GENETIC ANALYSIS AND GENOME-WIDE ASSOCIATION MAPPING OF CAROTENOID AND DRY MATTER CONTENT IN CASSAVA

by

WILLIAMS ESUMA

Submitted in fulfilment of the requirements in respect of the Doctoral Degree in Plant Breeding in the Department of Plant Sciences in the Faculty of Natural and Agricultural Sciences at the University of the Free State, Bloemfontein

Promoter:	**Prof. Maryke Tine Labuschagne**
Co-promoters:	**Prof. Liezel Herselman**
	Dr. Robert Sezi Kawuki

June 2016

DECLARATION

I, Williams Esuma, declare that the thesis that I herewith submit for the Doctoral Degree in Plant Breeding at the University of the Free State, is my independent work, and that I have not previously submitted it for a qualification at another institution of higher education.

I, Williams Esuma, hereby declare that I am aware that the copyright is vested in the University of the Free State.

I, Williams Esuma, declare that all royalties as regards to intellectual property that was developed during the course of and/or in connection with the study at the University of the Free State will accrue to the university.

I, Williams Esuma, hereby declare that I am aware that the research may only be published with the promoter's approval.

…… ……

Williams Esuma Date

DEDICATION

I dedicate this work to my entire family: my father Augusto Sodra and mother Regina Onia; my brothers (Mr. Ezama Jimmy, RIP) and sisters; my wife Betty and children (Flora and Dominic). Collectively, they inspired and supported me immensely throughout my life and education.

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my promoter, Prof. Maryke Labuschagne, for the exceptional support of my PhD study. She showed unlimited motivation, patience and immense knowledge that duly guided me to accomplish this study. Her dedication was complemented by unreserved commitments from my co-promoters, Prof. Liezel Herselman and Dr. Robert Kawuki, whose general guidance and critical double-checking approaches helped me throughout the research period and writing of this thesis. I could never imagine having any better mentors for my PhD study.

I am grateful to the National Agricultural Research Organisation (NARO) of Uganda for awarding me the PhD scholarship. My tuition, stipend and research funds were provided timely through the East African Agricultural Productivity Project (EAAPP) funded by a grant from the World Bank to the Government of Uganda. In this regard, I thank the Director General of NARO and the EAAPP coordination unit for the absolute commitments in providing the financial support for my study.

The Buckler Laboratory at Cornell University, USA, provided extra support to this study: firstly by paying the cost for genotyping-by-sequencing for genome-wide association study and then offering me a cost-free attachment in the laboratory, which enabled me to learn bioinformatics and other analytical tools for analysing next-generation sequencing data. I thank Prof. Edward Buckler for this incredible support and Dr. Ramu Punna together with Dr. Fei Lu for guiding me patiently through various analytical tools for my molecular research.

I thank the entire community of the National Crops Resources Research Institute (NaCRRI), for the supportive environment exhibited during the time of research for this study. I am most humbled by the commitment of Dr. James A. Ogwang (the former Director of Research at NaCRRI) who tirelessly ensured that funds were available whenever requested for the research work. I am equally grateful to the coherent team of scientists of the Root Crops Research Programme at NaCRRI for unreserved professional and social guidance that would uplift my life beyond this study period. In this spirit, I say thank you to Dr. Anton Bua, Dr. Yona Baguma, Dr. Titus Alicai, Dr. Christopher Omongo, Dr. Robert Kawuki, Dr. Benard Yada, Mr. Anthony Pariyo and Mr. Milton Otema. I remain grateful to all other technical and support staff of the Root Crops Research Programme for the solidarity and moral support accorded to me. I thank Mr. Nuwamanya Ephraim, Mr. Kaweesi Tadeo, Mr. Orone Joseph, Mr. Atwijukire Evans, Ms. Achan Sharon, Mr. Ozimati Alfred, Mr. Kirumira George and Mr. Jagen Solom for their

personal contributions during field or laboratory experimentation and data analysis. Equal thanks go to the leadership at Abi and Bulindi Zonal Agricultural Research Institutes of NARO. These institutes hosted the field trials conducted outside NaCRRI. They offered immense support for security and maintenance of these trials.

I thank the entire academic staff and graduate students of the Department of Plant Sciences, Plant Breeding division at the University of the Free State. Special thanks to my classmates Dr. Saidu Bah, Dr. Charles Mutimaamba, Mr. Pepukai Manjeru, Mr. Dibanzilua Nginamau, Mr. Obed John Mwenye, Mr. Fortunus Kapinga and Mr. Kabamba Mwansa for kind encouragements and support during the period of course work and thesis writing. It is difficult to summarise, in simple terms, the excellent work of Mrs. Sadie Geldenhuys who expertly ensured that my university registrations, accommodation and other logistics were taken care of in a timely manner. This was the best academic environment I ever encountered in my career; sincere thanks to all department members for the constructive interactions and wonderful social moments we shared.

I acknowledge inputs from individuals in the external research community whose insightful thoughts on my work provided additional thinking for enhancing the research output from this study. In particular, I had an opportunity to share my research in detail with Prof. Michael Gore (Cornell University, USA), Dr. Peter Kulakow, Dr. Melaku Gedil and Dr. Ismail Rabbi from the International Institute of Tropical Agriculture (IITA, Nigeria), who in turn provided great ideas that improved this research.

My heartfelt gratitude to the Fortuna Frontiers family for the social and moral support offered to my family during my study period. Sincere thanks to the families of Mr. Etrima Sunday, Mr. Onama Victor, Mr. Sadadi Ojoatre and Mr. Zaki Alfred. Together, you are a family to cherish.

I thank my wife, Betty Amakaru, for her love, support, encouragement, patience and the immense care she provided to my children Flora Asianzu and Dominic Feta.

Above all, I am humbled and thankful to the Almighty God for granting me a healthy life until now, most especially during the time of this study. May His will be done in me. To God be the glory.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abi-ZARDI	Abi Zonal Agricultural Research and Development Institute
AFLP	Amplified fragment length polymorphism
AMMI	Additive main effects and multiplicative interaction
ANOVA	Analysis of variance
ASV	AMMI stability value
BLUP	Best linear unbiased prediction
bp	Base pair(s)
BR	Baker's ratio
Bu-ZARDI	Bulindi Zonal Agricultural Research and Development Institute
°C	Degrees Celsius
CBSD	Cassava brown streak disease
CCD	Carotenoid cleavage dioxygenase
CIAT	International Centre for Tropical Agriculture
CMD	Cassava mosaic disease
CV	Coefficient of variation
DArT	Diversity array technology
DH	Doubled haploid
DMC	Dry matter content
DNA	Deoxyribonucleic acid
EAAPP	East African Agricultural Productivity Project
EDTA	Ethylenediaminetetraacetic acid
EST	Expressed sequence tag
RFC	Root flesh colour
F_1	First filial generation
FDR	False discovery rate
FRW	Fresh root weight
FSW	Fresh shoot weight
g	Gram(s)
g	Centrifugal force
GAPIT	Genome association and prediction integrated tool
GB	Giga byte
GBS	Genotyping-by-sequencing
GCA	General combining ability
GEBV	Genomic estimated breeding value

GEI	Genotype by environment interaction
GGE	Genotype plus genotype by environment interaction
GS	Genomic selection
GSI	Genotype selection index
GWA	Genome-wide association
GWAS	Genome-wide association study
h	Hour(s)
h^2	Narrow sense heritability
H^2	Broad sense heritability
ha	Hectare(s)
HI	Harvest index
IBS	Identical by state
IITA	International Institute of Tropical Agriculture
IPCA	Interaction principal component analysis
kb	Kilobase(s)
kg	Kilogram(s)
LD	Linkage disequilibrium
M	Molar(s)
m	Metre(s)
MAF	Minor allele frequency
MANOVA	Multivariate analysis of variance
MAP	Months after planting
MAS	Marker-assisted selection
Mb	Megabase(s)
min	Minute(s)
ml	Mililitre(s)
MLM	Mixed linear model
mM	Milimolar(s)
MS	Mean squares
NaCRRI	National Crops Resources Research Institute
NARO	National Agricultural Research Organisation
NCED	9-cis-epoxycarotenoid dioxygenase
ng	Nanogram(s)
NGS	Next generation sequencing
nm	Nanometer(s)
PBTools	Plant breeding tools

PC	Principal component
PCA	Principal component analysis
PCR	Polymerase chain reaction
pH	Potenz of hydrogen
PPD	Postharvest physiological deterioration
ppm	Parts per million
PSY	Phytoene synthase
PVP	Polyvinyl pyrrolidone
QTL	Quantitative trait loci
Q-Q	Quantile-quantile
RAM	Random-access memory
RAPD	Random amplified polymorphic DNA
RFC	Root flesh colour
RFLP	Restriction fragment length polymorphism
S_0	Non-inbred progenitor
S_1	Progeny of first selfing generation
S_2	Progeny of second selfing generation
SCA	Specific combining ability
SDS	Sodium dodecyl sulfate
SI	Selection index
SNP	Single nucleotide polymorphism
SS	Sum of squares
SSA	Sub-Saharan Africa
SSR	Simple sequence repeat
TAE	Tris-acetatediaminetetraacetic acid
TASSEL	Trait analysis by association, evolution and linkage
TCC	Total carotenoid content
TE	Tris-ethylenediaminetetraacetic acid
Tris-HCl	Tris (hydroxymethyl) aminomethane hydrochloride
V	Volt(s)
UV	Ultaviolet
VAD	Vitamin A deficiency
v/v	Volume per volume
w/v	Weight per volume
σ^2_A	Additive genetic variance
σ^2_D	Dominance genetic variance

σ^2_E	Environmental variance
σ^2_ε	Residual variance
σ^2_G	Genetic variance
σ^2_{GEI}	Genotype by environment interaction variance
σ^2_P	Phenotypic variance
µg	Microgram(s)
µl	Microlitre(s)
%	Per cent

CHAPTER 1

General introduction

Cassava (*Manihot esculenta* Crantz), second to maize as most important source of dietary energy in sub-Saharan Africa (SSA), continues to gain prominence as a food security crop across the world (Salvador et al. 2014; Tan 2015). This starchy root crop is grown and consumed widely in tropical regions of Africa, Asia and Latin America, where it dependably provides household food security in resource-poor farming systems (Monfreda et al. 2008). Globally, it is estimated that more than 800 million people derive the bulk of their dietary energy from cassava on a daily basis and over 500 million of these people live in SSA (FAOSTAT 2009; Montagnac et al. 2009; Burns et al. 2010).

In Uganda, cassava is second to bananas both in terms of production and consumption and the crop also ranks highly in most eastern and central African countries (Chipeta and Bokosi 2013; Salvador et al. 2014). Despite the apparent drop in farm yields and production of cassava over the last decade, which may be partly attributed to the threat of new diseases (Alicai et al. 2007), the harvested area (acreage) of the crop continues to increase in Uganda (Figure 1.1) (FAOSTAT 2014). This trend depicts an increasing importance of cassava in the economic welfare of people in Uganda.

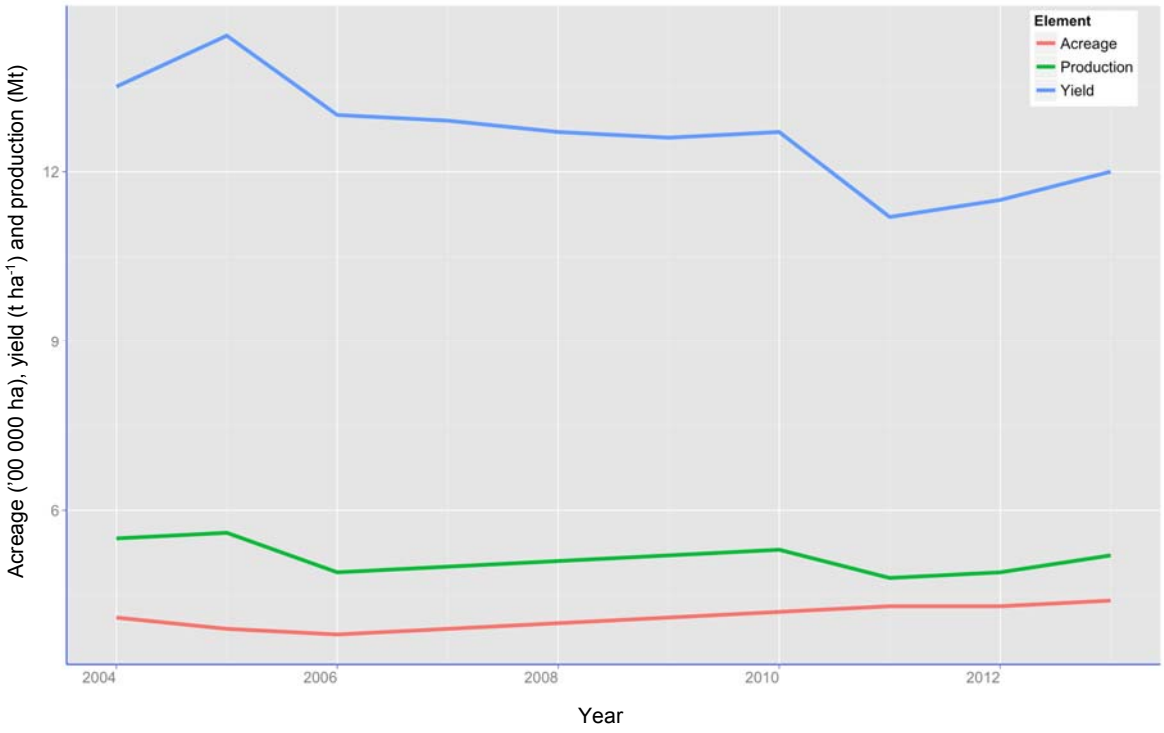


Figure 1.1 Changes in production and yield of cassava in Uganda between 2004-2013 (FAOSTAT 2014).

Current environmental challenges associated with global climatic changes have been predicted to escalate and cause a significant decrease in production for most food crops over the coming decades (Parry et al. 2005; Knox et al. 2012). Fortunately, various projections position cassava as a crop with immense ability to resist these challenges and sustain calorie demands in developing countries of SSA, which are currently experiencing rapid increases in both population size and urbanisation (Burns et al. 2010). The growing prominence of cassava is attributed primarily to the crop's competitive advantage to produce reasonable yields under adverse environments where other crops would fail or where resource-poor farmers simply cannot afford modern inputs required for meaningful production under such conditions (Daellenbach et al. 2005).

In the context of subsistence farming systems commonly practiced across SSA, cassava's suitability to intercrop with many other crop species and flexibility to time of harvesting makes it an appropriate choice of crop for production by peasants (Bamidele et al. 2008). This feature of cassava is complemented by its vegetative propagation method, which enables farmers to plant new gardens using planting materials saved from their own farms. Such practices increase efficiency of farm operations, especially planting at the onset of rains, while saving costs associated with seed purchases (Taiwo et al. 2014). These attributes make cassava an ideal crop for food production and income generation, particularly among resource-poor farmers in tropical regions of the world (Afolami et al. 2015). Additionally, cassava can be processed into a wide variety of food, feed, biofuel and starch that has numerous industrial applications (Kang et al. 2014; Okudoh et al. 2014). In fact, cassava is the second most important source of starch worldwide, after maize (Dufour et al. 1996) and its starch is the most traded worldwide (Norton 2014). This particular property elaborates a huge potential of cassava for commercialisation, which is likely to increase its production to meet the increasing demands for food and industrialisation (Abdoulaye et al. 2014).

However, heavy dependence on cassava for food has important nutritional drawbacks. The crop has relatively low nutritional quality, limiting it to providing only dietary energy (Montagnac et al. 2009). The limited nutritional value of cassava roots has dire implications for millions of people in SSA who depend on this staple crop. In particular, vitamin A deficiency (VAD) commonly afflicts people whose diets are constituted mainly by starchy staples (Rice et al. 2004; Sanghvi et al. 2007). Across SSA, an estimated 43% preschool children show clinical signs of VAD, of which 20% reside in eastern and central Africa (WHO 2009), with similar trends also observed in southern Asia (Akhtar et al. 2013). In Uganda, the WHO (2009) estimated 20% of preschool children and 19% of reproductive-age women

to be vitamin A deficient. Collectively, this information strongly correlates to the proportion of people afflicted by VAD with dependence on starchy staples. For example, Stephenson et al. (2010) reported that consuming cassava as a staple food places children 2-5 years old at a risk of inadequate protein and vitamin A intake in both Kenya and Nigeria.

VAD has several severe health and economic consequences, including early mortality and reduced productivity. This nutritional deficiency can lead to irreversible blindness in children under the age of five (Rice et al. 2004). Globally, over 450 000 children that suffer from VAD are rendered blind every year and such children have a 50% chance of dying within a year from preventable diseases like measles, diarrhoea and malaria (Sommer 2008). Several factors including inadequate medical care and poor sanitation contribute to micronutrient deficiencies in the developing world (Tulchinsky 2010), but a poor diet is the primary cause among resource-constrained communities. People in such communities tend to consume disproportionately high amounts of staples like cassava, which are relatively low in micronutrients compared to fruits, vegetables and animal products that provide the essential micronutrients for optimal health (Rice et al. 2004; Okeke et al. 2009).

Interventions to prevent vitamin A and other nutrient deficiencies have been applied across the world using three main traditional strategies: food fortification, supplementation and dietary diversification. These strategies can effectively reduce micronutrient malnutrition, but their implementation in developing countries is costly with low impact due to diverse reasons, including poor social infrastructure and high poverty levels (Boy et al. 2009; Iannotti et al. 2014). Subsequently, food-based approaches to combat VAD are emphasised to provide sustainable solutions to micronutrient malnutrition (Thompson and Amoroso 2011).

Based on this premise, a novel effort, referred to as the HarvestPlus Challenge Programme of the CGIAR, is being coordinated jointly by a consortium of research institutes including the International Centre for Tropical Agriculture (CIAT), IITA, the International Maize and Wheat Improvement Center and the International Potato Center to support the genetic improvement of the nutritional quality of staple crops (Mayer et al. 2008). This initiative, referred to as crop biofortification, is achievable through conventional breeding techniques that take advantage of the genetic variability for micronutrients in different crop genetic resources. Cassava, sweet potato, maize, rice, wheat, barley and beans are the priority crops being biofortified (Pfeiffer and McClafferty 2007). These crops form the major food staples for the majority of people often at a high risk of micronutrient deficiencies worldwide (Bouis and Welch 2010). Biofortification represents a sustainable strategy that aims at addressing the primary cause of micronutrient malnutrition, which is a nutrient-deficient diet

(Welch and Graham 2004). Therefore, deploying nutrient-rich crop varieties would significantly positively impact on people living in remote areas that rarely benefit from food fortification and vitamin supplementation programmes (Hefferon 2015).

Because cassava takes a central role in diets of millions of people vulnerable to VAD, substantial efforts and resources have been committed towards developing varieties enriched with provitamin A carotenoids (Nassar and Ortiz 2010). These efforts have led to the generation of cassava genetic stocks accumulating up to 25 μg g^{-1} of ß-carotene in roots (Ceballos et al. 2013) and a better understanding of the impact of cassava root processing on the bioavailability of carotenoids (Tanumihardjo et al. 2010; Ceballos et al. 2011; Ceballos et al. 2013). Efforts to biofortify cassava with provitamin A carotenoids have recently been boosted by reports that consumption of roots of such varieties increases the concentration of ß-carotene and retinyl palmitate triacylglycerol-rich lipoprotein plasma in adult women (Frano et al. 2013). While provitamin A cassava would be a new product to most farmers and consumers due to the characteristic yellow root pigmentation, it is sensory and culturally acceptable for consumption in eastern Africa (Talsma et al. 2013).

An additional benefit of carotenoid enrichment in cassava roots is the positive impact of carotenoids on extension of shelf life of fresh roots (Sánchez et al. 2006; Morante et al. 2010). Fresh cassava roots deteriorate within 24-48 hours after harvest, but roots enriched with ß-carotene appear to show reduced or delayed postharvest physiological deterioration (PPD) (Morante et al. 2010). Cassava varieties tolerant to PPD would be ideal for commercial production of the crop, because they would guard against the lost revenue in production and marketing of the crop (Abass et al. 2013; Nzeh and Ugwu 2014).

Accordingly, the cassava research programme in Uganda initiated a breeding pipeline for genetic improvement of the crop for provitamin A carotenoids. To kick-start this breeding initiative, a diverse set of improved germplasm with varying levels of ß-carotene was introduced from CIAT and IITA (Esuma et al. 2012). However, this genetic resource was only marginally used in the breeding programme, largely because of inadequate genetic information to guide systematic improvement of cassava for ß-carotene and associated quality traits (Akinwale et al. 2010). Such information would guide decisions on use of appropriate strategies for realising meaningful genetic gains through breeding and selection (Acquaah 2012; Ceballos et al. 2012).

Another technical challenge in cassava biofortification is the tendency of low dry matter content (DMC) in roots with high carotenoid content (Njoku et al. 2015). More often than not, farmers prefer cassava varieties with high DMC (Tumuhimbise et al. 2012; Ojo and

Ogunyemi 2014), indicating the need for breeders to develop provitamin A varieties in the genetic background of high DMC. For conventional breeding to produce meaningful gains in combining these traits and increase adoption of the resultant varieties, careful selection of parental genotypes on the basis of their combining ability is required (Ceballos et al. 2015). An efficient breeding approach would be to use marker-assisted selection (MAS) given the wide segregation observed in cassava for both carotenoid content and DMC (Akinwale et al. 2010; Esuma et al. 2012). Molecular markers would be an economically reliable tool for cassava breeding to facilitate efficient and timely selection of recombinants expressing these root quality traits in the background of superior agronomic performance, accelerating the variety development process (Rudi et al. 2010). However, the molecular breeding approach has been less utilised for cassava genetic improvement (Ceballos et al. 2015). Against this background, this study was conducted with the overall aim of developing improved provitamin A cassava genetic resources in Uganda. Specific objectives of the study were:

1. To determine the combining ability of provitamin A genotypes and the mode of gene action in inheritance of carotenoid content in cassava;
2. To assess the effect of genotype by environment interaction on accumulation of carotenoid content in cassava;
3. To identify genomic regions and polymorphisms associated with natural variation for DMC and carotenoid content in cassava.

References

Abass A, Mlingi N, Ranaivoson R, Zulu M, Mukuka I, Abele S, Bachwenkizi B, Cromme N (2013) Potential for Commercial Production Marketing of Cassava: Experiences from the Small-Scale Cassava Processing Project in East and Southern Africa. International Institute of Tropical Agriculture, Ibadan, Nigeria

Abdoulaye T, Abass A, Maziya-Dixon B, Tarawali G, Okechukwu R, Rusike J, Alene A, Manyong V, Ayedun B (2014) Awareness and adoption of improved cassava varieties and processing technologies in Nigeria. Journal of Development and Agricultural Economics 6:67-75

Acquaah G (2012) Principles of Plant Genetics and Breeding. Second edition. John Willey and Sons Ltd, UK

Afolami CA, Obayelu AE, Vaughan II (2015) Welfare impact of adoption of improved cassava varieties by rural households in south western Nigeria. Agricultural and Food Economics 3:1-17

Akhtar S, Ahmed A, Randhawa MA, Atukorala S, Arlappa N, Ismail T, Ali Z (2013) Prevalence of vitamin A deficiency in South Asia: causes, outcomes, and possible remedies. Journal of Health, Population and Nutrition 4:413-423

Akinwale MG, Aladesanmwa RD, Akinyele BO, Dixon AGO, Odiyi AC (2010) Inheritance of ß-carotene in cassava (*Manihot esculenta* Crantz). International Journal of Genetics and Molecular Biology 2:198-201

Alicai T, Omongo CA, Maruthi MN, Hillocks RJ, Baguma Y, Kawuki R, Bua A, Otim-Nape GW, Colvin J (2007) Re-emergence of cassava brown streak disease in Uganda. Plant Disease 91:24-29

Bamidele FS, Babatunde RO, Rasheed A (2008) Productivity analysis of cassava-based production sytems in the Guinea Savannah: case study of Kwara State, Nigeria. American-Eurasian Journal of Scientific Research 3:33-39

Bouis HE, Welch RM (2010) Biofortification - a sustainable agricultural strategy for reducing micronutrient malnutrition in the global south. Crop Science 50:S20-S32

Boy E, Mannar V, Pandav C, de Benoist B, Viteri F, Fontaine O, Hotz C (2009) Achievements, challenges and promising new approaches in vitamin and mineral deficiency control. Nutrition Reviews 67:24-30

Burns A, Gleadow R, Cliff J, Zacarias A, Cavagnaro T (2010) Cassava: the drought, war and famine crop in a changing world. Sustainability 2:3572-3607

Ceballos H, Kawuki RS, Gracen VE, Yencho GC, Hershey CH (2015) Conventional breeding, marker-assisted selection, genomic selection and inbreeding in clonally propagated crops: a case study for cassava. Theoretical and Applied Genetics 128:1647-1667

Ceballos H, Kulakow P, Hershey C (2012) Cassava breeding: current status, bottlenecks and the potential of biotechnology tools. Tropical Plant Biology 5:73-87

Ceballos H, Luna J, Escobar AF, Ortiz D, Pérez JC, Sánchez T, Pachón H, Dufour D (2011) Spatial distribution of dry matter in yellow fleshed cassava roots and its influence on carotenoid retention upon boiling. Food Research International 45:52-59

Ceballos H, Morante N, Sánchez T, Ortiz D, Aragón I, Chávez AL, Pizarro M, Calle F, Dufour D (2013) Rapid cycling recurrent selection for increased carotenoids content in cassava roots. Crop Science 53:23-42

Chipeta MM, Bokosi JM (2013) Status of cassava (*Manihot esculenta*) production and utilization in Malawi. International Journal of Agronomy and Plant Production 4:3637-3644

Daellenbach GC, Kerridge PC, Wolfe MS, Frossard E, Finckh MR (2005) Plant productivity in cassava-based mixed cropping systems in Colombian hillside farms. Agriculture, Ecosystems & Environment 105:595-614

Dufour D, O'Brien GM, Best R (1996) Cassava Flour and Starch. International Centre for Tropical Agriculture, Cali, Colombia

Esuma W, Rubaihayo P, Pariyo A, Kawuki R, Wanjala B, Nzuki I, Harvey JJW, Baguma Y (2012) Genetic diversity of provitamin A cassava in Uganda. Journal of Plant Studies 1:60-71

FAOSTAT (2009) Why cassava? http://www.fao.org/ag/AGP/agpc/gcds/index_en. Accessed 15 June 2015

FAOSTAT (2014) Food and Agriculture Organization of the United Nations Statistics. http://data.fao.org/ref/262b79ca-279c-4517-93de-ee3b7c7cb553. Accessed 4 September 2015

Frano MR, Woodhouse LR, Burnett DJ, Burri BJ (2013) Biofortified cassava increases beta-carotene and vitamin A concentrations in the TAG-rich plasma layer of American women. British Journal of Nutrition 110:310-320

Hefferon KL (2015) Nutritionally enhanced food crops; progress and perspectives. International Journal of Molecular Sciences 16:3895-3914

Iannotti LL, Trehan I, Manary JM (2014) Review of the safety and efficacy of vitamin A supplementation in the treatment of children with severe acute malnutrition. Nutrition Journal 12:125

Kang Q, Appels L, Baeyens J, Dewil R, Tan T (2014) Energy-efficient production of cassava-based bio-ethanol. Advances in Bioscience and Biotechnology 5:925-939

Knox J, Hess T, Daccache A, Wheeler T (2012) Climate change impacts on crop productivity in Africa and South Asia. Environmental Research Letters 7:034032

Mayer JE, Pfeiffer WH, Beyer P (2008) Biofortified crops to alleviate micronutrient malnutrition. Current Opinion in Plant Biology 11:166-170

Monfreda C, Ramankutty N, Foley JA (2008) Farming the planet: geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Global Biogeochemical Cycles 22:1-19

Montagnac AJ, Davis CR, Tanumuhardjo SA (2009) Nutritional value of cassava for use as a staple food and recent advances for improvement. Comprehensive Reviews in Food Science and Food Safety 8:181-194

Morante N, Sánchez T, Ceballos H, Calle F, Pérez JC, Egesi C, Cuambe CE, Escobar AF, Ortiz D, Chávez AL, Fregene M (2010) Tolerance to postharvest physiological deterioration in cassava roots. Crop Science 50:1333-1338

Nassar NMA, Ortiz R (2010) Breeding cassava to feed the poor. Scientific American Journal 302:78-82

Njoku DN, Gracen VE, Offei SK, Asante IK, Egesi CN, Kulakow P, Ceballos H (2015) Parent-offspring regression analysis for total carotenoids and some agronomic traits in cassava. Euphytica 206:657-666

Norton R (2014) Global Starch Market Outlook and Feedstock Economics. Cassava World. Centre for Management Technology (CMT), Lusaka

Nzeh EC, Ugwu JN (2014) Economic analysis of production and marketing of cassava in Akoko North-West local government area of Ondo State, Nigeria. International Journal of Agricultural Policy and Research 2:234-237

Ojo SO, Ogunyemi AI (2014) Analysis of factors influencing the adoption of improved cassava production technology in Ekiti state, Nigeria. International Journal of Agricultural Sciences and Natural Resources 1:40-44

Okeke EC, Eneobong HN, Uzuegbunam AO, Ozioko AO, Umeh SI, Kuhnlein H (2009) Nutrient composition of traditional foods and their contribution to energy and nutrient intakes of children and women in rural households in Igbo culture area. Pakistan Journal of Nutrition 8:304-312

Okudoh V, Trois C, Workneh T, Schmidt S (2014) The potential of cassava biomass and applicable technologies for sustainable biogas production in South Africa: a review. Renewable and Sustainable Energy Reviews 39:1035-1052

Parry M, Rosenzweig C, Livermore M (2005) Climate change, global food supply and risk of hunger. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 360:2125-2138

Pfeiffer WH, McClafferty B (2007) HarvestPlus: breeding crops for better nutrition. Crop Science 47:88-105

Rice A, West K, Black R (2004) Vitamin A deficiency. In: Ezzati M, Lopez AD, Rodgers A, Murray CJL (eds) Comparative Quantification of Health Risks vol 1. World Health Organisation Geneva, Switzerland, pp 211-256

Rudi N, Newton GW, Alwang J, Asumugha G (2010) Economic impact analysis of marker-assisted breeding for resistance to pests and postharvest deterioration in cassava. African Journal of Agricultural Research 4:110-122

Salvador EM, Steenkamp V, McCrindle CME (2014) Production, consumption and nutritional value of cassava (*Manihot esculenta* Crantz) in Mozambique: an overview. Journal of Agricultural Biotechnology and Sustainable Development 6:29-38

Sánchez T, Chávez AL, Ceballos H, Rodriguez-Amaya DB, Nestel P, Ishitani M (2006) Reduction or delay of post-harvest physiological deterioration in cassava roots

with higher carotenoid content. Journal of the Science of Food and Agriculture 86:634-639

Sanghvi T, Ameringen M, Baker J, Fiedler J (2007) Vitamin and mineral deficiencies technical situation analysis: a report for the ten year strategy for the reduction of vitamin and mineral deficiencies. Food and Nutrition Bulletin 28:157-216

Sommer A (2008) Vitamin A deficiency and clinical disease: an historical overview. The Journal of Nutrition 138:1835-1839

Stephenson K, Amthor R, Mallowa S, Nungo R, Maziya-Dixon B, Gichuki S, Mbanaso A, Manary M (2010) Consuming cassava as a staple food places children 2-5 years old at risk for inadequate protein intake, an observational study in Kenya and Nigeria. Nutrition Journal 9:9

Taiwo O, Dayo OS, Bolariwa KO (2014) Technical efficiency analysis of cassava production in Nigeria; implication for increased productivity and competitiveness Research Journal of Agriculture and Environmental Management 3:569-576

Talsma EF, Melse-Boonstra A, de Kok BP, Mbera GN, Mwangi AM, Brouwer ID (2013) Biofortified cassava with pro-vitamin A is sensory and culturally acceptable for consumption by primary school children in Kenya. PLoS One 8:e73433

Tan SL (2015) Cassava - silently, the tuber fills: the lowly cassava, regarded as a poor man's crop, may help save the world from the curse of plastic pollution. Utar Agriculture Science Journal 1:12-24

Tanumihardjo SA, Palacios N, Pixley KV (2010) Provitamin A carotenoid bioavailability: what really matters? International Journal for Vitamin and Nutrition Research 80:336-350

Thompson B, Amoroso L (2011) Combating Micronutrient Deficiencies: Food-Based Approaches, vol 4. CAB International and FAO, Oxfordshire, UK

Tulchinsky TH (2010) Micronutrient deficiency conditions: global health issues. Public Health Reviews 32:243-255

Tumuhimbise R, Melis R, Shanahan P, Kawuki R (2012) Farmers' perceptions on early storage root bulking in cassava (*Manihot esculenta* Crantz) in east and central Uganda and their implication for cassava breeding. World Journal of Agricultural Sciences 8:403-408

Welch RM, Graham RD (2004) Breeding for micronutrients in staple food crops from a human nutrition perspective. Journal of Experimental Botany 55:353-364

WHO (2009) Global Prevalence of Vitamin A Deficiency in Populations at Risk 1995–2005: WHO Global Database on Vitamin A Deficiency. World Health Organisation, Geneva

CHAPTER 2

Literature review

2.1 The cassava crop: its origin and diversity

Cassava belongs to the family Euphorbiaceae, genus *Manihot* and species *esculenta* (Allem et al. 2001). The genus *Manihot* contains 98 species, which are all classified as diploids (2n = 36 chromosomes), although a high number of duplicated nuclear chromosomes seen at metaphase 1 and at anaphase of meiosis suggests the crop to be a segmental allotetraploid (Hashimoto-Freitas and Nassar 2013). The cultivated species of cassava evolved from wild populations of *M. esculenta* ssp. *flabellifolia* (Allem 1999). There are no genetic and cytological barriers within species of the *Manihot* genus (Nassar 2002; Nassar 2003), allowing for crosses between species within the genus.

General views put the origin of cassava to be South America, with centres of diversity reported to be within central Brazil (Allem 2002; Nassar 2002; Nassar 2003). With the aid of molecular markers, Olsen (2004) undertook a more detailed examination of the origin of cassava, leading to the conclusions that cassava was domesticated from the wild *M. esculenta* ssp. *flabellifolia* and that the crop originated from the southern Amazon basin. Historical records of the arrival of cassava to Africa are unclear, but the first standard study of cassava in Africa by Jones (1959) indicated that Portuguese sailors first introduced the crop to parts of West Africa from Brazil in the 16th century. Cassava was gradually integrated into the traditional food systems across tropical Africa, arriving in Uganda in the early 1890s (Langlands 1966).

Since the time of its domestication and introduction to Africa, cassava has been cultivated primarily as source of dietary carbohydrates. However, large diversity exists within *M. esculenta* for nutritional and other quality traits. In particular, genetic variability has been reported for carotenoids (Chávez et al. 2000; Chávez et al. 2005; Nassar et al. 2007) and protein (Akinbo et al. 2011) in cassava. This diversity for nutritional traits has widened perspectives and approaches for cassava genetic improvement to make the crop more reliable for both food and nutritional security (Nassar and Ortiz 2010; Ceballos et al. 2013).

2.2 Critical considerations for accelerating cassava breeding

Plant breeding efforts, more often than not, target the identification and development of superior individuals and families. Conventional wisdom in plant breeding suggests that in order to transfer characteristics into a breeding population, the starting point is the identification of gene sources, which constitute genotypes that express the trait of interest at a high level. If such traits are heritable, then genes controlling their expression can be transferred to the progeny, hopefully to achieve a similar high level of expression (Acquaah 2012). Accordingly, cassava genetic improvement programmes often begin with the assembly and evaluation of target germplasm, which, in programmes other than those in Latin America, tend to be introductions from foreign gene pools (Ceballos et al. 2012a). The source germplasm is used to generate new recombinant genotypes through hybridisation among a selected panel of elite introductions (Kawano 2003). To increase the mean performance of breeding populations, individual plants with higher than average performance are selected and recombined in a recurrent selection fashion. This increment is higher for traits with high narrow sense heritability and increases parent-offspring resemblance and response to selection (Falconer and Mackay 1996).

Genetic gains achievable from a breeding programme are shaped by four modifiable components: narrow sense heritability constituted by additive genetic and phenotypic variance, selection intensity, parental control and time (Falconer and Mackay 1996; Fuente et al. 2013). Narrow sense heritability is a measure of the proportion of phenotypic variance explained by additive genetic variance (σ^2_A). σ^2_A is the component transferrable to the next generation and is affected by the choice of germplasm for developing segregating populations. Phenotypic variance is affected by the choice and management of selection environments (Bos and Caligari 2008). Selection intensity is influenced by a combination of the additive genetic and phenotypic variance components (Falconer and Mackay 1996; Acquaah 2012). The relationship between the change in mean performance of the breeding population before and after selection (the response to selection), R, and the within-generation change in the mean due to selection (selection deferential), S, is expressed by the linear relationship $R = h^2S$, where h^2 is the narrow sense heritability of the trait (Falconer and Mackay 1996; Bos and Caligari 2008). This relation is commonly referred to as the breeders' equation (Fehr 1993). Another component that can easily be modified is the selection intensity, which corresponds to the percentage of individuals advanced after a cycle of selection. Thus, it is practical to optimise the aforementioned factors through knowledge of the germplasm and use of predictive tools (Fehr 1993; Bos and Caligari 2008).

The most critical remaining factor to maximise genetic gain is time. While shorter selection cycles are preferred for increasing genetic gains over a given period of time, the inherent long growing cycle of cassava limits the number of generations per year. Cassava's long growing cycle makes it impractical to increase genetic gains with regard to selection cycle time, except for the use of off-season nurseries (Iglesias and Hershey 1994; Ceballos et al. 2004) and the potential eventual use of the doubled haploid (DH) technology in the future (Perera et al. 2014; Yan et al. 2014; Ceballos et al. 2015).

2.2.1 Selecting parental lines for hybridisation

Hybridisation facilitates transfer of genes for desired traits between specific pairs of parental lines (Fukuda et al. 2002). This practice allows for generation of recombinant gametes through meiosis, which is the principle basis for genetic variation arising through sexual reproduction (Bengtsson 2003). The heterozygous nature of cassava necessitates careful selection of parental genotypes for hybridisation. As such, prospective parents are selected on the basis of performance of their progeny in hybridisation programmes (Ceballos et al. 2004). Outstanding S_1 and/or S_2 progeny have also been generated through inbreeding cassava.

However, choosing parental lines based on their phenotypic performance *per se* can result in production of poor recombinants in the segregating population, which illustrates the need for using genotypes with high breeding values for hybridisation (Cowling and Léon 2013). The breeding value of a genotype indicates its ability to combine well with other genotypes and transmit genetic factors controlling useful traits to the progeny (Falconer and Mackay 1996). Thus, controlled pollination by hand is emphasised for production of full-sib families in cassava genetic improvement, which enables breeders to generate useful genetic information alongside the development of breeding populations with known pedigree information (Kawano et al. 1978; Ceballos et al. 2004).

2.2.2 Flowering and seed formation

Botanically, cassava is classified as a monoecious crop as it bears separate male (staminate) and female (pistillate) flowers on the same plant (Alves 2002). On average, a cassava plant flowers 3-4 months after planting, save for genotypes that never produce flowers (Kawano et al. 1978; Alves 2002). Male and female flowers are borne on the same panicle, but female flowers in every inflorescence mature 10-14 days earlier than male flowers (Halsey et al. 2008). Mature flowers open naturally, a practice that allows for cross pollination by insects and/or wind. Male flowers often open when female flowers

on the same inflorescence have been fertilised or have aborted. Nevertheless, a single plant may flower over a prolonged period of time, provided it continues to branch. Some cassava genotypes like NASE 14 (officially released in Uganda) can produce up to six tiers. Thus, pollen from one inflorescence can fertilise flowers from other branches on the same plant, essentially making self-pollination and cross-pollination practical mating techniques for cassava (Alves 2002; Halsey et al. 2008).

In order to produce sufficient amounts of seeds to raise a required number of progeny from any cross combination, a breeder needs to consider the number of pollinations carried out between the given pair of parents (Fukuda et al. 2002). Controlled pollination typically produces an average of one viable botanical seed out of the potential three seeds from a tri-locular ovary per fruit (Jennings and Iglesias 2002; Yan et al. 2014). Cassava breeders indicate a success rate of 40-50% for fruit formation from all crosses made in a crossing programme (personal communication with Mr. Pariyo Anthony, cassava breeder, NaCRRI) and 50-60% for seed formation from successfully formed fruits (personal communication with Dr. Mark Halsey, Donald Danforth Plant Science Centre, USA). Mature fruits freely dehisce 2.5-3 months after fertilisation to release seeds (Halsey et al. 2008), requiring mature fruits to be picked before dehiscence. The harvested seeds require a dormancy period of 2-3 months storage at ambient temperatures to achieve full physiological maturation before they can be germinated at optimum temperatures of 30-35ºC (Ellis et al. 1982).

However, the ability of a genotype to flower is the primary factor that affects hybridisation in cassava. Low rates of flower production, male sterility and the physiological state of the anther and/or stigma are particularly important problems hindering successful hybridisation (Kawano et al. 1978; Ceballos et al. 2012b). Types of male sterility reported in cassava include anther deformation, cytological abnormalities and functional male sterility, reflected by absence of anther dehiscence (Jos et al. 1990). However, for most crossing events, the genotype of the female parent appears to be more important than the pollen parent in determining the success of hybridisation (Kawano et al. 1978).

There are on-going research efforts by the Next Generation Cassava Breeding Project (http://nextgencassava.org) to explore avenues for flower induction and seed set in cassava. Options being examined under this initiative include grafting of shy-flowering genotypes (genotypes that delay flowering or produce insignificant number of flowers) onto those that produce profuse flowers and application of plant hormones and growth regulators. During the Next Generation Cassava Breeding Project review meeting in

Kampala, Uganda (February 2012), Dr. Hernan Ceballos (CIAT cassava breeder) presented some unpublished data on grafting experiments. This information indicated that using planting material generated from grafted branches that initially failed to produce flowers in the graft-union state resulted in plants that produced sufficient numbers of flowers. If validated, such information would allow use of high-value poor-flowering genotypes for hybridisation. Additional experiments are being set up to (1) assess the time for a stigma to remain receptive after anthesis, (2) determine how long the pollen takes to travel from the stigma to the embryo sac and (3) develop protocols that limit contamination after pollination (personal communication with Dr. Robert Kawuki, cassava breeder, NaCRRI). Such efforts could generate information to aid development of efficient methods to increase pollination success.

2.3 Important concepts guiding genetic analyses for crop improvement

2.3.1 Gene action and inheritance of traits

In higher plants, most traits of agronomic importance are quantitatively inherited. Expression of phenotypes of such polygenic traits is a result of one or more of the following gene actions: additive, dominance, overdominance or epistasis. As described by Falconer and Mackay (1996), (1) a given phenotype resulting from expression of a set of additive genes is the cumulative effect of each of the individual genes, (2) dominance gene effects are deviations from additive effects, (3) epistatic effects are a result of interaction between non-allelic genes at two or more loci resulting in one gene masking the phenotypic expression of another gene and (4) overdominance occurs when the combined effect of alleles exceeds the individual allelic effects.

Inheritance describes the transmission of genetic information to succeeding generations (Falconer and Mackay 1996). In terms of the theory of classical Mendelian genetics, inheritance implies expression of a dominant gene in a phenotype when two contrasting characters are combined (Acquaah 2012). Knowledge about inheritance of the gene is particularly essential when aiming to recover and maintain desirable donor genes in the progeny (Falconer and Mackay 1996). Such genetic information guides breeders in selecting appropriate designs that can improve breeding efficiency and precision to enhance genetic gains (Crossa et al. 2010).

Gene action and heritability are intrinsic components of the breeder's equation, which shapes the genetic gain that can be realised from a breeding programme (Falconer and Mackay 1996; Acquaah 2012). These phenomena provide the core basis for selection of

desirable genotypes in a breeding programme (Poehlman and Sleper 2006). Other factors that come into play when undertaking genetic improvement of crops include breeding priorities, number of varieties developed, return on investment and tools for increasing selection efficiency in breeding programmes (Ramalho et al. 2013).

