives. A case in point is the sharp contrast between maize and teosinte (Lauter and Doebley, 2002).

For cassava, studies have established that a significant portion of genetic variation of *M. esculenta* spp. *flabellifolia* exists within varieties (Olsen, 2004). This variation exists in the form of root, leaf and stem qualitative and quantitative traits on which selection has been done (Hershey, 2010). Cassava, also considered as a woody shrub, can grow erect up to 4 m in height. The plant is monoecious with both male and female flowers found on the same plant. However, female flowers open before male flowers (protogyny), a mechanism that enhances out-crossing (IITA, 1990). Because male and female flowers on different branches or on different plants of the same genotype can open simultaneously, self-pollination can occur (Kawano et al., 1978; Ceballos et al., 2004).

Flowering in cassava depends on the genotype, time of planting and environment (Jennings and Iglesias, 2002). Based on the flowering habit, varieties can be grouped into non-flowering, poor flowering, moderate flowering, profuse flowering with poor fruit setting or profuse flowering with high fruit set (Indira et al., 1977). On average, one to two seeds are obtained per pollination (Ceballos et al., 2004), which results in a low seed rate. Cassava seeds display dormancy periods of about two months and require temperatures in the range of 30^{o}-35^{o}C for germination (Ellis et al., 1982; IITA, 1990). Fruit maturation can take up to three months after pollination (IITA, 1990) and the female parent is more important in determining genetic progress (Jennings, 1963).

Cassava may be propagated either vegetatively (stem cuttings) or sexually (true seed). Vegetative propagation permanently conserves a genotype, whatever the level of heterozygosity. This has the obvious advantage that a superior plant, if identified at any evaluation stage, can be cloned indefinitely, maintaining genotypic integrity through successive generations. This fixation of the genotype forms the basis of all commercial plantings, while propagation by seed is largely limited to formal breeding purposes. However, some farmers are known to make selections from volunteer seedlings if their properties are deemed desirable (Balyejusa Kizito, 2006).

The out-crossing nature of the crop ensures that high levels of heterozygosity are maintained (Kawano et al., 1978; Meireles da Silva et al., 2003). The wide segregation observed in progeny from any cross combination (Kawano et al., 1978) is evidence for the highly heterozygous nature of cassava. These high levels of heterozygosity are however, a major hindrance to sustainable cassava improvement, for example in the implementation of a successful backcrossing scheme and identification of useful recessive traits. It is partly for these reasons that inbreeding and production of double haploids in cassava has been advocated (Ceballos et al., 2004).

2.2 Cassava utilisation and products

Cassava is characterised by high rates of starch accumulation and inherently high adaptability to drought prone and/or marginal tropical areas. Cassava owes its popularity in the tropics to the diversity of uses of its starchy roots (Hershey, 2010). As early as six weeks after planting, some of the fibrous roots begin to thicken rapidly, laying down large quantities of xylem parenchyma that are packed with starch granules and by 12 months after planting (MAP) most roots are saturated with starch (Howeler and Cadavid, 1983; Hershey, 2010).

Starch-based products fall into three main categories: 1) native or unmodified starch, 2) modified starch, which is manuplated by physical, chemical or biological means and 3) sweeteners, including high fructose syrup and glucose. Accordingly, cassava starch has diversified uses in the food, papermaking and textile industries and in the production of alcohol (Hershey, 2010). However, a major limitation of alcohol production from cassava is the limited energy balance as compared to crops like sugarcane because the sugars in cane stems can easily be converted into fuel, whereas cassava stems are needed for propagation of subsequent crops and the starch in cassava roots needs to be initially degraded (Ceballos et al., 2008).

Cassava utilisation patterns vary with the region of the world. In Africa, food consumption (fresh or processed) predominates (Nweke et al., 2002). Cassava utilisation as feed and/or as a raw material in the starch industry is far more progressed in some Asian and Latin American countries (Chang, 2000). The recent identification in Latin America of amylose-free cassava starch (Ceballos et al., 2007), high-amylose starch (Ceballos et al., 2008), sugary cassava (Carvalho et al., 2004) and yellow cassava roots with high levels of carotenes (Hershey, 2010), opens up new vistas for industrial and nutritional utilisation of cassava. In addition the potential for cassava to produce biomass for renewable energy has been recognised (Sinha and Swaminathan, 1984).

Besides the starchy roots, cassava leaves are an important vegetable in some communities (Lutaladio and Ezumah, 1981). Leaves are high in protein, vitamin C, iron and calcium and are used both as human food and in animal feeds. When used for human consumption, leaves are, however, cooked (Hershey, 2010). Traditional processing methods that include cooking, pounding, grating, drying and fermentation result in the liberation of poisonous hydrogen cyanide from cassava and hence making it safe for human consumption. It does suffice to note that cassava hugely appeals to low income households because it can be “banked” in the soil as a reserve food. This, coupled with flexible and low input requirements, make it popular to women in the rural communities (Nweke et al., 2002).

2.3 Challenges to optimal cassava productivity in Africa

After the introduction of cassava in Africa, it rapidly spread within the farming systems (Jones, 1959; Carter et al., 1992; Hershey, 2010) and is now a well established crop over vast agro-ecologies on the continent. Just like other introduced crops, increased cultivation resulted into the onset of production constraints. Attacks by insects, mites and pathogens often increase especially when natural control systems have been disrupted (Odongo and Otim-Nape, 1984).

In Latin America, where cassava has had a long history, there has been a co-evolution of the crop and its pests over a long period of time. Low to intermediate levels of resistance to prevalent pests are common there. However, in Africa and Asia there have been more serious new encounters of the crop with pests (Hershey, 2010). The long growth period of cassava that can extend up to 12 months in the field (IITA, 1990), inevitably leaves it vulnerable to overlapping attack by pathogens and insects.

Cassava production in Africa is constrained by a number of biotic (insects pests, diseases, weeds and nematodes) and abiotic stresses (soil fertility problems, drought and post-harvest deterioration), with their distribution and impact varying across the continent (IITA, 1990; Hershey, 2010). Viruses are among the most devastating pests of cassava. CMD, first reported in Africa in 1884, is one of the major biotic constraints that has been associated with the crop for a long time (Legg and Fauquet, 2004). To date, distinct species of these viruses are reported to infect cassava in Africa and India, where the species can interact synergistically, making its control difficult. Nonetheless, use of resistant varieties is the mainstay in defence against CMD (IITA, 1990; Hershey, 2010).

The recent emergence and spread of another viral disease, cassava brown streak disease (CBSD) is causing significant yield losses in the crop (Hillocks et al., 2001; Alicai et al., 2007). The disease causes a dry necrotic rot in the storage roots leading either to complete spoilage or significant reductions in quality. CBSD, first described in east Africa close to 60 years ago (Nichols, 1950), was thought to be restricted to coastal areas of Kenya, Tanzania and Mozambique. However, CBSD spread to mid altitude areas including Uganda (Alicai et al., 2007) and is now an eminent threat to cassava productivity in the region. Although some cassava genotypes with high levels of tolerance to CBSD have been identified in Tanzania (Edward Kanju, personal communication), tolerance/resistance to CBSD is yet to be detected in germplasm from other countries where CBSD is increasingly becoming a problem. Several other significantly important biotic constraints, particularly insect pests, bacterial and fungal pathogens have been noted in several African countries (IITA, 1990; Hershey, 2010).

Cassava roots, the principle economic part of the plant, have a short shelf life. Within one or two days after harvesting, there is rapid initiation of post-harvest physiological deterioration (PPD), which is associated with the synthesis of phenolic compounds (Beeching et al., 1998). PPD remains a huge challenge in the commercialisation of cassava. Most research on PPD has been conducted in Latin America, where genetic variability has been reported in populations including some inter-specific hybrids (Hershey, 2010). Stem storability, which is broadly defined as the capacity of stems to withstand long storage periods (that can extend up to two months after harvesting) is another major challenge to cassava productivity, particularly in areas with relatively long dry spells or erratic rainfall (Ceballos et al., 2004).

Though cassava is considered a drought tolerant crop, its resilience to weather extremes in the current face of global warming, will increasingly become a major abiotic constraint in the not too far distant future. All production constraints of cassava can not be highlighted here. The above narration presents some of the major constraints to optimal cassava productivity. Of hope however, is the premise that solutions to some of these challenges can be addressed through breeding interventions.

2.4 Genetic variation: a tool for cassava improvement

The array of cassava challenges highlighted above necessitate that concerted efforts be made to address them with the overall goal of increasing cassava productivity in Africa. However, the success of any breeding programme, whether customised towards hybrid or variety development, will require that maximum diversity of parental lines exists to either exploit heterosis or provide variability for additive variance for selection (Wricke and Weber, 1986; Maunder, 1992). Heterogeneous populations are a useful buffer against biotic and/or abiotic extremes. The increase of more than 50% cereal yield production between the periods 1960-1990 that was ascribed to adoption of better crop cultivars, is a testimony to this fact (Frey, 1992).

However, genetic variation alone is practically worthless in germplasm unless it harbours genes that are useful either singly or in combination with other previously evaluated germplasm (Smith and Duvick, 1989). In cassava, this is strongly illustrated by international breeding efforts conducted by the International Centre for Tropical Agriculture (CIAT) in Latin America and Asia that began with the collection and evaluation of over 2000 cassava varieties, mainly from farmer's fields in Latin America. Selected varieties were hybridised to generate progeny for further advancement from which outstanding commercial genotypes were officially released and widely adopted (Kawano, 2003). In Africa, the classic example of the use of genetic variation is illustrated by the germplasm derived from the east African breeding programme in the 1930s (Nichols, 1947; Jennings, 1957; Hershey, 2010). This programme searched for CMD resistant clones and selections with higher levels of resistance were intercrossed to get highly resistant hybrids that were distributed in the region to reduce ravages of CMD (Jennings, 1957).

Chang (1992) noted that a complete array of germplasm in a crop will consist of: 1) wild relatives, 2) unimproved cultivars or local varieties and 3) improved germplasm already in production. In cassava, this broad categorisation does exist and has been used in hybridisation programmes to generate new genetic variability (Jennings 1959; Hahn et al., 1980; Jennings and Iglesias, 2002; Kawano, 2003; Ojulong et al., 2008). Naturally occurring genetic variation in key agronomic and root quality traits of unimproved local cassava genotypes from Latin America has been reported (Chávez et al., 2005; Sánchez et al., 2009). The contribution of formal breeding initiatives by the International Institute of Tropical Agriculture (IITA), CIAT and the National Agricultural Research Systems (NARS), combined with the heterozygous nature of the crop and the inherently natural variation in cassava, inevitably resulted in broadened genetic variation in cassava, some of which is represented in the eastern, central and southern African regions. This genetic variation needs to be systematically quantified for optimal utilisation.

2.5 Quantification of genetic variation

Within the realms of genetics, quantification of variation is specifically important to obtain insights into evolutionary forces (mutation, selection, migration, recombination and random genetic drift) that shape today's population structure (Hartl, 2000; Klug et al., 2005). Genetic variation manifests itself at various levels of biological organisation and/or expression including ecological adaptation, chromosome structure and behaviour, biochemical pathways, morphological traits (qualitative), agronomic and consumer related traits (quantitative) and molecular variation (Hershey, 2010). Ecological adaptation largely describes the distribution of species within the genus (Rogers and Appan, 1973). Chromosome structure and behaviour that mainly involves utilisation of cytogenetics to infer the organisation of genetic diversity has provided conflicting results (Magoon et al., 1969; Umannah and Hartman, 1973; Chavarriaga-Aguirre et al., 1998). Because of their limitations, both ecological adaptation and cytogenetics have not been used in examining genetic diversity in cassava. Hence, hereafter quantification of variation using biochemical, morphological, agronomic and molecular approaches is discussed.

2.5.1 Biochemical variation

Principally, biochemical variation relies on protein polymorphisms and was first used to study populations of *Drosophila* in the 1960s (Klug et al., 2005). Considering a structured gene, if a nucleotide change results in the substitution of a charged amino acid, such as glutamic acid, for an uncharged amino acid, such as glycine, the net electrical charge on the protein will be altered (Klug et al., 2005). This difference in charge and size is used to separate protein molecules through an electric field. Isozyme proteins have been used to study genetic diversity in cassava (Ramirez et al., 1987) with phenomenal findings. For instance, the isozyme techniques were able to detect intermediate genotypes between *M. esculenta* and *M. glaziovii*, which can be attributed to evidence of gene flow between species. Seed storage proteins in the genus *Manihot* were used to compare 19 *Manihot* species of Brazilian origin (Grattapaglia et al., 1987).

There are a limited number of enzymes that can be studied and because many substitutions do not change the net electric charge on the molecule, only about 30% of the variation is detected (Klug et al., 2005). This severely limits the utility of protein-based polymorphisms in diversity assessment. Besides, protein synthesis depends on a particular gene being active, which is a function of plant age, origin and environmental factors. These extraneous factors either singly or in combination, if not well attended to, can lead to largely biased results. This could partly explain the limited utility of protein polymorphisms in current diversity assessment.

2.5.2 Morphological variation

The phenotype of a plant is of great agricultural and economic importance. Morphological traits, which are largely qualitative, are frequent in nature and considerable variation exists in them (Chawla, 2002). These traits, that display distinct phenotypes, were the earliest genetic markers to be used in cassava (Graner, 1942). These markers and/or descriptors are largely under monogenic gene control, with limited environmental influences and are often used in the certification of new varieties. However, under some exceptional cases modifier genes may cause some slight variation in the phenotype (Graner, 1942).

The International Plant Genetic Resources Institute (IPGRI) defined a set of relatively stable morphological traits useful for cassava characterisation (Gulick et al., 1983). This set comprised of 11 traits: apical leaf colour, colour of the petiole, stem epidermis colour, root flesh colour, root peduncle, shape of central leaf lobes, apical pubescence, stem periderm colour, root surface colour, flowering and root cortex colour, with distinct classifications within each. Another comprehensive descriptor list developed by Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), written in Portuguese (Documentos – CNPMF No. 78, ISSN 0101-5171-JUNHO/1998), has also been used to characterise cassava.

Genetics of a few morphological traits in cassava have been documented (Granar, 1942; Hershey and Ocampo, 1989). The authors came to the following conclusions about cassava traits: 1) leaf shape - narrow is dominant over broad, 2) stem periderm - light green is dominant over dark green, 3) stem growth habit - straight is dominant over zigzag, 4) root periderm - dark is dominant over white, 5) leaf margin - pandurate is dominant over entire and 6) parenchyma colour - yellow is dominant over white. Partial dominance has also been reported for parenchyma colour.

Some of the qualitative traits are of major agronomic importance. For example, root shape is a valuable indicator for maturity, as some short-rooted varieties will produce roots of commercial value within a shorter time compared to long-rooted varieties (Hershey, 2010). The cassava variety CMC-40 is widely grown in southern Brazil owing to its short roots that thicken quickly. Canopy characters also have agronomic relevance as they significantly influence the quality of planting material. For example, clones with limited branching produce more uniform stakes than those that are highly branched (Hershey, 2010).

Other morphological traits like leaf-vein colour have been used to establish whether or not progeny have resulted from cross-pollination or self-pollination (Kawano et al., 1978), but some are probably evolutionary neutral (e.g. stem periderm colour and colour of petiole). Ceballos et al. (2004) observed that good cooking quality is usually associated with other morphological traits such as colour of the peel and that farmers frequently reject changes in morphological traits. In other crops, three morphological markers were used in addition to molecular markers for genetic linkage mapping of diploid wheat *Triticum monococcum* (Link) Thell.) (Dubcovsky et al., 1996).

Various studies have employed cassava morphological traits with the objective of elucidating patterns of genetic variation (Benesi, 2005; Balyejusa Kizito, 2006; Zacarias, 2008). A deficiency of most of these studies was that they largely focused on above-ground qualitative traits (e.g. leaf pubescence, leaf shape, leaf colour, stem growth, branching habit etc.), with a few root traits.

However, to obtain the most from diversity assessment, it is necessary to broaden the traits for characterisation and provide for incorporation of indigenous technical knowledge. A thorough understanding of monogenic traits in cassava is desirable, as it can indirectly help the selection process, where breeders handle several highly heterozygous individuals at a given time in the breeding scheme.

2.5.3 Quantitative variation

Just like for other plants, most useful agronomic traits of cassava exhibit quantitative inheritance (IITA, 1990; Ceballos et al., 2004; Hershey, 2010). These traits have been manipulated in cultivated species for their adaptive and/or commercial value. Quantitative genetic variation is the basis of productive and reproductive traits and monitoring it may therefore reveal variation closely related to fitness. In order to understand variation, efforts should be made to obtain accurate phenotypic records, as this is pivotal in obtaining true genetic progress. Even with the availability of molecular techniques, visible trait expression still remains the most practical means of evaluating a phenotype (Hershey, 2010).

Quantitative traits are well known to be influenced by genotype by environment interactions (Wricke and Weber, 1986). In cassava, the extent of environmental influences varies among agronomic traits (Kawano, 2003; Ojulong, 2006; Ssemakula and Dixon, 2007). For example, in the analysis of 175 clones at two locations, significant genotype by environment interactions were only observed for roots per plant and not for storage root weight, harvest index (HI), root dry matter content (DMC), fresh root yield and dry root yield (Ojulong, 2006). However, genotype differences were observed for all traits evaluated. In carotenoid-rich cassava clones evaluated at five locations in Nigeria, genotypic effects had the highest impact on DMC, location effects had highest impact on dry root yield, while genotype x location effects were significant for DMC and all other traits analysed (Ssemakula and Dixon, 2007). Kawano (1987) observed that heritabilities of agronomically important traits including DMC and HI are sufficiently high to warrant predictable performance of hybrids.

Key agronomic traits that have frequently been studied in cassava include plant height, storage roots per plant, HI, DMC, PPD, hydrogen cyanide content and several plant health and quality-related traits (Ceballos et al., 2004; Chávez et al., 2005; Sánchez et al., 2009; Hershey, 2010). Plant height in cassava has been examined for various objectives including compatibility with intercropping systems (Eke-Okoro et al., 2008), drought response (Aina et al., 2007), mapping quantitative trait loci (QTL) (Okogbenin and Fregene, 2003) and for understanding its genetics and variation in populations (Calle et al., 2005; Cach et al., 2006; Rojas et al., 2009). Several other quantitative traits including storage roots per plant, root weight, fresh shoot yield and fresh root yield have been examined with the aim of understanding various aspects of cassava productivity (Pinho et al., 1995; Okogbenin et al., 2003; Ojulong, 2006).

Cassava, being a starchy crop, suggests that since its domestication, both human and natural evolutionary selections forces have played a significant role in determining the root DMC. For instance, in Colombia, high DMC varieties are more often associated with communities where cassava is processed, while intermediate to high DMC varieties are associated with regions where the roots are directly consumed (Hershey, 2010). DMC has been extensively studied in cassava for quantifying: 1) the extent of genetic variation (Kawano, 2003; Chávez et al., 2005), 2) genotype x environment effects (Kawano et al., 1987; Tan and Mak, 1995; Benesi, 2005; Ojulong, 2006; Ssemakula and Dixon, 2007), 3) nature of its inheritance (Jaramillo et al., 2005; Cach et al., 2006), 4) extent of gene transfer from wild relatives to cassava (Ojulong et al., 2008), 5) quantitative trait loci (Okogbenin and Fregene, 2003; Kizito et al., 2007) and 6) response of cassava to drought stress (Okogbenin et al., 2003). With the exception of co-workers of Kawano as well as Chávez, most of the abovementioned studies largely examined DMC in a few elite genotypes and/or breeding populations with limited consideration of local varieties. Most of these studies involved analysis of cassava germplasm from Latin America and/or Asia.

The distribution of biomass to economically useful plant parts is measured by the HI and in cassava, this is defined as the ratio of storage root weight to the total weight of the plant on a fresh weight basis (Kawano, 1990). HI is one of the agronomical traits that can substantially increase cassava productivity. Another important finding from these breeding experiments was the fact that in single row trials, indirect selection for yield through HI was more effective than direct selection for yield (Kawano, 2003).

Several studies examined HI in cassava with the objective of: 1) mapping QTL (Okogbenin et al., 2008), 2) understanding how it is affected by the soil tillage systems (Otsubo et al., 2008), 3) establishing relationships among cassava clones (Nick et al., 2008), 4) quantifying its variation among genotypes and in making early selections (Ramanujam et al., 1989; Tan, 1992), 5) quantifying environmental influences (Asadu et al., 2002; Egesi et al., 2007; Vidigal et al., 2007), 6) quantifying its heritability at different selection stages (Kawano et al., 1998), 7) quantifying fertiliser response (Pellet and El-Shakawy, 1993) and 8) determining its inheritance (Calle et al., 2005; Jaramillo et al., 2005; Cach et al., 2006). These studies largely examined HI in a few elite genotypes and/or breeding populations, with limited consideration of local genotypes. As for DMC, no systematic evaluation has been done for HI in cassava germplasm available within the NARS of eastern and southern Africa.

Compared to other staple crops in SSA, cassava performs relatively better on marginal soils characterised by uncertain rainfall patterns. The ability of cassava to tolerate these harsh conditions is purported to be an inter-play of several physiological traits (Hershey, 2010). The crop experiences simultaneous growth and development of the roots and the photosynthetic machinery, the leaves. Some studies have suggested that increased longevity of leaves and/or improved leaf retention could be important in attaining high yields in cassava (Cock and El-Sharkawy, 1988; Lenis et al., 2006) and that a simple visual evaluation of leaf retention can be applied in the field (Lenis et al., 2006). The attainment of stable cassava yields in marginal areas will require that physiological related traits like leaf retention be included during the selection process.

Lenis et al. (2006) evaluated 1350 clones under stress conditions and observed that clones containing the leaf retention trait produced more total fresh biomass and yielded 33% more root DMC than clones without the trait. Unfortunately, no systematic evaluation has been made for leaf retention in cassava germplasm available within the NARS of eastern and southern Africa.

2.5.4 Molecular variation

Molecular variation reflects naturally occurring DNA polymorphisms, which can be detected using several techniques. Irrespective of the technique used, information generated on DNA polymorphisms can be used to infer differences and/or similarities among individuals (Burr, 1994). DNA polymorphisms are more frequent than charge changes in proteins and/or phenotypic differences. A wide range of molecular techniques are available that enable detection of DNA polymorphisms. This detected variation has found diverse applications in genetic diversity analysis, construction of genetic maps, diagnostics, detection of linkage disequilibrium, gene cloning and whole genome scanning (Gut, 2001; Buckler and Thornsberry, 2002; Rafalski, 2002).

In practice, most DNA-based marker systems employ either the polymerase chain reaction (PCR) technique (Mullis, 1990) or the DNA-DNA hybridisation technique, which was adopted following the construction of a genetic linkage map in humans using the restriction fragment length polymorphism (RFLP) technique (Botstein et al., 1980). SNP markers (Raflaski, 2002; Flint-Garcia et al., 2003; Shastry, 2003) and microarray platforms (Wenzl et al., 2004; Hurtado et al., 2008) are recent additions to the array of molecular marker systems used in studying population variations. These DNA-based markers can either have dominant or co-dominant inheritance and are able to detect single locus and/or multiple locus differences (Chawla, 2002).

2.5.4.1 Molecular variation at genotype level

Genotype, referring to the combination of alleles situated on corresponding chromosomes, can be used to infer differences among individuals based on information generated from genetic markers (Hartl, 2000). Several molecular marker techniques have been developed and utilised to meet various objectives, but with the bulk of the work skewed towards quantification of genetic variation.

In the past three decades, most commonly used marker systems have included: 1) RFLPs; Botstein et al., 1980; Tanksley et al., 1989; Livini et al., 1992), 2) random amplified polymorphic DNA (RAPD; Williams et al., 1990; Stojsin et al., 1996; Sun et al., 2001), 3) amplified fragment length polymorphism (AFLP; Vos et al., 1995; Lin et al., 1996), SSRs (Litt and Lutty, 1989; Morgante and Olivieri, 1993; Powell et al., 1996) and DNA sequencing (Kreitman, 1983; Burr, 1994). Several derivatives of these marker systems including non-PCR based, PCR-based and targeted PCR techniques have emerged and continue to evolve (Burr, 1994; Hartl, 2000; Chawla, 2002).

Beeching et al. (1993) used cloned cassava genes involved in the cyanogensis pathway as RFLP markers to obtain genetic relationships between cassava and its wild relatives. From that study three distinct clusters that depicted three *Manihot* species (*M. esculenta*, *M. glaziovii* and *M. caerulescens* Pohl) were revealed, with inter-specific hybrids clustering in intermediate positions between the parent species. Other studies employing RFLPs have focussed on phylogeny (Haysom et al., 1994; Colombo et al., 2000) and in the construction of a cassava linkage map (Fregene et al., 1997). RAPDs have been used to study the genetic diversity of cassava in Latin America (Colombo et al., 2000), Africa (Marmey et al., 1994; Tonukari et al., 1997; Zacarias et al., 2004) and in the construction of a cassava linkage map (Fregene et al., 1997). AFLPs have been used to establish genetic relationships among *Manihot* species (Second et al., 1997), to provide evidence of introgression between cassava and wild relatives (Morillo et al., 2001) and in genetic diversity assessment among cassava varieties (Sanchez et al., 1999; Fregene et al., 2000; Benesi, 2005).

SSR makers that comprise of tandem repeats of short [2-6 base pairs (bp)] DNA sequences are abundant in the genome, co-dominantly inherited, highly polymorphic and reproducible (Morgante and Olivieri, 1993; Powell et al., 1996). This high level of polymorphism arising from site-specific length variation of the repeat units (Morgante and Olivieri, 1993) makes them ideal for studying populations. Moreover, a comparison of RFLP, RAPD, AFLP and SSR marker systems for germplasm analysis has confirmed the superiority of SSR markers (Powell et al., 1996).

Because of their informativeness, SSRs have had various applications in cassava: 1) assessment of genetic diversity (Chavarriaga-Aguirre et al., 1998; Balyejusa Kizito et al., 2005; Moyib et al., 2007; Hurtado et al., 2008; Siqueira et al., 2009), 2) construction of linkage maps (Fregene et al., 1997; Mba et al., 2001; Okogbenin et al., 2006), 3) mapping of QTL (Akano et al., 2002; Okogbenin and Fregene, 2003; Ojulong, 2006; Kizito et al., 2007; Okogbenin et al., 2008) and 4) tracing the origin of cassava (Olsen, 2004). From these studies it was established that: 1) cassava is highly diverse and native to the southern Amazon basin and 2) with more fine-mapping, SSRs have the potential to be used in the selection of agronomically important quantitative traits. DNA sequencing that enables identification of SNPs and/or haplotypes has also been used in cassava (Olsen, 2004; Lopez et al., 2005; Kawuki et al., 2009). Evidently, these molecular markers have contributed immensely to the understanding of cassava evolution and genetics and still offer tremendous scope to the understanding of cassava genetics and its improvement.

2.5.4.2 Molecular variation at sequence level

The development of cloning and DNA sequencing techniques provided opportunities for direct analysis of sequence variations of individuals in a population (Klug et al., 2005). Nuclear sequence variations in the form of SNPs provide excellent opportunities for studying phenotypes in populations, as they are the most common form of genetic variation (Brookes, 1999).

SNP polymorphisms can either be in the form of transitions (purine to purine or pyrimidine to pyrimidine) or in the form of transversion (i.e. a purine to a pyrimidine), a scenario which has been reported both in plants (Van et al., 2005) and humans (Wang et al., 1998). This kind of genetic variation provides an opportunity for direct analysis of sequence differences (both in genic and regulatory regions) between many individuals at a large number of loci, providing more insights into population diversity (Rafalski, 2002).

Considerable progress in SNP technology has been achieved within the realms of human genetics (Davignon et al., 1988; Fullerton et al., 2000; Gut, 2001; Collins et al., 2004; Klug et al., 2005). Key findings in some of these studies that could be of relevance to plants were that SNPs: 1) occur at a high frequency in the human genome (one in every 1000 bp), 2) provide the basis of understanding how genetic differences influence a phenotype, i.e. an individual's susceptibility to disease and response to drugs, 3) do not necessary cause disease, but can act as markers for populations at risk of developing a disease, 4) found particularly in the untranslated regions and/or non-coding regions, do not necessarily associate with a phenotype and 5) frequency varies greatly among genes.

In plants, the most practical utilisation of SNPs will primarily be in understanding crop genetics for eventual crop improvement through mapping of traits, construction of high resolution genetic maps, genetic diagnostics, analysis of population structure, phylogenetic analysis and association mapping (Buckler and Thornsberry, 2002; Flint-Garcia et al., 2003; Neale and Savolainen, 2004). Moreover, DNA sequence-based diversity will provide insights into plant selection, migration, recombination and mating systems (Buckler and Thormsberry, 2002). For example, studies of gene sequence diversity have established that some species like maize are more polymorphic, while others like melon are less polymorphic (Shattuck-Eidens et al., 1990). Related studies involving analysis of single genes were able to identify polymorphic sites undergoing selection and even relate polymorphisms to useful agronomic phenotypes (Buckler and Thornsberry, 2002).

SNPs have further been utilised in the estimation and examination of linkage disequilibrium (LD) i.e. the non-random association of alleles at different loci (Remington et al., 2001). In that study, intragenic and genome wide LD between SNPs in a diverse set of maize inbred lines across six genes established a rapid decay ($r^2 < 0.1$; within 1500 bp). However, there were loci e.g. the *sugary1* where decay in LD extended over 12 kb. These findings have practical relevance towards the implementation of SNPs for association mapping i.e. whether to select a whole genome scan or a candidate gene approach (Buckler and Thornsberry, 2002).

SNPs have been used in a few crops to assess diversity (grapevine, Salmaso et al., 2004; rice, Bao et al., 2006 and maize, Hamblin et al., 2007) and haplotype structure (grapevine, Salmaso et al., 2004 and sugar beet, Schneider et al., 2001). SNPs have been applied in phylogeographic analyses (Brumfield et al., 2003; Olsen, 2004). Other studies have characaterised the frequency of SNPs in genomes, for example one SNP per 78 bp for grapevine (Salmaso et al., 2004), one SNP per 31 bp in non-coding regions and one SNP per 124 bp in the coding regions for maize (Ching et al., 2002), one SNP per 2038 bp in the coding sequence and one SNP per 191 bp in the non-coding regions in soybean (Van et al., 2005). A key distinction between SNPs and other DNA-based marker systems, particularly AFLPs and genomic SSRs, is that the latter marker systems are limited to indirect analysis of DNA sequence variations. SNPs offer a direct way to sequence variation analysis, which to a large extent helps to explain observed phenotypes. Moreover, SSR markers are subject to homoplasy, which is the occurrence of SSR alleles of identical size, but of different evolutionary origin (Viard et al., 1998).

SNPs associated with an agronomic phenotype will considerably boost plant breeding efficiency, similar to what has been achieved in human diagnostics and therapy recommendations. However, SNP information is limited in cassava (Olsen, 2004; Lopez et al., 2005). Given the heterozygous nature and complexity associated with selection for some of the traits like CBSD, initiation of SNP based research will make a significant contribution to cassava improvement.

2.6 Cassava breeding

Plant breeding, defined as the art and science of changing heredity of plants, has immensely contributed towards human welfare. Technologies generated through plant breeding have played a significant role in providing food, feed and fibre to the continuously increasing human population (Frey, 1992). Regardless of the crop, plant breeding progress will largely depend on the selection and utilisation of the most appropriate germplasm and the particular breeding methods which can be applied (Wricke and Weber, 1986; Baenziger and Peterson, 1992; Hallauer, 1992; Maunder, 1992).

In cassava, the domestication process, which in part involved selections by subsistence farmers, began the era of cassava breeding. However, the advent of Mendelian and quantitative genetics, where more systematic selection and genetic progress could be attained, motivated the initiation of formal breeding in cassava. Involvement of trained plant breeders in cassava improvement began in the early twentieth century, with the breeding programmes that were established in India, Brazil, Madagascar, Tanzania, Nigeria and Indonesia (Hershey, 2010). Subsequent breeding activities in Africa were introduced by the French and Belgians in Senegal, DRC and Côte d'Ivoir. Although the Madagascar breeding programme was based on high input conditions and involved a wide range of germplasm and large numbers of seedlings being tested each year, it terminated before any major impact could be made. The most successful early breeding in Africa was that based at Amani in Tanzania under the guidance of Drs. H.H. Storey and R.F.W. Nichols (Hershey, 2010).

This regional programme in Tanzania, established in the mid 1930s, involved breeding for CMD and CBSD resistance, with outstanding clones being distributed among the neighbouring countries (Storey and Nichols, 1938; Nichols, 1947). In 1956, one year before the Amani station was closed, segregating populations were distributed to several African countries.

Later, in the 1960s, two international centers, IITA and CIAT were established and furthered cassava breeding in partnership with the NARS of Africa, Asia and Latin America (Kawano, 2003; Ceballos et al., 2004; Hershey, 2010). The principal output of most cassava breeding programmes has been information and genetically improved germplasm. Most of the information generated is intended for use by breeders and other scientists, while improved germplasm, usually in the form of new varieties, provides a measurable benefit to producers and consumers (Hershey, 2010).

The success of a developed cassava variety can be examined in two ways. Firstly, when the variety is predominantly grown over a large contiguous area. And secondly, when the variety may not predominate in any one area, but is moderately successful in several countries, i.e. have wide geographical adaptation and adoption (Hershey, 2010). In Latin America, the variety *Mantiqueira* (syn. CMC 40), bred by the Instituto Agronômico de Campinas (IAC), São Paulo, Brazil, has been moderately successful in southern Brazil, Cuba, Colombia, Dominican Republic, Haiti and the Philippines.

In Africa, the IITA bred variety TMS 30572 that combines both CMD resistance and desirable agronomic quality was successfully adopted in Nigeria and in several other African countries to mitigate the effects of CMD. For the Asian continent, the Thai national programme selected Rayong I from among local varieties and released it in 1975. For a period of over 20 years, nearly all Thailand's cassava acreage was planted with this variety (Hershey, 2010). Currently, the Thai variety KU50 is grown on more than one million hectares in Asia and is the most widely grown variety in Thailand. At a national level, NARS have developed and released several varieties. In Africa alone, more than 200 varieties have been released by the different NARS. However, because of limited documentation on area planted to specific varieties, it is extremely difficult to classify the degree of success enjoyed by most of these released varieties in Africa (Hershey, 2010).

2.6.1 Breeding objectives

Breeding objectives across countries appear to be relatively similar as documented by the several workshops and symposia that bring together scientists from IITA, CIAT and NARS (Hershey, 2010). Nearly all programmes include among their objectives high yield, high DMC, early maturity, tolerance/resistance to local pests and diseases and adaptation to local environmental conditions. However, the relevance of other quality-related cassava traits (i.e. PPD, hydrogen cyanide levels, beta-carotene and amylose/amylopectin content) vary greatly according to processing requirements and end-user needs (Ceballos et al., 2004). Because intercropping is practised in many rural communities, varieties compatible with intercropping also constitute a breeding objective in some communities (Ceballos et al., 2004). It does suffice to note that specific growing, processing and marketing situations require customised objectives for individual countries and/or regions. Principally, attainment of these breeding objectives will require that sufficient genetic variation exists and that appropriate breeding techniques be applied (Frey, 1992; Maunder, 1992).

2.6.2 The selection process

Breeding methods developed for cross-pollinated crops can practically be applied to cassava (Ceballos et al., 2004; Hershey, 2010). The breeding methodology in cassava involves selection of parents (based on complimentary traits), crossing (via controlled or open pollinations) and simple phenotypic selection of individual clones based on performance across years and locations. The vegetative nature allows maintenance of heterozygotes throughout the selection process. However, selection of parents for hybridisation should not be based solely on their *per se* performance, but rather on their combining ability, whose estimation will require utilisation of specific mating designs (Ceballos et al., 2004; Ojulong, 2006).

A modified cassava selection scheme that comprises six major steps, with each activity lasting a year, has been suggested (Ceballos et al., 2004). Year one involves generation of crosses to produce about 100000 genotypes. Year two involves evaluation of the F_1's (17500-100000 genotypes) on a single plant basis. During year three unreplicated clonal evaluation trials (1800-3000 genotypes) are established. Replicated and/or unreplicated single-row plots in a preliminary yield trial (100-300 genotypes) are established during year four. Year five involves the establishment of a replicated advanced yield trial (18-100 genotypes) and year six the establishment of replicated regional trials (5-30 genotypes).

Only with the initiation of replicated trials does the emphasis shift from high heritability traits to those of low heritability, such as yield. Ceballos et al. (2004) further discussed the utilisation of selection indices and estimation of parental combining abilities from the collected data. One particular feature of cassava breeding is that it requires considerable time, resources and logistics. It is partly for these reasons that marker-assisted selection (MAS) in cassava is being strongly advocated to augment the efficiency of the cassava selection process (Fregene et al., 2001; Setter and Fregene, 2007).

The relative efficiency of MAS compared to phenotypic-based selection is high when the trait of interest has low narrow sense heritability (h^2) under field conditions and when the ratio variance explained by the molecular marker compared to total additive genetic variance is high (Setter and Fregene, 2007; Hershey, 2010). MAS can be justified for: 1) evaluations based on a single plant, 2) screening disease resistance when the pathogen pressure is absent or low, as is the case with CMD in the Neo-Tropics, 3) using highly variable experimental fields like the case of largescale drought trials and 4) traits that are affected by the plant's growth stage, like DMC and CBSD. Besides enhancing h^2, markers considerably reduce the size of breeding populations by eliminating undesirable genotypes at seedling stage (Hershey, 2010). This selection considerably reduces the field establishment and evaluation logistics.

An example of the utility of MAS in cassava was that associated with the discovery of a single dominant gene, designated as *CMD2* (Akano et al., 2002). The authors identified three markers of which RME1 and NS158 at respective distances of 2 cM (centiMorgan) and 5 cM, were more strongly associated with the *CMD2* gene. These findings inspired CIAT and IITA to verify the utility of MAS for CMD resistance screening using the SSR marker NS158. This was done using six families with progeny sizes ranging from 36-840. IITA evaluated a total of 2490 genotypes in unreplicated field trials for resistance to CMD, while CIAT assayed the genotypes with NS158 using polyacrylamide gel analysis (Hershey, 2010). Results of the marker analysis and phenotypic evaluations revealed that NS158 offered excellent prediction for CMD resistance in some crosses, but not in others (Hershey, 2010). Further scrutiny of the marker alleles for the parents revealed that an allele from the susceptible parent had the same size as the allele associated with *CMD2* in the resistant parent. Though these findings demonstrated the potential of MAS in cassava, they also illustrated the need to develop many markers around a gene of interest to increase the precision of MAS. Another study employed MAS to transfer useful genes from wild relatives to cassava (Hershey, 2010).

Several other related studies have identified QTL in cassava. Jorge et al. (2000) identified eight QTL for resistance to cassava bacterial blight disease. In another study, two QTL on two different linkage groups controlling cyanogenic glucosides and six QTL on four different linkage groups controlling DMC, were identified (Kizito et al., 2007). Genetic mapping of QTL affecting productivity and plant architecture have been examined and 30 primary QTL and 84 secondary QTL were detected (Okogbenin and Fregene, 2003).

2.6.3 New approaches to cassava breeding

Since the inception of cassava breeding, utilisation of heterozygous parents to generate F_1's for onward evaluation and selection has been routine. Limitations on attainment of sustainable genetic progress due to the inherently high heterozygosity in cassava have been highlighted including: 1) masking allelic differences in segregating populations, 2) permitting a sizeable genetic load of deleterious alleles to persist in populations and 3) limiting the transfer of desirable traits from one genotype to another.

In response to these challenges, introduction of double haploids and/or inbreeding were proposed as solutions (Ceballos et al., 2004). These two approaches, both of which result in increased homozygosity, comprise part of the new approaches to cassava breeding. Complete or partial homozygosity have several advantages including: 1) elimination of deleterious recessive genes, 2) fixation of dominant alleles for future recurrent selection breeding, 3) precise production of hybrids, 4) ensuring consistent cumulative genetic progress, 5) expression of useful recessive traits, 6) making the backcrossing scheme possible as a breeding method, 7) enhanced facilitation for germplasm exchange and 8) containment of viral pathogens (Ceballos et al., 2004).

Maize, a heterozygous crop, stands out as a testimony for the commercial use of inbreds. Historical developments in maize breeding have by far overridden the initial discouraging results of introducing inbreeding in maize. Begg (1959) highlighted that at the inception of inbreeding in maize, of the original hand pollinated plants, only four lines survived by 1912, some of which yielded only two bushels per acre (125.4 kg/ha). This is by far less as compared to today's hybrids which yield up to 7247 kg/ha (Fabijanac et al., 2006). This commitment to promote inbreeding in maize despite the earlier discouraging results is needed for cassava.

Some scientists expected that cassava, being highly heterozygous, will be sensitive to inbreeding (Hershey, 2010). However, current quantitative evidence suggests that inbreeding depression (ID), general loss in fitness and/or vigour amongst inbred progeny, is not uniform across cassava traits and families (Rojas et al., 2009) and hence can still be considered. From that study, average estimates of ID were 63.9% for fresh root yield, 37.9% for fresh foliage yield, 26.5% for HI, 10.1% for plant height and 5.3% for root DMC. To date, the principal outputs from cassava inbreeding have been the discovery of an amylose-free starch mutant (Ceballos et al., 2007) and identification of induced mutants (Ceballos et al., 2008).

Increasing homozygosity through undertaking conventional inbreeding provides an opportunity for selection. However, it will require between six to seven years for acceptable levels of homozygosity to be achieved. The double haploid option which theoretically can lead to attainment of homozygosity within one generation is another option which is being considered for cassava (Hershey, 2010).

2.7 Conclusions

It is apparent from the literature review that since the introduction of cassava into Africa in the early 1700s, various research activities have been conducted on the continent. These studies have attempted to address several objectives including breeding for pest and disease resistance, understanding epidemiology and population dynamics of pests, socio-economic issues within the cassava pipeline and genetic relations in cassava populations. Despite this progress that has substantially increased cassava productivity as compared to wild relatives, some research gaps do exist, either due to limited attention by previous studies or due to new interventions. These gaps form the basis of the four research objectives in this thesis.

2.8 References

Aina, O.O., A.G.O. Dixon, and E.A. Akinrinde, 2007. Effect of soil moisture stress on growth and yield of cassava in Nigeria. Pakistan Journal of Biological Science 10:3085-3090.

Akano, A.O., A.G.O. Dixon, C. Mba, E. Barrera, and M. Fregene, 2002. Genetic mapping of a dominant gene conferring resistance to the cassava mosaic disease (CMD). Theoretical and Applied Genetics 105:521-525.

Alicai, T., C.A. Omongo, M.N. Maruthi, R.J. Hillocks, Y. Baguma, R. Kawuki, A. Bua, G.W. Otim-Nape, and J. Colvin, 2007. Re-emergence of cassava brown streak disease in Uganda. Plant Disease 91:24-29.

Allem, A.C., 1994. The origin of *Manihot esculenta* Crantz (Euphobiaceae). Genetic Resources and Crop Evolution 41:133-150.

Allem, A.C., 1999. The closest wild relatives of cassava (*Manihot esculenta* Crantz). Euphytica 107:123-133.

Asadu, C.L.A., A.G.O. Dixon, and R. Okechukwu, 2002. Comparative evaluation of the contributions of soil physicochemical properties to variations in the yields of four major staple food crops in eastern Nigeria. Soil and Tillage Research 65:141-155.

Baenziger, P.S., and C.J Peterson, 1992. Genetic variation: its origin and use for breeding self pollinated species. In: Stalker, H.T., and J.P. Murphy, (Eds.), Plant breeding in the 1990s. Proceedings of the symposium on plant breeding in the 1990s held at North Carolina State University Raleigh, NC, pp. 69-92. CABI International Wallington Oxon OX108DE, UK.

Balyejusa Kizito, E., 2006. Genetic and root growth studies in cassava (*Manihot esculenta* Crantz): implications for breeding. PhD Thesis, Swedish University of Agricultural Sciences, Uppsala, Sweden, pp. 127.

