GENETIC BASIS OF TEXTURE AND ASSOCIATED TRAITS IN CASSAVA

BY

ENOCH WEMBABAZI

10704460

THIS THESIS IS SUBMITTED TO THE UNIVERSITY OF GHANA, LEGON IN PARTIAL FULFILMENT OF REQUIREMENT FOR THE AWARD OF PhD IN PLANT BREEDING DEGREE

WEST AFRICA CENTER FOR CROP IMPROVEMENT

COLLEGE OF BASIC AND APPLIED SCIENCE

SCHOOL OF AGRICULTURE

UNIVERSITY OF GHANA

November 2025

DECLARATION

I declare that this thesis is my own original work and has not been presented before to any university or institution for academic purposes leading to the award of a degree,

Signed.....
Enoch Wembabazi

(Student)

.....

Prof. Kwadwo Ofori

(Supervisor)

.....

Prof. Eric Yirenkyi Danquah

(Supervisor)

.....

Dr. Daniel Kwadjo Dzidzienyo
(Supervisor)

.....

Dr. Robert Kawuki Sezi

(Supervisor)

Abstract

Cassava breeding efforts in Uganda have in the past largely focused on improving agronomic traits such as yield and disease resistance but little has been done to improve consumer traits. This has limited adoption of improved varieties among consumers. In Uganda, cassava is predominantly consumed in boiled form and varieties that have a soft texture are preferred. To boost integration of texture traits in routine breeding operations, this study was conducted with specific objectives to (1) assess the phenotypic variability for texture traits (softness, toughness, stiffness) and water absorption (WAB) in a cycle 2 (C2) genomic selection diverse panel, (2) determine the prediction accuracy of Near Infrared Spectroscopy (NIRS) for phenotyping texture traits and WAB30 and (3) identify genomic loci and polymorphisms linked to variation for texture traits and further determine genomic prediction accuracy for these traits.

From study one, significant differences ($p<0.001$) were observed among 250 accessions for all texture traits and WAB30, though environment effects also significantly influenced ($p<0.001$) the variation observed across all traits. Broad sense heritability ranged from low (Soft_p; H^2 = 0.38) to moderate (Gradient; H^2 = 0.62) indicating substantial genetic control of these traits.

Near infrared spectroscopy (NIRS) quantitative predictions for texture traits and WAB30 based on partial least squares regression (PLSR) were all low. Using qualitative machine learning models based on binary classification with support vector machines (SVM), WAB had the highest prediction accuracy and reliability (r^2_{cv} = 0.78, κ_{cv} = 0.46) when spectra were pre- treated with standard normal variate (SNV) in combination with gap segment derivatization (GAP). This finding is promising for the integration of NIRS in routine phenotyping activities for consumer preferred cooking traits in cassava breeding.

Through marker-trait association mapping, 24 SNPs that explained a substantial proportion of phenotypic variation were found significantly associated with textural traits and WAB in boiled

cassava. Most SNPs were found on chromosomes 4, 8, 17, 18, and a survey of the cassava genome v7.1 positioned these SNPs in the vicinity of several genes coding for cell wall modifying proteins including *Manes.04G139200* (Soft_T; *Betagalactosidase* 1) and Manes.18G044401 (WAB30; *Glycine rich ell wall structural protein*).

Assessment of genomic prediction accuracy for texture traits and WAB30 found that all predictions across traits remained low ($r^2_{cv} \leq 0.27$). Also, the Bayes A, Bayesian Ridge Regression (BRR) and Reproducing Kernel Hilbert Spaces (RKHS) models generally gave better prediction accuracies than GBLUP model. These findings though promising, point at the need for optimal training population size and composition to achieve higher prediction accuracies.

Overall, this work demonstrated that tools for improving cassava for consumer preferred traits in Uganda are available and with improvements, could be directly integrated into breeding operations. This initiative is crucial for developing relevant cassava varieties combining end-user preferences and superior agronomic performances for the millions of actors in the cassava value chain in Uganda.