2.3.2 Estimates of genetic variances

Present-day advances in crop genetic improvement can be greatly attributed to the impact of early quantitative genetics theory, which profoundly influenced the evolution of modern theoretical and applied statistics and facilitated the development of the theory behind regression and correlation analyses and principles upon which the analysis of variance is based (Venkocsky et al. 2012). Estimates of variance reflect the amount of variation for a character being measured in a population. Phenotypic variance (σ^2_P) of a given character is the sum of its genetic variance (σ^2_G) and environmental variance (σ^2_E), which is that part of the phenotypic variance attributed to prevailing environmental conditions (Falconer and Mackay 1996). The total σ^2_G, also referred to as the genotypic value, is partitioned into additive genetic variance (σ^2_A), dominance genetic variance (σ^2_D) and epistatic genetic variance. σ^2_A is the most important component for a plant breeder because it is the variance of breeding values from which genetic gain is derived (Acquaah 2012). This component is the heritable variance; thus, a major determinant of the observable genetic properties of a population and the response of that population to selection. The primary goal of every breeding programme is to generate and select high performing genotypes from a set of progeny. Thus, populations with greater σ^2_A are expected to produce larger numbers of superior transgressive segregants than those with narrow genetic variances (Dudley and Moll 1969; Jiang et al. 2014).

Different mating designs and associated statistical algorithms can be used to estimate variance components. Traditional analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) are by far the most powerful and commonly used tools to estimate variance components (Rencher 2002; Gelman 2005). An example of MANOVA is the tensor method, a form of an additive genetic variance-covariance matrix used to summarise multivariate genetic relationships among a set of traits (Aguirre et al. 2014). By elegantly capturing all of the variation in genetic variance among populations, the method allows the identification of the trait combinations that differ most in genetic variance. The expected values of these variances are used to estimate components of genetic variation by equating them with the observed values. However, at the early stages of a breeding programme, σ^2_A is more important as it reflects heritable effects that

can be effectively selected (Sprague and Tatum 1942). Non-additive effects become more important at later selection stages because the selected material has greater similarity, thereby largely eliminating the additive effects (Hallauer et al. 2010).

Cassava breeders commonly develop full-sib and/or half-sib families, followed by phenotypic mass selection for identification of varieties that can be released for cultivation by farmers (Kawano 2003). Using this method, breeders tend to focus on evaluating and selecting individual genotypes regardless of the family origin, a process that essentially disregards the family structure. Ceballos et al. (2015) reported that the relative importance of the variance components varies with the type of family, the stage of inbreeding and within and between families (Table 2.1). In fact, the genetic effects are asymmetrically distributed between the between- and within-family components in the half-sib or full-sib breeding systems. For instance, half of the additive variance present in the parental generation is expressed as differences between the full-sib families and the other half expressed as the within-family variation. Because cassava is clonally propagated, individual genotypes can be multiplied in such a way that environmental and genetic factors affecting their performance can be separated through multi-environment evaluations. This practice allows for more accurate estimation of within-family genetic effects. This way it is possible to overcome the tendency by breeders to neglect the within-family variance in the phenotypic mass selection method (Ceballos et al. 2015).

Table 2.1 Partitioning of additive and dominance genetic effects in different types of families

Family type	Inbreeding coefficient	Between families		Within families		Total	
		σ^2_A [c]	σ^2_D [d]	σ^2_A	σ^2_D	σ^2_A	σ^2_D
HS[a]	0	1/4	0	3/4	1	1	1
FS[b]	0	1/2	1/4	1/2	3/4	1	1
S_1/F_3	1/2	1	1/4	1/2	1/2	3/2	3/4
S_2/F_4	3/4	3/2	3/16	1/4	1/4	7/4	7/16
S_3/F_5	7/8	7/4	7/64	1/8	1/8	15/8	15/46
S_∞/F_∞	1	2	0	0	0	2	0

[a]Half-sib; [b]Full-sib; [c]Additive genetic variance component; [d]Dominance genetic variance component. Table based on Ceballos et al. (2015).

2.3.3 Mating designs commonly used for crop improvement

The purpose of using mating designs in plant breeding is to (1) furnish the breeder with

information on genetic control of the character under investigation and (2) generate segregating populations as new sources of potential varieties (Nduwumuremyi et al. 2013). Such information guides the breeder to use an appropriate strategy for assessing the genetic gain that can be attained for a given selection intensity (Singh et al. 2004). Various mating designs have been described and used for crop improvement, which include (1) bi-parental mating, (2) polycross, (3) top cross design, (4) North Carolina designs (North Carolina Design I, North Carolina Design II, North Carolina Design III), (5) diallel design and (6) line x tester design (Bernardo 2010; Acquaah 2012; Nduwumuremyi et al. 2013).

However, it is the breeder's responsibility to carefully consider the suitability of a particular design for development of an appropriate population for estimation of the variance components (Hallauer et al. 2010). Some of the decision-guiding factors for choice of a mating design for plant breeding include (1) type of pollination (self- or cross-pollinated), (2) type of crossing to be used (controlled or open), (3) type of pollen dissemination (wind or insect), (4) presence of a male-sterility system, (5) objective of the experiment (variety development or genetic studies) and (6) the required size of the population (Hill et al. 1998; Singh et al. 2004). The following section expands on the diallel mating design described by Griffing (1956), which was the choice of mating design for the genetic study presented in this thesis.

Diallel mating design

The concept of diallel mating has been defined as making a set of all possible crosses between several genotype pairs (Hayman 1954; Griffing 1956). Though initially used in animal breeding, Sprague and Tatum (1942) introduced the concept of diallel mating to the field of plant breeding by making all possible cross combinations among a set of maize inbred lines. Since then, this mating design has gained favour among breeders of different crop species to (1) obtain information on genotypes as parental lines, (2) assess gene actions in inheritance of traits and (3) develop appropriate selection procedures in a breeding programme (Egesel et al. 2003; Hallauer et al. 2010; Nduwumuremyi et al. 2013). Griffing (1956) described the critical assumptions for treating parents as fixed or random factors and the crossing methods used for diallel analyses.

Analysis and interpretation of information from diallel experiments are based on estimates of combining ability, defined as the performance of a line in hybrid combinations (Arunachalam 1976). Four methods to analyse combining ability using genetic estimates of the parent and hybrid components of a diallel cross have been

proposed (Griffing 1956). Variation in the methods depends on whether on not the parents, F_1 progeny or reciprocals are included in the analysis. Briefly, method 1 includes analysis with parents, F_1 and reciprocals so that there are p^2 cross combinations; in method 2, only parents and F_1 are included in the analysis so that there are $1/2p(p+1)$ cross combinations; method 3 includes F_1 and reciprocals only so that there are $p(p-1)$ combinations; and method 4 involves analysis with F_1 only, excluding parents and reciprocals so that there are $1/2p(p-1)$ cross combinations. The analysis partitions the combining ability into general combining ability (GCA) and specific combining ability (SCA). The GCA is expressed as the average performance of a line in hybrid combinations while SCA is the relative performance of a cross combination compared to the average performance of the lines involved (Sprague and Tatum 1942).

GCA and SCA are used for making inferences about additive and non-additive genetic effects of a trait and the extent of genetic gain that can be realised from the breeding programme. Large GCA:SCA variance ratios suggest the relative importance of additive genetic effects over non-additive genetic effects arising from dominance and/or epistatic gene effects (Griffing 1956; Viana and Matta 2003). The practical implication of combining ability effects is that smaller values of SCA relative to the GCA indicate the possibility of predicting the performance of single cross progeny on the basis of the GCA of the parents (Singh et al. 2004). Thus, when the GCA effect of a set of genotypes is important, a small number of such genotypes can be used as parents for hybridisation. On the other hand, if SCA is more important, a large number of parents will be required to produce a large number of the F_1 families from which superior recombinants will be selected (Poehlman and Sleper 2006).

Several studies have used the diallel mating design to understand the genetics of various traits in cassava, including the three parallel diallel crosses developed and tested in three contrasting environmental conditions in Colombia to study traits of commercial importance in cassava (Cach et al. 2005; Calle et al. 2005; Jaramillo et al. 2005). These studies indicated lower GCA:SCA ratios (< 5.0) for fresh root yield, while the same ratio was relatively higher (> 5.8) for DMC. Later reports by Kamau et al. (2010) and Parkes et al. (2013) similarly indicated higher GCA effects for root dry matter than for fresh root yield. Collectively, these results indicate a generally higher realised heritability for DMC than for fresh root yield, which also reflects the relative ease of improving the former (Kawano et al. 1998). In the case of resistance to diseases in Africa, diallel studies indicated wide variation for GCA:SCA ratios. For instance, Kamau et al. (2010) reported a GCA:SCA variance ratio of 1.1 for cassava mosaic disease (CMD) while Were et al.

(2012) reported a variance ratio of 12.0 for the same disease. On the other hand, Zacarias and Labuschagne (2010) reported a GCA:SCA variance ratio of 0.8 for the deadly cassava brown streak disease (CBSD), which was sharply contrasted by the ratio of 21.5 reported by Kulembeka et al. (2012). It becomes apparent that more studies are warranted to resolve such disparities in a crop where information on traditional genetics considerably lags behind molecular breeding efforts (Ceballos et al. 2015).

It is worth noting that the low capacity of cassava seedlings to generate sufficient planting materials for clonal trials tends to limit field experiments to one or a few locations and restricts the number of replications. In such scenarios, confounding effects of the environment on expression of quantitative traits may estimate the genetic effects with low precision (Viana et al. 1999; Ortiz et al. 2001; Ceballos et al. 2004). An additional drawback of diallel analysis is failure of the method to estimate non-allelic interactions, resulting in underestimation of the genetic nature of the character under study (Viana 2000; Singh et al. 2004). For example, significant epistatic effects have been reported for fresh root yield and performance of cassava in acidic conditions (Cach et al. 2005; Pérez et al. 2005). Another practical challenge of the diallel method is the large number of crosses that can be generated in the mating scheme. In this case, requirements for space, seed and labour involved in performing crosses and managing field trials limit the number of parents to no more than 8-10 (Stuber 1980). Despite these limitations, the diallel method offers breeders a practical avenue to identify and use parental lines with superior genetic values, which increases the chances of generating progeny with increased levels of traits (Nduwumuremyi et al. 2013).

2.4 Outline of cassava breeding scheme

Through various research consortia, cassava breeders across countries and continents have shown outstanding consistency in the general areas of priority for breeding, most of which point to the need for increased yield potential and resistance to biotic stresses. Frequently mentioned breeding objectives revolve around improvement of fresh and dry root yield, root DMC, resistance to principal local pests and diseases, tolerance to adverse soil and climatic conditions, good plant type and stake quality and other quality traits targeting increased adoption and utilisation (Fukuda et al. 2002; Ceballos et al. 2004). Considerable research is now being directed to enhance traits of nutritional value (such as provitamin A carotenoids) and industrial use (such as starch), which, respectively, are envisioned to overcome vitamin A deficiencies common among

resource-poor peasants dependent on cassava and enhance commercial production of the crop (Nassar and Ortiz 2010; Carvalho et al. 2011).

A typical breeding scheme in which full-sib or half-sib families form the baseline population for selection requires seven fundamental stages that are outlined in Figure 2.1. Although CIAT and IITA use different conventional breeding schemes, general features are identical to that in Figure 2.1 (IITA 1990; Ceballos et al. 2012a). This breeding scheme has been widely adopted by national cassava breeding programmes across SSA.

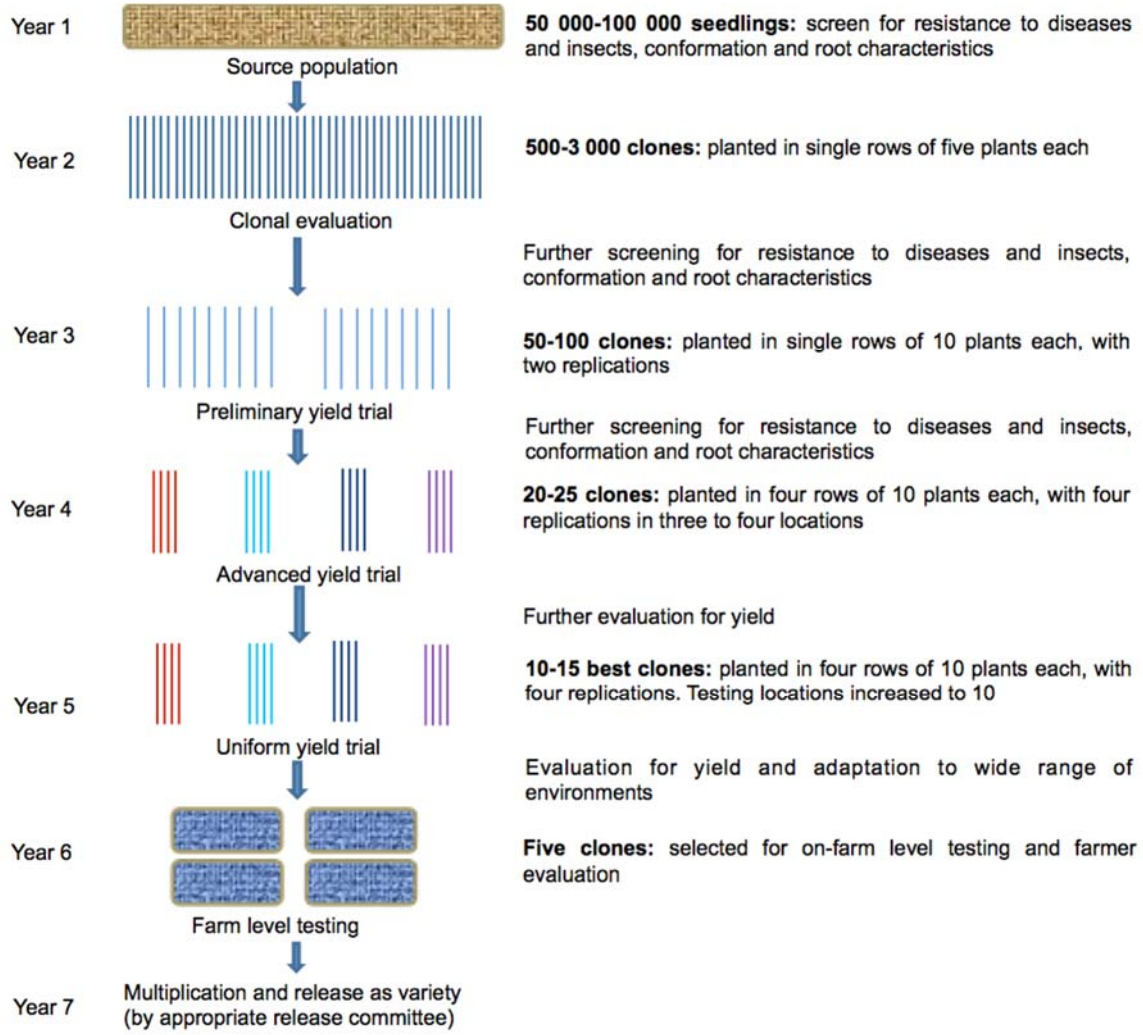


Figure 2.1 Conventional cassava breeding scheme (IITA 1990).

Common features of these breeding schemes are (1) reduced number of clones per advanced evaluation stage, (2) farmer participation in the final stages, (3) selection of clones with broad adaptability and non-location-specific selection and (4) differential experiment layouts, that is, use of un-replicated trials in earlier stages at single locations and then replicated trials in later stages at different locations. Selection stages begin with

a large number of genotypes represented by a few plants and progress toward a few genotypes planted in large plots.

A notable feature of the breeding scheme in Figure 2.1 is the long duration (≥ seven years) required to get material ready for release. This scheme illustrated in Figure 2.1 does not include time required for generating breeding populations in the form of botanical seeds through crossing, which requires no less than 12 months, essentially meaning at least eight years are required to develop and release a cassava variety. This situation, though undesirable, arises from the fact that several cycles of clonal propagation are required to raise sufficient planting material for conducting replicated trials (Ceballos et al. 2004). The replicated trials provide grounds for increasing selection precision and level of confidence so that valid decisions are made on clones advanced.

Breeding programmes with limited resources face an immense uphill task in implementing this scheme, with regard to logistical requirements, including finances (Nassar and Ortiz 2006). It is also possible that outbreaks of new biotic challenges may appear at advanced selection stages, to which the elite material being advanced may succumb. An example was the re-emergence of CBSD in Uganda in 2004 (Alicai et al. 2007). Before CBSD was noted, five candidate varieties had been selected from on-farm trials and earmarked for release to farmers in 2005. However, when these clones were screened for their reaction to the disease, only one of them was tolerant enough to be released (personal communication, Dr. Robert Kawuki, cassava breeder, Uganda). Such unforeseen misfortunes in cassava breeding require vigilance on the part of the breeder to minimise consequences associated with losses in finances and time (Rudi et al. 2010).

To overcome the challenge of the long breeding cycle, Kawuki et al. (2011) proposed a modified version of the above scheme to speed up the evaluation process and enhance local adoption of varieties in relatively shorter time. The most salient features of this modified scheme are (1) omission of the preliminary yield trial which is often conducted at a single location, (2) location-specific selection and (3) participation of farmers in evaluation and selection at both modified preliminary yield trials and modified uniform yield trials. Compared to the traditional breeding approach, the specific modifications in this scheme are: (1) the preliminary yield trial is conducted at several locations (six, in this case) and (2) the uniform yield trial is conducted with unequal number of genotypes depending on the location-specific selections and a larger number of replications (in this case, four replications per site). Through this scheme, the authors accomplished a

variety development process, including the participatory selection with farmers, in five years, which is a shorter period compared to the traditional CIAT and IITA schemes (Ceballos et al. 2012a).

Further research efforts have been made toward overcoming the long breeding cycle of cassava by using various tools to aid the conventional breeding approach. For example, the transgenic approach (Taylor et al. 2004) has been used to transform cassava for resistance to CBSD (Yadav et al. 2011; Ogwok et al. 2012) and increased expression of provitamin A carotenoids in roots (Adenle et al. 2012; Telengech et al. 2014). Genetically transformed genotypes are expected to retain their desirable agronomic attributes so that the transgenic plants would be rapidly multiplied for on-farm evaluation and subsequent selection for release in a remarkably short time (Adenle et al. 2012). However, for most countries expected to be the primary beneficiaries of the genetically modified cassava, particularly in Africa, several issues such as poor public perceptions and lack of functional laws to enable commercial production of transgenic plants are yet to be fully resolved (Adenle 2014). This situation has limited research with transgenic materials to laboratories and confined field trials.

Another potential strategy to save time in cassava breeding is the use of DH technology. The DH approach is a shorter route, potentially taking 2-3 years as opposed to the conventional 9-10 years, to develop completely homozygous cassava lines (Ceballos et al. 2007). The use of inbred progenitors in cassava breeding would make it possible to shift the current phenotypic recurrent selection method to line improvement, which would make practical development of outstanding hybrids a practical venture, as opposed to finding them by trial and error (Rojas et al. 2009; Ceballos et al. 2015). However, the DH technology for cassava is at a proof of concept level, with the current efforts dedicated to achieving a breakthrough in the required tissue culture protocols (Perera et al. 2014).

On the other hand, a team of researchers implementing the Next Generation Cassava Breeding Project (http://nextgencassava.org) is currently optimising genomic selection (GS) tools to enhance the efficiency of cassava breeding (Tecle et al. 2014). The on-going efforts focus on developing and validating the GS prediction models through associating marker information with phenotypic information for estimating genomic estimated breeding values (GEBVs) of genotypes (Jannink et al. 2010). With such prediction models at hand, a breeder would only need to genotype a seedling population and select genotypes with superior GEBVs for further recombination at clonal evaluation stage or for rapid multiplication and on-farm testing for possible release (Oliveira et al.

2012). This approach would dramatically reduce the entire breeding cycle to less than five years, which inevitably would save costs while accelerating genetic gains (Jannink et al. 2010; Ceballos et al. 2015). This advanced breeding approach has proved useful in animal breeding (Zhang et al. 2011) and crops such as wheat (Poland et al. 2012b), maize (Crossa et al. 2013) and rice (Xu et al. 2014).

2.5 Breeding for high dry matter and carotenoid content in cassava

Along with fresh root yield, breeding for high DMC is a typical objective of most cassava breeding programmes because of the effect of these traits on adoption of new varieties for food, feed and industrial raw materials (Awotide et al. 2014). DMC in cassava roots varies from 10-57%, depending on the genotype, age of the crop at harvest and prevailing environmental conditions (Babayoko et al. 2009; Ceballos et al. 2011; Ebah-Djedji et al. 2012). It is this wide variability for DMC that provides potential for improving the trait through simple breeding techniques such as phenotypic mass selection. On average, about 90% of root dry matter is carbohydrate, with 4% crude fibre, 3% ash, 2% crude protein and 1% fat (Salvador et al. 2014). Consequently, dry matter becomes an important trait for cassava producers because the crop is grown largely for its carbohydrate content (Awotide et al. 2014). Iglesias and Hershey (1994) reported DMC and fresh root yield as uncorrelated traits. However, the study by Ntawuruhunga et al. (2001) suggested a significantly negative correlation (r = -0.25) between root weight and DMC, supporting the uncertainty on whether high DMC levels can be maintained when yields are high (Kawano et al. 1987; Kawano 2003).

Genetic improvement of cassava for nutritional traits is a recent shift in the breeding paradigm, aimed at bridging the gap between agricultural research and nutrition (Mayer et al. 2008; Nassar and Ortiz 2010; Hefferon 2015). Initial efforts on breeding cassava for high carotenoid content focused on identification and characterisation of the useful genetic variation for carotenoids in the crop (Iglesias et al. 1997; Chávez et al. 2000). For example, Chávez et al. (2005) reported significant variation for carotenoid content (ranging from 1.02-10.4 µg g^{-1}) in a large collection (2 457 clones) of cassava sourced from different geographical regions. These pioneering efforts revealed important variations for carotenoids in cassava, with Nassar et al. (2007) reporting high levels of ß-carotene (4 µg g^{-1}) and lycopene (5 µg g^{-1}) in some landraces in Brazil, the centre of diversity for cassava. These landraces, including the carotene-rich UnB 400, have acquired large diversity in relation to quality traits, providing the basis for genetic improvement of cassava through hybridisation and selection (Nassar et al. 2009).

During the last 15 years, tremendous breeding efforts coordinated by CIAT and IITA have focused on exploiting the useful genetic variability for developing cassava varieties that accumulate high levels of provitamin A carotenoids (Pfeiffer and McClafferty 2007; Njoku et al. 2011). These efforts, developed under the HarvestPlus Project, have led to release of provitamin A cassava varieties to farmers in Nigeria (Njoku et al. 2011). Through a rapid cycling recurrent selection programme, remarkable genetic gains have been realised for carotenoid content, with Ceballos et al. (2013) reporting a maximum total carotenoid content of 25.8 μg g^{-1} in cassava roots. Some of this improved germplasm has been shared with cassava breeding programmes in SSA, including the provitamin A germplasm being utilised for breeding in Uganda (Esuma et al. 2012).

It is important to note that, compared to breeding efforts targeting release of cassava varieties with increased agronomic performance, progress in deployment of provitamin A varieties has been slow. The primary reason for this slow pace appears to be the tendency of low DMC to co-occur with high carotenoids content in roots (Njoku et al. 2015). The genetic basis of this undesirable relationship is unknown (Welsch et al. 2010). Efforts to underpin this hurdle through alternative breeding approaches, including genetic transformation (Zhu et al. 2007; Telengech et al. 2014), do not seem to point towards reversing the trend. Such a scenario presents some difficult situations for breeders to deal with because increasing carotenoid content would compromise DMC and reduce farmers' preference for provitamin A cassava varieties, especially where high DMC is preferred (Njukwe et al. 2013; Awotide et al. 2014). Nonetheless, Ceballos et al. (2013) have demonstrated the use of rapid cycling recurrent selection for exploiting the transgressive segregation pattern for both DMC and carotenoid content, which indicates an opportunity for introducing these two traits into a single genetic background.

2.6 Inheritance of dry matter and carotenoid content in cassava

The situation for cassava with respect to genetics of useful traits is unique in that while the molecular map based on whole genome sequencing is almost complete (Prochnik et al. 2012), knowledge of traditional genetics lags considerably behind (Ceballos et al. 2004; Ceballos et al. 2015). What is apparent is that most economically important characteristics in cassava appear to be quantitatively inherited and strongly affected by the environment. However, articles that comprehensively describe inheritance of these traits are few (Ceballos et al. 2004; Nassar and Ortiz 2006; Ceballos et al. 2012b).

Various studies have reported inheritance of DMC to be under the control of polygenic additive factors, but reports remain inconsistent. For example, Jaramillo et al. (2005)

observed higher GCA than SCA effects for DMC, harvest index (HI) and plant architecture, which indicated that additive genetic effects were more important than non-additive genetic factors for inheritance of these traits. Contrary to this, Pérez et al. (2005) and Kamau et al. (2010) reported non-significant GCA effects for DMC. Similarly, Easwari-Amma et al. (1995) reported that non-additive gene effects controlled inheritance of DMC and storage root yield. In sweet potato, additive gene action is slightly more predominant than non-additive gene action in predicting progeny performance for high DMC (Rukundo et al. 2013; Shumbusha et al. 2014). Collectively, reports indicate the need for further studies to understand the type of gene action controlling DMC for adopting more systematic approaches for enhancement of the trait in cassava.

Studies on inheritance of carotenoid content in cassava also appear to draw varying conclusions. For instance, Armstrong and Hearst (1996) suggested a two-gene system in inheritance of ß-carotene in cassava: one gene with complete dominance implicated in transportation of ß-carotene at high levels to roots and the other with partial dominance involved in accumulation of the trait in roots. Meanwhile, Iglesias et al. (1997) observed a pattern of quantitative variability within the root colour grades, leading to a suggestion that a number of genes with smaller effects are involved in the ß-carotene accumulation process. More recently, a population of 224 F_1 progeny and 160 F_1 reciprocals from crosses between white and yellow-flesh cassava genotypes segregated in the classical Mendelian ratio of 9:3:3:1 for white:light-yellow:yellow:deep-yellow (Akinwale et al. 2010). This segregation pattern suggested qualitative inheritance of carotenoid content in cassava, possibly controlled by two genes, with the recessive alleles increasing accumulation of carotenoids in roots. Welsch et al. (2010) identified a single allelic polymorphism in a phytoene synthase (PSY) gene associated with accumulation of coloured provitamin A carotenoids in cassava storage roots, which further indicated that few genes control expression of the trait in cassava. Similar qualitative inheritance patterns have been reported for carotenoid content in other crops like watermelon, *Cucumis melo* L. (Cuevas et al. 2009) and cucumber, *C. sativus* L. (Cuevas et al. 2010). Considering the segregation pattern for carotenoid content in cassava, it becomes imperative to adopt a recurrent selection approach for developing breeding populations using parental lines with superior GCA (Ceballos et al. 2013).

2.7 Phenotypic variation as influenced by genotype and environment

In the context of genetic improvement of a crop, heritable variation is defined as the

phenotypic variance attributable to genes that encode specific traits and can be transmitted to the following generation (Falconer and Mackay 1996). Particular environments in which genes are expressed will influence the phenotypic expression of the trait. Thus, the relationship between σ^2_P, σ^2_G, σ^2_E and the genotype by environment interaction (GEI) variance (σ^2_{GEI}) is a linear relationship expressed as: $\sigma^2_P = \sigma^2_G + \sigma^2_E + \sigma^2_{GEI}$ (Falconer and Mackay 1996; Acquaah 2012). The additive component of genetic variation is more important for a crop improvement programme as it reflects the cumulative effect of alleles from all gene loci influencing a character. σ^2_E is the non-heritable variation often associated with the heterogeneous nature of test environments or inconsistences in weather conditions such as temperature, rainfall and humidity (Falconer and Mackay 1996). Because such environmental factors are difficult to control in a routine plant breeding programme, they should be managed in a manner that reduces their confounding effect in estimating the genetic variability (Ceccarelli 1994).

The concept of GEI illustrates differential performance of genotypes with varying environmental conditions (Eberhart and Russell 1966). Highly quantitative traits such as yield and disease resistance are particularly influenced by environmental differences because these traits are controlled by collective expression of several genes, each of which is uniquely influenced by the different environments. This phenomenon implies that quantitative traits are characterised by low heritability (Falconer and Mackay 1996). Thus, selection for polygenic traits would require evaluation of test genotypes in diverse environmental conditions to determine the extent and nature of GEI involved in expression of a given character (Egesi et al. 2007; Aina et al. 2009). A significant GEI, especially where ranking of genotypes changes with the environment, is important in plant breeding (Flores et al. 1998).

Various statistical models for adaptability and stability analyses have been developed for identifying and selecting superior genotypes based on effects of GEI (Flores et al. 1998; Rencher 2002; Smith et al. 2005). For example, Cornelius et al. (1993) proposed the shifted multiplicative method based on clustering genotypes into groups within which crossover interactions (interactions involving rank changes) do not exist. But crossover interactions are frequently found to be an important component of GEI, a reason for which Lin and Butler (1990) proposed a cluster analysis for analysing two-way GEI data. The best linear unbiased prediction (BLUP) has also been used to shrink environmental effects in gauging the performance of genotypes in multi-location trials (Piepho 1998; Piepho and Mohring 2005; Piepho et al. 2007). However, the strong analytical attributes of the additive main effects and multiplicative interaction (AMMI) and genotype plus

genotype by environment interaction biplots (GGE biplots) have made them preferred tools for graphical visualisation and interpretation of performance of genotypes in the test environment. With these tools, it is easy to depict the discriminating ability of test environments, which facilitates selection of stable and adapted genotypes (Yau 1995; Yan and Kang 2002; Yan et al. 2007). A stable genotype will have a relatively narrow range of phenotypes in different environments due to buffering capacity of a genotype to environmental fluctuations, referred to as genetic homeostasis (Yan and Tinker 2006).

Cassava is subject to considerable GEI (Ceballos et al. 2012b; Ndung'u et al. 2012). Studies using different cassava genotypes tested in contrasting environments over the past several years have shown that yield and yield components are subject to strong GEI (Egesi et al. 2007; Tumuhimbise et al. 2014). DMC, an important consumer preference trait in Africa, has been reported to show high cultivar by year and cultivar by soil moisture content interactions (Huhn 1996; Boakye et al. 2013). Ssemakula and Dixon (2007) noted low influence of GEI on carotenoid content in cassava roots at harvest, but recommended further evaluation in more diverse environments. Recently, Maroya et al. (2012) reported significant interaction between 18 cassava genotypes and test environments for carotenoid content in Nigeria, which is similar to earlier reports on GEI for ß-carotene content in sweet potato (Bwanga et al. 2007). The collective take-home message from these studies for cassava breeding is that selection and recommendation of advanced clones for production should be based on data from multi-location evaluation trials, which provides the ability to identify genotypes with broad and specific adaptability (Acquaah 2012).

2.8 Biosynthetic pathway for carotenoids in plants

Carotenoids are a diverse group of coloured pigments naturally found in plants, algae, fungi and bacteria (Cazzonelli 2011; Eldahshan and Singab 2013). These compounds have been implicated to play essential roles in development, photosynthesis, root-mycorrhizal interactions and production of phytohormones in plants (Delgado-Vargas et al. 2000). Major mechanisms involved in carotenoid biosynthesis include localisation of biosynthetic enzymes in amyloplasts, catabolism and degradation, which occur throughout the life cycle of a plant. These processes are influenced by dynamic changes in composition matched to prevailing developmental requirements as well as response to external stimuli (Hannoufa and Hossain 2012; Shumskaya and Wurtzel 2013).

The carotenoid biosynthetic pathway contains major regulatory nodes that play the role of controlling the flux of metabolites into the pathway and alter flux through the pathway

(Armstrong and Hearst 1996; Hannoufa and Hossain 2012). Along the pathway, certain biosynthetic steps represent rate-limiting factors, which form critical target points for exploiting options of genetic engineering of carotenoid pathways in many crop species, including the case of golden rice (Giuliano et al. 2008; Giuliano 2014). An important biosynthetic mechanism is catabolism, which helps maintain carotenoids at physiologically important levels in photosynthetic tissues (Cazzonelli and Pogson 2010). Carotenoid cleavage dioxygenase enzymes (CCD1, CCD4, CCD7 and CCD8) and the cis-epoxycarotenoid dioxygenases (NCED2, NCED3, NCED5, NCED6 and NCED9) known to mediate catabolism, affect carotenoid composition and content in seeds of the highly studied Arabidopsis (Gonzalez-Jorge et al. 2013). Figure 2.2 illustrates an array of major genes and enzymes involved in the carotenoid biosynthetic pathway.

The *CCD* family of genes have been associated with depletion of the carotenoid pool in Arabidopsis seeds, chrysanthemum flowers and strawberries, (Auldridge et al. 2006; Garcia-Limones et al. 2008). In maize endosperm, high levels of *CCD1* transcripts have been reported to correlate with lower levels of carotenoids and a pronounced dosage effect resulting from copy number variation (Vallabhaneni et al. 2009; Silva et al. 2014). These *CCD* gene families act downstream to the activity pathway of the phytoene synthase gene known to control accumulation of provitamin A carotenoids in cassava (Welsch et al. 2010). Plant breeding could benefit from the rapid advance in biotechnology by integrating some of the molecular tools to dissect functional units within such carotenoid biosynthetic pathways, for exploring the possibility of pathway breeding to increase ß-carotene content in plants (Morandini and Salamini 2003; Silva et al. 2014).

2.9 Molecular marker technologies and cassava breeding

The intrinsic genetic nature of cassava, characterised by high heterozygosity, severe inbreeding depression and low seed production, constrains breeding efforts to improve the crop using conventional methods (Ceballos et al. 2004; Nassar and Ortiz 2006; Ceballos et al. 2015). The crop's phenology and several quantitative traits are highly influenced by the environment, which is a serious limitation to breeding programmes that rely on selection of recombinants based entirely on their phenotypes. Application of improved molecular technologies is expected to overcome these limitations by increasing selection efficiency while gaining time, which would substantially increase the success of cassava breeding programmes (Ferguson et al. 2011; Ceballos et al. 2015).

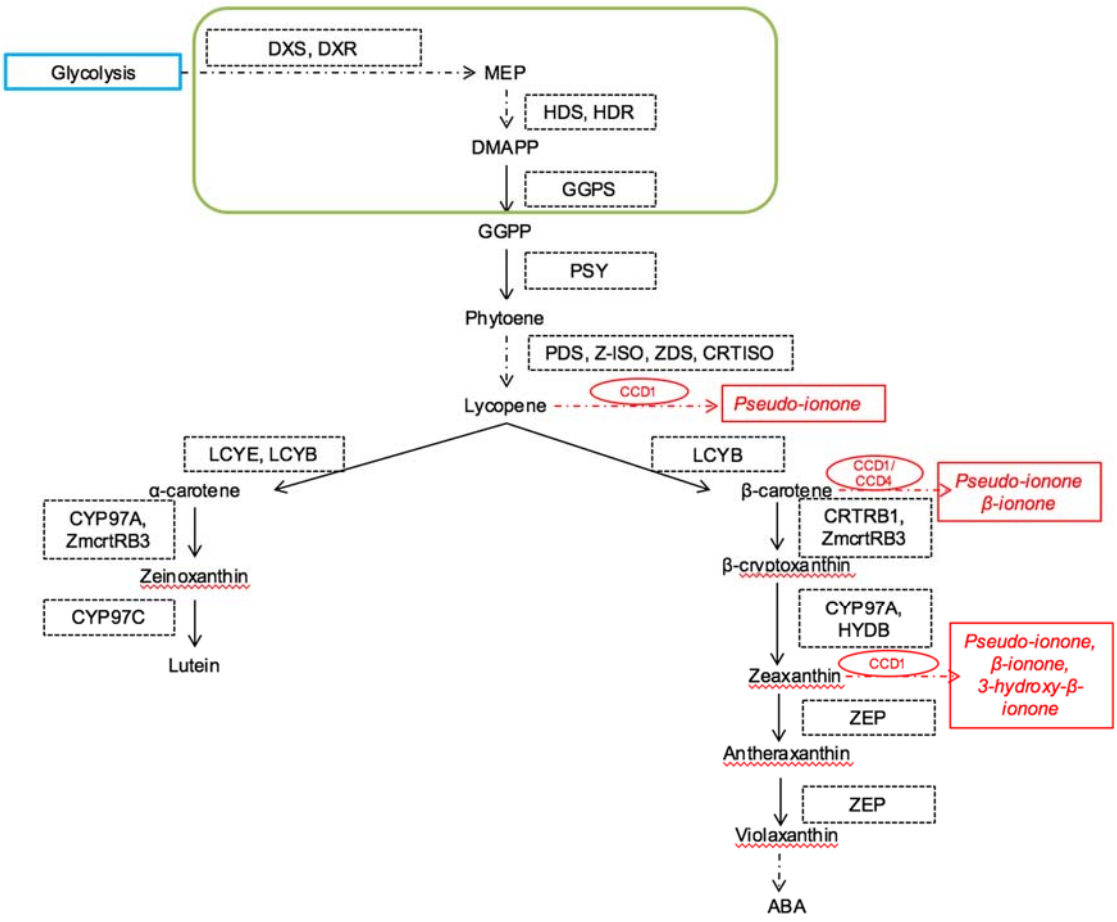


Figure 2.2 Carotenoid biosynthetic pathway based on Babu et al. (2013), Gonzalez-Jorge et al. (2013), Kandianis et al. (2013) and Arango et al. (2014).

Carotenoids are derived from products of glycolysis and the isoprenoid pathway (green blox). Provitamin A carotenoids are highlighted in red. CCD1 in red circles indicates the carotenoid substrates of CCD1 (Sun et al. 1996), and the corresponding catabolism products are highlighed in red. CCD4 is included as per the recent findings by Gonzalez-Jorge et al. (2013). Abbreviated intermediates: MEP = methyl-erythriol 4-phosphate, DMAPP = dimethylallyl diphosphate, GGPP = geranyl geranyl pyrophosphate, ABA = abcisic acid. Enzymes are in dashed boxes and are defined as: DXS = 1-deoxy-D-xtylulose-5-phosphate, DXR = deoxy-D-xtylulose-5-phosphate reductoisomerase, HDS = 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase, HDR = 4-hydroxy-3-methilbut-2-en-1-yl diphosphate reductase, GGPS = geranyl geranyl pyrophosphate synthase, PSY = phytoene synthase, PDS = phytoene desaturase, Z-ISO = 15-*cis*-zeta carotene isomerase, ZDS = ζ-carotene desaturase, CRTISO = carotenoid isomerase, LCYE = lycopene epsilon cyclase, LCYB = lycopene beta cyclase, CYP97A = carotenoid beta-ring hydroxilase, CYP97C = carotenoid ε-hydroxylase, CRTRB1 = ß-carotene hydroxylase, CCD1 = carotenoid cleavage dioxygenase 1, CCD4 = carotenoid cleavage dioxygenase 4, HYDB = ß-carotene hydroxylase, ZEP = zeaxanthin epoxidase.

2.9.1 First generation molecular markers

Successful use of molecular markers in cassava breeding dates back three decades when random amplified polymorphic DNA (RAPD) and restriction fragment length polymorphism (RFLP) markers were first used to study the genetic diversity within the genus *Manihot* (Beeching et al. 1993; Marmey et al. 1994). In later years, amplified fragment length polymorphism (AFLP) markers gained considerable importance and became extensively used to understand genetic differentiation in cassava (Colombo et al. 2000; Elias et al. 2000; Fregene et al. 2000). Another important application of AFLPs and RAPDs was in development of the first genetic linkage map of cassava, which opened the door to gain further insight into cassava genomics (Fregene et al. 1997). However, these markers had low throughput due to the high labour and time requirements they command, meaning they would gradually be replaced by simple sequence repeat (SSR) markers.

SSRs are short tandem repeat units of nucleotides that represent genomic points of variation within a species (Park et al. 2009). SSRs became more preferred for studying most crop species due to three major reasons: their (1) co-dominant and multi-allelic nature, (2) distribution throughout the genome and (3) ability for multiplexing on semi-automated systems (Varshney et al. 2009). Over the years, different research groups have developed a few thousand neutral SSR markers and SSRs derived from expressed sequence tags (ESTs) (Mba et al. 2001; Raji et al. 2009; Sraphet et al. 2011).

SSR markers have been effective for studying genetic diversity in the global cassava genetic resources (Ribeiro et al. 2011; Pariyo et al. 2013) and developing a cassava genetic map (Okogbenin et al. 2006; Rabbi et al. 2012). Perhaps the most remarkable application of SSR markers in cassava breeding is the marker-assisted introgression of resistance to cassava mosaic disease (CMD2) into Latin American germplasm for its genetic improvement in Africa (Okogbenin et al. 2007). Further validation of the SSR markers associated with the *CMD2* resistance gene in a set of African germplasm revealed usefulness of markers for selecting resistant cassava genotypes (Okogbenin et al. 2012). This was a major success story for MAS in cassava breeding (Ferguson et al. 2012). However, high costs of library construction and nucleotide sequencing in developing SSR markers led to the use of ESTs derived SSRs. This idea saw an increased amount of ESTs deposited in databases for various plants (Liang et al. 2009), which would allow rapid development of genic SSRs by low cost mining, for more efficient genetic mapping and MAS (Liang et al. 2009; Rabbi et al. 2012).

In the early 2000s, a micro-array DNA hybridisation method was developed as a novel marker technique to overcome the problem of low marker density and the high cost per data point associated with RFLPs, AFLPs, RAPDs and SSRs (Jaccoud et al. 2001). The technique, called diversity array technology (DArT), enables genotyping of hundreds of polymorphisms across a large number of individual plants over a short time and at low cost (Xia et al. 2005). However, DArT markers have limited usefulness for genetic characterisation of cassava due to their low discrimination power (Hurtado et al. 2008).

2.9.2 Next generation molecular markers

During the last decade, single nucleotide polymorphism (SNP) markers have gained significant popularity in plant molecular breeding. SNPs have been defined single nucleotide base changes as well as small insertions and deletions representing the most frequent form of naturally occurring genetic variation in populations (Mammadov et al. 2012). They typically occur in > 1% of the population because a single nucleotide variant that occurs at a frequency of ≤ 1% is considered a point mutation (Brookes 1999). The bi-allelic nature and evolutionary stability of SNPs make them individually less informative than SSRs (Syvanen 2001; Jehan and Lakhanpaul 2006), but their relatively high abundance within the genome and suitability for ultra-high throughput genotyping techniques have made them more informative and useful than the early generation markers (Rafalski 2002; Truong et al. 2012).

Initial studies focused on discovery of SNPs in cassava and their application to understanding sequence variation in the crop's genome. For example, Lopez et al. (2005) identified 136 SNPs from EST sequences and 50 SNPs from bacterial artificial chromosome end sequences. Kawuki et al. (2009) also studied genetic diversity in a panel of African cassava germplasm and identified 26 informative SNPs. These studies reported high frequency of SNPs: on average one SNP within every 62 and 121 bp for populations studied by Kawuki et al. (2009) and Lopez et al. (2005), respectively. The high frequency of SNPs in cassava is comparable with those in other higher crop species like maize (Ching et al. 2002) and grapevine (Salmaso et al. 2004). Ferguson et al. (2012) also identified and validated > 1 000 EST-derived SNP markers in cassava and these were incorporated into the crop's genome sequence v4.1.

The rapid advancement in next generation sequencing (NGS) technologies has reduced the cost of DNA sequencing to the point that genotyping-by-sequencing (GBS) is now practical for high diversity crops with large genome sizes (Varshney et al. 2009; Poland and Rife 2012). The basic advantage of GBS is the simplicity and the highly multiplexed

nature which allows construction of reduced representation libraries for the Illumina NGS platform (Elshire et al. 2011). This feature allows the generation of large numbers of SNPs for subsequent genotyping (Beissinger et al. 2013). The GBS protocol uses restriction enzymes to reduce genome complexity and DNA barcodes to allow parallel sequencing of scores of genotypes, thereby not only increasing the SNP calling accuracy but also decreasing cost (Elshire et al. 2011; Poland et al. 2012a). Compared to other molecular techniques, cost of GBS is low for a large number of genotypes, making it an affordable platform for researchers with limited financial resources (Poland and Rife 2012). Other advantages of the platform include reduced sample handling, fewer polymerase chain reactions (PCR) and purification steps, no size fractionation, no reference sequence limits, efficient barcoding and the ease to scale up (Davey et al. 2011).