Balyejusa Kizito, E., A. Bua, M. Fregene, T. Egwang, U. Gullberg, and A. Westerbergh, 2005. The effect of cassava mosaic disease on the genetic diversity of cassava in Uganda. Euphytica 146:45-54.

Bao, J.S, H. Corke, and M. Sun, 2006. Nucleotide diversity in *starch synthase IIa* and validation of single nucleotide polymorphism in relation to starch gelatinisation temperature and other physiochemical properties in rice (*Oryza sativa* L.). Theoretical and Applied Genetics 113:1171-1183.

Beeching, J.R., H. Yuanhuai, R. Gomez-Vazquaz, R.C. Day, and R.M. Cooper, 1998. Wound and defence responses in cassava as related to post-harvest physiological deterioration. In: Romeo, J.T., K.R. Downum, and R. Verpoorte (Eds.), Recent advances in Phytochemistry 23:231-248. Phytochemical signals in plant-microbe interactions, Plenum Press, New York.

Beeching, J.R., P. Marmey, M.C. Gavalda, M. Noirot, H.R. Haysom, M.A. Hughes, and A. Charrier, 1993. An assessment of genetic diversity within a collection of cassava (*Manihot esculenta* Crantz) germplasm using molecular markers. Annals of Botany 72:515-520.

Begg, C.M.M., 1959. An introduction to genetics. The English Universities Press Ltd, 102 Newgate Street, London, UK, pp. 291.

Benesi, I.R.M., 2005. Characterisation of Malawian cassava germplasm for diversity, starch extraction and its native and modified properties. PhD Thesis, University of the Free State, South Africa, pp. 199.

Botstein, D., R.L. White, M. Skolnick, and R.W. Davis, 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. American Journal of Human Genetics 32:314-331.

Brookes, A.J., 1999. The esscence of SNP's. Gene 234:177-186.

Brumfield, R.T., P. Beerli, D.A. Nickerson, and S.V. Edwards, 2003. The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology and Evolution 8:249-256.

Buckler, E.S., and J.M. Thornsberry, 2002. Plant molecular diversity and applications to genomics. Current Opinion in Plant Biology 5:107-111.

Burr, J., 1994. Some concepts and new methods for molecular mapping in plants. In: Phillips, R.L., and I.K. Vasil (Eds.), DNA-based markers in plants, pp. 1-7. Kluwer Academic Publishers.

Cach, N.T., J.I. Lenis, J.C. Perez, N. Morante, F. Calle, and H. Ceballos, 2006. Inheritance of useful traits in cassava grown in subhumid conditions. Plant Breeding 125: 177-182.

Calle, F., J.C. Perez, W. Gaitan, N. Morante, H. Ceballos, G. Llano, and E. Alvarez, 2005. Diallel inheritance of relevant traits in cassava (*Manihot esculenta* Crantz) adapted to acid-soil savannas. Euphytica 144:177-186.

Carter, S.E., L.O. Fresco, P.G. Jones, and J.N. Fairbairn, 1992. An atlas of cassava in Africa: Historical, agroecological and demographic aspects of crop distribution. CIAT, Cali, Columbia, pp. 85.

Carvalho, L.J.C.B., C.R.B. de Souza, J.C.M. Junior, and C.B. Campos, 2004. Identification and characterisation of novel cassava (*Manihot esculenta* Crantz) clone with high free sugar content and novel starch. Plant Molecular Biology 56:503-515.

Ceballos, H., C.C. Iglesias, J.C. Pérez, and A.G.O. Dixon, 2004. Cassava breeding: opportunities and challenges. Plant Molecular Biology 56:503-515.

Ceballos, H., T. Sánchez, A.P. Tofiño, E.A. Rosero, K. Denyer, A.M. Smith, D. Dufour, N. Morante, J.C. Pérez, and B. Fahy, 2008. Induction and identification of small-granule, high-amylose mutant in cassava (*Manihot esculenta* Crantz). Journal of Agricultural Food Chemistry 15:2715-2722.

Ceballos, H., T. Sánchez, N. Morante, M. Fregene, D. Dufour, A.M. Smith, K. Denyer, J.C. Perez, F. Calle and C. Mestres, 2007. Discovery of an amylose-free starch mutant in cassava (*Manihot esculenta* Crantz). Journal of Agricultural and Food Chemistry 55: 7469-7476.

Chang, L., 2000. New developments in the use of starch. In: Howeler, R.H., C.G. Oates, and G.M. O'Brien (Eds.), Cassava, starch and starch derivatives, pp. 5-11. Proceedings of the International Symposium held in Nanning, Guangxi, China.

Chang, T.T., 1992. Crop Genetic sources. In: Stalker, H.T., and J.P. Murphy (Eds.), Plant Breeding in the 1990s. Proceedings of the symposium on plant breeding in the 1990s held at North Carolina State University Raleigh, NC, pp. 83-103. CABI International Wallington Oxon OX108DE, UK.

Chavarriaga-Aguirre, P., M.M. Maya, M.W. Bonierbale, S. Kresovich, M.A. Fregene, J. Tohme, and G. Kochert, 1998. Microsatellites in Cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. Theoretical and Applied Genetics 97:493-501.

Chávez, A.L., T. Sánchez, G. Jaramillo, J.M. Bedoya, J. Echeverri, E.A. Bolaños, H. Ceballos, and C.A. Iglesias, 2005. Variation of quality traits in cassava roots evaluated in landraces and improved clones. Euphytica 143:125-133.

Chawla, H.S., 2002. Introduction to plant biotechnology. 2nd Edition. Science Publishers, Inc USA, pp. 538.

Ching, A., K.S. Caldwell, M. Jung, M. Dolan, O.S. Smith, S. Tingey, M. Morgante, and A.J. Rafalski, 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics 3:1-14.

Cock, J.H., and M.A. El-Sharkawy, 1988. Physiological characteristics for cassava selection. Experimental Agriculture 24:443-448.

Collins, A, W. Lau, and F.M. De la Vega, 2004. Mapping genes for common diseases: the case for genetic (LD) maps. Human Heredity 58:2-9.

Colombo, C., G. Second, and A. Charrier, 2000. Genetic relatedness between cassava (*Manihot esculenta* Crantz) and *M. flabellifolia* and *M. peruviana* based on both RAPD and AFLP markers. Genetics and Molecular Biology 23:417-423.

Davignon, J., R.E. Gregg, and C.F. Sing, 1988. Apolipoprotein E polymorphism and atherosclerosis. Arteriosclerosis 8:1-21.

Dubcovsky, J., M.-C. Luo, G.-Y. Zhong, R. Bransteiter, A. Desai, A. Kilian, A. Kleinhofs, and J. Dvorak, 1996. Genetic map of diploid wheat, *Triticum monococcum* L., and its comparison with maps of *Hordeum vulgare* L. Genetics 143:983-999.

Egesi, C.N., P. Ilona, F.O. Ogbe, M. Akoroda, and A. Dixon, 2007. Genetic variation and genotype x environment interaction for yield and other agronomic traits in cassava in Nigeria. Agronomy Journal 99:1137-1142.

Eke-Okoro, O.N., D.N. Njoku, and J.G. Ikeorgu, 2008. Evaluation of the productivity of low cyanide cassava varieties in cassava/soybean intercrop. Scientific Research and Essays 3:495-499.

Ellis, R.H., T.D. Hong, and E.H. Roberts, 1982. An investigation of the influence of constant and alternating temperature on the germination of cassava seed using a two-dimensional temperature gradient plate. Annuals of Botany 49:241-246.

Fabijanac, D., B. Varga, Z. Svecnjak, and D. Grbesa, 2006. Grain yield and quality of semiflint maize hybrids at two sowing dates. Agriculturae Conspectus Scientificus 71:45-50.

Flint-Garcia, S.A., J.M. Thornsberry, and E.S. Buckler, 2003. Structure of linkage disequilibrium in plants. Annual Review of Plant Biology 54:357-374.

Fregene, M., A. Bernal, M. Duque, A. Dixon, and J. Tohme, 2000. AFLP analysis of African cassava (*Manihot esculent* Crantz) germplasm resistant to the cassava mosaic disease (CMD). Theoretical and Applied Genetics 100:678-685.

Fregene, M., E. Okogbenin, C. Mba, F. Angel, M.C. Suarez, G. Janneth, P. Chavarriaga, W. Roca, M. Bonierbale, and J. Tohme, 2001. Genome mapping in cassava improvement: challenges, achievements and opportunities. Euphytica 120:159-165.

Fregene, M., F. Angel, R. Gomez, F. Rodriguez, P. Chavarriaga, W. Roca, J. Tohme, and M. Bonierbale, 1997. A molecular genetic map of cassava (*Manihot esculenta* Crantz). Theoretical and Applied Genetics 95:431-441.

Frey, K.J., 1992. Plant breeding perspectives for the 1990s. In: Stalker, H.T., and J.P. Murphy (Eds.), Plant breeding in the 1990s, Proceedings of the symposium on plant breeding in the 1990s held at North Carolina State University Raleigh, NC, pp. 1-13. CAB International Wallingford Oxon OX10 8DE, UK.

Fullerton, S.M., A.G. Clark, K.M. Weiss, D.A. Nickerson, S.L. Taylor, J.H. Stengard, V. Salomaa, E. Vartiainen, M. Perola, E. Boerwinkle, and C.F. Sing, 2000. Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. American Journal of Human Genetics 67:881-900.

Gibbons, A., 1990. New view of early Amazonia. Science 248:1488-1490.

Graner, E.A., 1942. Genetica de Manihot. I. Heretariedade da forma da folha e da coloracao de pelicula externa das raices en *Manihot utilissima* Pohl. Bragantia 2:13-22.

Grattapaglia, D., N.M.A. Nassar, and J.C. Dianese, 1987. Biossistematica de especies brasileiras do genero *Manihot* baseada em padroes de proteina da semente. Ciencia e Cultura 39:294-300.

Gulick, P., C. Hershey, and J. Esquinas-Alcazar, 1983. Genetic resources of cassava and wild relatives. Rome, International Board for Plant Genetic Resources Rome: IBPGR, 1983, pp. 56.

Gut, G.I., 2001. Automation in genotyping of single nucleotide polymorphism. Human Mutation 17:475-492.

Hahn, S.K, E.R. Terry, and K. Leuschner, 1980. Breeding cassava for resistance to cassava mosaic disease. Euphytica 29:673-683.

Hallauer, A.R., 1992. Use of genetic variation for breeding populations in cross pollinated species. In: Stalker, H.T., and J.P. Murphy, (Eds.), Plant Breeding in the 1990s. Proceedings of the symposium on plant breeding in the 1990s held at North Carolina State University Raleigh, NC, pp. 37-67. CABI International Wallington Oxon OX108DE, UK.

Hamblin, M.T., M.L. Warburton, and E.S. Buckler, 2007. Empirical comparison of simple sequence repeats and single nucleotide polymorphism in assessment of maize diversity and relatedness. Plos One 2: e1367. doi:10.1371/journal.pone.0001367.

Hartl, D.L., 2000. A primer of population genetics. Third Edition. Sinauer Associates, Inc, pp. 221.

Haysom, H.R., T.L.C. Chan, S. Liddle, and A. Hughes, 1994. Phylogenetic relationships of *Manihot* species revealed by restriction fragment length polymorphism. In: Roca, W., and A.M. Thro (Eds.), Working document No. 150, Proceedings of the Second international scientific meeting for the cassava biotechnology network, pp. 125-130. Bogor, Indonesia.

Hershey, C., 2010. Cassava breeding: theory and practice. A publication by Food and Agriculture Organisation (FAO), in press.

Hershey, C., and C. Ocampo, 1989. New marker genes found in cassava. Cassava Newsletter 13:1-5.

Hillocks, R.J., M.D. Raya, K. Mtunda, and H. Kiozia, 2001. Effects of brown streak virus disease on yield and quality of cassava in Tanzania. Journal of Phytopathology 149:389-394.

Howeler, R.H., and L.F. Cadavid, 1983. Accumulation and distribution of dry matter and nutrient during a 12-month growth cycle of cassava. Field Crops Research 7:123-139.

Hurtado, P., K.M. Olsen, C. Buitrago, C. Ospina, J. Marin, M. Duque, C. de Vicente, P. Wongtiem, P. Wenzel, A. Killian, M. Adeleke, and M. Fregene, 2008. Comparison of simple sequence repeat (SSR) and diversity array technology (DArT) markers for assessing genetic diversity in cassava (*Manihot esculenta* Crantz). Plant Genetic Resources: Characterisation and Utilisation 6:208-214.

IITA, 1990. Cassava in tropical Africa. A reference manual. International Institute of Tropical Agriculture, Ibadan, Nigeria, pp. 176.

Indira, P., T. Kurian, and S.B. Maini, 1977. Flowering behaviour in cassava (*Manihot esculenta* Crantz) as influenced by growth regulators. Indian Journal of Root Crops 2:37-40.

Jaramillo, G., N. Morante, J.C. Perez, F. Calle, H. Ceballos, B. Arias, and A.C. Bellotti, 2005. Diallel analysis in cassava adapted to the mid altitude valleys environment. Crop Science 45:1058-1063.

Jennings, D., 1957. Further studies in breeding cassava for virus resistance. East African Agricultural Journal 22:213-219.

Jennings, D.L., 1959. *Manihot melanobasis* Muell. Arg. - a useful parent for cassava breeding. Euphytica 8:157-162.

Jennings, D.L., 1963. Variation in pollen and ovule fertility in varieties of cassava, and the effects of interspecific crossing on fertility. Euphytica 12:69-76.

Jennings, D.L., and C.A. Iglesias, 2002. Breeding for crop improvement. In : Hillocks, R.J., J.M. Thresh, and A.C. Bellotti (Eds.), Cassava : biology, production, and utilisation, pp. 149-166. CABI Publishing.

Jones, W.O., 1959. *Manioc* in Africa. Standard University Press, Stanford, California, pp. 315.

Jorge, V., M.A. Fregene, M.C. Duque, M.W. Bonierbale, J. Tohme, and V. Verdier, 2000. Genetic mapping of resistance to bacterial blight disease in cassava (*Manihot esculenta* Crantz). Theoretical and Applied Genetics 101:865-872.

Kawano, K., 1987. Inherent and environmental factors related to cassava varietal selection. In: Hershey, C. (Ed.), Cassava breeding: a multidisciplinary review, pp. 207-226. CIAT, Cali, Colombia.

Kawano, K., 1990. Harvest index and evolution of major food crop cultivars in the tropics. Euphytica 46:195-202.

Kawano, K., 2003. Thirty years of cassava breeding for productivity-biological and social factors for success. Crop Science 43:1325-1335.

Kawano, K., A. Amaya, P. Daza, and M. Rios, 1978. Factors affecting efficiency of hybridization and selection in cassava. Crop Science 18:373-376.

Kawano, K., K. Narintaraporn, P. Narintaraporn, S. Sarakarn, A. Limsila, J. Limsila, D. Suparhan, V. Sarawat, and W. Watananonta, 1998. Yield improvement in a multi-stage breeding program for cassava. Crop Science 38:325-332.

Kawano, K., W.M.F. Goncalves, and U. Cenpukdee, 1987. Genetic and environmental effects on dry matter content of cassava root. Crop Science 27:69-74.

Kawuki, R.S., M. Ferguson, M. Labuschagne, L. Herselman, and D.J. Kim. 2009. Identification, characterisation and application of single nucleotide polymorphisms (SNPs) in cassava (*Manihot esculenta* Crantz). Molecular Breeding 23:669-684.

Kizito, E.B., A.C. Ronnberg-Wastliung, T. Egwang, T. Gullberg, M. Fregene, and A. Westerbergh, 2007. Quantitative trait loci controlling cyanogenic glucoside and dry matter content in cassava (*Manihot esculenta* Crantz). Hereditas 144:129-136.

Klug, W.S., M.R. Cummings, and C. Spencer, 2005. Concepts of Genetics. Eigth Edition, pp. 677.

Kreitman, M., 1983. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature 304:412-417.

Langlands, B.W., 1966. Cassava in Uganda 1860-1920. Uganda Journal 30:211-218.

Lauter, N., and J. Doebley, 2002. Genetic variation for phenotypically invariant traits detected in teosinte: implications for the evolution of novel forms. Genetics 160:333-342.

Legg, J.P., and C.M. Fauquet, 2004. Cassava mosaic germiniviruses in Africa. Plant Molecular Biology 56:585-599.

Lenis, J.I., F. Calle, G. Jaramillo, J.C. Perez, H. Ceballos, and J.H. Cock, 2006. Leaf retention and cassava productivity. Field Crops Research 95:126-134.

Lin, J., J.J.M. Kuo, J.A. Saunder, H.S. Beard, M.H. MacDonald, W. Kenworth, G. Ude, and B.F. Mathews, 1996. Identification of molecular markers in soybean comparing RFLP, RAPD and AFLP DNA mapping techniques. Plant Molecular Biology Reporter 14:156-169.

Litt, M, and J.A. Lutty, 1989. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. American Journal of Human Genetics 44:397-401.

Livini, C., M.P. Ajmine, A.E. Melchinger, M.M. Messmer, and M. Motto, 1992. Genetic diversity of maize inbred lines within and between heterotic groups revealed by RFLP. Theoretical and Applied Genetics 84:17-25.

Lopez, C., B. Piėgu, R. Cooke, M. Delseny, J. Tohme, and V. Verdier, 2005. Using cDNA and genomic sequences as tools to develop SNP strategies in cassava (*Manihot esculenta* Crantz). Theoretical and Applied Genetics 110:425-431.

Lutaladio, N.B., and H.C. Ezumah, 1981. Cassava leaf harvesting in Zaire. In: Tenry, E.R., K.A. Oduri, and F. Cavaness (Eds.), Tropical root crops: research strategies for the 1980's. pp. 134-136. IDRC. Ohawa, Canada.

Magoon, M.L., R. Krishnan, and K. Vijaybai, 1969. Morphology of pachytene chromosomes and meoisis in *Manihot esculanta* Crantz. Cytologia 34:612-618.

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

Maunder, A.B., 1992. Identification of useful germplasm for practical plant breeding programmes. In: Stalker, H.T., and J.P. Murphy (Eds.), Plant Breeding in the 1990s. Proceedings of the symposium on plant breeding in the 1990s held at North Carolina State University Raleigh, NC, pp. 147-169. CABI International Wallington Oxon OX108DE, UK.

Mba, R.E.C., P. Stephenson, K. Edwards, S. Melzer, J. Nkumbira, U. Gullberg, K. Apel, M. Gale, J. Tohme, and M. Fregene, 2001. Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. Theoretical and Applied Genetics 102:21-31.

Meireles da Silva, R., G. Bandel, and P.S. Martins, 2003. Mating system in an experimental garden composed of cassava (*Manihot esculenta* Crantz) ethnovarieties. Euphytica 134:127-135.

Morgante, M., and A.M. Olivieri, 1993. PCR-amplified microsatellites as markers in plant genetics. Plant Journal 3:175-182.

Morillo, E., P. Fuennayor, C. de Carvalho, and G. Second, 2001. AFLP and SSR polymorphism: Evidence of significant levels of introgression from *Manihot glaziovii* and *M. carthagenensis* into traditional varieties of cassava in their area of origin. In: Fauquet, C.M., and N.J. Taylor (Eds.), Cassava: an ancient crop for modern times. Proceedings of the fifth international meeting of the CBN. Held in November 4-9, 2001 at St Louis Missouri, USA. CD3, CBN-V Video Archive S6-18.

Moyib, O.K., O.A. Odunola, and A.G.O. Dixon, 2007. SSR markers reveal genetic variation between improved cassava cultivars and landraces within a collection of Nigerian cassava germplasm. African Journal of Biotechnology 6:2666-2674.

Mullis, K.B., 1990. The unusual origin of the polymerase chain reaction. Scientific American April:36-43.

Nasser, N.M.A., 2005. Cassava: some ecological and physiological aspects related to plant breeding. An article published online with Gene Conserve. URL http:\www.geneconserve.pro.br. October 2009.

Neale, B.B., and O. Savolainen, 2004. Association genetics of complex traits in conifers. Trends in Plant Science 9:325-330.

Nichols, R.F.W., 1947. Breeding cassava for virus resistance. East African Agricultural Journal 15:154-160.

Nichols, R.F.W., 1950. The brown streak disease of cassava: distribution climatic effects and diagnostic symptoms. East African Agricultural Journal 15:154-160.

Nick, C., M. Carvallio, L.H.B. de Assis, and S.P. de Carvalho, 2008. Genetic dissimilarity in casssava clones determined by multivariate techniques. Crop Breeding and Applied Biotechnology 8:104-110.

Nweke, F., D. Spencer, and J. Lynam, 2002. The cassava transformation: Africa's best kept secret. Michigan State University Press, East Lansing, USA, pp. 273.

Odongo, B., and G.W. Otim-Nape, 1984. Control of cassava green mite in Uganda. In: Terry, E.R., E.V. Doku, O.B. Arene, and N.M. Mahungu (Eds.), Tropical Root Crops: Production and uses in Africa, pp. 101-102. Proceedings of the Second Triennial Symposium of the International Society for Tropical Root Crops-Africa Branch. Douala, Cameroon, 14-19 August 1983. IDRC, Ottawa.

Ojulong, H.F., 2006. Quantitative and molecular analyses of agronomic traits in cassava (*Manihot esculenta* Crantz). PhD Thesis, University of the Free State, Bloemfontein, South Africa, pp. 190.

Ojulong, H., M.T. Labuschagne, L. Herselman, and M. Fregene, 2008. Introgression of genes for dry matter content from wild cassava species. Euphytica 164:163-172.

Okogbenin, E., and M. Fregene, 2003. Genetic mapping of QTLs affecting productivity and plant architecture in a full-sib cross from non-inbred parents in cassava (*Manihot esculenta* Crantz). Theoretical and Applied Genetics 107:1452-1462.

Okogbenin, E., I.J. Ekanayake, and M.C.M. Porto, 2003. Genotypic variability in adaptation responses of selected clones of cassava to drought stress in the Sudan savanna zone of Nigeria. Journal of Agronomy and Crop Science 189:376-389.

Okogbenin E., J. Marin, and M. Fregene, 2006. An SSR-based molecular genetic map of cassava. Euphytica 147:433-440.

Okogbenin, E., J. Marin, and M. Fregene, 2008. QTL analysis for early yield in a pseudo-F2 population of cassava. African Journal of Biotechnology 7:131-138.

Olsen, K.M., 2004. SNPs, SSRs and inferences on cassava origin. Plant Molecular Biology 56:517-526.

Olsen, K.M., and B.A. Schaal, 1999. Evidence of the origin of cassava: phylogeography of *Manihot esculenta*. Proceedings of the National Academy of Sciences (USA) 96:5586-5591.

Olsen, K.M., and B.A. Schaal, 2001. Microsatellite variation in cassava (*Manihot esculenta*, Euphorbiaceae) and its wild relatives: further evidence for a southern Amazonian origin of domestication. American Journal of Botany 88:131-142.

Otsubo, A.A., F.M. Mercante, R.F. da Silva, and C.D. Borges, 2008. Soil tillage systems, cover crops and productivity in cassava. Pesquisa Agropecuaria Brasileira 43:327-332.

Pellet, D., and M.A. El-Shakawy, 1993. Cassava varietal response to phoshorus fertilization 1. Yield biomass and gas-exchange. Field Crops Research 35:1-11.

Pinho, J.L.N., J.A.F. Tavora, O.I.F. Melo, and G.M. Queiroz, 1995. Yield components and partitioning characteristics of cassava in the coastal area of Ceara. Revista Brasileira de Fisiologia Vegetal 7:89-96.

Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski, 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding 2:225-238.

Rafalski, J.A., 2002. Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology 5:94-100.

Ramanujam, T., J.S. Jos, and K.R. Lakshmi, 1989. Biomass production as an index for rapid screening of cassava genotypes for productivity. Indian Journal of Plant Physiology 32:31-34.

Ramirez, H., A. Hussein, W. Roca, and W. Bushiuk, 1987. Isozyme electropherograms of sixteen enzymes in five tissues of cassava (*Manihot esculenta* Crantz) varieties. Euphytica 36:39-48.

Remington, D.L., J.M. Thornsberry, Y. Matsuoka, L.M. Wilson, S.R. Whitt, J. Doebley, S. Kresovich, M.M. Goodman, and E.S. Buckler, 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proceedings of the National Academy of Science (USA) 98:11479-11484.

Rogers, D.J., and S.G. Appan, 1973. *Manihot* and *manihotoides* (Euphorbiaceae): a computer assisted study. Organisation for Flora Neotropica. Monograph 13. Hafner, New York, pp. 278.

Rojas, M.C., J.C. Pérez, H. Ceballos, D. Baena, N. Morante, and F. Calle, 2009. Analysis of inbreeding depression in eight S1 cassava families. Crop Science 49:543-548.

Salmaso, M., G. Faes, C. Segala, M. Stefanini, I. Salakhutdinov, E. Zyprian, R. Toepfer, M.S. Grando, and R. Velasco, 2004. Genome diversity and gene haplotypes in the grapevine (*Vitis vinifera* L) as revealed by single nucleotide polymorphisms. Molecular Breeding 14:385-395.

Sanchez, G., S. Restrepo, M.C. Duque, M. Fregene, M. Bonierbale, and V. Verdier, 1999. AFLP assessment of genetic variability in cassava accessions (*Manihot esculenta*) resistant and susceptible to the cassava bacterial blight (CBB). Genome 42:163-171.

Sánchez, T., E. Salcedo, H. Ceballos, D. Dufour, G. Mafla, N. Morante, F. Calle, J.C. Pèrez, D. Debouck, G. Jaramillo, and I.X. Moreno, 2009. Screening of starch quality traits in cassava (*Manihot esculenta* Crantz). Starch 61:12-19.

Schneider, K., B. Weisshaar, D. Borchardt, and F. Salamini, 2001. SNP frequency and allelic haplotype structure of *Beta vulgaris* expressed genes. Molecular Breeding 8:63-74.

Second, G.A.C., L. Allem, C. Empérarie, C. Ingram, R.A. Columbo, L.J. Mendes, and C.B. Carvalho, 1997. AFLP based Manihot and cassava numerical taxonomy and genetic structure analysis in progress: Implications for dynamic conservation and genetic mapping in Africa. African Journal of Root and Tuber Crops 2:140-146.

Setter, T.L., and M. Fregene, 2007. Recent advances in molecular breeding of cassava for improved drought stress tolerance. Chapter 28, Advances in molecular breeding toward drought and salt tolerant crops, pp. 701-711.

Shastry, B.S., 2003. SNPs and haplotypes: Genetic markers for disease and drug response. International Journal of Molecular Medicine 11:379-382.

Shattuck-Eidens, D.M., R.N. Bell, S.L. Neuhausen, and T. Helentjaris, 1990. DNA sequence variation within maize and melon: observations from polymerase chain reaction amplification and direct sequencing. Genetics 126:207-217.

Sinha, S.K., and M.S. Swaminathan, 1984. New parameters and selection criteria in plant breeding. In: Vose, P.B., and S.G. Blixt (Eds.), Crop breeding a contemporary basis, pp. 1-31. Pergamon Press Ltd, Germany.

Siqueira, M.V.B.M., J.R. Queiroz-Silva, E.A. Bressan, A. Borges, J.C.K. Pereira, J.G. Pinto, and A.E. Veasey, 2009. Genetic characterisation of cassava (*Manihot esculenta*) landraces in Brazil assessed with simple sequence repeats. Genetics and Molecular Biology 32:104-110.

Smith, J.S.C., and D.N. Duvick, 1989. Germplasm collections and the private plant breeder. In: Brown, A.H.D., D.R. Marshall, O.H. Frankel, and J.T. Williams (Eds.), The use of plant genetic resources, pp. 17-31. Cambridge University Press.

Ssemakula, G., and A. Dixon, 2007. Genotype x environment interaction, stability and agronomic performance of carotenoid-rich cassava clones. Scientific Research and Essays 2:390-399.

Stojsin, D., L.W. Kannenberg, J. Rajnpreht, P.K. Pauls, and R. Stojsin, 1996. Genetic relationships among commercial corn hybrids and parents based on RAPD analysis of pericarp and embryo DNA. Genetics 28:137-150.

Storey, H.H., and R.F.W. Nichols, 1938. Studies of the mosaic disease of cassava. Annals of Applied Biology 25:790-806.

Sun, G.L., J. Williams, J. Liu, K.J. Kasha, and K.P. Pauls, 2001. Microsatellite and RAPD polymorphism in Ontario corn hybrids are related to the commercial sources and maturity ratings. Molecular Breeding 7:13-24.

Tan, S.L., 1992. Selection for early cassava clones. MARDI Research Journal 20:1-14.

Tan, S.L., and C. Mak, 1995. Genotype x environment influence on cassava performance. Field Crops Research 42:111-123.

Tanksley, S.D., N.D. Young, A.H. Paterson, and M.W. Bonierbale, 1989. RFLP mapping in plant breeding, new tools for old science. Bio/Technology 7:257-262.

Tonukari, N.J., G. Thottappilly, N.Q. Ng, and H.D. Mignouna, 1997. Genetic polymorphism of cassava within the Republic of Benin detected with RAPD markers. African Crop Science Journal 5:219-228.

Umannah, E.E., and R.W. Hartman, 1973. Chromosome numbers and karyotypes of some *Manihot* species. Journal of American Society of Horticultural Science 98:272-274.

Van, K., E.Y. Hwang, Y.M. Kim, H.J. Park, S.H. Lee, and P.B. Cregan, 2005. Discovery of SNPs in soybean genotypes frequently used as the parents of mapping populations in the United States and Korea. Journal of Heredity 96:529-535.

Viard, F., P. Franck, M.P. Dubois, A. Estoup, and P. Jarne, 1998. Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. Journal of Molecular Evolution 47:42-51.

Vidigal, P.S., M.G. Pequeno, M.V. Kvitschal, F. Rimoldi, M.C. Goncalves-Vidigal, and G.C. Zuni, 2007. Yield stability of sweet cassava cultivars from Parana State. Semina-Ciencias Agrarias 28:551-562.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van de Lee, M. Hornes, A. Frijters, L. Pot, J. Peleman, M. Kuiper, and M. Zabeau, 1995. AFLP: A new technique for DNA fingerprinting. Nucleic Acids Research 23:4407-4414.

Wang, D.G., J.B. Fan, C.J. Siao, A. Berno, P. Young, R. Sapolsky, G. Ghandour, N. Perkins, E. Winchester, J. Spencer, L. Kruglyak, L. Stein, L. Hsie, T. Topaloglou, E. Hubbell, E. Robinson, M. Mittmann, M.S. Morris, N. Shen, D. Kilburn, J. Rioux, C. Nusbaum, S. Rozen, T.J. Hudson, R. Lipshutz, M. Chee, and E.S. Lander, 1998. Large scale identification, mapping and genotyping of single nucleotide polymorphisms in the human genome. Science 280:1077-1082.

Wenzl, P., J. Carling, D. Kudrna, D. Jaccound, E. Huttner, A. Kleinhof, and A. Kilian, 2004. Diversity array technology (DArT) for whole-genome profiling of barley. Proceedings of the National Academy of Sciences (USA) 101:9915-9920.

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

Wricke, G., and W.E. Weber, 1986. Quantitative genetics and selection in plant breeding. Walter de Gruyter, Berlin, pp. 406.

Zacarias, A.M., 2008. Breeding potential of cassava (*Manihot esculenta* Crantz) in Mozambique. PhD Thesis University of the Free State, South Africa, pp. 164.

Zacarias, A.M., A.M. Botha, M.T. Labuschagne, and I.R.M, Benesi, 2004. Characterisation and genetic distance analysis of cassava (*Manihot esculenta* Crantz) germplasm from Mozambique using RAPD fingerprinting. Euphytica 138:49-53.

CHAPTER 3

VARIATION IN QUALITATIVE AND QUANTITATIVE TRAITS OF CASSAVA GERMPLASM FROM SELECTED NATIONAL BREEDING PROGRAMMES IN AFRICA

3.1 Introduction

The phenotype of a plant is of great agricultural and economic importance, as it represents the coordinated set of traits that are manifested by an individual over a range of environments (Sultan, 2003). Phenotypic traits often show large variability, with ecological significance of certain phenotypes being apparent and easily correlated with environmental conditions (Sinha and Swaminathan, 1984; Hartl and Clarke, 1989). This phenotypic variation plays a key role in defining and attainment of breeding objectives (Baenziger and Peterson, 1992; Hallauer, 1992) and in plant species differentiation (Linhart and Grant, 1996). Since the advent of agriculture, humans have applied directional selection on a series of plant traits resulting in varieties adapted to specific agricultural environments, with disparate phenotypes exhibiting symptomatic patterns of quantitative variation in response to environmental stimuli (Alonso-Blanco et al., 2005).

In practice, this variation is made possible because of the existence of many extreme minor variants in the base population and/or frequently arising mutations of either small or large effects that can potentially affect different phenotypes (Barton and Keightley, 2002; Klug et al., 2005). Whether created by humans and/or by natural evolutionary forces, this variation provides for heterogeneous populations, which have great capacity to stabilise productivity over a range of changing environments (Chang et al., 1979). Chang (1992) observed that with the unrelenting human population growth and environmental extremes, phenotypic variation in germplasm needs to be thoroughly exploited to meet the inevitable expansion in future food, feed and fibre needs.

Cassava is a monoecius, naturally out-crossing species believed to have been domesticated several years ago in the Amazon region (Olsen and Schaal, 1999; 2001). The starchy roots are the principal economic product of the cassava plant. Cassava shares a significant portion of its genome with its wild progenitors (Olsen and Schaal, 1999; Olsen, 2004), with genetic variation exhibited in the form of root, leaf and stem characteristics that display either qualitative or quantitative variation. It does suffice to note that phenotypic variation is expected to reduce with domestication of the crop, particularly under monoculture. From the Amazon, cassava was introduced to the east, central and southern African regions through two routes; through the west African coastline in the 1700s and through the east African coastline in the 1750s (Jones, 1959).

This position therefore invites a view that the present day cassava phenotypes in the east, central and southern African regions are a product of evolutionary forces. The evolutionary process is dependant on rare mutations that create novel morphologies and stabilising selection, which acts on the phenotypes (Klug et al., 2005). Moreover, since the introduction of cassava, farmers have for several years been undertaking individual plant selections (i.e. for shorter stature, higher root DMC, early maturity and preferred culinary qualities), which could further be restructuring the genetic make-up of cassava in the region.

In parallel, cassava breeding teams in the region, notably by IITA, have also been actively developing and disseminating cassava genotypes (IITA, 1990). With this in mind, the general picture is that cassava in the region comprises of a spectrum of phenotypes including cultivars released through formal breeding and local farmer selections, whose genetic structure and functional utility needs to be examined. The functional utility will however require systematic characterisation and/or evaluation of the agronomically useful traits of cassava.

Phenotypic characterisation and evaluation is critically important, because upon defining breeding goals and selection of evaluation environments, true genetic gain will only be achieved if substantial genetic diversity of the parental material exists to provide for additive improvement (Maunder, 1992). Besides, without systematic evaluation of existing germplasm, its potential utilisation cannot be realised to the full. For example, in the selection of parents for crossing without evaluating progeny, it was observed that the multivariate method, which requires comprehensive phenotypic data on each genotype, provided the best results for identifying wheat crosses with transgressive segregants (Bhatt, 1973). Furthermore, in the Americas, country based phenotypic evaluation of all maize collections in ten countries provided informative data and establishment of freely available databases (Smith and Duvick, 1989). Regrettably, most national cassava breeding programmes in the east, central and southern Africa region have limited or no phenotypic information on their germplasm collections.

An equally important justification for phenotyping is to obtain insight into cassava plasticity. Phenotypic plasticity involving the morphological or physiological response of organisms to the environment (Schlichting, 1989), will be particularly important in the current face of rapidly changing agricultural ecosystems resulting from climate change. These plastic responses can manifest themselves by making shifts in resource allocation and morphological patterns. Indeed, extensive and purposeful phenotypic evaluations have been undertaken for some major starchy crops like maize (Smith and Duvick, 1989) and wheat (Fischbeck, 1989), with phenomenal findings and establishment of phenotypic databases.

In cassava, phenotypic variation has been measured experimentally for some qualitative traits (Benesi, 2005; Balyejusa Kizito, 2006) but with limited efforts devoted towards ecologically-adaptive and/or consumer-related quantitative traits, particularly in local cassava varieties. Grando et al. (2000) illustrated that landraces are not only genetic resources to be conserved for future needs, but are breeding material to be used today, particularly in breeding for stressful environments.

It is against this backdrop that this study examined the phenotypic diversity of cassava germplasm (both local varieties and elite genotypes) available within the national cassava breeding programmes of Tanzania, Uganda, Kenya, Rwanda, DRC and Madagascar. Specifically, phenotypic variation was quantified for four quantitative traits including HI, root DMC, leaf retention (LR) and root cortex thickness as well as for 29 qualitative traits.

3.2 Materials and methods

3.2.1 Cassava germplasm

The germplasm included in this study represented a spectrum of genotypes either released through formal breeding (hereafter referred to as elite genotypes) or farmer selections (hereafter referred to as local genotypes) that were available within national breeding programmes of six countries. Qualitative traits were scored on cassava germplasm from Tanzania (130 genotypes), Uganda (317 genotypes), Kenya (97 genotypes), Rwanda (177 genotypes), DRC (182 genotypes) and Madagascar (188 genotypes). Genotypes included for the qualitative analysis had < 10% missing data.

Quantitative traits were measured on germplasm from Tanzania (110-148 genotypes), Uganda (320-326 genotypes), Kenya (76-99 genotypes), Rwanda (117-177 genotypes), DRC (131-220 genotypes) and Madagascar (143-186 genotypes). For quantitative traits, the number of genotypes for which data was collected varied among traits and hence the differences in the number of genotypes evaluated within a country. For example, data on root DMC were only collected for genotypes that were able to provide 3-5 kg of root weight as required by the specific gravity method, indicating that genotypes without DMC data could have LR and/or HI data. The number of genotypes included in this study was selected such that they would be representative of the country's cassava germplasm collection. This selection process was jointly done with cassava breeders of the respective countries. It was therefore inevitable to have a different number and set of genotypes evaluated and/or characterised per country.

3.2.2 Establishment of trial sites

Trials were established at six sites in the six countries: Mwanza (northern Tanzania), Serere (eastern Uganda), Kakamega (western Kenya), Rubona (southern Rwanda), Mvuazi (south-western DRC) and Ambohitsilaozana (eastern Madagascar).With the exception of the Rwanda trial that was planted in 2008, all other trials were planted in 2007. The selected sites are within major cassava growing regions in the respective countries. At each site, each genotype was established in a single-row plot comprising of 10 plants. Plant spacing was 1 m within rows and 1.5 m between rows, to limit inter-plot interferences. Cuttings of mature woody stakes were planted in a horizontal orientation. Because trials were unreplicated, a standard genotype TME 14 was planted at each site after every 20 entries to enable the estimation of the standard error associated with quantitative trait evaluations. No fertilisers, irrigation or pesticides were applied to the crop. The trials were kept weed free by regular manual weeding. To remove border effects, all assessments were done on the six central plants. These trials were used for collecting data on both qualitative and quantitative traits.

3.2.3 Phenotypic characterisation of qualitative traits

At each site, characterisation was done using a standardised descriptor list. Apparently, no universally accepted descriptor list exists for cassava. The most comprehensive list was developed by EMBRAPA in Portuguese (Documentos – CNPMF No. 78, ISSN 0101-5171-JUNHO/1998). Although this publication is comprehensive, it needed revision to incorporate some of the important farmer qualitative traits. On the other hand, the IPGRI descriptor list encompasses many of the EMBRAPA descriptors but lacks some visual images to aid field work. Thus, before undertaking characterisation, these two descriptor lists were compared and a consensus descriptor list compiled. The consensus list was largely based on the EMBRAPA list. This consensus descriptor list adopted for this study has been submitted for publication (Morag Ferguson, personal communication).

Characterisation was done on four occasions, at three, six, nine and 12 MAP. A total of 29 qualitative traits were characterised in 1091 genotypes across the six countries. The first three assessments were primarily devoted to above ground traits, while the assessments at 12 MAP were for both root and above-ground qualitative traits (Table 3.1).

Table 3.1 A list of qualitative traits used in the characterisation of the cassava germplasm from the six African NARS

Trait	Abbreviation	Assessment date	Assessment scale[1]
Colour of apical leaf	AL	3 MAP	3, 5, 7 or 9
Pubescence on apical leaf	P	3 MAP	0 or 1
Shape of central leaf	CLS	6 MAP	1-10
Petiole colour	PC	6 MAP	1, 2, 3, 5, 7 or 9
Leaf colour	LC	6 MAP	3, 5, 7 or 9
Number of leaf lobes	LL	6 MAP	3, 5, 7, 9 or 11
Lobe margins	LM	6 MAP	3 or 7
Colour of leaf vein	CLV	6 MAP	3, 5, 7 or 9
Orientation of petiole	OP	6 MAP	1 , 3, 5 or 7
Prominence of foliar scars	FS	9 MAP	3 or 5
Colour of stem cortex	CSC	9 MAP	1, 2 or 3
Colour of stem epidermis	CSE	9 MAP	1, 2, 3 or 4
Colour of stem exterior	CS	9 MAP	3, 4, 5, 6, 7, 8 or 9
Growth habit of stem	SG	9 MAP	1 or 2
Colour of end branches	CEB	9 MAP	3, 5 or 7
Length of stipule	LS	9 MAP	3 or 5
Stipule margin	SM	9 MAP	1 or 2
Levels of branching	LB	12 MAP	0-7
Branching habit	BH	12 MAP	1, 2, 3 or 4
Shape of plant	PS	12 MAP	1, 2, 3 or 4
Extent of root peduncle	RP	12 MAP	0, 3 or 5
Root constrictions	RC	12 MAP	1, 2 or 3
Root shape	RS	12 MAP	1, 2, 3 or 4
External colour of root	ERC	12 MAP	1, 2, 3 or 4
Colour of root pulp	CRP	12 MAP	1, 2, 3, 4 or 5
Colour of root cortex	CRC	12 MAP	1, 2, 3 or 4
Cortex ease of peeling	EP	12 MAP	1 or 2
Texture of root epidermis	TRE	12 MAP	3, 5 or 7
Hydrogen cyanide levels	HCN	12 MAP	1-9

[1] Each phenotypic trait had distinct phenotypes which were depicted by the values ranging from 0 to 10. Images associated with these scale values can be found in the EMBRAPA and/or the IPGRI descriptor list. Hydrogen cyanide content (HCN) was determined using the picric acid test that uses the pictorial qualitative scale of 1-9. MAP = months after planting.

3.2.4 Phenotypic evaluation of quantitative traits

Data on the three quantitative traits, including LR, DMC and HI were collected from single-row unreplicated trials. These traits, due to their moderate to high heritability, could be measured using unreplicated trials. However, traits of low heritability i.e., fresh root yield that require replicated trials, were excluded from analysis. A common feature of most cassava breeding programmes is that the initial stages of evaluation and selection are usually unreplicated (Kawano, 2003; Ceballos et al., 2004). Unreplicated single-row trials have previously been used to evaluate quantitative traits in over 1500 cassava clones (Kawano, 2003; Chávez et al., 2005), results of which have directly contributed to the genetic improvement of cassava. Lack of adequate good quality planting material and the considerable logistical complications are the justification for this tradeoff of having more genotypes being evaluated in unreplicated trials as opposed to having fewer genotypes evaluated in replicated trials. Once selections have been made at single-row trials, selected clones can be evaluated in replicated trials with bigger plot sizes. Since this study aimed at exploring the extent of variation in agronomically important traits in order to define future breeding objectives, a large number of clones were established in unreplicated trials in the six countries.