DEDICATION

I dedicate this work to my parents, my dear wife Florence, 'The Tribe' and my children (Eugene, Elsa and Ellah).

ACKNOWLEDGEMENT

I thank my God who has given me life and this opportunity to undertake this study and sustained me throughout the process. To Him be the glory, honor, and power forever.

I am very grateful to the Next Generation (NEXTGEN) cassava funding agencies, the Bill and Melinda Gates Foundation and the UK Department for International Development for the opportunity to pursue my Ph.D. I also thank the leadership of the RTBfoods project and RU-FORUM for funding specific aspects of this work and providing special training opportunities that helped me in completing this research.

I thank the leadership of the National Agricultural Research Organization (NARO) especially the Director General Dr. Yona Baguma for the tremendous support accorded to me in my career. Your words of wisdom and financial support have shaped my career. God bless you.

I thank my supervisory committee, Prof. Kwadwo Ofori, Prof. Eric Yirenkyi Danquah and Dr Daniel Dzidzienyo for the unwavering support rendered to me during this work. I particularly thank Dr. Kawuki Sezi Robert for the mentorship, patience, encouragement and technical input accorded to me in all seasons of this work. God bless you.

I thank the Director of the National Crops Resources Research Institute (NaCRRI) in Uganda, Dr Godfrey Asea for providing the facilities in which this research work was conducted, I also thank Dr Nuwamanya Ephraim (NaCRRI) for the all the support accorded to me in the lab. I am grateful to all the Nutrition lab technicians (NaCRRI), Ivan Lyatumi, Gaudensia Nakimera, Muhumuza Nicholas, Abubaker Sserubidde, Joseph Nsubuga, Betty Nalukwago, Martha Aciko and Grace Arayo for the tremendous contribution towards data collection for this work. God bless you all in your endeavors.

I specially thank my colleagues and mentors Dr. Williams Esuma, Dr. Angele Ibanda, Dr.

Ismail Kayondo, Dr. Alfred Ozimati, Dr. Paula Iragaba, Michael Kanaabi, Dr. Leah Nandudu, Fatumah Babirye and Florence Ndagire all of NaCRRI, Uganda for the technical and non-technical support and advice availed towards various aspects of this work. I also value the technical input and advise received from Assoc. Prof Jenna Hershberger (Clemson University, USA) and Dr. Fabrice Davrieux (CIRAD, FRANCE) on phenomics in this work. Thank you so much.

I also thank my cohort 12 colleagues at WACCI, Molly Allen, Jacinta Adoma Opoku, Keneth Obuobi Opare, Emmanuel Mrema, Faty Diaw, Georgina Lala Ehemba, Joseph Okpani Mbe, Simon Peter Abah, Abigail Asare Tweneboah, Rodrigue Ajibogoun, Cynthia Aghogho and Emmanuel Gilbert Omiat for great team spirit that kept me going.

I also thank the field technicians at NaCRRI, George kirumira, Asia Nabukalu, Samaly Nesigamye, Majara Charles, and Orone Joseph who tirelessly assisted in the experiment set ups, maintenance and harvesting. May God reward you abundantly.

I thank my parents Hon Justice Simon Byabakama Mugenyi and Mrs. Dorothy Kasaija Byabakama who started me off on this education path. May God bless you abundantly. I also thank the *Tribe* (Edgar, Justine, Edison, Edwin, Joy, Elijah, Fridah, Emmanuel) and our *Triblings* for spiritual and emotional support.