To fully exploit benefits of the GBS platform, a bioinformatics pipeline referred to as trait analysis by association, evolution and linkage - GBS (TASSEL-GBS) has been designed and optimised for various crop species for the efficient processing of raw GBS sequence data into SNP genotype calls (Glaubitz et al. 2014). The TASSEL-GBS pipeline is user-friendly to small research programmes because it can run on modest computing resources such as desktop or laptop machines with only 8 GB of RAM, which are commonly available to such programmes. Another advantage of the pipeline is its ability to offer scalability from small to extremely large studies, making it possible to score hundreds of thousands or even millions of SNPs in up to 100 000 individuals. Such large-scale genotyping is typically required for large breeding programmes or genetic surveys (Poland and Rife 2012). Thus, the TASSEL-GBS pipeline helps to accelerate the breeding process, requiring rapid turnover from tissue collection to SNP genotypes.

In the near future, cassava genetic improvement could benefit from the powerful features of NGS technologies that would allow implementation of (1) genomic diversity studies (Peterson et al. 2014), (2) genome-wide association studies (GWAS) (He et al. 2014; Ceballos et al. 2015), (3) genetic linkage analysis and (4) genomic selection in cassava under large-scale breeding programmes (Oliveira et al. 2012). A case in time is when Rabbi et al. (2014) used the TASSEL-GBS platform to map 772 SNPs across 19 linkage groups and anchored 313 unique scaffolds in cassava genome sequence v4.1.

Several analytical tools have been developed to facilitate GWAS. The TASSEL software is inbuilt with bioinformatics algorithms that allow for computation of marker-trait association statistics by fitting genotype and phenotype data into mixed linear models,

with corrections for population structure (Glaubitz et al. 2014). Some of the challenges associated with TASSEL, such as a need for computers with RAM capacity ≥ 8 GB, are overcome by the genome association prediction integrated tool (GAPIT) software (Lipka et al. 2012) developed in R (R Development Core Team 2010). Other statistical tools for performing GWAS in R include qqman (Turner 2014), EMMA (Kang et al. 2008), GenAbel (Aulchenko et al. 2007) and Plink (Rentería et al. 2013), which vary in terms of the computational efficiency or quality of data outputs.

These analytical tools typically generate the quantile-quantile (Q-Q) and Manhattan plots for diagnosis of marker-trait association signals (Luo et al. 2011). The Q-Q plot helps in assessment of how well a model used in GWAS accounts for population structure and familial relatedness. It is generated by plotting negative logarithms of *P*-values from the model fitted in GWAS against their expected values under the null hypothesis of no association. The Manhattan plot is generated by plotting the negative logarithms of the *P*-values against the genomic position of each SNP, so that large peaks passing certain thresholds (e.g., 5% Bonferroni correction) suggest strong association between a genomic region and the trait studied (Balding 2006).

In a highly heterozygous crop like cassava, marker-assisted breeding inevitably increases response to selection, especially for complex traits (Ceballos et al. 2015). In such a case, practical application of GBS would be in uncovering the genetic basis of the important traits for identification of genes and pathways associated with traits (He et al. 2014). An application of the NGS platform in cassava genomics research is the high-resolution mapping of resistance to cassava mosaic geminiviruses, in which the GBS method was used to generate SNP data for mapping that led to identification of a single locus depicting monogenic resistance to CMD2 in the crop (Rabbi et al. 2014). Elsewhere, a study involving comprehensive genotyping of 2 815 maize inbred accessions revealed strong association between some SNPs and known candidate genes for kernel colour, sweetness and flowering time (Romay et al. 2013). Relatedly, Owens et al. (2014) and Suwarno et al. (2015) used GWAS to identify functional loci within the maize genome that explain phenotypic variation in carotenoid content. These studies are motivations for the application of NGS technologies to accelerate genetic improvement of cassava.

2.10 Summary

The literature discussed in this chapter provides further insight into the progress, challenges and opportunities within cassava breeding communities across the world.

Overall, biofortification of staple crops is a new paradigm shift in research that seeks to close the gap between agriculture and health by enhancing nutrition. Biofortified cassava would undoubtedly alleviate some of the chronic nutritional challenges afflicting resource-poor peasants, but would require more research efforts committed towards generating varieties that meet the needs of farmers. In the context of breeding cassava for provitamin A carotenoids, there is an apparent need for more information on classical genetics of carotenoid content and how it relates with other quality traits in the crop. Systematic genetic studies would generate such useful information as the basis for unravelling functional genetic factors controlling the traits.

2.11 References

Acquaah G (2012) Principles of Plant Genetics and Breeding. Second edition. John Willey and Sons Ltd, UK

Adenle AA (2014) Stakeholders' perceptions of GM technology in West Africa: assessing the responses of policy makers and scientists in Ghana and Nigeria. Journal of Agricultural & Environmental Ethics 27:241-263

Adenle AA, Aworh OC, Akhromah R, Parayil G (2012) Developing GM super cassava for improved health and food security: future challenges in Africa. Agriculture and Food Security 1:11

Aguirre JD, Hine E, McGuigan K, Blows MW (2014) Comparing G: multivariate analysis of genetic variation in multiple populations. Heredity 112:21-29

Aina OO, Dixon AG, Ilona P, Akinrinde EA (2009) GxE interaction effects on yield and yield components of cassava (landraces and improved) genotypes in the savannah regions of Nigeria. African Journal of Biotechnology 8:4933-4945

Akinbo O, Labuschagne M, Fregene M (2011) Increased storage protein from interspecific F1 hybrids between cassava (*Manihot esculenta* Crantz) and its wild progenitor (*M. esculenta* ssp. *flabellifolia*). Euphytica 185:303-311

Akinwale MG, Aladesanmwa RD, Akinyele BO, Dixon AGO, Odiyi AC (2010) Inheritance of ß-carotene in cassava (*Manihot esculenta* Crantz). International Journal of Genetics and Molecular Biology 2:198-201

Alicai T, Omongo CA, Maruthi MN, Hillocks RJ, Baguma Y, Kawuki R, Bua A, Otim-Nape GW, Colvin J (2007) Re-emergence of cassava brown streak disease in Uganda. Plant Disease 91:24-29

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

Allem AC (2002) The origin and taxonomy of cassava. In: Hillocks RJ, Thresh JM, Bellotti AC (eds) Cassava: Biology, Production and Utilization. vol 1. CAB International, UK, pp 1-16

Allem AC, Mendes RA, Salomão AN, Burle ML (2001) The primary gene pool of cassava (*Manihot esculenta* Crantz subspecies *esculenta,* Euphorbiaceae). Euphytica 120:127-132

Alves AAC (2002) Cassava botany and physiology. In: Hillocks RJ, Thresh JM, Bellotti AC (eds) Cassava: Biology, Production and Utilization, vol 1. CAB International, UK, pp 67-89

Arango J, Jourdan M, Geoffriau E, Beyer P, Welsch R (2014) Carotene hydroxylase activity determines the levels of both alpha-carotene and total carotenoids in orange carrots. Plant Cell 26:2223-2233

Armstrong GA, Hearst JE (1996) Genetics and molecular biology of carotenoid pigment synthesis. Serial Review 10:228-238

Arunachalam A (1976) The utility of covariance of combining ability in plant breeding. Theoretical and Applied Genetics 47:303-306

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R library for genome-wide association analysis. Bioinformatics 23:1294-1296

Auldridge ME, Block A, Vogel JT, Dabney-Smith C, Mila I, Bouzayen M, Magallanes-Lundback M, DellaPenna D, McCarty DR, Klee HJ (2006) Characterization of three members of the *Arabidopsis* carotenoid cleavage dioxygenase family demonstrates the divergent roles of this multifunctional enzyme family. Plant Journal 45:982-993

Awotide BA, Abdoulaye T, Alene A, Manyong VM (2014) Assessing the extent and determinants of adoption of improved cassava varieties in south-western Nigeria. Journal of Development and Agricultural Economics 6:376-385

Babayoko S, Tschannen A, Nindjin C, Dao D, Girardin O, Assa A (2009) Impact of water stress on fresh tuber yield and dry matter content of cassava (*Manihot esculenta* Crantz) in Cote d'Ivoire. African Journal of Agricultural Research 4:21-27

Babu R, Rojas NP, Gao S, Yan J, Pixley K (2013) Validation of the effects of molecular marker polymorphisms in LcyE and CrtRB1 on provitamin A concentrations for 26 tropical maize populations. Theoretical and Applied Genetics 126:389-399

Balding DJ (2006) A tutorial on statistical methods for population association studies. Nature Reviews Genetics 7:781-791

Beeching JR, Marmey P, Gavalda MC, Noirot M, Haysome HR, Hughes MA, Charrier A (1993) An assessment of genetic diversity within a collection of cassava (*Manihot*

esculenta Crantz) germplasm using molecular markers. Annals of Botany 72:515-520

Beissinger TM, Hirsch CN, Sekhon RS, Foerster JM, Johnson JM, Muttoni G, Vaillancourt B, Buell CR, Kaeppler SM, de Leon N (2013) Marker density and read depth for genotyping populations using genotyping-by-sequencing. Genetics 193:1073-1081

Bengtsson BO (2003) Genetic variation in organisms with sexual and asexual reproduction. Journal of Evolutionary Biology 16:189-199

Bernardo R (2010) Breeding for Quantitative Traits in Plants. Second edition. Stemma Press, Woodbury, Minnesota

Boakye B, Ofori K, Asante IK, Parkes EY (2013) Performance of nine cassava (*Manihot esculanta* Crantz) clones across three environments. Journal of Plant Breeding and Crop Science 5:48-53

Bos I, Caligari P (2008) Selection Methods in Plant Breeding. Second edition. Springer, Dordrecht, The Netherlands

Brookes AJ (1999) The essence of SNPs. Gene 234:177-186

Bwanga Z, Mataa M, Msabaha M (2007) Quality and yield stability of orange fleshed sweet potato (*Ipomoea batatas*) varieties grown in different agroecologies. African Crop Science Conference Proceedings 8:339-345

Cach NT, Pérez JC, Lenis JI, Calle F, Morante N, Ceballos H (2005) Epistasis in the expression of relevant traits in cassava (*Manihot esculenta* Crantz) for subhumid conditions. Journal of Heredity 96:586-592

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

Carvalho LJ, Vieira EA, Fialho JF, Souza CRB (2011) A genomic assisted breeding program for cassava to improve nutritional quality and industrial traits of storage root. Brazilian Society of Plant Breeding 11:289-296

Cazzonelli CI (2011) Carotenoids in nature: insights from plants and beyond. Functional Plant Biology 38:833-847

Cazzonelli CI, Pogson BJ (2010) Source to sink: regulation of carotenoid biosynthesis in plants. Trends in Plant Science 15:266-274

Ceballos H, Fregene M, Pérez JC (2007) The use of doubled-haploids in cassava breeding. In: Proceedings of the 13th ISTRC Symposium, Arusha, Tanzania. pp 102-108

Ceballos H, Hershey CH, Becerra-López-Lavalle LA (2012a) New approaches to cassava breeding. Plant Breeding Reviews 36:427-504

Ceballos H, Iglesias CA, Perez J, Dixon AGO (2004) Cassava breeding: opportunities and challenges. Plant Molecular Biology 56:503-516

Ceballos H, Kawuki RS, Gracen VE, Yencho GC, Hershey CH (2015) Conventional breeding, marker-assisted selection, genomic selection and inbreeding in clonally propagated crops: a case study for cassava. Theoretical and Applied Genetics 128:1647-1667

Ceballos H, Kulakow P, Hershey C (2012b) Cassava breeding: current status, bottlenecks and the potential of biotechnology tools. Tropical Plant Biology 5:73-87

Ceballos H, Luna J, Escobar AF, Ortiz D, Pérez JC, Sánchez T, Pachón H, Dufour D (2011) Spatial distribution of dry matter in yellow fleshed cassava roots and its influence on carotenoid retention upon boiling. Food Research International 45:52-59

Ceballos H, Morante N, Sánchez T, Ortiz D, Aragón I, Chávez AL, Pizarro M, Calle F, Dufour D (2013) Rapid cycling recurrent selection for increased carotenoids content in cassava roots. Crop Science 53:23-42

Ceccarelli S (1994) Specific adaptation and breeding for marginal conditions. Euphytica 77:205-219

Chávez AL, Bedoya JM, Sánchez T, Iglesias CA, Ceballos H, Roca W (2000) Iron, carotene, and ascorbic acid in cassava roots and leaves. Food and Nutrition Bulletin 21:410-414

Chávez AL, Sánchez T, Jaramillo G, Bedoya JMI, Echeverry J, Bolaños EA (2005) Variation of quality traits in cassava roots evaluated in landraces and improved clones. Euphytica 143:125-133

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics 3:14

Colombo C, Second G, Charrier A (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

Cornelius PL, Van Sanford DA, Seyedsadr MS (1993) Clustering cultivars into groups without rank-change interactions. Crop Science 33:1193-1200

Cowling WA, Léon J (2013) Sustainable plant breeding. Plant Breeding 132:1-9

Crossa J, Beyene Y, Kassa S, Pérez P, Hickey JM, Chen C, Campos G, Burgueno J, Windhausen VS, Buckler E, Jannink JL, Lopez Cruz MA, Babu R (2013) Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3: Genes, Genomes, Genetics 3:1903-1926

Crossa J, Campos L, Pérez P, Gianola D, Burgueno J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan J, Arief V, Banziger M, Braun HJ (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713-724

Cuevas HE, Song H, Staub JE, Simon PW (2009) Inheritance of beta-carotene-associated flesh color in cucumber (*Cucumis sativus* L.) fruit. Euphytica 171:301-311

Cuevas HE, Staub JE, Simon PW (2010) Inheritance of beta-carotene-associated mesocarp color and fruit maturity of melon (*Cucumis melo* L.). Euphytica 173:129-140

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12:499-510

Delgado-Vargas F, Jimenez AR, Paredes-Lopez O (2000) Natural pigments: carotenoids, anthocyanins, and betalains - characteristics, biosynthesis, processing, and stability. Critical Reviews in Food Science and Nutrition 40:173-289

Dudley JW, Moll RH (1969) Interpretation and use of estimates of heritability and genetic variances in plant breeding. Crop Science 9:257-262

Easwari-Amma CS, Sheela MN, Thankamma PK (1995) Combining ability analysis in cassava. Journal of Root Crops 21:65-71

Ebah-Djedji BC, Dje KM, Zohouri GP, Amani NG (2012) Effect of harvest period on starch yield and dry matter content from the tuberous roots of improved cassava (*Manihot esculenta* Crantz) varieties. Pakistan Journal of Nutrition 11:414-418

Eberhart SA, Russell WA (1966) Stability parameters for comparing varieties. Crop Science 6:36-40

Egesel CO, Wong JC, Lambert RJ, Rocheford TR (2003) Combining ability of maize inbreds for carotenoids and tocopherols. Crop Science 43:818-823

Egesi CN, Ilona P, Ogbe FO, Dixon AG (2007) Genetic variation and genotype x environment interaction for yield and other agronomic traits in cassava in Nigeria. Agronomy Journal 99:1137-1142

Eldahshan OA, Singab NAB (2013) Carotenoids. Journal of Pharmacognosy and Phytochemistry 2:225-234

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

Ellis RH, Hong TD, Roberts EH (1982) An investigation of the influence of constant and alternating temperature on the germination of cassava seed using a two-dimensional temperature gradient plate. Annals of Botany 49:241-246

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379

Esuma W, Rubaihayo P, Pariyo A, Kawuki R, Wanjala B, Nzuki I, Harvey JJW, Baguma Y (2012) Genetic diversity of provitamin A cassava in Uganda. Journal of Plant Studies 1:60-71

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics. Fourth edition. Longman, Essex, UK

Fehr WR (1993) Principles of Cultivar Development: Theory and Technique, vol 1. Macmillan Publishing Company, New York, USA

Ferguson ME, Hearne SJ, Close TJ, Wanamaker S, Moskal WA, Town CD, de Young J, Marri PR, Rabbi IY, de Villiers EP (2012) Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. Theoretical and Applied Genetics 124:685-695

Ferguson ME, Rabbi I, Kim D-J, Gedil M, Lopez-Lavalle LAB, Okogbenin E (2011) Molecular markers and their application to cassava breeding: past, present and future. Tropical Plant Biology 5:95-109

Flores F, Moreno MT, Cubero JI (1998) A comparison of univariate and multivariate methods to analyse GxE interaction. Field Crops Research 56:271-286

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

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

Fuente GN, Frei UK, Lubberstedt T (2013) Accelerating plant breeding. Trends in Plant Science 18:667-672

Fukuda WMG, Silva SO, Iglesias CA (2002) Cassava breeding. Crop Breeding and Applied Biotechnology 2:617-638

Garcia-Limones C, Schnabele K, Blanco-Portales R, Luz Bellido M, Caballero JL, Schwab W, Munoz-Blanco J (2008) Functional characterization of FaCCD1: a carotenoid cleavage dioxygenase from strawberry involved in lutein degradation during fruit ripening. Journal of Agricultural and Food Chemistry 56:9277-9285

Gelman A (2005) Analysis of variance - why it is more important than ever. The Annals of Statistics 33:1-53

Giuliano G (2014) Plant carotenoids: genomics meets multi-gene engineering. Current Opinion in Plant Biology 19:111-117

Giuliano G, Tavazza R, Diretto G, Beyer P, Taylor MA (2008) Metabolic engineering of carotenoid biosynthesis in plants. Trends in Biotechnology 26:139-145

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One 9:e90346

Gonzalez-Jorge S, Ha SH, Magallanes-Lundback M, Gilliland LU, Zhou A, Lipka AE, Nguyen YN, Angelovici R, Lin H, Cepela J, Little H, Buell CR, Gore MA, Dellapenna D (2013) Carotenoid cleavage dioxygenase4 is a negative regulator of beta-carotene content in *Arabidopsis* seeds. Plant Cell 25:4812-4826

Griffing B (1956) Concept of general and specific combining ability in relation to diallel crossing systems. Australian Journal of Experimental Biology & Medical Science 9:463-493

Hallauer AR, Carena MJ, Filho JBM (2010) Quantitative Genetics in Maize Breeding, vol 6. Springer, New York, USA

Halsey ME, Olsen KM, Taylor NJ, Chavarriaga-Aguirre P (2008) Reproductive biology of cassava (*Manihot esculenta* Crantz) and isolation of experimental field trials. Crop Science 48:49-58

Hannoufa A, Hossain Z (2012) Regulation of carotenoid accumulation in plants. Biocatalysis and Agricultural Biotechnology 1:198-202

Hashimoto-Freitas DY, Nassar NM (2013) Cytogenetic and anatomic behavior of cytochimeras and total polyploids in cassava. Genetics and Molecular Research 12:4879-4894

Hayman BI (1954) The theory and analysis of diallel crosses. Genetics 39:789-809

He J, Zhao X, Laroche A, Lu ZX, Liu H, Li Z (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in Plant Science 5:484

Hefferon KL (2015) Nutritionally enhanced food crops; progress and perspectives. International Journal of Molecular Sciences 16:3895-3914

Hill J, Becker HC, Tigerstedt PMA (1998) Quantitative and Ecological Aspects of Plant Breeding. Chapman and Hall, UK, London

Huhn M (1996) Non-parametric analysis of genotype x environment interactions by ranks. In: Kang MS, Gauch HG (eds) Genotype by Environment Interaction. CRC Press, Boca Raton, New York, pp 213-228

Hurtado P, Olsen KM, Buitrago C, Ospina C, Marin J, Duque M, de Vicente C, Wongtiem P, Wenzel P, Killian A, Adeleke M, Fregene M (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: Characterization and Utilization 6:208-214

Iglesias CA, Hershey CH (1994) Cassava breeding at CIAT: Heritability estimates and genetic progress in the 1980's In: Ofori F, Hahn SK (eds) Tropical Root Crops in a Developing Economy. ISTRC/ISHS, Wageningen, Netherlands, pp 149-163

Iglesias CA, Meyer J, Chávez AL, Calle F (1997) Genetic potential and stability of carotene content in cassava roots. Euphytica 94:367-373

IITA (1990) Cassava in Tropical Africa: A Reference Manual. International Institute of Tropical Agriculture, Ibadan, Nigeria

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Research 29:1-7

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Briefings in Functional Genomics 9:166-177

Jaramillo G, Morahan G, Calle F, Ceballos H, Arias B, Bellotti AC (2005) Diallel analysis in cassava adapted to the mid-altitude valleys environment. Crop Science 45:1058-1063

Jehan T, Lakhanpaul S (2006) Single nucleotide polymorphism (SNP) - methods and applications in plant genetics. Indian Journal of Biotechnology 5:435-459

Jennings DL, Iglesias C (2002) Cassava: biology, production and utilization In: Hillocks RJ, Thresh JM, Bellotti AC (eds) Breeding for Crop Improvement. vol 8. CAB International, UK, pp 149-166

Jiang Q, Webb SL, Yesudas CR, Bhandari HS, Narasimhamoorthy B, Bouton JH, Saha MC (2014) Variance components and heritability of biomass yield in switchgrass (*Panicum virgatum* L.) grown in the Southern Great Plains. Field Crops Research 168:148-155

Jones WO (1959) Manioc in Africa. Standard University Press. Stanford

Jos JS, Nair RB, Sreekumari MT (1990) Diverse types of male sterility in cassava. The Indian Journal of Genetics and Plant Breeding 50:19-25

Kamau J, Melis R, Liang M, Derera J, Shanahan P, Ngugi E (2010) Combining the yield ability and secondary traits of selected cassava genotypes in the semi-arid areas of Eastern Kenya. Journal of Plant Breeding and Crop Science 2:181-191

Kandianis CB, Stevens R, Liu W, Palacios N, Montgomery K, Pixley K, White WS, Rocheford T (2013) Genetic architecture controlling variation in grain carotenoid

composition and concentrations in two maize populations. Theoretical and Applied Genetics 126:2879-2895

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709-1723

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

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

Kawano K, Fukuda WMG, Cenpukdee U (1987) Genetic and environmental effects on dry matter content of cassava roots. Crop Science 27:69-74

Kawano K, Narintaraporn K, Narintaraporn P, Sarakarn S, Limsila A, Limsila J, Suparhan D, Sarawat V, Watananonta W (1998) Yield improvement in a multistage breeding program for cassava. Crop Science 38:225-332

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

Kawuki RS, Pariyo A, Amuge T, Nuwamanya E, Ssemakula G, Tumwesigye S, Bua A, Baguma Y, Omongo C, Alicai T, Orone J (2011) A breeding scheme for local adoption of cassava (*Manihot esculenta* Crantz). Journal of Plant Breeding and Crop Science 3:120-130

Kulembeka HP, Ferguson M, Herselman L, Kanju E, Mkamilo G, Masumba E, Fregene M, Labuschagne MT (2012) Diallel analysis of field resistance to brown streak disease in cassava (*Manihot esculenta* Crantz) landraces from Tanzania. Euphytica 187:277-288

Langlands BW (1966) Cassava in Uganda 1880-1920. Uganda Journal 30:211-218

Liang X, Chen X, Hong Y, Liu H, Zhou G, Li S, Guo B (2009) Utility of EST-derived SSRs in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species. BMC Plant Biology 9:35

Lin CS, Butler G (1990) Cluster analyses for analyzing two-way classification data. Agronomy Journal 82:344-348

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397-2399

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

Luo L, Boerwinkle E, Xiong M (2011) Association studies for next-generation sequencing. Genome Research 21:1099-1108

Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S (2012) SNP markers and their impact on plant breeding. International Journal of Plant Genomics 2012:1-11

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

Maroya NG, Kulakow P, Dixon AGO, Maziya-Dixon B, Bakare MA (2012) Genotype x environment interaction of carotene content of yellow-fleshed cassava genotypes in Nigeria. Journal of Life Sciences 6:595-601

Mayer JE, Pfeiffer WH, Beyer P (2008) Biofortified crops to alleviate micronutrient malnutrition. Current Opinion in Plant Biology 11:166-170

Mba REC, Stephenson P, Edwards K, Melzer S, Nkumbira J, Gullberg U, Apel K, Gale M, Tohme J, Fregene M (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

Morandini P, Salamini F (2003) Plant biotechnology and breeding: allied for years to come. Trends in Plant Science 8:70-75

Nassar NMA (2002) Cassava, *Manihot esculenta* Crantz, genetic resources: origin of the crop, its evolution and relationships with wild relatives. Genetics and Molecular Research 1:298-305

Nassar NMA (2003) Cassava, *Manihot esculenta* Crantz genetic resources: anatomy of a diversity center. Genetics and Molecular Research 2:214-222

Nassar NMA, Junior OP, Sousa MV, Ortiz R (2009) Improving carotenoids and amino acids in cassava. Recent Patents on Food, Nutrition & Agriculture 1:32-38

Nassar NMA, Ortiz R (2006) Cassava improvement: challenges and impacts. Journal of Agricultural Science 145:163-171

Nassar NMA, Ortiz R (2010) Breeding cassava to feed the poor. Scientific American Journal 302:78-82

Nassar NMA, Schwartz CA, Junior OP (2007) Cassava diversity in Brazil: the case of carotenoid-rich landraces. Genetics and Molecular Research 6:116-121

Ndung'u JN, Wachira FN, Kinyua MG, Lelgut DK, Okwaro H, Njau P, Obiero H (2012) Influence of the environment on cassava quality traits in central rift valley of Kenya. American Journal of Plant Sciences 3:1504-1512

Nduwumuremyi A, Tongoona P, Habimana S (2013) Mating designs: helpful tool for quantitative plant breeding analysis. Journal of Plant Breeding and Genetics 1:117-129

Njoku DN, Gracen VE, Offei SK, Asante IK, Egesi CN, Kulakow P, Ceballos H (2015) Parent-offspring regression analysis for total carotenoids and some agronomic traits in cassava. Euphytica 206:657-666

Njoku DN, Vernon G, Egesi CN, Asante I, Offei SK, Okogbenin E, Kulakow P, Eke-okoro ON, Ceballos H (2011) Breeding for enhanced ß-carotene content in cassava: constraints and accomplishments. Journal of Crop Improvement 25:560-571

Njukwe E, Hanna R, Kirscht H, Araki S (2013) Farmers perception and criteria for cassava variety preference in Cameroon. African Study Monographs 34:221-234

Ntawuruhunga P, Rubaihayo P, Whyte JBA, Dixon AGO, Osiru DSO (2001) Inter-relationships among traits and path analysis for yield components of cassava: a search for storage root yield indicators. African Crop Science Journal 9:599-606

Ogwok E, Odipio J, Halsey ME, Gaitan-Solis E, Bua A, Taylor NJ, Fauquet C, Alicai T (2012) Transgenic RNA interference (RNAi)-derived field resistance to cassava brown streak disease. Molecular Plant Pathology 13:1019-1031

Okogbenin E, Egesi CN, Olasanmi B, Ogundapo O, Kahya S, Hurtado P, Marin J, Akinbo O, Mba C, Gomez H, de Vicente C, Baiyeri S, Uguru M, Ewa F, Fregene M (2012) Molecular marker analysis and validation of resistance to cassava mosaic disease in elite cassava genotypes in Nigeria. Crop Science 52:2576-2586

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

Okogbenin E, Porto MCM, Egesi C, Mba C, Espinosa E, Santos LG (2007) Marker-assisted introgression of resistance to cassava mosaic disease into Latin American germplasm for the genetic improvement of cassava in Africa. Crop Science 47:1895-1904

Oliveira EJ, Resende MDV, Silva Santos V, Ferreira CF, Oliveira GAF, Silva MS, Oliveira LA, Aguilar-Vildoso CI (2012) Genome-wide selection in cassava. Euphytica 187:263-276

Olsen MK (2004) SNPs, SSRs and inferences on cassava's origin. Plant Molecular Biology 56:517-526

Ortiz R, Madsen S, Wagoire WW, Hill J, Chandra S, Stolen O (2001) Additive main effect and multiplicative interaction model for a diallel-cross analysis. Theoretical and Applied Genetics 102:1103-1106

Owens BF, Lipka AE, Magallanes-Lundback M, Tiede T, Diepenbrock CH, Kandianis CB, Kim E, Cepela J, Mateos-Hernandez M, Buell CR, Buckler ES, DellaPenna D, Gore MA, Rocheford T (2014) A foundation for provitamin A biofortification of maize: genome-wide association and genomic prediction models of carotenoid levels. Genetics 198:1699-1716

Pariyo A, Tukamuhabwa P, Baguma Y, Kawuki R, Alicai T, Gibson P, Kanju E, Wanjala B, Harvey J, Nzuki I, Rabbi IY, Ferguson M (2013) Simple sequence repeats (SSR) diversity of cassava in south, east and central Africa in relation to resistance to cassava brown streak disease. African Journal of Biotechnology 12:4453-4464

Park YJ, Lee JK, Kim NS (2009) Simple sequence repeat polymorphisms (SSRPs) for evaluation of molecular diversity and germplasm classification of minor crops. Molecules 14:4546-4569

Parkes EY, Fregene M, Dixon A, Peprah BB, Labuschagne MT (2013) Combining ability of cassava genotypes for cassava mosaic disease and cassava bacterial blight, yield and its related components in two ecological zones in Ghana. Euphytica 194:13-24

Perera PI, Ordonez CA, Lopez-Lavalle LA, Dedicova B (2014) A milestone in the doubled haploid pathway of cassava (*Manihot esculenta* Crantz): cellular and molecular assessment of anther-derived structures. Protoplasma 251:233-246

Pérez JC, Ceballos H, Calle F, Morante N, Gaitán W, Llano G, Alvarez E (2005) Within-family genetic variation and epistasis in cassava (*Manihot esculenta* Crantz) adapted to the acid-soils environment. Euphytica 145:77-85

Peterson G, Dong Y, Horbach C, Fu Y-B (2014) Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. Diversity 6:665-680

Pfeiffer WH, McClafferty B (2007) HarvestPlus: breeding crops for better nutrition. Crop Science 47:88-105

Piepho HP (1998) Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. Theoretical and Applied Genetics 97:195-201

Piepho HP, Mohring J (2005) Best linear unbiased prediction of cultivar effects for subdivided target regions. Crop Science 45:1151-1159

Piepho HP, Mohring J, Melchinger AE, Buchse A (2007) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209-228

Poehlman JM, Sleper DA (2006) Breeding Field Crops. Fifth edition. Blackwell Publishing Professional, Iowa, USA

Poland J, Brown PJ, Sorrells ME, Jannink JL (2012a) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7:e32253

Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-Villeda H, Sorrells M, Jannink J-L (2012b) Genomic selection in wheat breeding using genotyping-by-sequencing. The Plant Genome Journal 5:103-113

Poland J, Rife T (2012) Genotyping-by-sequencing for plant breeding and genetics. The Plant Genome Journal 5:92-102

Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, Rodriguez F, Fauquet C, Tohme J, Harkins T, Rokhsar DS, Rounsley S (2012) The cassava genome: current progress, future directions. Tropical Plant Biology 5:88-94

R Development Core Team (2010) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Australia

Rabbi I, Hamblin M, Gedil M, Kulakow P, Ferguson M, Ikpan AS, Ly D, Jannink J-L (2014) Genetic mapping using genotyping-by-sequencing in the clonally propagated cassava. Crop Science 54:1384-1396

Rabbi IY, Kulembeka HP, Masumba E, Marri PR, Ferguson M (2012) An EST-derived SNP and SSR genetic linkage map of cassava (*Manihot esculenta* Crantz). Theoretical and Applied Genetics 125:329-342

Rafalski AJ (2002) Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology 5:94-100

Raji AA, Anderson JV, Kolade OA, Ugwu CD, Dixon AG, Ingelbrecht IL (2009) Gene-based microsatellites for cassava (*Manihot esculenta* Crantz): prevalence, polymorphisms, and cross-taxa utility. BMC Plant Biology 9:118

Ramalho MAP, Carvalho BL, Nunes JAR (2013) Perspectives for the use of quantitative genetics in breeding of autogamous plants. ISRN Genetics 2013:1-6

Rencher AC (2002) Methods of Multivariate Analysis. Second edition. John Wiley & Sons, New York

Rentería ME, Cortes A, Medland SE (2013) Using PLINK for genome-wide association studies (GWAS) and data analysis. Methods in Molecular Biology 1019:193-213

Ribeiro MNO, Carvalho SP, Santos JB, Antonio RP (2011) Genetic variability among cassava accessions based on SSR markers. Crop Breeding and Applied Biotechnology 11:263-269

Rojas MC, Pérez JC, Ceballos H, Baena D, Morante N, Calle F (2009) Analysis of inbreeding depression in eight S1 cassava families. Crop Science 49:543-548

Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, Elshire RJ, Acharya CB, Mitchell SE, Flint-Garcia SA, McMullen MD, Holland JB, Buckler ES,

Gardner CA (2013) Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biology 14:55

Rudi N, Newton GW, Alwang J, Asumugha G (2010) Economic impact analysis of marker-assisted breeding for resistance to pests and postharvest deterioration in cassava. African Journal of Agricultural Research 4:110-122

Rukundo P, Shimelis H, Liang M, Gahakwa D (2013) Storage root formation, dry matter synthesis, accumulation and genetics in sweet potato. Australian Journal of Crop Science 7:2054-2061

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

Salvador EM, Steenkamp V, McCrindle CME (2014) Production, consumption and nutritional value of cassava (*Manihot esculenta* Crantz) in Mozambique: an overview. Journal of Agricultural Biotechnology and Sustainable Development 6:29-38

Shumbusha D, Tusiime G, Edema R, Gibson P, Adipala E, Mwanga ROM (2014) Inheritance of root dry matter content in sweet potato. African Crop Science Journal 22:69-78

Shumskaya M, Wurtzel ET (2013) The carotenoid biosynthetic pathway: thinking in all dimensions. Plant Science 208:58-63

Silva MR, Galli V, Dos Anjos ESSD, Rombaldi CV (2014) Carotenoid biosynthetic and catabolic pathways: gene expression and carotenoid content in grains of maize landraces. Nutrients 6:546-563

Singh RK, Pooni HS, Singh M, Bandopadhyaya A (2004) Mating designs and their implications for plant breeding. In: Jain HK, Kharkwal MC (eds) Plant Breeding: Mendelian to Molecular Approaches, vol 12. Springer, New York, USA, pp 523-534

Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. Journal of Agricultural Science 143:449-462

Sprague GF, Tatum LA (1942) General vs. specific combining ability in single crosses of corn. Journal of American Society of Agronomy 34:923-932

Sraphet S, Boonchanawiwat A, Thanyasiriwat T, Boonseng O, Tabata S, Sasamoto S, Shirasawa K, Isobe S, Lightfoot DA, Tangphatsornruang S, Triwitayakorn K (2011) SSR and EST-SSR-based genetic linkage map of cassava (*Manihot esculenta* Crantz). Theoretical and Applied Genetics 122:1161-1170

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

Stuber CW (1980) Mating designs, field nursery layouts and breeding records. In: Fehr WR, Hadley HH (eds) Hybridization of Crop Plants. American Society of Agronomy, Madison, WI, USA, pp 83-103

Sun Z, Gantt E, Cunningham FX (1996) Cloning and functional analysis of the ß-carotene hydroxylase of *Arabidopsis thaliana*. The Journal of Biological Chemistry 4:24349-24352

Suwarno WB, Pixley KV, Palacios-Rojas N, Kaeppler SM, Babu R (2015) Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. Theoretical and Applied Genetics 128:851-864

Syvanen A (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Reviews 2:930-942

Taylor N, Chavarriaga P, Raemakers K, Siritunga D, Zhang P (2004) Development and application of transgenic technologies in cassava. Plant Molecular Biology 56:671-688

Tecle IY, Edwards JD, Menda N, Egesi E, Rabbi IY, Kulakow P, Kawuki R, Jannink J, Mueller LA (2014) solGS: a web-based tool for genomic selection. BMC Bioinformatics 15:398

Telengech PK, Maling'a JN, Nyende AB, Gichuki ST, Wanjala BW (2014) Gene expression of beta carotene genes in transgenic biofortified cassava. Biotechnology 5:465-472

Truong HT, Ramos AM, Yalcin F, de Ruiter M, van der Poel HJ, Huvenaars KH, Hogers RC, van Enckevort LJ, Janssen A, van Orsouw NJ, van Eijk MJ (2012) Sequence-based genotyping for marker discovery and co-dominant scoring in germplasm and populations. PLoS One 7:e37565

Tumuhimbise R, Melis R, Shanahan P, Kawuki R (2014) Genotype x environment interaction effects on early fresh storage root yield and related traits in cassava. The Crop Journal 2:329-337

Turner SD (2014) qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. http://dx.doi.org/10.1101/005165. Accessed 12 December 2014

Vallabhaneni R, Gallagher CE, Licciardello N, Cuttriss AJ, Quinlan RF, Wurtzel ET (2009) Metabolite sorting of a germplasm collection reveals the *hydroxylase3* locus as a new target for maize provitamin A biofortification. Plant Physiology 151:1635-1645

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends in Biotechnology 27:522-530

Venkocsky R, Ramalho MAP, Toledo FHR (2012) Contribution and perspectives of quantitative genetics to plant breeding in Brazil. Crop Breeding and Applied Biotechnology S2:7-14

Viana JMS (2000) The parametric restrictions of the Griffing diallel analysis model: combining ability analysis. Genetics and Molecular Biology 23:877-881

Viana JMS, Cruz CD, Cardoso AA (1999) Theory and analysis of partial diallel crosses. Genetics and Molecular Biology 22:591-599

Viana JMS, Matta FP (2003) Analysis of general and specific combining abilities of popcorn populations, including selfed parents. Genetics and Molecular Biology 26:465-471

Welsch R, Arango J, Bar C, Salazar B, Al-Babili S, Beltran J, Chavarriaga P, Ceballos H, Tohme J, Beyer P (2010) Provitamin A accumulation in cassava (*Manihot esculenta*) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. Plant Cell 22:3348-3356

Were WV, Shanahan P, Melis R, Omari OO (2012) Gene action controlling farmer preferred traits in cassava varieties adapted to mid-altitude tropical climatic conditions of western Kenya. Field Crops Research 113:113-118

Xia L, Peng K, Yang S, Wenzl P, Carmen MV, Fregene M, Kilian A (2005) DArT for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. Theoretical and Applied Genetics 110:1092-1098

Xu S, Zhu D, Zhang Q (2014) Predicting hybrid performance in rice using genomic best linear unbiased prediction. Proceedings of the National Academy of Science, USA 111:12456-12461

Yadav JS, Ogwok E, Wagaba H, Patil BL, Bagewadi B, Alicai T, Gaitan-Solis E, Taylor NJ, Fauquet CM (2011) RNAi-mediated resistance to cassava brown streak Uganda virus in transgenic cassava. Molecular Plant Pathology 12:677-687

Yan H, Lu L, Alzate A, Ceballos H, Hershey CH, Chen S, Li K (2014) Fruit, seed and embryo development of different cassava (*Manihot esculenta* Crantz) genotypes and embryo rescue. African Journal of Biotechnology 13:1524-1528

Yan W, Kang MS (2002) GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists and Agronomists. CRC Press, New York

Yan W, Kang MS, Ma B, Woods S, Cornelius PL (2007) GGE biplot vs. AMMI analysis of genotype by environment data. Crop Science 47:643

Yan W, Tinker N (2006) Biplot analysis of multi-environment trial data: principles and applications. Canadian Journal of Plant Science 86:623-645

Yau SK (1995) Regression and AMMI analyses of genotype x environment interactions: an empirical comparison. Agronomy Journal 87:121-126

Zacarias AM, Labuschagne MT (2010) Diallel analysis of cassava brown streak disease, yield and yield related characteristics in Mozambique. Euphytica 176:309-320

Zhang Z, Zhang Q, Ding X (2011) Advances in genomic selection in domestic animals. Chinese Science Bulletin 56:2655-2663

Zhu C, Naqvi S, Gomez-Galera S, Pelacho AM, Capell T, Christou P (2007) Transgenic strategies for the nutritional enhancement of plants. Trends in Plant Science 12:548-555

CHAPTER 3

Diallel analysis of provitamin A carotenoid and dry matter content in cassava

3.1 Introduction

Cassava, the fifth most important staple crop in the world, is a widely grown and consumed root crop in SSA (Tan 2015). The crop is central to food and income security throughout SSA, especially among resource-poor farmers (Fermont et al. 2010). This popularity of cassava is attributed to its ability to produce reasonable yields under poor soil conditions and extended droughts, where other food crops would practically fail (El-Sharkawy 2007). In Uganda, the area harvested under cassava has steadily increased from 405 000 ha in 2003 to 440 000 ha in 2013 (FAOSTAT 2014), despite the known enormous threat from CBSD and CMD during this period. This trend in cultivation of cassava is a reflection of the growing relevance of the crop in people's livelihoods.

However, compared to efforts in other important food crops, genetic improvement of cassava has lagged behind, keeping the crop's production below its potential (Ceballos et al. 2015). For example, average fresh root yield of cassava reported for Uganda is 12 t ha^{-1}, which is lower than that for Ghana (18.3 t ha^{-1}), Thailand (21.8 t ha^{-1}) and the Caribbean (22.1 t ha^{-1}) (FAOSTAT 2014). Formal cassava breeding in Africa began in the 1930s at Amani research station, Tanzania, when scientists Storey and Nichols guided the first systematic efforts to breed for CMD and CBSD resistance (Storey and Nichols 1938; Nichols 1947). Subsequent cassava breeding activities only intensified after the colonial rule when an organised programme was instituted at IITA, Ibadan, in the 1970s (Umanah 1977). Although efforts are now reshaping breeding strategies to exploit the yield potential of cassava, they lag considerably behind what has been achieved in, for example, maize since the 1900s when a breakthrough was recorded in use of inbred lines to exploit heterosis for increasing grain yield (Shull 1908).

Most agronomic traits including yield, DMC and disease resistance are quantitatively inherited in cassava, and information about their mode of inheritance helps breeders to use methods that increase genetic gain (Calle et al. 2005; Kulembeka et al. 2012). During early stages of cassava breeding, open pollination schemes were the most predominant mating designs used to generate breeding populations upon which selection could be imposed (Kawano 2003). These designs involved establishing selected parental lines of complementary traits *per se* in crossing nurseries to generate

half-sib progeny. Although these mating schemes are simple to implement, they are limited in that they do not provide for estimation of SCA, which is important in inheritance of key traits like fresh root yield (Ceballos et al. 2004; Crossa et al. 2010; Ceballos et al. 2015). Subsequent shifts in cassava breeding schemes have seen an increased production of full-sib progeny (Nassar and Ortiz 2006; Ceballos et al. 2012). The full-sib crossing schemes employ controlled pollinations, where selected mating designs are used to generate full-sib families from specific parental combinations, facilitating genetic studies alongside production of breeding populations (Nduwumuremyi et al. 2013).

The diallel mating design, specifically, has become popular for cassava breeding because it allows the identification of parents with superior combining ability for developing breeding populations, while facilitating generation of useful information on genetics of key agronomic traits (Zacarias and Labuschagne 2010; Kulembeka et al. 2012; Tumuhimbise et al. 2014). It is this genetic information that guides breeders to deploy appropriate methods for crop improvement (Acquaah 2012; Nduwumuremyi et al. 2013). Knowledge of GCA of parental lines is particularly helpful for predicting genetic gains in a breeding programme (Falconer and Mackay 1996). Hayman (1954) and Griffing (1956) elaborated on the procedure for statistical analyses based on diallel data, which partitions total variation into GCA of the parents and SCA of crosses.

Recently, the national cassava breeding programme of Uganda initiated a breeding pipeline with an objective of developing high-yielding provitamin A cassava varieties (Esuma et al. 2012). Deployment of such carotene-rich cassava would sustainably improve nutrition and reduce prevalence of VAD in communities that primarily depend on cassava (Mayer et al. 2008; Nassar and Ortiz 2010). Ideally, provitamin A cassava varieties are also expected to have high DMC. Varieties that combine these two traits (provitamin A and DMC) are more likely to be adopted for subsistence agriculture (Abele et al. 2007; Njukwe et al. 2013). The strong negative correlation that has so far been reported between DMC and carotenoid content in the African cassava germplasm (Akinwale et al. 2010; Njoku et al. 2015) could present a potential challenge for cassava breeding programmes tasked to improve both traits. Contrastingly, other studies analysing Latin American cassava germplasm indicated that correlations between carotenoid content and DMC are not strong enough to reach statistical significance (Chávez et al. 2005; Sánchez et al. 2014). It is against this background that this study was undertaken. Specifically, the objectives of this study were to (1) generate F_1 populations segregating for both DMC and total carotenoid content (TCC), (2) estimate GCA of six cassava parental lines and their SCA through their progeny for DMC and

TCC and (3) determine the types of gene action controlling DMC and TCC in cassava.