3.2.4.1 Leaf retention evaluation

Cassava experiences simultaneous growth and development of the economic plant parts (roots) and the photosynthetic sites, the leaves. This phenomenon could suggest that greater leaf longevity will result in stable and/or higher yields, particularly under drought stress as observed in previous studies (Lenis et al., 2006; Hershey, 2010). To genetically improve this trait, genetic variation is needed. Hence, variation in LR was assessed in trials established in DRC, Kenya, Madagascar and Rwanda. LR was assessed at six MAP. This period normally coincides with drought stress in most cassava growing regions in eastern and southern Africa. Data could not be collected from Tanzania, while assessments were done later (at seven MAP) in Uganda. It is for these reasons that LR data for both Tanzania and Uganda were not included in the analysis.

LR was visually scored on a plot (row) basis using a scale of 1-5, where 1 = very poor LR; 2 = LR worse than average retention; 3 = average LR; 4 = LR greater than average; 5 = out standing LR (Lenis et al., 2006).

3.2.4.2 Root dry matter content and harvest index evaluation

At harvest, which coincided with 12 MAP, six plants per entry were uprooted and used for phenotypic assessments. Roots were separated from the harvestable biomass (leaves, stems and original planting stake) and HI, defined as the proportion of root weight to total biomass (on a fresh weight basis), was computed for each entry following the procedure outlined by Kawano (1990). Estimation of DMC in the root samples was based on the specific gravity method (Kawano et al., 1987), which is frequently used in cassava studies (Cach et al., 2005; Chávez et al., 2005; Ojulong et al., 2008b). Approximately 3-5 kg of roots was weighed using a hanging scale to provide weight in air (Wa). The same sample was weighed with the roots submerged in water using a Scout^{R} pro-balance (Ohaus Corporation, USA) to get the weight in water (Ww). DMC (%) was estimated using the formulae:

$$\% \text{ DMC} = \left(\frac{Wa}{Wa-Ww} \times 158.3 - 142 \right)$$

3.2.4.3 Root cortex thickness evaluation

The cassava peel is a composite of the periderm and cortex. This is usually removed as a unit from the fresh root (parenchyma) prior to cassava utilisation. The implications of different root peel thickness are not well understood. Measurements on peel thickness were made using root samples harvested in trials established in the DRC, Kenya, Madagascar, Tanzania and Uganda. Data for root peel thickness in Rwanda was not collected.

For each genotype, a random sample of three roots were measured at the proximal (near the penducle), mid and distal (end of the root) root positions using an absolute digimatic calliper (Mitutoyo Corporation, Japan) and the means recorded.

3.2.5 Data analysis

3.2.5.1 Qualitative traits

Data collected from the six countries across the 29 qualitative traits was used to generate a distance matrix that was used to infer relationships between: 1) cassava genotypes on a country basis and 2) the local and elite genotypes studied. Because the dataset contained categorical data with several unordered modalities including 0/1 data, the Rogers and Tanimoto (1960) distance was used to compute the dissimilarity matrix using the formula: $d_{ij} = 2u/(m+2u)$; where d_{ij} is the dissimilarity between individual i and j, u the number of unmatching variables and m the number of matching variables. Clustering was performed using the weighted neighbour-joining algorithm and relationships displayed as a phenogram. This analysis was done using the DARwin software version 5.0.153 (Perrier and Jackquemound-Collet, 2006).

The dimensionality of the 29 qualitative traits was examined by subjecting the data to multidimensional scaling (MDS) analysis. The MDS represents a set of units in a few dimensions on a map using the similarity/distance matrix between them, such that inter-individual proximities in the map nearly match the original similarities/distances and hence highlighting units that are similar (Mohammadi and Prasanna, 2003). The non-metric algorithm which considers rank order was used and this analysis was done using the number cruncher statistical system (NCSS) software (Hintze, 2001).

3.2.5.2 Quantitative traits

Quantitative traits analysed included LR, DMC, HI and root cortex thickness. Data collected across the six countries (DMC and HI), four countries (LR) and five countries (root cortex thickness), were combined separately for each trait and subjected to general analysis of variance with no blocking option in Genstat statistical software version 7.2.2.222. The countries and groups (local and elite genotypes) were considered as treatments. From this analysis, variation in DMC, HI, LR and root cortex thickness were statistically tested both at country and group levels.

Additionally, principle component analysis (PCA) was performed for DMC, HI, root cortex thickness and LR. The PCA analysis was run on the correlation matrix because the variables were of different scales (Mohammadi and Prasanna, 2003). From the pre-analysis of data, the Gleason-Staelin redundancy measure (Phi = 0.311) and the Bartlett's sphericity test (138.61; Prob = 0.00) were indicative of interrelationships among the variables and hence PCA was appropriate for the data as required by the NCSS statistical software (Hintze, 2001). Phenotypic correlations among variables were done to establish their respective relationships.

3.3 Results

3.3.1 Qualitative traits

Genetic relationships based on qualitative data of the cassava germplasm from the six different countries are presented in Figure 3.1. Eight major groupings were observed. Some of these groupings were, however, heterogeneous comprising of cassava genotypes from the different countries, but some consisted primarily of germplasm from a single or two countries. It was evident that: 1) Rwanda germplasm was mainly exclusive in one group, 2) germplasm from DRC formed two exclusive groups, 3) germplasm from Tanzania and Madagascar formed a few exclusive groups and 4) germplasm from Uganda was the most widespread across the phenogram (Figure 3.1).

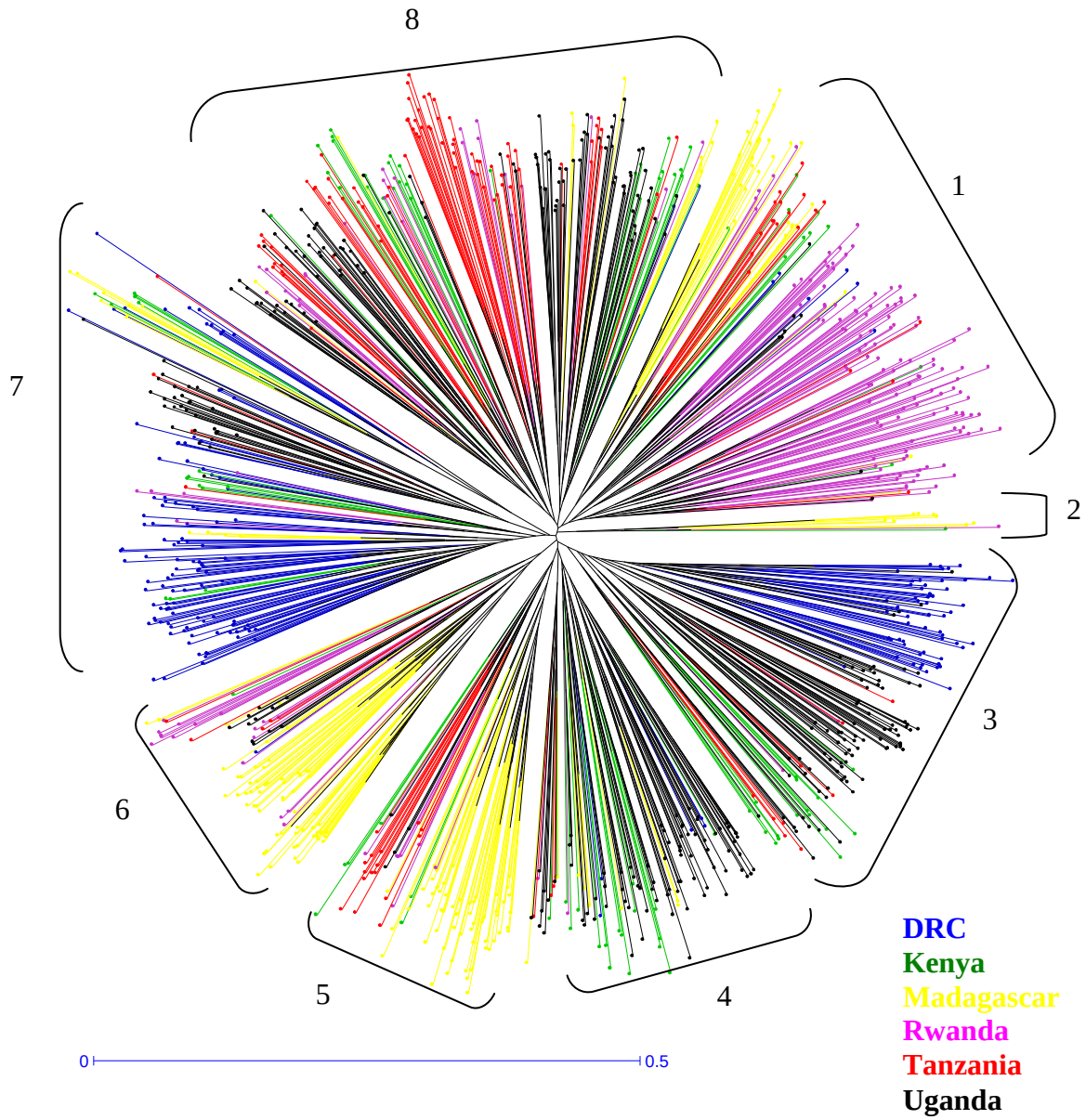


Figure 3.1 **Phenogram generated from 29 cassava qualitative morphological traits displaying genetic relationships between cassava germplasm available within the national breeding programmes of Uganda (317 genotypes), Kenya (97), Tanzania (130), DRC (182), Rwanda (177) and Madagascar (188).**

The genetic relationship between local and elite germplasm is presented in Figure 3.2. No apparent pattern was observed in the genetic relationships between local and elite genotypes based on 29 qualitative morphological traits. All the major clusters were heterogeneous comprising of both local and elite genotypes.

The dimensionality of the 29 qualitative traits is shown in Figure 3.3. The goodness-of-fit measured by the stress value was 0.177 and the total variation explained was only 26% when three dimensions were considered. Over 90% of the variation could be accounted for by 22 dimensions which can not be presented herein. This observation suggested limited relatedness amongst the analysed morphological traits.

For ease of presentation, only two plots were displayed, both of which indicated limited relatedness amongst the morphological traits (Figure 3.3). For example, for dimensions 2 and 1, notably two morphological traits, number of leaf lobes (LL) and colour of root pulp (CRP) were closely related, while in dimensions 3 and 1, texture of root epidermis (TRE) and colour of stem exteriors (CS) were the only closely related morphological traits (Figure 3.3). The changes in clustering as observed in Figures 3.3a and 3.3b are indicative of limited relatedness of the morphological traits.

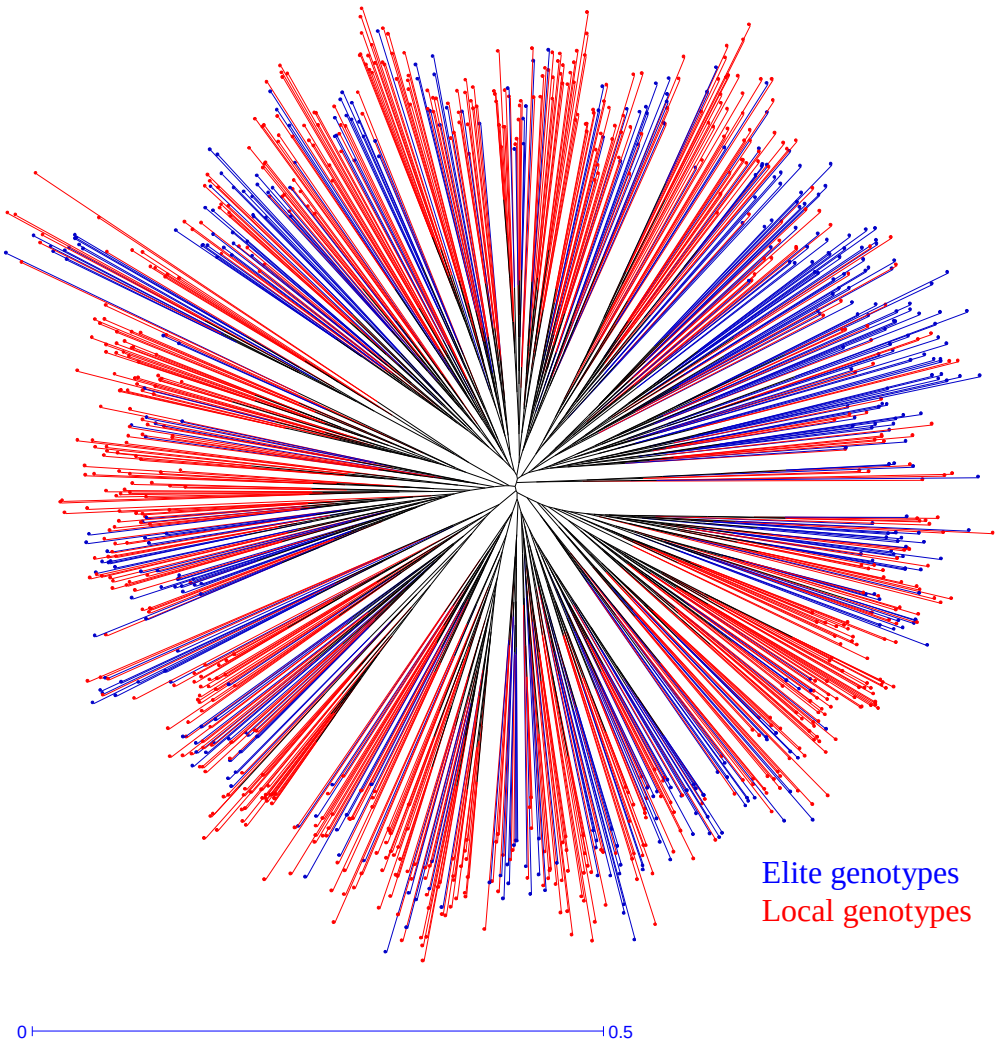
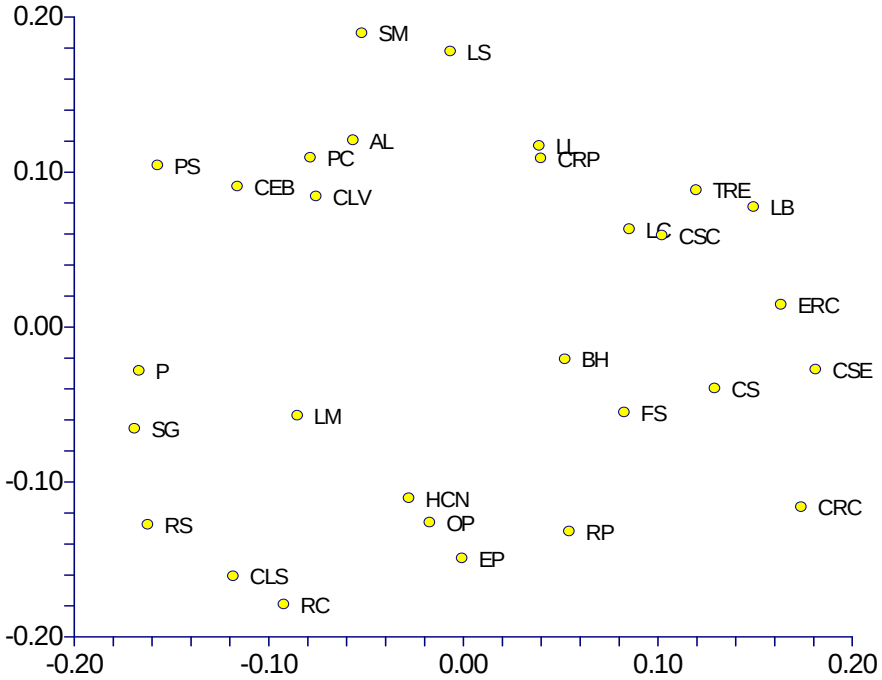


Figure 3.2 **Phenogram generated from 29 cassava qualitative morphological traits across six countries displaying genetic relationships between elite (386 genotypes) and local (705 genotypes) germplasm.**

a



b

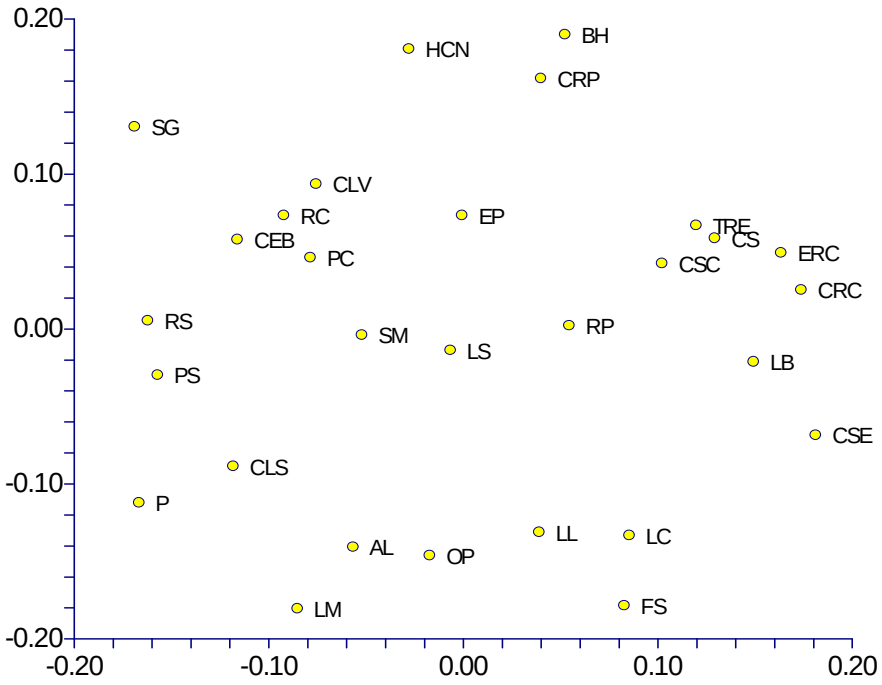


Figure 3.3 **Dimensions of the 29 qualitative traits analysed in 1091 cassava genotypes: a) dimensions 2 and 1 and b) dimensions 3 and 1. Abbreviations as defined in Table 3.1**

3.3.2 Quantitative traits

3.3.2.1 Leaf retention

Analysis of variance (ANOVA) for LR indicated significant differences among the countries' breeding germplasm, groups (local and elite genotypes) and their interactions (Table 3.2). The standard error (estimated from values recorded on the standard genotype TME 14) associated with LR evaluations in the different countries was 0.104. Among the elite genotypes, highest average LR was recorded in Madagascar (4.6) and lowest (2.4) in the DRC (Table 3.3). For local genotypes, the highest average leaf retention (4.5) was recorded in Madagascar and lowest (2.9) in Kenya (Table 3.3). Generally, local varieties had higher LR than elite genotypes. Results indicated that most (>44.5%) of the genotypes from Kenya and DRC had an average LR (score of 3) (Table 3.3). Most cassava genotypes from Rwanda had above average LR (score of 4), while most genotypes from Madagascar had outstanding LR (score of 5) (Table 3.3). LR was negatively correlated with HI ($r = -0.288$; $P < 0.001$) and DMC ($r = -0.274$; $P < 0.001$).

3.3.2.2 Dry matter content

Significant differences in DMC were observed between the countries' national breeding germplasm, groups and their interactions (Table 3.2). The standard error (estimated from values recorded on the standard genotype TME 14) associated with evaluations in the different countries was 0.858. DMC varied significantly between the breeding programmes for both elite and local cassava genotypes (Table 3.4). However, DMC content of elite genotypes from Madagascar, Rwanda and Tanzania was not significantly different. Among the elite genotypes, the highest average DMC (39.3%) was recorded in Uganda, while the lowest (31.2%) was recorded in Tanzania. On the other hand, highest average DMC in local genotypes was reported in Kenya (38.3%) and lowest (30.1%) from Tanzania (Table 3.4). In the entire data set (952 genotypes), the lowest DMC (16.3%) was recorded on an elite genotype "H 71" from Madagascar, while the highest DMC (49.6%) was recoded on an elite genotype "Nase 1" from Uganda (Table 3.4).

Some local genotypes like "Zanzibar" from Madagascar and "Nyantansingizi" from Tanzania registered the highest DMC of 42.4% and 35.5% in the respective countries. Overall, the elite genotypes had a relatively higher DMC content than the local genotypes (Table 3.4). A graphical display of the DMC across different countries is presented in a dot plot, which indicated that most of the DMC ranged between 25-40% (Figure 3.4). It is also apparent that: 1) genotypes from DRC and Kenya had DMC values that tended towards the upper range, 2) genotypes from Rwanda had DMC values that were evenly distributed without apparent clustering, 3) genotypes from Tanzania had DMC values that tended towards the lower range and 4) genotypes from Uganda had clustering of DMC values in the upper range (Figure 3.4)

Table 3.2 Mean squares for leaf retention, dry matter content and harvest index of cassava germplasm available within selected national cassava breeding programmes[1]

Source of variation	MS LR	MS DMC	MS HI	MS root cortex
Country (C)	84.20*	1405.9*	2.389*	19.795*
Group (G)	6.91*	239.6*	0.823*	6.178*
C x G	14.13*	141.5*	0.160*	1.802*
Residual	0.51	21.5	0.011	0.224

[1]Countries represent the national cassava breeding programmes; Groups represent the elite and local cassava genotypes; * $P \leq 0.05$. MS = mean square. LR = leaf retention in DRC, Kenya, Madagascar and Rwanda. DMC = dry matter content assessed in Kenya, Madagascar, Rwanda, Tanzania, Uganda and DRC. HI = harvest index assessed in Kenya, Madagascar, Tanzania, Uganda and DRC.

Table 3.3 A comparison of leaf retention in cassava germplasm available within selected national cassava breeding programmes[1]

Country	Elite genotypes	Local genotypes	Percentage of genotypes described by 1-5 scale[2]			
			2	3	4	5
DRC	2.4 (85)	3.4 (135)	25.0	44.5	30.4	0
Kenya	3.3 (63)	2.9 (36)	18.1	50.5	27.2	4.0
Madagascar	4.6 (44)	4.5 (142)	1.0	10.2	20.9	67.7
Rwanda	3.7 (122)	3.5 (55)	8.4	29.9	47.4	14.1
Mean	**3.4**	**3.8**				
CV (%)	20.3	19.2				
LSD	0.24	0.26				

[1]Figures in parentheses indicate number of genotypes evaluated; [2]Leaf retention scored on a scale of 1-5; 1 = very poor retention; 2 = less than average retention; 3 = average leaf retention; 4 = better than average retention; 5 = outstanding leaf retention (Lenis et al., 2006). No genotypes scored on the scale of 1. CV = coefficient of variation; LSD = Least significant difference at * $P \leq 0.05$.

Table 3.4 A comparison of dry matter content (%) in cassava germplasm available within selected national cassava breeding programmes[1]

Country	Elite genotypes	Local genotypes	Min DMC	Max DMC
DRC	37.5 (82)	33.6 (78)	22.4	48.9
Kenya	37.2 (50)	38.3 (26)	27.9	44.8
Madagascar	32.7 (33)	33.0 (110)	16.3	42.4
Rwanda	33.0 (74)	34.7 (43)	20.7	48.1
Tanzania	31.2 (3)	30.1 (133)	24.3	35.5
Uganda	39.3 (78)	37.2 (242)	16.4	49.6
Mean	**36.32**	**34.4**		
CV (%)	14	12.7		
LSD	5.9	1.77		

[1]Figures in parentheses indicate the number of genotypes evaluated; Min DMC and Max DMC = minimum and maximum dry matter content recorded respectively. CV = coefficient of variation; LSD = Least significant difference at 5%.

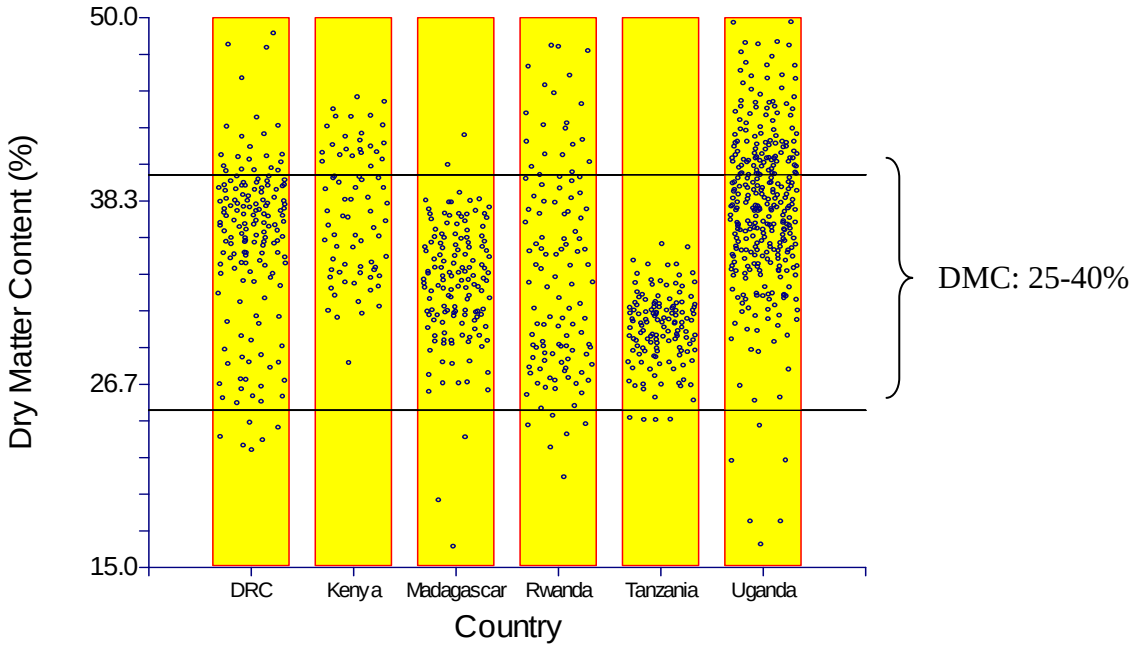


Figure 3.4 Dot plot display of dry matter content in cassava germplasm available within six national cassava breeding programmes.

3.3.2.3 Harvest index

HI varied significantly between the different countries' breeding germplasm, groups and their interactions (Table 3.2). The standard error associated with HI evaluations in the different countries was 0.0021. Among the elite germplasm, the highest average HI (0.66) was recorded in Uganda, while the lowest (0.32) was recorded in Madagascar (Table 3.5). For local genotypes, the highest average (0.59) was again recorded in Uganda and the lowest (0.26) in Kenya (Table 3.5). Across the entire data set (1071 genotypes), the lowest HI (0.04) was recorded on a local genotype "Sakay no 2" from Madagascar, while the highest (0.90) was also recorded on a local genotype "Mwaihwa" from Uganda. A graphical display of HI across different countries is presented in a dot plot, which indicated that most of the genotypes evaluated, had HI values ranging between 0.25-0.75 (Figure 3.5).

It is apparent from the dot plot that: 1) genotypes from Kenya and Madagascar tended to have HI values in the lower range, 2) genotypes from Uganda tended to have HI values in the upper range and 3) genotypes from DRC, Tanzania and Rwanda had HI values that were more evenly distributed (Figure 3.5).

Table 3.5 A comparison of harvest index of cassava germplasm available within selected national cassava breeding programmes[1]

Country	Elite genotypes	Local genotypes	Min HI	Max HI
DRC	0.57 (89)	0.42 (100)	0.06	0.83
Kenya	0.35 (56)	0.26 (35)	0.06	0.77
Madagascar	0.32 (34)	0.32 (109)	0.04	0.68
Tanzania	0.62 (4)	0.53 (144)	0.14	0.75
Uganda	0.66 (77)	0.59 (249)	0.20	0.90
Rwanda	0.43 (124)	0.46 (50)	0.24	0.71
Mean	**0.49**	**0.48**		
CV (%)	20.2	23.3		
LSD	0.09	0.04		

[1]Figures in parentheses indicate number of genotypes evaluated; Min and Max HI indicate minimum and maximum harvest index recorded. CV = coefficient of variation; LSD = Least significant difference at 5%.

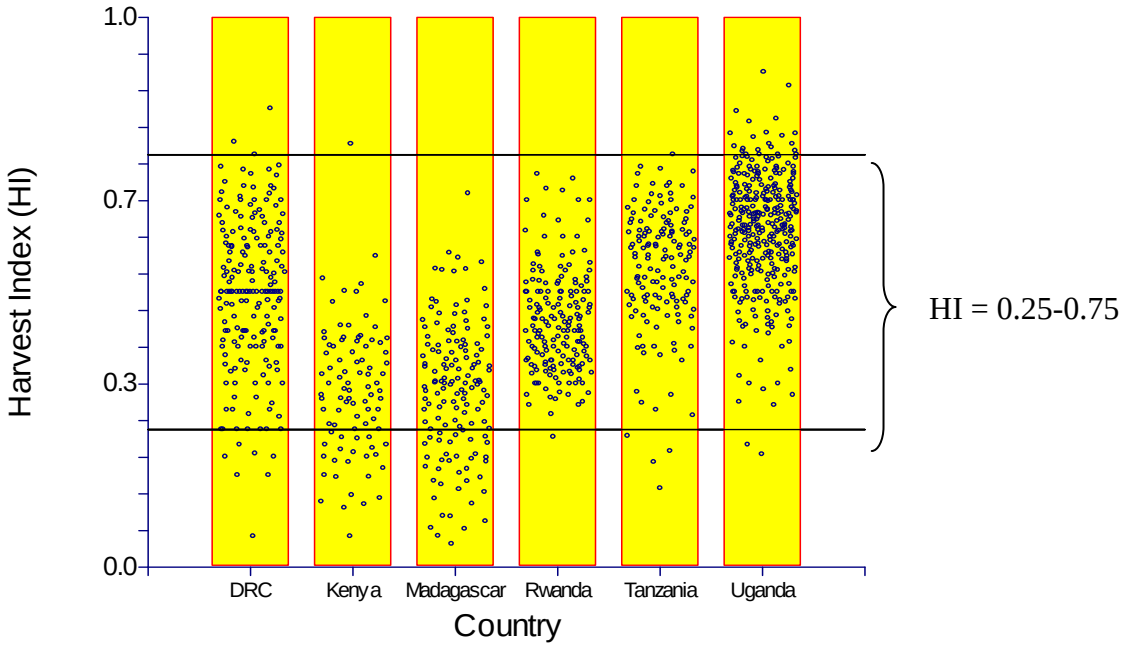


Figure 3.5 Dot plot display of harvest index of cassava genotypes across the six national cassava breeding programmes.

3.3.2.4 Root cortex thickness

Root cortex thickness varied significantly between the cassava germplasm available in the different countries, between local and elite genotypes and their interactions (Table 3.2). Elite genotypes had higher root cortex thickness than the local genotypes (Table 3.6). Among the elite genotypes, highest average root cortex thickness (2.24 mm) was recorded in the DRC, while the lowest (1.13 mm) was recorded in Tanzania (Table 3.6). For local genotypes, DRC (1.88 mm) and Tanzania (1.24 mm) had the highest and lowest root cortex thichness, respectively (Table 3.6). Across the entire dataset (825 genotypes), the lowest root cortex thickness (0.34 mm) was recorded on a local genotype "Mangi no 2" from Madagascar, while the highest root cortex thickness (4.89 mm) was recorded on a local genotype "Mayombe" from DRC. The regression of root cortex thickness on DMC established that the R^2 value, the proportion of variation in root cortex thickness accounted for by variation in DMC, was 0.0693 (Figure 3.6).

Table 3.6 A comparison of root cortex thickness (mm) of cassava germplasm available within selected national cassava breeding programmes[1]

Country	Elite genotypes	Local genotypes	Min cortex	Max cortex
DRC	2.24 (46)	1.88 (85)	0.55	4.89
Kenya	2.22 (64)	1.61 (34)	0.83	4.44
Madagascar	1.26 (37)	1.33 (126)	0.34	2.30
Tanzania	1.13 (5)	1.24 (105)	0.62	2.26
Uganda	1.98 (83)	1.84 (240)	0.95	3.87
Mean	**1.96**	**1.62**		
CV (%)	27.3	27.5		
LSD	0.48	0.16		

[1]Figures in parentheses indicate number of genotypes evaluated; Min and Max cortex indicate minimum and maximum root cortex thickness recorded. CV = coefficient of variation; LSD = Least significant difference at * $P \leq 0.05$.

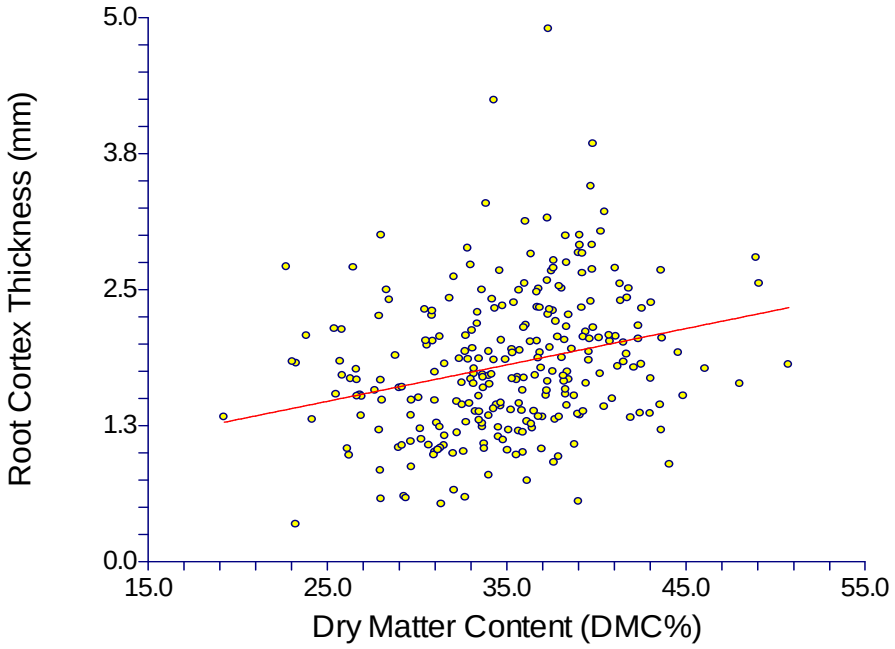


Figure 3.6 Regression plot of root cortex thickness (y) on root dry matter content (x); Y = 0.6375 + 0.0334X. The R^2 value = 0.0693.

Principal component analysis revealed that the first three principal components (PCs) explained 87.35% of the total variation (Table 3.7). Results indicated that PC1 with eigen value of 1.89 and accounting for 47.34% of the variation, had LR and peel thickness as the most important variables. Principal components two and three had respective eigen values of 0.94 and 0.66, indicating that one variable was contributing most of the variation. Hence, in PC2 and PC3, DMC and HI were respectively the most important variables. The correlation between DMC and HI was low ($r = 0.016$; $P < 0.001$).

Table 3.7 Principal component coefficients of four agronomic cassava traits evaluated in 270 cassava genotypes

Trait	PC1	PC2	PC3	PC4
Harvest index	-0.427	-0.644	-0.627	0.096
Dry matter content	-0.378	0.762	-0.524	0.007
Leaf retention	0.575	-0.028	-0.466	-0.670
Peel thickness	-0.585	-0.051	0.337	-0.735
Eigen value	1.89	0.94	0.66	0.50
Individual percentage	47.34	23.50	16.50	12.65
Cumulative percentage	47.34	70.84	87.35	100

3.4 Discussion

The objectives of this research study were two-fold. Firstly, to establish the genetic relationship between cassava germplasm available within the national breeding programmes based on morphological traits. Secondly, to examine the extent of genetic variation in some agronomically important traits in order to define future breeding objectives in the region. Genetic relationships were inferred from 29 morphological traits, which resulted in eight major clusters. Some clusters consisted of germplasm overwhelmingly from one or two countries. This indicated some discrimination of cassava according to country of origin based on morphological traits.

The phenogram illustrated that morphologically germplasm from Rwanda had a narrow genetic base as the majority of germplasm occurred in a single cluster. It may thus be necessary for breeders in Rwanda to increase morphological diversity in the breeding programme so that they can make significant gains in any of the morphological traits studied. Germplasm from DRC occurred in two main clusters. Again, a narrow genetic base is indicated within these two germplasm groups. It would however be interesting to examine the cause of this difference. It may be adviseable to increase both genetic diversity as well as generate crosses among germplasm groups to increase chances of facilitating genetic gain.

A cluster of germplasm from Tanzania also exists; it is possible that this germplasm was derived from the Amani breeding programme. Uganda and Madagasacar seem to have a range of germplasm represented from across the region, with a few small clusters. The factors that define these groups should be investigated.

In this study, the different states or classes of the 29 qualitative traits were represented in most countries. Observed differences were due to differences in frequencies as opposed to presence or absence differences. For example, considering a trait like colour of apical leaves which had four categories: 3 = light green; 5 = dark green; 7 = purplish green and 9 = purple. All these phenotypes were present in the germplasm studied in each of the countries, but at slightly different frequencies. This indicated limited discriminatory power of this set of morphological markers in this set of germplasm. It could also indicate a lack of diversity among germplasm from the different countries which could be a result of unrestricted inter-country movement of material or the result of a 'genetic bottleneck' caused by the introduction of restricted cassava diversity. The limited clustering of morphological traits as depicted by the MDS indicated that traits were not closely related and/or are unlinked, which is desirable. In an effort to increase discriminatory power, the germplasm was further examined using SSR markers as described in Chapter 4.

In a study conducted in Malawi, 12 morphological traits did not uniquely classify 93 cassava accessions, which the author attributed to a limited number of traits used in the analysis (Benesi, 2005). The author noted that the morphological traits typified many accessions, a finding which is consistent with the present study. A major distinction of the current study from previous studies (Benesi, 2005; Balyejusa Kizito, 2006) is the number of traits and genotypes analysed. The general trend however, is that morphological traits are limited in discriminatory power. Nonetheless, in the absence of molecular markers, the studied 29 qualitative traits can still be used to infer a general picture on the genetic relationship among cassava accessions.

The analysis of variance indicated significant differences in LR between 1) the germplasm from the countries' breeding programmes and 2) local and elite genotypes. The standard error associated with the evaluations was modest (0.104) and broad sense heritability has been reported to be 0.55 in evaluated germplasm (Lenis et al., 2006), providing confidence in the generated results. Cassava experiences simultaneous growth and development of the roots and leaves. Thus, if longevity of leaves is increased to maintain high photosynthetic rates, it could be possible to maintain a given leaf area index with less distribution of assimilates for leaf development and hence more to the root development (Hershey, 2010). Hence, this trait may present an additional opportunity to increase cassava yield.

Highest LR in both local and elite genotypes was recorded in Madagascar. Local genotypes had higher LR than elite genotypes. This is the first analysis of LR in cassava germplasm from these countries, indicating that it is not a trait that has been extensively selected for in most elite genotypes. This could explain the relatively high LR in local compared to elite genotypes. In Madagascar, cassava leaves are a major vegetable in subsistence communities, which could perhaps explain the higher leaf retention in both elite and local genotypes.

Assignment of different genotypes to the varying phenotypes on the 1-5 scale for LR is indicative of the quantitative nature of this trait and its high frequency in cassava. Lenis et al. (2006) observed LR in 37 out of 62 cassava families evaluated, with some families having a higher frequency of clones with the LR trait than others. However, phenotypic correlations based on 110 cassava clones indicated non-significant positive correlation between LR and HI (0.15) and between LR and DMC (0.17) (Lenis et al., 2006). In the current study, phenotypic correlations were negative with HI ($r = -0.288$; $P < 0.0001$) and DMC ($r = -0.274$; $P < 0.001$). These findings certainly present a discrepancy when compared to earlier studies, a phenomenon that needs to be settled once more detailed studies are conducted.

Root DMC varied significantly between 1) the cassava germplasm from the different countries' breeding programmes and 2) local and elite genotypes. The trial sites were not under severe biotic stress and only DMC data for genotypes that provided 3-5 kg of root weight, as required by the specific gravity method, were included in the analysis. In addition, DMC evaluations were done before onset of the rainy seasons, as DMC is reduced with onset of rains (Tan and Mak, 1995; Lenis et al., 2006). The standard error associated with the DMC evaluations was relatively low (0.858), which further provided confidence in the obtained results.

The highest DMC was recorded in the elite germplasm from Uganda, which comprised of elite genotypes from IITA and hybrids generated by the Uganda National Cassava Breeding Programme using elite parental genotypes from IITA. The breeding scheme at IITA largely utilises the recurrent selection procedure (IITA, 1990) and hence most elite genotypes from IITA will have high DMC. Previous studies have established significant general combining ability (GCA) estimates for DMC (Cach et al., 2005; 2006; Jaramillo et al., 2005), indicating preponderance of additive variance, which can further explain in part, the relatively high DMC in Ugandan hybrids generated by hybridising elite IITA lines. Furthermore, in Uganda, direct selection for DMC is routine, while in most other countries selection for fresh root yield is routine. This discrepancy could, in part, explain the observed differences in DMC among elite genotypes.

The presence of relatively high DMC in some local genotypes i.e. "Zanzibar" from Madagascar with 42% DMC, further demonstrated the importance of immediate exploitation of local genotypes. Local varieties like "Mbundumali" with up to 43% DMC have been reported in Malawi (Benesi, 2005). Cassava is primarily a starchy crop, hence the current local varieties could have been positively selected for by subsistence farmers, who have cultivated the crop since its introduction, explaining the high DMC levels in some of the local genotypes. In the evaluation of 2022 genotypes from Latin America, DMC ranged from 10.72 to 57.23%, with a mean of 34.27% (Chávez et al., 2005).

However, in a BC_2 population, DMC ranged from 34.3 to 42.7% (Ojulong et al., 2008b). When 672 cassava genotypes were evaluated, DMC ranged from 20.6 to 41.2% (Ojulong et al., 2008a). DMC recorded in the present study ranged from 16.3 to 49.6% and is comparable with previous studies that mainly evaluated different cassava populations from Latin America. These findings suggest that enough DMC genetic variability exists in the region to initiate a breeding scheme, once frameworks for germplasm exchange have been finalised. At a national level, hybridisation schemes involving "elite x local" or "local x local" can be initiated in countries that have not yet started breeding for increased DMC.

HI varied significantly between 1) the germplasm from the different countries' breeding programmes and 2) local and elite genotypes. Relatively high broad sense heritability values for HI of up to 0.8 have been reported (Ojulong et al., 2008a). Kawano (2003) observed that in single row trials, indirect selection for yield through HI was more efficient than direct selection for yield itself. In this study, the standard error associated with HI evaluations was low (0.0021), which further gave confidence in the obtained results. The highest HI was observed in Ugandan elite germplasm, which can partly be explained by the same reasoning purported for DMC, as both high HI and DMC are specific breeding objectives of IITA (IITA, 1990). Local genotypes from Rwanda had exceptionally high HI compared to the elite genotypes, a finding which cannot be explicitly explained, as positive selection for HI involves methodical procedures, which can hardly be done by subsistence farmers. However, significant specific combining ability that is indicative of dominance variance has been reported for HI (Cach et al., 2005; 2006; Calle et al., 2005), which may partly explain this scenario.

In a BC_2 population, HI ranged from 0.17 to 0.55. The authors noted a reduction of HI with backcrossing (Ojulong et al., 2008b). When 1950 genotypes were evaluated in single row trials at CIAT, HI ranged from 0.0 to 0.75 (Kawano, 2003). Recent trials conducted in Nigeria at three sites indicated that HI ranged from 0.06 to 0.92 (Egesi et al., 2007).

Results from the present study indicated that HI ranged from 0.04 to 0.90 which is comparable to previous evaluations done in both Latin America and west Africa, indicating variability in HI. The doubling of fresh root yield in cassava within a short period since the inception of cassava breeding at CIAT, was largely due to improvement in HI (Kawano, 2003). This should be a motivation to start utilising HI in the national breeding programmes of Africa. Ceballos et al. (2004) noted that HI could have been over-exploited by now in most Latin American germplasm. This however, is not the case in most African cassava breeding programmes, which largely make selections based on fresh root yield and not HI.