I thank my dear wife Florence Nakasujja for supporting, encouraging, and believing with me during the course of this work. God bless you Queen - size.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AGRA	Alliance for a Green Revolution in Africa
ANOVA	Analysis of Variance
AMMI	Additive main effects and multiplicative interaction
AYT	Advanced Yield Trial
Bayes A	Bayesian A
BLUP	Best Linear Unbiased Predictor
BRR	Bayesian Ridge Regression
BC	Baseline Correction
Ca^{2+}	Calcium ions
C2	Cycle two
CE	Clonal Evaluation
CET	Clonal Evaluation Trial
CGM	Cassava Green Mite
CD	Cassava Mosaic Disease
CBSD	Cassava Brown Streak Disease
CBSV	Cassava Brown Streak Virus
CIAT	International Centre for Tropical Agriculture
CIRAD	Centre de cooperation internationale en recherche agronomique pour le développement
CBLUP	Compressed Best Linear Unbiased Prediction
CMLM	Compressed Mixed Linear Model
Chr	Chromosome
CV	Coefficient of Variation

CWM	Cell wall material
DArT-seq	Diversity Array Technology sequencing
EA	East Africa
EMBRAPA	Brazilian Agricultural Research Corporation
EU	European Union
FRY	Fresh root yield
FHT	Fruit Hardness Tester
GBLUB	Genomic Best Linear Unbiased Prediction
GWAS	Genome Wide Association Analysis
GS	Genomic Selection
GDP	Gross Domestic Product
GEBV	Genomic Estimated Breeding Value
GCV	Genotypic Coefficient of Variation
GA	Genetic Advance
GAM	Genetic Advance as a Percentage of mean
GAP	Gap segment derivatization
G×E	Genotype by Environment
GEI	Genotype by Environment Interaction
GD	Genetic Diversity
H^2	Broad sense heritability
HCN	Hydrogen cyanide
HI	Harvest Index
H_o	Observed heterozygosity

IITA	International Institute of Tropical Agriculture
iPLS	Interval Partial Least Squares regression
κ_{cv}	kappa coefficient of cross validation
KASP	Kompetitive allele specific PCR markers
LD	Linkage Disequilibrium
LSD	Last Significant Differences
MGDI	multi-trait Genotype-Ideotype Distance Index
Max	Maximum
Min	Minimum
MAS	marker assisted selection
MPLS	Modified partial least squares regression
MSC	Multiple Scatter Correction
MAE	Mean Absolute Error
MAF	Minor allele frequency
MLM	Mixed linear model
NIRS	Near infrared spectroscopy
NARS	National Agriculture Research System
NaCRRI	National Crops Resources Research Institute
NCBI	National Center for Biotechnology Information
PPD	Post-harvest physiological deterioration
PLSR	Partial Least Squares Regression
PYT	Preliminary Yield Trial

PCV	Phenotypic Coefficient of Variation
PCA	Principal Component Analysis
PC	Principal Component
PEV	Prediction error variance
PVE	Phenotypic Variance Explained
PIC	Polymorphic Information Content
Q-Q	Quantile-Quantile plot
RDMC	Root Dry Matter Content
Root_no	Root number
RPD	Ratio of performance to deviation
RKHS	Reproducing kernel Hilbert Spaces
RTB	Roots Tubers and Bananas
RFLPs	Restriction Fragment Length Polymorphisms
RRBLUP	Ridge Regression Best Linear Unbiased Prediction
RMSE	Root mean square error
RBF	Radial Bias Function
RUFORUM	Regional Universities Forum for Capacity Building in Agriculture
Soft_p	Softness assessed by penetrometer
Soft_T	Softness assessed by texture analyzer
SVM	Support Vector Machine
SNV	Standard normal Variate
SNP	Single Nucleotide Polymorphism
SSA	Sub-Saharan Africa
SBLUP	Super Best Linear Unbiased Prediction

SCN	Soybean cyst nematode
SDS	Sudden death syndrome
SECV	Standard error of cross validation
SOV	Source of variation
SVG	Savitzky-Golay smoothing filter
TA	Texture analyzer
TE	Test population
TCC	Total carotenoid content
UYT	Uniform Yield Trial
WAB	Water absorption assessed after 30 minutes of boili