3.2 Materials and methods

3.2.1 Experimental sites

A crossing block for the diallel study was established in November 2011 at NaCRRI in Namulonge, Uganda. The seedling trial for the diallel population was established at Abi Zonal Agricultural Research Development Institute (Abi-ZARDI), which is a satellite research station for NARO located in north western Uganda. Abi-ZARDI was chosen for the seedling trial because of low CBSD and CMD pressure. CBSD causes cassava roots to become necrotic (Hillocks and Jennings 2003; Nuwamanya et al. 2015), thus rendering them unsuitable for carotenoid quantification. CMD, in situations of high severity, can reduce a plant's ability to produce a reasonable number of stakes. Therefore, conducting the seedling trial under low disease pressure was a critical consideration to ensure planting materials were disease-free. Upon generation of sufficient planting materials (6-12 stakes per genotype) from the seedling trial clonal trials for the selected study materials were conducted at Namulonge and Abi-ZARDI.

Abi-ZARDI is located at 31°1'28.4"E and 2°36'33.3"N at 1 060 m above sea level, with predominantly sandy-loam soils. Namulonge is located at 32°37'36.0"E and 0°31'13.7"N at 1 164 m above sea level, with sandy-clay-loam soils. Both locations experience a bimodal rainfall pattern, with two distinct rainy and dry seasons of nearly equal length. The first season of rainfall peaks between March and mid-June while the second season rains peak from August to November.

3.2.2 Parental selection and hybridisation

Six genetically diverse clones in advanced selection stages were used as parental lines. Three of these genotypes were clones enriched with provitamin A carotenoids (pVAC) introduced from IITA, two were pVAC clones introduced from CIAT and one was a local white-flesh officially released variety popularly grown by farmers in Uganda (Table 3.1). The six parents were planted in the crossing block at Namulonge under rain-fed conditions in paired rows to facilitate generation of the 15 F_1 families of a 6 x 6 half-diallel design. Planting was done at a spacing of 1.5 m between pairs of parents for crossing and 2 m alleys between subsequent parental pairs, which collectively provided the additional space to ease movement during the pollination process. Five plants represented each parental line and the planting was done once.

Table 3.1 List of progenitors used in the 6 x 6 half-diallel study

Genotype	Code	Source	RFC[a]	Salient traits
NASE 3	P1	IITA[b]	White	High DMC[c]; CMD[d] and CBSD[e] tolerance
CPCR24B-10	P2	CIAT[f]	Light yellow	pVAC[g], CMD resistance
MH05-2870	P3	IITA	Yellow	pVAC, CMD resistance
MH05-0233	P4	IITA	Yellow	pVAC, CMD resistance
CPCR15B-26	P5	CIAT	Yellow	pVAC, CMD resistance
MH02-073HS	P6	IITA	Deep yellow	pVAC, CMD resistance

[a]Root flesh colour; [b]International Institute for Tropical Agriculture; [c]Dry matter content of roots; [d]Cassava mosaic disease [e]Cassava brown streak disease; [f]International Centre for Tropical Agriculture; [g]Provitamin A carotenoid.

Controlled pollinations were performed by hand following the standard procedures described by Kawano (1980). Briefly, female flowers targeted for pollination were covered with nylon-meshed pollination bags 2-3 days before they reached anthesis (i.e. before opening). To speed up the pollination process, mature male flowers were harvested in a clearly labelled Falcon tube of capacity 50 ml and carried between designated female parents. The harvested male flowers were used within a period of less than six hours. Only mature and fully opened female flowers of an inflorescence were pollinated and the rest of the un-pollinated flowers plucked off. Newly pollinated flowers were re-covered immediately with the pollination bag for 3-4 days. Covering mature flowers before and after pollination was a precaution to avoid contamination with pollen carried by wind and insects that forage on cassava flowers (Alves 2002; Halsey et al. 2008).

Only mature fruits were picked within 2.5-3 months after pollination. Fruits were allowed to dry in seed bags placed on branches of female plants. Botanical seeds were extracted from these bags and stored in labelled paper bags for two months to break seed dormancy. At least 100 seeds from each family were germinated in a screen house in plastic pots filled with natural forest soil. After germination, seedlings were watered routinely whenever deemed necessary to ensure vigorous growth.

3.2.3 Seedling trial design

The seedling trial was planted in June 2013. Forty-five vigorous seedlings were randomly selected from each of the 15 cross combinations. Selections from each family were

planted within a plot; every family was represented by three rows, each containing 15 seedlings. Planting was done at a spacing of 1 x 1 m providing a density of 10 000 plants ha^{-1}. To overcome environmental shock commonly experienced by newly transplanted cassava seedlings in the first few weeks, plants were watered by hand every two days during the first three weeks to ensure good establishment. Weeding was done as necessary. The seedling trial was harvested 12 months after planting (MAP) and selections made for clonal evaluation. No data were collected from the seedling trial, as its purpose was to generate planting material for replicated clonal trials.

3.2.4 Clonal trial

The purpose of the clonal trial was to assess performance of the test material at two locations with contrasting environmental conditions, which is important for genetic analyses as confounding effects of environments on expression of phenotypes are minimised (Falconer and Mackay 1996). Before seedlings were uprooted, 20 genotypes with the ability to generate ≥12 standard-size cuttings (i.e. 4-6 nodes) were randomly selected from each F_1 family and tagged to constitute the population for clonal evaluation. At each site, trials were laid out in an incomplete block design with two replications. A single row plot of three plants represented each entry, with 2 m alleys between blocks. Genetic analysis trials with similar plot sizes have been previously used for cassava (Ojulong et al. 2008). Planting was done at a spacing of 1 x 1 m, giving a total population of 10 000 plants ha^{-1}. Weeding was done as necessary. These clonal trials were planted in May 2014 and harvested in April 2015.

3.2.5 Data collection

F_1 clonal trials were phenotyped at 12 MAP, the average age for physiological maturity of cassava when the most important traits are optimally expressed (Alves 2002). Data of clonal evaluation was preferred in this study for two reasons: (1) quantification of carotenoids and DMC requires considerable amounts of root tissue, which may not be obtained in sufficient quantities from seedlings (Rodriguez-Amaya and Kimura 2004; Chávez et al. 2008) and (2) DMC in seedlings is less stable than in clones from the same seedlings (Ojulong et al. 2010). During harvesting, all plants in a plot from each replication were uprooted and the biomass bulked to estimate yield components by separately weighing roots (kg $plant^{-1}$) and foliage (kg $plant^{-1}$) using a Salter Brecknell suspended weighing scale [model: 23510S(SHFSB-0404)] calibrated in kilograms. HI was computed from the measure of fresh root weight (FRW) and fresh shoot weight (FSW), using the formula:

$$HI = \frac{FRW}{(FRW+FSW)}$$

Three roots (with the ability to provide ≥ 250 g parenchyma tissue) were randomly selected from each plot, labelled and processed for measurement of DMC and TCC. These roots were peeled, washed under running water and dried with a paper towel. Dried roots were cut longitudinally into quarters. The opposite quarters of each of the three roots were pooled, chopped into small pieces and homogenised. Homogenous samples (200 g) were used for measurement of DMC by drying the samples in an oven (SMO14-2 SHEL LAB FORCED AIR OVEN, USA) to constant weights at a temperature of 105°C for 24 hours. Dried samples were reweighed to obtain their DMC as:

$$DMC(\%) = \frac{DSW}{FSW} \times 100$$

where
DSW = dry sample weight
FSW = fresh sample weight.

Approximately 90% of the TCC in cassava is known to be ß-carotene, which is the most active form of provitamin A carotenoids (Rodriguez-Amaya and Kimura 2004; Nassar et al. 2007). Therefore, a measure of TCC using the iCheck analytical kit supplied by BioAnalyt Laboratory, Berlin, Germany (http://www.bioanalyt.com) was deemed sufficient to reflect the amount of ß-carotene in genotypes (Ceballos and Parkes 2014). Briefly, 5 g of the homogenous root sample was pounded and ground into a smooth and fine paste using a mortar and pestle. To aid grinding of the sample, 20 ml of distilled water was added gradually and the resultant solution transferred into a 50 ml calibrated tube. The tube content was shaken thoroughly and 0.4 ml of the solution injected into the iEx™ CAROTENE vial using the syringe and needle provided with the kit**.** Vials were placed on a solid surface for approximately 5 min, shaken again and allowed to stand until two solution phases appeared inside the vial: a clear upper phase and a turbid lower phase. At this point, the absorbance of the vial content (the upper solution phase) was measured using the iCheck™ CAROTENE device in the iCheck kit. TCC was calculated as:

$$TCC\ (\mu g\ g^{-1}) = \frac{V_s}{W_s} \times A$$

where:

V_s = volume of solution transferred to the tube
W_s = weight of sample (in this case the 5 g)
A = absorbance of the iEx™ CAROTENE vial content at a wavelength of 450 nm.

Each sample was extracted and measured for TCC once. All procedures for carotenoid quantification were performed in a dark room. All harvested root samples were analysed within 12 hours for both TCC and DMC. In addition to quantitative measurement of carotenoid content, root flesh colour (RFC) was scored for all genotypes using visual inspection following the standard colour scale developed by CIAT (Appendix 1). Based on the chart, colour of the root parenchyma can vary on a scale of 1-8, where 1 = white and 8 = pink. The proximal portion of the root was cut transversally with a sharp knife and the observed colour score assigned. This part of the root accumulates more carotenoids (Ceballos et al. 2011) and thus reflects true potential of the root for TCC. The colour chart was obtained from Dr. Egesi Chiedozie, cassava breeder and Assistant Director at the National Root Crops Research Institute, Nigeria. Figure 3.1 is a pictorial summary of major events during implementation of research activities for the diallel study.

3.2.6 Data analysis

Analysis of variance (ANOVA) was done using the Plant Breeding Tools software (PBTools) (PBTools 2014). Diallel analysis was conducted according to the Griffing (1956) method 2, model I for fixed effects, to estimate the GCA and SCA effects. Briefly, method 2 is a diallel, where parents and one set of F_1s, but not reciprocals, are included in the statistical analyses. Model I considers parents as fixed effects and interpretations of the genetic effects are limited to the specific set of parents used. Thus, GCA and SCA effects were estimated as:

$$Y_{ijk} = \mu + g_i + g_j + s_{ij} + e_{ijk}$$

where:
Y_{ijk} = observed value for the ij^{th} cross in the k^{th} replication/environment combination
μ = overall mean
g_i = GCA effect for the i^{th} parent
g_j = GCA effect for the j^{th} parent
s_{ij} = SCA of the cross between the i^{th} and j^{th} parents
e_{ijk} = error term associated with the ij^{th} cross in the k^{th} replication/environment.

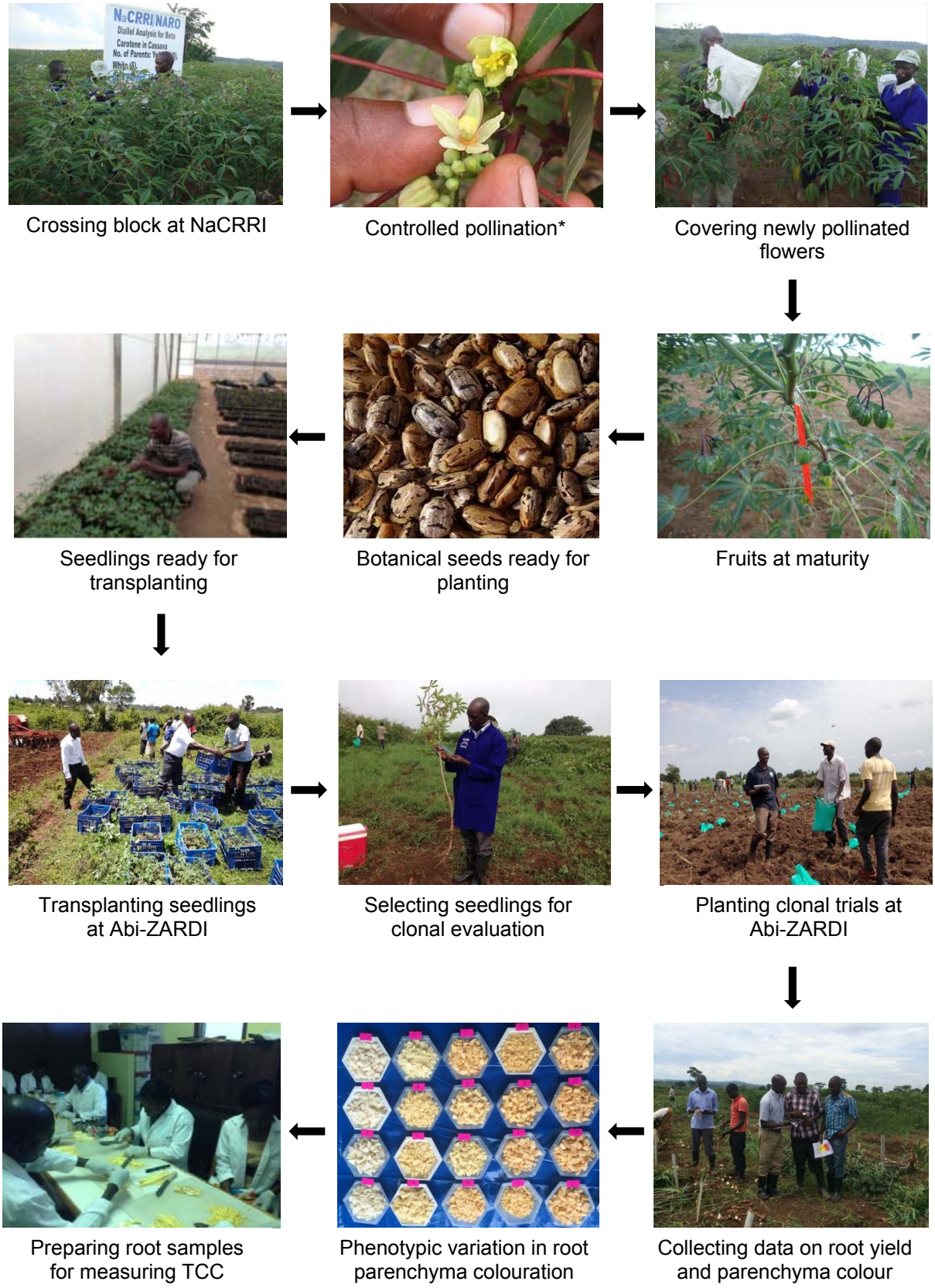


Figure 3.1 **Photographic summary of major activities undertaken during the diallel study**. NaCRRI = National Crops Resources Research Institute; Abi-ZARDI = Abi Zonal Agricultural Research and Development Institute; *Performed by dusting physiologically mature pollen onto stigma; TCC = total carotenoid content; Arrows indicate the sequence of the activities from 2011-2015.

PBTools provides statistical output specifying estimates of the genetic variance components from which heritability estimates and dominance ratios were derived for each of the traits phenotyped. Estimates of GCA (σ^2_{GCA}) variance and SCA (σ^2_{SCA}) variance for each trait were derived from the mean squares (MS) from the ANOVA output and used to estimate σ^2_A and σ^2_D variance components as $\sigma^2_A = 2(\sigma^2_{GCA})$ and $\sigma^2_D = \sigma^2_{SCA}$. These genetic parameters were used to estimate σ^2_G and σ^2_P as $\sigma^2_G = \sigma^2_A + \sigma^2_D$ and $\sigma^2_P = \sigma^2_G + \sigma^2_\varepsilon$, where σ^2_ε is the residual variance. Narrow (h^2) and broad (H^2) sense heritability were calculated from estimates of the variance components according to Falconer and Mackay (1996), so that $h^2 = \sigma^2_A/\sigma^2_P$ and $H^2 = \sigma^2_G/\sigma^2_P$. The relative importance of GCA and SCA effects for each trait was determined from the proportions of a family's sum of squares (SS) due to GCA and SCA. Further statistical inference was made using Baker's ratio (BR) calculated from the genetic parameters as:

$$BR = \frac{2MS_{GCA}}{2MS_{GCA} + MS_{SCA}}$$

where:
MS_{GCA} = variance of mean squares due to GCA
MS_{SCA} = variance of mean squares due to SCA effects.

The R statistical programme (R Development Core Team 2010) was used to calculate Pearson's correlation coefficients for phenotypes from the combined data across environments and replications using the following method:

$$r_p = \frac{cov_{xy}}{\sigma_x \sigma_y}$$

where:
r_p = phenotypic correlation between traits *x* and *y*
cov_{xy} = phenotypic covariance between traits *x* and *y*
σ_x = standard deviation of *x*
σ_y = standard deviation of *y*.

Genetic correlation coefficients among traits were calculated from the GCA effects as described by Hohls and Clarke (1995). Selection was imposed on progeny from the 15 F_1 families to identify superior clones for further breeding. The base selection index of Brim-Williams (Williams 1962; Brim et al. 1995) was used for selection due to its

efficiency, simplicity of calculation, ease of result interpretation and freedom from errors of parameter estimation. BLUPs for each trait were computed and economic weights assigned to TCC (3.5), DMC (2.5) and FRW (2.0), so that the selection index (SI) was calculated as:

$$SI = (B_{TCC} \times 3.5 \times h^2_{TCC}) + (B_{DMC} \times 2.5 \times h^2_{DMC}) + (B_{FRW} \times 2.0 \times h^2_{FRW})$$

where B and h^2 are the BLUP and heritability of the traits of traits, respectively.

3.3 Results

3.3.1 Environmental conditions at experimental locations

Weather data on rainfall and temperature were collected at the two locations during the period of the clonal trials. The amount of rainfall received during this period varied between the experimental locations (Table 3.2), but it was sufficient for supporting the growth and development of cassava (Hauser et al. 2014). Soil nutrient content was analysed prior to planting the clonal trials at the two sites and the nutrient levels were within an ideal range for cassava production (Cadavid 2012), except for phosphorus.

Table 3.2 Description of weather and soil conditions at experimental sites for clonal evaluation of the population for the diallel study

Parameter	Unit of measure	Critical value	Abi-ZARDI[g]	Namulonge
pH		4.0 - 8.0*	6.1	6.0
OM[a]	%	3.0	6.6	7.8
N	%	0.2	0.3	0.4
P	ppm	10.0*	3.0	2.4
Ca	ppm	50.0*	1 141.6	1 248.9
Mg	ppm	14.3*	336.2	734.2
K	ppm	58.5*	336.2	443.2
Rainfall	mm		884	1264
T (min)[b]	°C		15.9	17.4
T (max)[c]	°C		31.4	29.8
AEZ[d]			NWSG[e]	LVC[f]

[a]Organic matter; [b]Minimum temperature; [c]Maximum temperature; [d]Agroecological zone; [e]Northwestern Savannah Grasslands; [f]Lake Victoria Crescent; [g]Abi Zonal Agricultural Research and Development Institute; Values in asterisks are critical for cassava (Cadavid 2012).

3.3.2 Mean performance of the 15 F_1 families and their parents

Table 3.3 presents data on the performance of 276 genotypes distributed across the 15 F_1 families. Families with < 20 genotypes were due to failure of establishment of some genotypes in the clonal trials. For all genotypes evaluated across the 15 F_1 families, TCC values varied from 0 to 11.0 µg g^{-1}, with the highest mean (5.8 µg g^{-1}) recorded for family P5×P6 and the lowest (2.1 µg g^{-1}) for family P1 x P2 (Table 3). Individual DMC values for the evaluated genotypes ranged from 15.9 to 45.9%.

At family level, DMC ranged from 22.2% for family P5×P6 to 34.4% for P1×P2. Family P5×P6 recorded the highest mean FRW (2.4 kg $plant^{-1}$), while family P2 x P3 recorded the lowest mean FRW (1.2 kg $plant^{-1}$). The highest mean value of FSW (5.5 kg $plant^{-1}$) was recorded for family P1 x P5 and the lowest value (2.4 kg $plant^{-1}$) for P2 x P4. Meanwhile, family P2 x P3 had the lowest mean value (0.28) for HI and family P5 x P6 had the highest mean value (0.74) for HI.

At progenitor level, genotype MH02-073HS recorded the highest level of both TCC (10.4 µg g^{-1}) and FRW (6.3 kg $plant^{-1}$), but had the lowest DMC (22.2%). NASE 3 recorded the highest DMC (37.3), but had very low TCC (0.2 µg g^{-1}).

3.3.3 Analysis of variance

ANOVA summaries for the analysed traits are presented in Table 3.4. Based on mean squares (MS), differences in performance of families were very significant ($P \leq 0.01$) for all the studied traits. Mean performance varied of crosses significantly across environments for DMC, FRW and FSW, but such variations were non-significant for TCC and HI. Meanwhile, GCA effects were highly significant ($P \leq 0.001$) for TCC, very significant ($P \leq 0.01$) for DMC, significant ($P \leq 0.05$) for HI, but non-significant for FRW and FSW. SCA effects were highly significant FRW and FSW, very significant for DMC and HI and non-significant for TCC. The coefficient of variation (CV) associated with traits ranged from 7.5% for TCC to 21.2% for HI; these are within the acceptable range, and thus give confidence in the generated datasets.

TCC had the highest value for Baker's ratio (0.89) followed by DMC (0.69), while FRW had the lowest value (0.36). A Baker's ratio above 0.5 implies that additive genetic effects are proportionately more important than dominance effects (Baker 1978).

Table 3.3 Performance of parents and their respective F_1 progeny across two locations in Uganda during 2014-2015

Parent/Family	Number[a]	TCC[b]	DMC[c]	FRW[d]	FSW[e]	HI[f]
NASE 3 (P1)	-	0.2	37.3	2.5	2.6	0.51
CPCR24B-10 (P2)	-	4.4	32.3	2.0	2.9	0.44
MH05-2870 (P3)	-	4.3	33.4	1.5	5.7	0.32
MH05-0233 (P4)	-	4.9	29.0	2.4	1.7	0.58
CPCR15B-26 (P5)	-	5.3	30.7	2.3	2.1	0.49
MH02-073HS (P6)	-	10.4	22.2	6.3	2.5	0.72
P1 x P2	17	2.1	34.4	1.5	4.1	0.33
P1 x P3	18	3.1	31.0	1.8	4.0	0.42
P1 x P4	19	3.1	31.3	1.9	3.7	0.45
P1 x P5	20	2.4	33.9	1.9	5.5	0.34
P1 x P6	16	2.6	28.1	1.9	5.3	0.56
P2 x P3	15	2.7	32.4	1.2	3.8	0.28
P2 x P4	15	3.1	32.1	1.5	2.4	0.37
P2 x P5	20	3.2	32.1	1.8	3.4	0.43
P2 x P6	17	3.7	29.6	1.5	4.1	0.54
P3 x P4	20	2.3	33.0	2.1	3.7	0.43
P3 x P5	16	3.2	32.9	2.0	2.9	0.38
P3 x P6	20	3.7	29.7	2.1	4.1	0.57
P4 x P5	20	4.3	33.6	2.0	2.8	0.52
P4 x P6	20	4.4	29.0	2.0	3.8	0.41
P5 x P6	19	5.8	22.2	2.4	3.4	0.74
Minimum[g]		0.0	15.9	0.0	0.1	0.00
Maximum[g]		11.0	45.9	13.5	52.0	0.85
Mean		3.8	30.9	2.1	3.5	0.44
SE[h]		0.062	0.150	0.150	0.322	0.025
$LSD_{0.05}$[i]		0.127	0.304	0.303	0.650	0.050
Number of F_1[j]		261	255	272	272	272

[a]Number of genotypes evaluated per F_1 family; [b]Total carotenoid content ($\mu g\ g^{-1}$); [c]Dry matter content of roots (%); [d]Fresh root weight (kg $plant^{-1}$); [e]Fresh shoot weight (kg $plant^{-1}$); [f]Harvest index; [g]Values based on all F_1 genotypes evaluated; [h]Standard error; [i]Least significant difference at 5% confidence level; [j]Number of F_1 genotypes evaluated: reduction from the total population (272) indicates proportion of genotypes whose roots were not sufficient for measuring TCC and/or DMC.

Table 3.4 Mean squares of crosses and combining ability effects of five traits evaluated at two locations in 15 F_1 families and parents

Source of variation	DF[a]	TCC[b]	DMC[c]	FRW[d]	FSW[e]	HI[f]
Environment (E)	1	0.03	9.31**	4.63***	31.29***	0.13**
Crosses (C)	14	4.23***	16.87**	1.75**	2.73**	0.07**
C x E	14	0.05	3.09***	0.26*	1.12*	0.01
GCA[g]	5	42.26***	16.66**	2.86	3.85	8.49*
SCA[h]	14	10.81	15.24**	10.30***	12.66***	16.21*
GCA x E	5	0.06	15.71***	0.73*	9.08**	0.05
SCA x E	14	1.01	6.47*	8.81*	15.95**	0.48*
% SS[i] due to GCA		71.31	64.34	21.29	12.84	51.27
% SS due to SCA		24.48	31.28	56.84	48.11	28.51
Residual	34	0.02	0.36	0.36	1.66	0.01
CV[j] (%)		7.50	16.31	11.21	19.42	21.20
Baker's ratio		0.89	0.69	0.37	0.38	0.56

[a]Degrees of freedom; [b]Total carotenoid content ($\mu g\ g^{-1}$); [c]Dry matter content of roots (%); [d]Fresh root weight (kg $plant^{-1}$); [e]Fresh shoot weight (kg $plant^{-1}$); [f]Harvest index; [g]General combining ability; [h]Specific combining ability; [i]Sum of squares; [j]Coefficient of variation; * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$.

When SS of families were partitioned into SS due to parents (GCA effects) and interaction between parents (SCA effects), the GCA effects accounted for more than 64% of the total variation expressed by the families for TCC and DMC, while FRW, FSW and HI each contributed less than 22% of variability in GCA effects. The SCA effects were important for FRW and FSW.

3.3.4 General combining ability of progenitors

Traits evaluated in this study were measured with preference for high scores (i.e. higher positive values for combining ability estimates were preferred for each of the evaluated traits). Genotype MH02-073HS, with the highest TCC, showed the highest GCA effect of 1.93 for the trait (Table 3.5). On the other hand, NASE 3, a white-fleshed progenitor with negligible TCC, had significant negative GCA effect of -1.59 for TCC. Negative GCA for TCC indicates unsuitability of specific progenitors as combiners when targeting high carotenoid content in the progeny. However, NASE 3 was the best general combiner for DMC with positive and significant GCA of 1.75, while progenitor MH02-073HS had negative and significant GCA of -3.72 for DMC. Only progenitor MH02-073HS had positive and significant GCA for FRW. Meanwhile, MH02-073HS was the best general combiner for HI, with a GCA effect of 0.16.

Table 3.5 General combining ability effects of cassava parental lines used in a 6x6 half-diallel analysis for five traits

Parents	TCC[a]	DMC[b]	FRW[c]	FSW[d]	HI[e]
NASE 3	-1.59***	1.75***	-0.11	0.38*	-0.12
CPCR24B-10	-0.38*	0.87**	-0.42**	-0.14*	-0.04
MH05-2870	-0.27	0.82**	-0.33*	0.63**	-0.05
MH05-0233	0.03	-0.17*	-0.06	-0.63**	0.03
CPCR15B-26	0.28*	0.44**	-0.03	-0.34**	0.02
MH02-073HS	1.93***	-3.72***	0.95**	0.12	0.16*
$LSD_{0.05}$[f]	0.253	0.607	0.605	1.300	0.020
SE[g]	0.041	0.097	0.097	0.208	0.014

[a]Total carotenoid content (μg g^{-1}); [b]Dry matter content of roots (%); [c]Fresh root weight (kg $plant^{-1}$); [d]Fresh shoot weight (kg $plant^{-1}$); [e]Harvest index; [f]Least significant difference at 5% confidence level; [g]Standard error; * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$.

3.3.5 Specific combining ability of crosses

Three families (P1 x P5, P3 x P4, and P4 x P6) showed positive significant SCA effects for TCC (Table 3.6). The other 12 families had either negative or non-significant positive SCA effects. Overall, there was high variation in SCA effects for DMC, with five families showing positive significant SCA effects. For this trait, family P1 x P2 had the most positive and significant SCA effect of 1.60. For FRW, there were four F_1 families with positive significant SCA effects, with P1 x P5 as the best specific combination. Meanwhile, none of the 15 F_1 families showed positive significant SCA effects for FSW. On the other hand, three families (P2 x P5, P3 x P5 and P4 x P6) had positive and significant SCA effect for HI.

3.3.6 Genetic parameters

The genetic analyses revealed that σ^2_A was higher than σ^2_D for TCC (Table 3.7). Furthermore, TCC had a moderate narrow sense h^2 of 0.48. Similarly, a larger proportion of σ^2_A than σ^2_D was obtained for DMC, with h^2 of 0.41. HI had slightly larger σ^2_A than σ^2_D, with a moderate h^2 of 0.34. Both FRW and FSW showed larger proportions of σ^2_D than σ^2_A, with very low h^2 (< 0.17). TCC, DMC, FRW and HI showed high H^2 (> 0.6) while FSW had the lowest H^2 (0.15).

Table 3.6 Specific combining ability effects for a 6x6 half diallel analysis of five traits evaluated at two locations in Uganda

Family	TCC[a]	DMC[b]	FRW[c]	FSW[d]	HI[e]
P1 x P2	0.77	1.60*	-0.43	-0.49*	-0.04
P1 x P3	-0.62	-1.17*	-0.87*	0.03	0.06
P1 x P4	-0.98	-1.09	-0.23	0.12	-0.09
P1 x P5	1.65*	1.22*	1.90**	0.16	-0.10
P1 x P6	-0.82	-0.56	-0.37	0.17	0.16
P2 x P3	-0.04	-0.43	-0.33	0.02	-0.25*
P2 x P4	0.46	-0.04	0.52	0.06	-0.08
P2 x P5	-0.18	-1.34*	0.83*	0.20	0.32*
P2 x P6	-1.01*	0.21	-0.58	0.22	0.05
P3 x P4	1.00*	0.54	0.71*	-0.04	0.12
P3 x P5	-0.19	0.16	0.00	0.03	0.24*
P3 x P6	-0.15	0.90	0.49	-0.04	-0.16
P4 x P5	-1.87*	0.55	-2.09**	-0.08	-0.18
P4 x P6	1.39*	0.04	1.09*	-0.05	0.22*
P5 x P6	0.59	-0.59	-0.64	-0.31*	-0.28*
SE[f]	0.77	1.60*	-0.43	-0.49*	-0.04

[a]Total carotenoid content (μg g^{-1}); [b]Dry matter content (%); [c]Fresh root weight (kg $plant^{-1}$); [d]Fresh shoot weight (kg $plant^{-1}$); [e]Harvest index; [f]Standard error; P1, P2, P3, P4, P5 and P6 are parental genotypes defined in Table 3.1; * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$.

Table 3.7 Genetic parameter estimates for five traits of 6x6 half diallel F_1 families evaluated at two locations in Uganda

Component	TCC[a]	DMC[b]	FRW[c]	FSW[d]	HI[e]
σ^2_A[f]	6.47	12.01	0.57	0.01	1.06
σ^2_D[g]	4.46	11.94	2.40	1.00	0.93
σ^2_A x E[h]	0.01	0.87	0.07	1.56	0.02
σ^2_D x E	0.00	3.71	0.20	2.40	0.06
h^2[i]	0.48	0.41	0.16	0.08	0.34
H^2[j]	0.94	0.82	0.63	0.15	0.77

[a]Total carotenoid content (μg g^{-1}); [b]Dry matter content of roots (%); [c]Fresh root weight (kg $plant^{-1}$); [d]Fresh shoot weight (kg $plant^{-1}$); [e]Harvest index; [f]Additive genetic variance; [g]Dominance genetic variance; [h]Environmental effect; [i]Narrow sense heritability; [j]Broad sense heritability.

3.3.7 Phenotypic and genetic correlation among traits

The most positive and highly significant correlation was between root flesh colour (RFC) and TCC, with a coefficient of 0.94 (Table 3.8). This relationship indicated that higher intensities of root pigmentation reflected higher levels of TCC in roots. There was negative and significant correlation (r = -0.44) between the RFC and DMC. Similarly, a negative correlation was noted between TCC and DMC (r = -0.45). These relationships indicate that cassava roots with higher levels of TCC have low DMC. HI correlated positively and significantly with FRW but had significantly negative correlation with FSW, which is expected, given that HI is derived from the two traits.

The most positive and highly significant genetic correlation was between RFC and TCC (r = 0.87). DMC had significant negative genetic correlation with TCC (r = -0.82). These genetic correlations (RFC with TCC and DMC with TCC) were consistent with patterns of phenotypic correlations between the same trait pairs.

These correlations are of practical importance in cassava breeding and selection for TCC and DMC. The scatter plot in Figure 3.2 reveals genotypes combining high levels of both TCC and DMC, which could form genetic resources for advancement along the cassava breeding pipeline.

Table 3.8 Phenotypic (lower diagonal) and genetic (upper diagonal) correlation coefficients for six traits in 6x6 half-diallel families evaluated at two locations in Uganda

Trait	RFC[a]	FRW[b]	DMC[c]	FSW[d]	TCC[e]	HI[f]
RFC		0.04	-0.62**	0.31	0.87***	0.02
FRW	0.14		-0.23	-0.03	0.83**	0.81**
DMC	-0.44**	0.07		0.08	-0.82**	-0.08
FSW	-0.02	0.07	-0.01		-0.05	-0.13
TCC	0.94***	0.132	-0.45**	-0.02		0.86**
HI	0.09	0.57**	0.09	-0.47**	0.08	

[a]Root flesh colour; [b]Fresh root weight; [c]Dry matter content of roots; [d]Fresh shoot weight; [e]Total carotenoid content; [f]Harvest index; ** $P \leq 0.01$; *** $P \leq 0.001$.

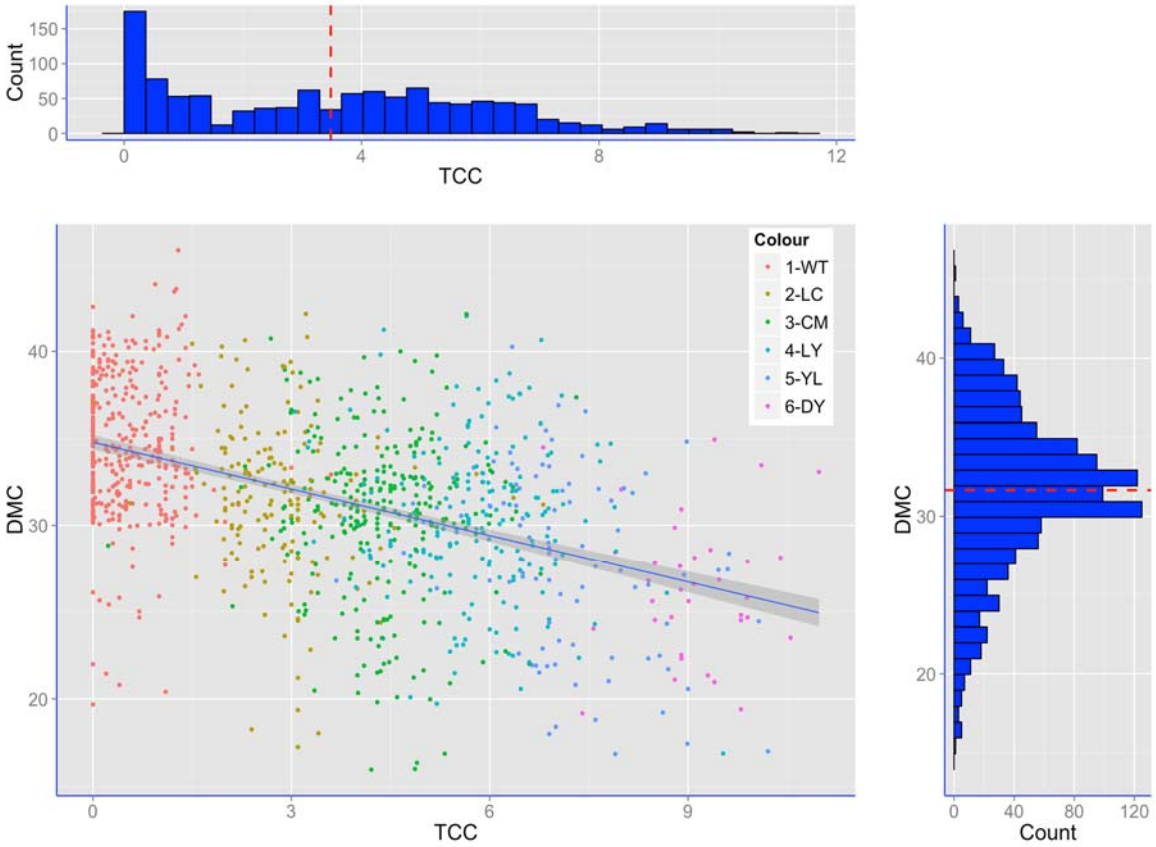


Figure 3.2 Scatter plot of dry matter content (%) and total carotenoid content (μg g^{-1}) scaled by root flesh colour. Colour 1-WT = white; 2-LC = light cream; 3-CM = cream; 4-LY = light yellow; 5-YL = yellow and 6-DY = deep yellow. Red dashed lines in the side histograms indicate means of the respective traits.

3.3.8 Selection of breeding material for advancement

Based on the SI, 61 genotypes (Appendix 2) were selected across the 15 F_1 families. Family P4 x P5 (MH05-0233 x CPCR15B-26) contributed the highest number of individual clones (12 genotypes) to the selected individuals while no genotype was selected for advancement from families P1 x P5 (NASE 3 x CPCR15B-26) and P1 x P6 (NASE 3 x MH02-073HS) (Table 3.9). Selected genotypes had higher means for TCC (5.83 μg g^{-1}), DMC (31.9 kg plant^{-1}) and FRW (2.7 kg plant^{-1}) than means for all genotypes studied.

Family P5 x P6 (CPCR15B-26 x MH02-073HS) had the highest mean values for TCC (7.05 μg g^{-1}) and FRW (2.9 kg plant^{-1}) for all genotypes selected. Highest mean DMC (36.3%) was recorded for family P1 x P3 (NASE 3 x MH05-2870) (Table 3.9).

Table 3.9 Number and means for total carotenoid content, dry matter content and fresh root weight of genotypes selected from the 6x6 half diallel breeding population for advancement

Family	Number[a]	TCC[b] mean	DMC[c] mean	FRW[d] mean
P1 x P2	1	5.46	28.8	2.3
P1 x P3	1	5.40	36.3	1.5
P1 x P4	2	5.72	30.0	3.0
P1 x P5	0	-	-	-
P1 x P6	0	-	-	-
P2 x P3	1	4.92	32.8	2.5
P2 x P4	4	6.41	31.7	2.0
P2 x P5	6	5.32	31.2	3.0
P2 x P6	2	6.35	29.3	2.7
P3 x P4	9	5.70	35.0	3.1
P3 x P5	3	4.98	31.2	3.2
P3 x P6	7	5.90	28.6	2.4
P4 x P5	12	5.33	34.4	2.8
P4 x P6	3	5.29	29.5	2.7
P5 x P6	10	7.05	29.4	2.9
Mean of individuals selected		5.83	31.9	2.7
Mean of all individuals		3.80	31.0	2.1
SE[e]		0.31	0.6	1.9
CV[f] (%)		15.2	21.4	24.8

[a]Number of genotypes selected; [b]Total carotenoid content (μg g^{-1}); [c]Dry matter content of roots (%); [d]Fresh root weight (kg $plant^{-1}$); [e]Standard error of the mean for selected individuals; [f]Coefficient of variation; P1, P2, P3, P4, P5 and P6 are parental genotypes defined in Table 3.1.

3.4 Discussion

3.4.1 Phenotypic variability and correlations among traits evaluated

TCC in the F_1 progeny evaluated for this study varied from 0.0-11.0 μg g^{-1}, with a mean of 3.8 μg g^{-1}. Values for TCC in the progenitors varied from 0.2-10.4 μg g^{-1}. The observed TCC mean for the F_1 progeny was comparable to 3.6 μg g^{-1} and 5.0 μg g^{-1} reported, respectively, by Maroya *et al.* (2012) and Ssemakula and Dixon (2007) for breeding populations evaluated at IITA, but quite lower than the mean (14.7 μg g^{-1}) reported for populations at CIAT (Ceballos et al. 2013). The high level of TCC in cassava breeding populations at CIAT is a result of 10-year cyclic selection process imposed to

primarily advance carotenoid-rich clones, targeted to attain levels above 15 μg g^{-1}. The population evaluated in the current study arose from a single round of recombination and fell short of this nutrient level targeted by HarvestPlus program (Pfeiffer and McClafferty 2007).

It is worth noting that this study was strategically designed to generate pVAC clones expressing both CMD and CBSD resistance by including the white-fleshed NASE 3 as a progenitor. Nonetheless, the wide level of segregation for TCC observed in the current breeding population could provide a basis for implementing a recurrent selection scheme for developing cassava varieties with increased provitamin A carotenoid content in future. The rapid cycling scheme used by CIAT for increasing carotenoids content in the roots (Ceballos et al. 2013) could be adopted for this endeavour. With this scheme maximum carotenoids levels was increased from 10 to 26 μg g^{-1} within a period of nine years.

DMC varied from 15.9-45.9%, with a mean of 30.9%. The mean DMC in the pVAC populations developed in this study is somewhat below the DMC levels (mean = 35%) of most improved white-fleshed varieties currently grown by farmers in Uganda (Kawuki et al. 2011). Importantly, some of the carotenoid-rich genotypes evaluated here had higher DMC than that of varieties commonly grown by farmers, for example, 81 genotypes had DMC values ≥ 35%. These genotypes are of interest to breeding, as they qualify to be used as progenitors for the next round of recombination. High DMC in cassava roots is a trait highly preferred by farmers in SSA (Tumuhimbise et al. 2012, Njukwe et al. 2013). To enhance adoption of pVAC varieties in future, breeders will need to focus deliberate efforts on increasing DMC in the carotene-rich breeding populations. Fresh root yield is another critical trait that influences adoption of new varieties by farmers; however, in the current study, this trait was measured in the form of FRW and HI. FRW potential was assessed on plant basis and it varied from 0 to 13.5 kg $plant^{-1}$. Genotypes with the ability to yield ≥ 3 kg $plant^{-1}$ would be preferred for cultivation, as current varieties popularly grown by farmers in Uganda have average yield of 25 t ha^{-1} (2.5 kg $plant^{-1}$) (Kawuki et al. 2011). However, such genotypes identified from the current study will require further verification using larger plot sizes.

Two phenotypic correlations in this study are of special importance for developing pVAC cassava varieties. Firstly, the strong positive correlation between root flesh colour and TCC is a good incentive for screening large early-generation breeding populations by visual assessment. At such stage of breeding, genotypes with higher intensity of root

pigmentation can be selected for advancement, which saves time and the high costs associated with quantification of carotenoids. Iglesias et al. (1997) and Chávez et al. (2000) have previously demonstrated the effectiveness of using root colour to select for high carotenoid content in cassava, especially if increasing carotenoids content is the sole breeding objective. However, Sánchez et al. (2014) suggest that the intensity of pigmentation, though useful for identifying high pVAC clones, may result in lower DMC. Perhaps a simplistic explanation is that high DMC has lots of white starch which tends to dilute the intensity of pigmentation. This is a relevant issue as it can explain why carotenoids and DMC even have a positive correlation in Latin American cassava breeding populations (Sánchez et al. 2014) compared with those in Africa (Njoku et al. 2015). Thus, it would be important to make a mild initial selection based on colour intensity and then a stronger one based on quantified carotenoids levels.