Root peel thickness varied significantly between 1) the germplasm from the different countries' breeding germplasm and 2) local and elite genotypes. The standard error associated with measurements of root peel thickness was 0.085. Peel thickness could be involved in resistance and/or tolerance to: root-feeding insects, root pathogens and post-harvest handling damage. Another hypothesis could be that a thicker peel will probably mean a higher proportion of dry matter partitioned to a non-usable product, or at least a less valuable product (Hershey, 2010). Most of these hypotheses have not been tested. This is the first quantitative study on this trait in cassava, from which it was established that cassava genotypes display naturally varying levels of root peel thickness ranging from 0.34 to 4.89 mm. In this study, the phenotypic correlation between root cortex and DMC was 0.263, which indicated that selection for increased DMC will increase root cortex thickness. However, the R^2 value between root cortex thickness and DMC was low (0.0693). This suggested that more assimilates will be apportioned to the root parenchyma than to the root cortex.

Sinha and Swaminathan (1984) reviewed parameters for plant breeding for the developing world, which included finding sources of biomass and renewable energy. This is because, with the depletion of fossil fuel reserves, energy is becoming a major limiting factor in economic development and hence alternatives are needed, which may include plant residues.

The cassava root cortex currently has limited use and hence can be one alternative to use. This study has established that genetic variability for this trait exists naturally and can support a breeding scheme devoted to its improvement.

3.5 Conclusions

This is the first comprehensive regional study on cassava phenotypic variability. For countries where initial work has been done, this study is a continuation of that pioneering work. The 29 morphological traits provided limited discrimination of the cassava germplasm. It was therefore important to examine the germplasm using the more robust SSR markers to get a better picture of cassava diversity within the NARS. Results based on SSR analysis are presented in Chapter 4. Some of the phenotypic traits examined in this study i.e. DMC and LR, are important for fitness in the environment. Information on phenotypic plasticity which is environment dependent (Schlichting, 1989), will be particularly important in breeding for climatic uncertainty and extreme environments.

This study has established that some local cassava genotypes have reasonably high amounts of DMC, HI and LR. One of the most successful international cassava breeding programmes coordinated by CIAT began with the collection and evaluation of over 2000 cassava varieties mainly from farmers' fields in Latin America. The selected varieties were hybridised to generate progeny for further advancement from which outstanding commercial genotypes were officially released (Kawano, 2003). Results of this study can therefore be tailored to achieve the same goal in SSA.

Upon defining breeding objectives and selection of target environments, breeders are tasked to establish if adequate genetic variation is available and what breeding scheme to adopt (Hallauer, 1992; Maunder, 1992). This study has found substantial genetic variation in some of the agronomically important cassava traits. The recurrent selection scheme as demonstrated by IITA appears to be a reliable breeding scheme to attain genetic progress in these traits.

It is therefore strongly recommended that national breeding programmes define their respective breeding objectives and begin undertaking cassava breeding to increase cassava productivity at both national and regional level. Selection, if undertaken appropriately with controls and selection indices, should lead to attainment of true genetic progress. These efforts will certainly redress the limited utilisation of germplasm by plant breeders as already observed (Baenziger and Peterson, 1992).

3.6 References

Alonso-Blanco, C., B. Mendez-Vigo, and M. Koornneef, 2005. From phenotypic to molecular polymorphism involved in naturally occurring variation of plant development. International Journal of Developmental Biology 49:715-732.

Baenziger, P.S., and C.J. Peterson, 1992. Genetic variation: its origin and use for breeding self pollinated species. In: Stalker, H.T., and J.P. Murphy (Eds.), Plant breeding in the 1990s. Proceedings of the symposium on plant breeding in the 1990s held at North Carolina State University Raleigh, NC, pp. 69-92. CABI International Wallington Oxon OX108DE, UK.

Balyejusa Kizito, E., 2006. Genetic and root growth studies in cassava (*Manihot esculenta* Crantz): implications for breeding. PhD Thesis, Swedish University of Agricultural Sciences, Uppsala, Sweden, pp. 127.

Barton, N.H., and P.D. Keightley, 2002. Understanding quantitative genetic variation. Nature Reviews 3:11-21.

Benesi, I.R.M., 2005. Characterisation of Malawian cassava germplasm for diversity, starch extraction and its native and modified properties. PhD Thesis, University of the Free State, South Africa, pp. 199.

Bhatt, G.M., 1973. Comparison of various methods of selecting parents for hybridization in common bread wheat (*Triticum aestivum* L.). Australian Journal of Agricultural Research 24:457-464.

Cach, N.T., J.C. Perez, J.I. Lenis, F. Calle, N. Morante, and H. Ceballos, 2005. Epistasis in the expression of relevant traits in cassava (*Manihot esculenta* Crantz) for subhumid conditions. Journal of Heredity 96:586-592.

Cach, N.T., J.I. Lenis, J.C. Perez, N. Morante, F. Calle, and H. Ceballos, 2006. Inheritance of useful traits in cassava grown in subhumid conditions. Plant Breeding 125: 177-182.

Calle, F., J.C. Perez, W. Gaitan, N. Morante, H. Ceballos, G. Llano, and E. Alvarez, 2005. Diallel inheritance of relevant traits in cassava (*Manihot esculenta* Crantz) adapted to acid-soil savannas. Euphytica 144:177-186.

Ceballos, H., C.C. Iglesias, J.C. Pérez, and A.G.O. Dixon, 2004. Cassava breeding: opportunities and challenges. Plant Molecular Biology 56:503-515.

Chang, T.T., 1992. Crop Genetic sources. In: Stalker, H.T., and J.P. Murphy (Eds.), Plant Breeding in the 1990s. Proceedings of the symposium on plant breeding in the 1990s held at North Carolina State University Raleigh, NC, pp. 83-103. CABI International Wallington Oxon OX108DE, UK.

Chang, T.T., W.L. Brown, J.G. Boonman, J. Sneep, and H. Lamberts, 1979. Crop genetics resources. In: Sneep, J., A.J.T. Hendriksen, and O. Holbek (Eds.), Plant breeding perspective: Centennial publication of Koninklijk Kweekbedrijf en Zaadhandel D.J. van der Have 1879-1979, pp. 83-103. Centre for Agricultural publishing and documentation, Wageningen, Netherlands.

Chávez, A.L., T. Sánchez, G. Jaramillo, J.M. Bedoya, J. Echeverri, E.A. Bolaños, H. Ceballos, and C.A. Iglesias, 2005. Variation of quality traits in cassava roots evaluated in landraces and improved clones. Euphytica 143:125-133.

Egesi, C.N., P. Ilona, F.O. Ogbe, M. Akoroda, and A. Dixon, 2007. Genetic variation and genotype x environment interaction for yield and other agronomic traits in cassava in Nigeria. Agronomy Journal 99:1137-1142.

Fischbeck, G., 1989. Evaluation of cereals in Europe. In: Brown, A.H.D., D.R. Marshall, O.H. Frankel, and J.T. Williams (Eds.), The use of plant genetic resources, pp. 197-211. Cambridge University Press, New York.

Grando, S., R. von Bothmer, and S. Ceccarelli, 2000. Genetic diversity of barley: use of locally adapted germplasm to enhance yield and yield stability of barley in dry areas. In: Cooper, H.D., C. Spillane, and T. Hodgkin (Eds.), Broadening the genetic base of crop production. pp. 351-371. CABI, Wallingford, FAO, Rome, IPGRI, Rome.

Hallauer, A.R., 1992. Use of genetic variation for breeding populations in cross pollinated species. In: Stalker, H.T., and J.P. Murphy (Eds.), Plant Breeding in the 1990s. Proceedings of the symposium on plant breeding in the 1990s held at North Carolina State University Raleigh, NC, pp. 37-67. CABI International Wallington Oxon OX108DE, UK.

Hartl, D.L., and A.G. Clarke, 1989. Principles of population genetics. Second Edition. Sinauer Associates Inc., Publishers, pp. 682.

Hershey, C., 2010. Cassava breeding: theory and practice. A publication by Food and Agriculture Organisation (FAO), in press.

Hintze, J., 2001. NCSS and PASS. Number cruncher statistical systems. Kayville, Utah., www.NCSS.com.

IITA, 1990. Cassava in tropical Africa. A Reference Manual. International Institute of Tropical Agriculture, Ibadan, Nigeria, pp. 176.

Jaramillo, G., N. Morante, J.C. Perez, F. Calle, H. Ceballos, B. Arias, and A.C. Bellotti, 2005. Diallel analysis in cassava adapted to the mid altitude valleys environment. Crop Science 45:1058-1063.

Jones, W.O., 1959. *Manioc* in Africa. Standard University Press, Stanford, California, pp. 315.

Kawano, K., 1990. Harvest index and evolution of major food crops cultivars in the tropics. Euphytica 46:195-202.

Kawano, K., 2003. Thirty years of cassava breeding for productivity-biological and social factors for success. Crop Science 43:1325-1335.

Kawano, K., W.M.F. Goncalves, and U. Cenpukdee, 1987. Genetic and environmental effects on dry matter content of cassava root. Crop Science 27:69-74.

Klug, W.S., M.R. Cummings, and C. Spencer, 2005. Concepts of Genetics. Eighth Edition, pp. 677.

Lenis, J.I., F. Calle, G. Jaramillo, J.C. Perez, H. Ceballos, and J.H. Cock, 2006. Leaf retention and cassava productivity. Field Crops Research 95:126-134.

Linhart, Y.B., and M.C. Grant, 1996. Evolutionary significance of local genetic differentiation in plants. Annual Review of Ecological Systems 27:237-277.

Maunder, A.B., 1992. Identification of useful germplasm for practical plant breeding programmes. In: Stalker, H.T., and J.P. Murphy (Eds.), Plant Breeding in the 1990s. Proceedings of the symposium on plant breeding in the 1990s held at North Carolina State University Raleigh, NC, pp. 147-169. CABI International Wallington Oxon OX108DE, UK.

Mohammadi, S.A., and B.M. Prasanna, 2003. Analysis of genetic diversity in crop plants-Salient statistical tools and considerations. Crop Science 43:1235-1248.

Ojulong, H., M.T. Labuschagne, L. Herselman, and M. Fregene. 2008b. Introgression of genes for dry matter content from wild cassava species. Euphytica 164:163-172.

Ojulong, H., M.T. Labuschagne, M. Fregene, and L. Herselman, 2008a. A cassava clonal evaluation trial based on a new cassava breeding scheme. Euphytica 160:119-129.

Olsen, K.M., 2004. SNPs, SSRs and inferences on cassava origin. Plant Molecular Biology 56:517-526.

Olsen, K.M., and B.A. Schaal, 1999. Evidence of the origin of cassava: phylogeography of *Manihot esculenta*. Proceedings of the National Academy of Sciences (USA) 96:5586-5591.

Olsen K.M., and B.A. Schaal, 2001. Microsatellites variation in cassava (*Manihot esculenta*, Euphorbiaceae) and its wild relatives: further evidence for a southern Amazonian origin of domestication. American Journal of Botany 88:131-142.

Perrier X., and J.P. Jacquemound-Collet, 2006. DARwin software. http://darwin.cirad.fr/darwin

Rogers, D. J., and T.T. Tanimoto, 1960. A computer program for classifying plants. Science 132:1115-1118.

Schlichting, C.D., 1989. Phenotypic integration and environmental change. Bioscience 39:460-464.

Sinha, S.K., and M.S. Swaminathan, 1984. New parameters in plant breeding. In: Vose, P.B., and S.G. Blixt (Eds.), Crop breeding a contemporary basis, pp. 1-31. Pergamon Press Ltd, Germany.

Smith, J.S.C., and D.N. Duvick, 1989. Germplasm collections and the private plant breeder. In: Brown, A.H.D., D.R. Marshall, O.H. Frankel, and J.T. Williams (Eds.), The use of plant genetic resources, pp. 17-31. Cambridge University Press.

Sultan, S.E., 2003. Phenotypic plasticity in plants: a case study in ecological development. Evolution and Development 5:25-33.

Tan, S.L., and C. Mak, 1995. Genotype x environment influence on cassava performance. Field Crops Research 42:111-123.

CHAPTER 4

PATTERNS OF ALLELE FREQUENCY DISTRIBUTION IN CASSAVA GERMPLASM AVAILABLE WITHIN SELECTED NATIONAL BREEDING PROGRAMMES IN AFRICA

4.1 Introduction

Studying allele frequencies is particularly important for plant breeding because allelic variation is the raw material on which true genetic gain can be attained. Selection limits largely depend on the initial number of segregating alleles in the population rather than the heterozygosity (Hill and Rasbash, 1986). Examining genetic variation at representative marker loci allows genetic classification of populations and more importantly, provides insights into heritable traits that can be used in crop improvement. Sustaining genetic gains in yield, increased tolerance to abiotic and biotic stress and diversified product utilisation, will depend on existing genetic variation within the species of interest. Moreover, estimates of gene diversity and population differentiation (F_{ST}) can be used as indirect ways of measuring adaptive polygenic traits (Toro et al., 2009).

Studies conducted in barley over several generations established that high grain yield, high seed numbers per plant and other characteristics associated with high reproductive capacity were always associated with the most frequent allele of each polymorphic locus (Allard, 1988). In wheat, a comparison of the genetic profile of three ancient accessions and 45 modern cultivars indicated no appreciable inheritance of allele signatures from the ancient germplasm (Maccaferri et al., 2003), a finding which can be used to define future wheat breeding objectives. The above narration clearly portrays how information on allele distribution in cassava, if generated, can help define and/or set priorities for national cassava breeding teams in SSA.

In terms of computation, allele frequencies can be computed from alleles detected using a range of molecular marker techniques. SSRs which are densely interspersed in eukaryotic genomes (Tautz and Renz, 1984) have in the past decade and until now, continually been used to study genetic variation in plants and animal species (Powell et al., 1996; Matsuoka et al., 2002; Morgante et al., 2002; Maccaferri et al., 2003; Toro et al., 2009). When these genomic regions are amplified using a pair of unique flanking oligonucleotides primers, they invariably show extensive polymorphism (Morgante and Olivieri, 1993). The resultant high level of allelic diversity, coupled with stability and co-dominant nature of inheritance, make genomic SSRs ideal for studying neutral genetic variation both within and between populations. Moreover, highly significant association of microsatellite frequency and single copy DNA in some plant genomes has been established (Morgante et al., 2002). This finding further makes SSRs attractive for genetic analysis in orphan crops like cassava, where methodical analysis of genetic variation of germplasm available within NARS in SSA has been limited.

Most cassava acreage in Africa lies in SSA particularly in countries of west, central and east Africa (Nweke et al., 2002; Hershey, 2010). The crop is popular in this region because of its high starch yield per unit area and its recently renewed interest as a potential biofuel crop. The broad genetic composition of cassava in the east, central and southern African regions comprises of local "unimproved" varieties and elite varieties that have undergone intensive selection by either IITA or the NARS in the region. Generally, this germplasm collection has an array of attributes ranging from yield potential to culinary qualities, which are social-cultural specific, with a loose-fitting classification as either "sweet" or "bitter" cassava (Hershey, 2010).

Vellvė (1993) reported that formal breeding significantly reduced genetic diversity in European agriculture. Other studies have strongly demonstrated increased genetic variability with formal breeding (Maccaferri et al., 2003). The effect of formal breeding on the genetic diversity of cassava is not known.

However, what remains relevant for the future is the need for plasticity in cassava to respond to changes in climate, quality requirements and alien pests. These factors require that the genetic structure of cassava be well-known.

Cassava genetic diversity assessments in SSA have been done using a range of molecular markers (Fregene et al., 2000; Zacarias et al., 2004; Balyejusa Kizito et al., 2005; Benesi, 2005), which inevitably provided varying outputs, involved different analysis methods and applications. For instance, the Cassava Molecular Diversity Network (MOLCAS), whose overall goal is to enhance the dissection and utilisation of diversity of local landraces, employed SSR markers to study diversity in northern Malawi, southern Tanzania, Uganda and some selected sites in Latin America (Anonymous, 2001). Other key cassava producing countries in the region, notably the DRC, the island of Madagascar and Rwanda, have hardly undertaken any methodical cassava genetic analysis studies. Detailed knowledge on genetic structure and variability of the cassava germplasm available within the NARS breeding programmes is critically important for effective regional and national conservation prioritisation and/or defining of breeding objectives.

Besides quantifying genetic variability within cassava germplasm, this study aimed at obtaining initial insights into allelic contribution of local varieties to the modern elite varieties developed by IITA for SSA. Ideally, the coefficient of parentage and/or co-ancestry (*f*) should serve for this purpose because it reflects the degree of relatedness between two individuals based on their pedigree. Unfortunately, pedigree information in cassava is scanty. Besides, *f* is an indirect measurement that relies on the expected proportion of alleles identical by descent in the absence of selection, mutation and drift (Melchinger et al., 1991), a situation which does not necessarily always hold. It is for these reasons that allelic contribution of local varieties was studied using the genetic distance based on highly informative and heritable SSRs. In this research chapter, genotypic data generated from 1401 cassava genotypes assayed at 26 SSR loci is presented.

The first objective was to understand the nature, extent, distribution and hierarchical organisation of genetic variation that exists within the NARS of seven countries: Tanzania, Uganda, Kenya, Rwanda, DRC, Madagascar and Mozambique, so as to develop science-based breeding strategies. The second objective was to quantify the extent of allelic similarity between selected popular local varieties grown in each of these countries and a selection of elite IITA varieties that have been widely adopted in the east, central and southern African regions.

4.2 Materials and methods

4.2.1 Cassava germplasm

The germplasm included in this study represented a spectrum of cultivars released through formal breeding (hereafter referred to as elite genotypes) and farmer unimproved varieties (hereafter referred to as local genotypes). Leaf samples were collected from the field trials that were analysed in Chapter 3: Tanzania (279), Uganda (270), Kenya (239), Rwanda (192), DRC (192), Madagascar (189) and Mozambique (82).

The discrepancy between the number of genotypes analysed in Chapter 3 and Chapter 4 was due to the fact that while leaf samples could easily be collected at once from most genotypes, all phenotypic data could not be collected on all genotypes and hence less genotypes were included in the analysis for Chapter 3. The harvested young fresh leaf samples were freeze dried in dry ice and total genomic DNA extracted using the miniprep protocol as described by Dellaporta et al. (1983). Of the 1443 cassava genotypes sampled originally, a total of 1401 were successfully assayed and used in the analysis (Table 4.1). Because these cassava genotypes were collected from seven different countries, and in each country both elite and local genotypes were sampled, they constituted hierarchical levels whose genetic structure was investigated. It was hypothesised that cassava genotypes in each country had acquired different allelic frequencies and hence formed the hierarchical groups.

Table 4.1 Cassava germplasm used for genetic analysis

Country	No. of local varieties	No. of elite varieties	Total number
Tanzania	249	21	270
Uganda	203	65	268
Kenya	37	197	234
Rwanda	47	137	184
DRC	104	73	177
Madagascar	134	52	186
Mozambique	73	9	82
Total	**847**	**554**	**1401**

4.2.2 Microsatellite genotyping and allele calls

Cassava genotypes were assayed with 26 highly polymorphic SSR markers (Appendix 1). These markers represented 13 linkage groups, with six markers being unmapped (Morag Ferguson, personal communication). Map positions of these SSRs are available (Mba et al., 2001; Okogbenin et al., 2006). Twenty four of these SSRs had previously been used in a global diversity assessment of more than 2000 cassava genotypes and found to be highly informative and provided a similar overall population structure to that provided by 30 SSR loci (Morag Ferguson, personal communication). These loci were furthermore selected based on single-locus amplification, high degree of polymorphism and reproducibility.

Amplifications with SSR primers were performed in 10 µl reactions containing 50 ng of DNA, 1 pmole of each primer, 1x *Taq* polymerase buffer, 2 mM $MgCl_2$, 0.2 mM deoxynucleotide triphosphates (dNTPs) and 0.375 U *Taq* polymerase (New England Biolabs Inc.). The PCR profile was 95^{o}C for 2 min, followed by 30 cycles of 95^{o}C for 30 sec, 55-57^{o}C for 1 min and 72^{o}C for 1 min and a final extension at 72^{o}C for 30 min. Because amplicons were of different sizes and the forward primers were fluorescently labelled (by manufacturers, Mwg-Biotech), co-loading of amplicons from the same individual, but at different loci was therefore possible. Seven co-loading sets were optimised and used for the entire analysis (Table 4.2). For each co-loading set 1 to 2 µl of the different amplicons were mixed and briefly vortexed.

Aliquots of 1 µl of the mixture were added to 9 µl of a master-mix containing HiDi formamide and GeneScan 500 LIZ size standard (1 ml of HiDi + 12 µl of 500-LIZ). Amplicons were denatured (95^{o}C for 3 min) and subjected to capillary electrophoresis using the ABI 3730 DNA sequencer (Applied Biosystems) and allele calls made using the GENEMAPPER® software version 3.7 (Applied Biosystems). All genotyping was done at the Biosciences Eastern and Central Africa (BecA) hub in Nairobi, Kenya.

Table 4.2 Optimised co-loading sets of 26 SSR primers based on amplicon size and dye colour

SSRY loci	Co-loading set	Amplicon size range (bp)	Dye label
SSRY21	1	120-230	NED
SSRY38	1	80-140	6-FAM
SSRY59	1	130-180	PET
SSRY69	1	180-270	VIC
SSRY5	2	70-150	6-FAM
SSRY161	2	200-240	NED
SSRY52	2	230-280	PET
NS911	2	90-150	VIC
SSRY155	3	130-180	NED
SSRY151	3	160-240	6-FAM
SSRY110	3	230-260	PET
SSRY9	3	245-285	VIC
SSRY12	4	220-290	NED
SSRY100	4	170-280	PET
SSRY102	4	160-200	6-FAM
SSRY147	4	80-140	VIC
SSRY63	5	240-340	PET
SSRY148	5	90-140	VIC
SSRY182	5	190-260	NED
SSRY181	5	130-230	6-FAM
SSRY135	6	230-265	PET
SSRY171	6	240-330	NED
SSRY169	6	70-130	6-FAM
SSRY19	7	190-250	6-FAM
SSRY51	7	230-330	PET
SSRY64	7	160-230	VIC

4.2.3 Data analysis

4.2.3.1 Gene diversity and allelic richness

Genetic variability of a sample can be measured by the frequencies of genotypes and alleles, proportion of polymorphic loci, observed and expected heterozygosity and the allelic richness (Toro et al., 2009). In this study expected heterozygosity (gene diversity; H_e), defined by Nei (1973) as the probability that two alleles chosen at random from a population are different, was used. H_e was calculated for each locus using the formulae $H_e = 1-\sum p_{il}^2$ where p_i is the frequency of the i^{th} allele at the l^{th} locus. The H_e-values were computed for each locus for the cassava genotypes constituting the sample.

In addition, allelic richness (number of alleles segregating in the population) was examined because: 1) selection limits are determined by the initial number of alleles (Hill and Rasbash, 1986), 2) it is of relevance in conservation programmes and 3) unlike H_e, it is sensitive to bottlenecks and gives a better reflection on past population fluctuations (Toro et al., 2009). Because allelic richness is largely dependant on sample size, it was necessary to compensate for differences in sample sizes. For this purpose, the rarefaction index as suggested by El Mousadik and Petit (1996) was adapted for the computation of allelic richness. The principle for the rarefraction is: if N_{ik} represents the number of copies of the k^{th} allele from the sample of population i and N_i represents the total number of genes sampled from that population, the allelic richness at one locus is denoted as the expected number of different alleles that a sample would have had if the sample size had been g number of genes (the smallest sample size) instead of N_i. Computation of allelic richness and gene diversity were done using the FSTAT software programme version 2.9.3.2 (Goudet, 2001).

$$\text{Allelic richness} = \sum_k \left\{ 1 - \frac{\begin{bmatrix} N_i - N_{ik} \\ g \end{bmatrix}}{\begin{bmatrix} N_i \\ g \end{bmatrix}} \right\}$$

Because movement of genes and individuals counters divergence (Hartl and Clarke, 1989), estimates of gene flow based on variances in allele frequencies among local and elite cassava within each NARS were computed. Gene flow was estimated by *Nm*, the number of migrants coming into a population per generation: $Nm = 0.25(1 - F_{ST})/ F_{ST}$ as described by Nei (1987), where F_{ST}, also considered as the Wahlund effect, is the measure of heterozygote deficit among populations. This analysis was done using Popgene software version 1.31 (Yeh et al., 1999).

4.2.3.2 Genetic relationships and variance

Genetic relationships between: 1) elite and local genotypes and 2) local genotypes from the different breeding programmes were obtained by subjecting the genotypic data to phenetic analysis. Phenetic analyses uses distance matrix-based methods to construct trees based on overall similarity (Mohammadi and Prasanna, 2003). These analyses are often assumed to approximate phylogenetic relationships and thus, phylogenetic analyses attempt to determine evolutionary pathways, whereas phenetic analyses depict present day differences. The simple matching Euclidean distance (D_E) was used to compute the distance matrix and clustering was done using the weighted neighbour-joining algorithm and relationships displayed as a dendrogram. This analysis was respectively done using PowerMarker version 3.25 (Liu and Muse, 2005) and DARwin software (Perrier and Jackquemound-Collet, 2006). PowerMarker makes provision for calculating genetic distance for groups of individuals, while DARwin is largely limited to genetic distance among individuals.

The organisation of genetic variation was revealed by partitioning the variance among the hypothesised hierarchical levels. Each cassava collection per country was considered a group; hence seven groups were sampled. The hypothesis to consider cassava from each country as a group is based on the fact that cassava farmers make different selections (Balyejusa Kizito, 2006) for a range of attributes across the region, depending on the agro-ecological environment, organoleptic preferences and other traits deemed important. These factors may most likely be shaping the cassava genetic structure.

Moreover, the east and central African regions had two independent cassava introductions, firstly through the east African coast in the 1750s (Jones, 1959) and then through west Africa via Sudan on its way to northern Uganda in the 1870s (Langlands, 1966). These separate introductions could have caused different founder effects (Ladizinsky, 1985).

It has been observed that populations often display differences in allele and genotypic frequencies which vary from one geographic region to another (Hartl and Clarke, 1989). In chapter 3, phenotypic differences in quantitative traits were observed among cassava germplasm from different countries. Furthermore, within a country, cassava genotypes were classified as either local or elite thus constituting sub-groups within a group. Partitioning of variance among hierarchical levels provides for an estimation of the components of genetic variance (Weir, 1990). Accordingly, the genotypic data set was subjected to analysis of molecular variation (AMOVA) using the Arlequin version 3.1 software (Excoffier et al., 2005) to get insights into this hierarchical genetic organisation.

4.2.3.3 Genetic differentiation and structure

Quantification of genetic differentiation resulting from either genetic sampling and/or drift was examined using F-statistics (Weir, 1990). The Weir and Cockerham (1984) F-statistics which weighs allele frequencies according to sample size were used for this purpose, as it considers sampling error. F-statistics (F_{ST}, F_{IS} and F_{IT}) also known as fixation indices, describe the level of heterozygosity in a population when compared to Hardy–Weinberg expectation. The three hierarchical F-statistics are inbreeding coeficients: 1) F_{ST} - inbreeding of individuals in sub-populations relative to the total populations to which they belong; 2) F_{IS} - inbreeding in individuals relative to the sub-population to which they belong and 3) F_{IT} - inbreeding due to combined effects of non-random mating within sub-populations and random genetic drift among sub-populations (Hartl and Clarke, 1989).

F_{ST} values are always positive and when zero, it indicates that the sub-populations are identical in allele freqencies and when it is one, it indicates that the sub-populations are fixed for different alleles. On the other hand, F_{IS} and F_{IT} values can either be negative indicating an excess of heterozygotes or positive indicating deficit of heterozygotes (Mohammadi and Prasanna, 2003). F-statistics have previoulsy been used to study genetic differentiation in cassava (Balyejusa Kizito, 2006).

Because of the mutational dynamics of microsatellites, care must be taken in estimating genetic differentiation using the correct genetic models (Colson and Goldstein, 1999; Laval et al., 2002), which can either be the stepwise mutation model (SMM) or infinite allele model (IAM). Under the SMM, an allele carrying *i* repetitions can mutate to an allele carrying $j = i \pm 1$ repetitions to create allelic states over time (Kimura and Crow, 1964). Alleles of similar size are assumed to be more closely related to each other than those of different size. For these reasons, SMM is expected to be more accurate for populations that diverged long ago because current genetic differentiation reflects mutations accumulated since divergence (Goldstein and Pollock, 1997).

On the other hand, the IAM model assumes that a mutation generates new alleles not previously present in the population (Ohta and Kimura, 1973). This model is more appropriate for recently diverged populations, as genetic differentiation among such populations will reflect more of the sorting of ancestral variation than mutation divergence (Goldstein and Pollock, 1997). Hence, the F-statistics for estimating genetic differentiation were computed under the IAM model, as cassava is a recently diverged crop (Olsen and Schaal, 1999; 2001). Jacknifing over samples and loci and bootstrapping over loci were performed for the computation of all F-statistics and all analyses were done using FSTAT version 2.9.3.2 (Goudet, 2001).

To assign individuals (cassava genotypes) to populations on the basis of their genotypes, the data set was subjected to structure analysis using the admixture model. This Bayesian based model assumes that each individual inherited some portion of its ancestry from each one of the *K* populations (Pritchard et al., 2000). Initially, the length of the burn-in period and Markov Chain Monte Carlo (MCMC) replications were set at 10000 for $K = 1$ to 17 (number of putative populations in the model); this was repeated ten times. The ΔK statistic, which is based on the rate of change in the log probability of the data between successive *K*-values, was then used to detect the true number of *K* populations in the dataset (Evanno et al., 2005). The determined population *K* was re-run at a burn-in period and MCMC replications of 100000 to verify its consistency.

4.2.3.4 Allele sharing between popular local and elite varieties

In this study allele similarity between popular local varieties grown in each of the selected African countries and a selection of popular elite cassava varieties developed and disseminated by IITA, was done. For this purpose 15 elite IITA cassava varieties that were released between the period 1990 and 2002 (I91/2327, 92/00057, I92/00067, 95/SE00036, MM95/0104, MM96/7688, MH97/2961, MM96/1419, MM01/0014, MM 02/0993, TMS30572, TME14, NASE 1, NASE 4 and NASE 12) were selected. The varieties have been fairly adopted in SSA. For local cassava varieties, Nyaraboke (Uganda), Liongo Kwimba (Tanzania), Kibandameno (Kenya), Boma (DRC), Gitamisi (Rwanda), Nakasoga (Madagascar) and Nikwaha (Mozambique) were selected.

Each of the local varieties was compared to the 15 IITA elite varieties. This comparison was done by computing Rogers's genetic distance (d_R) using allelic frequency data generated from the 26 loci using PowerMarker version 3.25 software (Liu and Muse, 2005). A methodical study illustrated that Roger's genetic distance is linearly related to the coefficient of co-ancenstry and for this reason it is desirable for uncovering pedigree relationships among operational taxonomic units (Reif et al., 2005).

4.3 Results

4.3.1 Gene diversity and allelic richness

Genotypic data generated from 1401 cassava genotypes assayed at 26 SSR loci was analysed. Fourty two genotypes had more than 10% missing data and were excluded from this analysis. The missing alleles could not be accurately scored and/or were due to repetitive PCR failure. Data on allelic richness, gene diversity and gene flow among the cassava germplasm available within the NARS of Tanzania, Uganda, Kenya, Rwanda, DRC, Madagascar and Mozambique is presented in Table 4.3. With the exception of cassava from Tanzania and Rwanda, allelic richness was higher in the elite genotypes compared to local cassava varieties. Cassava from the DRC had the highest number of segregating alleles (local cassava = 93.5; elite cassava = 97.5).

Lowest number of segregating alleles in local (81.9) and elite (82.5) genotypes were respectively observed in Uganda and Tanzania cassava germplasm (Table 4.3). Overall, the lowest average segregating alleles (84.8) were observed in Uganda germplasm and the highest (95.5) in DRC germplasm (Table 4.3). A similar trend was observed for gene diversity, with the highest (0.63) being observed in elite cassava from the DRC and the lowest (0.53) in local varieties of Uganda (Table 4.3). However, mean gene diversity between local and elite cassava was similar. Cassava from Mozambique had relatively similar genetic diversity and allelic richness for both elite and local cassava genotypes.

However, with no correction for varying sample size, the total number of alleles sampled in the cassava within the NARS presented a different outlook (Table 4.4). The data suggested that local cassava varieties from Tanzania, DRC, Madagascar and Mozambique had higher allelic richness than their elite counterparts, with the highest diversity being reported in cassava from local varieties from Tanzania and the lowest from elite varieties from Mozambique (Table 4.4). The data indicated that the local genotypes had the highest (125.8) and elite the lowest (119.4) number of segregating alleles (Table 4.4).

Table 4.3 Trends in allelic richness, gene diversity and gene flow of cassava germplasm available within the NARS[1]

Country	Allelic richness			Gene diversity				
	Loc.	Eli.	Mean	Loc.	Eli.	Mean	F_{ST}	*Nm*
Tanzania	89.58	82.52	86.05	0.60	0.55	0.56	0.017	14.14
Uganda	81.94	87.84	84.89	0.53	0.57	0.55	0.044	5.36
Kenya	82.24	92.35	87.30	0.58	0.61	0.59	0.038	6.21
Rwanda	88.36	85.71	87.04	0.59	0.56	0.57	0.040	5.98
DRC	93.51	97.53	95.52	0.61	0.63	0.62	0.024	9.84
Madagascar	84.47	86.95	85.71	0.58	0.60	0.59	0.010	24.41
Mozambique	85.29	85.68	85.49	0.55	0.55	0.55	0.074	3.10
Mean	**86.40**	**88.30**		**0.57**	**0.58**			

[1]Allelic richness was computed using the rarefaction method as described by El Mousadik and Petit (1996). F_{ST} and gene flow (*Nm*) estimates compared local and elite cassava genotypes for each national breeding programme. Loc. = local genotypes; Eli. = elite genotypes.

Table 4.4 Trends in total number of alleles sampled across 26 SSR loci in cassava germplasm without compensation for varying sample sizes

Country	Total number of alleles	
	Elite	Local
Tanzania	106	143
Uganda	125	123
Kenya	134	114
Rwanda	122	117
DRC	138	140
Madagascar	114	123
Mozambique	97	121
Mean	**119.4**	**125.8**

Allele frequency distribution in the studied germplasm is presented in Appendix 2 and 3. Locus SSRY38 had the highest frequency of the major allele across the analysed germplasm with a value of 0.772 in cassava from Madagascar and 0.9268 in cassava from Mozambique (Appendix 2). On the other hand, locus SSRY100 had the frequency of the major allele ranging from 0.2432 in germplasm from Madagascar to 0.2953 in germplasm from Kenya (Appendix 2). Generally, alleles with frequencies ranging between 0.02 and 0.09 were most prevalent in all studied cassava germplasm (Appendix 3). Germplasm from Madagascar had equal distribution of alleles with frequencies of ≤ 0.01 and those with frequecies ranging between 0.02 and 0.09. Alleles with frequencies > 0.70 were generally few (Appendix 3).

A comparison of allele frequencies is illustrated with data from eight loci (Figure 4.1). The comparison was made between four loci (SSRY102, SSRY5, SSRY63 and SSRY110) which had $F_{ST} > 0.2$ and four loci (SSRY64, SSRY100, SSRY161 and SSRY171) with $F_{ST} < 0.03$. Loci with lower F_{ST} values generally had: 1) a relatively more even distribution of allele frequency than loci with higher F_{ST} values and 2) a higher number of alleles with frequency of less than 0.1% (Figure 4.1).

Highest gene flow, depicting highest number of individuals moving between local and elite cassava were observed in Madagascar (24.4) and the lowest in Mozambique (3.1). Other NARS genotypes generally had low levels of less than ten individuals migrating except Tanzania (Table 4.3). However, across the entire dataset low levels of gene flow (1.97) were observed (data not shown).

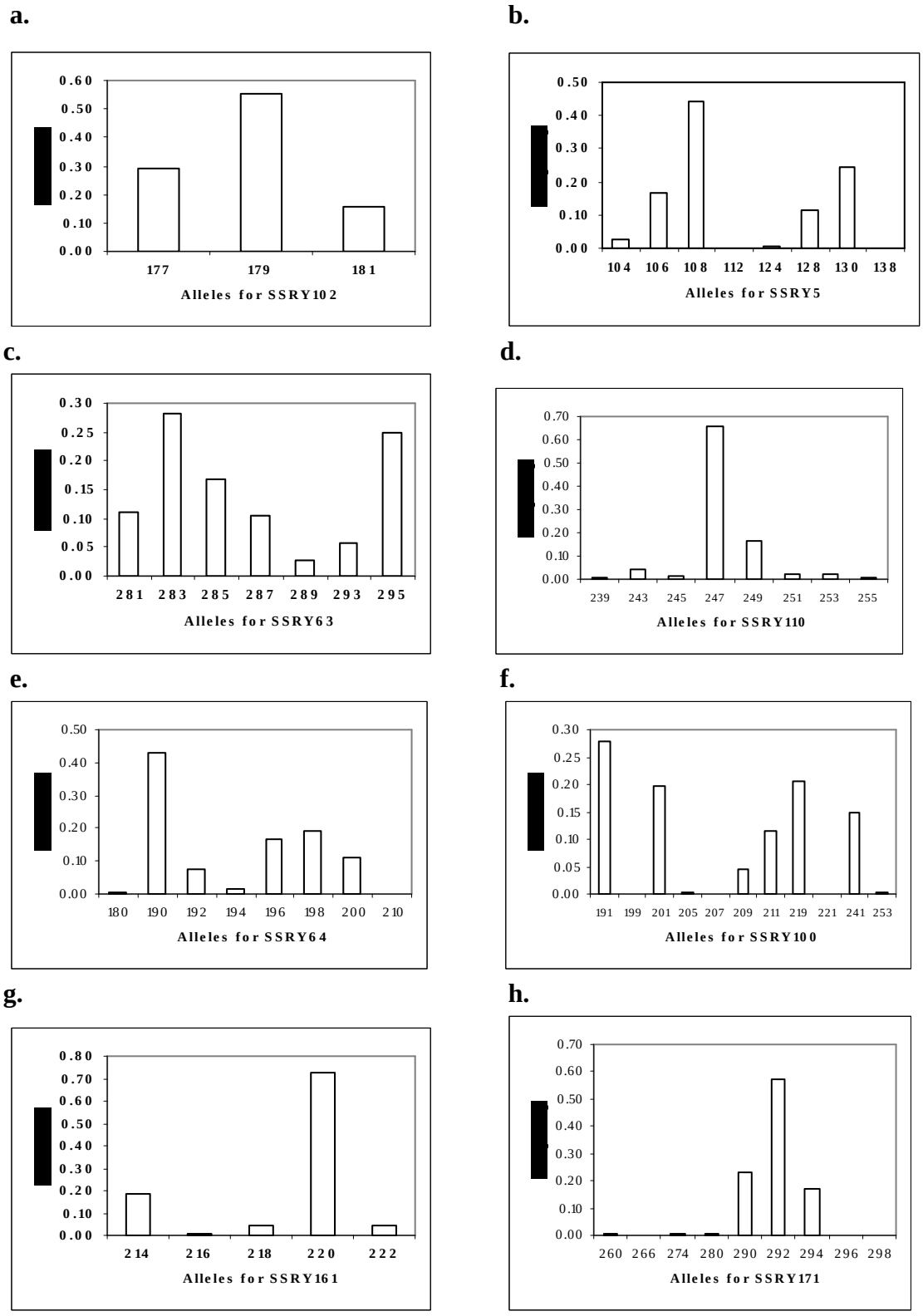


Figure 4.1 **A comparison of allele frequency distribution across cassava germplasm for eight loci: a) SSRY102, b) SSRY5, c) SSRY63 and d) SSRY110) with F_{ST} > 0.2 and e) SSRY64, f) SSRY100, g) SSRY161 and h) SSRY171) with F_{ST} < 0.03.**

4.3.2 Phenetic relationships

A phenetic display of the country specific local and elite genotypes is presented in Figure 4.2. The associated genetic distances are presented in Appendix 4. Three major groupings were noted. Group I comprised of Madagascar elite, Madagascar local, Mozambique local, Tanzania elite and Tanzania local genotypes. Group II comprised of DRC elite, Uganda elite, DRC local, Kenya elite, Rwanda elite, Mozambique elite, Uganda local and Rwanda local genotypes. Group III comprised exclusively of Kenyan local genotypes. Most of the elite genotypes clustered in group II (Figure 4.2). The elite and local genotypes from Madagascar and Tanzania, associated with relatively higher gene flow amongst themselves (Table 4.3), clustered together, while elite and local genotypes from Mozambique that had the lowest gene flow, were relatively distinct from each other.

When all elite genotypes from the seven countries were analysed as a single group and analysed together with local genotypes from each country, it was observed that most elite genotypes clustered together (Figure 4.3). It was also evident that a few local genotypes from Rwanda, DRC and Uganda clustered together with the elite genotypes (Figure 4.3). Phenetic analysis of only local genotypes established that most local genotypes from a specific country clustered together (Figure 4.4). In addition, it was evident that some local genotypes from Tanzania, Uganda and DRC clustered with local genotypes from the other countries (Figure 4.4).