CHAPTER ONE

1.1 GENERAL INTRODUCTION

Cassava (*Manihot esculenta* Crantz) is a major source of affordable calories in Sub-Saharan Africa (Pérez *et al*., 2011., Adebayo, 2023). Production in Africa is concentrated in the tropics where the crop is a prized food security commodity, owing to its resilience to climate change and its ability to remain productive even in marginal environments (Ferris *et al*., 2002., Pradyawong *et al*., 2018). Lately, cassava use has expanded into several industries (Pradyawong *et al*., 2018) and is projected to become a critical stimulus for rural development and poverty alleviation in the tropics (FAO., 2018).

The growing demand for cassava starch and food-based products locally and globally is also incentivizing increased production among cassava farmers who are realizing increased incomes (Graffham *et al* 2017). For instance, export of cassava chips from Ghana and Uganda is reported to be steadily growing, providing these countries vital foreign exchange (Spencer and Ezedinma, 2017). Overall, global production has been steadily rising over the years (Supp. Figure 1) and is currently estimated at 277 million tons (FAO stat, 2022).

However, cassava production in Uganda has tremendously fluctuated since the early 2000s (Supp. Figure 2), mostly because of the re-emergence of cassava viral diseases, cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) (Waigumba *et al*., 2016).

Considering this challenge, cassava breeding has largely focused on addressing the two viral diseases and several varieties with resistance to CMD, varying tolerance levels to cassava brown streak disease (CBSD) and relatively stable yield have been developed and released to farmers (Kawuki *et al*., 2016). However, adoption of improved varieties among

farmers remains low and, in some instances, farmers have reverted to their indigenous cultivars citing undesirable culinary and processing qualities in the improved varieties (Nakabonge *et al*., 2018). This negative pattern arose out of concerted efforts to prioritize agronomic traits over quality traits during cassava breeding processes (Dufour *et al*., 2020).

Global led initiatives to address this anomaly rallied different institutions i.e. the International Institute of Tropical Agriculture (IITA), International Centre for Tropical Agriculture (CIAT, Colombia), Brazilian Agricultural Research Corporation (EMBRAPA), Centre de coopération internationale en recherche agronomique pour le développement (CIRAD, France) and the national programs in Uganda and Nigeria to mainstream end-user qualities in the cassava variety development process (Thiele *et al*., 2020).

This goal is achievable through conventional breeding that would make use of available genetic variability of cooking qualities in existing breeding populations, provided that reliable phenotyping methods are available. To this end, the Root Tubers and Banana (RTB) foods project (https://rtbfoods.cirad.fr/) in partnership with the NextGen project (https://www.nextgencassava.org/) among other goals focused on developing high throughput methods for assessing end-user cassava qualities including texture, cooking time and starch in order to equip breeders with effective selection tools (Dufour *et al*. 2020; Tran *et al*. 2020).

Through such efforts, cassava breeding in Uganda has also been sensitized on market segmentation (Dufour *et al*., 2020). Within this context, end-user defined traits such as taste, texture and starch content are being benchmarked for inclusion in selection indices to increase acceptability of improved cassava varieties among end-users in Uganda (Iragaba *et al*. 2020).

However, breeding efforts in this direction are still slowly bearing fruit, in part due to the phenotypic and genetic complexity of these traits and lack of appropriate methodologies for screening large breeding populations (Iragaba *et al.,* 2019).

Traits such as texture are difficult to phenotype and are associated with several physicochemical characters in the root such as starch, sugar content, cell wall properties and crude fiber (Franck *et al*., 2011). This has left breeders relying on subjective sensory score methods (Persley and Anthony, 2017) which are often reserved towards the final selection stages when the number of clones being evaluated has reduced (Iragaba *et al*., 2019). This phenotyping handicap needs to be addressed urgently to limit genetic erosion and speed up selections for texture and other consumer preferred traits in early selection stages.