Continuous selection for both DMC and carotenoid content in Latin America has been underway much longer than in Africa and weak initial negative correlations could have been broken during the several cycles of recombination. For example, Ceballos et al. (2013) reported an interesting result of increased pVAC levels in breeding populations upon which several cycles of recurrent selection were imposed for high levels of carotenoid content, yet DMC increased along with this selection strategy. In that study, the authors suggested that continuous recombination in a recurrent selection scheme could generate clones that combine high levels of both DMC and TCC, with values of up to 37% and 25 $\mu g\ g^{-1}$, respectively. Compared to other crops, Grisales et al. (2014) reported a strong negative correlation between carotene content and fruit DMC in butternut squash (*Cucurbita moschata* D.), which is similar to reports by Vimala et al. (2011) for studies on sweet potato (*Ipomea batatas* L.). As further studies are undertaken, more useful genetic information relevant for cassava breeding will be generated and hopefully, this discrepancy clarified. For now, it is important to continue to pursue this approach of combining both traits.

3.4.2 Combining ability estimates of evaluated traits

In genetic analysis, GCA and SCA are important parameters that illustrate the importance of additive and non-additive genetic effects in inheritance of economically important traits. GCA is the average performance of all progeny from a specific parent, expressed as a deviation from the overall mean of crosses. SCA is the deviation from the expected value from a cross, which is the sum of the GCA of two parental lines involved

in a cross. Thus, GCA is the main genetic effect, while SCA is an interaction (Falconer and Mackay 1996).

Interpretation of the genetic nature of traits evaluated in this study was based on the mean squares for GCA and SCA effects. The proportion of SS for crosses explained by GCA components gives an estimate of the relative importance of additive effects in expression of traits (Falconer and Mackay 1996). GCA accounted for a significantly larger SS than SCA for TCC, DMC and HI, explaining more than 64% of the total variation for DMC and TCC. These results suggest that additive gene effects are more important in controlling accumulation of TCC and DMC in cassava. Similarly, 51.3% of the variation in HI was attributed to GCA. The relative importance of additive genetic factors for TCC, DMC and HI was reflected by the higher Baker's ratios for these traits. Baker's ratio above 0.5, as was the case with these three traits, means these traits are under the control of additive genetic factors. The implication of these findings for cassava breeding is that a recurrent mass selection method can be an efficient breeding method for improving TCC and DMC, as it would enhance the exploitation of additive genetic effects for the traits (Ceballos et al. 2013).

FRW had non-significant GCA effects, suggesting that this trait is largely under control of non-additive genetic factors. This deduction is consistent with previous reports by Calle et al. (2005), Zacarias and Labuschagne (2010) and Kulembeka et al. (2012). In this case, breeders targeting the increase of fresh root yield in cassava would consider crossing progenitor combinations with superior SCA for FRW, a strategy that would increase the chances of selecting high-yielding clones from the segregating progeny (Mal 2013). However, the relatively high importance of GCA effects for HI suggests a possibility of increasing FRW by crossing progenitors with high HI. Although Ojulong et al. (2010) demonstrated the effectiveness of using progenitors with high HI for improving FRW in cassava, the practice does not necessarily result in gains for fresh root yield. In reality, it is common to find a genotype with low plant vigour (low foliage weight) showing a high HI when its actual root yield is very low (Hay 2008). Perhaps it would be a better practice to use HI as a complementary trait to FRW when selecting for high yield in cassava, which agrees well with suggestions by Hay (2008) that selections based on HI alone can be less effective.

Generally, progenitors with higher levels of TCC showed positive GCA, suggesting their contribution towards enhancing TCC in the progeny. Progenitor MH02-073HS had the highest positive and significant GCA effect for TCC, FRW and HI. Therefore, an

appropriate breeding design would be to cross MH02-073HS to a genetic background of high DMC to increase chances of generating clones that accumulate favourable alleles for expressing higher levels of both traits (Ceballos et al. 2013). Similar relative importance of GCA in inheritance of carotenoid content has been reported in *Cucumis sativus* L. (Navazio and Simon 2001) and *Zea mays* L. (Senete et al. 2011), suggesting the trait is controlled by genes that act additively. The highest negative GCA effects for TCC shown by NASE 3 is unsurprising, given it is a white-fleshed genotype with negligible levels of TCC. Conversely, the same genotype had the highest positive GCA effect for DMC, which also correlates well with its high level of DMC. NASE 3, also referred to as TMS 30572, is an IITA bred variety that is characterized with high levels of DMC (mean of > 35%). This clone was officially released in Uganda in the 1990s and has remained popular among farmers, largely because of the high DMC and tolerance to both CMD and CBSD.

The non-significant SCA effects for TCC indicate that SCA is less important than GCA for inheritance of TCC in cassava, which is similar to deductions made by Senete et al. (2011) for inheritance of carotenoids in maize. Meanwhile, five F_1 families with positive and significant SCA effects for DMC could present possibility for generating clones with high DMC when such genotypic combinations are used for hybridization. Overall, SCA effects were more important than GCA effects for inheritance of fresh root yields, which agrees with previous by Kulembeka et al. (2012) and Tumuhimbise et al. (2014). Genotypic differences of the progenitors used in these studies may account for this disparity in genetic control of FRW in cassava.

Although populations evaluated in this study have gone through one recombination cycle, there were individual genotypes selected for combined superior levels of TCC, DMC and FRW. Some of the selected genotypes had DMC and FRW comparable to some cultivars currently grown by farmers in Uganda and could be further evaluated on-farm with the view to identifying candidate varieties for the future. However, additional recombination cycles are required to increase TCC to levels sufficient enough to impact positively on nutrition of targeted beneficiaries. On-going international collaborations for agricultural research could offer platforms for introducing improved provitamin A cassava genotypes from CIAT and IITA to increase carotenoid content of the current populations in Uganda through further hybridisation.

3.5 Conclusion

This study reports on a pioneer effort to breed cassava for high pVAC content, targeting to benefit the entire eastern Africa. Both breeding products in the form of genotypes and genetic information have been generated, all of which are useful for follow-up studies and future breeding programmes. The generated datasets suggested that GCA effects are more important than SCA effects in the genetic control of carotenoid content in cassava, indicating that additive genetic effects largely control inheritance of carotenoids in cassava. The additive nature of carotenoids content provides scope for its improvement under the recurrent selection scheme. The negative correlation between root carotenoid content and DMC is important situation to deal with as a matter of priority. In particular, future breeding efforts that can uncover the genetic basis of the negative correlation between DMC and TCC are warranted. A breakthrough in this case could offer more realistic prospects for developing pVAC varieties acceptable to farmers. Such a breakthrough can be attained by at least two different approaches (that can hopefully be pursued together): 1) introducing germplasm from Latin America for combining high TCC and DMC and 2) strengthening the selection of segregating progenies based on the quantification of carotenoids as opposed to visual pigment selection.

3.6 References

Abele S, Twine E, Ntawuruhunga P, Baguma Y, Kanobe C, Bua A (2007) Development and dissemination of improved cassava varieties in Uganda: analysis of adoption rates, variety attributes and speed of adoption. In: AAAE Ghana Conference, Accra, pp 479-482

Acquaah G (2012) Principles of Plant Genetics and Breeding. Second edition. John Willey and Sons Ltd, UK

Akinwale MG, Aladesanmwa RD, Akinyele BO, Dixon AGO, Odiyi AC (2010) Inheritance of ß-carotene in cassava (*Manihot esculenta* Crantz). International Journal of Genetics and Molecular Biology 2:198-201

Alves AAC (2002) Cassava botany and physiology. In: Hillocks RJ, Thresh JM, Bellotti AC (eds) Cassava: Biology, Production and Utilization, vol 1. CAB International, UK, pp 67-89

Brim CA, Johnson HW, Cockerham CC (1995) Multiple selection criteria in soybeans. Agronomy Journal 51:42-46

Cach NT, Pérez JC, Lenis JI, Calle F, Morante N, Ceballos H (2005) Epistasis in the expression of relevant traits in cassava (*Manihot esculenta* Crantz) for subhumid conditions. Journal of Heredity 96:586-592

Cadavid LF (2012) Soils and fertilizers for the cassava crop. In: Ospina B, Ceballos H (eds) Cassava in the Third Millennium: Modern Production, Processing, Use, and Marketing Systems, vol 377. CIAT, Cali, Colombia, pp 113-137

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

Ceballos H, Hershey CH, Becerra-López-Lavalle LA (2012) New approaches to cassava breeding. Plant Breeding Reviews 36:427-504

Ceballos H, Iglesias CA, Perez J, Dixon AGO (2004) Cassava breeding: opportunities and challenges. Plant Molecular Biology 56:503-516

Ceballos H, Kawuki RS, Gracen VE, Yencho GC, Hershey CH (2015) Conventional breeding, marker-assisted selection, genomic selection and inbreeding in clonally propagated crops: a case study for cassava. Theoretical and Applied Genetics 128:1647-1667

Ceballos H, Luna J, Escobar AF, Ortiz D, Pérez JC, Sánchez T, Pachón H, Dufour D (2011) Spatial distribution of dry matter in yellow fleshed cassava roots and its influence on carotenoid retention upon boiling. Food Research International 45:52-59

Ceballos H, Morante N, Sánchez T, Ortiz D, Aragón I, Chávez AL, Pizarro M, Calle F, Dufour D (2013) Rapid cycling recurrent selection for increased carotenoids content in cassava roots. Crop Science 53:23-42

Ceballos H, Parkes EY (2014) Measuring provitamin A content in crops. The 2nd Global Conference on Biofortification: Getting Nutritious Foods to People. Kigali, Rwanda

Chávez AL, Bedoya JM, Sánchez T, Iglesias CA, Ceballos H, Roca W (2000) Iron, carotene, and ascorbic acid in cassava roots and leaves. Food and Nutrition Bulletin 21:410-414

Chávez AL, Ceballos H, Rodriguez-Amaya BD, Pérez JC, Sánchez T, Calle F, Morante N (2008) Sampling variation for carotenoids and dry matter contents in cassava roots. Journal of Root Crops 34:43-49

Chávez AL, Sánchez T, Jaramillo G, Bedoya JMI, Echeverry J, Bolaños EA (2005) Variation of quality traits in cassava roots evaluated in landraces and improved clones. Euphytica 143:125-133

Crossa J, Campos L, Pérez P, Gianola D, Burgueno J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan J, Arief V, Banziger M, Braun HJ (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713-724

El-Sharkawy MA (2007) Physiological characteristics of cassava tolerance to prolonged drought in the tropics: implications for breeding cultivars adapted to seasonally dry and semiarid environments. Brazialian Journal of Plant Physiology 19:257-286

Esuma W, Rubaihayo P, Pariyo A, Kawuki R, Wanjala B, Nzuki I, Harvey JJW, Baguma Y (2012) Genetic diversity of provitamin A cassava in Uganda. Journal of Plant Studies 1:60-71

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics. Fourth edition. Longman, Essex, UK

FAOSTAT (2014) Food and Agriculture Organization of the United Nations Statistics. http://data.fao.org/ref/262b79ca-279c-4517-93de-ee3b7c7cb553. Accessed 4 September 2015

Fermont AM, Babirye A, Obiero HM, Abele S, Giller KE (2010) False beliefs on the socio-economic drivers of cassava cropping. Agronomy for Sustainable Development 30:433-444

Griffing B (1956) Concept of general and specific combining ability in relation to diallel crossing systems. Australian Journal of Experimental Biology & Medical Science 9:463-493

Grisales O, Sanín VR, Piedad M, Cabrera V, Alirio F, García B (2014) Genetic correlations and path analysis in butternut squash *Cucurbita moschata* Duch. Revista Facultad Nacional de Agronomía 68:7399-7409

Halsey ME, Olsen KM, Taylor NJ, Chavarriaga-Aguirre P (2008) Reproductive biology of cassava (*Manihot esculenta* Crantz) and isolation of experimental field trials. Crop Science 48:49-58

Hauser S, Wairegi L, Asadu CLA, Asawalam DO, Jokthan G, Ugbe U (2014) Cassava System Cropping Guide: Africa Soil Health Consortium. CAB International, Nairobi

Hay RKM (2008) Harvest index: a review of its use in plant breeding and crop physiology. Annals of Applied Biology 126:197-216

Hayman BI (1954) The theory and analysis of diallel crosses. Genetics 39:789-809

Hillocks RJ, Jennings DL (2003) Cassava brown streak disease: a review of present knowledge and research needs. International Journal of Pest Management 49:225-234

Hohls T, Clarke GPY (1995) Additive and non-additive genetic correlations: properties of sampling distribution Properties of sampling distribution. Biometrical Journal 37:919-935.

Iglesias CA, Meyer J, Chávez AL, Calle F (1997) Genetic potential and stability of carotene content in cassava roots. Euphytica 94:367-373

Jaramillo G, Morahan G, Calle F, Ceballos H, Arias B, Bellotti AC (2005) Diallel analysis in cassava adapted to the mid-altitude valleys environment. Crop Science 45:1058-1063

Kawano K (1980) Cassava. In: Fehr WR, Hadley HH (eds) Hybridisation of Crop Plants. America Society of Agronomy and Crop Science Society of America, WI, USA, pp 225-233

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

Kawuki RS, Pariyo A, Amuge T, Nuwamanya E, Ssemakula G, Tumwesigye S, Bua A, Baguma Y, Omongo C, Alicai T, Orone J (2011) A breeding scheme for local adoption of cassava (*Manihot esculenta* Crantz). Journal of Plant Breeding and Crop Science 3:120-130

Kulembeka HP, Ferguson M, Herselman L, Kanju E, Mkamilo G, Masumba E, Fregene M, Labuschagne MT (2012) Diallel analysis of field resistance to brown streak disease in cassava (*Manihot esculenta* Crantz) landraces from Tanzania. Euphytica 187:277-288

Mal M (2013) Combining ability and mode of gene action in cassava for resistance to cassava green mite and cassava mealy bug in Malawi. Journal of Plant Breeding and Crop Science 5:195-202

Maroya NG, Kulakow P, Dixon AG, Maziya-Dixon B, Baker KS (2012) Genotype x environment interaction of carotene content of yellow-fleshed cassava genotypes in Nigeria. Journal of Life Sciences 6:595-601

Mayer JE, Pfeiffer WH, Beyer P (2008) Biofortified crops to alleviate micronutrient malnutrition. Current Opinion in Plant Biology 11:166-170

Nassar NMA, Ortiz R (2006) Cassava improvement: challenges and impacts. Journal of Agricultural Science 145:163-171

Nassar NMA, Ortiz R (2010) Breeding cassava to feed the poor. Scientific American Journal 302:78-82

Nassar NMA, Schwartz CA, Junior OP (2007) Cassava diversity in Brazil: the case of carotenoid-rich landraces. Genetics and Molecular Research 6:116-121

Navazio JP, Simon PW (2001) Diallel analysis of high carotenoid content in cucumbers. American Society of Horticultural Science 126:100-104

Nduwumuremyi A, Tongoona P, Habimana S (2013) Mating designs: helpful tool for quantitative plant breeding analysis. Journal of Plant Breeding and Genetics 1:117-129

Nichols RFW (1947) Breeding cassava for virus resistance. East African Agricultural Journal 15:154-160

Njoku DN, Gracen VE, Offei SK, Asante IK, Egesi CN, Kulakow P, Ceballos H (2015) Parent-offspring regression analysis for total carotenoids and some agronomic traits in cassava. Euphytica 206:657-666

Njukwe E, Hanna R, Kirscht H, Araki S (2013) Farmers perception and criteria for cassava variety preference in Cameroon. African Study Monographs 34:221-234

Nuwamanya E, Baguma Y, Atwijukire E, Acheng S, Alicai T (2015) Effect of cassava brown streak disease (CBSD) on cassava (*Manihot esculenta* Crantz) root storage components, starch, starch quantities and starch quality properties. International Journal of Plant Physiology and Biochemistry 7:12-22

Ojulong H, Labuschangne MT, Fregene M, Herselman L (2008) A cassava clonal evaluation trial based on a new cassava breeding scheme. Euphytica 160:119-129

Ojulong HF, Labuschagne MT, Herselman L, Fregene M (2010) Yield traits as selection indices in seedling populations of cassava. Crop Breeding and Applied Biotechnology 10:191-196

PBTools (2014) Biometrics and Breeding Informatics, Version 1.4. PBGB Division. International Rice Research Institute, Los Banos, Laguna

Pfeiffer WH, McClafferty B (2007) HarvestPlus: breeding crops for better nutrition. Crop Science 47:88-105

R Development Core Team (2010) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Australia

Rodriguez-Amaya BD, Kimura M (2004) HarvestPlus Handbook for Carotenoid Analysis, vol 1. International Food Policy Research Institute (IFPRI) and International Center for Tropical Agriculture (CIAT), Washington, DC and Cali

Sánchez T, Ceballos H, Dufour D, Ortiz D, Morante N, Calle F, Zum Felde T, Dominguez M, Davrieux F (2014) Prediction of carotenoids, cyanide and dry matter contents in fresh cassava root using NIRS and hunter color techniques. Food Chemistry 151:444-451

Senete CT, Guimarães PEO, Paes MCD, Souza JC (2011) Diallel analysis of maize inbred lines for carotenoids and grain yield. Euphytica 182:395-404

Shull GH (1908) The composition of a field of maize. Journal of Heredity 4:296-301

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

Storey HH, Nichols RFW (1938) Studies of the mosaic disease of cassava. Annals of Applied Biology 25:790-806

Tan SL (2015) Cassava - silently, the tuber fills: the lowly cassava, regarded as a poor man's crop, may help save the world from the curse of plastic pollution. Utar Agriculture Science Journal 1:12-24

Tumuhimbise R, Melis R, Shanahan P (2014) Diallel analysis of early storage root yield and disease resistance traits in cassava (*Manihot esculenta* Crantz). Field Crops Research 167:86-93

Tumuhimbise R, Melis R, Shanahan P, Kawuki R (2012) Farmers' perceptions on early storage root bulking in cassava (*Manihot esculenta* Crantz) in east and central Uganda and their implication for cassava breeding. World Journal of Agricultural Sciences 8:403-408

Umanah EE (1977) Cassava research in Nigeria before 1972. In: Cock J, Machntyre R, Graham M (eds) Proceedings of the Fourth Symposium of the International Society for Tropical Root Crops, Ottawa, IDRC, pp 137-141

Viana JMS, Cruz CD, Cardoso AA (1999) Theory and analysis of partial diallel crosses. Genetics and Molecular Biology 22:591-599

Vimala B, Sreekanth A, Binu H, Wolfgang G (2011) Variability in 42 orange-fleshed sweet potato hybrids for tuber yield and carotene and dry matter content. Geneconserve 40:190-200

Williams JS (1962) The evaluation of a selection index. Biometrics 18:375-393

Zacarias AM, Labuschagne MT (2010) Diallel analysis of cassava brown streak disease, yield and yield related characteristics in Mozambique. Euphytica 176:309-320

CHAPTER 4

Genotype by environment interaction of carotenoid and dry matter content in cassava in Uganda

4.1 Introduction

Cassava is the second most widely produced and consumed crop in Uganda. The crop is grown by most smallholder farmers throughout the country due to its ability to yield better than other staple food crops under conditions of extended drought and poor soils (El-Sharkawy 2007, Ceballos et al. 2011). Despite this important feature, it is common for cassava cultivars to display high sensitivity to differences in environmental conditions (Akinwale et al. 2011). The phenomenon of differential genotypic responses under varying environments is referred to as GEI. Abiotic and biotic stresses influence expression of genes that control key agronomic traits, which gives rise to GEI (Kang 2002). For example, during a typical 12-month growing period, cassava can experience overlapping and/or contrasting environmental stresses, thus exacerbating the extent of GEI. As such, GEI remains of interest in most plant breeding programmes. Subsequently, systematic evaluation of GEI effects for a given trait is useful for understanding varietal stability and hence strategic deployment of varieties (Acquaah 2012).

It is for these reasons that several univariate and multivariate statistical models have been developed for stability analyses and/or understanding GEI (Eberhart and Russell 1966, Gauch et al. 2008, Gauch 2013). Over the past years, excellent reviews highlighting weaknesses, strengths and best practices of these stability and/or GEI models, have been undertaken (Crossa 1990, Piepho 1994, Ye et al. 2013). It suffices to note that a number of studies on cassava have opted for AMMI for assessment GEI effects on, among other traits, carotenoid content (Maroya et al. 2012), early bulking of storage roots (Agyeman et al. 2015) and resistance to CBSD (Pariyo et al. 2015). AMMI allows exhaustive data analysis by performing regular ANOVA and estimating interaction effects through principal component analysis (PCA), which somewhat increases precision in trait estimates and enables reliable selections (Gauch et al. 2008, Hongyu et al. 2014).

A complementary analytical tool to visualise GEI is the genotype plus genotype by environment (GGE) biplot (Yan and Tinker 2006). The polygon view of a GGE biplot is the best way to assess the interaction patterns between genotypes and environments

and to effectively interpret a biplot (Yan and Kang 2002). Genotypes that occupy vertices of the polygon are the best performers for a given trait in a specific environment. The GGE biplot allows identification of stable and best performing genotypes in test environments, which is an important decision-making tool for identifying crop varieties for subsequent release (Rao et al. 2011, Farshadfar et al. 2013).

It is commonplace for cassava breeders to evaluate advanced breeding lines (as many as 30) in several environments (as many as 10) to account for GEI when identifying genotypes with high and stable performance (Akinwale et al. 2011, Maroya et al. 2012). Studies by Akinwale et al. (2011), Tumuhimbise et al. (2014) and Agyeman et al. (2015) have indicated considerable variation in fresh root yield across varying environmental conditions. Ssemakula and Dixon (2007) noted low influence of GEI on carotenoid content in cassava roots at harvest, based on analysis of 28 genotypes in five environments evaluated over two growing cycles. A much later study on performance of 18 provitamin A clones across five environments in Nigeria indicated significant interaction between genotypes and test environments for carotenoid content (Maroya et al. 2012). Advancement of improved cassava clones for on-farm production would require subjecting such clones to systematic evaluation under diverse environments to identify better adapted genotypes (Fukuda et al. 2002, Nassar and Ortiz 2006).

Recently, the national cassava breeding programme in Uganda initiated a breeding objective tailored towards developing provitamin A cassava that expresses high levels of other farmer preference traits, especially DMC (Esuma et al. 2012). It is envisioned that this initiative will culminate into deployment of provitamin A cassava varieties for purposes of improving nutrition among populations vulnerable to vitamin A deficiency. This food-based intervention to alleviate micronutrient deficiency could have sustainable impact in developing countries, including Uganda, where food fortification and supplementation have been less impacting due to poor social infrastructures and high poverty levels (Mayer et al. 2008, Boy et al. 2009, Thompson and Amoroso 2011). Therefore, this study was undertaken to (1) assess the GEI for DMC and carotenoid content, (2) evaluate the effect of crop age on DMC and carotenoid content in cassava roots and (3) identify stable genotypes for high carotenoid and DMC levels.

4.2 Materials and methods

4.2.1 Genotypes

Thirteen genotypes of diverse genetic background were evaluated in this study (Table

4.1). Ten of the genotypes were selected from sets of germplasm previously acquired from CIAT and IITA. The CIAT and IITA materials were clones at advanced stages of selection for fresh root yield and pVAC. The other three genotypes were yellow-flesh landraces obtained from farmer fields in Uganda.

Table 4.1 Provitamin A cassava genotypes used to study genotype by environment interaction for carotenoid and dry matter content

Genotype	Code	Status	Source
91-01730	G1	Improved	IITA[a]
ANDIFEKU	G2	Landrace	Uganda
CPCR15B-26	G3	Improved	CIAT[b]
MAYAYA	G4	Landrace	Uganda
MH02-073HS	G5	Improved	IITA
MH04-2757	G6	Improved	IITA
MH05-0452	G7	Improved	IITA
MH07-0529	G8	Improved	IITA
MM01-0014	G9	Improved	IITA
MM01-1003	G10	Improved	IITA
MM06-0466	G11	Improved	IITA
MM06-2862	G12	Improved	IITA
BUSIA	G13	Landrace	Uganda

[a]International Institute of Tropical Agriculture; [b]International Centre for Tropical Agriculture

4.2.2 Experimental sites

Trials were conducted over two growing seasons between May 2012 and December 2014, with each cropping season lasting 15 months. Trials for the 2012/2013 season were planted in May 2012 while 2013/2014 trials were planted in September 2013. Experiments were conducted at three sites, each located at the Bulindi Zonal Agricultural Research and Development Institute (Bu-ZARDI), Abi-ZARDI and Namulonge (at NaCRRI). Bu-ZARDI is located in the western savannah grasslands, while Abi-ZARDI and Namulonge (at NaCRRI) are as described section 3.2.1. Each cropping season was considered an environment, giving a total of six environments (Table 4.2). Weather instruments available at these research stations were used for recording temperature and rainfall data during experimentation. Figure 4.1 shows the agroecological zones where the experimental sites were located, which are major cassava regions of Uganda.

Table 4.2 Geographical characteristics of environments for the genotype by environment interaction study on accumulation of carotenoids and dry matter content in cassava

Location	Latitude	Longitude	Altitude	Cropping season	E[a]
Abi[b]	2°36'33.3"N	31°1'28.4"E	1 060 m	May 2012-Aug 2013	E1
Bulindi[c]	1°27'58.9"N	31°26'39.1"E	1 157 m	May 2012-Aug 2013	E2
Namulonge[d]	0°31'13.7"N	32°37'36.0"E	1 164 m	May 2012-Aug 2013	E3
Abi				Sep 2013-Dec 2014	E4
Bulindi				Sep 2013-Dec 2014	E5
Namulonge				Sep 2013-Dec 2014	E6

[a]Environments in which GEI trials were conducted; [b]Abi Zonal Agricultural Research and Development Institute; [c]Bulindi Zonal Agricultural Research and Development Institute; [d]National Crops Resources Research Institute (NaCRRI).

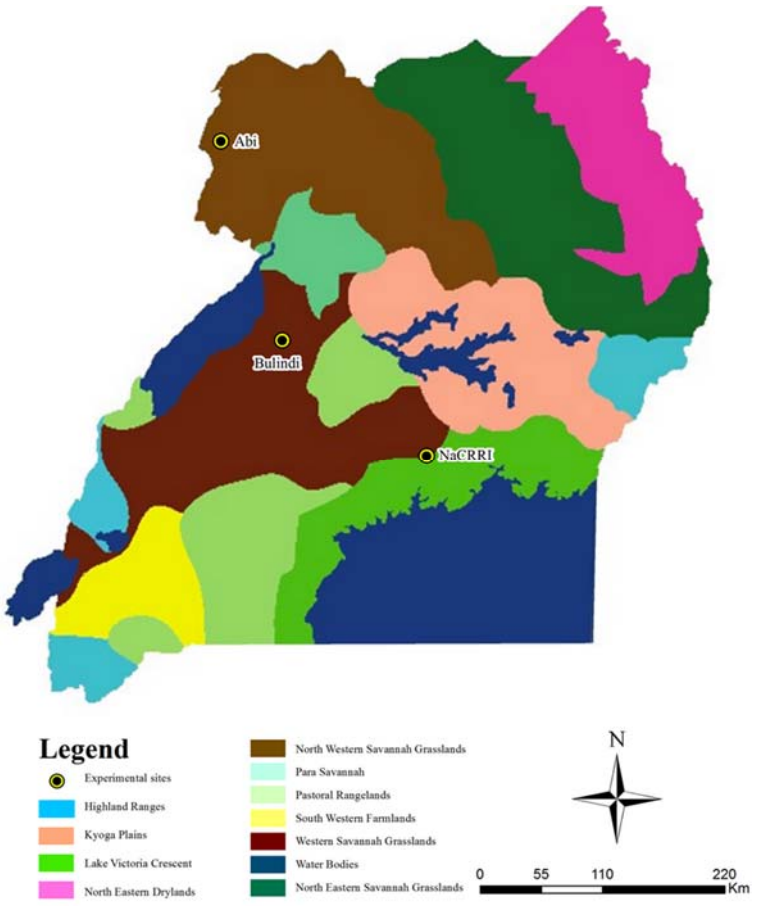


Figure 4.1 Experimental sites for the genotype by environment interaction study on total carotenoid content and dry matter content in cassava in Uganda.

NaCRRI = Namulonge experimental site. Legend shows agroecological zones of Uganda.

4.2.3 Experimental design

Each trial was laid out in a randomised complete block design with three replications. In every replication, a plot consisted of a genotype planted as seven rows of seven plants each, giving a plot size of 49 plants. Planting was done at a spacing of 1 × 1 m, giving a density of 10 000 plants ha^{-1}. To increase chances of sprouting and uniform plant establishment, all stakes used for planting were generated from middle portions of mature stems. Adjacent plots were separated by 2 m alleys in order to avoid inter-genotypic competitions for space. Weeding was done as necessary. Experiments were entirely rain fed and no fertilizers were applied to plants in the field.

4.2.4 Data collection

Traits measured in this study were TCC, DMC, FRW and HI. These traits were measured at 6, 9, 12 and 15 MAP. Inner five rows of each experimental plot constituted a net plot of 25 plants so that six random plants were uprooted during every sampling time to take measurement on the four traits. DMC, FRW, FSW and HI were estimated as described in section 3.2.5. The iCheck analytical kit for carotenoid analysis became available after this study had commenced. Thus, for consistence in datasets, TCC was measured using the ultraviolet (UV)/visible spectrophotometry method described by Rodriguez-Amaya and Kimura (2004). A summary of this method is presented below.

Three roots, each randomly picked from three of the six harvested plants, were used for preparing homogenous samples (as described in section 3.2.5) for measurement of TCC. Approximately 10 g of the homogenous root sample was weighed and transferred into a mortar. To aid grinding, 3 g Hyflosupercel (celite) was added to the sample and the mixture ground in 50 ml cold acetone, using a pestle. The resultant solution was filtered into a conical flask through glass wool in a thistle funnel. This procedure was repeated 2-3 times until the residue was free of any colour. The extract was transferred into a 500 ml separating funnel with a Teflon stopcock, containing about 40 ml of petroleum ether. To remove acetone from the extract, double distilled water was added gently along the slanting surface of the funnel so that emulsion formation was avoided. The aqueous phase was discarded and the procedure repeated 3-4 times to get rid of acetone residues. The petroleum ether phase was transferred to a 50 ml volumetric flask through a funnel containing 15 g of anhydrous sodium sulphate to remove the residual water and the extract made up to 50 ml with petroleum ether. Absorbance of this extract was measured at 450 nm using a spectrophotometer (Specord 210, Analytikjena model Torre Boldone BG, Italy) and TCC calculated as:

$$TCC\ (\mu g\ g^{-1}) = \frac{A \times V \times 10^4}{2\ 592 \times W}$$

where:

A = absorbance at 450 nm

V = total extract volume (ml)

W = sample weight (g)

2 592 = ß-carotene absorption coefficient in petroleum ether.

All procedures for carotenoid extraction and measurement were performed in a dark room as described in section 3.2.5. Roots harvested from the Namulonge trials, located on the NaCRRI premises, were analysed for TCC and DMC within 12 h of harvesting. Roots sampled for analysis from trials at Abi-ZARDI and Bu-ZARDI were prepared into homogenous samples and lots for analyses were wrapped in aluminium foil, placed in cool boxes with ice blocks and transported to NaCRRI on the same day. On arrival at NaCRRI, these samples were kept at -80ºC overnight and were analysed for TCC and DMC the following day. Keeping samples covered in aluminium foil and low temperatures were precautions to avoid photo-oxidation and enzymatic degradation of carotenoid in samples, respectively (Rodriguez-Amaya and Kimura 2004).

4.2.5 Data analysis

Datasets for each environment and sampling point were initially analysed independently and error variances tested for homogeneity using Hartley's Fmax test (Hartley 1950), but differences were non-significant ($P < 0.05$). Therefore, un-weighted combined AMMI analysis was performed across environments by exploiting the features of interaction principal axis component analysis (IPCA) (Gauch 2006, Gauch et al. 2008, Gauch 2013) in version 36.5.1 of the Agrobase software (Agronomix Software 2013), using the model:

$$Y_{ge} = \mu + \alpha_g + \beta_e + \sum_{n}^{N} \lambda_n \gamma_{gn} \sigma_{en} + \varepsilon_{ge}$$

where:

Y_{ge}= trait value of genotype *g* in environment *e*

μ = grand mean

α_g = genotype deviation from the grand mean

β_e = environment deviation from the grand mean

N = number of interaction principal components (IPC) considered

λ_n = singular value for the IPC *n*

γ_{gn} = the element of eigenvector for genotype *g* and IPC *n*

σ_{en} = the element of eigenvector for environment *e* and IPC *n*

ε_{ge} = random error.

The AMMI analysis showed that mean squares for interaction principal component axis two (IPCA2) were non-significant for all traits evaluated at 12 MAP, which is the optimal crop age for phenotypic evaluation of cassava. Thus, the AMMI1 model was adopted and biplots of the IPCA1 scores versus genotype and environment means were presented for measured traits. The AMMI analyses were complemented with GGE biplot analysis. The first two principal components were used to obtain GGE biplots using the PBTools software (PBTools 2014). To generate a biplot for visual analysis of multi-environment data, the singular values were partitioned into genotype and environment eigenvectors so that the GGE biplot model described by Yan and Kang (2002) and Gauch et al. (2008) was used as:

$$Y_{ge} = \mu + \beta_e + \sum_{n}^{N} \lambda_n \gamma_{gn} \sigma_{en} + \varepsilon_{ge}$$

where:

Y_{ge} = trait value of genotype *g* in environment *e*

μ = grand mean

β_e = environment deviation from the grand mean

N = number of interaction principal components (IPC) considered

λ_n = singular value for the IPC *n*

γ_{gn} = the element of eigenvector for genotype *g* and IPC *n*

σ_{en} = the element of eigenvector for environment *e* and IPC *n*

ε_{ge} = random error.

Collectively, AMMI and GGE biplots were used to assess the performance and interaction patterns of genotypes and environments. Based on AMMI, a genotype with absolute IPCA1 scores close to zero indicated low interaction and was considered to be stable. Based on biplots, genotypes with broad or specific adaptation to target agroecologies or environments for traits evaluated were identified.

In stability analysis, it is possible to find a highly stable genotype that is not necessarily

the best performer for traits of interest. To overcome this challenge, the genotype selection index (GSI) was adopted, which simultaneously selects for performance and stability (Farshadfar et al. 2013). For a given genotype, GSI is the sum of the corresponding rankings for mean performance and the AMMI stability value (ASV). The ASV is a measure of the stability of a genotype based on weighted IPCA1 and IPCA2 scores. Lower values of ASV indicate greater stability of the genotype (Purchase et al. 2000). For this study, IPCA2 axes were non-significant for all traits at 12 MAP, the optimum age of physiological maturity of cassava; therefore, GSI was modified such that ranking was only based on IPCA1 as indicated below:

$$GSI_g = RIPCA1_g + RY_g$$

where:
GSI_g = genotype stability index for genotype *g* across locations for each trait
$RIPCA1_g$ = rank of genotype *g* across environments based on IPCA1
RY_g = rank of genotype *g* based on mean performance across locations.

Subsequently, genotypes with the lowest GSI for a given trait were considered to have the highest combined performance and stability (Farshadfar et al. 2013).

Meanwhile, estimates of variance components were used to calculate heritability of traits, such that:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{g\times e}^2 + \sigma_e^2}$$

where:
H^2 = broad sense heritability
σ_g^2 = variance component for genotype effects
$\sigma_{g\times e}^2$ = variance component for interaction between genotype and environment
σ_e^2 = variance component for residual effects.

4.3 Results

4.3.1 Soil and weather conditions at experimental locations

Soil nutrient content in each experimental field was analysed prior to planting the GEI trials (Table 4.3). Although levels of phosphorus, boron and copper were low, the overall

nutrient profiles of the trial fields were within the range reported to be effective enough to support cassava production (Cadavid 2012; Hauser et al. 2014). Weather data collected during the experimentation period indicated variations in temperatures and amount of rainfall received at the experimental locations. Nonetheless, the weather conditions remained within the range that could sufficiently support cassava growth and production (Hauser et al. 2014).

Table 4.3 Soil and weather characteristics of the six environments of the genotype by environment interaction trials

Parameter	Critical[d]	E1[e]	E2	E3	E4	E5	E6
pH	4.00 - 8.00*	6.10	6.20	5.70	5.60	6.0	6.10
OM[a] (%)	3.00	3.12	2.12	5.60	5.10	3.60	3.90
N (%)	0.20	0.31	0.36	0.41	0.24	0.29	0.19
P (ppm)	10.00*	1.90	2.21	2.10	0.90	4.80	4.80
Ca (ppm)	50.00*	3 724	3 689	4 833	4 933	4 224	3 724
Mg (ppm)	14.30*	742	644	1 215	1 221	681	581
K (ppm)	58.50*	356	389	255	164	630	639
Zn (ppm)	1.00	2.22	2.01	1.70	1.30	4.10	3.41
B (ppm)	0.20*	0.04	0.04	0.02	0.01	0.06	0.08
Cu (ppm)	5.00	3.13	3.00	2.00	2.01	3.10	3.10
Fe (ppm)	50.00*	189	172	182	191	172	189
Mn (ppm)	20.00	144	138	189	180	156	165
Rainfall mm		1 278	1 300	1 306	1 343	1 370	1 396
Min T[b] (°C)		17.6	18.2	18.4	18.7	18.0	19.0
Max T[c] (°C)		30.4	31.1	29.2	29.5	28.3	29.1

[a]Organic matter content; [b]Minimum temperature; [c]Maximum temperature; [d]Critical values for levels of nutrients required for crop growth; Values with asterisks are critical for cassava (Cadavid 2012); [e]Six environments (E1-E6) as defined in Table 4.2.

4.3.2 Additive main effect and multiplicative interaction analysis

Combined AMMI analysis showed varying levels of significance for the mean squares of traits measured across the crop age (Table 4.4). Genotype mean squares were highly significant ($P \leq 0.001$) for all traits evaluated at different crop ages, indicating wide phenotypic variability in genotypes used in this study. There were significant ($P \leq 0.05$) GEI mean squares for all traits measured at every sampling age of the crop, except for DMC and FRW at 15 MAP. However, environmental differences were nonsignificant for TCC at 12 and 15 MAP.

IPCA1 mean squares were significant, with varying levels of significance, for all traits at all crop ages. IPCA2 mean squares were highly significant ($P \leq 0.001$) for TCC at 6 MAP and DMC at 9 MAP, very significant ($P \leq 0.01$) for DMC at 9 MAP, significant ($P \leq 0.05$) for FRW at 9 MAP and TCC at 15 MAP ($P \leq 0.01$), but non-significant for HI at all crop ages. For all traits studied, both IPCA1 and IPCA2 accounted for more than 80% of the total variation observed in GEI, which was confirmed by the significant GEI effects for traits (with exception of HI). Overall, H^2 was high, with the smallest value of 0.64 observed for HI. CV ranged from 2.98-20.37%, which was within the acceptable range.

Table 4.4 AMMI analysis of 13 cassava genotypes phenotyped in six environments in Uganda

Source of variation	DF[a]	TCC6[b]	DMC6[c]	FRW6[d]	HI6[e]
Environment (E)	5	2.82***	4.53*	3.81***	0.02**
Genotype (G)	12	8.40***	26.72***	3.16***	0.03***
GEI[f]	60	0.41***	0.77**	0.04***	0.04*
IPCA1[g]	16	1.37***	1.56**	0.10***	0.02**
IPCA2[h]	14	0.12***	1.27	0.02	0.00
Residual	144	0.03	0.75	0.01	0.00
CV[i] (%)		4.46	4.68	9.16	5.34
H^2[j]		0.71	0.89	0.68	0.64
%GEI due to IPCA1		88.46	54.43	78.47	58.10
%GEI due to IPCA2		6.87	38.76	12.14	30.99
		TCC9	DMC9	FRW9	HI9
E	5	0.39***	4.99**	7.55***	0.08**
G	12	21.03***	49.47***	36.23***	0.13***
GEI	60	0.23***	2.36***	0.21***	0.11**
IPCA1	16	0.75***	6.40***	0.63***	0.06**
IPCA2	14	0.07	2.21***	0.13*	0.00
Residual	144	0.05	0.68	0.07	0.01
CV (%)		3.97	2.99	6.76	20.24
H^2		0.81	0.89	0.74	0.71
%GEI due to IPCA1		85.20	72.32	78.98	78.26
%GEI due to IPCA2		6.91	21.89	14.16	9.32

[a]Degrees of freedom; [b]Total carotenoid content; [c]Dry matter content; [d]Fresh root weight; [e]Harvest index; [f]Genotype by environment interaction; [g]Interaction principal component axis 1; [h]Interaction principal component axis 2; [i]Coefficient of variation; [j]Broad sense heritability; Numbers after each trait acronym refer to the crop age (months after planting) of data collection; *$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.001$.

Table 4.4 AMMI analysis of 13 cassava genotypes phenotyped in six environments in Uganda *(continued)*

Source of variation	DF[a]	TCC12[b]	DMC12[c]	FRW12[d]	HI12[e]
Environment (E)	5	1.15	11.09**	2.93***	0.15***
Genotype (G)	12	29.37***	33.53***	51.06***	0.26***
GEI[f]	60	0.32***	1.90***	0.37***	0.13**
IPCA1[g]	16	0.75***	4.19***	0.98***	0.08**
IPCA2[h]	14	0.27	1.81	0.29	0.18
Residual	144	0.13	0.83	0.08	0.01
CV[i] (%)		4.84	2.98	5.64	20.37
H^{2j}		0.78	0.94	0.71	0.68
%GEI due to IPCA1		61.87	58.89	71.77	89.74
%GEI due to IPCA2		19.51	22.28	18.41	7.94
		TCC15	DMC15	FRW15	HI15
E	5	0.29	40.59***	2.91	0.18***
G	12	28.46***	57.57***	51.46***	0.32***
GEI	60	0.39***	1.11	0.37	0.11***
IPCA1	16	0.82***	2.44**	0.98**	0.03*
IPCA2	14	0.48**	1.38	0.29	0.00
Residual	144	0.17	1.06	0.08	0.01
CV (%)		5.48	3.34	5.63	19.99
H^2		0.86	0.91	0.78	0.71
%GEI due to IPCA1		55.37	58.77	71.37	85.43
%GEI due to IPCA2		28.71	29.21	18.56	8.32

[a]Degrees of freedom; [b]Total carotenoid content; [c]Dry matter content; [d]Fresh root weight; [e]Harvest index; [f]Genotype by environment interaction; [g]Interaction principal component axis 1; [h]Interaction principal component axis 2; [i]Coefficient of variation; [j]Broad sense heritability; Numbers after each trait acronym refer to the crop age (months after planting) of data collection; *$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.001$.

4.3.3 Variation in root traits with crop age

ANOVA was performed for root traits where crop age was treated as a factor. In this case, the mean squares for crop age varied significantly for all traits measured (Table 4.5). Environmental effects were significant for all traits, except for TCC, which was consistent with results of the AMMI analysis for TCC at 12 and 15 MAP.

A comparison of mean values of traits across crop ages indicated significant increments in trait values at different cassava growth ages, but differences between mean values at

12 and 15 MAP were non-significant (Table 4.6), except for HI. Nonetheless, there were some genotypes that showed significant increase in levels of TCC (G1, G7, G8 and G12) and DMC (G2, G4, G6 and G10) from 6-15 MAP. There was no increment in FRW from 12-15 MAP.

Table 4.5 Combined analysis of variance of 13 cassava genotypes evaluated at different crop ages

Source of variation	DF[a]	TCC[b]	DMC[c]	FRW[d]	HI[e]
Environment (E)	5	1.24	25.32**	383.23***	0.01
Genotype (G)	12	83.53***	139.42***	2677.18***	0.44***
Crop age (A)	3	808.38***	7 978.41***	13 890.12***	1.68***
GEI[f]	60	0.83***	2.47***	17.12***	0.01***
GAI[g]	36	2.04***	7.13***	149.83***	0.03***
GEAI[h]	195	0.22***	2.23***	5.42***	0.02**
Residual	612	0.14	1.34	1.03	0.00
CV[i] (%)		14.3	22.2	24.3	28.3
H^{2j}		0.79	0.88	0.67	0.70

[a]Degrees of freedom; [b]Total carotenoid content; [c]Dry matter content; [d]Fresh root weight; [e]Harvest index; [f]Genotype by environment interaction; [g]Genotype by age interaction; [h]Genotype by environment by age interaction; [i]Coefficient of variation; [j]Heritability; **$P \leq 0.01$; ***$P \leq 0.001$.