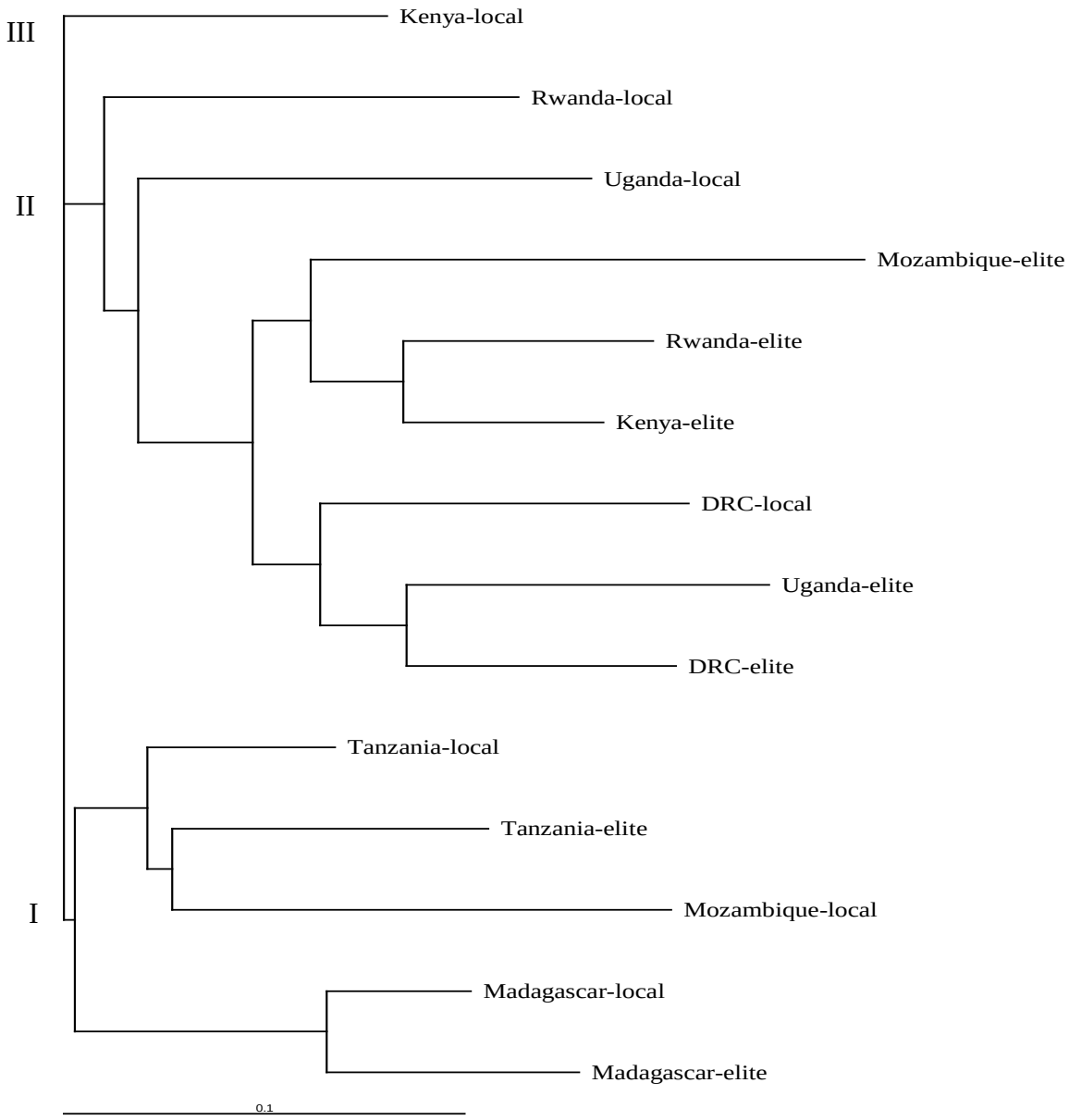


Figure 4.2 **Phenetic relationships of elite and local cassava genotypes from the NARS of Madagascar, Mozambique, Tanzania, DRC, Uganda, Rwanda and Kenya based on genotypic data at 26 SSR loci using Roger's genetic distance and neighbour-joining clustering in PowerMarker.**

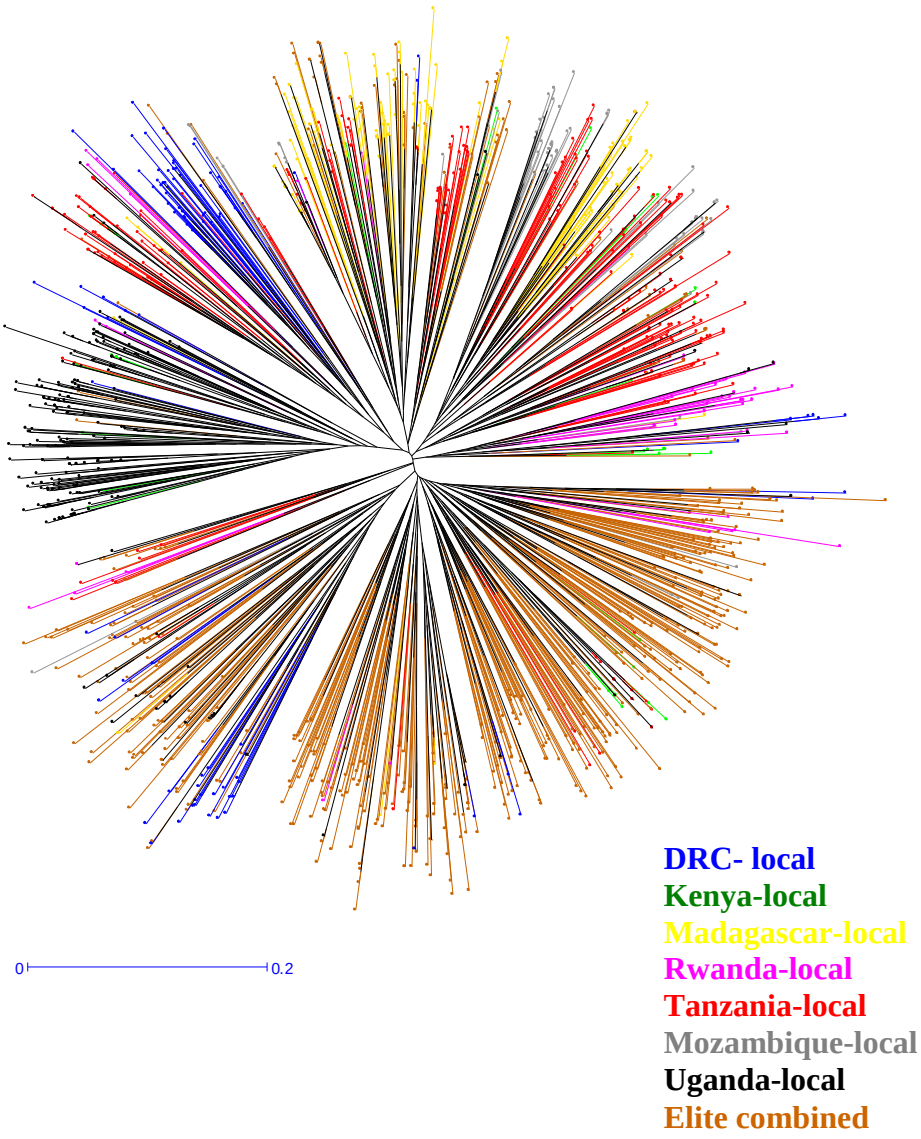


Figure 4.3 **A comparison of local cassava genotypes from the seven NARS (DRC, Kenya, Madagascar, Rwanda, Tanzania, Mozambique and Uganda) and combined elite genotypes based on genotypic data at 26 SSR loci using simple matching coefficient and the neighbour-joining clustering in DARwin.**

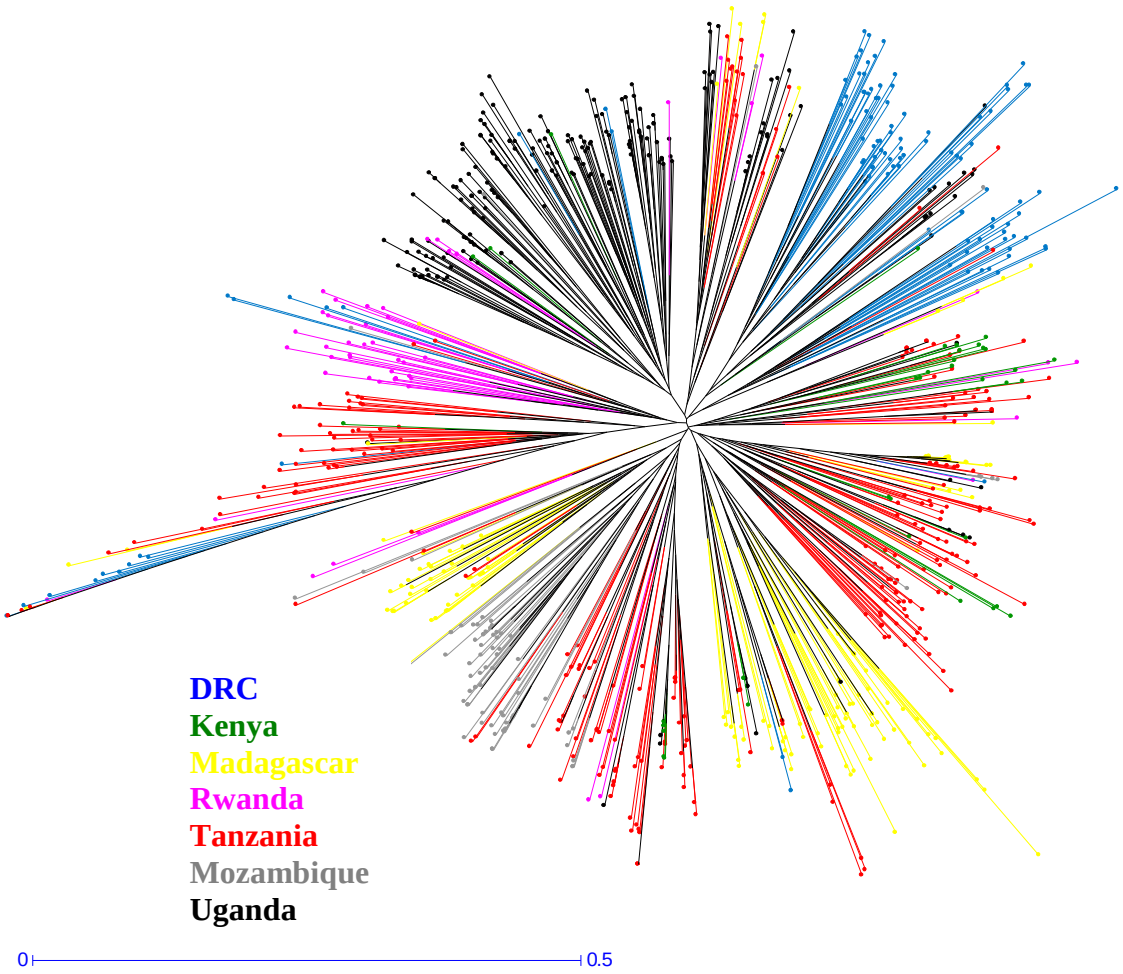


Figure 4.4 **A comparison of local cassava genotypes from the NARS of DRC, Kenya, Madagascar, Rwanda, Tanzania, Mozambique and Uganda based on genotypic data at 26 SSR loci using simple matching coefficient and the neighbour-joining clustering in DARwin.**

4.3.3 Genetic variance distribution

Three different AMOVAs were done. The first AMOVA was done with seven main groups for seven countries, with two subgroups per country (elite and local). Genetic variation varied among the hierarchy that comprised of groups (country of origin), among populations within groups (local and elite cassava genotypes) and within populations (individual genotypes) (Table 4.5). Most of the genetic variation was distributed between individuals (89.36%), with marginal variation quantified among countries (5.48%) and among populations within countries (5.15%).

The second AMOVA was done with eight instead of seven groups. The first seven groups represented the local genotypes from the seven countries, while the eighth group contained seven subgroups representing the elite genotypes from all seven countries. This new grouping was done based on the results obtained in Figure 4.3, where most elite varieties tended to group together. Again, most of the variation was distributed between individuals (89.33%) followed by among populations within groups (8.30%; between elite genotypes from seven different countries) and lastly among groups, which was only 2.35% (data not shown). The third AMOVA was done with only local genotypes (i.e. seven subgroups) and this was based on the results obtained in Figure 4.4. Again, most of the variation was between individuals (89.3%) compared to among groups (10.7%; among countries) (data not shown). It is therefore apparent that most of the genetic variance was distributed within individuals irrespective of whether elite and/or local genotypes were analysed.

Table 4.5 Analysis of molecular variation (AMOVA) based on 1401 elite and local cassava genotypes averaged over 26 SSR loci

Source of variation	Sum of squares	Variance component	Percentage variation
Among country of origin	1735.400	0.4662	5.48
Among populations[1] within countries	418.777	0.4379	5.15
Within populations	19755.803	7.5964	89.36
Total	**21909.98**	**8.500**	

[1]Populations represent either local or elite genotypes within each country

4.3.4 F-statistics and genetic differentiation

Data on F-statistics is presented in Table 4.6. Although 11 loci (SSRY100, SSRY12, SSRY181, SSRY182, SSRY52, SSRY69, SSRY135, SSRY151, SSRY169, SSRY171 and SSRY51) indicated an excess of heterozygosity within populations as depicted by the negative F_{IS} values, the overall estimate was positive (F_{IS} = 0.026). This finding indicated a slight deficit of heterozygotes within populations. The overall F_{IT} value (0.113) also indicated a slight deficit of heterozygotes in the total population.

No appreciable genetic differentiation was observed among cassava populations (groups) as depicted by the overall F_{ST} value of 0.089 (Table 4.6). However, four loci (SSRY102, SSRY5, SSRY63 and SSRY110) showed some gene differentiation among cassava populations with F_{ST} values > 0.2 (Table 4.6). When analysis was based on the 10 loci with F_{ST} values > 0.1 (SSRY102, SSY147, SSRY148, SSRY182, SSRY21, SSRY5, SSRY63, SSRY19, SSRY110 and SSRY169), gene differentiation increased slightly (overall F_{ST} = 0.157) (data not shown). The variance components associated with the F-statistics indicated that within individual sample variance was the highest (15.05), with marginal variances among populations (1.52) and between individuals within population variance (0.41) (Table 4.6).

4.3.5 Population structure

Though cassava does not have true defined populations, it was hypothesised that cassava genotypes in each country had acquired different allelic frequencies and hence formed hierarchical groups. Results from population structure analysis indicated that the genotyped 1401 individuals largely belonged to three populations i.e. K = 3 (Figure 4.5). Assignment of individuals to the three populations is presented in Table 4.7. Elite genotypes from DRC, Mozambique, Kenya and Rwanda had over 70% of their individuals assigned to population K = 1. Local cassava varieties from Madagascar, Tanzania, Mozambique and Kenya had over 70% of their individuals assigned to population K = 2 (Table 4.7).

These groupings (with the exception of local varieties from Kenya) are somewhat similar to that established in Figure 4.2. The latter is limited in robustness since it does not provide estimates of number of individuals assigned to a cluster.

Table 4.6 Estimates of F_{IT}, F_{ST}, F_{IS} and associated variance components across the 26 loci surveyed in 1401 cassava genotypes[1]

Loci	F_{IS}	F_{IT}	F_{ST}	Sig_a	Sig_b	Sig_w
NS911	0.006	0.045	0.040	0.017	0.002	0.400
SSRY100	-0.065	-0.037	0.026	0.021	-0.051	0.835
SSRY102	0.076	0.310	0.253	0.154	0.035	0.420
SSRY12	-0.061	-0.012	0.046	0.034	-0.042	0.732
SSRY147	0.387	0.473	0.139	0.069	0.166	0.262
SSRY148	0.024	0.149	0.128	0.095	0.015	0.627
SSRY161	0.151	0.169	0.021	0.009	0.065	0.363
SSRY181	-0.032	-0.003	0.028	0.015	-0.017	0.551
SSRY182	-0.001	0.117	0.118	0.093	-0.001	0.693
SSRY21	0.012	0.112	0.102	0.064	0.006	0.554
SSRY38	0.082	0.115	0.035	0.008	0.018	0.205
SSRY5	0.021	0.244	0.227	0.165	0.012	0.551
SSRY52	-0.023	0.022	0.044	0.031	-0.015	0.694
SSRY59	0.523	0.543	0.042	0.025	0.298	0.273
SSRY63	0.017	0.283	0.271	0.227	0.010	0.600
SSRY69	-0.052	0.000	0.050	0.040	-0.040	0.808
SSRY19	0.038	0.145	0.111	0.093	0.028	0.715
SSRY110	0.041	0.233	0.201	0.110	0.018	0.421
SSRY135	-0.039	-0.008	0.030	0.023	-0.029	0.769
SSRY151	-0.058	-0.004	0.051	0.042	-0.045	0.825
SSRY155	0.044	0.072	0.028	0.015	0.024	0.507
SSRY169	-0.154	-0.003	0.131	0.081	-0.083	0.620
SSRY171	-0.068	-0.038	0.027	0.016	-0.039	0.613
SSRY51	-0.017	0.023	0.039	0.029	-0.012	0.706
SSRY64	0.032	0.048	0.017	0.012	0.023	0.699
SSRY9	0.093	0.129	0.039	0.028	0.063	0.611
Overall	**0.026 ± 0.025**	**0.113 ±0.029**	**0.089 ± 0.016**	**1.516**	**0.409**	**15.05**

[1] F_{IS} = heterozygote deficit within populations; F_{IT} = global deficit of heterozygotes; F_{ST} = heterozygote deficit among populations (a measure of the Wahlund effect); Sig_a = the among sample variance component; Sig_b = the between individual within sample variance component; Sig_w = the within individual variance component

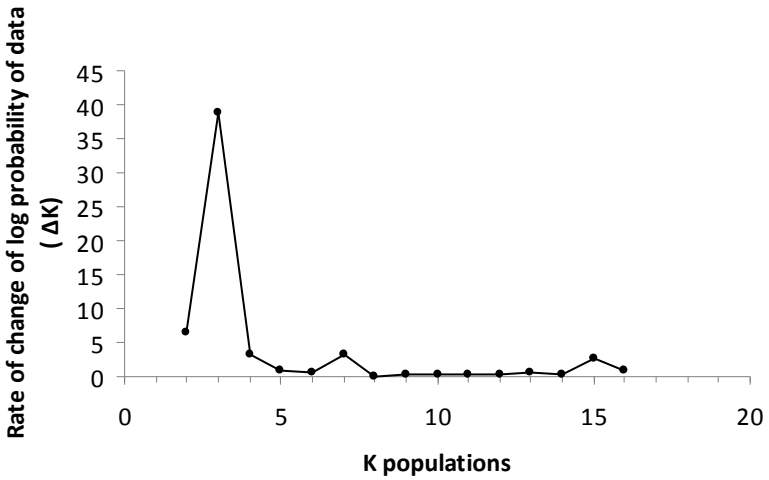


Figure 4.5 Determination of the *K* populations in the genotyped 1401 local and elite individuals following procedures of Evanno et al. (2005). Analysis was based on 10000 burn-in and MCMC replications for *K* = 1 to 17 and 10 replications per run.

Over 80% of individuals in Uganda local germplasm were assigned to population *K* = 3 (Table 4.7). This made it the only local group clustering with over 70% of the individuals in *K* = 3. Local genotypes from Uganda also formed a somewhat single cluster in Figure 4.2. The Ugandan elite genotypes were mainly distributed in two populations, *K* = 3 (56.9%) and *K* = 1 (39.9%) (Table 4.7). A relatively similar number of individuals from the local cassava genotypes from Rwanda were assigned between populations *K* = 1 and *K* = 2, while local genotypes from the DRC, were assigned between populations *K* = 1 and *K* = 3 (Table 4.7).

Table 4.7 Assignment of the 1401 cassava genotypes to three populations as revealed by structure analysis

Presumed population structure	Proportion of individuals assigned			Number of individuals
	***K* = 1**	***K* = 2**	***K* = 3**	
Uganda elite	0.399	0.032	**0.569**	65
Uganda local	0.057	0.070	**0.873**	203
DRC elite	**0.756**	0.051	0.192	73
DRC local	**0.418**	0.155	**0.427**	104
Madagascar elite	0.107	**0.857**	0.037	52
Madagascar local	0.057	**0.900**	0.043	134
Tanzania elite	0.093	**0.855**	0.052	21
Tanzania local	0.064	**0.881**	0.055	249
Mozambique elite	**0.716**	0.211	0.073	9
Mozambique local	0.058	**0.894**	0.048	73
Kenya elite	**0.833**	0.131	0.036	197
Kenya local	0.159	**0.747**	0.093	37
Rwanda elite	**0.927**	0.054	0.019	137
Rwanda local	0.432	**0.501**	0.068	47

When genotypic data of only the local varieties were analysed (10000 burn-in and MCMC replicationss) to have seven assumed populations, individuals were assigned to two populations i.e. K = 2 (Figure 4.6), which is the minimum observed number of populations using the procedures of Evanno et al. (2005). It is thus possible that the local varieties could all have belonged to one population. Local genotypes from Uganda and the DRC had 84% and 60% of their individuals, respectively assigned to population K = 1, while local genotypes from Madagascar, Tanzania, Mozambique, Kenya and Rwanda had 66-77% of their individuals assigned to population K = 2 (data not shown). The bar plots that show assignment of all genotypes to three populations and the assignment of local genotypes to two populations, are respectively presented in Figures 4.7 and 4.8.

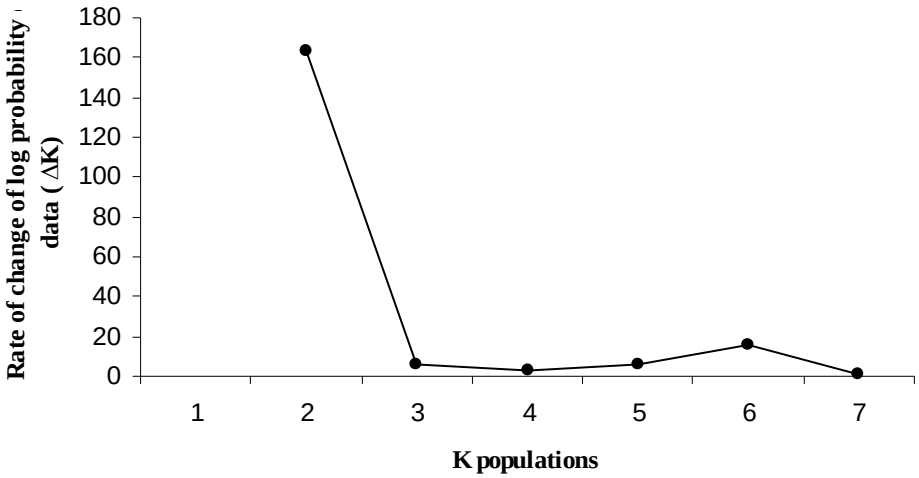


Figure 4.6 **Determination of the *K* populations in the genotyped local genotypes following procedures of Evanno et al. (2005). Analysis was based on 10000 burn-in and MCMC replications for *K* = 1 to 7 and 10 replications per run.**

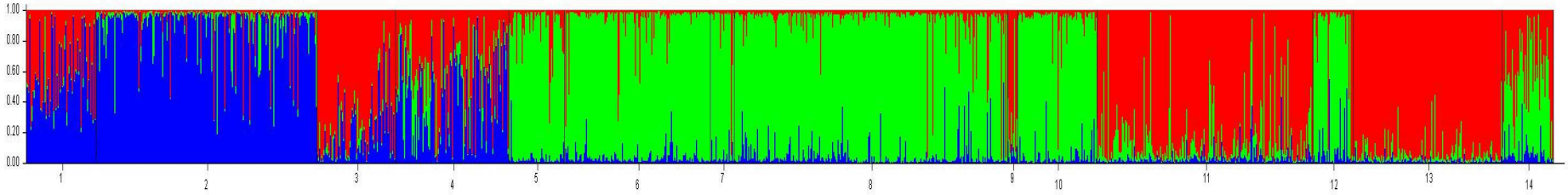


Figure 4.7 **Assignment of 1401 individuals (local and elite) to three populations (*K* = 1, red; *K* = 2, green; *K* = 3 blue) using the admixture model of population structure: 1 = Uganda elite; 2 = Uganda local; 3 = DRC elite; 4 = DRC local; 5 = Madagascar elite; 6 = Madagascar local; 7 = Tanzania elite; 8 = Tanzania local; 9 = Mozambique elite; 10 = Mozambique local; 11= Kenya elite; 12 = Kenya local; 13 = Rwanda elite; 14 = Rwanda local.**

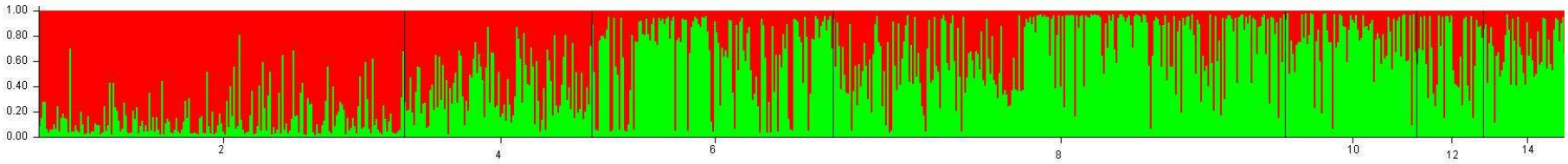


Figure 4.8 **Assignment of 847 local genotypes to two populations (*K* = 1, red; *K* = 2, green) using the admixture model of population structure: 2 = Uganda local; 4 = DRC local; 6 = Madagascar local; 8 = Tanzania local; 10 = Mozambique local; 12 = Kenya local; 14 = Rwanda local.**

4.3.6 Allele sharing between popular local and elite varieties

Data on the comparison of genetic profiles of popular local varieties grown in each of the NARS with 15 elite genotypes from IITA is presented in Table 4.8. Local varieties Gitamis (Rwanda), Nikwaha (Mozambique) and Liongo Kwimba (Tanzania), had genetic profiles that were different (d_R > 0.6) from the 15 elite cassava varieties from IITA, indicating their distinctiveness and different ancestral backgrounds. However, the local varieties Nakasoga (Madagascar), Kibadameno (Kenya), Nyaraboke (Uganda) and Boma (DRC) had genetic profiles that were moderately similar (d_R = 0.47 to 0.55) to those of IITA varieties (Table 4.8), which could point to some degree of shared ancestry. The IITA elite clones had genetic profiles that were moderately similar, with d_R ranging between 0.41 for genotype I91/2327 to 0.5 for genotype MM01/0014, while for local varieties d_R ranged from 0.48 for genotype Nakasoga to 0.71 for Gitamis (data not shown). Gitamis and Nikwaha were different (d_R > 0.7) from all other local varieties.

Table 4.8 A comparison of genetic profiles based on Roger's genetic distance of selected IITA elite lines and popularly grown local varieties of the seven countries

IITA elite lines	Popular local varieties[1]						
	Git.	**Nik.**	**Lio.**	**Nak.**	**Kib.**	**Nya.**	**Bom.**
9200057	0.825	0.785	0.573	0.466	0.401	0.466	0.464
95/SE00036	0.832	0.792	0.563	0.443	0.492	0.506	0.482
I91/2327	0.826	0.797	0.658	0.544	0.448	0.494	0.411
I92/00067	0.824	0.822	0.639	0.465	0.501	0.425	0.370
MH972961	0.793	0.805	0.607	0.546	0.505	0.442	0.406
MM010014	0.783	0.809	0.603	0.526	0.605	0.560	0.529
MM02/0993	0.783	0.814	0.605	0.599	0.514	0.593	0.472
MM95/0104	0.825	0.795	0.605	0.568	0.477	0.426	0.534
MM96/1419	0.800	0.821	0.628	0.634	0.576	0.539	0.487
MM96/7688	0.828	0.819	0.637	0.463	0.606	0.473	0.508
NASE 1	0.802	0.769	0.538	0.547	0.451	0.517	0.504
NASE 12	0.804	0.870	0.731	0.701	0.554	0.572	0.436
NASE 4	0.867	0.821	0.605	0.592	0.443	0.545	0.420
TMS 30572	0.854	0.828	0.748	0.593	0.596	0.554	0.503
TME 14	0.792	0.797	0.600	0.534	0.466	0.506	0.459
Mean	**0.816**	**0.810**	**0.623**	**0.548**	**0.509**	**0.508**	**0.466**

[1]Local varieties commonly grown in the seven sub-Saharan African countries: Rwanda (Git. = Gitamisi); Mozambique (Nik. = Nikwaha); Tanzania (Lio. = Liongo Kwimba); Madagascar (Nak. = Nakasoga); Kenya (Kib. = Kibandameno); Uganda (Nya. = Nyaraboke) and Democratic Republic of Congo (Bom.= Boma).

4.4 Discussion

A major objective of this research was to study the extent of genetic variation in cassava available within the cassava breeding programmes of Tanzania, Uganda, Kenya, Rwanda, DRC, Madagascar and Mozambique and ascertain whether or not it has a hierarchical organisation. It was hypothesised that the two independent introductions of cassava into the east African region (Jones, 1959) and the contrasting farmer-selection needs across the region, could in part, have contributed to shaping current day cassava into distinct groups. For this purpose, selectively neutral microsatellites which are suited for understanding population structure or history were used (Toro et al., 2009). It is important to note that genetic models used for analysis are based on several assumptions, including random mating and sexual reproduction (Weir, 1990). Data presented in this research chapter was generated from representative samples of local and elite cassava varieties, which at one stage of their development, involved recombination.

Furthermore, in populations, both genetic and statistical sampling is likely to influence quantification of genetic variation (Weir, 1990). In this study a representative sample from each country was selected and quantified for allelic richness using the rarefaction method (El Mousadik and Petit, 1996; Foulley and Ollivier, 2006). Data on allelic richness and total number of alleles observed across samples presented a somewhat different picture on genetic variability. Because different numbers of genotypes were collected across countries and comparisons were made between countries, allelic richness was preferred because it compensates for differences in sample size. The discussion hereafter is thus limited to allelic richness and not total number of alleles.

Allelic richness, defined as alleles segregating in the population, varied among cassava samples from the seven countries, with the highest number of segregating alleles being registered in the DRC and lowest in Uganda. This finding suggested that cassava from the DRC is genetically more diverse than cassava from the other six surveyed countries.

It was observed that elite cassava genotypes on average had a relatively higher number of alleles than local cassava genotypes. So often, local cassava varieties are cautiously and/or noncautiously generated and spread within a distinct geographical zone, primarily by farmers. Mating between relatives is thus possible, which could consequently reduce genetic variability due to increased homozygosity. The presence of volunteer seedlings has been observed in cassava fields (Balyejusa Kizito, 2006). Contrastingly, elite varieties generally result from controlled hybridisation, sometimes with diverse exotic parents, providing high levels of genetic variability. This could therefore explain differences observed in allelic richness between local and elite genotypes from the different NARS.

This premise however, does not hold true for all scenarios, as observed by higher allelic richness in local varieties from Rwanda and Tanzania, which could point to outbreeding rather than inbreeding. Similar observations have been noted in Guyana where local varieties grown at one location had the same level of genetic diversity (Shannon index = 4.293) as a sample from the core collection at CIAT (Shannon index = 4.289) (Elias et al., 2000). In an analysis of 31 elite IITA cassava varieties and five Nigerian local varieties, it was observed that elite lines had higher levels of diversity than the local lines, which the authors attributed to a narrow genetic base of local varieties (Moyib et al., 2007). In that study, despite the skewed sample size (31 elite versus five local), no correction for differences in sample size was made, leading to largely biased results. Nonetheless, their findings are consistent with data presented in this study.

In a related study, involving the analysis of 195 cassava genotypes from Uganda and 19 from Tanzania across 35 loci, it was observed that Ugandan cassava had a total of 183 alleles compared to 137 alleles for Tanzanian cassava. This finding could in itself indicate higher diversity in cassava from Uganda than Tanzania, though no appreciable differentiation was observed (F_{ST} = 0.051) (Balyejusa Kizito et al., 2005). In the present study where differences in sample size have been corrected for, it was noted that cassava from Tanzania (270 genotypes) and Uganda (268 genotypes) had relatively similar levels of diversity as reflected by allelic richness and gene diversity across 26 loci.

It is thus important that correction for sample size should always be done when comparing samples of varying census. Foulley and Ollivier (2006) have noted that the often reported number of alleles per loci in a population are not comparable because of differences in sample sizes and hence recommended specific analysis to rectify this limitation including either the rarefaction or the extrapolation methods.

Despite some differences in allelic diversity and the observed clustering, most variation was detected within individuals, which accounted for 89% of the total genetic variation. This value remained the same irrespective of whether local and elite genotypes were analysed together and/or separately. The inherent heterozygous nature of cassava coupled with formal breeding schemes that involve hybridisation to generate new recombinants could explain this phenomenon. A single cross in cassava that involves two heterozygous parents can generate F_1 recombinant individuals that display considerable variation. Since most of the NARS that participated in this research are involved in continuous generation and/or evaluation of introduced F_1 individuals, it is inevitable to have high genetic variability amongst individuals. Even on the island of Madagascar, where theoretically reduced genetic diversity would be expected, genetic diversity was high both in local and elite clones, despite the relatively high gene flow rate (Nm = 24.4). The higher variability within individuals as opposed to between groups (countries of origin), significantly limited the establishment of structured populations.

However, it does suffice to note that cassava is a recently diverged crop that was introduced onto the African continent from Latin America (Jones, 1959). As a species is brought into domestication, founder effects (Ladizinsky, 1985), population bottlenecks (Eyre-Walker et al., 1998) and selection (Buckler and Thornsberry, 2002) are all expected to reduce genetic diversity relative to the wild progenitor. It is therefore possible that the cassava founder introductions did not have distinct genetic variability and hence, the subsequent trans-border movements that happened in the region, transferred relatively similar allelic combinations leading to most variation being within individuals as opposed to between countries of origin (groups).

Moreover, the vegetative nature of cassava causes the genotypes to be fixed and hence specific allelic combinations can be easily transferred by humans. This is supported by the observed phenetic relationships which depicted: 1) the combined clustering of some local genotypes from Tanzania, Madagascar and Mozambique and 2) the distinct clustering of some elite and local genotypes. The presence of a regional cassava breeding programme in Tanzania in the 1930s (Nichols, 1947), that disseminated cassava varieties (from the same gene pool) to neighbouring countries is another possible explanation for the observed clustering and distribution of genetic variation. Bonnin et al. (1996) working on *Medicago truncatula* observed that even in the absence of gene flow, lack of structure can result when populations have a common origin.

No appreciable gene differentiation (F_{ST} = 0.089) was observed. However, it is important to note that classical F-statistics used in estimation of population differentiation are influenced by both genetic and statistical sampling procedures (Mohammadi and Prasanna, 2003). For instance, markers derived from coding versus non-coding regions of the genome infer different estimates and interpretations of population differentiation (Karl and Avise, 1992; Pogson et al., 1995). The nature of inheritance of DNA markers also influence the estimates, as noted by different differentiation estimates from microsatellites and diversity array technology (DArT) markers in cassava (Hurtado et al., 2008). In this research however, the selectively neutral microsatellites were used, from which no appreciable differentiation was observed. No comparisons were, however, made in this study with DNA markers targeting coding regions.

Under-dominance and local adaptation serve to inflate genetic differentiation, while over-dominance and spatially uniform selection tend to reduce genetic differentiation (Whitlock and McCauley, 1999). Furthermore, when selection is strict enough, it can result in substantial increases in genetic differentiation (Charlesworth et al., 1997). Indications of selection have however, been observed for some cassava genes (Kawuki et al., 2009; results Chapter 5). It is therefore possible that some genetic differentiation does exist in cassava, despite having a low overall estimate (F_{ST} = 0.089) based on the 26 selectively neutral microsatellites.

Indeed, when analysis was based on 10 of the 26 loci (SSRY102, SSRY5, SSRY63, SSRY110, SSRY169, SSRY19, SSRY21, SRRY182, SSRY148 and SSY147), which were selected based on F_{ST} values > 0.1, a slight increase in gene differentiation was observed among populations (overall F_{ST} = 0.161). These loci depicted a skewed distribution of alleles across the entire population, a finding which could suggest that evolutionary forces, notably random genetic drift and/or selection that shape gene differentiation, could within limits, be underway within some parts of the cassava genome, despite the short domestication history of the crop. This finding is consistent with studies conducted in *M. truncatula* where of the 22 loci studied, only two loci showed strong differentiation among populations (Bonnin et al., 1996).

Balyejusa Kizito et al. (2005), though working with fewer cassava genotypes (ranging between 19-195), observed no appreciable gene differentiation among cassava from Uganda (east Africa) and Ghana (west Africa) (F_{ST} = 0.04) and between cassava from Ghana and Tanzania (east Africa) (F_{ST} = 0.05), a finding which is somewhat consistent with the present study. Peroni et al. (2007) analysed sweet (76 entries) and bitter cassava (60 entries) and observed higher genetic diversity within populations (higher within sweet cassava), but with no appreciable genetic differentiation (R_{ST} = 0.057) among the two populations. Several other studies using neutral microsatellites observed no appreciable gene differentiation within different cassava populations (Asante and Offei, 2003; Lokko et al., 2006; Siqueira et al., 2009).

However, compared to the wild taxa comprising *M. esculenta* ssp. *flabellifolia* and *M. pruinosa,* genetic differentiation based on microsatellite data with respective F_{ST} values of 0.42 and 0.28 had been observed (Olsen and Schaal, 2001). For other out-crossing plants like pitch pine, F_{ST} values ranging from 0.003 to 0.069, with a mean of 0.024 across all 21 allozyme loci were observed, indicating no appreciable genetic differentiation among populations (Guries and Ledig, 1982). Bonnin et al. (1996), when studying inbreeding populations of *Medicago truncatula* with 22 DNA markers, noted a large within population variation (69.9%) compared to among populations (30.1%), which the authors attributed to out-crossing and/or migration events.

With the exception of the wild relatives, the low estimates of gene differentiation reported in this study are consistent with previous studies in both out-crossing and inbreeding species highlightened above.

The F_{IS} statistic measures the extent of nonrandom mating within populations i.e. inbreeding of individuals relative to the subpopulation to which they belong (Hartl and Clark, 1989). Values that are negative or zero, indicate an excess of heterozyogotes within the population, while values greater than zero, indicate possible inbreeding within populations. Averaged over the entire dataset F_{IS} was positive (0.026), indicating heterozygosity deficit within populations. Because some farmers have been reported to make selections from volunteer seedlings (Lebot et al., 2005; Balyejusa Kizito, 2006), it is possible that partial inbred lines resulting from spontaneous inbreeding (mating of relatives) could have been selected and fixed within localities and hence the slight heterozygosity deficit.

Besides inbreeding, high frequencies of null alleles (Brookfield, 1996) and unrecognised genetic structure in the form of fine-scale differentiation within populations (Gibbs et al., 1997) have been reported to cause heterozygosity deficits within populations. However, positive confirmation of null alleles will require direct sequence analysis of the genomic region where null alleles are hypothesised to be present. On the other hand, unrecognised genetic variation within populations can be examined at a nucleotide level to establish whether or not signatures of selection have occurred. This was, however, beyond the scope of this study. Nonetheless, the estimate on the frequency of null alleles was computed using the method of Brookfield (1996), which assumes that null homozygotes are present in the sample. The estimated frequency of null alleles was low (0.04), suggesting that inbreeding could be a major cause of the heterozygosity deficit. Heterozygosity deficits have been reported for both cultivated and wild relatives of cassava (Olsen and Schaal, 2001). The authors suggested that both the cassava inflorescence architecture and seed dispersal mechanisms limit long-distance gene flow, but favour mating among relatives, which in essence reduces heterozygosity. This reasoning is certainly applicable to the cassava analysed in this study.

When the robust Bayesian model-based analysis was done, the 1401 local and elite individuals were assigned into three distinct populations (groups). Most elite genotypes from DRC, Mozambique, Kenya and Rwanda were assigned to the same population indicating a shared ancestry. The cassava breeding programmes of these countries are routinely involved in the evaluation and selection of elite genotypes developed by IITA and hence mainly originate from the same gene pool.

Most elite and local genotypes from Madagascar and Tanzania together with local genotypes from Mozambique, Kenya and Rwanda were assigned to the same population, indicating shared ancestry. Screening of cultivars and inter-specific hybrids in both Tanzania and Madagascar in the 1930s and 1940s (Jennings, 1994) began the advent of cassava breeding in Africa, which could in part explain this observed clustering. The cassava breeding programme in Tanzania also disseminated cassava varieties from the same gene pool to neighbouring countries (Nichols, 1947). Based on the results, it is possible that some elite varieties may be hybrids from elite x local varieties and/or germplasm indicated as local may in fact be a local x elite cross. Structure analysis provides a robust and definitive picture of a population. The finding of only three populations further supported the observed limited gene differentiation observed in the cassava samples available within the NARS breeding programmes.

Genetic profiles of IITA elite cassava varieties and popular local varieties grown in each of the countries established that: 1) the popular local varieties grown in Rwanda (Gitamisi), Mozambique (Nikwaha) and Tanzania (Liongo Kwimba) did not have a shared ancestry history with the IITA elite varieties; 2) the popular local varieties grown in Uganda (Nyaraboke), Madagascar (Nakasoga), Kenya (Kibadameno) and DRC (Boma) shared almost half of their pedigree with the IITA elite varieties and 3) the elite varieties from IITA had a somewhat similar pedigree, perhaps originating from the same gene pool. This finding could therefore indicate limited utilisation of local varieties in the development of elite cassava varieties. This is further supported by the different grouping and/or population assignment of local and elite genotypes as specifically observed in germplasm from Kenya and Mozambique.

Local varieties have specific allelic combinations that have been fixed following several recombination and selection events. Through both formal and informal interactions with farmers it is increasingly becoming apparent that farmers prefer local varieties owing to their excellent culinary properties, which are not matched by most of the elite cassava varieties. This situation has, in turn, resulted in limited adoption of elite varieties, which in itself, is an unfortunate development. Maccafferi et al. (2003) examined the contribution of three native land races to the genetic diversity of 45 modern durum wheat varieties. Low levels of similarity were observed, which the authors suggested to be due to limited inheritance of alleles found in the ancient germplasm. Their study established that a substantial amount of total genetic variability present in the modern durum germplasm (92.7%) could be accounted for by alleles introduced from a limited number of founders. Thus, for cassava, a similar approach can be adopted to tap into the allelic diversity present in the local varieties by crossing them with elite parental lines.

4.5 Conclusions

Results from this chapter highlighted two important points. Firstly, cassava germplasm available within the NARS of Tanzania, Kenya, Uganda, DRC, Madagascar, Mozambique and Rwanda contained appreciable genetic variation. This variation, based on selectively neutral microsatellite alleles, was distributed in a manner which suggested that little differentiation has occurred. The majority of variation was detected amongst individuals, which is a consequence of clonal propagation that maintains variation among individuals. This is the first methodical assessment of genetic variation in the NARS of DRC, Madagascar and Rwanda. Information generated should be useful in setting conservation priorities in all the NARS included in this study.

Toro et al. (2009) proposed that the best way of assessing the genetic value of a breed (population) is to consider how much diversity it adds to a core set constituted of commercial lines or breeds that are already under successful conservation. This sounds like a reasonable strategy to replicate for cassava. A major limitation, however, is the lack of national cassava conservation strategies in the NARS.

Data generated from this study could therefore provide a starting point towards the establishment of functional national conservation programmes. By then, the maximum-diversity strategy should be adopted to guarantee cassava's plasticity to environmental extremes, biotic stresses, nutritional and culinary qualities.

Secondly, with the exception of germplasm from Tanzania and Madagascar, the genetic profiles of sampled local varieties appeared somewhat different from the elite cassava varieties, a finding which could indicate limited inheritance of alleles from local varieties while generating new elite cassava varieties. This therefore calls for systematic inclusion of local varieties into the hybridisation schemes. Through this course of action, basic and important information on the combining ability of local varieties will be established. Furthermore, this initiative will jumpstart the current need to breed for locally adapted cassava varieties that combine both preferred culinary quality (sourced from local varieties) and optimal yield (sourced from elite varieties).

4.6 References

Allard, R.W., 1988. Genetic changes associated with the evolution of adaptedness in cultivated plants and their wild progenitors. The Journal of Heredity 79:225-238.

Anonymous, 2001. The cassava molecular diversity network (MOLCAS). http://isa.ciat.cgiar.org/molcas October 2009.

Asante, I.K., and S.K. Offei, 2003. RAPD-based genetic diversity study of fifty cassava (*Manihot esculenta* Crantz) genotypes. Euphytica 131:113-119.

Balyejusa Kizito, E., 2006. Genetic and root growth studies in cassava (*Manihot esculenta* Crantz): implications for breeding. PhD Thesis, Swedish University of Agricultural Sciences, Uppsala, Sweden, pp. 127.

Balyejusa Kizito, E., A. Bua, M. Fregene, T. Egwang, U. Gullberg, and A. Westerbergh, 2005. The effect of cassava mosaic disease on the genetic diversity of cassava in Uganda. Euphytica 146:45-54.

Benesi, I.R.M., 2005. Characterisation of Malawian cassava germplasm for diversity, starch extraction and its native and modified properties. PhD Thesis, University of the Free State, South Africa., pp. 199.

Bonnin, I., J. Prosperi, and I. Olivieri, 1996. Genetic markers and quantitative genetic variation in *Medicago truncatula* (Leguminosae): A comparative analysis of population structure. Genetics 143:1795-1805.

Brookfield, J.F.Y., 1996. A simple new method for estimating null frequency from heterozygous deficiency. Molecular Ecology 5:453-455.

Buckler, E.S., and J.M. Thornsberry, 2002. Plant molecular diversity and applications to genomics. Current Opinion in Plant Biology 5:107-111.

Charlesworth, B., M. Nordborg, and D. Charlesworth, 1997. The effects of local selection, balanced polymorphism, and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genetical Research 70:155-174.

Colson, I., and D.B. Goldstein, 1999. Evidence for complex mutations at microsatellite loci in *Drosophila*. Genetics 152:617-627.

Dellaporta S.J., J. Wood, and J.B. Hicks, 1983. A plant DNA mini preparation. Version II. Plant Molecular Biology Reporter 1:19-21.

El Mousadik, A., and R.J. Petit, 1996. Chloroplast DNA phylogeography of the argan tree of Morocco. Molecular Ecology 5:547-555.

Elias M., O. Panaud, and T. Robert, 2000. Assessment of genetic variability in a traditional cassava (*Manihot esculenta* Crantz) farming system using AFLP markers. Heredity 85:219-230.

Evanno, G., S. Regnaut, and J. Goudet, 2005. Detecting the number of clusters of individuals using software STRUCTURE: a simulation study. Molecular Ecology 14:2611-2620.

Excoffier, L., G. Laval, and S. Schneider, 2005. Arlequin version 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics online 1:47-50.