More recently, Near Infrared Spectroscopy (NIRS) has been popularized as an accurate, reliable, fast and convenient phenotyping tool that cassava breeders could adopt for screening large populations for quality traits of interest (Sánchez *et al*., 2014; Davrieux *et al*., 2016).

NIRS has already been used for phenotyping dry matter, total carotenoids and starch content in fresh cassava roots and a high correlation (r^2=0.95) with reference methods has been achieved (Ikeogu *et al*., 2017; Mbanjo *et al*., 2022). Other studies have demonstrated the effectiveness of NIRS in predicting physicochemical constituents in cassava (Lebot *et al*., 2009). However, very little has been done to test NIR on sensory attributes such as texture in cassava.

It is also not clear what physiochemical properties influence sensory attributes in cassava and how they may be used routinely in breeding. In potatoes for example, NIRS has shown

good predictions for texture by relating it with dry matter using Partial Least Square Regression (PLSR) (van Dijk *et al*., 2002). This strategy could be exploited in cassava through the generation of texture reference and spectra data from a phenotypically diverse population for regression to determine the accuracy of prediction models (Balabin and Lomakina 2007). Developed models would be cross validated to estimate their reliability in comparison with an available reference method (García-Sánchez *et al.*, 2017).

Another aspect critical to improving end-user traits is a need to understand their genetic architecture, which would inform the cassava improvement strategy to deploy (Holland, 2007). Currently, there is paucity of information on the genetic mechanisms that regulate texture traits in cassava and how the environment influences expression of these traits. Consequently, this has limited the adoption of genomic based breeding strategies which could accelerate genetic gains for these traits in stable agronomic backgrounds.

As a result, cassava improvement for end user qualities is still lengthy and inefficient due to continued use of conventional breeding (Ceballos *et al*., 2020), which is often hindered by asynchronous or delayed flowering in potential breeding lines, long breeding cycles and poor seed set (Fathima *et al*., 2022).

Overall, concerted and strategic efforts are required to determine and utilize available genetic diversity of cassava to improve cooking traits in order to address consumer preferences in Uganda and to improve variety adoption. Such strategies require development of high throughput phenotyping methods to equip breeders with tools to enhance section for cooking traits at all selection stages. Improved phenotyping would also enable mapping of genetic loci linked to cooking traits, for wider adoption into genomic assisted breeding strategies

for accelerating genetic gains. The combination of these crop improvement tools would provide impetus for designing improved breeding pipelines that utilize advancements in phenomics and genomics to accelerate genetic gains for end-user preferences in cassava.

Objectives

General objective

To contribute to the development of cassava varieties that meet consumer preferences in Uganda.

Specific Objectives

(i) To determine the variability in texture and associated characteristics in cassava roots,

(ii) To develop and validate near infrared spectroscopy (NIRS) prediction models for texture and associated traits for cassava roots and

(iii) To identify genomic regions linked to texture in the cassava genome and determine genomic prediction accuracy for texture-related traits

CHAPTER TWO

Literature Review

2.1 Cassava production trends

The importance of cassava globally cannot be overstated. Cassava (*Manihot esculenta* Crantz) is a diploid (2n=36) highly heterozygous and non-inbred clonally propagated crop that is widely cultivated in the tropics for its starchy roots which provide affordable calories. Its foliage is also a nutrient dense vegetable that can be consumed by both humans and animals (Ceballos *et al*., 2017), (Okareh *et al*., 2021). The crop is also highly adaptable to several environments, making it relatively easy to cultivate even under stressed environments as evidenced in majority of Sub-Saharan Africa (SSA) that experiences an array of climatic conditions from heavy rainfall to sustained drought periods (de Oliveira *et al*., 2017).