4.3.4 Mean performance of genotypes

4.3.4.1 Total carotenoid content

Based on GSI for TCC, genotype MH02-073HS ranked highest, while ANDIFEKU ranked lowest for the trait (Table 4.7). The improved genotypes introduced from CIAT and IITA had higher TCC values compared to Ugandan landraces (ANDIFEKU, MAYAYA and BUSIA). MM06-2862 was the most stable genotype across all test environments, based on its low value of 0.015 for absolute IPCA1 score. Meanwhile, MH07-0529 ranked lowest for stability based on the high value (0.980) for IPCA1 score. However, all the absolute values of IPCA1 scores were close to zero, which indicated a generally stable performance across the six environments. Based on GSI, genotype MM06-0466 ranked as the best performer for TCC while genotypes ANDIFEKU, MAYAYA and MH07-0529 performed poorest. The test environments showed low absolute IPCA1 scores for TCC: E1 = 0.774, E2 = 0.309, E3 = 0.486, E4 = -0.255, E5 = -0.916 and E6 = -0.397. The AMMI1 biplot scattered these environments close to the axis for mean TCC, suggesting low interaction effects for environments (Figure 4.2).

Table 4.6 Means of four traits measured at different crop ages in 13 genotypes across six environments in Uganda

Trait	Age[a]	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	Mean
TCC[b]	6	3.5d	2.6c	3.5c	3.3c	5.1c	4.1c	4.1d	4.0d	4.9c	4.0c	3.9c	3.9d	3.0c	3.8c
	9	5.0c	3.5b	5.1b	4.7b	7.6b	6.4b	5.6c	5.5c	6.7b	6.1b	5.8b	5.4c	4.2b	5.5b
	12	6.8b	5.0a	7.5a	6.5a	9.8a	8.5a	7.3b	7.3b	8.7a	8.4a	7.8a	6.7a	5.9a	7.4a
	15	7.2a	5.2a	7.5a	6.7a	9.9a	8.5a	7.6a	7.7a	8.9a	8.5a	7.8a	7.3b	5.8a	7.6a
DMC[c]	6	17.5c	20.8d	19.1c	20.1d	16.4c	18.8d	18.7c	18.7c	18.2c	19.9d	17.7c	18.3c	17.3c	18.6c
	9	26.0b	29.8c	28.0b	29.1c	24.5b	28.0c	28.9b	26.9b	25.9b	30.1c	26.6b	26.5b	27.7b	27.5b
	12	29.6a	31.6b	32.3a	32.0b	27.7a	29.7b	30.1a	31.0a	30.4a	32.5b	29.4a	29.7a	31.0a	30.5a
	15	29.8a	33.6a	31.8a	32.9a	27.7a	31.6a	30.1a	30.4a	30.7a	33.8a	28.8a	29.9a	30.9a	30.9a
FRW[d]	6	1.7c	0.9c	0.8c	0.4c	1.3c	1.2c	1.5c	1.8c	1.2c	1.8c	1.2c	1.8c	1.0c	1.3c
	9	5.7b	1.8b	2.4b	1.2b	4.6b	3.2b	4.1b	4.9b	2.7b	5.0b	4.2b	5.6b	2.8b	3.8b
	12	6.9a	2.9a	3.0a	1.6a	6.1a	5.4a	5.1a	6.0a	4.7a	6.3a	5.8a	7.1a	3.7a	5.0a
	15	7.0a	2.9a	3.0a	1.7a	6.1a	5.5a	5.1a	6.0a	4.8a	6.3a	5.9a	7.1a	3.8a	5.0a
HI[e]	6	0.3c	0.2b	0.2b	0.1b	0.2c	0.2c	0.2c	0.2d	0.2c	0.3c	0.2b	0.2c	0.2c	0.2d
	9	0.4b	0.4a	0.2b	0.2a	0.4b	0.3b	0.3b	0.4c	0.4b	0.4b	0.4a	0.4b	0.4b	0.3c
	12	0.5a	0.4a	0.3a	0.2a	0.5a	0.3b	0.3b	0.5b	0.5a	0.5a	0.4a	0.4b	0.4b	0.4b
	15	0.5a	0.4a	0.3a	0.2a	0.6a	0.4a	0.4a	0.7a	0.5a	0.5a	0.4a	0.5a	0.6a	0.5a

[a]Crop age (months after planting) at which traits were measured; [b]Total carotenoid content (μg g^{-1}); [c]Dry matter content or roots (%); [d]Fresh root weight (kg $plant^{-1}$); [e]Harvest index; For each genotype, means with the same letter are not statistically different; G1-G13 are genotypes defined in Table 4.1; With the exception of HI, no significant difference was observed between root traits phenotyped at 12 and 15 MAP.

Table 4.7 Ranking of 13 cassava genotypes based on the genotype selection index for total carotenoid content

Genotype	TCC[a]	RTCC[b]	IPCA1[c]	RIPCA1[d]	GSI[e]	RGSI[f]
91-01730	6.76	9	0.185	6	15	8
ANDIFEKU	4.95	13	-0.237	8	21	11
CPCR15B-26	7.49	6	0.146	4	10	4
MAYAYA	6.52	11	0.393	10	21	11
MH02-073HS	9.81	1	0.450	11	12	7
MH04-2757	8.48	3	0.651	12	15	8
MH05-0452	7.32	7	-0.023	2	9	2
MH07-0529	7.26	8	-0.980	13	21	11
MM01-0014	8.68	2	-0.238	9	11	5
MM01-1003	8.36	4	-0.162	5	9	2
MM06-0466	7.82	5	0.060	3	8	1
MM06-2862	6.73	10	-0.015	1	11	5
BUSIA	5.91	12	-0.231	7	19	10

[a]Total carotenoid content (µg g^{-1}); [b]Rank of genotypes solely based on the mean TCC; [c]Interaction principal component axis 1; [d]Rank of genotypes based on absolute value of IPCA1; [e]Genotype selection index, computed as summation of the ranking as described by Farshadfar et al. 2013; [f]Rank of genotypes based on GSI.

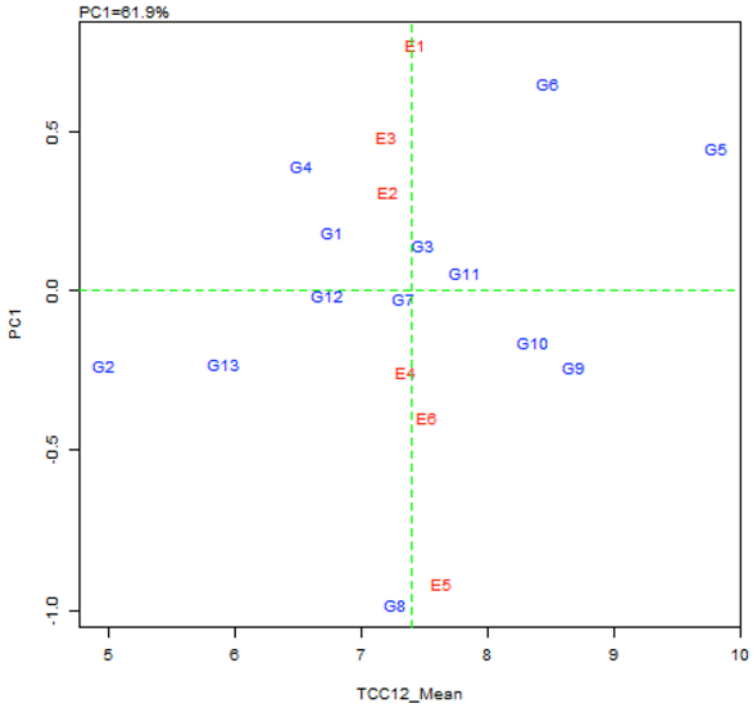


Figure 4.2 AMMI1 biplot for mean total carotenoid content (TCC) and PC1 scores for 13 cassava genotypes evaluated in six environments in Uganda. TCC12_Mean = mean of TCC (µg g^{-1}) at 12 months after planting; genotype (G1-G13) and environment (E1-E6) names are as defined in Tables 4.1 and 4.2, respectively.

4.3.4.2 Dry matter content

DMC was highest in genotype MM01-1003 and lowest in MH02-073HS (Table 4.8). Overall, landraces ANDIFEKU, MAYAYA and BUSIA had higher DMC compared to the carotene-rich genotypes introduced from CIAT and IITA. BUSIA had the lowest absolute value for IPCA1 (-0.115), ranking as the most stable genotype across the test environments. MH04-2757 ranked lowest for stability, based on the absolute IPCA1 score of 1.183. With the exception of MH04-2757, all other genotypes had absolute values for IPCA1 scores close to zero, which indicated that the performance of genotypes was generally stable for DMC across the six test environments. Interestingly, BUSIA ranked highest based on GSI for DMC, which matched its rank based on stability. Therefore, BUSIA was the best performer for DMC, whereas MH04-2757 ranked as the poorest performer for the trait.

Table 4.8 Ranking of 13 cassava genotypes based on the genotype selection index for dry matter content

Genotype	DMC[a]	RDMC[b]	IPCA1[c]	RIPCA1[d]	GSI[e]	RGSI[f]
91-01730	29.60	11	0.503	8	19	12
ANDIFEKU	31.64	4	-0.493	6	10	3
CPCR15B-26	32.25	2	-0.562	9	11	4
MAYAYA	31.97	3	0.578	10	13	5
MH02-073HS	27.72	13	0.361	2	15	8
MH04-2757	29.69	10	1.183	13	23	13
MH05-0452	30.13	8	-0.494	7	15	8
MH07-0529	30.95	6	-0.370	3	9	2
MM01-0014	30.41	7	-0.654	11	18	11
MM01-1003	32.48	1	0.925	12	13	5
MM06-0466	29.42	12	-0.443	5	17	10
MM06-2862	29.74	9	-0.420	4	13	5
BUSIA	31.04	5	-0.115	1	6	1

[a]Dry matter content (%); [b]Rank of genotypes based on the mean DMC; [c]Interaction principal component axis 1 for DMC; [d]Rank of genotypes based on absolute IPCA1 for DMC; [e]Genotype selection index, computed as a summation of the rankings as described by Farshadfar et al. 2013; [f]Rank of genotypes based on GSI.

Environments also showed low absolute IPCA1 scores for DMC: E1 = -1.211, E2 = -0.397, E3 = -0.949, E4 = 0.956, E5 = 0.856 and E6 = 0.744. The absolute values were generally close to zero, except for E1, indicating low interaction between genotypes and

test environments, which was consistent with the corresponding AMMI biplot display pattern (Figure 4.3).

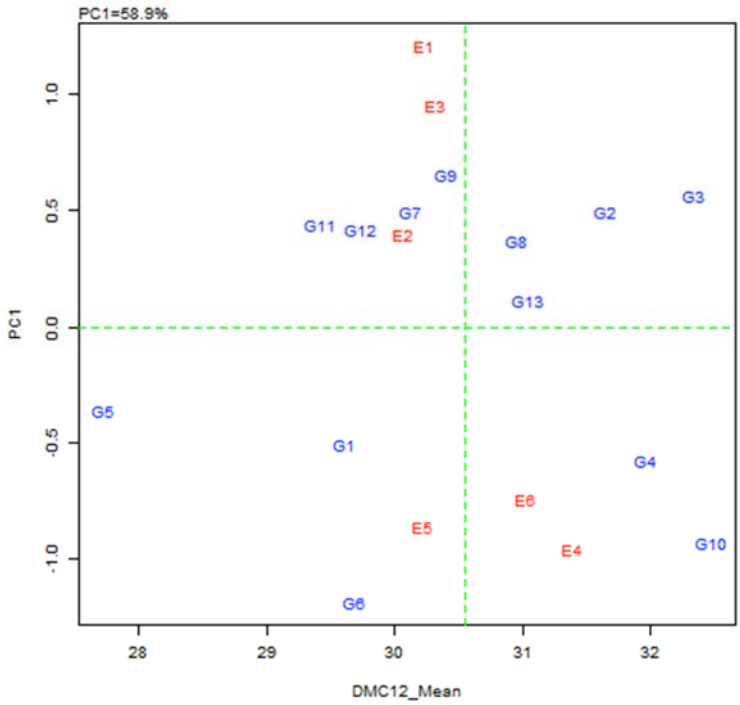


Figure 4.3 AMMI1 biplot for mean dry matter content (DMC) and PC1 scores for 13 cassava genotypes evaluated in six environments in Uganda. DMC12_Mean = mean of DMC (%) at 12 months after planting; genotype (G1-G13) and environment (E1-E6) names are as defined in Tables 4.1 and 4.2, respectively.

4.3.4.3 Fresh root weight

MM06-2862 ranked highest for mean FRW, while MAYAYA ranked lowest for the trait (Table 4.9). Overall, landraces ANDIFEKU, MAYAYA and BUSIA had low ranks for FRW compared to introductions from CIAT and IITA. MH07-0529 had the lowest absolute value for IPCA1 score (0.042), making it the most stable genotype for FRW; but MAYAYA had the highest absolute value of IPCA1 score (0.985), implying it was the most unstable and specifically adapted genotype. Meanwhile, genotype MH07-0529 and MM01-1003 were the best overall performers for FRW, based on the GSI. MAYAYA, which is a landrace, had the lowest rank for all parameters used for assessing the genotypes' performance. The IPCA1 scores for environments were also relatively low for FRW: E1 = 0.675, E2 = 0.696, E3 = 0.320, E4 = -0.873, E5 = -0.136 and E6 = -0.682, indicating low environmental interaction effects for this trait (Figure 4.4).

Table 4.9 Ranking of 13 cassava genotypes based on the genotype selection index for fresh root weight

Genotype	FRW[a]	RFRW[b]	IPCA1[c]	RIPCA1[d]	GSI[e]	RGSI[f]
91-01730	6.91	2	-0.538	11	13	6
ANDIFEKU	2.88	12	0.193	6	18	11
CPCR15B-26	2.98	11	0.624	12	23	12
MAYAYA	1.60	13	0.985	13	26	13
MH02-073HS	6.10	4	-0.404	9	13	6
MH04-2757	5.42	7	-0.223	7	14	8
MH05-0452	5.09	8	0.112	4	12	5
MH07-0529	6.00	5	-0.042	1	6	1
MM01-0014	4.73	9	0.074	2	11	4
MM01-1003	6.25	3	-0.084	3	6	1
MM06-0466	5.83	6	-0.428	10	16	10
MM06-2862	7.10	1	-0.402	8	9	3
BUSIA	3.71	10	0.131	5	15	9

[a]Fresh root weight (kg plant^{-1}) based on average of 10 plants plot^{-1}; [b]Rank of genotypes based on the mean FRW; [c]Interaction principal component axis 1 for FRW; [d]Rank of genotypes based on IPCA1 for FRW; [e]Genotype selection index; [f]Rank of genotypes based on GSI.

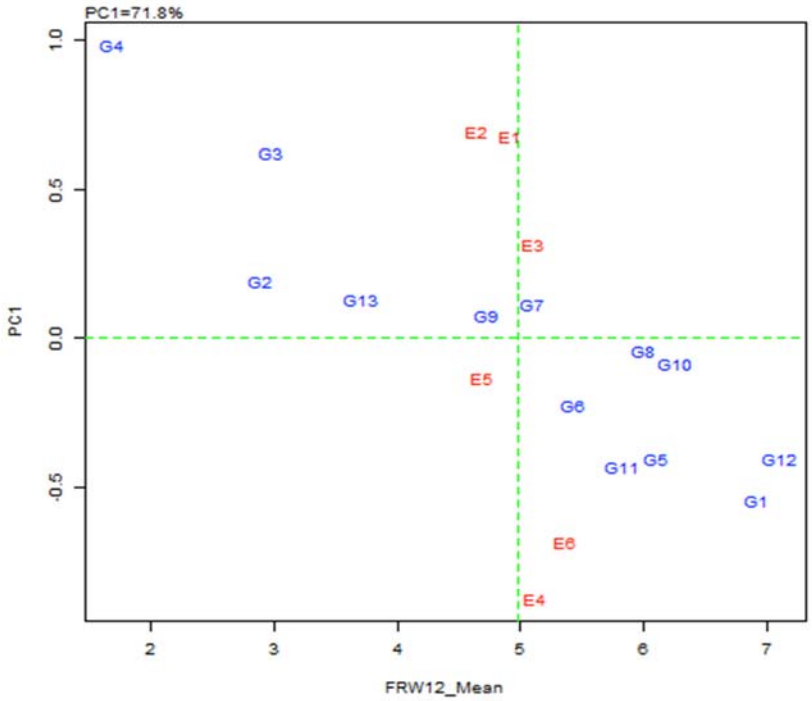


Figure 4.4 AMMI1 biplot for mean fresh root weight (FRW) and PC1 scores for 13 cassava genotypes evaluated in six environments in Uganda. FRW12_Mean = Mean of FRW at 12 months after planting (kg plant^{-1}); genotype (G1-G13) and environment (E1-E6) names are as defined in Tables 4.1 and 4.2, respectively.

4.3.5 Winning genotypes and mega-environments

For this section, results for TCC, DMC and FRW at 12 MAP are presented, as these were the focus traits. The following were the vertex genotypes: G2 (ANDIFEKU), G4 (MAYAYA), G5 (MH02-073HS), G6 (MH04-2757) and G9 (MM01-0014) for TCC (Figure 4.5 A); G3 (CPCR15B-26), G5 (MH02-073HS), G6 (MH04-2757), G9 (MM01-0014), G10 (MM01-1003) and G11 (MM06-0466) for DMC (Figure 4.5 B); G1 (91-01730), G2 (ANDIFEKU), G4 (MAYAYA), G10 (MM01-1003) and G12 (MM06-2862) for FRW (Figure 4.5 C).

Another important feature of GGE biplots is that they indicate environmental groupings, suggesting possible existence of different mega-environments. For example, two mega-environments are suggested for each trait evaluated in this study (Figure 4.5). For TCC, the first mega environment had environments E1, E2 and E3, with genotype G5 (MH02-073HS) and G6 (MH04-2757) as the best performer and the second mega environment had environments E4 and E6, with genotype G9 (MM01-0014) performing best.

For DMC, the first mega environment was constituted by E2 in which G3 (CPCR15B-26) was the most superior genotype and the second mega environment contained E4, E5 and E6, with genotype G10 (MM01-1003) as the best performer. The first mega environment for FRW was the group of E1, E3 and E6 in which G10 (MM01-1003) and G12 (MM06-2862) performed best, while the second mega environment had E4 with G1 as the best yielder. In practice, it is difficult to have a mega-environment per trait; data generated in this study could suggest (within limits) E4 as a candidate mega environment for provitamin A cassava trials.

4.3.6 Phenotypic correlations among traits studied

Phenotypic values at 12 and 15 MAP were used for the correlation analysis because they reflected optimal levels of the traits. For traits, there were positively significant correlations between mean values at different crop ages, suggesting temporal accumulation of phenotypes studied. DMC had a significant negative correlation with TCC (Table 4.10), with *r* values comparable to those reported in section 3.3.7.

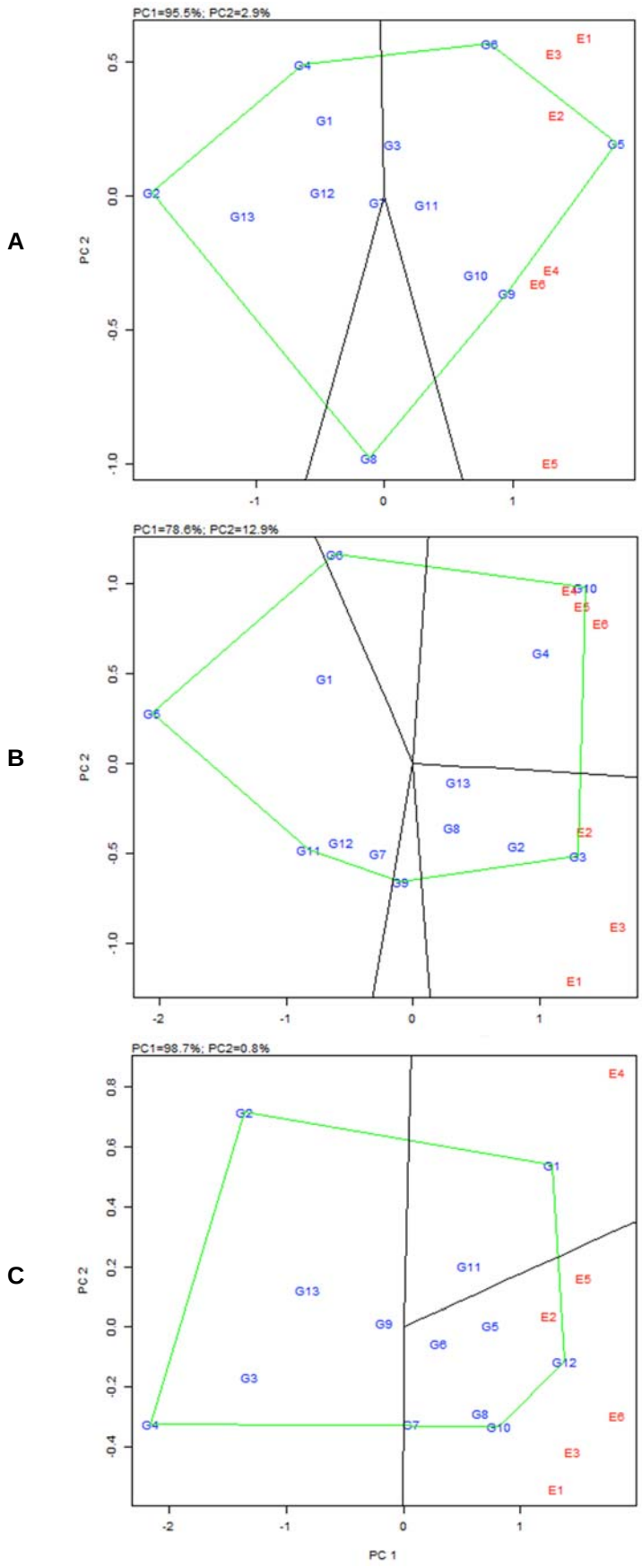


Figure 4.5 Polygon views of the GGE biplots based on symmetrical scaling for the which-won-where pattern of genotypes and environments for total carotenoid content (A), dry matter content (B) and fresh root weight (C). G1-G13 and E1-E6 are genotypes (Tables 4.1) and environments (Table 4.2), respectively.

Table 4.10 Spearman correlation coefficients among three traits phenotyped for 13 cassava genotypes in six environments in Uganda

Trait	TCC12[a]	TCC15	DMC12[b]	DMC15	FRW12[c]
TCC15	0.91***				
DMC12	-0.35**	-0.36**			
DMC15	-0.39**	-0.41**	0.68***		
FRW12	0.14	0.15	-0.11*	-0.14	
FRW15	0.13	0.18	-0.04	-0.08	0.89***

[a]Total carotenoid content; [b]Dry matter content; [c]Fresh root weight; Numbers after the trait acronyms indicate crop age at harvest (months after planting); *$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.001$.

4.4 Discussion

The overall objective of this study was to assess the stability of 13 provitamin A clones by examining their performance for TCC, DMC, FRW and HI in six environments. The significant variation for TCC, DMC and FRW in these genotypes presents an important opportunity to exploit in cassava breeding in Uganda. This variability could form the basis for making forward progress in genetic improvement of cassava for these traits through hybridisation and selection. Environmental effects were non-significant for TCC at 12 and 15 MAP but highly significant for both DMC and FRW, indicating significant variation in mean performance of genotypes for the latter traits in different environments. This could suggest that carotene content in roots reaches a plateau when a cassava plant attains physiological maturity.

An obvious deduction from the low environmental effect on TCC is that selection for the trait can effectively be achieved by evaluating target genotypes in one location. However, a typical cassava variety selection scheme involves screening no less than five candidate genotypes for fresh root yield and DMC that are key drivers of variety adoption (Fukuda et al. 2002, Owusu and Donkor 2012, Abdoulaye et al. 2014). Therefore, breeding programmes targeting development of provitamin A rich cassava varieties could use initial on-station trials for identifying carotene-rich genotypes that can later be subjected to multi-locational evaluations where focus shifts to other traits highly influenced by environmental effects. This strategy would save costs while increasing precision to identify best performers for root yield and DMC.

The AMMI analysis for all traits showed that more than 50% of the variation in GEI sum of squares (SS) was accounted for by IPCA1. Subsequently fitted IPCAs, notably IPCA2,

were non-significant, indicating that they largely captured random noise. These results compare well with those of Gauch (2006), which showed that significant IPCA1 and subsequent axes in AMMI capture interaction exclusively in a monotonic sequence that decreases from the largest component in the first axis to the smallest component in the last axis. Therefore the significance of IPCA1 scores provided the necessary confidence for considering use of AMMI biplots for visual assessment of the genotype and location performances and their interactions (Gauch et al. 2008).

It was evident in this study that ranking of genotypes based on stability alone was not consistent with their mean performance for traits. In this regard, GSI proved to be a more reliable selection criterion for identifying best performers when conducting multi-location evaluation trials. Using this selection criterion, MM06-0466 and BUSIA were identified as the best yielders for TCC and DMC, respectively. Landraces used in this study were generally the poorest performers for TCC and FRW, but best performers for DMC. Landraces offer the advantage of having alleles that enhance adaptation to local environments. Therefore, landraces studied here constitute invaluable cassava genetic backgrounds for introgressing TCC and FRW from the introductions.

In this study, genotypes were evaluated for 15 months, which means natural growth conditions in the first three environments differed from those in the last three. The first three environments (E1, E2 and E3) were the 2012-2013 trials evaluated in the three experimental sites while the last three environments (E4, E5 and E6) were a repeat of the trials evaluated during the 2013-2014 season. Actually, 2012-2013 trials were planted in May while 2013-2014 trials were planted in September, which suggests differences environmental conditions experienced by the plants during the entire growth cycle. This observation appeared to account for the characteristic clustering of environments in AMMI1 biplots for all traits studied, in which E1, E2 and E3 tended to group together and distant from E4, E5 and E6 that also grouped together.

The pattern of temporal variation in levels of TCC, DMC and FRW are important to note. Mean values for these traits progressively increased up to 12 MAP, suggesting that cassava harvested at this age would give optimal levels for these traits. The quest for early maturing cassava is gathering rapid pace as the crop's relevance for food security becomes more prominent (Tumuhimbise et al. 2012, Bassey and Harry 2013), but this is likely to be met with costs associated with less than optimal production for most of the traits before 12 MAP. Besides, other biotic constraints such as CMD and CBSD impact negatively on cassava's potential for early bulking (Tumuhimbise et al. 2014). In an

earlier study, Ngeve (2003) noted continuous increase in root yield between 6-16 MAP, with the fresh root yield increasing by up to 9.3 t ha^{-1} from 8 MAP to 12 MAP. As breeders pursue the development of early bulking cassava varieties, it is imperative that selection be made to exploit the crop's potential to accumulate other important traits, including TCC and DMC at an equally early age.

However, synthesis and accumulation of TCC and DMC in cassava roots appear to be driven by genetic factors expressed along the growth stages of cassava, which are significantly influenced by prevailing environmental conditions (Asafu-Agyei and Osafo 2000). Sagrilo et al. (2008) associated the steady increase of DMC in cassava roots with temporal partitioning and accumulation of assimilates into the storage parts, with the peak influenced by the amount of vegetative growth. They indicated that the highest carbohydrate proportions were allocated to the storage roots during periods of low vegetative growth. The current study suggested 12 MAP as the age at which optimal levels of TCC, DMC and FRY can be realised in cassava roots. In comparison to findings of the current study, Tumuhimbise et al. (2014) reported the possibility of achieving economically meaningful fresh root yield (up to 25 t ha^{-1}) at 9 MAP, but some of the genotypes evaluated in that study had a genetic background of early bulking. Nonetheless, such genotypes could provide the genetic resource for combining TCC and DMC with early bulking, through hybridisation. Based on genotypes evaluated in the current study, it would make economic sense to harvest cassava at 12 MAP, which eliminates maintenance costs required for further weeding, releases the land for production of other crops and guarantees good quality of planting material for the coming season as their storage period would be short (Ngeve 2003).

The strong negative correlation between DMC and TCC in the current study is undesirable. It is worth noting that combined selection for both DMC and carotenoid content in Latin America has been underway much longer than in Africa. Thus, such negative correlations could have been broken during the several cycles of recombination (Ceballos et al. 2013). Going forward, high DMC would be an important feature for cassava breeding efforts targeting generation of provitamin A varieties that are acceptable to farmers. DMC in the 13 genotypes studied here was less than that in varieties commonly grown by farmers in Uganda (Kawuki et al. 2011). To translate investments in cassava biofortification research into impact on human nutrition, breeding efforts will need to focus on developing varieties that combine high levels of both DMC and TCC in high-yielding genetic backgrounds. Nonetheless, the best performing

genotypes identified in this study could form the material for such genetic improvement through hybridisation.

On a positive note, best performing genotypes were identified for each of the traits studied. For example, overall performance of the genotypes based on combined raking indicated that genotypes MM01-1003, MM06-2862, MH07-0529 and MH05-0452 were top four performers for all the traits studied. Such genotypes could be of immediate importance for further evaluation and/or use in breeding. It suffices to note that the improved genotypes evaluated in this study were a set drawn from an advanced breeding population, which means they could have attained stability for important agronomic traits including FRW and DMC. Therefore, it is recommended that such genotypes be further screened for resistance to CBSD and other biotic stresses, for the possible advancement of the best clones to on-farm production.

4.5 Conclusion

The study revealed (1) significant GEI effects for TCC, DMC and FRW in the cassava genotypes evaluated, (2) continuous increment in the levels of TCC, DMC and FRW in cassava during the first 12 MAP and (3) significant genotypic variation for TCC, DMC and FRW, which could provide resources for making selections for further genetic improvement through hybridisation. The observed GEI effects for TCC did not result in changes in genotype ranks in varying environments, indicating high stability of the trait and broad adaptability of genotypes across the test environments. Therefore, it would be possible to make reliable selections for TCC based on single location data, which saves time and high costs associated with carotenoid quantification. To increase the acceptance of provitamin A cassava varieties in future, breeding efforts need to focus on consolidating strategies, including the rapid cycling recurrent selection scheme already used at CIAT, to overcome the undesirable negative correlation between DMC and TCC. To this effect, genotypes MM01-1003, MM06-2862, MH07-0529 and MH05-0452 that the best combined performance for these traits could be used for further genetic improvement through hybridisation.

4.6 References

Abdoulaye T, Abass A, Maziya-Dixon B, Tarawali G, Okechukwu R, Rusike J, Alene A, Manyong V, Ayedun B (2014) Awareness and adoption of improved cassava varieties and processing technologies in Nigeria. Journal of Development and Agricultural Economics 6:67-75

Acquaah G (2012) Principles of Plant Genetics and Breeding. Second edition. John Willey and Sons Ltd, UK

Agronomix Software (2013) AGROBASE Generation II User's Guide. Release version 36.5.1 in MS-SQL. Agronomix Software Inc, Winnipeg, MB, Canada

Agyeman A, Parkes EY, Peprah BB (2015) AMMI and GGE biplot analyses of root yield performance of cassava genotypes in forest and coastal ecologies. International Journal of Agricultural Policy and Research 3:122-132

Akinwale MG, Akinyele BO, Odiyi AC, Dixon AGO (2011) Genotype x environment interaction and yield performance of 43 improved cassava (*Manihot esculenta* Crantz) genotypes at three agro-climatic zones in Nigeria. British Biotechnology Journal 1:68-84

Asafu-Agyei JN, Osafo DM (2000) Dry matter production and distribution model in intercropped cassava. Ghana Journal of Agricultural Science 34:27-37

Bassey EE, Harry GI (2013) Screening cassava (*Manihot esculenta* Crantz) genotypes for tuber bulking, early maturity and optimum harvesting time in Uyo, south eastern Nigeria. Peak Journal of Agricultural Sciences 1:83-88

Boy E, Mannar V, Pandav C, de Benoist B, Viteri F, Fontaine O, Hotz C (2009) Achievements, challenges and promising new approaches in vitamin and mineral deficiency control. Nutrition Reviews 67:24-30

Cadavid LF (2012) Soils and fertilizers for the cassava crop. In: Ospina B, Ceballos H (eds) Cassava in the Third Millennium: Modern Production, Processing, Use, and Marketing Systems, vol 377. CIAT, Cali, Colombia, pp 113-137

Crossa J (1990) Statistical analyses of multilocation trials. Advances in Agronomy 44:55-85

Eberhart SA, Russell WA (1966) Stability parameters for comparing varieties. Crop Science 6:36-40

Egesi CN, Ilona P, Ogbe FO, Dixon AG (2007) Genetic variation and genotype x environment interaction for yield and other agronomic traits in cassava in Nigeria. Agronomy Journal 99:1137-1142

El-Sharkawy MA (2007) Physiological characteristics of cassava tolerance to prolonged drought in the tropics: implications for breeding cultivars adapted to seasonally dry and semiarid environments. Brazialian Journal of Plant Physiology 19:257-286

Esuma W, Rubaihayo P, Pariyo A, Kawuki R, Wanjala B, Nzuki I, Harvey JJW, Baguma Y (2012) Genetic diversity of provitamin A cassava in Uganda. Journal of Plant Studies 1:60-71

Farshadfar E, Rashidi M, Mahdi MJ, Zali H (2013) GGE biplot analysis of genotype x environment interaction in chickpea genotypes. European Journal of Experimental Biology 3:417-423

Fukuda WMG, Silva SO, Iglesias CA (2002) Cassava breeding. Crop Breeding and Applied Biotechnology 2:617-638

Gauch GH (2006) Statistical analysis of yield trials by AMMI and GGE. Crop Science 46:1488-1500

Gauch GH (2013) A simple protocol for AMMI analysis of yield trials. Crop Science 53:1860-1869

Gauch GH, Piephob H, Annicchiarico P (2008) Statistical analysis of yield trials by AMMI and GGE: further considerations. Crop Science 48:866-889

Hartley HO (1950) The use of range in analysis of variance. Biometrika 37:271-280

Hauser S, Wairegi L, Asadu CLA, Asawalam DO, Jokthan G, Ugbe U (2014) Cassava System Cropping Guide: Africa Soil Health Consortium. CAB International, Nairobi

Hongyu K, García-Peña M, Araújo LB, Santos DCT (2014) Statistical analysis of yield trials by AMMI analysis of genotype x environment interaction. Biometrical Letters 51:89-102

Kang MS (2002) Quantitative Genetics, Genomics, and Plant Breeding. CABI Publishing, New York, USA

Kawuki RS, Pariyo A, Amuge T, Nuwamanya E, Ssemakula G, Tumwesigye S, Bua A, Baguma Y, Omongo C, Alicai T, Orone J (2011) A breeding scheme for local adoption of cassava (*Manihot esculenta* Crantz). Journal of Plant Breeding and Crop Science 3:120-130

Maroya NG, Kulakow P, Dixon AG, Maziya-Dixon B, Baker KS (2012) Genotype x environment interaction of carotene content of yellow-fleshed cassava genotypes in Nigeria. Journal of Life Sciences 6:595-601

Mayer JE, Pfeiffer WH, Beyer P (2008) Biofortified crops to alleviate micronutrient malnutrition. Current Opinion in Plant Biology 11:166-170

Nassar NMA, Ortiz R (2006) Cassava improvement: challenges and impacts. Journal of Agricultural Science 145:163-171

Ngeve JM (2003) Cassava root yields and culinary qualities as affected by harvest age and test environment. Journal of the Science of Food and Agriculture 83:249-257

Oliveira EJ, Aidar ST, Morgante CV, Chaves ARM, Cruz JL, Coelho FMA (2015) Genetic parameters for drought-tolerance in cassava. Pesquisa Agropecuária Brasileira 50:233-241

Owusu V, Donkor E (2012) Adoption of improved cassava varieties in Ghana. Agricultural Journal 7:146-151

Pariyo A, Baguma Y, Alicai T, Kawuki R, Kanju E, Bua A, Omongo C, Gibson P, Osiru DS, Mpairwe D, Tukamuhabwa P (2015) Stability of resistance to cassava brown streak disease in major agro-ecologies of Uganda. Journal of Plant Breeding and Crop Science 7:67-78

PBTools (2014) Biometrics and Breeding Informatics, Version 1.4. PBGB Division. International Rice Research Institute, Los Banos, Laguna

Piepho HP (1994) Best linear unbiased prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis. Theoretical and Applied Genetics 89:647-654

Purchase JL, Hatting H, Van Deventer SC (2000) Genotype x environment interaction of winter wheat (*Triticum aestivum* L.) in South Africa: II. Stability analysis of yield performance. South African Journal of Plant and Soil 17:101-107

Rao PS, Reddy PS, Rathore A, Reddy BVS, Panwar S (2011) Application of GGE biplot and AMMI model to evaluate sweet sorghum (*Sorghum bicolor)* hybrids for genotype x environment interaction and seasonal adaptation. Indian Journal of Agricultural Science 81:438-444

Rodriguez-Amaya BD, Kimura M (2004) HarvestPlus Handbook for Carotenoid Analysis, vol 1. International Food Policy Research Institute (IFPRI) and International Center for Tropical Agriculture (CIAT), Washington, DC and Cali

Sagrilo E, Filho PSV, Pequeno MG, Gonçalves-Vidigal MC, Kvitschal MV (2008) Dry matter production and distribution in three cassava (*Manihot esculenta* Crantz) cultivars during the second vegetative plant cycle. Brazilian Archives of Biology and Technology 51:1079-1087

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

Thompson B, Amoroso L (2011) Combating Micronutrient Deficiencies: Food-Based Approaches, vol 4. CAB International and FAO, Oxfordshire, UK

Tumuhimbise R, Melis R, Shanahan P, Kawuki R (2012) Farmers' perceptions on early storage root bulking in cassava (*Manihot esculenta* Crantz) in east and central Uganda and their implication for cassava breeding. World Journal of Agricultural Sciences 8:403-408

Tumuhimbise R, Melis R, Shanahan P, Kawuki R (2014) Genotype x environment interaction effects on early fresh storage root yield and related traits in cassava. The Crop Journal 2:329-337

Yan W, Kang MS (2002) GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists and Agronomists. CRC Press, New York

Yan W, Tinker N (2006) Biplot analysis of multi-environment trial data: principles and applications. Canadian Journal of Plant Science 86:623-645

Ye G, McNeil DL, Hill DL (2013) Methods for analysing multi-site plant variety trials II. Selection for yield and stability. Agronomy Society of New Zealand 31:25-33

CHAPTER 5

Genome-wide association study of carotenoid and dry matter content in cassava

5.1 Introduction

As the world population size approaches the 9 billion mark, cassava is projected to be a cornerstone on which to rest aspirations for meeting increased calorie demands in developing countries of SSA, which are currently experiencing the highest population growth rates in the world (Edgerton 2009; Burns et al. 2010). However, more than 90% of the edible portion of cassava roots consists of carbohydrates (Montagnac et al. 2009), making diets that heavily depend on cassava to be deficient in essential micronutrients such as vitamin A, iron and zinc. Consequently, people that over depend on cassava for food are exposed to deficiencies of such micronutrients that cause, for example, corneal blindness, reduced immunity to preventable diseases, stunted growth, anaemia and reduced capacity for physical activity (Gegios et al. 2010; Stephenson et al. 2010).

It was in response to this special challenge that a special breeding agenda, referred to as biofortification, was initiated in the early 2000s under the HarvestPlus project to develop cassava varieties enriched with provitamin A carotenoids (Pfeiffer and McClafferty 2007). Consumption of provitamin A cassava varieties is expected to minimise VAD that is widespread among resource-poor farmers that primarily depend on cassava (Mayer et al. 2008; Talsma et al. 2013). A key obstacle towards deployment of provitamin A cassava varieties to farmers is the negative correlation between root DMC and carotenoid content (Njoku et al. 2011). High DMC is a critical trait that influences farmers' decisions to adopt new cassava varieties (Owusu and Donkor 2012; Njukwe et al. 2013; Ojo and Ogunyemi 2014).

The intrinsic heterozygous nature (Kawano et al. 1978; Ceballos et al. 2015) and high sensitivity to environmental variations make cassava a difficult crop to breed conventionally. Nonetheless, adoption of advanced molecular breeding technologies demonstrated in crops such as wheat (Poland et al. 2012), maize (Ersoz et al. 2009; Crossa et al. 2013) and rice (Yuan-yuan et al. 2014) could increase genetic gains for complex traits in cassava (Varshney et al. 2009). A classic example was the pioneering advance of the marker-assisted introgression of CMD resistance into Latin American germplasm prior to its introduction into Africa (Okogbenin et al. 2007). This is a classical form of MAS, in which either single genes or quantitative trait loci (QTL) were selected.

Patterns of phenotypic segregation for carotenoid content and DMC in cassava appear to suggest that these traits are controlled by more than one gene (Nassar et al. 2007; Akinwale et al. 2010). Efforts that can pinpoint the allelic basis of such variability would, for example, enhance chances of using gene-based selection to facilitate the rapid recurrent selection scheme that has already been demonstrated to increase carotenoid content in cassava (Ceballos et al. 2013; Ceballos et al. 2015).

Recent advances in NGS technologies have made linkage disequilibrium (LD)-based GWAS a favourable tool for mapping complex traits in higher plants (Davey et al. 2011). It is now possible to use high-density genotyping platforms such as GBS to quickly generate millions of marker data points that are distributed throughout a genome (Elshire et al. 2011; He et al. 2014; Peterson et al. 2014). The ability of GBS to provide high marker density and extensive genome coverage makes GWAS an ideal tool for unravelling the genic basis of complex traits in the highly heterozygous cassava (Oliveira et al. 2012; Ceballos et al. 2015). Owens et al. (2014) and Suwarno et al. (2015) have used these NGS platforms to identify polymorphisms linked to carotenoids within the maize genome, which could provide a foundation for systematic provitamin A biofortification in the crop.

With more than 97% of the cassava genome sequenced (Prochnik et al. 2012), it is now practical to use GBS and subsequently GWAS to precisely anchor SNPs linked to traits of interest to specific regions of the genome and infer functions of such polymorphisms based on annotations of genomic regions where they are positioned. Rabbi et al. (2014) have used SNP data generated from the GBS platform to develop a high-resolution genetic map that identified genomic regions of cassava controlling resistance to cassava mosaic geminiviruses.

The national cassava breeding programme recently initiated a breeding objective to develop cassava genetic resources that combine high levels of provitamin A carotenoid content, virus resistance and farmer-preferred traits such as DMC. It suffices to note that the inherent heterozygous nature of cassava limits the use of conventional breeding approaches to combine these traits, due to reduced accuracy and efficiency to select superior clones from a segregating population. This hurdle could be overcome by employing molecular breeding tools such as MAS or GS, but these tools have been least used for cassava breeding due to inadequate information on molecular markers for carotenoid content and DMC (Ceballos et al. 2015). Therefore, the current study was conducted to identify genomic regions and SNPs linked with natural variations for

carotenoid content and DMC in cassava as a basis for more systematic biofortification of the crop.

5.2 Materials and methods

5.2.1 Genotypes

A panel of 655 genotypes was used in this study. These genotypes included an array of S_1 and S_2 partial inbreds generated from eight S_0 progenitors of diverse genetic background (Table 5.1). S_0 progenitors were ß-carotene-rich lines selected from a set of CIAT and IITA germplasm previously introduced by the national cassava breeding programme of Uganda for developing provitamin A varieties (Esuma et al. 2012). S_0 progenitors were self-pollinated to generate S_1 progeny during the 2009-2010 season, through a prior hybridisation programme.

Thirty nine clones selected from the S_1 progeny were planted in a crossing clock in August 2011 and self-pollinated to generate S_2 plants. These S_1 plants were selected based on (1) TCC in their roots, (2) flowering ability and (3) tolerance to CBSD. Standard agronomic practices (as described in section 3.2.2) were applied in the crossing block to ensure uniform plant establishment and growth. Controlled pollination described in section 3.2.2 was used to generate both the S_1 and S_2 seeds.

Table 5.1 Pedigree and number of cassava genotypes used for the genome-wide association study

S_0 progenitor	Source	S_1	S_2	Total[c]
MH05-2575	IITA[a]	3	64	68
MH04-2767	IITA	5	47	53
CPCR15B-12	CIAT[b]	11	71	83
01/1335	IITA	7	103	111
CPCR11B-27	CIAT	4	95	100
MH05-0442	IITA	3	77	81
MH05-2757	IITA	4	114	119
CPCR-27B-17	CIAT	2	37	40
Total		39	608	655

[a]International Institute for Tropical Agriculture; [b]International Centre for Tropical Agriculture; [c]Total includes the eight S_0 progenitors.