Eyre-Walker, A., R.L. Gaut, H. Hilton, D.L. Feldman, and B.S. Gaut, 1998. Investigation of the bottleneck leading to the domestication of maize. Proceedings of the National Academy of Sciences (USA) 95:4441-4446.

Foulley, J.-L., and L. Ollivier, 2006. Estimating allelic richness and its diversity. Livestock Science 101:150-158.

Fregene, M., A. Bernal, M. Duque, A. Dixon, and J. Tohme, 2000. AFLP analysis of African cassava (*Manihot esculenta* Crantz) germplasm resistant to the cassava mosaic disease (CMD). Theoretical and Applied Genetics 100:678-685.

Gibbs, H.L., K.A. Prior, P.J. Weatherhead, and G. Johnson, 1997. Genetic structure of populations of the threatened eastern massasauge rattlesnake, *Sistrurus c. catenatus*: evidence from microsatellite DNA markers. Molecular Ecology 6:1123-1132.

Goldstein, D.B., and D.D. Pollock, 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. Journal of Heredity 88:335-342.

Goudet, J., 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html.

Guries, R.P., and F.T. Ledig, 1982. Genetic diversity and population structure in pitch pine (*Pinus rigida* Mill.). Evolution 23:387-402.

Hartl, D.L., and A.G. Clarke, 1989. Principles of population genetics. Second Edition. Sinauer Associates, Inc., Publishers, pp. 682.

Hershey, C., 2010. Cassava breeding: theory and practice. A publication by Food and Agriculture Organisation (FAO), in press.

Hill, W.G., and J. Rasbash, 1986. Models of long-term artificial selection in finite populations. Genetical Research 48:41-50.

Hurtado, P., K.M. Olsen, C. Buitrago, C. Ospina, J. Marin, M. Duque, C. de Vicente, P. Wongtiem, P. Wenzel, A. Killian, M. Adeleke, and M. Fregene, 2008. Comparison of simple sequence repeat (SSR) and diversity array technology (DArT) markers for assessing genetic diversity in cassava (*Manihot esculenta* Crantz). Plant Genetic Resources: Characterisation and Utilisation 6:208-214.

Jennings, D.L., 1994. Breeding for resistance to African cassava mosaic germiniviruses in East Africa. Tropical Science 34:110-122.

Jones, W.O., 1959. *Manioc* in Africa. Standard University Press, Stanford, California, pp 315.

Karl, S.A., and J.C. Avise, 1992. Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. Science 256:100-102.

Kawuki, R.S., M. Ferguson, M. Labuschagne, L. Herselman, and D.J. Kim, 2009. Identification, characterisation and application of single nucleotide polymorphisms (SNPs) in cassava (*Manihot esculenta* Crantz). Molecular Breeding 23:669-684.

Kimura, M., and J.T. Crow, 1964. Number of alleles that can be maintained in a finite population. Genetics 49:725-738.

Ladizinsky, G. 1985. Founder effect in crop-plant evolution. Economic Botany 39:191-199.

Langlands, B.W., 1966. Cassava in Uganda 1860-1920. Uganda Journal 30:211-218.

Laval, G., M. SanCristobal, and C. Chevalet, 2002. Measuring genetic distances between breeds: use of some distances in various short term evolution models. Genetics Selection and Evolution 34:481-507.

Lebot, V., A. Ivancic, and K. Abraham, 2005. The geographical distributions of allelic diversity, a practical means of preserving and using minor root crop genetic resources. Experimental Agriculture 41:475-489.

Liu, K., and S.V. Muse, 2005. PowerMarker. Integrated analysis environment for genetic marker data. Bioinformatics 21:2128-2129.

Lokko, Y., A. Dixon, S. Offei, E. Danquah, and M. Fregene, 2006. Assessment of genetic diversity among African cassava *Manihot esculenta* Crantz accessions resistant to the cassava mosaic virus disease using SSR markers. Genetic Resources and Crop Evolution 53:1441-1453.

Maccaferri, M., M.C. Sanguineti, P. Donini, and R. Tuberosa, 2003. Microsatellite analysis reveals a progressive widening of the genetic basis in the elite durum wheat germplasm. Theoretical and Applied Genetics 107:783-797.

Matsuoka, Y., S.E. Mitchell, S. Kresovich, M. Goodman, and J. Doebley, 2002. Microsatellites in Zea-variability, patterns of mutation, and use for evolutionary studies. Theoretical and Applied Genetics 104:436-450.

Mba, R.E.C., P. Stephenson, K. Edwards, S. Melzer, J. Nkumbira, U. Gullberg, K. Apel, M. Gale, J. Tohme, and M. Fregene, 2001. Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. Theoretical and Applied Genetics 102:21-31.

Melchinger, A.E., M.M. Messmer, M. Lee, W.L. Woodman, and K.R. Lamkey, 1991. Diversity and relationships among U.S. maize inbreds revealed by restriction fragment length polymorphism. Crop Science 31:669-678.

Mohammadi, S.A., and B.M. Prasanna, 2003. Analysis of genetic diversity in crop plants-Salient statistical tools and considerations. Crop Science 43:1235-1248.

Morgante, M., and A.M. Olivieri, 1993. PCR-amplified microsatellites as markers in plant genetics. Plant Journal 3:175-182.

Morgante, M., M. Hanafey, and W. Powell, 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genetics 30:194-200.

Moyib, O.K., O.A. Odunola, and A.G.O. Dixon, 2007. SSR markers reveal genetic variation between improved cassava cultivars and landraces within a collection of Nigerian cassava germplasm. African Journal of Biotechnology 6:2666-2674.

Nei, M., 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences (USA) 70:3321-3323.

Nei, M., 1987. Molecular evolutionary genetics. Columbia University Press, New York, pp. 512.

Nichols, R.F.W., 1947. Breeding cassava for virus resistance. East African Agricultural Journal 12:184-194.

Nweke, F., D. Spencer, and J. Lynam, 2002. The cassava transformation: Africa's best kept secret. Michigan State University Press, East Lansing, USA, pp. 273.

Ohta, T., and M. Kimura, 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genetic Research 22:201-204.

Okogbenin, E., J. Marin, and M. Fregene, 2006. An SSR-based molecular genetic map of cassava. Euphytica 147:433-440.

Olsen, K.M., and B.A. Schaal, 1999. Evidence of the origin of cassava: phylogeography of *Manihot esculenta*. Proceedings of the National Academy of Sciences (USA) 96:5586-5591.

Olsen K.M., and B.A. Schaal, 2001. Microsatellites variation in cassava (*Manihot esculenta*, Euphorbiaceae) and its wild relatives: further evidence for a southern Amazonian origin of domestication. American Journal of Botany 88:131-142.

Peroni, N., P.Y. Kageyama, and A. Begossi, 2007. Molecular differentiation, diversity, and folk classification of "sweet" and "bitter" cassava (*Manihot esculenta*) in Caicara and Caboclo management systems (Brazil). Genetic Resources and Crop Evolution 54:1333-1349.

Perrier, X., and J.P. Jacquemound-Collet, 2006. DARwin software. http://darwin.cirad.fr/darwin.

Pogson, G.H., K.A. Mesa, and R.G. Boutilier, 1995. Genetic population structure and gene flow in the Atlantic cod *Gadus morhua*: A comparison of allozyme and nuclear RFLP loci. Genetics 139:375-385.

Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski, 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding 2:225-238.

Pritchard, J.K., M. Stephens, and P. Donnelly, 2000. Inferences of population structure from multilocus genotypic data. Genetics 155: 945-959.

Reif, J.C., A.E. Melchinger, and M. Frisch, 2005. Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. Crop Science 45:1-7.

Siqueira, M.V.B.M., J.R. Queiroz-Silva, E.A. Bressan, A. Borges, J.C.K. Pereira, J.G. Pinto, and A.E. Veasey, 2009. Genetic characterisation of cassava (*Manihot esculenta*) landraces in Brazil assessed with simple sequence repeats. Genetics and Molecular Biology 32:104-110.

Tautz, D., and M. Renz, 1984. Simple sequences are ubiquitous components of the eukaryotic genomes. Nucleic Acids Research 12:4127-4138.

Toro A.A., J. Fernàndez, and A. Caballero, 2009. Molecular characterization of breeds and its use in conservation. Livestock Science 120:174-195.

Vellvẻ, R., 1993. The decline of diversity in European agriculture. Ecologist 23:64-69.

Weir, B.S., 1990. Genetic data analysis. Methods for discrete population genetic data. Sinauer Associates, Inc. Publishers, pp. 377.

Weir, B.S., and C.C. Cockerham, 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370.

Whitlock, M.C., and D.E. McCauley, 1999. Indirect measures of gene flow and migration: $F_{st} \neq 1/(4Nm+1)$. Heredity 82:117-125.

Yeh, F.C., Y. Rongcai, and T. Boyle, 1999. Popgene version 1.31. Microsoft windows-based free software for population genetic analysis. http:www.valberta.ca/fyeh.

Zacarias, A.M., A.M. Botha, M.T. Labuschagne, and I.R.M, Benesi, 2004. Characterisation and genetic distance analysis of cassava (*Manihot eculenta* Crantz) germplasm from Mozambique using RAPD fingerprinting. Euphytica 138:49-53.

CHAPTER 5

IDENTIFICATION, CHARACTERISATION AND APPLICATION OF SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs) FOR DIVERSITY ASSESSMENT IN CASSAVA (*Manihot esculenta* Crantz)[1]

5.1 Introduction

Genetic diversity in the field is important for maintaining ecological balance, particularly in low input agricultural systems in which farmer-varieties predominate. To monitor this diversity it is important that it can be measured accurately. At the farm or field level, varieties are likely to be relatively closely related with only a few highly heritable morphological differences that can be scored. Molecular markers allow the measurement of diversity at a larger number of loci that are not affected by the environment. At the moment SSR markers are the markers of choice for diversity assessments, although other marker systems are also in use such as AFLP, RFLP, RAPD and their derivatives. A limitation of many of these systems is the number of markers that are found to be polymorphic in closely related individuals. SNPs and small insertions and deletions (indels) represent the most frequent form of naturally occurring genetic variation in populations (Kruglyak, 1997) and may provide sufficient variation to discriminate between closely related individuals and thus become the marker of choice for dissecting diversity in the field.

In the context of population genetics, SNPs are single base pair positions at defined loci in genomic DNA at which different sequence alternatives (alleles) exist, with the rarer allele, in this case, being defined as more common than 1% in the population (Brookes, 1999). A single nucleotide variant that occurs at a frequency of 1% or less in a population is generally referred to as a point mutation.

[1] Molecular Breeding (2009) 23:669-684

Thus a T/C polymorphism defines two alleles for which there could be three genotypes including homozygous T-A, homozygous C-G and heterozygous with either T or A in one molecule and C or G at the corresponding site in the homologous chromosome (Hartl, 2000).

SNPs account for 90% of all human DNA polymorphisms (Collins et al., 1998). They have been shown to be mutationally stable and generally bi-allelic being scored co-dominantly (Cho et al., 1999; Lindblad-Toh et al., 2000; Gupta et al., 2001). These characteristics make them ideal for genetic analyses as well as amenable to automation. SNPs have been shown to be more frequent in intron regions (Rafalski, 2002a) and tend to associate as robust intragene haplotypes (Schneider et al., 2001), which is of relevance to diversity assessment. SNPs have been used in a few studies to assess diversity (grapevine, Salmaso et al., 2004) and haplotype structure (grapevine, Salmaso et al., 2004; *Beta vulgaris* L., Schneider et al., 2001) and have been applied to phylogeographic analyses (Brumfield et al., 2003). Olsen (2004) used SNP and indel variation in portions of two low copy nuclear genes and five SSR loci to infer the origin of cassava. SNPs are now the focus of large-scale genotyping projects in humans, model organisms and increasingly so in crop plants.

In humans, SNP frequencies are estimated at one per 1000 nucleotides when two homologous DNA sequences are compared (Wang et al., 1998), while in plants contrasting SNP frequencies have been reported between inbreeding and out-crossing species. For example, one SNP per 78 bp in grapevine (Salmaso et al., 2004); one SNP per 31 bp in non-coding regions and one SNP per 124 bp in the coding regions in maize (Ching et al., 2002) and one SNP per 191 bp in the non-coding regions and one SNP per 2038 bp in the coding regions in soybean (Van et al., 2005). In clonally propagated, predominantly out-crossing cassava, Lopez et al. (2005) found an overall frequency of one SNP per 62 bp in expressed sequence tags (ESTs), 3'-EST and bacterial artificial chromosomes (BAC) end sequences of a number of cultivars.

In some cases SNPs derived from ESTs correlate with non-conservative amino acid changes which may lead to phenotypic variation. This allows the association of genotypic and phenotypic variation (McCallum et al., 2000; Rafalski, 2002a). This strategy has been exploited in humans where SNPs have been associated with genetic diseases (Collins et al., 2004), including risks of developing cardiovascular diseases (Davignon et al., 1988) and Alzheimer's disease (Fullerton et al., 2000).

In plants, SNPs located within the putative *betaine aldehyde dehydrogenase-2 gene*, have been associated with the rice fragrance trait (Bradbury et al., 2005). Those within the *starch synthase IIa gene* have been associated with gelatinising temperatures (Bao et al., 2006) and those within the *waxy* gene (*Wx*) associated with amylose content in rice (Gupta et al., 2001). Not only are SNPs likely to find utilisation in crop improvement through genotype-phenotype associations, but in the construction of high resolution genetic maps, diagnostics, analysis of populations, cultivar identification and detection of linkage disequilibrium (Gut, 2001; Buckler and Thornsberry, 2002; Rafalski, 2002b; Neale and Savolainen, 2004). Indeed, SNP discovery and utilisation is advancing rapidly in a number of crops including maize (Tenaillon et al., 2001; Ching et al., 2002; Hamblin et al., 2007), rye (Varshney et al., 2007), barley (Kanazin et al., 2002), sugar beet (Schneider et al., 2001), rice (Bradbury et al., 2005) and soybean (Zhu et al., 2003; Van et al., 2005). These studies revealed the frequent occurrence of SNPs and the existence of intragene haplotypes.

In terms of detection, a broad range of SNP-based genotyping chemistries have been developed, each with their own advantages. Some of these include single strand conformation polymorphism (SSCP), heteroduplex analysis (HA), denaturing gradient gel electrophoresis (DGGE), RNase mismatch cleavage, dideoxy fingerprinting, molecular beacon, protein truncation test, restriction of PCR products, allele specific oligo hybridisation, primer extension, electronic SNP (eSNP), oligonucleotide microarray based analysis and direct sequencing of PCR products. In-depth reviews on most of these methodologies have been compiled (Gupta et al., 2001; Gut, 2001; Kwok, 2001).

Fundamentally, many of these methods involve either post-PCR enzymatic reactions or oligonucleotide hybridisation, with no single genotyping methodology ideally suited to all applications. In this study the direct sequencing approach (Rafalski, 2002a) was used to examine SNPs in selected cassava genes.

Previous hypotheses favoured a hybrid origin for cassava through interbreeding *Manihot* species (Rogers, 1963; Rogers and Appan, 1973). However, relatively recent evidence suggests cassava has a single wild progenitor, *M. esculenta* ssp. *flabellifolia* and that it originates from the southern Amazon basin (Olsen, 2004). Cassava is a monoecius and naturally out-crossing clonally propagated plant. Meiotic behaviour studies depict the 18 bivalent formation of chromosomes typical of its diploid (2n = 36) nature (Jos and Nair, 1979). Cassava is a staple crop for most subsistence communities in the tropics, with increasing utilisation in the industrial sector owing to its diversity in starch functionality including its potential use as a biofuel. Partly due to this emerging commercialisation, there is an increasing need to monitor its diversity at farm level. In this study, insights into the utility of SNP markers for measuring genetic diversity in cassava were provided. SNPs were identified and characterised at 26 loci from nine genes. Polymorphism were scored at these loci in 74 diverse cassava genotypes and considered in relation to SSR diversity at 12 loci. In addition the frequency and informativeness of SNPs were examined as well as the allelic haplotypes and nucleotide diversity.

5.2 Materials and methods

5.2.1 Plant materials

A highly diverse set of 74 cassava genotypes representing three geographical regions: South America (30 genotypes); Asia (four genotypes) and Africa (40 genotypes) were selected based on prior analysis of 2494 genotypes at 22 polymorphic SSR loci (Morag Ferguson, personal communication) (Table 5.1). Total genomic DNA samples for both South America and Asia cassava genotypes were kindly provided by CIAT, while the African DNA samples were provided by IITA. Both CIAT and IITA extracted DNA from young leaf samples using the miniprep protocol as described by Dellaporta et al. (1983).

Table 5.1 A list of cassava genotypes selected from Asia, America and Africa used in the study

Country/Region	Genotype	Country/Region	Genotype
Asia		**Africa**	
Thailand	TAI18	Guinea Conakry	TME 1239
Thailand	TAI1	Guinea Conakry	TME 1219
Thailand	MTAI 8	Guinea Conakry	TME 1286
Malaysia	MAL29	Guinea Conakry	TME 598
		Cape Verde	TME 125
America		Cape Verde	TME 135
Cuba	CUB5	Cape Verde	TME 131
Cuba	CUB70	Democratic Republic of Congo	TME 1353
Mexico	MEX41	Benin	TME 443
Mexico	MEX55	Kenya	TME 786
Argentina	ARG12	Cameroon	TME 214
Argentina	ARG74	Cameroon	TME 152
Paraguay	PAR23	Ghana	TME 1567
Paraguay	PAR81	Ghana	TME 1683
Costa Rica	CR19	Ghana	TME 1700
Costa Rica	CR65	Ghana	TME 1384
Venezuela	VEN164	Ghana	TME 633
Venezuela	VEN173	Ghana	TME 1368
Venezuela	VEN81	Nigeria	TME 287
Peru	PER196	Nigeria	TME 290
Peru	PER178	Nigeria	TME 1
Brazil	BRA1292	Nigeria	TME 34
Brazil	BRA1061	Nigeria	TME 14
Brazil	BRA192	Nigeria	TME 17
Brazil	BRA125	Nigeria	TME 5
Colombia	MCol1522	Nigeria	TME 342
Colombia	MCOL 2206	Nigeria	TME 1050
Colombia	C9-489-34	Nigeria	TME 9
Colombia	COL2532	Nigeria	TMS 30572
Colombia	CM2217-2	Nigeria	TME 117
Columbia	CM5306-8	Nigeria	TME 3
Columbia	CM3306-9	Tanzania	Albert
USA	USA7	Tanzania	TME 1311
Panama	PAN38	Tanzania	Namikonga
Ecuador	Ecu72	Tanzania	Kibaha
Guatemala	GUA59	Tanzania	Kalolo
		Togo	TME 232
		Togo	TME 396
		Uganda	TME 539
		Sierra Leone	TME 857

5.2.2 Candidate loci, primer design, PCR and sequencing reactions

To maximise SNP discovery, primers targeting the non-coding regions, including the 3' and 5' untranslated regions and the intron regions of the genome were designed (Rafalski, 2002b). Core nucleotides in the National Centre for Biotechnology Information (NCBI) database were surveyed to identify full or partial sequences of cassava genes. BLASTn was used to identify genes of low copy number, which are desirable for precise amplification. To make useful interpretations of the BLASTn output, frequency tallies for appearance of respective sequences in either rice or *Arabidopsis* were made considering sequences with a threshold statistical E-value of > e^{-20}. Based on the BLASTn results, 19 gene sequences were selected and primers designed using Primer3 (Rozen and Skaletsky, 2000). In some cases, more than one primer targeting a different part of the gene was designed. In total, 28 primer pairs were designed (Table 5.2). The selected genes are involved in starch metabolism, cyanogenesis, drought and/or stress responses.

Primers were optimised using the AccuPower^{TM} PCR Premix (1 U *Taq* polymerase, 0.25 mM dNTPs, 10 mM Tris-HCl, 40 mM KCl and 1.5 mM $MgCl_2$ following the manufacturer's recommendations (Bionexus Inc., USA) with 50 ng DNA and 0.16 pmole primer. This was denatured at 94^{o}C for 2 min followed by 30 cycles of denaturation at 94^{o}C for 1 min, annealing at 48^{o}-63^{o}C for 1 min and extension at 72^{o}C for 2.5 min and a final extension at 72^{o}C for 30 min. The annealing temperatures were established by running a temperature gradient PCR ranging from 42^{o}-63^{o}C using the Techne TC-512 thermocycler. Products of primers which produced single and clear PCR products were selected for sequencing. PCR products from 14 primer pairs that produced single amplicons were sequenced directly from 74 test genotypes (Table 5.2). Initially sequencing was done using both forward and reverse primers to identify primers that produced good quality sequences. All sequencing reactions were performed using the BIGDYE Terminator Cycle Sequencing system according to manufacturer's recommendations and visualised on the ABI 3730 (Applied Biosystems). Trace files were aligned using Sequencher 4.6 (Gene Codes Corporation, USA). Good quality sequences were used for SNP identification.

Table 5.2 Candidate gene selection, gene bank accession number, primer sequence, annealing temperature and PCR products

Primer [1]	Gene Bank accession No.	Forward sequence (5'-3')	Reverse sequence (5'-3')	Annealing Temp.	PCR product (bp)	PCR product
G3pdh A	AF136149	CGGGGCTGCCAAGGTAAG	TGAGATGCCAAAAGAAAAAGGAGT	57°C	350	Single
G3pdh B	AF136149	TACCGGTAAGCTCATCATTCGTC	CGCGCAGATCAAGTAACCAAAAC	50°C	-	Multiple
PAL2 B	AF383152	CATCCTGGGTGAGAAGTCGTAGC	TTCGCCCTGATCCAATAAAATGA	-	-	No product
PAL2 A	AF383152	CATGCCGGCCTGACGAGATA	TCAGCGGCAAGCAAGGAGTAAT	59°C	700	Single
CYP79D2 A	AF140614	CGTCTCCACCACCGCAACCA	TCGTCAAGTATCCGCCGCTCATT	59°C	400	Single
CYP79D2 B	AF140614	CCTTCAGGCTCCACCCAGTTG	CTCACATTCATCCCTTCCCATCTT	54°C	450	Single
GBSSII A	AF173900	GTGAGGGGATGGGTTGGTTTTTC	TGGCCCCCTTACTAGCATCAGAG	63°C	400	Single
GBSSII B	AF173900	CTTTGGCTCGGCGTGGAC	CCGACCCTGGTGAGCTATGTTATG	53°C	-	Multiple
MeEF1 A	AF041463	CTCCCGTGATTTAGATTTACTTC	AGCACCCAGGCATACTTGA	46°C	-	Multiple
MeEF1 B	AF041463	GTCTAGTACTGGGGTTGGTCTGGT	CTGGCATCCGTTTCATTAGG	49°C	500	Single
MeEF1 C	AF041463	ATACGCCAGGGTTTTCATAGGT	GAAATCGCAAAATCAACTCAACAA	50°C	400	Single
bglA A	X94986	TTCGCGTCTATCTATTTACTGTTA	ACTTCTTGGCAATGCTCACTTTA	59°C	-	Multiple
bglA B	X94986	CATTGCCAAGAAGTGAAAACC	ATCCGCAACGTCTCCAGTG	49°C	-	Multiple
pLIN A	U95298	ATAGTTGCCCATAATTTGCTTCTT	ATCTATCTCCGGCTAAATCTACCA	50°C	450	Single
pLIN B	U95298	AATGCGTTTAGAATGTCCA	ATCCCAATGAAAAATAGTGA	48°C	-	Multiple
HRGP A	AF239615	AACGGGATCATGGCGAACTGC	TATGGGGGAGGAGGAGATGGAGAC	-	-	No product
HRGP B	AF239615	AGTTTACATTTACGCCTCACCATA	AAACCAAACATAAACACCCACTTC	-	-	No product
CI	AY101376	ACAACCGAGGCTGCAACTGAGG	TACCCAAAACATCGCCATACTG	59°C	300	Single
TK	AY745770	ACGTGGATATACATGTGAAGAAGC	GTGGGGTAAAAGCGAAGAGAAA	59°C	250	Single
APX3	AY973623	GGTGCCACAAGGAACGCTCTG	CTTCCGACCATCATCACATTCAAC	63°C	500	Single
RUBISCO	AF101233	CGCCAAGTCCAGTGCATCAGTTT	CATCCGCATCTTCCATTCCATTTC	63°C	-	Multiple
ERF	AY973613	GAAAGCAATGGAAAATCTCTATCA	CAACCCAAGTCCAGTGTCTCA	50°C	450	Single
CAT 2	AY973614	ACTTGGGCGTCTGGTGTTGA	GTGATGCGAGGGTCTGATAAGG	56°C	550	Single
COEEP	AY818396	GCTGGTTTGGCTTCTGGGTCTTTC	TTCTCTGCCTCTGGGGTGCTCTTA	57°C	600	Single
SOD	AY642137	TGCCCCTGAGGACGACATTC	AAAACAAACGGGGCAAACCAG	50°C	-	Multiple
AKR	AY973615	CTCATGCCGTTCACCCGATAA	ATGCAAAGACAGCCCTGAAATACA	53°C	-	Multiple
CPI	AF265551	CGCCGTCGATGATTACAACAAGA	CGGGTACCATGGCACAGATAAGAT	-	-	No Product
Amy A	AY944583	ACCCGCTCTGATGGATTTTGTGTC	TGTGCCTCGAGTTGATCCAGTGTC	-	-	No product

[1] *G3pdh = glyceraldehyde 3-phosphate dehydrogenase*; *PAL2 = phenylalanine ammonia-lyase 2* gene; *CYP79D2 = N-hydroxylating cytochrome P450* gene; *GBSSII = granule bound starch synthase II precursor*; *MeEF1 = Manihot esculenta elongation factor 1-alpha* gene; *bglA = beta glucosidase*; *pLIN = linamarase* gene; *HRGP = hydroxyproline-rich glycoprotein; CI= allergenic-related protein Pt2L4; TK = Pto-like serine/threonine kinase; APX3 = ascorbate peroxidise; RUBISCO = ribulose 1,5-bisphosphate carboxylase small chain precursor; ERF = ethylene response factor; CAT2 = catalase; COEEP = chloroplast oxygen-evolving enhancer protein; SOD = copper/zinc superoxide dismutase; AKR = aldo/keto reductase; CPI = cysteine protease inhibitor; AmyA = alpha-amylase gene.* The letters ABC adjacent to the same genes represent primers targeting different parts of the gene.

5.2.3 Characterisation of SNPs

The identified SNPs were classified as either being a transition (substitution between purines or between pyrimidines) or transversion (substitution of purine with pyrimidine or *vice versa*) and their frequencies were calculated on a gene fragment basis. Nucleotide diversity at the level of the gene fragment represents the proportion of nucleotides that differ between two sequences, averaged over all available pairs of genotype comparisons. This was computed from the number of polymorphic segregating sites on a base pair basis (Nei, 1987). Tajima's *D*-statistic (Tajima, 1989) was computed to obtain insights into the hypothesis of selective neutrality. Non-significant values indicated no evidence for evolutionary selection. The above analysis was performed using DNA sequence polymorphism (DNASP) version 4.10.9 (Rozas et al., 2003).

The distribution of SNPs along gene fragments was examined to understand if nucleotide variability was distributed randomly or organised in haplotypes. The high level of heterozygotes in cassava meant that the phase of adjacent alleles could not be determined and thus allelic haplotypes had to be inferred. A similar situation was experienced with grapevine (Salmaso et al., 2004). Haplotypes were inferred using PHASE software version 2.1.1 (Stephens et al., 2001; Stephens and Donnelly, 2003) which uses a Bayesian statistical approach. Haplotype based gene diversity was computed for each polymorphic gene fragment using DNASP software version 4.10.9 (Rozas et al., 2003).

Haplotype-based polymorphic information content (PIC) values for each gene fragment were computed as described by Botstein et al. (1980). PIC = $1-\sum p_i^2 - \sum 2p^2_i\ p_j^2$, where $\sum p_i^2$ is the sum of each squared i^{th} haplotype frequency. During computation, each haplotype was considered to be an allele. Observed heterozygosity (H_o) was calculated as the proportion of heterozygous individuals at each locus. Nei's index of gene diversity (expected heterozygosity) and PIC were calculated for each SNP within a gene fragment as described Botstein et al. (1980). This analysis was done using PowerMarker version 3.25 software (Liu and Muse, 2005).

5.2.4 Comparison of diversity patterns revealed by SSRs and SNPs

To compare the information content and diversity revealed by SNPs and SSRs, the 74 test cassava varieties were genotyped using 12 polymorphic SSR markers distributed across nine linkage groups (SSRY5, SSRY9, SSRY21, SSRY69, SSRY102, SSRY135, SSRY 147, SSRY148, SSRY161, SSRY181, SSRY182 and NS911) (Mba et al., 2001; Okogbenin et al., 2006). The numbers of loci to be sampled in a genetic study is not absolute. It will depend on the objectives, the genetic relationship among individuals being studied (the more diverse individuals are, the fewer markers are required) and the informativeness of the SSR. Olsen (2004) used sequenced portions of two low copy genes and five SSR loci to infer the origin of cassava using highly diverse germplasm.

Amplifications with SSR primers were carried out in 10 µl reactions containing 50 ng of DNA, 1 pmole of each primer, 1x *Taq* polymerase buffer, 2 mM of $MgCl_2$, 0.2 mM dNTPs and 0.375 U of *Taq* polymerase. The PCR profile was 95^{o}C for 2 min followed by 30 cycles of 95^{o}C for 30 sec, 55°-57^{o}C for 1 min and 72^{o}C for 1 min and a final extension at 72^{o}C for 30 min. The SSR amplicons were subjected to capillary electrophoresis as described in Chapter 4 and allele calls made using the GENEMAPPER® software version 3.7 (Applied Biosystems). Data was analysed as described above.

DARwin (Perrier and Jackquemound-Collet, 2006) was used to compute a distance matrix from the SSR and SNP data independently using the simple matching Euclidean distance. Cluster analysis was performed using the weighted neighbour-joining algorithm and relationships displayed as a dendrogram. The Mantel test (Mantel, 1967) with 1000 permutations was performed to assess the correlation between the distance matrices of the SSR and SNP genotypic data. The Mantel test uses randomisation to estimate the *P*-value for the null hypothesis of no correlation.

5.3 Results

5.3.1 Characterisation of SNPs

The approach used here for cassava SNP identification and characterisation was based on the direct sequencing of gene fragments amplified by primers designed from NCBI cassava core nucleotides, which targeted the intron regions of the selected genes. Of the designed 28 primer pairs (targeting 19 genes), 14 (50%) produced single and clear PCR products, nine (32%) produced multiple products and five (18%) produced no PCR products (Table 5.2). The PCR products ranged from 250-700 bp. Of the 14 primer pairs that produced single PCR products, high quality sequences appropriate for SNP characterisation were obtained from nine gene fragments (Table 5.3). In total, 3165 bp of genomic DNA from nine gene fragments or 176269 bp from all samples were sequenced. The discrepancy from the expected 234000 bp (74 x 3165) resulted from different samples across the different gene fragments.

The number of SNPs per gene fragment varied from one in the *Ethylene response factor* (*ERF*) and *allergenic-related protein Pt2L4* (*CI*) to five in the *catalase* gene fragment (*CAT2*) (Table 5.3). Overall 26 SNPs were identified in the 3165 bp, giving a frequency of one SNP every 121 nucleotides. Sequence alignments using Spidey NCBI (http://www.ncbi.nlm.nih.gov/spidey/) revealed that of the 26 SNPs, 14 were located within the coding region spanning 2534 bp (one SNP every 181 bp or a frequency of 5 x 10^{-3}) and 12 in non-coding regions spanning 631 bp (one SNP evey 53 bp or a frequency of 1.9 x 10^{-2}) (Table 5.3). The frequency of transitions and transversions were 50% with the most frequent transition and transversion being A/G and A/T, respectively (Table 5.3). Upon translation of the sequenced gene fragments, one amino acid change from alanine (GCA) to serine (TCA) was noted at polymorphic site 225 within the *N-hydroxylating cytochrome P450 (CYP79D2)* gene.

Table 5.3 Summary of SNP characteristics in the sequenced cassava gene fragments[1]

Gene	Sequenced length (bp)	SNP position (bp)	Homozygotes	Heterozygotes	Context of mutation	SNP location
CAT2	570	63	C (3); T (22)	C/T (13)	Transition	Coding region
		200	A (4); G (38)	A/G (26)	Transition	Non-coding region
		259	G (3); C (29)	G/C (25)	Transversion	Coding region
		269	A (4); T (35)	A/T (22)	Transversion	Coding region
		454	C (21); T (48)	-	Transition	Non-coding region
COEEP	497	98	G (7); A (31)	G/A (21)	Transition	Coding region
		248	A (1); T (47)	A/T (8)	Transversion	Coding region
		349	G (3); A (27)	G/A (2)	Transition	Coding region
PAL2	356	34	G (4); A (37)	-	Transition	Non-coding region
		55	G (2); A (44)	G/A (21)	Transition	Non-coding region
		79	A (6); T (23)	A/T (17)	Transversion	Non-coding region
		188	G (4); A (46)	-	Transition	Non-coding region
GBSSII	282	22	T (1); G (55)	T/G (5)	Transversion	Coding region
		82	C (1); A (67)	C/A (5)	Transversion	Coding region
		274	T (72)	C/T (2)	Transition	Non-coding region
CYP79D2	290	170	G (2); A (19)	G/A (13)	Transition	Coding region
		187	A (3); G (17)	A/G (19)	Transition	Coding region
		225	T (2); G (17)	T/G (16)	Transversion	Coding region
		239	A (3); T (18)	A/T (14)	Transversion	Coding region
pLIN	253	37	A (44)	A/T (12)	Transversion	Coding region
		42	T (23)	T/C (25)	Transition	Coding region
		152	A (50)	A/T (4)	Transversion	Non-coding region
MeEF1	283	147	A (36)	A/G (17)	Transition	Non-coding region
		166	G (55); T (1)	G/T (3)	Transversion	Non-coding region
ERF	428	152	T (6)	T/G (7)	Transversion	Non-coding region
CI	206	11	A (7); T (18);	T/A (2)	Transversion	Non-coding region

[1]*CAT2 = catalase*; *COEEP = chloroplast oxygen-evolving enhancer protein*; *PAL2 = phenylalanine ammonia-lyase 2* gene; *GBSSII = granule bound starch synthase II precursor*; *CYP79D2 = N-hydroxylating cytochrome P450* gene; *pLIN = linamarase* gene; *MeEF1 = Manihot esculenta elongation factor 1-alpha* gene; *ERF = ethylene response factor*; *CI = allergenic-related protein Pt2L4.* The numbers in parenthesis indicate scored gene fragments in the contig. Homozygous SNPs include both the more commonly occurring nucleotide (major alleles) and the less common nucleotide (minor alleles), while heterozygous SNPs contained both the major and minor nucleotides (alleles) at a specified position along the gene fragment.

5.3.2 Nucleotide diversity, selection, haplotypes and polymorphic information content

Nucleotide diversity varied among the gene fragments, from 7.8 x 10^{-4} in the *granule bound starch synthase II precursor gene (GBSSII)* to 5.63 x 10^{-3} *in CYP79D2* gene (Table 5.4), representing a seven-fold difference in nucleotide diversity. All gene fragments had non-significant Tajima's *D*-statistic values with the exception of the *catalase* (*CAT2*) gene (D = 2.433; $P < 0.05$) and the *CYP79D2* gene (D = -2.116; $P < 0.05$). However, *GBSSII* and *Manihot esculenta elongation factor 1-alpha (MeEF1)* genes had respective negative *D*-values of -0.991 and -0.369 (Table 5.4). Significant negative *D*-statistic values are indicative of negative or purifying selection against genotypes carrying the less frequent alleles and/or are indicative of a recent population bottleneck eliminating less frequent alleles, whereas significantly positive values are indicative of balancing or diversifying selection for two or more alleles (Tajima, 1989; Hartl, 2000).

The number of haplotypes varied among the gene fragments from two in *CI* and *ERF* genes to 14 in the *CAT2* gene, with a mean of seven (Table 5.4). In the absence of intragenic recombination, repeated mutation or back mutation, the maximum number of haplotypes expected from segregating sites (s) is $s + 1$ (Clarke et al., 1998). Based on the results only three gene fragments *ERF*, *CI* and *GBSSII* were consistent with this rule (Table 5.4). Most gene fragments had major haplotypes accompanied by a series of low frequency haplotypes. The frequency of the minor (rare) alleles varied from 0.013 in *GBSSII* to 0.320 in the *CYP79D2* gene (Table 5.5). Although the number of SNPs and haplotypes varied considerably among the genes studied, a linear relationship (r = 0.914) was evident between the number of SNPs identified for a gene fragment and the number of inferred haplotypes.

Haplotype-based gene diversity (H_d) or heterozygosity was highest in the *CAT2* gene (H_d = 0.814) and lowest in the *GBSSII* gene (H_d = 0.191) (Table 5.4). On the other hand, haplotype-based PIC values ranged from 0.113 for the *GBSSII* gene to 0.634 for the *phenylalanine ammonia-lyase 2* gene *(PAL2A)* gene (Table 5.4). Marginal differences were observed between haplotype-based PIC values computed from "best reconstruction" haplotypes and expected haplotype frequencies.

The observed heterozygosity per individual SNP locus ranged from zero in *PAL2-34*, *PAL2-188* and *CAT2-454* to 0.62 at *ERF* (Table 5.5). With the exception of *PAL2-34*, *PAL2-188* and *CAT2-454*, all other loci were heterozygous as expected in an out-crossing species. The sequence data clearly showed the heterozygosity in cassava (Figure 5.1). Expected heterozygosity (H_e) ranged from 0.026 (*GBSII-274*) to 0.435 (*CYP79D2-187*). PIC values varied from 0.026 (*GBSII-274*) to 0.340 (*CYP79D2-187)* (Table 5.5). SNPs within the *CAT2* and *CYP79D2* genes had relatively similar PIC values along the gene fragments, as compared to other gene fragments whose PIC values varied (Table 5.5). PIC values increased substantially when calculated on the basis of haplotypes as opposed to individual SNPs (Tables 5.4 and 5.5).

Table 5.4 Total haplotypes, haplotype diversity (H_d), nucleotide diversity, neutrality, D-statistics and haplotype based polymorphism information content of polymorphic cassava genes

Gene	Haplotypes	H_d ± SD	Nucleotide diversity ± SD	*D*-statistic	PIC
CAT2	14	0.814 ± 0.025	0.00351 ± 0.00017	2.433*	0.602
COEEP	7	0.661 ± 0.029	0.00213 ± 0.00015	0.886	0.594
PAL2	11	0.806 ± 0.017	0.00387 ± 0.00021	1.701	0.634
GBSSII A	4	0.191 ± 0.043	0.00078 ± 0.00019	-0.991	0.113
CYP79D2	7	0.692 ± 0.039	0.00563 ± 0.00046	2.116*	0.375
pLIN	8	0.428 ± 0.048	0.00199 ± 0.00026	0.452	0.481
MeEF1	4	0.233 ± 0.050	0.00153 ± 0.00036	-0.369	0.263
ERF	2	0.409 ± 0.083	0.00098 ± 0.00020	0.918	0.335
CI	2	0.455 ± 0.040	0.00285 ± 0.00025	1.449	0.330
Mean	**7**	**0.521**	**0.00258**		**0.414**

[1]*CAT2 = catalase; COEEP = chloroplast oxygen-evolving enhancer ptrotein; PAL2 = phenylalanine ammonia-lyase 2 gene; GBSSII = granule boundstarch synthase II precursor; CYP79D2 = N-hydroxylating cytochrome P450 gene; PLIN = linamarase gene; MeEF1 = Manihot esculenta elongation factor 1-alpha gene; ERF = ethylene response factor; CI = allergenic-related protein Pt2L4*. Haplotypes computed from PHASE. H_d = haplotype based gene diversity; SD = standard deviavtion; PIC = polymorphic information content; * Significant at $P < 0.05$

Table 5.5 Gene diversity, heterozygosity and polymorphic information content (PIC) of the identified cassava SNPs[1]

Gene fragment	SNP position (bp)	No. of observations	SNP context	q	Gene diversity (H_e)	Observed heterozygosity (H_o)	PIC
CAT2	63	38	C/T	0.250 (C)	0.375	0.342	0.304
	200	68	A/G	0.250 (A)	0.375	0.382	0.304
	259	57	C/G	0.271 (G)	0.396	0.438	0.317
	269	61	A/T	0.245 (A)	0.370	0.360	0.302
	454	69	C/T	0.304 (C)	0.423	0.000	0.333
	Mean				**0.388**	**0.304**	**0.312**
COEEP	98	59	A/G	0.296 (G)	0.417	0.355	0.330
	248	56	A/T	0.089 (A)	0.162	0.142	0.149
	349	32	A/G	0.125 (G)	0.218	0.062	0.194
	Mean				**0.266**	**0.187**	**0.224**
PAL2	34	41	A/G	0.097 (G)	0.176	0.000	0.160
	55	67	A/G	0.186 (G)	0.303	0.313	0.257
	79	46	A/T	0.315 (A)	0.431	0.369	0.338
	188	50	A/G	0.080 (G)	0.147	0.000	0.136
	Mean				**0.264**	**0.170**	**0.223**
GBSSII	22	61	G/T	0.057 (T)	0.108	0.082	0.102
	82	73	A/C	0.047 (C)	0.091	0.068	0.087
	274	74	C/T	0.013 (C)	0.026	0.027	0.026
	Mean				**0.075**	**0.059**	**0.071**
CYP79D2	170	34	A/G	0.250 (G)	0.375	0.382	0.304
	187	39	A/G	0.320 (A)	0.435	0.487	0.340
	225	35	G/T	0.285 (T)	0.402	0.457	0.324
	239	35	A/T	0.285 (A)	0.408	0.400	0.324
	Mean				**0.406**	**0.431**	**0.323**
pLIN	37	56	A/T	0.107 (T)	0.191	0.214	0.173
	42	48	T/C	0.260 (T)	0.385	0.520	0.311
	152	54	T/A	0.037 (T)	0.071	0.074	0.068
	Mean				**0.216**	**0.269**	**0.184**
MeEF1	147	53	A/G	0.160 (G)	0.269	0.321	0.233
	166	59	G/T	0.042 (T)	0.081	0.050	0.077
	Mean				**0.175**	**0.185**	**0.155**
ERF	152	13	G/T	0.308 (G)	0.426	0.615	0.335
CI	11	27	T/A	0.297 (A)	0.417	0.074	0.330

[1]*CAT2 = catalase; COEEP = chloroplast oxygen-evolving enhancer ptrotein; PAL2 = phenylalanine ammonia-lyase 2 gene; GBSSII = granule boundstarch synthase II precursor; CYP79D2 = N-hydroxylating cytochrome P450 gene; PLIN = linamarase gene; MeEF1 = Manihot esculenta elongation factor 1-alpha gene; ERF = ethylene response factor; CI = allergenic-related protein Pt2L4*. q = frequency of the minor allele and the letters in parenthesis represent the minor nucleotide (allele) at a specific polymorphic site along the seqeunced gene fragment. PIC = polymorphic information content.

Figure 5.1 Heterozygous and homozygous SNPs at position 200 of the *CAT2* gene. Sequences were compared for five individuals.

5.3.3 Comparison of SSR and SNP data

The number of alleles at SSR loci ranged from three for SSRY102 to 11 for SSRY135. The PIC values ranged from 0.358 for SSRY102 to 0.759 for SSRY6, with a mean of 0.571. This average is higher than that of individual SNPs (0.228) and comparable to that based on haplotypes (0.414). Average gene diversity was generally lower in cassava genotypes from Asia (SNP-based = 0.251; SSR-based = 0.553), followed by African cassava (SNP-based = 0.271; SSR-based = 0.556) and American cassava (SNP-based = 0.285; SSR-based = 0.615). The Mantel test showed a significant correlation between the distance matrices of SNP and SSR genotypic data ($r = 0.219$; $P < 0.001$), indicating correlation between the SSR and SNP genotypic data.