Most cassava production is done by small holder farmers and Africa alone accounts for over 55% of the global production (Legg *et al*., 2014). Nigeria is ranked as the leading global producer, while Thailand is the leading exporter and China the leading importer of Cassava products. In Africa, Uganda is ranked 14th in production of cassava (Borku *et al*; 2015) with an estimated annual production of 4.2 million tons and average yield of 3.33MT/ha (FAO stat, 2024).

Cassava growing in Uganda is also predominantly done by small holder farmers with average parcel sizes in the range 0.4 – 0.8 hectares under rain fed system and harvesting is done from 6 to 12 months, though roots may even be kept in soil for up to two three years for some cultivars (Kilimo Trust, 2013). The Eastern region in Uganda tops production, followed by the North, Western and Central regions. Also, the area under cassava cultivation in

Uganda has steadily been rising from 405,000 hectares in 2003 to 1,262,121 hectares in 2020 (FAO stat, 2024), reflecting the growing importance of the crop in people's livelihoods.

Majority of cassava produced in Uganda is consumed directly by boiling, steaming, frying and roasting or processed first into flour and developed into a range of food products. Of these, consumption of fresh cassava roots predominates other end uses (Henry, *et al* 1998; Nanyonjo et al., 2019). Very little is advanced for industrial processing resulting in wastage due to the fast onset of post-harvest physiological deterioration (PPD) and saturation of the market with fresh cassava.

This is in part due to lack of processing facilities but also an inherent lack of required qualities for industrial processing among Ugandan varieties. This is confirmed by the large importation of corn starch by the country annually to feed pharmaceutical, paperboard and laundry industries. However, this narrative is slowly changing with the growing European Union (EU) and China market for cassava chips and starch (Ferris *et al*., 2002, Parmar *et al*., 2017).

Consequently, there is an appreciation of cassava as an industrial crop that has led to its bench marking in several industries including beverage, confectionary and ethanol industries (Ilukor *et al*., 2022). However, there is need to tap into the high-end industries such as pharmaceuticals and adhesives, which offer premium prices for starch. If achieved, such industries could encourage commercial cultivation of cassava by offering competitive prices for fresh or processed products, thus saving the country vital foreign exchange. However, whereas the economic potential of cassava is apparent, several constraints limit the realization of the full benefits associated with the crop.

2.2 Constraints to cassava production and the value chain in Uganda

Pests and Diseases

Production of cassava in Uganda was documented to grow from 3.42 million tons (MT) from 412,000 ha in 1990 to 5.5MT from 407,000ha 14 years later. However, this trend has since fluctuated due to pest and disease epidemics such cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) (Esuma *et al*., 2019).

CBSD majorly lowers the fresh weight and quality of the precious tuberous roots, thus making them inconsumable and unmarketable, and causing estimated financial loss in excess of US$ 1 billion in Africa annually (Amuge *et al*., 2017). As such, the diseases pose a threat to the food security situation in Africa, as it is observed to be spreading from East to West of the continent (Legg *et al*., 2015). In Uganda, CBSD has ravaged the crop in the major cassava growing areas of East and parts of Central where the disease pressure is highest, leading to estimated economic losses over US$ 30 million (Katono *et al*., 2015).

Breeding efforts in Uganda have subsequently targeted the introduction of resistance genes into local germplasm through conventional approaches, with tremendous success recorded for CMD resistance and considerable success achieved with CBSD tolerance (Kawuki *et al*, 2016). Genetic engineering has also been tested and demonstrated as an efficient tool for conferring resistance to cassava against CBSD successfully, though no variety has been released for production yet (Wagaba *et al*., 2017).

Other efforts are currently targeting reduction of whitefly populations which are the primary vectors for the CBSVs, though no variety has been released to this effect (Katono *et al*., 2023). In the long run, all these efforts are expected to provide lasting resistance against

the rapidly evolving cassava viruses and reduce their vectors, to conserve the cassava genetic resource in Uganda as well as ensuring food and income security for the populace.