5.2.2 Phenotyping

A total of 1 120 S_2 seedlings were generated and grown in the field at Abi-ZARDI to produce planting materials sufficient for replicated clonal trials. Seed germination and seedling evaluation (for purpose of generating planting materials) were conducted as described in sections 3.2.2 and 3.2.3, respectively. The association panel of 655 genotypes was a random selection from the total seedling population, with family numbers varying to capture adequate genetic variability. An additional criterion for selecting the association panel was the ability of the S_2 genotype to generate ≥ 12 standard-size cuttings, as described in section 3.2.4.

Clonal trials for these genotypes were conducted from June 2014 to May 2015 at two locations: Abi-ZARDI and NaCRRI research stations in Uganda, which have different agroecological characteristics, as described in section 3.3.1. Therefore, soil and weather conditions at these experimental stations were similar to those described for the clonal experiments for the diallel study presented in section 3.3.1. At each site, trials were laid out in an incomplete block design (Clewer and Scarisbrick 2001) with two replications and 20 incomplete blocks per replication. Plots in a block consisted of single rows of three plants each. Planting and weeding were done as described in section 3.2.4. Trials were harvested 12 MAP. DMC and TCC were measured using the oven and iCheck methods, respectively, as described in section 3.2.5.

5.2.3 Genotyping

5.2.3.1 DNA extraction

Genomic DNA was extracted from newly expanded plant leaves picked from seedling plants, according to the procedure described by Dellaporta et al. (1983). Approximately 0.25 g of fresh leaf tissue was placed in extraction tubes and dried using a freeze dryer (Labconco FreeZone®, MO, USA). In order to extract DNA from tissue cells, freeze-dried leaf samples were ground using a genogrinder (GenoGrinder SP-2000-115, NJ, USA). Samples were ground in a 96-well plate design, which enabled high-throughput extraction of many samples in parallel.

About 500 µl of Dellaporta extraction buffer [100 mM Tris-HCl (Tris (hydroxymethyl) aminomethane hydrochloride) pH 8, 50 mM EDTA (ethylenediaminetetraacetic acid) pH 8, 500 mM NaCl, 10 mM ß-mercaptoethanol, 2% (w/v) PVP (polyvinyl pyrrolidone) and 20% (w/v) SDS (sodium dodecyl sulfate)] was added to the tissue powder to extract cell

contents. In order to remove lipid membranes, 33 µl of 20% (w/v) laury sulphate was added to each sample, mixed thoroughly and incubated in a water bath at 65°C for 10 min. To these samples, 160 µl of 5 M potassium acetate was added and the mixture furthermore incubated at -20°C for 10 min. At this point, samples were centrifuged in an Allegra® X-12 centrifuge (Beckman Coulter, CA, USA) at 4 000 *g* for 10 min. A volume of 450 µl of the supernatant was transferred to a new microfuge tube. Cold isopropanol (450 µl) was added to the supernatant and centrifuged at 4 500 *g* for 20 min to precipitate DNA. The DNA pellet was washed of excess salt by adding 500 µl of 70% (v/v) ethanol and centrifuged at 4 000 *g* for 5 min. The supernatant from these samples was decanted and DNA pellets air-dried for 30-60 min until ethanol evaporated completely. Air-dried DNA was resuspended in 200 µl low salt TE (tris-ethylenediaminetetraacetic acid) buffer (10 mM Tris-HCl pH 7.5 and 1 mM EDTA) containing 3 µl of 10 mg μl^{-1} RNAse A and incubated at 37°C for 1-2 h.

The concentration and purity of the DNA were checked using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and agarose gel electrophoresis, respectively. To do electrophoresis, 2 ng μl^{-1} DNA solution was mixed with 2 µl loading dye (Blue/Orange 6x Loading Dye, Promega, WI, USA), loaded onto a 1.5% agarose gels stained with 0.15 µg ml^{-1} ethidium bromide and electrophoresed in 1x TAE (tris-acetatediaminetetraacetic acid) buffer (40 mM Tris-HCl pH 7.6, 20 mM acetic acid and 1 mM EDTA) at 100 V for 20 min. DNA samples were diluted to 100 ng μl^{-1}, freeze-dried and shipped to the Institute for Genomic Diversity at Cornell University, NY, USA for genotyping.

5.2.3.2 SNP genotyping

The GBS platform described by Elshire et al. (2011) was used to genotype the association panel. Briefly, restriction enzyme *Ape*KI, which is a type II restriction endonuclease that recognises a degenerate 5 bp sequence (GCWGC, where W is A or T), was used to digest DNA samples. *Ape*KI was selected as the restriction enzyme because of two reasons: it (1) has partial sensitivity to DNA methylation and (2) cuts genomic DNA at low frequency, thus avoiding repetitive genomic regions and enhancing adapter ligation to insert DNA (Peterson et al. 2014). GBS sequencing libraries (96-plex, including a blank control well) were prepared by ligating the digested DNA to barcodes (unique nucleotide adapters) followed by standard PCR. Libraries were sequenced using 1-lane 86 bp reads using a Genome Analyzer 2000 (Illumina HiSeq2000 Inc, San Diego, CA, USA). Sequence reads from different genotypes were de-convoluted using barcodes

and aligned to the cassava genome sequence v6.1 as the reference genome (Glaubitz et al. 2014). SNP calling was done using the TASSEL-GBS pipeline.

5.2.3.3 Processing of raw sequence data and SNP calling

A detailed description of related bioinformatics for processing raw reads to SNP calls can be found in Elshire et al. (2011) and Glaubitz et al. (2014). The bioinformatics procedure in the TASSEL-GBS pipeline (Glaubitz et al. 2014) was used to process the FASTQ sequence data (nucleotide sequence and its corresponding quality scores stored in a text-based format) into SNP calls based on data from the cassava genome sequence v6.1. The entire association panel of 655 genotypes was assayed using 635 247 SNPs, which were previously optimised and found to be informative for cassava GBS (Dr. Ramu Punna, Bioinformatician at Cornell University).

Genotypic data were further processed to ensure high quality for subsequent analyses. The dataset originally had close to 70% missing data points, which necessitated imputation of the non-genotyped markers, as described in earlier studies (Romay et al. 2013; Swarts et al. 2014). To achieve this, indels were first removed from the raw sequence data using TASSEL v5.2.9 (Bradbury et al. 2007) and the cleaned dataset imputed using Beagle software v4.0 based on an algorithm which searches for the closest neighbour in small SNP windows across the entire genome (Browning and Browning 2013; Swarts et al. 2014). To remove non-segregating and uninformative sites, the imputed data were further filtered at the following thresholds: minor allele frequency (MAF) = 0.01, minimum count of genotypes = 50 and minimum heterozygous proportion = 0.01. This filtered dataset represented a reduced marker set comprising 179 310 SNPs with robust MAF desirable for estimating population structure and kinship in the subsequent statistical analyses.

5.2.4 Statistical analysis

A total of 591 genotypes had phenotypic data across the two locations and were therefore considered for subsequent statistical analyses to test marker-trait associations. A mixed linear model (MLM) fitted across environments was used to conduct ANOVA using the *lme4* package of the R software (Vazquez et al. 2010; R Development Core Team 2010). The model used was:

$$Y_{ijk} = \mu + ß_i + R_{ij} + G_k + (ß_i \times G_k) + \varepsilon_{ijkm}$$

where:

Y_{ijk} = phenotypic value

μ = overall phenotypic mean

$ß_i$ = effect of environment *i*

R_{ij} = effect of block j in environment *i*

G_k = effect of genotype *k*

$(ß_i \times G_k)$ = effect of interaction between environment i and genotype *k*

ε_{ijkm} = residual.

Both the phenotypic means and BLUPs extracted from the MLM were used for genome-wide association (GWA) testing to benefit from the additional power that this might offer.

Estimates of the variance components were used to calculate heritability of traits according to Clewer and Scarisbrick (2001), such that:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{g\times e}^2 + \sigma_e^2}$$

where:

H^2 = broad sense heritability

σ_g^2 = variance component for genotype effects

$\sigma_{g\times e}^2$ = variance component for interaction between genotype and environment

σ_e^2 = variance component for residual effects.

Unless stated, all statistical analyses of the genotypic data were performed using TASSEL v5.2.9 (Bradbury et al. 2007). PCA through eigenvalue decomposition of the covariance matrix was performed using a random set of 20 000 markers to infer population structure. The first two axes of the PCA were used to draw a scatter plot to visualise genetic differentiation among genotypes. In execution of GWAS, the first three principal components constituted the population structure.

A kinship matrix was constructed using the scaled_IBS method (IBS = identical by state). This method produces a matrix that is scaled to give a reasonable estimate of the additive genetic variance, which effectively controls statistical errors arising from population structure in a MLM for GWAS. The kinship matrix was calculated according to Bishop and Williamson (1990):

$$\text{IBS distance} = \frac{[\text{No. of markers IBS2} + 0.5(\text{No. of markers IBS1})]}{\text{Number of non-missing markers}}$$

where:

IBS1 = state in which two genotypes share one allele at a locus

IBS2 = state in which two genotypes share two alleles at a locus.

Genome-wide and chromosome-wise LD was assessed based on adjacent pairwise r^2 values (the squared correlation coefficients among alleles at two adjacent SNP markers) and physical distances among these SNPs (Remington et al. 2001). The nonlinear model with r^2 as response and pairwise marker distance was fitted into the genome-wide and chromosome-wise LD data using the *nlin* function in R software. To ensure accuracy, LD was estimated using the unimputed dataset consisting of 194 322 SNPs, considering only markers with < 25% missing data and MAF > 0.05. Average pairwise distances in which LD decayed at r^2 = 0.2 and r^2 = 0.1 were then calculated based on the model, whereby the expected value (E) of r^2 was:

$$E(r^2) = \left[\frac{10 + C}{(2 + C)(11 + C)}\right]\left[1 + \frac{(3 + C)(12 + 12C + C^2)}{n(2 + C)(11 + C)}\right]$$

where:

r^2 = squared correlation coefficient

n = sample size

C = model coefficient for the distance variable (Hill and Weir 1988).

After removing monomorphic and low-quality SNPs, a total of 179 310 SNPs were available for the 591 member association panel. The MLM was used for testing individual SNP-based associations, following the correlation/trend method described by Weir (2008). The method was chosen for statistical analyses due to its ability to achieve consistently good control over false positives while yielding the highest power among other structure correction methods (Wang et al. 2012). The MLM is illustrated as:

$$Y = (SNP \times ß) + (PC \times \alpha) + K + \varepsilon$$

where:

Y = phenotypic response of the dependent variable (means or BLUPs of TCC and DMC)

SNP = SNP marker (fixed effect)

ß = fixed effect model coefficient for SNP

PC = first three components of PCA (which accounted for 97.6% variation) to present population structure (fixed effect)

α = fixed effect model coefficient for PC (population structure)

K = kinship matrix (random effect)

ε = residual.

Population structure was incorporated into this model to correct for their confounding effects on marker-trait associations, thus controlling false positives (Li et al. 2014). Evaluations of the association mapping model were based on the Q-Q plot, which is a plot of observed $-\log_{10}P$ values versus expected $-\log_{10}P$ values under the null hypothesis that there is no association between a SNP and the phenotype. The approach used for identifying GWA signals was based entirely on the smallest *P* values obtained from the MLM, where SNPs were ranked based on ascending order of their *P* values. SNPs with *P* values less than the 5% Bonferroni threshold were considered to be significantly associated with phenotypes.

Furthermore, chromosome-wise association signals were visualised from Manhattan plots generated using the *ggman* package of R software (Turner 2014). Because the statistical power in detecting QTL with smaller effects decreases in smaller association panels (Park et al. 2010), a genome-wide false discovery rate (FDR) of 10% was used to search for polymorphisms with relatively smaller effects (Gyorffy et al. 2005). FDR was computed using the GAPIT genome association tool (Lipka et al. 2012) in R.

Physical positions of GBS-SNPs were identified using the cassava genome sequence v6.1 (Goodstein et al. 2012; Prochnik et al. 2012). To overcome issues of non-collinearity that may exist within this genome version (given it is only partially complete), broader intervals (up to 0.5 Mb) were considered while searching for pathway genes in the vicinity of SNPs showing significant associations. The proportion of phenotypic variation explained by a combination of candidate SNPs was estimated by running multiple linear regressions for each trait using phenotype values as a response variable and candidate SNPs as predictors. This step was necessary because phenotypic differences between two homozygous classes of a given significant SNP for a trait are presented as an effect size, which is not corrected for by population structure or kinship.

5.3 Results

5.3.1 Phenotypic variability and correlations

Out of the 655 individuals initially considered for this study, 591 genotypes had both phenotypic and genotypic data and were used for subsequent statistical analyses. The average TCC was 4.9 µg g^{-1}, varying from 0-13.45 µg g^{-1}, which presented considerable variability for carotenoid content. RFC varied from 1 (white) to 6 (deep yellow). Distribution of the genotypes based on RFC showed 135 clones were white-fleshed, 37 light cream, 99 cream, 134 light yellow, 129 yellow and 57 deep yellow. DMC averaged 24.3%, varying from 10.7-31.2%.

ANOVA showed significant genotypic effects for both TCC and DMC among genotypes studied. Environmental effects were significant for DMC ($P \leq 0.01$) but non-significant for TCC (Table 5.2). Broad-sense heritability was higher for TCC (0.91) than for DMC (0.82). Phenotypic values and BLUPs of TCC showed a bimodal distribution (Figure 5.1a) while DMC was normally distributed around the mean value (Figure 5.1b). The Pearson correlation coefficient between TCC and DMC was negative and very significant (r = -0.38; $P \leq 0.01$) while RFC correlated positively with TCC (r = 0.84; $P \leq 0.01$).

Table 5.2 Analysis of variance of 591 cassava genotypes evaluated in two environments in Uganda

Source of variation	DF[a]	TCC[b]	DMC[c]
Environment (E)	1	10.09	45.57**
Genotype (G)	590	23.53***	21.26***
GEI[d]	247	1.87**	13.20***
Residual	766	1.29	11.74
CV[e] (%)		24.7	31.2
H^{2f}		0.91	0.82

[a]Degrees of freedom; [b]Total carotenoid content; [c]Dry matter content; [d]Genotype by environment interaction; [e]Coefficient of variation; [f]Heritability; **$P \leq 0.01$; ***$P \leq 0.001$.

5.3.2 Marker coverage and missing data

A total of 635 247 SNP markers were used to genotype the panel of 655 clones. Overall, there was a low SNP call rate, with an average proportion of missing data of 71%. The average SNP call rate per sample was 28.2%, varying from 1.3-81.4%. Following removal of monomorphic markers and imputation of the genotype subset, 179 310 SNPs

were selected and used for subsequent analyses. These SNPs were distributed along the 18 chromosomes and two linkage groups of unanchored scaffolds of the cassava genome sequence v6.1. A comparison of the imputed GBS data with results from the unprocessed GBS SNP data for the 591 member association panel showed a median discrepancy rate of 8.2% for all calls. When heterozygous calls were excluded, the median error rate reduced to 4.3%, which highlighted high accuracy of imputation. The imputed data were used to perform GWAS.

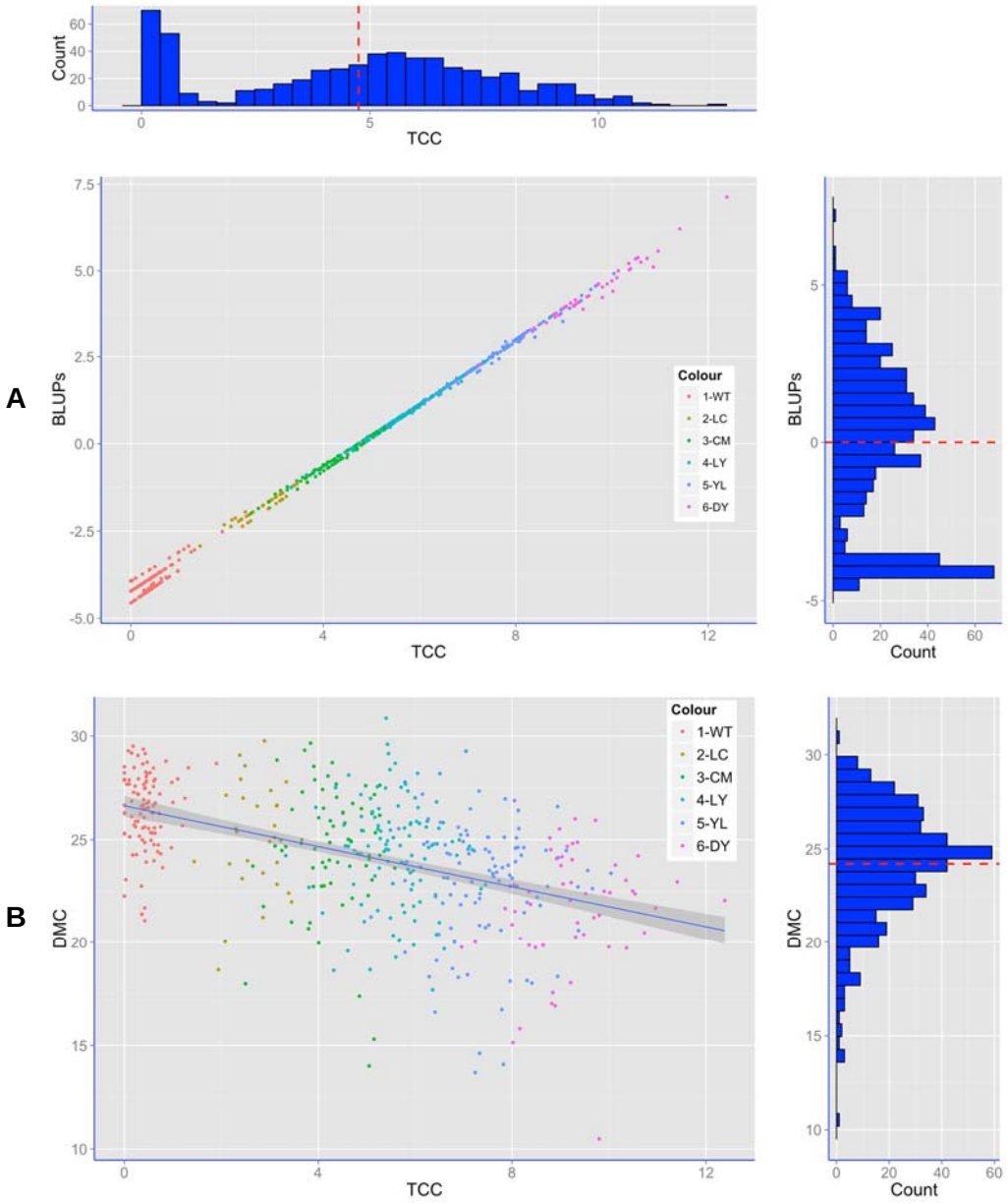


Figure 5.1 Scatter plot and histograms of total carotenoid content (TCC) vs. best linear unbiased predictions of TCC (A) and TCC vs. dry matter content (DMC) (B) for 591 cassava genotypes used for the genome-wide association study. Colour 1-WT = white; 2-LC = light cream; 3-CM = cream, 4-LY = light yellow; 5-YL = yellow; 6-DY = deep yellow; red dashed lines on the side histograms indicate means.

5.3.3 Population structure, allele frequency and linkage disequilibrium

Cassava genotypes used in this study were drawn from breeding populations developed using eight parental lines from IITA and CIAT, thus they presented a set of genotypes from diverse genetic backgrounds, which would be a possible source of population structure. PCA was used to describe the population structure for the association panel for this study. The first principal component explained 64.8% of genotypic variation while the second component accounted for 28.3%, cumulatively explaining 93.1% genotypic variation. Grouping of genotypes along the axis of the first two principal components scattered genotypes around the centre, without clear-cut differentiation between them. However, classifying genotypes according to their origin revealed some level of genetic differentiation, with the IITA group drifting to the left along the first principle component axis (Figure 5.2). This apparent genetic differentiation provided the basis for using both PCA and the kinship matrix in the MLM for association analysis, which was an important strategy to improve prediction accuracy by eliminating more false positives in the mapping result. Overall, 82% of the SNPs typed in the entire population had MAF < 0.01 (Figure 5.3), with more than 71% of the alleles shared between the CIAT and IITA accessions.

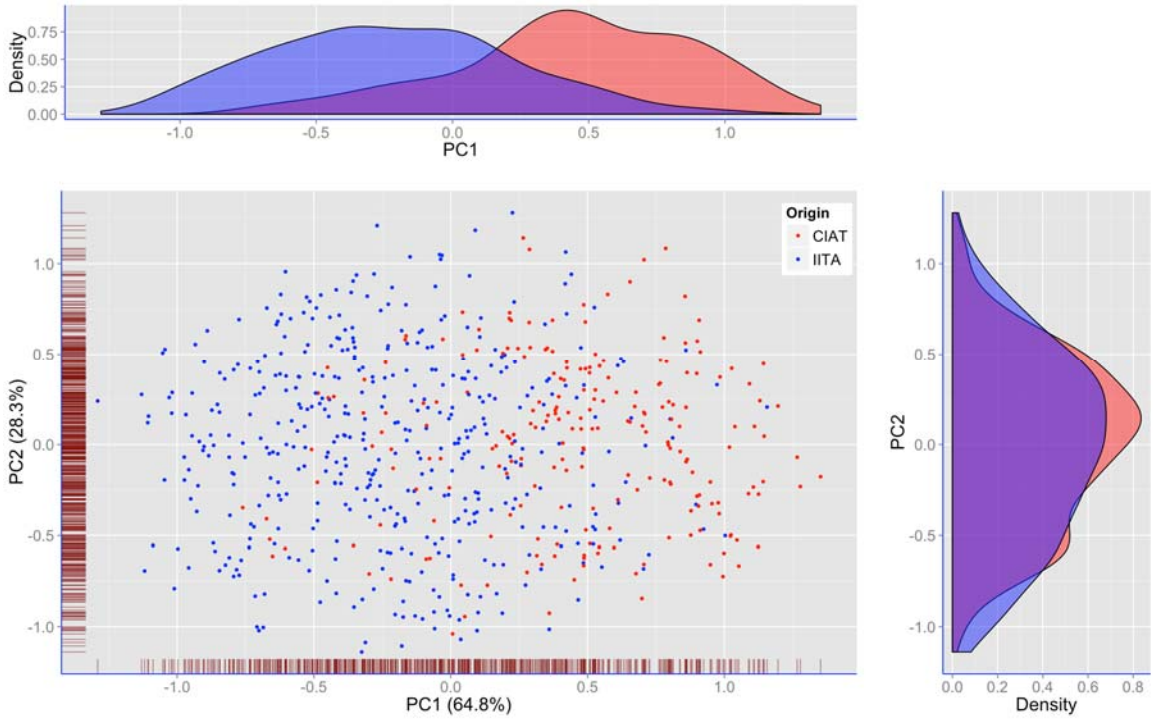


Figure 5.2 Scatter plot of 591 genotypes based on principle component analysis. PC1 = principle component 1; PC2 = principle component 2; CIAT = International Centre for Tropical Agriculture; IITA = International Institute for Tropical Agriculture.

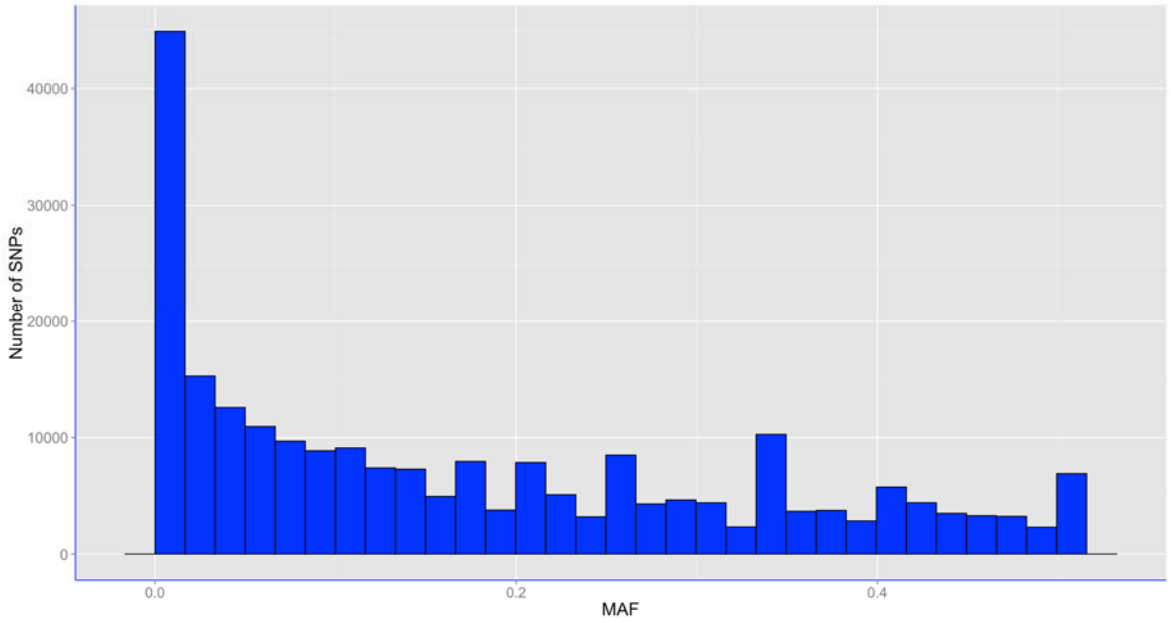


Figure 5.3 Distribution of minor allele frequency (MAF) based on unfiltered genotyping-by-sequencing data.

The average physical distance between pairs of markers was 23.4 kb and average genome-wide LD (r^2) obtained based on adjacent pairs of markers was 0.29. Genome-wide LD decay was 1.32 kb at r^2 = 0.2 and 3.21 kb at r^2 = 0.1 (Figure 5.4).

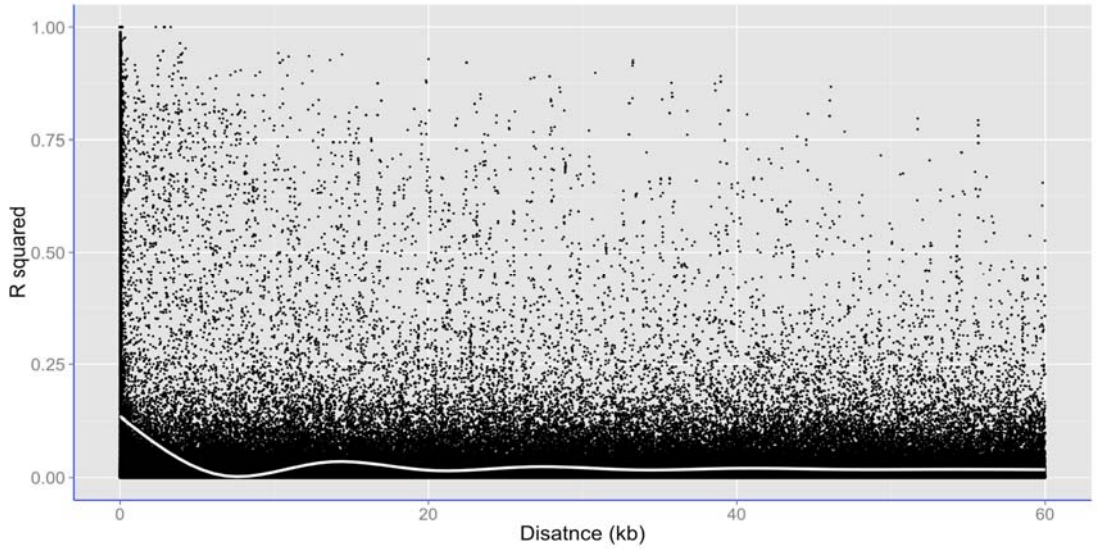


Figure 5.4 Plot of genome-wide linkage disequilibrium decay based on adjacent pairwise genetic and physical distance. Pairs with distance > 60 kb are not shown.

5.3.4 Association results

Two traits were considered for association analysis, TCC and DMC. However, all GWA for DMC showed non-significant marker-trait associations for the panel of SNPs used in this study. Therefore, only association results for TCC are presented in this section. Based on the MLM analysis for TCC, an association signal was revealed by significant deviation of observed *P* values of some SNPs from expected *P* values, based on means, BLUPs and RFC (Figure 5.5). When the Manhattan plot was used for chromosome-wise examination of association signals, one significant association peak was detected on chromosome 1 of the cassava genome sequence v6.1 (Figure 5.6). Within this peak, four SNPs had *P* values smaller than the 5% Bonferroni threshold for genome-wide significance. This association signal was consistent for analyses based on BLUPs and TCC means. This was the case with TCC means as phenotype.

When the association test was performed using RFC as a phenotype, an association hit similar to that obtained for BLUPs and TCC means was detected. This particular result corroborated the strong positive association between RFC and carotenoid content and Manhattan plots consistently reflected these results. In all these cases, the same set of SNPs was identified to show significant association signals. The four SNPs with *P* values below the 5% Bonferroni correction threshold were located within a stretch of 1.37 Mb on chromosome 1 (Figure 5.7).

Four SNPs identified under the 5% Bonferroni threshold were also detected when a 5% FDR criterion was used for GWA significance. Under a less conservative criterion of 10% FDR, additional three SNPs located in the vicinity of the most significant polymorphisms at the 5% Bonferroni threshold showed significant associations with TCC.

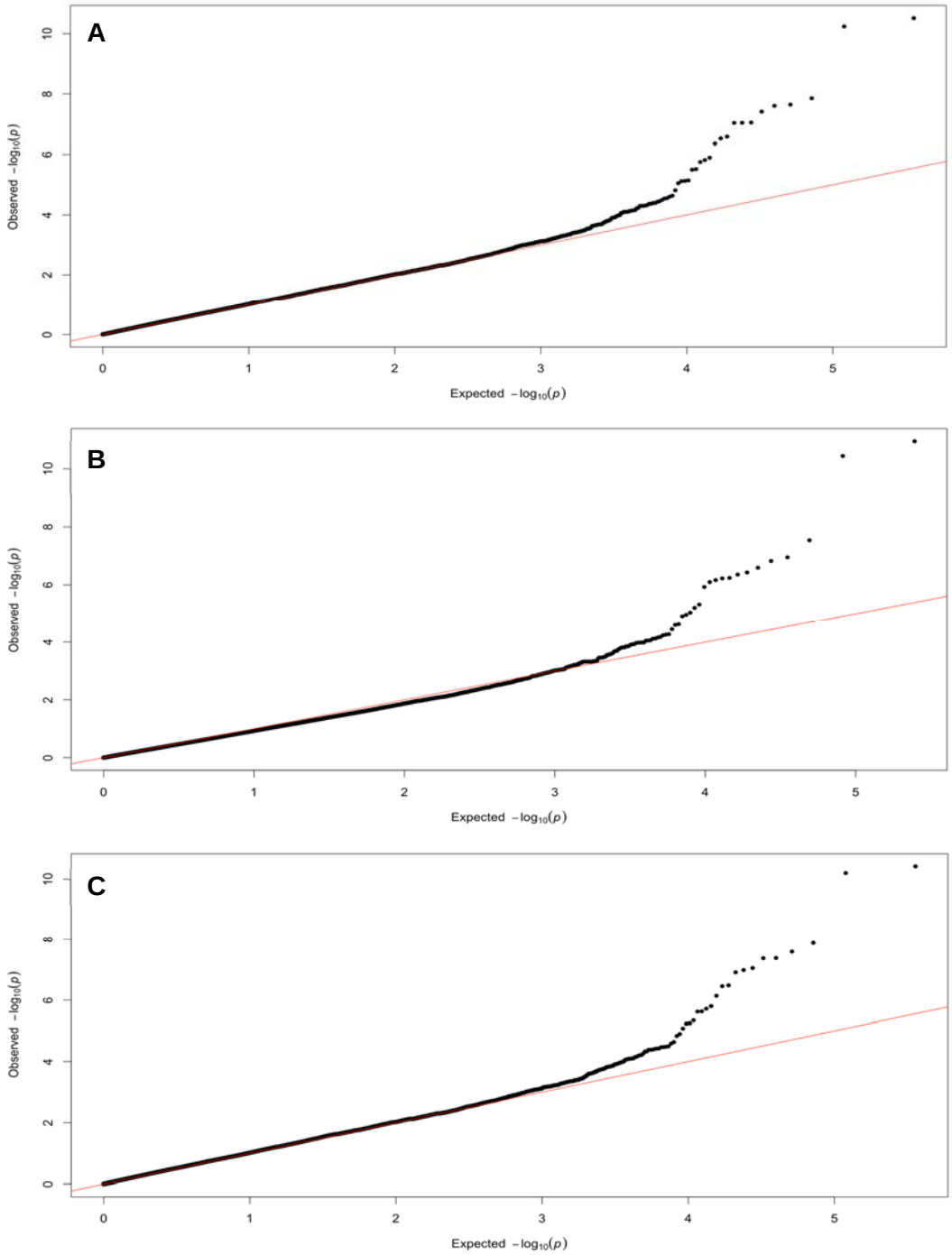


Figure 5.5 **Quantile-quantile plots for diagnosis of association signals based on best linear unbiased predictions for total carotenoid content (A), mean of total carotenoid content (B) and root flesh colour (C).**

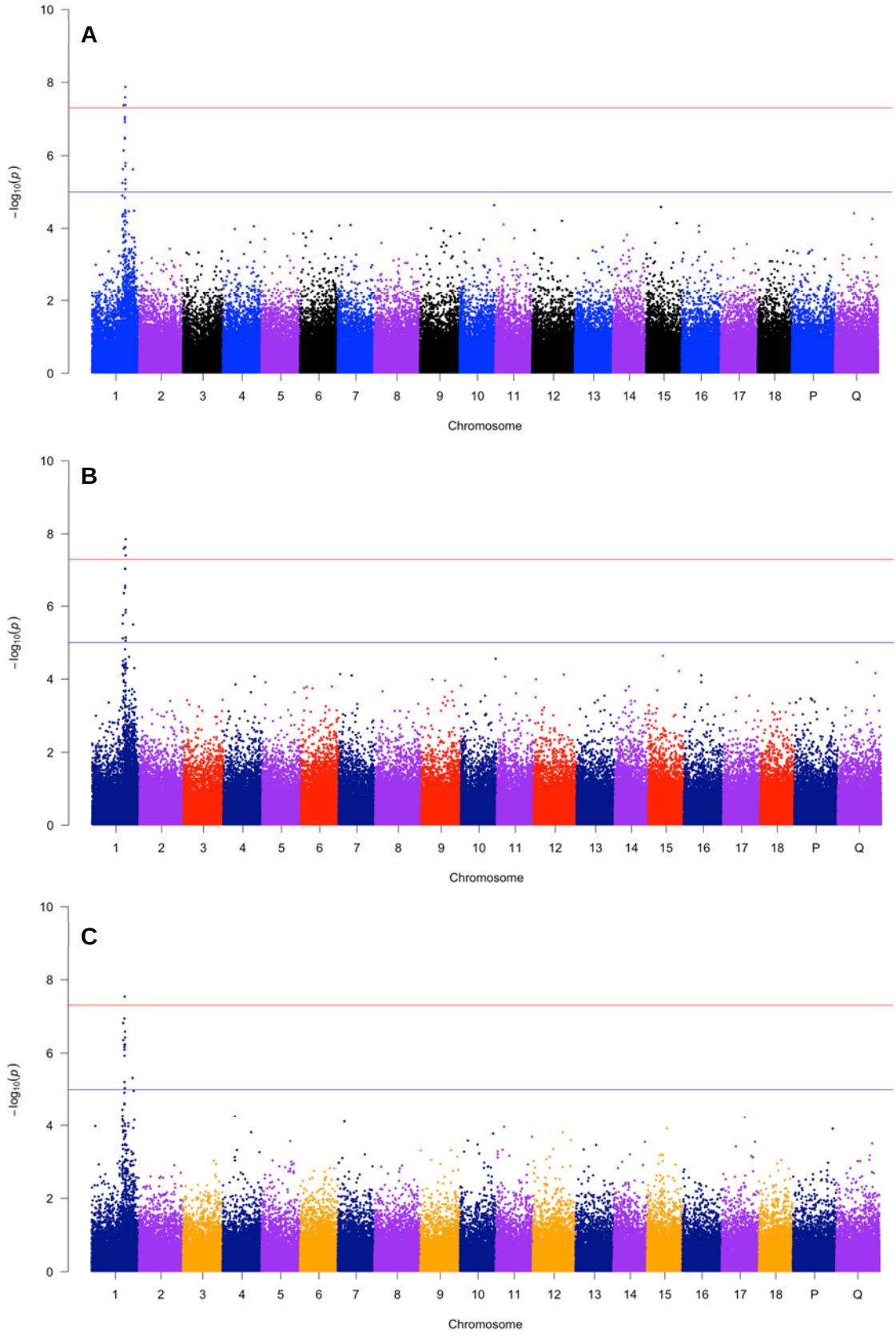


Figure 5.6 Manhattan plots for genome-wide diagnosis of association signals based on best linear unbiased predictions for total carotenoid content (A), means of total carotenoid content (B) and root flesh colour (C). P and Q are linkage groups of unanchored scaffolds in the cassava genome sequence v6.1. Horizontal blue and red lines are the suggestive and 5% Bonferroni threshold lines, respectively.

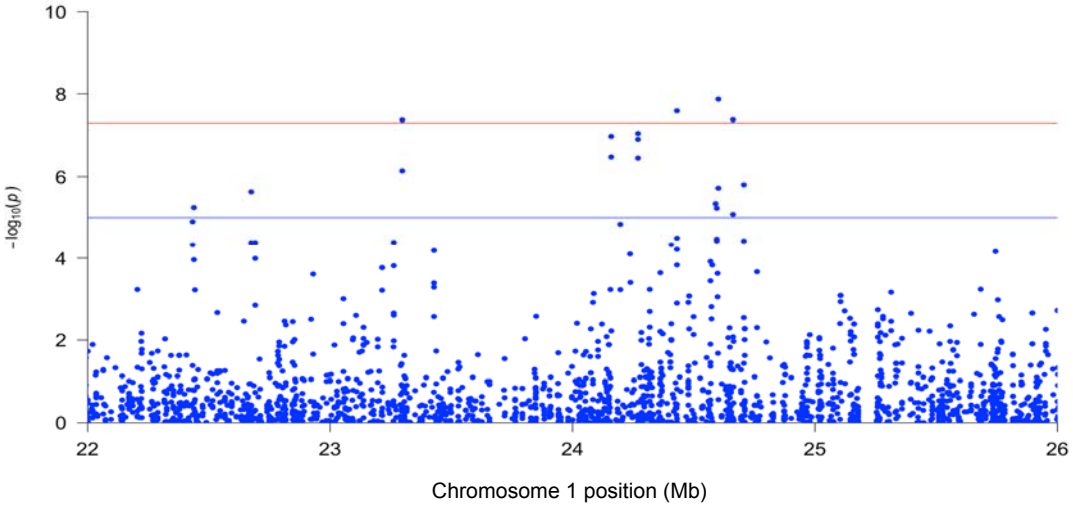


Figure 5.7 Manhattan plot of a portion of chromosome 1 with significant association signals for total carotenoid content. Horizontal blue and red lines are the suggestive and 5% Bonferroni threshold lines, respectively.

MLM analyses based on both BLUPs and TCC means indicated highest -$\log_{10}(P)$ for marker S1_24661940 (P = 4.4E-08) followed by marker S1_24654776 (P = 1.3E-08). These two SNPs accounted for the highest phenotypic variance (9%) for TCC and were separated by a distance of 0.01 Mb (Table 5.3). Marker S1_24270132 had the least significant P value (P = 1.6E-07), but was located closer to the best two SNPs than marker S1_23297833 that had a comparably higher P value (P = 6.2E-07). Together, these seven SNPs were located within a stretch of 1.37 Mb on chromosome 1 of the cassava genome sequence v6.1.

Table 5.3 List of SNPs with genome-wide association significance for total carotenoid content

Marker	Position[a] (Mb)	Allele		MAF[b]	*P* value	R^{2c}
		Major	Minor			
S1_24661940	24.66	T	C	0.32	4.4E-08	0.09
S1_24654776	24.65	C	G	0.34	1.3E-08	0.09
S1_24601429	24.60	C	A	0.32	8.3E-07	0.07
S1_23297833	23.29	G	A	0.23	6.2E-07	0.07
S1_24430613	24.43	G	A	0.45	5.5E-07	0.07
S1_24661956	24.66	T	C	0.32	5.1E-07	0.06
S1_24270132	24.27	T	C	0.48	1.6E-07	0.06

[a]Location of SNP on chromosome 1 of the cassava genome sequence v6.1: exact physical position of the SNP can be inferred from the marker's name, [b]Minor allele frequency; [c]Proportion of genetic trait variation explained by SNPs.

Reference was made to the cassava genome sequence v6.1 in *Phytozome v10.3* (Goodstein et al. 2012) to identify annotated genes within a distance of ± 0.5 Mb of the genomic region occupied by the significant SNPs. *Manes.01G124200.1*, which is a *PSY* gene known to increase accumulation of provitamin A carotenoids in cassava roots, was found within the genomic region occupied by significant SNPs. The genomic size of this gene is 3.3 kb stretching from 24 153 420-24 156 720 bp positions and falls within the 1.37 Mb portion occupied by association hit for TCC on chromosome 1. Functional annotation of this gene relates its biological activity to the enzyme PSY, which is strongly linked to carotenoid biosynthesis in plants (Goodstein et al. 2012). S1_23297833 and S1_24270132 were significant SNPs that flanked *Manes.01G124200.1* upstream and downstream, respectively. However, SNPs with the highest association significance for TCC were furthest downstream of the *PSY* locus. Other genes positioned within the actual locations of significant SNPs appeared to perform functions other than carotenoid biosynthesis (Table 5.4).

Table 5.4 Annotated genes within location of significant SNPs for total carotenoid content

Marker	Position[a] (Mb)	Loci[b]	Functional annotation[c]
S1_24661940	24.66	*Manes.01G131900.1*	N-Acetylglucosamine kinase
S1_24654776	24.65	*Manes.01G131700.1*	Protein binding
S1_24601429	24.60	None	-
S1_23297833	23.29	*Manes.01G112400.1*	None
S1_24430613	24.43	*Manes.01G128500.1*	Transcription regulation
S1_24661956	24.66	*Manes.01G131900.1*	N-Acetylglucosamine kinase
S1_24270132	24.27	*Maescv61001563m*	Protein phosphorylation

[a]Location of SNP on chromosome 1 of cassava genome sequence v6.1, [b]Name of characterised gene within the position of the significant SNP based on *Phytozome v10.3*, [c]Biological function of characterised genes found at the SNP location.

5.4 Discussion

Agricultural production across the world is being challenged by a rapid human population growth, limited arable land and adverse climate changes, which call for greater efforts to optimise the use of available resources needed to balance increased food production with reduced environmental effects. On their part, plant breeders could benefit from application of advanced genomic tools to discover and use functional genetic variations

for crop improvement (Moose and Mumm 2008). Diverse cassava genetic resources exist across the tropics for economically important traits (Nassar 2003; Nassar et al. 2007). Such useful genetic variation could be explored through GWAS to isolate and improve agriculturally important traits in a crop that is difficult to rapidly improve conventionally (Ceballos et al. 2015).

This study focused on ß-carotene, a trait of immense value for diets in SSA. Rapid and systematic genetic enhancement of cassava for provitamin A carotenoids is still limited. Thus, in this study, a GWA testing with 179 310 SNP markers and 591 cassava partial inbreds was conducted with the view of dissecting genetic causes of natural variation for carotenoid content and DMC in cassava. This study could provide a framework for undertaking more systematic genetic improvement of cassava for increasing levels of carotenoid content and DMC in roots.

It suffices to note that one serious consideration in defining an association panel is the extent of LD, which can greatly influence GWAS results because larger LD blocks and a slower rate of LD decay generally result in lower mapping resolutions (Grady et al. 2011). For example, Romay et al. (2013) observed more rapid LD decay in tropical germplasm of *Z. mays* compared to that in temperate germplasm, which means a higher resolution map would be achieved by using the tropical germplasm. In the current study, two broad accession groups (originating from CIAT and IITA) showed comparable LD decay, implying that similar mapping results could be achieved with one or a combination of these sets of genotypes.