Dendograms of both SSR and SNP genotypic data resulted in three major clusters (Figures 5.2a and 5.2b). The SNP data showed less discrimination (Figure 5.2a). Cluster one contained 84% of germplasm of African origin, a second cluster contained germplasm of predominantly American and Asian origin (62.5%), while the third cluster contained germplasm predominantly from America and Africa. Based on SSR data, 28 of the African accessions clustered together with a single accession from the Americas (ECU72). The second cluster represented accessions of varied origins, while the third cluster had germplasm from the Americas or Asia (Figure 5.2b). A notable exception of genetic relationships revealed by the individual genes was that of *GBSSII*, where two distinct clusters were produced, with each comprising of similar representation of genotypes from Africa and Latin America (data not shown).

Figure 5.2a Dendrogram of cassava genotypes based on pairwise genetic distances derived from 26 SNPs.

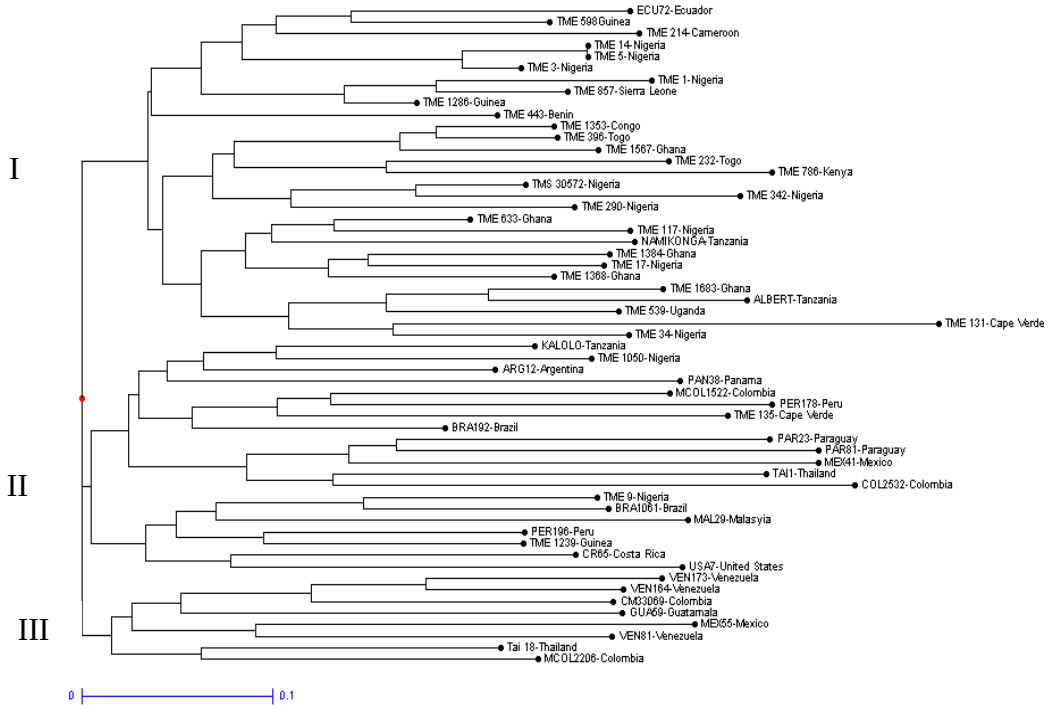


Figure 5.2b Dendrogram of cassava genotypes based on pairwise genetic distances derived from 12 SSRs.

5.4 Discussion

A major objective of this study was to characterise SNPs in selected cassava genes and to assess their utilisation for diversity assessment. The nature and frequency of SNPs in nine low copy genes involved in cyanogenesis, starch metabolism, stress and/or defense related pathways were studied. This study identified 26 SNPs, with an average of one SNP every 121 nucleotides. Lopez et al. (2005) found a higher frequency of one SNP in 66 bp in 33 EST-derived amplicons analysed in six cultivars and a frequency of one SNP in 49 bp when nine BAC end sequence amplicons were analysed. As expected, the frequency of SNPs in ESTs within a cassava cultivar was much lower, one in 905 bp, while between two or more cultivars it was one in 1032 bp (Lopez et al., 2005).

A similar high frequency of SNPs has been found in other crops, for example one SNP in 78 bp for grapevine (Salmaso et al., 2004), one SNP per 31 bp in non-coding regions and one SNP per 124 bp in the coding regions based on analysis of 18 maize genes in 36 inbred lines (Ching et al., 2002). For inbreeding crops like soybean, SNP frequencies have been reported to be one in 2038 bp in the coding sequence and one SNP in 191 bp in the non-coding regions based on the analysis of 35 gene fragments in 15 genotypes (Van et al., 2005). Undoubtedly, the frequency of detected SNPs in the genome appears to be largely dependant on the nature of the gene sampled and the individuals studied. No correlation between the length of the sequenced gene fragment and the number of SNPs identified ($R^2 = 0.079$; $P = 0.235$) was established, a scenario which is consistent with the findings of Lopez et al. (2005). Out-crossing and inbreeding-sensitive crops like cassava are expected to exhibit higher DNA polymorphisms in part due to the inherently high number of loci maintained in a heterozygous state.

The lower frequency of SNPs in the coding regions as opposed to the non-coding regions and the fact that only one polymorphic site in the *CYP79D2* gene was associated with the alanine (GCA) to serine (TCA) amino acid change can be explained in part by Kimura's neutral theory. This hypothesises that nonsynonymous substitutions are subjected to purifying selection more often than synonymous substitutions (Kimura, 1983).

Most studies conducted in both inbreeding (Van et al., 2005) and out-crossing (Ching et al., 2002; Salmaso et al., 2004; Lopez et al., 2005) species indicate a higher number of polymorphisms in the non-coding than in the coding regions. Substitutions in the coding region that lead to alterations in the amino acid sequences and/or early termination of the translation process may cause an altered phenotype (Klug et al., 2005). If the altered phenotype is of breeding value, then these coding SNPs could be valuable genetic markers for cassava.

Though this study did not associate sequence polymorphisms with phenotypes, this can in the future be undertaken through association tests between sequence variants in candidate genes and phenotypes, as illustrated by the GC/TT polymorphism in the rice *starch synthase IIa* gene and gelatinisation temperature (Bao et al., 2006). Alternatively, SNPs without any observable mutant phenotype can be used for the positional cloning of the gene in which they are located (Kruglyak, 1997).

Average nucleotide diversity across gene fragments was 2.58×10^{-3}, which is generally lower than that observed in maize (9.6×10^{-3}; Tenaillon et al., 2001 and 6.3×10^{-3}; Ching et al., 2002), but higher than for the inbreeding species soybean, which is reported to range from 5.3×10^{-4} to 1.1×10^{-3} (Zhu et al., 2003). In most eukaryotic genes, nucleotide diversity is less than 0.01. However, an exception is the *major histocompatibility complex* gene in humans and mice whose values were greater than 0.04 largely due to overdominant selection at this locus (Nei, 1990). In this study the nucleotide diversity (nd) of *GBSSII* with three SNPs was two to seven-fold lower than in the majority of other gene fragments. This observation may be due to a relatively recent genetic bottleneck, however this was not observed in our study (D = -0.991; $P > 0.1$), or positive selection for the major allele.

The greatest nucleotide diversity was found in the *CYP79D2* gene (nd = 5.63×10^{-3}), involved in cyanogenesis. Evidence for diversifying selection was observed for this gene (D = 2.116; $P < 0.05$), which also had a nonsynonymous mutation. Estimates of nucleotide diversity of two other gene fragments also involved with the cyanogenesis pathway, ß-*glucosidase* (*Bgla*) and ά-*hydroxynitrile* (*Hnl*) in 20 cassava varieties and 212 wild relatives were similarly high (4.1×10^{-3} and 4.5×10^{-3} respectively), although this may reflect the diversity of genotypes studied (Olsen, 2004). Two-fold lower levels of nucleotide diversity were observed within the *linamarase gene* (*pLIN)* (nd = 1.99×10^{-3}), also involved in the cyanogenesis pathway. All cassava genotypes studied so far contained cyanogenic glycosides, which are governed by several genes within the cyanogenesis pathway.

Since most cassava breeding programmes will advocate for reduced glycosides for health reasons, the heterozygous nature of the crop, ID and the involvement of several loci in cyanogenic glycosides, collectively make attainment of this breeding objective challenging. This could explain the high nucleotide diversity within and among the different genes involved in the cassava cyanogenic pathway reflecting varying selection intensities. Other studies have also found variable levels of nucleotide diversity and selection in the same biochemical pathway, for example in the *flavanone-3 hydroxylase* (*F3H1* and *F3H2*) genes involved in the flavonoid pathway in Douglas fir tree species (Krutovsky and Neale, 2005) and in the six genes involved in the starch pathway in maize (Whitt et al., 2002).

Key factors that are thought to influence nucleotide diversity include background selection which involves selection against recurrent deleterious mutations (Charlesworth et al., 1993), recombination rates (Begun and Aquadro, 1992) and positive-trait selection (Buckler and Thornsberry, 2002). It is possible that the observed differences in nucleotide diversity among the studied cassava genes are due to these factors acting either singly or in combination at different levels among the studied genes. However, significant Tajima's *D* statistic values were observed for *CAT2* and *CYP79D2* genes, indicating diversifying selection. Other genes had non-significant positive or negative *D* values, indicating no evidence of selection. In a related study in which GC/TT sequence polymorphism in the rice *starch synthase IIa* gene were associated with gelatinising, no detectable selection was detected despite having a TT allele frequency of 47% in 334 breeding lines as compared to only 1.2% TT allele frequency in 172 landraces, a finding which could in itself indicate directional selection in the breeding lines (Bao et al., 2006).

As expected from an out-crossing species, heterozygous loci were frequent and were found within all gene fragments, but not at all SNP loci. A relative excess of homozygosity was observed within the *GBSSII* gene compared to other genes. Cassava is primarily a starchy crop and *GBSSII* is one of the key genes involved in starch metabolism.

Though no detectable selection was noticed in the *GBSSII* gene based on Tajima's *D* statistic (-0.991; $P > 0.1$), the low PIC values (SNP-based = 0.071; haplotype-based = 0.113) and excess homozygosity are indicative of reduced genetic diversity within this gene. It is possible that this reduced genetic variability relates to functional versions of this gene. Indeed, Ceballos et al. (2007) reported a mutation that renders the *GBSS* enzyme inoperative in an S_1 cassava clone. In maize, which is also a starchy crop, low genetic diversity and a strong evidence of selection across three genes (*brittle2*, *sugary1* and *amylose extender1*) involved in the starch pathway have been reported (Whitt et al., 2002).

The informativeness of a genetic marker depends on the number of alleles it detects and allele frequencies and this is quantified by PIC. The highest PIC value of individual SNPs was 0.34 from the polymorphic site 187 bp of the *CYP79D2* gene. SNPs are bi-allelic markers, which can have a 50-50 distribution of the two alleles (Kruglyak, 1997) and hence their PIC values can not exceed 0.5. However, when linked alleles (haplotypes) are considered simultaneously, the level of polymorphism can increase (Kruglyak, 1997). Indeed, haplotype-based PIC values (0.113-0.634) were almost double when compared to PIC values based on individual SNPs (0.026-0.340), as several SNP loci are considered together. Similar trends were observed for the haplotype-based gene diversity.

In rye, haplotype-based PIC values ranged between 0.38-0.8 with a mean of 0.66, while SNP-based PIC values ranged between 0.22-0.5, with a mean of 0.32 (Varshney et al., 2007). Similar trends have been registered in grapevine (Salmaso et al., 2004) and maize (Ching et al., 2002). In a related study that involved analysis of 259 maize inbred lines with 847 SNPs and 554 SNP haplotypes, allele frequencies ranged from 0-0.5 and 0-0.9, respectively (Hamblin et al., 2007), a finding which further confirmed the discriminatory power of SNP haplotypes over individual SNPs, irrespective of the crop's genetic and/or mating system.

SSR and SNP data revealed a similar trend in data resulting in an overall significant correlation of genetic relationships. Both data sets revealed three major clusters, one predominantly African germplasm, the other predominantly germplasm of American and Asian origin and a third consisting of accessions from mixed origin. Indeed, both SSR and SNP data revealed similar trends in gene diversity according to region of origin, although SSRs consistently revealed greater diversity and discrimination. These results are similar to those obtained from an extensive SSR survey of over 2000 cassava accessions at 30 SSR loci (Morag Ferguson, personal communication). The genetic relationships as revealed by combined SNP data were similar to previous genetic analysis where African genotypes TME 1, TME 3, TME 5, TME 9 and TME 14 clustered together and differently from TMS 30572 based on AFLP analysis (Fregene et al., 2000).

Hamblin et al. (2007) tried to assign a known population to 259 maize inbred lines using 89 SSRs, 847 SNPs and 554 SNP haplotypes. The populations were stiff stalk, non-stiff stalk and tropical-semitropical and the analysis was run with *K* (number of putative populations in the model) ranging from 2-5. The assignment of individuals across the genetic markers was consistent with $K = 3$, but less consistent at higher values of *K*. This study also established that for all *K* values tested, SNP data did not contain sufficient information to resolve all relationships that were detected by SSR variation resulting in lower percentage of individuals assigned to a population (Hamblin et al., 2007).

These findings are generally consistent with the SNP and SSR datasets generated from this study on cassava, a highly heterozygous and out-crossing species. The similarity of American, particularly Latin American germplasm and Asian germplasm, reflects the movement of germplasm from South America, the centre of diversity and domestication of cassava (Allem, 1994; Olsen and Schaal, 1999; Olsen, 2004) to Asia. Clustering of some African germplasm with that from American/Asian germplasm indicated similarity and movement of germplasm among these regions.

Cassava was introduced to Africa by the Portuguese and spread rapidly to many agro-ecologies (Jones, 1959). The cluster containing predominantly African germplasm is likely to contain clones of ancient introductions that have undergone natural and/or artificial selection and adaptation.

The higher PIC values of SSRs, derived from their multi-allelism, rapid mutation rate and low probability of being affected by the narrowing influence of selection, appeared to provide greater discrimination among accessions from different regions. Individual SNPs had lower PIC values than SSRs resulting from their bi-allelism. This may have been compounded by the fact that SNPs were closely associated within genes, although evidence was found for recombination within gene fragments. Genic regions have a higher probability of being affected by directional selection than non-genic regions. In addition, only a small proportion of the genome was represented in this assessment. The above factors will tend to decrease the discriminatory power of SNPs. It is likely that as a result of their bi-allelic as opposed to multi-allielic nature, a larger number of SNPs than SSRs will be required for the same level of discrimination. If the objective is to study intra-genic variation, however, then SNPs may be more appropriate than SSRs due to their far higher frequency.

5.5 Conclusions

This research chapter has provided initial estimates of the nature, distribution and frequency of SNPs in selected cassava genes. It is clear that SNPs occur at a high frequency within cassava genes and can be targeted as useful genetic markers, particularly where an indication of intra-genic diversity is required. The relative frequency of SNPs within a gene compared to SSRs makes them suitable for this purpose. Furthermore, this study has shown that haplotype-based genetic diversity analysis is by far more informative than analysis based on individual SNPs. However, due to high levels of heterozygosity within the cassava genome, haplotypes had to be inferred, making this approach less reliable for diversity assessment.

SNPs and SSRs revealed similar overall relationships among individuals, confirming the utility of SNPs for diversity estimates. Individual SNPs had lower PIC values than SSRs. For this reason a larger number of SNPs will be necessary to achieve the same level of discrimination provided by SSRs.

5.6 References

Allem, A.C., 1994. The origin of *Mahihot esculenta* Crantz (Euphorbiaceae). Genetic Resources and Crop Evolution 41:133-150.

Bao, J.S., H. Corke, and M. Sun, 2006. Nucleotide diversity in *starch synthase IIa* and validation of single nucleotide polymorphism in relation to starch gelatinisation temperature and other physiochemical properties in rice (*Oryza sativa* L.). Theoretical and Applied Genetics 113:1171-1183.

Begun, D.J., and C.F. Aquadro, 1992. Levels of naturally occurring DNA polymorphisms correlate with recombination rate in *D. melanogaster*. Nature 256:519-520.

Botstein, D., R.L. White, M. Skolnick, and R.W. Davis, 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. American Journal of Human Genetics 32:314-331.

Bradbury, M.T., T.L. Fitzgerald, R.J. Henry, Q. Jin, and D.L.E. Waters, 2005. The gene for fragrance in rice. Plant Biotechnology Journal 3:363-370.

Brookes, A.J., 1999. The esscence of SNP's. Gene 234:177-186.

Brumfield, R.T., P. Beerli, D.A. Nickerson, and S.V. Edwards, 2003. The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology and Evolution 8:249-256.

Buckler, E.S., and J.M. Thornsberry, 2002. Plant molecular diversity and applications to genomics. Current Opinion in Plant Biology 5:107-111.

Ceballos, H., T. Sáchez, N. Morante, M. Fregene, D. Dufour, A.M. Smith, K. Denyer, J.C. Pèrez, F. Calle, and C. Mestres, 2007. Discovery of amylose-free starch mutant in cassava (*Manihot esculenta* Crantz). Journal of Agricultural and Food Chemistry 55: 7469-7476.

Charlesworth, B., M.T. Morgan, and D. Charlesworth, 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289-1303.

Ching, A., K.S. Caldwell, M. Jung, M. Dolan, O.S. Smith, S. Tingey, M. Morgante, and A.J. Rafalski, 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics 3:1-14.

Cho, R.J., M. Mindrinos, D.R. Richards, R.J. Sapolsky, M. Anderson, E. Drenkard, J. Dewdney, T.L. Reuber, M. Stammers, N. Federspiel, A. Theologis, W. Yang, E. Hubbell, M. Au, E.Y. Chung, D. Lashkari, B. Lemieux, C. Dean, R.J. Lipshutz, F.M. Ausubel, R.W. Davis, and P.J. Oefner, 1999. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Natural Genetics 23:203-207.

Clarke, A.G., K.M. Weiss, D.A. Nickerson, S.L. Taylor, A. Buchanan, J. Stengård, V. Salomaa, E. Vartiainen, M. Perola, E. Boerwinkle, and C.F. Sing, 1998. Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. American Journal of Human Genetics 63:595-612.

Collins, A., W. Lau, and F.M. De la Vega, 2004. Mapping genes for common diseases: the case for genetic (LD) maps. Human Heredity 58:2-9.

Collins, F.S., L.D. Brooks, and A. Chakravarti, 1998. A DNA polymorphism discovery resource for research on human genetic variation. Genome Research 8:1229-1231.

Davignon, J., R.E. Gregg, and C.F. Sing, 1988. Apolipoprotein E polymorphism and atherosclerosis. Arteriosclerosis 8:1-21.

Dellaporta, S.J., J. Wood, and J.B. Hicks, 1983. A plant DNA mini preparation. Version II. Plant Molecular Biology Reporter 1:19-21.

Fregene, M., A. Bernal, M. Duque, A. Dixon, and J. Tohme, 2000. AFLP analysis of African cassava (*Manihot esculenta* Crantz) germplasm resistant to the cassava mosaic disease (CMD). Theoretical and Applied Genetics 100:678-685.

Fullerton, S.M., A.G. Clark, K.M. Weiss, D.A. Nickerson, S.L. Taylor, J.H. Stengard, V. Salomaa, E. Vartiainen, M. Perola, E. Boerwinkle, and C.F. Sing, 2000. Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. American Journal of Human Genetics 67:881-900.

Gupta, P.K., J.K. Roy, and M. Prasad, 2001. Single nucleotide polymorphism: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Current Science 80:524-535.

Gut, G.I., 2001. Automation in genotyping of single nucleotide polymorphism. Human Mutation 17:475-492.

Hamblin, M.T., M.L. Warburton, and E.S. Buckler, 2007. Empirical comparison of simple sequence repeats and single nucleotide polymorphism in assessment of maize diversity and relatedness. Plos One 2: e1367. doi:10.1371/journal.pone.0001367. 26th December 2007.

Hartl, D.L., 2000. A primer of population genetics. Third Edition. Sinauer Associates, Inc., pp. 221.

Jones, W.O., 1959. *Manioc* in Africa. Standard University Press, Stanford, California, pp. 315.

Jos, J.S., and S.G. Nair, 1979. Pachytene pairing in relation to pollen fertility in five cultivars of cassava. Cytologia 44:813-820.

Kanazin, V., H. Talbert, D. See, P. DeCamp, E. Nevo, and T. Blake, 2002. Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). Plant Molecular Biology 48:529-537.

Kimura, M., 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, pp. 367.

Klug, W.S, M.R. Cummings, and C. Spencer, 2005. Concepts of Genetics. Eighth Edition, pp. 677.

Kruglyak, L., 1997. The use of a genetic map of biallelic markers in linkage studies. Nature Genetics 17:21-24.

Krutovsky, K.V., and D.B. Neale, 2005. Nucleotide diversity and linkage disequilibrium in cold-hardiness and wood quality related candidate genes in Douglas fir. Genetics 171:2029-2041.

Kwok, P.Y., 2001. Methods for genotyping single nucleotide polymorphisms. Annual Review of Genomics and Human Genetics 2:235-258.

Lindblad-Toh, K., E. Winchester, M.J. Daly, D.G. Wang, J.N. Hirschhorn, J.P. Laviolette, K. Ardlie, D.E. Reich, E. Robinson, P. Sklar, N. Shah, D. Thomas, J.B. Fan, T. Gingeras, J. Warrington, N. Partil, T.J. Hudson, and E.S. Lander, 2000. Large-scale discovery and genotyping of single-nucleotide polymorphism in the mouse. Nature Genetics 24:381-386.

Liu, K., and S.V. Muse, 2005. PowerMarker. Integrated analysis environment for genetic marker data. Bioinformatics 21:2128-2129.

Lopez, C., B. Piėgu, R. Cooke, M. Delseny, J. Tohme, and V. Verdier, 2005. Using cDNA and genomic sequences as tools to develop SNP strategies in cassava (*Manihot esculenta* Crantz). Theoretical and Applied Genetics 110:425-431.

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209-220.

Mba, R.E.C., P. Stephenson, K. Edwards, S. Melzer, J. Nkumbira, U. Gullberg, K. Apel, M. Gale, J. Tohme, and M. Fregene, 2001. Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. Theoretical and Applied Genetics 102:21-31.

McCallum, C.M., L. Comai, E.A. Greene, and S. Henikoff, 2000. Targeted screening for induced mutations. Nature Biotechnology 18:455-457.

Neale, B.B., and O. Savolainen, 2004. Association genetics of complex traits in conifers. Trends in Plant Science 9:325-330.

Nei, M., 1987. Molecular evolutionary genetics. Columbia University Press, New York, NY, pp. 512.

Nei, M., 1990. DNA polymorphism and adaptive evolution. In: Brown, A.H.D, M.T. Clegg, A.L. Kahler, and B.S. Weir (Eds.), Plant population genetics, breeding and genetic resources, pp. 128-142. Sinauer Associates Inc. Sunderland, Massachusetts, USA.

Okogbenin, E., J. Marin, and M. Fregene, 2006. An SSR-based molecular genetic map of cassava. Euphytica 147:433-440.

Olsen, K.M., 2004. SNPs, SSRs and inferences on cassava origin. Plant Molecular Biology 56:517-526.

Olsen, K.M., and B.A. Schaal, 1999. Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. Proceedings of the National Academy of Sciences (USA) 96:5586-5591.

Perrier, X., and J.P. Jacquemound-Collet, 2006. DARwin software. http://darwin.cirad.fr/darwin

Rafalski, J.A., 2002a. Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology 5:94-100.

Rafalski, J.A., 2002b. Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Science 162:329-333.

Rogers, D.J., 1963. Studies of *Manihot esculenta*, Crantz and related species. Bulletin of the Torrey Botany Club 90:43-54.

Rogers, D.J., and S.G. Appan, 1973. *Manihot manihotoides* (Euphorbiaceae) Flora Neotropica. Monograph no. 13, Hafner Press, New York, N.Y, pp. 272.

Rozas, J., J.C. Sanchez-Delbarrio, X. Messeguer, and R. Rozas. 2003. DnaSP. DNA polymorphism analysis by the coalescent and other methods. Bioinformatics 19:2496-2497.

Rozen, S., and H. Skaletsky, 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S., and S. Misener (Eds.), Bioinformatics methods and protocols: methods in molecular biology, pp. 365-386. Humana Press, Totowa, NJ.

Salmaso, M., G. Faes, C. Segala, M. Stefanini, I. Salakhutdinov, E. Zyprian, R. Toepfer, M.S. Grando, and R. Velasco, 2004. Genome diversity and gene haplotypes in the grapevine (*Vitis vinifera* L.) as revealed by single nucleotide polymorphisms. Molecular Breeding 14:385-395.

Schneider, K., B. Weisshaar, D. Borchardt, and F. Salamini, 2001. SNP frequency and allelic haplotype structure of *Beta vulgaris* expressed genes. Molecular Breeding 8:63-74.

Stephens, M., and P. Donnelly, 2003. A comparison of Bayesian methods for haplotype reconstruction. American Journal of Human Genetics 73:1162-1169.
Stephens, M., N.J. Smith, and P. Donnelly, 2001. A new statistical method for haplotype reconstruction from population data. American Journal of Human Genetics 68:978-989.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

Tenaillon, M.I., M.C. Sawkins, A.D. Long, R.L. Gaut, and J.F. Doebley, 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* spp.). Proceedings of the National Academy of Sciences (USA) 98:9161-9166.

Van, K., E.Y. Hwang, Y.M. Kim, H.J. Park, S.H. Lee, and P.B. Cregan, 2005. Discovery of SNPs in soybean genotypes frequently used as the parents of mapping populations in the United States and Korea. Journal of Heredity 96:529-535.

Varshney, R.K., U. Beier, E.K. Khlestkina, R. Kota, V. Korzun, A. Graner, and A. Börner, 2007. Single nucleotide polymorphisms in rye (*Secale cereale* L.): discovery, frequency and application for genome mapping and diversity studies. Theoretical and Applied Genetics 114:1105-1116.

Wang, D.G., J.B. Fan, C.J. Siao, A. Berno, P. Young, R. Sapolsky, G. Ghandour, N. Perkins, E. Winchester, J. Spencer, L. Kruglyak, L. Stein, L. Hsie, T. Topaloglou, E. Hubbell, E. Robinson, M. Mittmann, M.S. Morris, N. Shen, D. Kilburn, J. Rioux, C. Nusbaum, S. Rozen, T.J. Hudson, R. Lipshutz, M. Chee, and E.S. Lander, 1998. Large scale identification, mapping and genotyping of single nucleotide polymorphisms in the human genome. Science 280:1077-1082.

Whitt, S.R., L.M. Wilson, M.I. Tenaillon, B.S. Gaut, and E.S. Buckler, 2002. Genetic diversity and selection in the maize starch pathway. Proceedings of the National Academy of Sciences (USA) 99:12959-12962.

Zhu, Y.L., Q.J. Song, D.L. Hyten, C.P. Van Tassell, L.K. Matukumalli, D.R. Grimm, S.M. Hyatt, E.W. Fickus, N.D. Young, and P.B. Cregan, 2003. Single-nucleotide polymorphism in soybean. Genetics 163:1123-1134.

CHAPTER 6

SEGREGATION OF HARVEST INDEX, DRY MATTER AND AMYLOSE CONTENT IN S_1 CASSAVA FAMILIES

6.1 Introduction

Inbreeding, defined as the mating between individuals related by a shared ancestry, can occur in various forms (Falconer and Mackay, 1996). However, self-fertilisation in which two gametes (from the same individual) participate towards the formation of a new individual is the closest form of inbreeding (Altenburg, 1957; Begg, 1959). Inbreeding occurs naturally in self-pollinating plants, while in cross-pollinating plants self incompatibility mechanisms that take various forms, limit self-fertilisation (Glènmin et al., 2001). Once inbreeding occurs in cross-pollinated species, it can lead to varying levels of ID, which has been observed in nearly all cross-pollinated species (Wricke and Weber, 1986), including cassava (Rojas et al., 2009). This decline in fitness is well illustrated by experiments in maize (cross-pollinated species) that were conducted since the early 1900s (Begg, 1959). In that study, of the original hand pollinated plants, only four lines survived by 1912, some of which yielded only two bushels per acre (125.4 kg/ha).

Genetically, ID is caused by increased expression of homozygous recessive genes, which are concealed from expression (or full expression) when in a heterozygous form (Begg, 1959). On the other hand, inbreeding also presents enormous benefits when strategically employed in cross-pollinating species like cassava. Indeed, most successful breeding programmes involve inbreeding during at least one stage of cultivar development. The pioneering work on hybrid maize, based on the use of inbred lines in the early 1900s was a phenomenal initiative (East, 1908), which can be reciprocated in other cross-pollinated crops like cassava. The author recognised the need of three and double-way crosses, spurring the development of hybrid maize, which dominates today's maize acreage.

The superiority of the generated hybrids is purported to be explained by three hypotheses: 1) partial to full dominance for a large number of loci, 2) over-dominance of several loci and 3) several types of epistasis (Wricke and Weber, 1986). Cassava is an out-crossing, highly heterozygous plant that has not undergone intensive and systematic inbreeding to exploit its benefits, as done in maize. Nonetheless, it is encouraging to observe that interest in cassava inbreeding is beginning to gain momentum (Ceballos et al., 2004; 2007; Rojas et al., 2009). Certainly, to tap into the benefits of inbreeding as witnessed in maize, it is vital to initiate inbreeding in cassava.

Walsh (2005) observed that inbreeding provides an opportunity to exploit both additive and non-additive effects. The author demonstrated that under random mating a single diploid parent only contributes one allele per locus and hence cannot pass on its dominance component to its offspring. To exploit non-additive genetic effects, parents must contribute more than just their additive values, that is, they should pass on coordinated groups of alleles at different loci and/or whole genotypes at single or more loci, a phenomenon that will require parents to be related (Walsh, 2005). Inbreeding fixes both dominant and recessive alleles (Sneep et al., 1979), suggesting that if dominant alleles are not linked with recessive alleles on the same chromosome, exceptional vigour can be expected in inbreds (Altenburg, 1957; Begg, 1959).

Studies have indicated that inbreeding does not only express lethal recessive genes, but also good quality recessive genes (Altenburg, 1957). For instance, cassava inbreeding experiments conducted at CIAT observed higher yields of some selfed families (M Colombia 971) compared to the parents, with ID varying among families (Kawano et al., 1978). Waxy cassava starch was recently identified in a S_1 inbred clone (AM206-5) at CIAT (Ceballos et al., 2007). More recent studies have illustrated that ID varies among cassava agronomic traits: fresh root yield (63.9%), fresh foliage yield (37.9%), HI (26.5%), plant height (10.1%) and DMC (5.3%) (Rojas et al., 2009). These findings have practical relevance for cassava breeding.

It is envisaged that inbreeding in cassava will provide several advantages including: 1) reduction of genetic load which limits attainment of sustainable genetic progress, 2) increased probability of attaining useful recessive traits and 3) facilitation of the implementation of mutation breeding (Ceballos et al., 2004). These benefits coupled with the identification of a naturally occurring mutation on the *Wx* locus in a S_1 cassava clone (AM206-5) were a major motivation for this study. In this exploratory study, developed S_1 progeny from six cassava genotypes were examined for starch quality (amylose content) and two agronomic (HI and DMC) traits.

6.2 Materials and methods

6.2.1 Generation and field establishment of S_1 families

Six cassava genotypes (I92/00067, TMS 30572, 95/SE-00036, NASE 4, MH95/0469 and Bamunanika) were used as progenitors (S_0) to generate S_1 progeny. With the exception of Bamunanika which is a local variety, the rest are elite genotypes introduced from IITA. For each genotype 20 stem cuttings were planted at isolation plots at a spacing of 1 m x 0.9 m. The isolation plots were separated by a distance of 100 m from any neighbouring cassava to ensure that only natural self-pollination occurs. A separation of 30 m has been reported to be efficient to ensure genetic isolation in cassava (Kawano et al., 1978). This isolation strategy was used to generate S_1 botanical seed for each genotype.

Cassava is monoecius with female flowers opening before the male flowers on the same branch (IITA, 1990). However, because male and female flowers on different branches or on different plants of the same genotype can open simultaneously, self-pollination can occur (Jennings and Iglesias, 2002; Ceballos et al., 2004). This phenomena was used to produce S_1 botanical seed in the isolation plots. For each selfed genotype, mature fruits were carefully harvested, placed in labelled brown paper bags and left to shatter naturally. Harvested S_1 botanical seeds were not treated, but allowed a two month dormancy breakdown period before being established in nurseries.

After two months in the nursery, the S_1 seedlings were transplanted to a well-prepared field where they were grown until 10 months, after which they were cloned to generate at least 6-10 cuttings (middle section) per seedling. Each S_1 seedling (genotype) was represented by six plants, which were established in the field for evaluation. Because of modest number of progeny per family (8-30), field planting was done so that all progeny belonging to the same family, together with the parental genotype, were established in the same block. Each row represented a clone and spacing within rows was 1 m, with between row spacing of 1.5 m to minimise interplot interference. A variable number of S_1 seedlings were generated per family and these were evaluated at the National Crops Resources Research Institute (NaCRRI), Uganda.

6.2.2 Evaluation of S_1 progeny for harvest index and dry matter content

At harvest, which coincided with 11 MAP, four innermost plants per clone were uprooted and used for phenotypic assessments. Roots were separated from the vegetative harvestable biomass (leaves, stems and original planting stake) and HI was computed for each clone following the procedure outlined by Kawano (1990).

Estimation of DMC in the root samples was based on the oven dry method. Briefly, collected roots were washed, peeled and chopped into 1 cm thick pieces to a total weight of 200 g (fresh weight). Samples were dried to constant weight in an oven that was maintained at 72^{o}C. Upon attainment of constant weight (48 h), samples were immediately weighed (dry weight). Percentage DMC was computed by dividing the dry weight by the fresh weight and multiplying by 100. Depending on number of samples, logistics and objective of the experiment, DMC weight measurements can be done using peeled or unpeeled root samples. Because samples for evaluation were modest and DMC in parenchyma was of interest, peeled root samples were used for evaluations. The specific gravity method for determination of DMC (Kawano et al., 1987) was not used because some clones could not raise the required 3-5 kg per sample.

Evaluations for HI and DMC were based on unreplicated single row plots. As indicated in Chapter 3, lack of adequate quality planting material was the major reason for lack of replication, hence the decision to evaluate more individuals in unreplicated trials as opposed to evaluation of a few individuals in replicated trials. The ID phenotypically observed in some progeny significantly limited the number of individuals that generated sufficient and good quality planting material. This study's aim was to obtain initial insights into the effect of inbreeding on key agronomic traits of cassava in order to objectively define future cassava inbreeding activities and thus unreplicated single row field trials were established.

6.2.3 Extraction of starch from the S_1 cassava progeny

Harvested cassava roots were used for starch analysis. Cassava starch extraction was carried out using the method described by Benesi (2005). Fresh root samples were washed, peeled and chopped to about 1 cm^3 cubes. Chopped roots were pulverised in a 4 l capacity high speed industrial blender for 5 min. The generated pulp was filtered through a muslin cloth and the filtrate allowed to stand for 4 h to facilitate starch sedimentation. The top liquid was decanted and discarded. The starch was air-dried on aluminium pans overnight at room temperature. Each S_1 clone, including the parents, was represented by two samples. This cassava starch extraction process is simple, rapid and upon settling, the starch is free from any colour, impurities and contamination from proteins or fats and has been used in previous studies (Ceballos et al., 2007; Sánchez et al., 2009).

6.2.4 Determination of amylose content using the colorimetric method

Quantification of amylose in starch samples can take various forms (Gèrard et al., 2001). Initially, three methods, concanavalin A which is a commercial kit from Megazyme (Wicklow, Ireland), high performance size exclusion chromatography (HPSEC) and the iodine colorimetric methods were used. Since the iodine colorimetric method produced consistent results it was used for determining amylose content. The colorimetric amylose procedure has frequently been used in cassava (Ceballos et al., 2007; Randhika and Moorthy, 2008; Sánchez et al., 2009).

Using a sensitive balance, 20 mg of the starch flour sample was accurately weighed and dissolved by heating in 1 M sodium hydroxide for 30 min in a water bath maintained at 95^{o}C. When dissolved, this solution was diluted to a concentration of 5 mg/ml by addition of deionised water. Aliquots of this solution (0.1 ml) were diluted with 5 ml of trichloroacteic (v/v) acid (0.5% concentration) and 0.05 ml of iodine solution (0.01 M). The contents were mixed and the absorbance of a sample of this solution read at 620 nm using a spectrophotometer. Because of the lack of purified cassava starch standards, inferences on amylose in the samples had to be made from a standard curve generated from purified potato starch that contained 100% amylose.

Standard curves obtained from purified amylose and amylopectin extracted from potato tubers have previously been used to infer amylose content in cassava (Ceballos et al., 2007). In this study, purified amylose (100%) from potato was serially diluted and used to generate a standard curve for the estimation of the amylose content in the cassava starch samples. The standard curve was generated from different concentrations of amylose: 0, 10, 20, 30, 40, 50, 60, 70 and 80%, by diluting with trichloroacteic acid (0.5%; v/v). Absorbance readings were done as described above. Three readings were taken for each dilution and the mean used.

6.2.5 Data analysis

The number of progeny evaluated varied among families, making the data largely unbalanced. Summary statistics (range, mean, variance and skewness) for DMC and HI were computed for each family. For amylose content, the dataset was subjected to ANOVA using the unbalanced treatment structure in Genstat. The amylose data were further subjected to linear mixed model analysis for the estimation of variance components using the restricted maximum likelihood (REML) method. For this analysis, replicates were considered as fixed, while families and S_1 progeny were considered as random factors. ID was estimated for both HI and DMC as a percentage of the S_0 average. ID = [(s_0 mean – s_1 mean)/s_0 mean] x 100. Therefore, the lower the ID value, the lower the depression (Rojas et al., 2009).

6.3 Results

6.3.1 Segregation of harvest index and dry matter content in S_1 progeny

Data on HI of the different S_1 progeny is presented in Table 6.1. The lowest average HI (0.08) was observed in progeny derived from the parental genotype Bamunanika, while the highest (0.42) was observed in progeny derived from the parental genotype MH95/0469 (Table 6.1). Highest variability in HI as reflected by the variance, was observed in progeny derived from parental genotype 95/SE-00036, which ranged from 0-0.69, while the lowest was observed in progeny derived from TMS 30572, which ranged from 0.07-0.50 (Table 6.1).

Progeny derived from parental genotypes TMS 30572, I92/00067 and Bamunanika were positively skewed, while progeny derived from parental genotype 95/SE-00036, NASE 4 and MH95/0469 were negatively skewed (Table 6.1). Most of the progeny had HI values below 0.5. Only TMS 30572 and Bamunanika had HI values less than 0.5 for all their S_1 progeny (Table 6.1). For HI, ID varied among the different families ranging from 15.7 to 83.3% (Table 6.1). Highest ID was observed in progeny derived from Bamunanika (83.3%). Only progeny derived from parental genotypes 95/SE-00036, NASE 4 and MH95/0469 showed ID of less than 20% (Table 6.1).

The data for root DMC of the S_1 progeny in the different families is presented in Table 6.2. DMC varied both between and within the families. The lowest average DMC (28.5%) was recorded for the progeny derived from parental genotype Bamunanika, while the highest (35.1%) was recorded for progeny derived from TMS 30572. Highest variability as reflected by the within family variance in DMC content, was observed in progeny of 95/SE-00036 that ranged between 11-42% and lowest within progeny of parental genotype NASE 4, which ranged from 26-33% (Table 6.2).

Table 6.1 Variation in harvest index in S_1 cassava progeny generated from six genotypes[1]

Family	Parent	S_1 progeny	Min	Max	Mean	Variance	Skewness	ID
MH95/0469	0.52	16 (5)	0.14	0.66	0.42	0.017	-0.57	19.23
NASE 4	0.48	16 (5)	0.00	0.62	0.39	0.024	-1.11	18.75
TMS 30572	0.57	23 (0)	0.07	0.50	0.28	0.009	0.03	50.87
I92/00067	0.48	18 (2)	0.15	0.57	0.34	0.014	0.33	29.16
Bamunanika	0.48	28 (0)	0.00	0.35	0.08	0.016	0.96	83.33
95/SE-00036	0.38	30 (1)	0.00	0.69	0.32	0.027	-0.49	15.78

[1] Parent represents harvest index values for respective non-inbreds; S_1 progeny represents individuals evaluated; numbers in parentheses indicate number of progeny with harvest index values > 0.5; Min and Max indicate minimum and maximum harvest index, respectively. ID = inbreeding depression estimated as $[(s_0 \text{ mean} - s_1 \text{ mean})/s_0 \text{ mean}] \times 100$. The confidence interval associated with the data at 95% ranged from 0.042 to 0.082.

With the exception of progeny from the parental genotype MH95/0469 which had positive skewness (0.20), all other parental genotypes had progeny data that was negatively skewed (Table 6.2). The data further indicated (with the exception of S_1 progeny from TMS 30572) that over 60% of the progeny had DMC values of less than 35%. ID for DMC varied between 2.0 to 23.8%, with the highest (23.8%) observed in progeny from I92/00067; progeny of parental genotypes MH95/0469 and 95/SE-00036 had ID in DMC of only 2% (Table 6.2). Thus, within the S_1 progeny, the highest ID was observed in HI compared to DMC.

Table 6.2 Variation in root dry matter content in S_1 cassava progeny generated from six parental genotypes[1]

Family	Parent	S_1 progeny	Min	Max	Mean	Variance	Skewness	ID
MH95/0469	32.0	13 (4)	24.0	38.0	31.3	20.2	0.20	2.1
NASE 4	35.0	8 (0)	26.0	33.0	30.1	5.26	-0.39	14.0
TMS 30572	41.0	22 (11)	20.0	43.0	35.1	32.9	-0.89	14.3
I92/00067	39.0	17 (2)	15.0	40.0	29.7	41.0	-0.96	23.8
Bamunanika	37.0	9 (1)	18.0	38.0	28.5	32.5	-0.23	22.9
95/SE-00036	34.0	27 (11)	11.0	42.0	33.3	42.8	-1.67	2.0

[1] Parent represents DMC values for respective non-inbreds; S_1 progeny represents individuals evaluated; numbers in parentheses indicate number of progeny with DMC > 35%; Min and Max indicate minimum and maximum DMC respectively. ID = inbreeding depression estimated as $[(s_0 \text{ mean} - s_1 \text{ mean})/s_0 \text{ mean}] \times 100$. The confidence interval associated with the data at 95% ranged from 1.918 to 4.383.

6.3.2 Variation in amylose content in S_1 progeny

ANOVA data for amylose content in the S_1 cassava families is presented in Table 6.3. The amylose content varied significantly between cassava families and progeny, with most of the variation recorded within the progeny. The lowest amylose content (11.8%) was recorded in S_1 progeny from 95/SE-00036, while the highest (34.2%) was recorded from S_1 progeny from MH95/0469 (Table 6.4).

However, when family averages were compared, most of the progeny had relatively higher amylose content compared to the non-inbred parental genotypes (Table 6.4). The highest variability as reflected by the sample variance in amylose content was observed in progeny derived from MH95/0469, while the lowest was observed in progeny derived from NASE 4. With exception of progeny from parental genotype NASE 4 and 95/SE-00036, which had negative skewness, all other progeny had positive skewness for amylose content.

Table 6.3 Analysis of variance for amylose content in S_1 cassava families[1]

Source of variation	df	SS	MS	Variance component
Replication	1	4.548	4.548	-
Family	5	527.814	105.563*	2.651 (2.196)
Progeny	107	2707.260	25.301*	10.112 (1.772)
Residual	112	577.649	5.158	-

[1] df = degrees of freedom; SS = sum of squares; MS = mean squares; figures in parentheses are standard errors associated with the variance components; * $P \leq 5\%$.