Abiotic stress factors

Whereas cassava is generally resilient crop, drought and poor soil fertility are some of the major constraints to the crop's productivity in Uganda. Poor soil fertility has particularly been documented to reduce cassava production potential in Uganda by up to 6.7 tons/ha (Fermont *et al*., 2009). Drought on the other hand has been shown to negatively impact cassava yield as well as several growth characteristics (Brown *et al*., 2016; Fischer *et al*., 2019), which in turn directly affects the food security situation in some communities in Uganda which depend on the crop for daily nutritional needs.

Limited access to agro inputs and mechanization

Many small holder farmers in Uganda do not have access to vital inputs and information to help them augment cassava productivity. This is reflected in the consistent low usage of fertilizers, pesticides and limited use of mechanization in crop production (AGRA Uganda, 2018). Such challenges tend to restrict farmers to subsistence holdings which they can easily manage to provide food for the household but also limits industrial usage of cassava due to poor quality roots and general low productivity.

Low market absorption and low prices

Limited access to urban markets and industries mean that farmers get low prices for their produce due to limited bargaining power at farm gate. In part this may be attributed to poor quality cassava varieties (devoid of preferred qualities), poor handling, limited access to

processing facilities and misconceptions about cassava use in processed foods and other industrial products (Kleih *et al*, 2012).

Sometimes farmers grow local cultivars which have excellent culinary qualities but poor processing properties, hence cannot not meet standards of high-end industries such as pharmaceuticals, adhesives, drilling fluid, paper industries (Harry *et al*., 2016), (Naing *et al*., 2018). Such cultivars tend to fetch a low price or none at all depending on the intended end-use. This demoralizes farmers who choose not to prioritize cassava growing owing to its low contribution to household incomes.

Spoilage

Cassava is prone to post-harvest physiological deterioration (PPD) which sets in as early as 48 hours, causing reduction in quality and integrity of the root matrix. In Uganda, PPD is reported to be responsible for majority of the post-harvest losses incurred by farmers but especially retailers, causing economic loss in the excess of US$ 30 million (Waigumba *et al*., 2016). PPD in cassava has been attributed to the high moisture content, soft texture and high respiration rates all of which facilitate bacterial and fungal growth. This challenge majorly affects the fresh cassava market which provides the farmers with quick returns in the harvest season (USAID, 2017).

2.3 Constraints to cassava inclusion in industry in Uganda

Utilization of cassava in industry is incumbent upon the amount of starch therein and its properties. For instance, production of ethanol requires a variety with a high starch and sugar content (Nuwamanya *et al*., 2010) whereas highly digestible starch is important for making

glucose syrups. Other traits of industrial importance include paste viscosities, texture, starch granule size and morphology.

Whereas considerable progress has been made in characterizing several local and improved cultivars for their potential industrial use (Nuwamanya *et al*., 2010; Nuwamanya *et al*., 2011), to date many of these traits have not been accurately profiled in majority of cassava germplasm in Uganda, limiting their utilization in breeding and eventual application in industry.

Its therefore no wonder that Uganda continues to import large amounts of corn starch from neighboring countries owing to the high-quality specifications levied for pharmaceutical and food industries (Graffham *et al*., 2000). Also, prevailing perceptions about cassava starch tend to rank it lower than most cereal starches for application in industrial processes, thus limiting its adoption. Such perceptions may be persistent owing to the scanty documentation on Ugandan cultivars about their root quality and starch properties.

In addition, farmers in Uganda show preference for local cultivars for particular end uses (Tumuhimbise *et al*., 2012), yet these are at times low yielding, susceptible to pests and diseases and could also be unsuitable for processing.

However, through breeding, end use traits such as starch content and cooking qualities (Iragaba *et al*., 2021) can be introgressed and combined with key agronomic traits including yield and disease resistance. Notwithstanding, there would be need for adoption of high throughput phenotyping methods to aid rapid screening of lines under development to identify superior genotypes for advancement or subsequent recombination (Ceballos *et al*., 2021).