Seven SNPs with GWA significance for TCC were identified in this study. These polymorphisms were located in the vicinity of a previously characterised gene with functions linked to carotenoid biosynthesis in cassava. Accumulation of provitamin A carotenoids in cassava roots has been linked to a *PSY* gene (Welsch et al. 2010), which is also a major regulator of carotenoid accumulation in rice (Beyer et al. 2002), maize (Fu et al. 2010) and durum wheat (Giuliano 2014). According to Welsch et al. (2010), a SNP in a *PSY* gene that is highly expressed in cassava roots co-segregated with high ß-carotene levels. This polymorphism resulted in a single amino acid change in a highly conserved region of the protein which, when tested in *Escherichia coli*, resulted in increased catalytic activity. Findings of the current study strongly relate to the reports on carotenoid genes in cassava by Welsch et al. (2010) by identifying the same genomic region controlling carotenoid content in cassava roots, but could also pinpoint additional genetic variants to further explain natural variation for carotenoid content in the crop.

GWA hits for TCC stretched across the region occupied by *Manes.01G124200.1,* but the most significant SNPs within the hit were located furthest from this carotenoid gene. This apparent clustering of the significant SNPs could be a manifestation of more than one QTL for TCC within the association region, suggesting the possibility of discovering additional genes controlling accumulation of carotenoid content in cassava. This also confirms the additive genetic effects for TCC observed through the diallel study presented in Chapter 3 of this thesis. Welsch et al. (2010) identified provitamin A carotenoid polymorphisms using a bi-parental mapping population of 164 individuals developed from two landrace genotypes from Brazil. An important feature of bi-parental mapping populations is that alleles segregating in progeny are limited to allelic diversity in the two parents (Semagn et al. 2010). In the current study, a panel of diverse germplasm of both African and Latin American origins was used to survey polymorphisms linked to TCC across the cassava genome. This approach may have presented an opportunity for discovering additional polymorphisms that could furthermore explain the extent of phenotypic variation for carotenoid content. It is possible that the diverse genetic material assayed in this study increased the frequency and chances of detecting novel genetic variants explaining the global phenotypic diversity for TCC (Korte and Farlow 2013).

When GBS is used to generate genotypic data for conducting GWAS, the possibility of finding SNPs linked to the phenotype is highly dependent on the genetic architecture of the trait, germplasm in which the trait is studied and number of SNPs used for genotyping (Poland and Rife 2012; Donato et al. 2013; He et al. 2014). Length and number of haplotypes detected can vary enormously, depending on the region of the genome and the germplasm group (Truong et al. 2012; Korte and Farlow 2013). For example, when some germplasm groups are under-represented in the dataset, a polymorphism that is not present at an appreciable frequency can pass the GBS pipeline quality filters due to reduced haplotype diversity (Beissinger et al. 2013). Romay et al. (2013) observed that the efficiency of GWAS would decrease remarkably when GBS is performed with less than 700 000 SNP markers in a highly diverse set of tropical maize germplasm.

The heterozygous nature and large genome size of cassava are important factors to consider when applying NGS technologies to study the crop (Oliveira et al. 2012; Ceballos et al. 2015). Version 6.1 of the cassava genome assembly is approximately 582.25 Mb in length with a total of 33 033 loci containing protein-coding transcripts arranged on 18 chromosomes and 2 001 scaffolds that have not been anchored onto

specific chromosomes (Goodstein et al. 2012; Prochnik et al. 2012). To achieve reasonable coverage of such a large genome through GBS, a large number of SNPs distributed throughout the genome would be required (Deschamps et al. 2012). In this study, 635 247 SNPs were used to genotype 591 clones. Unprocessed GBS data showed more than 70% missing data points. After filtering to remove non-informative SNPs, 179 310 markers were used for GWAS, which may represent low genome coverage and thus reduced the chances of capturing more polymorphisms with statistical significance. As observed in barley (Pasam et al. 2012), SNPs in association testing commonly explain a low percentage of genetic trait variation (> 4%) compared to QTL effects (as high as 47%) detected through bi-parental mapping (Abdel-Haleem et al. 2010), which essentially limits the ability of GWAS to detect association signals with small effects. Failure of SNP effects to explain a complete phenotypic variance may support the hypothesis that the expression of a quantitative trait is caused by a large number of very small effects that escape detection (Ingvarsson and Street 2011). Such scenarios may have accounted for the failure to detect association signals for DMC in the current study (Davey et al. 2011; Narum et al. 2013). Another reason, to some extent, for lack of a significant association signal for DMC could be the low variability observed for the trait in the association panel. Low genetic variability for a trait under association test reduces the ability of GWAS to capture allelic polymorphisms that explain only small variations in phenotypes (Korte and Farlow 2013).

Currently, the Buckler laboratory at Cornel University in the USA, with support from the Next Generation Cassava Breeding Project (www.nextgencassava.org), is refining genotyping and bioinformatics tools to facilitate more efficient genomic studies on cassava. Before long, such resources could make GWAS a more powerful tool for dissecting genetic architecture of important agronomic and quality traits of cassava that have been too complex to exploit through conventional breeding methods (Ceballos et al. 2012; Ceballos et al. 2015). To increase the power of GWAS for uncovering causative loci for quantitative traits such as DMC in cassava, phenotyping of traits will need to be refined such that they are scored more proximal to the underlying genetics (Ingvarsson and Street 2011). This strategy would potentially reduce the number of loci that contribute to the trait and thus increase the statistical power to detect causative variants (Yang et al. 2014).

5.5 Conclusion

This study was one of the pioneering efforts to apply NGS technologies to elucidate

genetics of cassava at molecular level and has proved GWAS as a powerful tool that can be used to unravel genetic factors linked to natural variations for important traits in the crop. The GWAS identified seven SNPs strongly linked to carotenoid content in cassava and located within the vicinity of *Manes.01G124200.1*, a gene known to increase accumulation of TCC in roots. These SNPs linked to TCC could provide a refined insight into genetic architecture of carotenoid content and offer practical basis for developing cassava varieties with high levels of provitamin A carotenoids through marker-assisted breeding or GS. Sequencing the genomic region containing the significant association signal would uncover useful sequence information and facilitate the development of genic markers to aid selection for TCC. On the other hand, the set of SNPs significantly associated with TCC could be considered as fixed effects when implementing GS for improving cassava for carotenoid content in genetic backgrounds of relevant agronomic traits. Collectively, these strategies would enhance the chances of developing cassava varieties with increased levels of provitamin A carotenoids, which, when deployed to farmers, could sustainably reduce VAD that is currently prevalent where the crop is a major staple. Therefore, follow-up studies are recommended for validating these SNPs associated with carotenoid content, particularly those located furthest from the previously characterised *Manes.01G124200.1* gene*,* to ascertain their biological roles in the carotenoid biosynthetic pathway and usefulness in cassava genetic improvement for carotenoid content.

5.6 References

Abdel-Haleem H, Bowman J, Giroux M, Kanazin V, Talbert H, Surber L, Blake T (2010) Quantitative trait loci of acid detergent fiber and grain chemical composition in hulled x hull-less barley. Euphytica 172:405-418

Akinwale MG, Aladesanmwa RD, Akinyele BO, Dixon AGO, Odiyi AC (2010) Inheritance of ß-carotene in cassava (*Manihot esculenta* Crantz). International Journal of Genetics and Molecular Biology 2:198-201

Beissinger TM, Hirsch CN, Sekhon RS, Foerster JM, Johnson JM, Muttoni G, Vaillancourt B, Buell CR, Kaeppler SM, de Leon N (2013) Marker density and read depth for genotyping populations using genotyping-by-sequencing. Genetics 193:1073-1081

Beyer P, Al-Babili S, Ye X, Lucca P, Schaub P, Welsch R, Potrykus I (2002) Golden rice: introducing the beta-carotene biosynthesis pathway into rice endosperm by genetic engineering to defeat vitamin A deficiency. Journal of Nutrition 132:506-510

Bishop DT, Williamson JA (1990) The power of identity-by-state linkage methods of analysis. American Journal of Human Genetics 46:254-265

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635

Browning BL, Browning SR (2013) Improving the accuracy and efficiency of identity by descent detection in population data. Genetics 194:459-471

Burns A, Gleadow R, Cliff J, Zacarias A, Cavagnaro T (2010) Cassava: the drought, war and famine crop in a changing world. Sustainability 2:3572-3607

Ceballos H, Kawuki RS, Gracen VE, Yencho GC, Hershey CH (2015) Conventional breeding, marker-assisted selection, genomic selection and inbreeding in clonally propagated crops: a case study for cassava. Theoretical and Applied Genetics 128:1647-1667

Ceballos H, Kulakow P, Hershey C (2012) Cassava breeding: current status, bottlenecks and the potential of biotechnology tools. Tropical Plant Biology 5:73-87

Ceballos H, Morante N, Sánchez T, Ortiz D, Aragón I, Chávez AL, Pizarro M, Calle F, Dufour D (2013) Rapid cycling recurrent selection for increased carotenoids content in cassava roots. Crop Science 53:23-42

Clewer AG, Scarisbrick DH (2001) Practical Statistics and Experimental Design for Plant Crop Science. John Wiley & Sons, Ltd, England

Crossa J, Beyene Y, Kassa S, Pérez P, Hickey JM, Chen C, Campos G, Burgueno J, Windhausen VS, Buckler E, Jannink JL, Lopez Cruz MA, Babu R (2013) Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3: Genes, Genomes, Genetics 3:1903-1926

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12:499-510

Dellaporta S, Wood J, Hicks J (1983) A plant DNA minipreparation: version II. Plant Molecular Biology Reporter 1:19-21

Deschamps S, Llaca V, May GD (2012) Genotyping-by-sequencing in plants. Biology 1:460-483

Donato M, Peters SO, Mitchell SE, Hussain T, Imumorin IG (2013) Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. PLoS One 8:e62137

Edgerton MD (2009) Increasing crop productivity to meet global needs for feed, food, and fuel. Plant Physiology 149:7-13

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379

Ersoz ES, Yu J, Buckler ES (2009) Applications of linkage disequilibrium and association mapping in maize. In: Kritz AL, Larkins BA (eds) Molecular Genetic Approaches to Maize Improvement, vol 63. Springer, Berlin, pp 173-190

Esuma W, Rubaihayo P, Pariyo A, Kawuki R, Wanjala B, Nzuki I, Harvey JJW, Baguma Y (2012) Genetic diversity of provitamin A cassava in Uganda. Journal of Plant Studies 1:60-71

Fu Z, Yan J, Zheng Y, Warburton ML, Crouch JH, Li JS (2010) Nucleotide diversity and molecular evolution of the *PSY1* gene in *Zea mays* compared to some other grass species. Theoretical and Applied Genetics 120:709-720

Gegios A, Amthor R, Maziya-Dixon B, Egesi C, Mallowa S, Nungo R, Gichuki S, Mbanaso A, Manary MJ (2010) Children consuming cassava as a staple food are at risk for inadequate zinc, iron, and vitamin A intake. Plant Foods for Human Nutrition 65:64-70

Giuliano G (2014) Plant carotenoids: genomics meets multi-gene engineering. Current Opinion in Plant Biology 19:111-117

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One 9:e90346

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research 40:1178-1186

Grady BJ, Torstenson ES, Ritchie MD (2011) The effects of linkage disequilibrium in large scale SNP datasets for MDR. BioData Mining 4:11

Gyorffy B, Gyorffy A, Tulassay Z (2005) The problem of multiple testing and its solutions for genome-wide studies. Orvosi Hetilap 146:559-563

He J, Zhao X, Laroche A, Lu ZX, Liu H, Li Z (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in Plant Science 5:484

Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. Theoretical Population Biology 33:54-78

Ingvarsson PK, Street NR (2011) Association genetics of complex traits in plants. New Phytologist 189:909-922

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

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:1-9

Li M, Liu X, Bradbury P, Yu J, Zhang Y, Todhunter RJ, Buckler ES, Zhang Z (2014) Enrichment of statistical power for genome-wide association studies. BMC Biology 12:73

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397-2399

Mayer JE, Pfeiffer WH, Beyer P (2008) Biofortified crops to alleviate micronutrient malnutrition. Current Opinion in Plant Biology 11:166-170

Montagnac AJ, Davis CR, Tanumuhardjo SA (2009) Nutritional value of cassava for use as a staple food and recent advances for improvement. Comprehensive Reviews in Food Science and Food Safety 8:181-194

Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiology 147:969-977

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. Molecular Ecology 22:2841-2847

Nassar NMA (2003) Cassava, *Manihot esculenta* Crantz genetic resources: anatomy of a diversity center. Genetics and Molecular Research 2:214-222

Nassar NMA, Schwartz CA, Junior OP (2007) Cassava diversity in Brazil: the case of carotenoid-rich landraces. Genetics and Molecular Research 6:116-121

Njoku DN, Vernon G, Egesi CN, Asante I, Offei SK, Okogbenin E, Kulakow P, Eke-okoro ON, Ceballos H (2011) Breeding for enhanced ß-carotene content in cassava: constraints and accomplishments. Journal of Crop Improvement 25:560-571

Njukwe E, Hanna R, Kirscht H, Araki S (2013) Farmers perception and criteria for cassava variety preference in Cameroon. African Study Monographs 34:221-234

Ojo SO, Ogunyemi AI (2014) Analysis of factors influencing the adoption of improved cassava production technology in Ekiti state, Nigeria. International Journal of Agricultural Sciences and Natural Resources 1:40-44

Okogbenin E, Porto MCM, Egesi C, Mba C, Espinosa E, Santos LG (2007) Marker-assisted introgression of resistance to cassava mosaic disease into Latin American germplasm for the genetic improvement of cassava in Africa. Crop Science 47:1895-1904

Oliveira EJ, Resende MDV, Silva Santos V, Ferreira CF, Oliveira GAF, Silva MS, Oliveira LA, Aguilar-Vildoso CI (2012) Genome-wide selection in cassava. Euphytica 187:263-276

Owens BF, Lipka AE, Magallanes-Lundback M, Tiede T, Diepenbrock CH, Kandianis CB, Kim E, Cepela J, Mateos-Hernandez M, Buell CR, Buckler ES, DellaPenna D, Gore MA, Rocheford T (2014) A foundation for provitamin A biofortification of maize: genome-wide association and genomic prediction models of carotenoid levels. Genetics 198:1699-1716

Owusu V, Donkor E (2012) Adoption of improved cassava varieties in Ghana. Agricultural Journal 7:146-151

Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, Chatterjee N (2010) Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nature Genetics 42:570-575

Pasam RK, Sharma R, Malosetti M, van Eeuwijk FA, Haseneyer G, Kilian B, Graner A (2012) Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC Plant Biology 12:16

Peterson G, Dong Y, Horbach C, Fu Y-B (2014) Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. Diversity 6:665-680

Pfeiffer WH, McClafferty B (2007) HarvestPlus: breeding crops for better nutrition. Crop Science 47:88-105

Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-Villeda H, Sorrells M, Jannink J-L (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. The Plant Genome Journal 5:103-113

Poland J, Rife T (2012) Genotyping-by-sequencing for plant breeding and genetics. The Plant Genome Journal 5:92-102

Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, Rodriguez F, Fauquet C, Tohme J, Harkins T, Rokhsar DS, Rounsley S (2012) The cassava genome: current progress, future directions. Tropical Plant Biology 5:88-94

R Development Core Team (2010) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Australia

Rabbi IY, Hamblin MT, Kumar PL, Gedil MA, Ikpan AS, Jannink JL, Kulakow PA (2014) High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. Virus Research 186:87-96

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

Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, Elshire RJ, Acharya CB, Mitchell SE, Flint-Garcia SA, McMullen MD, Holland JB, Buckler ES,

Gardner CA (2013) Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biology 14:55

Semagn K, Bjørnstad A, Xu Y (2010) The genetic disection of quantitative traits in crops. Electronic Journal of Biotechnology 13:1-45

Stephenson K, Amthor R, Mallowa S, Nungo R, Maziya-Dixon B, Gichuki S, Mbanaso A, Manary M (2010) Consuming cassava as a staple food places children 2-5 years old at risk for inadequate protein intake, an observational study in Kenya and Nigeria. Nutrition Journal 9:9

Suwarno WB, Pixley KV, Palacios-Rojas N, Kaeppler SM, Babu R (2015) Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. Theoretical and Applied Genetics 128:851-864

Swarts K, Li H, Romero Navarro JA, An D, Romay MC, Hearne S, Acharya C, Glaubitz JC, Mitchell S, Elshire RJ, Buckler ES, Bradbury PJ (2014) Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. The Plant Genome 7:1-12

Talsma EF, Melse-Boonstra A, de Kok BP, Mbera GN, Mwangi AM, Brouwer ID (2013) Biofortified cassava with pro-vitamin A is sensory and culturally acceptable for consumption by primary school children in Kenya. PLoS One 8:e73433

Truong HT, Ramos AM, Yalcin F, de Ruiter M, van der Poel HJ, Huvenaars KH, Hogers RC, van Enckevort LJ, Janssen A, van Orsouw NJ, van Eijk MJ (2012) Sequence-based genotyping for marker discovery and co-dominant scoring in germplasm and populations. PLoS One 7:e37565

Turner SD (2014) qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. http://dx.doi.org/10.1101/005165. Accessed 12 December 2014

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends in Biotechnology 27:522-530

Vazquez A, Bates D, Rosa G, Gianola D, Weigel K (2010) Technical note: An R package for fitting generalized linear mixed models in animal breeding. Journal of Animal Science 88:497-504

Wang M, Yan J, Zhao J, Song W, Zhang X, Xiao Y, Zheng Y (2012) Genome-wide association study (GWAS) of resistance to head smut in maize. Plant Science 196:125-131

Weir BS (2008) Linkage disequilibrium and association mapping. Annual Review of Genomics and Human Genetics 9:129-142

Welsch R, Arango J, Bar C, Salazar B, Al-Babili S, Beltran J, Chavarriaga P, Ceballos H, Tohme J, Beyer P (2010) Provitamin A accumulation in cassava (*Manihot*

esculenta) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. Plant Cell 22:3348-3356

Yang W, Guo Z, Huang C, Duan L, Chen G, Jiang N, Fang W, Feng H, Xie W, Lian X, Wang G, Luo Q, Zhang Q, Liu Q, Xiong L (2014) Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. Nature Communications 5:5087

Yuan-yuan L, Hong-jian T, Xiang-qian Z, Jie X, Geng-mi L, Shi-kai H, Guo-jun D, Zheng-yuan S, Li-wen W, Jiang H, Guo-you Y, Long-biao G (2014) Molecular improvement of grain weight and yield in rice by using *GW6* gene. Rice Science 21:127-132

CHAPTER 6

General conclusions and recommendations

Cassava's ability for resilience in adverse environmental conditions could give the crop an increased prominence compared to other staple crops and become a prime source of dietary energy across the tropical world. However, diets that heavily depend on starchy staples such as cassava are highly vulnerable to VAD. Such micronutrient deficiency problems are widespread phenomena in peasant families where nutritious foods in the form of animal products, fruits and vegetables become unaffordable or unrealistic to produce from their own farms. Severe VAD in children and women of reproductive age increases their susceptibility to preventable diseases such as measles and diarrhoea, impairing their physical development. Such afflicted individuals become a drain on socio-economic development of their own communities. In remote parts of developing countries including Uganda, traditional interventions to minimise VAD using supplementation and food fortification are less effective due to poor social infrastructures and high poverty levels. Overall, such situations demonstrate an urgent need to breed and deploy nutrient-rich crop varieties to afflicted communities, as a sustainable approach to improving nutrition.

The broader motivation for this study was the need to develop cassava genetic resources with enhanced levels of ß-carotene content in roots as a primary step towards deployment of provitamin A varieties to farmers in Uganda, with possible spill over effects targeting the entire eastern Africa. Specific objectives contributing to this aim were to (1) determine combining ability of provitamin A genotypes and types of gene action in inheritance of carotenoid content in cassava using a diallel mating scheme, (2) assess the influence of genotype by environment interaction on accumulation of carotenoid content in cassava roots and (3) identify genomic regions and polymorphisms linked to natural variation for DMC and carotenoid content in cassava, through a GWAS.

The diallel study was conducted to understand the heritability and nature of gene actions in expression of DMC and carotenoid content and identify genotypes most ideal as parental lines for subsequent use in improving these traits in cassava through hybridisation. Such information would be helpful for systematic implementation of breeding programmes using appropriate breeding methods. ANOVA, GCA:SCA ratio and percentage contribution of combining ability effects to total SS of genotypes indicated that GCA effects were larger than SCA effects for TCC, DMC and HI, suggesting the

importance of additive gene effects in controlling these traits. The implication of this finding for cassava breeding is that recurrent mass selection would be an appropriate method to increase genetic gains when targeting improvement of cassava for carotenoid content along with DMC and HI.

Average levels of TCC in three sets of genotypes evaluated in this study were below the 15 µg g^{-1} target set by the HarvestPlus research community, sufficient to cause a positive nutritional impact on consumers. It is possible that provitamin A cassava genotypes analysed in this study presented a small fraction of the global diversity for TCC. However, as selection and cyclic recombination continue to be imposed on these populations, mean TCC will increase through the exploitation of additive genetic effects. The relatively low genetic variability in germplasm sets evaluated in this study would also imply a need for introduction of additional genetic resources from CIAT and/or IITA for further recombination. An interesting approach in this case would be to introduce some of the provitamin A varieties already released to farmers in West Africa for breeding in Uganda. Such varieties probably already possess common farmer-preferred traits that could be introgressed into provitamin A genetic resources in Uganda. This approach could furthermore exploit additive genetic effects to increase accumulation of TCC in cassava roots, while generating recombinants with superior agronomic traits through specific cross combinations.

An important challenge to overcome during the development of carotene-rich cassava varieties is the negative correlation between TCC and DMC, which implies a compromise of one trait when selecting for increased levels of the other. This relationship is undesirable as it would hinder adoption of provitamin A varieties and overshadow efforts in breeding for higher levels of ß-carotene in cassava. It was evident that the white-fleshed genotype (NASE 3) was the best combiner for DMC, but progeny from this parent showed low levels of TCC. This scenario indicates that combining TCC and DMC by crossing carotene-rich lines with white-fleshed genotypes would not necessarily generate desired results. Currently, the genetic basis of this negative correlation is unknown.

However, the relationship appears to be non-significant for provitamin A populations at CIAT. Unpublished data from CIAT presented at two scientific meetings: the Plant and Animal Genomics Conference in San Diego, USA (January 2015) and the Next Generation Cassava Breeding Project Annual Review Meeting in Kampala (February 2015), indicated non-significant correlations between TCC and DMC which were similar

to results reported in 2014. It is possible that a longer period (>10 years) of cyclic selection undergone by Latin American germplasm led to breakage in potential genetic linkages between these traits, resulting in transgressive recombination. As the recurrent cycling approach requires several years to achieve meaningful results, cassava breeders in SSA may consider acquiring improved provitamin A germplasm from Latin America to increase prospects of generating varieties that would be acceptable to farmers. Nonetheless, some parents and families with good combining ability for TCC, DMC and FRW were identified from genotypes studied and could form the genetic basis for future breeding programmes targeting development of provitamin A cassava varieties.

An important output from this study is the set of 61 genotypes that were identified to combine high levels of TCC, DMC and FRW. These genotypes could constitute an invaluable genetic resource for possible identification of provitamin A cassava varieties, which could be fast tracked through on-farm evaluations for subsequent deployment to farmers in Uganda. Besides, further hybridisation could be undertaken with these genetic materials to increase genetic gains for these traits by exploiting the additive gene effects reported for TCC and DMC. As a regional centre of excellence for cassava research in eastern Africa, the cassava breeding programme in Uganda would share these improved genotypes with fellow breeders as a foundation for cassava biofortification across the region. Such are the spill-over effects from this study that could trigger further research to VAD through development and deployment of carotene-rich cassava varieties.

The GEI study showed (1) significant variation in genotypic effects for all traits studied, (2) non-significant environmental effects for TCC at 12 MAP, (3) significant GEI effects for all traits studied and (4) significant increase in TCC and DMC with crop age over the first 12 MAP. In a nutshell, results depicting differential performance of genotypes across environments underline the need to conduct multi-locational trials in order to identify genotypes with specific or general adaptation to target environments. The non-significant environmental effect for TCC is a useful piece of information for cassava breeders targeting development of provitamin A varieties. This suggests a possibility of using one location, such as on-station seedling or clonal trials, to effectively select genotypes with desired carotenoid content for subsequent multi-location screening and selection for other important low-heritability traits. In fact, for breeding trials with large population sizes, initial selection can be based on visual assessment of the root pigmentation, which strongly correlated with carotenoid content. This practice would offer the advantage of saving costs and time as quantification of carotenoids is expensive with regard to reagents required for biochemical analyses to assay large breeding populations.

Through GWAS, seven SNPs were identified that co-segregate with high levels of TCC in cassava roots. These SNPs were distributed across a chromosomal region of a previously characterised *PSY* gene known to increase accumulation of carotenoids in cassava roots. Carotenoid content in cassava has been suggested as a qualitative trait, which the current study corroborated by identifying a single genomic region associated with TCC, within the vicinity of a locus known to increase carotenoid synthesis in cassava. However, wide phenotypic variation for TCC across genotypes studied may suggest involvement of other alleles with minor effects in the genetic control of TCC in cassava.

In particular, SNPs discovered for TCC in this study could provide an extra dimension for further dissection of the genetic basis of provitamin A carotenoids in cassava, as they may reveal additional potentially useful genes. For example, most significant SNPs were located furthest from the known locus for carotenoid content in cassava. Characterisation of biological functions of each of these SNPs is necessary for the development of molecular markers for applying MAS for carotenoid content. Cassava breeders using GS could incorporate the set of SNPs significantly associated with TCC as fixed effects in the GS prediction models to select for high levels of carotenoid content in breeding populations. These approaches would leverage development of provitamin A varieties with higher levels of TCC so that smaller and attainable quantities of the root can be consumed daily to provide a beneficial level of provitamin A carotenoid. In this way, the impact of high provitamin A cassava intervention programmes could be broadened. Thus, genetic research that can ensure continual increase in levels of provitamin A in roots is needed, provided such increase does not negate other important agronomic and quality traits.

GWAS was also undertaken to search for genetic polymorphisms linked to DMC within the cassava genome in view of finding leads to elucidate the negative correlation between DMC and TCC. However, the GWA tests for DMC did not show any association signal of statistical significance. It is possible for such a result to arise from the fact that DMC is a quantitative trait controlled by many genes, each of which may have smaller effects, potentially reducing the statistical power of GWAS to detect their GWA significance. Future research efforts are warranted to better understand the molecular genetics of DMC and how it relates to inheritance of TCC in cassava, as these two traits would be inseparable in the context of cassava biofortification for carotenoid content.

This study was a pioneer effort towards a comprehensive hybridisation programme to

develop provitamin A cassava aiming to benefit the entire eastern Africa, as previous efforts largely focused on germplasm acquisition. The study was a component of broader research targeting development of cassava genetic stocks that combine high levels of provitamin A carotenoids with virus resistance and farmer-preferred traits. The improved genetic stocks would be shared with cassava breeding programmes of Ethiopia, Kenya and Tanzania that jointly implemented the East African Agricultural Productivity Project, with Uganda as a regional centre of excellence for cassava research. Selections from the diallel population, along with the best performing genotypes identified from the GEI study, could form the genetic basis for implementing breeding programmes to enhance carotenoid content in cassava in the region.

It is imperative to recognise CBSD as the current most serious limitation to genetic improvement of cassava in eastern Africa. The disease poses a realistic threat to both food and income security in the region as it causes corky necrosis within edible portions of roots, rendering them unpalatable and unmarketable. For cassava biofortification to yield meaningful results and create the desired nutritional impact, combining CBSD resistance with provitamin A carotenoids should be pursued as a matter of priority for breeders targeting development of carotene-rich cassava varieties for eastern Africa. Fortunately, cassava genotypes with high levels of CBSD resistance are being identified and shared among breeding programmes in the region, thanks to the Cassava Varieties and Clean Seed to Combat CBSD and CMD project (5CP) funded by the Bill and Melinda Gates Foundation. These varieties possess farmer-preferred traits such as high DMC high fresh root yield and could provide the source of CBSD resistance genes to combine with provitamin A carotenoids for generating varieties that will be acceptable to farmers.

SUMMARY

Global efforts are underway to develop staple crops with improved levels of provitamin A carotenoids to help combat dietary vitamin A deficiency (VAD), which has afflicted the health of resource-poor farmers in the developing world. As a staple crop for more than 500 million people in sub-Saharan Africa, cassava (*Manihot esculenta* Crantz) enriched with provitamin A carotenoids could have a widespread nutritional impact. In Uganda, cassava is the second most consumed crop after bananas. Accordingly, the national cassava programme initiated a breeding pipeline to develop cassava genetic resources combining high carotenoid content with farmer-preferred traits. To advance this initiative, a study was conducted with specific objectives to (1) determine the combining ability of provitamin A genotypes and gene actions in inheritance of carotenoid content and dry matter content (DMC) in cassava, (2) assess the effect of genotype by environment interaction (GEI) on accumulation of carotenoid content and DMC in cassava roots and (3) identify genomic regions and polymorphisms linked to natural variation for carotenoid content and DMC in cassava.

A 6x6 partial diallel analysis was conducted to estimate combining ability of six provitamin A clones and gene actions controlling carotenoid content and DMC in cassava roots. Fifteen F_1 families generated from diallel crosses were evaluated in two environments in Uganda. General combining ability (GCA) effects were significant for total carotenoid content (TCC) and DMC, suggesting the relative importance of additive gene effects in controlling these traits in cassava. Genotype MH02-073HS was the best general combiner for TCC and fresh root weight (FRW) while NASE 3 was the best general combiner for DMC. Such genotypes with desirable GCA effects could form the genetic basis for breeding programmes targeting improvement of cassava for TCC and DMC. Importantly, 64 genotypes were selected from the 15 F_1 families for advancement because of their superior performance for TCC, DMC and FRW.

When GEI effects for carotenoid content and DMC were assessed by evaluating 13 advanced provitamin A clones in six environments, additive main effect and multiplicative interaction analysis showed significant variation among genotypes for TCC, FRW and DMC. Environmental effects were non-significant for TCC, but GEI effects were significant for all these traits. From this effort, best performing genotypes were identified for TCC, DMC and FRW and could constitute the genetic material for on-farm testing or further recombination through systematic hybridisation programmes.

To survey genetic factors associated with carotenoid content in cassava, a high throughput genotyping-by-sequencing platform was used to genotype a panel of 591 genotypes using 635 247 single nucleotide polymorphisms (SNPs) distributed across the cassava genome. Through marker-trait association testing, a genomic region stretching 1.37 Mb on chromosome 1 was identified to contain seven SNPs significantly associated with carotenoid content in cassava. A survey of the cassava genome sequence v6.1 positioned these SNPs in the vicinity of *Manes.01G124200.1*, a gene known to increase carotenoid concentration in cassava roots. These findings demonstrated the usefulness of GWAS for dissecting the genetic architecture of traits in cassava and could offer practical grounds for improving carotenoid content in cassava through MAS and genomic selection.

Overall, a significant negative correlation was found between TCC and DMC, indicating the need for breeders to adopt methods that can combine high levels of these traits, given farmers prefer high DMC. Nonetheless, the study suggested an opportunity for cassava biofortification in eastern Africa. Improved genotypes were identified and genetic information generated, which could form the primary resources for systematic development of carotene-rich cassava varieties for deployment to farmers. Consumption of such nutritious varieties would sustainably reduce VAD among resource-poor peasants, for the benefit of their own physical and economic development.

Key words: biofortification, cassava, combining ability, genetic improvement, GWAS, provitamin A carotenoid, stability, vitamin A deficiency.

OPSOMMING

Globale pogings is onderweg om stapelgewasse te ontwikkel met verbeterde vlakke van provitamien A karotenoïede om te help met die stryd teen dieet gebasseerde vitamien A tekorte (VAD), wat die gesondheid van hulpbron arm boere in die ontwikkelende wêreld affekteer. As ‘n stapelgewas vir meer as 500 miljoen mense in sub-Sahara Afrika, kan cassava (*Manihot esculenta* Crantz), verreik met provitamien A karotenoïede, ‘n wye voedings impak hê. In Uganda is cassava die tweede mees belangrike stapelgewas na piesangs. Daarom het die nasionale cassava program ‘n telingspyplyn begin om cassava genetiese hulpbronne te ontwikkel wat hoë karotenoïed inhoud kombineer met eienskappe wat deur boere verkies word. Om hierdie inisiatief te bevorder, is hierdie studie uitgevoer met die spesifieke doel om (1) die kombineervermoë van provitamien A in genotipes, en geenaksies in oorwerwing van karotenoïed inhoud en droëmassa inhoud (DMC) te bepaal in cassava (2) om die effek van genotipe by omgewing interaksie (GEI) op die akkumulasie van karotenoïed inhoud en DMC in cassava wortels te bepaal en (3) om die genomiese streke en polimorfismes gekoppel aan natuurlike variasie vir karotenoïed inhoud en DMC in cassava te bepaal.

‘n 6x6 gedeeltelike dialleel analise is gedoen om die kombineervermoë van ses provitamien A klone en geenaksies van karotenoïed inhoud en DMC inhoud in cassava wortels te bepaal. Vyftien F_1 families, gegenereer van die dialleel kruisings, is geëvalueer in twee omgewings in Uganda. Algemene kombineervemoë (GCA) effekte was betekenisvol vir totale karotenoïed inhoud (TCC) en DMC, wat die relatiewe belangrikheid van additiewe geeneffekte aandui in die beheer van hierdie eienskappe in cassava. Genotipe MH02-073HS was die beste algemene kombineerder vir TCC en vars wortel gewig (FRW) terwyl NASE 3 die beste algemene kombineerder was vir DMC. Sulke genotipes met gunstige GCA effekte kan die genetiese basis vorm vir teelprogramme vir die verbetering van cassava vir TCC en DMC. Belangrik is dat 64 genotipes geselekteer is van die 15 F_1 families vir verdere ontwikkeling op grond van hulle goeie TCC, DMC en FRW eienskappe.

Die GEI effekte vir karotenoïed inhoud en DMC is bepaal met die evaluasie van 13 gevorderde provitamien A klone in ses omgewings. Additiewe hoofeffek en multiplikatiewe interaksie analise het betekenivollve variasie tussen genotipes getoon vir TCC, FRW en DMC. Omgewingseffekte was nie betekenisvol vir TCC nie, maar GEI effekte was betekenisvol vir al hierdie eienskappe. Uit hierdie analise is die beste genotipes

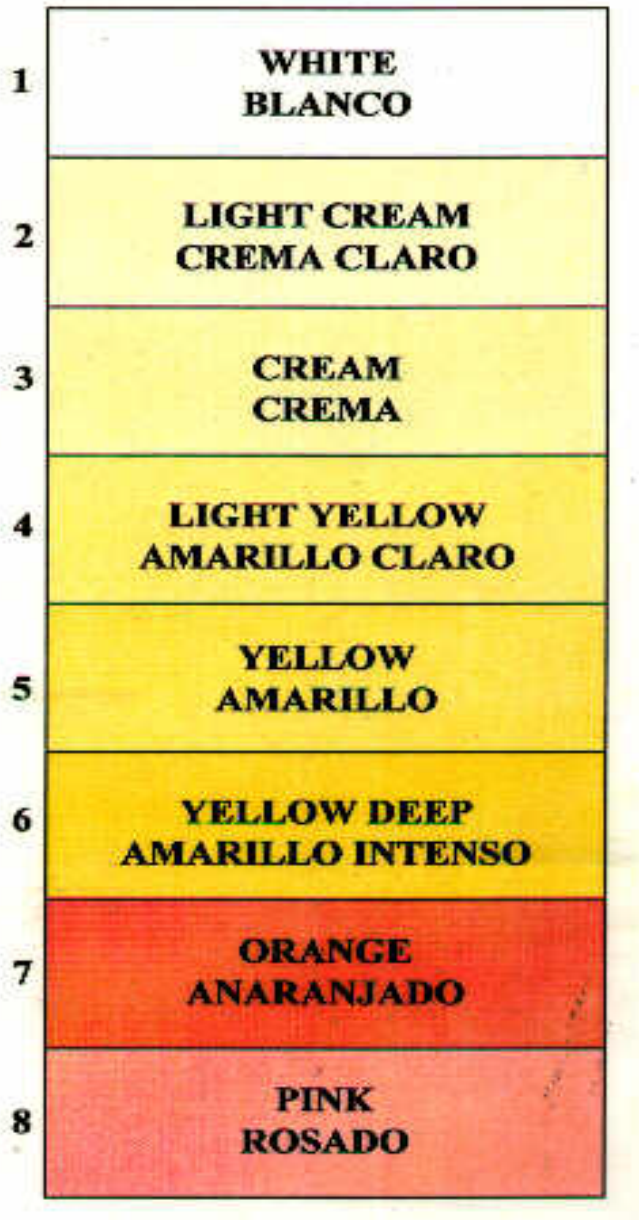
geïdentifiseer vir TCC, DMC en FRW en hulle kan die genetiese basis vorm vir evaluasie op die plase of vir verdere rekombinasie deur sistematiese hibridisasie programme.

Om die genetiese faktore geassosieer met karotenoïed inhoud in cassava te bepaal, is ‘n hoë deurvloei genotipering-by-volgordebepalings platform gebruik om ‘n paneel van 591 genotipes te genotipeer met 635 247 enkel nukleotied polimorfismes (SNPs) versprei oor die cassava genoom. Deur merker-eienskap assosiasie toetse, is ‘n genoom area oor 1.37 Mb op chromosoom 1 geïdentifiseer met sewe SNPs wat betekenisvol met karotenoïed inhoud in cassava geassosieer is. ‘n Opname van die cassava genoom volgorde v6.1 het hierdie SNPs in die omgewing van *Manes.01G124200.1*, ‘n geen bekend vir die verhoging van karotenoïed konsentrasie in cassava wortels, geplaas. Hierdie bevindings het die bruikbaarheid van GWAS bewys vir die bepaling van genetiese argitektuur van eienskappe in cassava en kan ‘n praktiese basis vorm vir die verbetering van karotenoïed inhoud in cassava deur MAS en genomiese seleksie.

In die algeheel gesien is ‘n negatiewe korrelasie gevind tussen TCC en DMC, wat die nodigheid wys vir telers om metodes te gebruik wat hoë vlakke van hierdie eienskappe kombineer, omdat boere hoë DMC verkies. Nietemin het die studie die moontlikheid van cassava biofortifikasie in oos Africa uitgewys. Verbeterde genotipes is geïndentifiseer en genetiese inligting is gegenereer wat die primêre hulpbron kan vorm vir sistematiese ontwikkeling van karotenoïed-ryk cassava variëteite vir verspreiding aan boere. Die inname van sulke voedingryke variëteite kan lei tot volhoubare vermindering van VAD onder hulpbron arm boere, wat tot die voordeel van hulle eie fisiese en ekonomiese ontwikkeling sal wees.

Sleutelwoorde: biofortifikasie, cassava, kombineervermoë, genetiese verbetering, GWAS, provitamien A karotenoïede, stabiliteit, vitamien A tekort

Appendix 1 Colour chart for visual assessment of carotenoid content based on pigmentation of root parenchyma



Source: Dr. Egesi Chiedozie, cassava breeder and Assistant Director, National Root Crops Research Institute, Umudike, Nigeria

Appendix 2 List of genotypes selected from 15 F_1 families from a 6x6 half diallel cross for advancement

Family	Genotype	TCC[a]	DMC[b]	FRW[c]
P1 x P2	UG10F39P040	5.5	38.8	2.3
P1 x P3	UG10F41P008	5.4	36.3	1.5
P1 x P4	UG10F40P024	5.2	30.8	2.0
P1 x P4	UG10F40P034	6.2	29.2	4.0
P2 x P3	UG10F16P023	4.9	32.8	2.5
P2 x P4	UG10F15P003	5.0	33.2	2.3
P2 x P4	UG10F15P004	5.5	39.2	1.3
P2 x P4	UG10F15P005	8.9	24.4	2.3
P2 x P4	UG10F15P006	6.2	30.1	1.9
P2 x P5	UG10F14P001	8.1	21.5	1.7
P2 x P5	UG10F14P003	4.4	33.2	2.1
P2 x P5	UG10F14P022	4.5	32.1	2.7
P2 x P5	UG10F14P030	6.1	32.0	5.0
P2 x P5	UG10F14P032	4.7	33.5	3.6
P2 x P5	UG10F14P033	4.0	35.1	1.2
P2 x P6	UG10F13P003	6.6	29.8	1.7
P2 x P6	UG10F13P030	6.1	28.9	3.7
P3 x P4	UG10F28P026	5.4	30.7	3.1
P3 x P4	UG10F28P050	7.6	30.0	1.4
P3 x P4	UG10F28P052	5.0	40.1	3.1
P3 x P4	UG10F38P012	3.9	33.8	4.9
P3 x P4	UG10F38P042	4.2	34.7	1.0
P3 x P4	UG10F38P048	3.7	38.5	3.5
P3 x P4	UG10F38P052	3.7	39.9	1.7
P3 x P4	UG10F38P053	9.0	32.9	4.7
P3 x P4	UG10F38P053	9.0	34.1	4.7
P3 x P5	UG10F26P016	3.7	31.6	3.1
P3 x P5	UG10F26P017	5.3	29.6	2.3
P3 x P5	UG10F26P033	5.9	32.5	4.3
P3 x P6	UG10F25P001	6.5	30.4	2.0
P3 x P6	UG10F25P008	5.5	32.5	1.7

[a]Total carotenoid content ($\mu g\ g^{-1}$); [b]Dry matter content (%); [c]Fresh root weight (kg $plant^{-1}$); P1, P2, P3, P4, P5 and P6 are parental genotypes defined in Table 3.1.

Appendix 2 List of genotypes selected from 15 F_1 families from a 6x6 half diallel cross for advancement *(continued)*

Family	Genotype	TCC[a]	DMC[b]	FRW[c]
P3 x P6	UG10F25P015	4.4	31.2	2.5
P3 x P6	UG10F25P016	3.6	34.0	2.5
P3 x P6	UG10F25P017	7.3	26.3	1.6
P3 x P6	UG10F25P022	7.4	28.1	1.7
P3 x P6	UG10F25P031	6.5	18.0	4.6
P4 x P5	UG10F20P004	5.6	33.4	2.1
P4 x P5	UG10F20P005	4.2	33.3	2.8
P4 x P5	UG10F20P007	7.2	31.3	2.9
P4 x P5	UG10F20P012	6.2	31.4	2.1
P4 x P5	UG10F20P013	6.4	39.2	3.2
P4 x P5	UG10F20P019	4.1	39.0	1.7
P4 x P5	UG10F20P022	4.5	34.2	1.4
P4 x P5	UG10F20P033	6.1	30.8	3.7
P4 x P5	UG10F20P037	4.9	34.5	1.3
P4 x P5	UG10F20P038	4.0	31.8	5.9
P4 x P5	UG10F20P042	5.1	38.4	3.4
P4 x P5	UG10F20P047	5.6	36.0	3.3
P4 x P6	UG10F19P014	6.7	27.7	2.7
P4 x P6	UG10F19P016	5.2	29.9	2.4
P4 x P6	UG10F19P017	4.0	30.9	3.0
P5 x P6	UG10F07P002	4.1	37.8	1.9
P5 x P6	UG10F07P005	6.8	32.4	4.5
P5 x P6	UG10F07P006	6.4	30.4	2.3
P5 x P6	UG10F07P007	6.5	30.7	2.7
P5 x P6	UG10F07P009	7.5	26.8	3.7
P5 x P6	UG10F07P012	6.2	30.3	2.9
P5 x P6	UG10F07P014	8.5	22.8	4.6
P5 x P6	UG10F07P015	6.2	28.7	3.8
P5 x P6	UG10F07P018	9.6	26.4	1.3
P5 x P6	UG10F07P018	8.7	27.6	1.3
Mean		5.8	31.9	2.7

[a]Total carotenoid content ($\mu g\ g^{-1}$); [b]Dry matter content (%); [c]Fresh root weight (kg $plant^{-1}$); P1, P2, P3, P4, P5 and P6 are parental genotypes defined in Table 3.1.