Table 6.4 Variation in amylose content in six S_1 cassava families[1]

Family	Parent	Progeny performance					
		S_1 No	Min	Max	Mean	Variance	Skewness
MH95/0469	19.1	11	17.2	34.2	23.7	29.50	0.75
NASE 4	18.3	15	12.5	22.1	18.1	5.96	-0.53
TMS 30572	14.7	22	12.2	27.4	19.0	16.20	0.45
I92/00067	17.2	21	12.9	25.6	19.2	9.41	0.016
Bamunanika	19.1	16	14.7	26.0	20.2	12.50	0.088
95/SE-00036	16.2	22	11.8	22.8	18.2	10.19	-0.595

[1] Parent represents amylose content of non-inbreds; S_1 No represents number of S_1 individuals evaluated; Min and Max indicate minimum and maximum amylose respectively. The confidence interval associated with the data at 95% ranged from 1.35 to 3.65.

6.4 Discussion

A major objective of this study was to obtain initial insights into the effects of inbreeding in cassava to guide future cassava inbreeding efforts. This study was not meant to quantify the extent of ID in cassava. Accurate quantification of ID will require that a S_0 population be self-pollinated to obtain a representative S_1 population that is evaluated and compared to the S_0 population, as conducted in maize (Lamkey and Smith, 1987; Pacheco et al., 2002). In this study however, six cassava genotypes (not populations) were self-pollinated and the resultant S_1 progeny evaluated for two fitness-related traits (HI and DMC) and one quality trait (amylose content).

HI varied both within and between families, with progeny from parental genotype Bamunanika having the lowest (0.08) and progeny from MH95/0469 having the highest (0.42) average values. HI reflects the efficiency of dry matter distribution (Cock et al., 1979) and thus higher HI values are desirable. With inbreeding an average reduction of 36% in HI was recorded when six non-inbred parents were compared to the S_1 progeny. This indicated that S_1 progeny yielded only 64% of the non-inbred HI. Inbreeding of inherently heterozygous plants results in general loss of vigour, a phenomenon called ID which is expected to be more severe in early than later generations because during the first inbreeding generation, 50% heterozygosity is lost (Altenburg, 1957; Wricke and Weber, 1986).

ID could therefore explain the low HI observed in the S_1 inbreds. However, its expression varied among the parental genotypes and within the inbreds as depicted by the range, mean and skewness values of HI. Based on the HI data, progeny from Bamunanika appeared to be most affected by ID. In the field, these progeny had poor vigour and general phenotype compared to other progeny.

Though there are limitations to compare the performance of S_1 progeny to a few selected non-inbred cassava individuals (a S_0 non-inbred cassava population is preferred because it does not overestimate ID) results indicate a general reduction in HI with inbreeding. Rojas et al. (2009) estimated ID in cassava using eight families. The authors observed that average ID for HI was 26.5% ranging from 16.6-43% in the different families. These authors acknowledged the limitation of not using a non-inbred S_0 population, which could have overestimated the ID. The current study recorded a slightly higher average reduction for HI (36%). This could be a result of overestimation or because totally different germplasm i.e. Latin American versus African is being compared. It could also be due to the fact that small family sizes were used for the estimation. This can only be resolved by further detailed studies that compare different cassava populations.

However, some S_1 inbreds from parental genotypes MH95/0469, NASE 4, 95/SE-00036 and I92/00067 had higher HI values than their respective non-inbred parents suggesting that they did not succumb to ID and/or tolerated inbreeding. Elsewhere, studies have indicated that although inbreeding does not cause change in gene frequencies, it changes genotypic frequencies in the offspring, which once changed, affects breeding values and dominance deviations (Wricke and Weber, 1986). This phenomenon could partly explain the relatively high HI in some of the inbreds. Walsh (2005) observed that inbreeding provides an opportunity to partially exploit non-additive (dominance and epistatic) variance, which could further explain higher HI in some of the inbreds.

Studies conducted in maize established that variance of dominance deviations of inbred lines were 1.6-3.3 times higher than the variance of dominance deviations for non-inbred maize individuals for key productivity traits (Edwards and Lamkey, 2002). Moreover, significant specific combining ability that is indicative of dominance variance has been reported for HI in cassava (Cach et al., 2005; 2006; Calle et al., 2005). Khehra et al. (1997) suggested that when most of the genetic variation is non-additive, it is desirable to use highly inbred parents for trait improvement.

In alfalfa, non-additive gene effects have been reported to be more important in inbred relatives than in non-inbred individuals (Gallais, 1984), a finding which further justifies the need to advance inbreeding in cassava to improve HI, as it is one trait which has not been selected for extensively by national breeding programmes in SSA.

Root DMC content varied both within and between inbreds, with progeny derived from Bamunanika having the lowest average DMC (28.5%) and progeny derived from TMS 30572 having the highest value (35.1%). Higher root DMC is a major breeding objective of many breeding programmes and significant progress in its improvement has been attained through hybridisation schemes involving both domesticated cassava (Kawano, 2003; Ceballos et al., 2004) and wild relatives (Ojulong et al., 2008).

Average DMC in the six non-inbred cassava genotypes was 36.3%, while in S_1 progeny it was 31.3%. Hence, with introduction of inbreeding, an average reduction of 13.2% in DMC was observed. Over 60% of the inbreds (with exception of progeny from TMS 30572) had DMC below 35%, a finding which could point to ID as already indicated above for HI. The differences in range, mean and skewness of DMC in the different families is indicative of their varied response to ID. Rojas et al. (2009) recorded an average ID for DMC of 5.3%, ranging between 0.3-8.7%. This difference between these values and the ones reported in the current study can be explained by the same reasoning purported for HI above.

In related studies conducted in maize, ID was observed in productivity traits that included grain yield, ear height, plant height, days to mid pollen and days to mid silk (Edwards and Lamkey, 2002). In this study however, some inbred lines within the families of I92/00067, TMS 30572, 95/SE-00036, MH95/0469 and Bamunanika, had DMC values well above those of their respective non-inbred parents indicating that they did not succumb to ID and/or tolerated inbreeding.

Since significant GCA estimates for DMC, indicating preponderance of additive variance, have been reported in cassava (Cach et al., 2005; 2006; Jaramillo et al., 2005), it is possible that increased additive gene effects accounted for the relatively higher DMC content in some of these S_1 inbreds. This is further supported by the fact that additive variance among progeny in cross-pollinated crops increases with inbreeding because additive genetic variance is the major component of the total genetic variance (Hallauer, 1992). This strategy has been utilised in grasses, where inbreeding to generate S_1 progeny is used in the selection of parents for hybridisation; the degree of ID reflects the amount of additive variance present (Sneep et al., 1979). It is therefore possible that inbred cassava can be used to further exploit the additive and/or dominance variance with the overall goal of increasing DMC in roots.

The amylose content in the cassava S_1 inbreds varied both within and between inbreds, with most of the variation within inbreds. Starch, which constitutes the largest amount of the dry root weight of cassava (IITA, 1990), is exclusively composed of amylose and amylopectin polymers (Roger et al., 1999). The relative proportion of these polymers greatly influences the physiochemical properties of starch and its nutritional and industrial significance (Matueev et al., 2001; Charles et al., 2005; Aryee et al., 2006). It is for this reason that analysis of amylose levels in cassava was done after one generation of inbreeding.

Firstly, the data seemed to suggest that inbreeding of parental genotypes TMS 30572, 95/SE-00036 I92/00067 and NASE 4, led to the generation of a few S_1 inbreds with unusually low (< 13%) levels of amylose. Secondly, with the exception of S_1 progeny from NASE 4, all other progeny had amylose levels relatively higher than the parental genotypes. Whether or not additive or non-additive gene effects were responsible for this increase in amylose content is unknown. This is an aspect which future studies can establish, as to date no studies have been conducted on the inheritance of proportion of amylose in cassava starch.

In parallel, it will be equally important to know the cause of the decrease in amylose content (in some S_1 progeny) with inbreeding, as it cannot simply be ascribed to ID because it is not a fitness-related trait. In related studies conducted at CIAT that involved generation of several partial inbreds from different genotypes, one S_1 clone (AM206-5) had amylose-free starch, which was the first reported naturally occurring mutation in cassava (Ceballos et al., 2007). Another study that evaluated over 4000 non-inbred cassava, found that no more than 1.5% of the samples had amylose values less than 17.5 or greater than 24.5% (Sánchez et al., 2009). In the present study, no amylose-free cassava was identified. The lowest amount of amylose registered in the S_1 progeny was 11.8%, a finding which should inspire further cassava inbreeding in the search for novel cassava starches. A major justification for this is based on the fact that only a few S_1 progeny were evaluated and yet most variability was found within the S_1 progeny. In addition, it would also be rational to induce and/or enhance expression of mutations that inhibit production of amylose, as these have huge impacts on starch phenotypes, are not influenced by environment and are easy to handle from a breeding point of view.

Amylose content in non-inbred cassava is variable as observed in previous studies: 15.2-26.5% (Sánchez et al., 2009) and 17.1-24.9% (Sanni et al., 2008). Cassava starch with up to 28.8% (Zaidul et al., 2007) and 44% amylose (Aryee et al., 2006) has been reported. Though these studies used different methodologies and genotypes, they all indicated considerable variation in amylose content in non-inbred cassava. The current study that evaluated partial inbreds also established considerable variation in amylose content, which certainly can be exploited to increase the competitiveness of cassava starch.

6.5 Conclusions

This study presented the first report on cassava inbreeding in Uganda, information of which will be important for the general cassava breeding community that is increasingly showing interest in introducing inbreeding in cassava.

Though the number of individuals evaluated were relatively few, they provided general trends on the effect of inbreeding on two environment-fitness traits (DMC and HI) and one quality trait, amylose content. ID was evident in some of the inbreds as demonstrated by the relatively lower values for HI and DMC. In parallel, inbreeding resulted in the generation of vigorous S_1 progeny (high DMC and HI), which appeared to have benefited from either additive or non-additive genetic effects or a combination of the two. Earlier studies have established that if dominant alleles are not linked with recessive alleles on the same chromosome, exceptional vigour can be expected in the inbreds (Altenburg, 1957; Begg, 1959).

Though results were based on the comparison of a few S_1 progeny they indicated that ID was more pronounced in HI than DMC, a finding which was consistent with other findings that evaluated a higher number of S_1 progeny (Rojas et al., 2009). A major limitation of using non-inbred parents as compared to a non-inbred population is the overestimation of ID. Systematic studies that will quantify ID in cassava should consider using representative S_0 populations, which should be selfed to generate representative S_1 populations that will be evaluated and compared to the S_0 populations. These populations can comprise of elite and local cassava genotypes from Africa, Asia and/or Latin America. Traits for evaluations can include fresh root yield, foliage yield, plant height and plant health. This approach will be a replica of what is used in maize and will provide fair estimates of ID in key agronomic traits of cassava.

Deliberate starch quality modification in the form of reconstituting amylose and amylopectin amounts in cassava through inbreeding, is another aspect that this study aimed at. The premise for this is based on the fact that composition of amylose and amylopectin will impart specific characteristics that will define starch quality and utilisation (Charles et al., 2005). For instance, foods with slow digestive starch that depend on the starch structure and composition (Zhang et al., 2006), are nutritionally desirable for food products (Ludwig, 2002).

Hence, through systematic inbreeding and evaluation, cassava starch composition can be drastically changed to lend itself to utilisation as a quality healthy food, which is currently dominated by maize starch.

6.6 References

Altenburg, E., 1957. Genetics. Revised Edition. Henry Holt and Company, Inc., pp. 496.

Aryee, F.N.A., I. Oduro, W.O. Ellis, and J.J. Afuakwa, 2006. The physicochemical properties of flour samples from the roots of 31 varieties of cassava. Food Control 17: 916-922.

Begg, C.M.M., 1959. An introduction to genetics. The English Universities. Press Ltd, 102 Newgate Street, London, UK, pp. 291.

Benesi, I.R.M., 2005. Characterisation of Malawian cassava germplasm for diversity, starch extraction and its native and modified properties. PhD Thesis, University of the Free State, South Africa, pp. 199.

Cach, N.T., J.C. Perez, J.I. Lenis, F. Calle, N. Morante, and H. Ceballos, 2005. Epistasis in the expression of relevant traits in cassava (*Manihot esculenta* Crantz) for subhumid conditions. Journal of Heredity 96:586-592.

Cach, N.T., J.I. Lenis, J.C. Perez, N. Morante, F. Calle, and H. Ceballos, 2006. Inheritance of useful traits in cassava grown in subhumid conditions. Plant Breeding 125:177-182.

Calle, F., J.C. Perez, W. Gaitan, N. Morante, H. Ceballos, G. Llano, and E. Alvarez, 2005. Diallel inheritance of relevant traits in cassava (*Manihot esculenta* Crantz) adapted to acid-soil savannas. Euphytica 144:177-186.

Ceballos, H., C.C. Iglesias, J.C. Pérez, and A.G.O. Dixon, 2004. Cassava breeding: opportunities and challenges. Plant Molecular Biology 56:503-515.

Ceballos, H., T. Sáchez, N. Morante, M. Fregene, D. Dufour, A.M. Smith, K. Denyer, J.C. Pèrez, F. Calle, and C. Mestres, 2007. Discovery of amylose-free starch mutant in cassava (*Manihot esculenta* Crantz). Journal of Agricultural and Food Chemistry 55: 7469-7476.

Charles, A.L., Y.H. Chang, W.C. Ko, K. Sriroth, and T.C. Huang, 2005. Influence of amylopectin structure and amylose content on the gelling properties of five cultivars of cassava starches. Journal of Agricultural and Food Chemistry 53:2717-2725.

Cock, H., D. Franklin, G. Sandoval, and P. Juri, 1979. The ideal cassava plant for maximum yield. Crop Science 19:271-279.

East, E.M., 1908. Inbreeding in corn. Annual report of Connecticut Agricultural Experiment Station 31:419-428.

Edwards, J.W., and K.R. Lamkey, 2002. Quantitative genetics of inbreeding in a synthetic maize population. Crop Science 42:1094-1104.

Falconer D.S., and T.F.C. Mackay, 1996. Introduction to quantitative genetics, 4th edition. Longman: Harlow, pp. 438.

Gallais, A., 1984. An analysis of heterosis vs. inbreeding effects with an autotetraploid cross-fertilised plant *Medicago sativa* L. Genetics 105:123-137.

Gèrard, C., C. Barron, P. Colonna, and V. Planchot, 2001. Amylose determination in genetically modified starches. Carbohydrate Polymers 44:19-27.

Glènmin, S., T. Bataillon, J. Ronfort, A. Mignot, and I. Olivieri, 2001. Inbreeding depression in small populations of self-incompatible plants. Genetics 159:1217-1229.

Hallauer, A.R. 1992. Use of genetic variation for breeding populations in cross pollinated species. In: Stalker, H.T., and J.P. Murphy (Eds.), Plant Breeding in the 1990s. Proceedings of the symposium on plant breeding in the 1990s held at North Carolina State University Raleigh, NC, pp. 37-67. CABI International Wallington Oxon OX108DE, UK.

IITA, 1990. Cassava in tropical Africa. A reference manual. International Institute of Tropical Agriculture, Ibadan, Nigeria, pp. 176.

Jaramillo, G., N. Morante, J.C. Perez, F. Calle, H. Ceballos, B. Arias, and A.C. Bellotti, 2005. Diallel analysis in cassava adapted to the midaltitude valleys environment. Crop Science 45:1058-1063.

Jennings, D.L., and C.A. Iglesias, 2002. Breeding for crop improvement. In: Hillocks, R.J., J.M. Thresh, and A.C. Belloti (Eds.), Cassava: Biology, Production and Unitilsation, pp. 149-166. CABI Publishing.

Kawano, K., 1990. Harvest index and evolution of major food crops cultivars in the tropics. Euphytica 46:195-202.

Kawano, K., 2003. Thirty years of cassava breeding for productivity-biological and social factors for success. Crop Science 43:1325-1335.

Kawano, K., A. Amaya, P. Daza, and M. Rios, 1978. Factors affecting efficiency of hybridization and selection in cassava. Crop Science 18:373-376.

Kawano, K., W.M.F. Goncalves, and U. Cenpukdee, 1987. Genetic and environmental effects on dry matter content of cassava root. Crop Science 27:69-74.

Khehra, A.S., B.S. Dhillon, and D.S. Multani, 1997. Maize. In: Bahl, P.N., P.M. Salimath, and A.K. Mandal (Eds.), Genetics, cytogentics and breeding of crop plants, Volume 2, Cereal and commercial crops, pp. 145-196. Science Publishers Inc., USA.

Lamkey, R.R., and O.S. Smith, 1987. Performance and inbreeding depression of populations representing seven eras of maize breeding. Crop Science 27:695-699.

Ludwig, D.S.J., 2002. Physiological mechanisms relating to obesity, diabetes, and cardiovascular disease. Journal of American Medical Association 287:2414-2423.

Matueev, Y.I., J.J.G. van Soest, C. Nieman, L.A. Wasserman, V.A. Protserov, M. Ezernitskaja, and V.P. Yuryev, 2001. The relationship between thermodynamic and structural properties of low and high amylose maize starches. Carbohydrate Polymers 44:151-160.

Ojulong, H., M.T. Labuschagne, L. Herselman, and M. Fregene, 2008. Introgression of genes for dry matter content from wild cassava species. Euphytica 164:163-172.

Pacheco, C.A.P., M.X. dos Santos, C.D. Cruz, S.N. Parentoni, P.E.G. de Oliveira, E.E. Gama, A.E. da Silva, H.W.L. de Carvalho, and P.A. Jr. Vieira, 2002. Inbreeding depression of 28 maize open pollinated varieties. Genetics and Molecular Biology 25: 441-448.

Randhika, G.S., and S.N. Moorthy, 2008. Effect of sodium dodecyl sulphate on the physicochemical, thermal and pasting properties of cassava starch. Starch 60:87-96.

Roger, P., L.A. Bello-Perez, and P. Colonna, 1999. Contribution of amylose and amylopectin to the light scattering behaviour of starches in aqueous solution. Polymer 40: 6897-6909.

Rojas, M.C., J.C. Pérez, H. Ceballos, D. Baena, N. Morante, and F. Calle, 2009. Analysis of inbreeding depression in eight S1 cassava families. Crop Science 49:543-548.

Sánchez, T., E. Salcedo, H. Ceballos, D. Dufour, G. Mafla, N. Morante, F. Calle, J.C. Pèrez, D. Debouck, G. Jaramillo, and I.X. Moreno, 2009. Screening of starch quality traits in cassava (*Manihot esculenta* Crantz). Starch 61:12-19.

Sanni, L.O., A.A. Adebowale, B. Maziya-Dixon, and A.G.O. Dixon, 2008. Chemical composition and pasting properties of CMD resistant cassava clones planted at different locations. Journal of Food Agriculture and Environment 6:97-104.

Sneep, J., B.R. Murty, and H.F. Utz, 1979. Current breeding methods. In: Sneep, J., A.J. T. Hendriksen, and O. Holbek (Eds.), Plant breeding perspective: Centennial publication of Koninklijk Kweekbedrijf en Zaadhandel D.J. van der Have 1879-1979, pp. 104-233. Centre for Agricultural publishing and documentation, Wageningen, Netherlands.

Walsh, B., 2005. The struggle to exploit non-additive variation. Australian Journal of Agricultural Research 56:873-881.

Wricke, G., and W.E. Weber, 1986. Quantitative genetics and selection in plant breeding. Walter de Gruyter, Berlin, pp. 406.

Zaidul, I.S.M., N.A. Nik Norulaini, A.K. Mohd Omar, H. Yamauchi, and T. Noda, 2007. RVA analysis of mixtures of wheat flour and potato, sweet potato, yam, and cassava starches. Carbohydrate Polymers 69:784-791.

Zhang, G., Z. Ao, and B.R. Hamaker, 2006. Slow digestion property of native cereal starches. Biomacromolecules 7:3252-3258.

CHAPTER 7

GENERAL DISCUSSION AND RECOMMENDATIONS

Cassava is a widely grown crop in SSA, where it plays a dual role of being a staple and an income generating crop. However, the crop's long growth period coupled with the tropical environment in SSA, provide for a plethora of biotic and abiotic constraints that limit its optimal productivity. As witnessed in other major crops where science-led initiatives offered solutions to production constraints, the same should be replicated in cassava. Data presented in this thesis revealed important aspects within the realms of cassava genetics, which can be of use for the improvement of the crop. With the help of multi-player interventions tailored towards cassava improvement, this information will most likely benefit rural communities that primarily depend on cassava for their livelihood in Uganda, Kenya, Tanzania, DRC, Madagascar, Rwanda, Mozambique and other countries not included in the study, but who are important cassava producers. The principal anticipated benefit will be development and adoption of cassava varieties that will enhance food security and/or generate income to rural communities.

This study analysed 29 morphological cassava traits in 1091 cassava genotypes from six countries. Unlike most previous studies that analysed only a few above-ground traits, in this study, both above-ground and below-ground traits were analysed and used to infer genetic relationships. Though the 29 qualitative traits did not provide good discriminatory power amongst the regional cassava germplasm collections compared to the genomic molecular markers, they can still be used at a national level to characterise germplasm and/or any new cassava varieties deemed suitable for release. The lack of good discriminatory power of the qualitative traits is not surprising considering that any two individuals being compared can have the same phenotype (i.e. green periderm) but different genotype (i.e. heterozygous versus homozygous at the same locus). Nonetheless, data generated during this and other studies should be used in the establishment of a regional cassava database. The regional cassava phenotypic database should not be static, but rather phenotypic data collected from new genotypes should be added.

This initiative should jumpstart the urgent need for the establishment of data retrieval and/or storage databases, as this has been largely ignored in most, if not all, NARS involved in this study. An efficient breeding programme should have a sound information management system, which can now be initiated in the NARS of Uganda, Kenya, Tanzania, DRC, Rwanda and Madagascar using the phenotypic data generated from their respective germplasm collections.

Quantification of variation in agronomically important traits is particularly important in designing breeding programmes. For instance, large variation indicates that selection can proceed immediately, while small variation justifies the need for finding germplasm sources elsewhere so as to increase genetic variation in the population. Data generated for HI, DMC and LR suggested that considerable variation for these traits in both local and elite germplasm is available within the NARS. Because local varieties pre-dominate cassava acreage in these countries, efforts to improve their productivity through genetic improvement of HI and DMC could be desirable. To begin with, progeny derived by crossing local x local and/or local x elite genotypes could form heterogeneous populations on which selection can be applied. The presence of additive variance and availability of refined evaluation protocols for both traits makes their selection relatively easy for most NARS breeding programmes.

Previous breeding efforts in some major crops largely depended on a relatively small sample of locally adapted cultivars resulting in a reduction of genetic variability as variety improvement advanced. It is therefore important that NARS cassava breeding teams consider the utilisation of local varieties in their respective hybridisation schemes. Among the positive results that this initiative will bring, is the combining of desirable agronomic traits with culinary qualities. However, selection of parents for hybrdisation will require that potential parental lines (selected based on data generated from this thesis) be evaluated in replicated trials by the respective NARS, as they will have been reduced to manageable numbers.

The genotypic data on 1401 cassava genotypes from seven NARS, established limited gene differentiation i.e. only three populations were quantified. The hypothesised minimum and maximum populations were respectively, seven when only the seven countries were considered and 14 when the sub-populations of local versus elite genotypes from each country were considered. Most variation (> 89%) was detected within individuals with the rest being distributed among populations (country of origin) and sub-populations (local and elite genotypes). This distribution of variation within individuals limited the structuring of cassava into distinct populations.

The observed clustering of some local with elite and local with local genotypes indicated shared ancestry, which could have resulted from the regional cassava breeding scheme that was established in the 1930s at Amani, Tanzania. This regional breeding programme selected germplasm resistant to CMD and distributed it to the neighbouring countries for further evaluation and selection. It is also possible that some elite varieties may be hybrids from elite x local varieties and/or germplasm indicated as local may in fact be a local x elite cross. This strongly illustrates the value of molecular markers towards both identification of unique variability and selection of parents with widest additive variation for hybridisation.

However, this high genetic variability within cassava is particularly important in this era where the natural resource base, specifically soil and underground water, is becoming depleted. Emerging biotic constraints coupled with climate change that may be associated with increased temperatures, greater evapotranspiration and increased drought incidence, add another gloomy picture. The detected high levels of genetic variability provide cassava with the necessary plasticity to respond to these imminent threats. It is however, important that efforts be made to conserve this genetic variability, as currently limited efforts have been made at both national and regional level. Most NARS largely rely on field-based conservation methods, which are highly risky.

At the inception of the IPGRI in 1974, there was hope that for each major crop, efforts at international, regional and national level could work towards the conservation, distribution and utilisation of germplasm. Today, various crop-specific initiatives have been made, but at different levels. For cassava, significant progress has been made at the international level by both CIAT and IITA, where cassava genotypes are held in trust by the Food and Agriculture Organisation (FAO). The genebank at CIAT contains germplasm predominantly from the Americas and Asia, whereas germplasm from Africa is conserved largely at IITA. Unfortunately, germplasm from southern, eastern and central Africa represents approximately 2% of cassava accessions within the IITA genebank. This highlights the urgency for regional and/or national conservation efforts. Much of the germplasm used in this study is not actively conserved under a specific conservation programme. The genotypic data generated in this study can be used in setting priorities for cassava conservation at either regional or national level. However, it is important to note that conservation and use of genetic resources are two indivisible components of varietal improvement and hence should be considered together when setting priorities for cassava conservation. The data presented in this thesis should form the basis of a regional cassava conservation strategy. Development of a core collection that capature the widest possible variability is another principal output that can be generated from this data set.

The characterisation of SNPs in cassava revealed a frequency of one SNP every 121 bp. This is a relatively high frequency compared to other crops like soybean with one SNP every 2038 bp and thus justifies efforts to utilise SNPs as genetic markers in cassava. This high frequency is most likely due to cassava's heterozygous nature and short domestication history. However, when modest SNPs and SSRs were compared for diversity assessment, it was evident that more SNPs would be required to provide the same discrimination power among cassava individuals as provided by SSRs. This appears to be consistent with what was found in other crops like maize, where the two marker systems have been compared. This is largely because SNPs are bi-allelic, while SSRs are multi-allelic and more informative. SNPs are limited to a maximum PIC value of 0.5, while SSRs can have PIC values greater then 0.5.

Nonetheless, because of their high frequency, SNPs within the cassava genome offer excellent opportunities to be used as genetic markers. Non-synonymous substitution SNPs that result into a desired phenotype would form gene based markers and would be particularly useful for MAS. This will, however, entail extensive studies that will involve undertaking high resolution genetic linkage and/or association tests between sequence polymorphisms and desired phenotypes. Certainly, these gene-based markers, if identified, will make MAS more accurate and efficient.

Data generated on S_1 cassava progeny is one of the few reports on deliberate cassava inbreeding in SSA. Although field data on HI and DMC were from single-row unreplicated trials, it provided two major findings: 1) cassava inbreeding did not only lead to reduced productivity, but also resulted in increased productivity of some partial inbred individuals and 2) that the extent of ID varied among families. Whether or not similar trends will be observed in advanced generations is not known and will need specific studies to provide answers. The S_1 progeny evaluated in this study have been advanced to the S_2 for further evaluation.

During the course of evolution, many organisms accumulate deleterious alleles, which are concealed from expression by dominant alleles. Inbreeding exposes these alleles so that they can be selected against. Conversely, inbreeding can also reveal or accentuate the effects of advantageous genes by placing them in a homozygous state. Waxy cassava identified by CIAT in a S_1 clone is testimony to this fact. For these reasons, quantification of amylose in S_1 inbreds was done as part of this study. The lowest amylose content reported (11.8%) was recorded in the S_1 progeny from 95/SE-00036, while the highest (34.2%) was recorded from the S_1 progeny from MH95/0469, producing respective amylopectin compositions of 89.2% and 65.8%. Relative proportions of amylopectin and amylose are critical in determining starch functionality including viscosity, resistance to high temperatures, solubility, swelling index and sensitivity to shear stress. This therefore, makes starch modification through inbreeding a promising intervention towards the commercialisation of cassava starch.

In conclusion, information generated in this thesis is well-suited for both immediate and future use. For example, cassava genotypes with high DMC and/or HI can be considered for release after satisfactory evaluations have been done, as required by the variety release committees in the respective countries. Alternatively, they can be used as progenitors as already highlightened. Furthermore, this information can be used as supporting proof for any policy lobbying towards the establishment of a regional gene bank and the information provided should be used as a basis for determining conservation strategies. In addition, diversity information should be used within breeding programmes to maximise diversity among crosses to enhance the probablility of making significant genetic gains. Information on SNPs offers good insights into future SNP-based research in cassava and the positive benefits associated with inbreeding should be a motivation to advance inbreeding in cassava.

SUMMARY

Science-driven interventions aimed at increasing cassava's (*Manihot esculenta* Crantz), productivity will positively impart on communities that primarily depend on it. Data presented in this thesis contributed towards this goal. Specifically, this thesis examined phenotypic variation in cassava germplasm available within selected national breeding programmes, analysed patterns of allele frequency distribution in the cassava germplasm, characterised single nucleotide polymorphisms (SNPs) in the cassava genome and examined variation for two agronomic [root dry matter content (DMC) and harvest index (HI)] and one quality trait (amylose content) in S_1 cassava inbreds. The 29 qualitantive traits provided limited discrimination of cassava germplasm from Uganda, Kenya, Tanzania, Rwanda, Democratic Republic of Congo and Madagascar. However, significant differences in root DMC, HI, leaf retention (LR) and root cortex thickess were observed in this germplasm. Highest average DMC was registered in Uganda (37.7%) and lowest in Tanzania (30.1%), with the elite genotypes having a relatively higher DMC than local genotypes. Similary, highest average HI was observed in Uganda (0.60) and lowest in Kenya (0.32). This germplasm (1401 local and elite cassava genotypes) were furthermore genotyped using simple sequence repeat (SSR) markers. Genotypes were assigned to three distinct populations, with no appreciable gene differentiation (F_{ST} = 0.089) observed. Most genetic variation (> 89%) was detected within individuals. A total of 26 SNPs were identified from quality sequences of nine genes, giving an estimated frequency of one SNP for every 121 nucleotides. Nucleotide diversity ranged from 7.8 x 10^{-4} to 5.63 x 10^{-3}. Average haplotype-based polymorphism information content (PIC = 0.414) was higher than for individual SNP (PIC = 0.228). HI and DMC varied both within and between the cassava S_1 families. The highest reduction in HI was observed in S_1 progeny from Bamunanika (83.3%), while the highest reduction in DMC (23.8%) was observed in S_1 progeny from I92/00067. Amylose content varied significantly between cassava families and S_1 progeny, with most of the variation recorded within progeny.

Keywords: cassava inbreeding, cassava starch, heterozygosity, population structure, quantitative and qualitative traits

OPSOMMING

Wetenskaplik-gedrewe intervensies wat ten doel het om cassava (*Manihot esculenta* Crantz) se produktiwiteit te verbeter, sal ‘n positiewe impak maak op gemeenskappe wat primêr daarop staatmaak. Data wat in hierdie tesis voorgehou is, het tot hierdie doel bygedra. Hierdie tesis het spesifiek die fenotipiese variasie in cassava kiemplasma beskikbaar binne geselekteerde nasionale teelprogramme ondersoek, patrone van alleel frekwensie verspreiding binne die cassava kiemplasma geanaliseer, enkel nukleotied polimorfismes (ENP) in die cassava genoom gekarakteriseer en die variasie vir twee agronomiese [wortel droë materiaal opbrengs (DMO) en oesindeks (OI)] en een kwaliteitseienskap (amilose inhoud) in S_1 cassava ingeteelde lyne ondersoek. Die 29 kwalitatiewe eienskappe het tot beperkte diskriminasie van die cassava kiemplasma van Uganda, Kenia, Tanzanië, Rwanda, Demokratiese Republiek van die Kongo en Madagaskar gelei. Betekenisvolle verskille in wortel DMO, OI, blaarretensie (BR) en wortel korteksdikte is egter in hierdie kiemplasma waargeneem. Die hoogste gemiddelde DMO is in Uganda (37.7%) en die laagste in Tanzanië (30.1%) waargeneem, met die elite genotipes wat ‘n relatiewe hoër DMO as die plaaslike genotipes gehad het. Net so is die hoogste gemiddelde OI in Uganda (0.60) en die laagste in Kenia (0.32) waargeneem. Hierdie kiemplasma (1401 plaaslike en elite cassava genotipes) is verder genotipies getipeer deur van eenvoudige volgorde herhalende (EVH) merkers gebruik te maak. Genotipes is in drie duidelike populasies gegroepeer, sonder enige beduidende geen differensiasie (F_{ST} = 0.089). Die grootste gedeelte van die genetiese variasie (> 89%) is tussen individue waargeneem. ‘n Totaal van 26 ENPs is van kwaliteitsvolgordes van nege gene geïdentifiseer, wat ‘n geraamde frekwensie van een ENP vir elke 121 nukleotide gee. Nukleotied diversiteit het van 7.8×10^{-4} tot 5.63×10^{-3} gevarieer. Gemiddelde haplotied-gebaseerde polimorfiese informasie inhoud (PII = 0.414) was hoër as vir individuele ENP (PII = 0.228). OI en DMO het beide binne en tussen die cassava S_1 families gevarieer. Die grootste verlaging in OI was in S_1 nageslag van Bamunanika (83.3%) en die grootste verlaging in DMO (23.8%) in die S_1 nageslag van I92/00067 waargeneem. Amilose inhoud het betekenisvol tussen cassava families en S_1 nageslag gevarieer, met die grootste gedeelte van die variasie binne die nageslagte.

Sleutelwoorde: cassava inteling, cassava stysel, heterosigositeit, kwantitatiewe en kwalitatiewe eienskappe, populasiestruktuur

Appendix 1 Assayed SSR loci, annealing temperatures and primer sequences

SSR loci	Linkage group[1]	Left primer sequence	Right primer sequence
SSRY5	J	TGATGAAATTCAAAGCACCA	CGCCTACCACTGCCATAAAC
SSRY161	E	AAGGAACACCTCTCCTAGAATCA	CCAGCTGTATGTTGAGTGAGC
SSRY51	-	AGGTTGGATGCTTGAAGGAA	GGATGCAGGAGTGCTCAACT
SSRY147	-	GTACATCACCACCAACGGGC	AGAGCGGTGGGGCGAAGAGC
SSRY59	M	GCAATGCAGTGAACCATCTTT	CGTTTGTCCTTTCTGATGTTC
SSRY19	K	TGTAAGGCATTCCAAGAATTATCA	TCTCCTGTGAAAAGTGCATGA
SSRY155	-	CGTTGATAAAGTGGAAAGAGCA	ACTCCACTCCCGATGCTCGC
SSRY52	H	GCCAGCAAGGTTTGCTACAT	AACTGTCAAACCATTCTACTTGC
SSRY148	-	GGCTTCATCATGGAAAAACC	CAATGCTTTACGGAAGAGCC
SSRY38	G	GGCTGTTCGTGATCCTTATTAAC	GTAGTTGAGAAAACTTTGCATGAG
SSRY12	F/H	AACTGTCAAACCATTCTACTTGC	GCCAGCAAGGTTTGCTACAT
SSRY64	J	CGACAAGTCGTATATGTAGTATTCACG	GCAGAGGTGGCTAACGAGAC
SSRY63	H	TCAGAATCATCTACCTTGGCA	AAGACAATCATTTTGTGCTCCA
SSRY135	G	CCAGAAACTGAAATGCATCG	AACATGTGCGACAGTGATTG
SSRY102	M	TTGGCTGCTTTCACTAATGC	TTGAACACGTTGAACAACCA
SSRY171	C	ACTGTGCCAAAATAGCCAAATAGT	TCATGAGTGTGGGATGTTTTTATG
NS911	S	TGTTGTTCAGACGATGTCCAA	TTGAAGCAGTTATGAACCGT
SSRY151	-	AGTGGAAATAAGCCATGTGATG	CCCATAATTGATGCCAGGTT
SSRY100	K	ATCCTTGCCTGACATTTTGC	TTCGCAGAGTCCAATTGTTG
SSRY69	I	CGATCTCAGTCGATACCCAAG	CACTCCGTTGCAGGCATTA
SSRY182	M	GGAATTCTTTGCTTATGATGCC	TTCCTTTACAATTCTGGACGC
SSRY169	-	ACAGCTCTAAAAACTGCAGCC	AACGTAGGCCCTAACTAACCC
SSRY21	B/D	CCTGCCACAATATTGAAATGG	CAACAATTGGACTAAGCAGCA
SSRY110	L	TTGAGTGGTGAATGCGAAAG	AGTGCCACCTTGAAAGAGCA
SSRY9	D	ACAATTCATCATGAGTCATCAACT	CCGTTATTGTTCCTGGTCCT
SSRY181	K	GGTAGATCTGGATCGAGGAGG	CAATCGAAACCGACGATACA

[1] These SSR loci represented 13 linkage groups, with six being unmapped **(-).**

Appendix 2 Frequency of the major allele in cassava germplasm available within the national breeding programmes[1]

Loci	Tanzania		Uganda		Kenya		Rwanda		DRC		Madagascar		Mozambique	
	n	***p***	***n***	***p***	***n***	***p***	***n***	***p***	***n***	***p***	***n***	***p***	***n***	***p***
NS911	3	0.8038	3	0.7884	3	0.6273	4	0.5895	3	0.6958	3	0.6250	3	0.8608
SSRY100	7	0.2823	7	0.2653	6	0.3953	9	0.3049	7	0.3026	9	0.2432	6	0.3598
SSRY102	3	0.5970	3	0.5522	3	0.7329	2	0.7432	2	0.6250	3	0.6703	3	0.7407
SSRY12	5	0.4839	5	0.6004	5	0.2846	5	0.3277	5	0.3872	5	0.6087	5	0.4136
SSRY147	4	0.6825	4	0.7989	3	0.7670	3	0.7363	4	0.6220	4	0.5565	3	0.8902
SSRY148	5	0.4119	4	0.4844	6	0.3293	4	0.4972	4	0.5183	5	0.5000	3	0.6098
SSRY161	5	0.7013	4	0.8130	5	0.6287	3	0.7722	5	0.6646	4	0.7326	5	0.7727
SSRY181	5	0.6071	5	0.5536	5	0.7207	5	0.7024	5	0.6034	3	0.5082	4	0.7622
SSRY182	5	0.3946	5	0.4377	5	0.4643	4	0.4581	6	0.4030	6	0.3895	6	0.3902
SSRY21	7	0.4869	5	0.6250	7	0.7885	6	0.6503	6	0.7983	6	0.3234	5	0.4615
SSRY38	5	0.9211	3	0.9073	4	0.8098	3	0.8743	3	0.9042	2	0.7737	4	0.9268
SSRY5	4	0.5933	4	0.5896	4	0.6606	4	0.6154	6	0.3588	3	0.6980	3	0.5679
SSRY52	5	0.4811	5	0.5996	5	0.2683	5	0.3258	5	0.5206	4	0.6250	5	0.4250
SSRY59	3	0.5067	7	0.5792	4	0.6650	5	0.5497	8	0.4186	4	0.5000	6	0.4753
SSRY63	3	0.5680	5	0.5284	4	0.3460	4	0.5267	4	0.5563	3	0.5528	4	0.5875
SSRY69	8	0.3829	6	0.2755	7	0.2414	6	0.2898	6	0.3394	6	0.4543	6	0.4868
SSRY 19	8	0.2918	8	0.4173	8	0.2593	7	0.6311	10	0.5094	11	0.4081	7	0.3133
SSRY110	7	0.8182	6	0.8234	7	0.6996	6	0.6607	9	0.3879	5	0.7585	5	0.7683
SSRY135	5	0.3781	4	0.3171	6	0.3586	4	0.3087	6	0.2862	5	0.3373	7	0.4451
SSRY151	8	0.4029	7	0.4156	8	0.2905	8	0.2297	7	0.2590	6	0.3455	7	0.3413
SSRY155	4	0.5332	4	0.6346	4	0.5847	4	0.6156	6	0.4048	3	0.6836	4	0.6341
SSRY169	6	0.4136	5	0.7143	5	0.5830	5	0.7482	6	0.5353	5	0.8108	5	0.5610
SSRY171	7	0.6577	6	0.6348	5	0.5223	7	0.4637	4	0.4260	5	0.6491	4	0.6154
SSRY51	7	0.2904	5	0.4847	5	0.3906	5	0.6458	6	0.4430	6	0.3432	5	0.4557
SSRY64	7	0.4455	6	0.3585	6	0.3922	6	0.4420	6	0.5337	6	0.4000	5	0.5500
SSRY9	7	0.3877	7	0.6681	7	0.4954	7	0.6149	7	0.4781	6	0.3933	6	0.3537

[1]n = number of alleles detected and p = frequency of the major alleles

Appendix 3 Proportion of allele frequency distribution in cassava germplasm available with seven national breeding programmes

	Proportion (%) of allele frequency distribution						
Allele frequency	**Tanzania**	**Uganda**	**Kenya**	**Rwanda**	**DRC**	**Madagascar**	**Mozambique**
≤ 0.01	21.6783	15.7895	14.5985	14.5038	8.9041	20.3125	8.7302
0.02 – 0.09	22.3776	27.0677	26.2774	25.1908	39.0411	20.3125	34.9206
0.10 – 0.20	18.1818	23.3083	23.3577	27.4809	19.1781	18.7500	22.2222
0.21 – 0.30	15.3846	11.2782	18.9781	12.9771	15.0685	14.8438	7.9365
0.31 – 0.40	6.2937	3.0075	5.8394	4.5802	6.8493	11.7188	8.7302
0.41 – 0.50	6.2937	5.2632	1.4599	3.0534	2.0548	3.1250	6.3492
0.51 – 0.60	3.4965	5.2632	2.9197	2.2901	4.7945	2.3438	2.3810
0.61 – 0.70	3.4965	3.7594	3.6496	6.1069	2.7397	5.4688	2.3810
0.71 – 0.80	0.6993	2.2556	2.1898	3.0534	0.6849	2.3438	3.1746
0.81 – 0.90	0.6993	1.5038	0.7299	0.7634	0.6849	0.7813	1.5873
0.91 – 1.00	0.6993	0.7519	0.0000	0.0000	0.6849	0.7813	0.7937

Appendix 4 Roger's genetic distance between elite and local cassava genotypes from the seven NARS breeding programmes

	DRC-elite	DRC-local	Ke-elite	Ke-Local	Mad-elite	Mad-local	Moz-elite	Moz-local	Rw-elite	Rw-local	Tz-elite	Tz-local	Ug-elite	Ug-local
DRC-elite	0.0000													
DRC-local	0.2281	0.0000												
Ke-elite	0.2421	0.3371	0.0000											
Ke-Local	0.3840	0.3776	0.2836	0.0000										
Mad-elite	0.3776	0.3781	0.3550	0.3040	0.0000									
Mad-local	0.3512	0.3679	0.3307	0.2517	0.1397	0.0000								
Moz-elite	0.3469	0.3888	0.2925	0.4300	0.4800	0.4566	0.0000							
Moz-local	0.3858	0.3767	0.3605	0.3434	0.4045	0.3698	0.3681	0.0000						
Rw-elite	0.2346	0.3423	0.1587	0.3407	0.3761	0.3475	0.3218	0.3806	0.0000					
Rw-local	0.3413	0.2999	0.3205	0.2632	0.3493	0.3054	0.4293	0.3915	0.2752	0.0000				
Tz-elite	0.3819	0.3537	0.3130	0.2727	0.3283	0.2550	0.3991	0.2877	0.3541	0.3337	0.0000			
Tz-local	0.3402	0.3127	0.2741	0.1951	0.2751	0.2202	0.3896	0.2545	0.3229	0.2500	0.1824	0.0000		
Ug-elite	0.2227	0.3155	0.2669	0.4112	0.3935	0.3687	0.3580	0.4452	0.2732	0.4180	0.4133	0.3825	0.0000	
Ug-local	0.3501	0.3256	0.3634	0.2904	0.3705	0.3092	0.4458	0.4083	0.3825	0.3315	0.3297	0.2903	0.2964	0.0000

DRC = Democratic Republic of Congo; Ke = Kenya; Mad = Madagascar; Moz = Mozambique; Rw = Rwanda; Tz = Tanzania; Ug = Uganda. Elite = elite genotypes; local = local genotypes