2.4 Breeding cassava root quality traits

Plant breeding usually involves three core interrelated activities including crossing, evaluation and selection (Barandica *et al*., 2016). Optimization of these processes guided by the breeder's equation are the foundation of plant breeding. Breeding often targets the development or identification of genotypes or families that express a target trait at a desired threshold. This may be achieved through generating new variability through recombination and imposing selection to identify superior types (Ceballos *et al*., 2004).

Typically, phenotypic recurrent selection is the breeding method of choice in cassava, but this takes a lot of time owing to the slow multiplication of cassava stems for multilocation evaluation (Ceballos *et al*., 2016). The entire process often starts with crossing among elite diploid progenitors to develop seed. The seed is then germinated in pots and a month later, are transplanted to the field to make a seedling nursery, whose size could get to 100,000 genotypes each with a unique combination of traits (Nassar and Ortiz, 2010).

Following maturity, the seedlings are evaluated for highly heritable traits such as yield and disease resistance which are relatively easy to measure, given the large population size of the seedling nursery. The average selection cycle may take 3 to 6 years from germination of seedlings to the establishment of multi-location trials and up to 8 years (Figure 1) before a variety is released (Ceballos *et al*., 2012). The additional years are usually required to screen promising genotypes in several environments prior to release of new varieties (Wolfe *et al*., 2017).

Phenotyping other traits such as root quality properties is often impractical at early selection stage gates owing to the large population size and lack of cost-effective methods for

screening (Ceballos *et al*., 2012). Besides, each genotype is normally represented by a single plant in seedling experiments, which may not give sufficient roots for evaluation. A clonal evaluation trial (CET) usually proceeds the seedling nursery and is associated with drastic reduction in genotype numbers to the few thousands following selection at seedling stage (Barandica *et al*., 2016).

At CET, each genotype is represented by about 10 plants in a row and there is usually no replication due to limitations in planting material and logistical challenges (Ceballos *et al*., 2004). Even with this, the trial remains quite large and impractical to screen for many root quality traits using conventional phenotyping methods.

However, advances in phenomics are increasingly making it possible to characterize root quality in minimal time, thus enabling high throughput data acquisition (Alamu *et al*., 2020a), (Hershberger *et al*., 2022). Conventionally, evaluation of mature plants is done after 12 months and selection is normally made based on fresh root yield (FRY), harvest index (HI) and disease scores, thus further reducing genotypes to be advanced (typically a few hundreds) (Ceballos *et al*., 2004).

The selections made are advanced to preliminary yield trials (PYT; Figure 1) where the plot size for each genotype tested increases with about 10 plants per plot and replication of each genotype is made possible allowing for estimation of more root quality traits due to abundance of materials and reduction of experimental error (Ceballos *et al*., 2004).

The selections from the PYT (50 – 80) are advanced to the advanced yield trial (AYT) which is also replicated and genotypes are evaluated on much bigger plots with several rows (~5)

for about two years with minimal selection pressure in multiple locations. Selection at this stage is done by evaluating about nine central plants within the plot while the border plants can be reserved for generating planting material (Ceballos *et al*., 2012).

This stage also presents opportunity for characterizing the selected genotypes for important cassava root quality traits in order to identify genotypes with unique combinations of traits that could be targeted for different end users, thus calling for rapid phenotyping approaches. The selections made are advanced to regional trials (RT) which are typically arranged the same way as AYTs but with fewer genotypes (20 – 25) and 5 – 8 commercial checks and these trials are usually conducted in several locations (5 – 10) for two years (Ceballos *et al*., 2012).

Thereafter, a cassava variety may be released to farmers depending on its performance relative to the checks. The presence of high throughput phenotyping platforms therefore would facilitate tracking genotype performance on several traits at every selection stage until variety release.

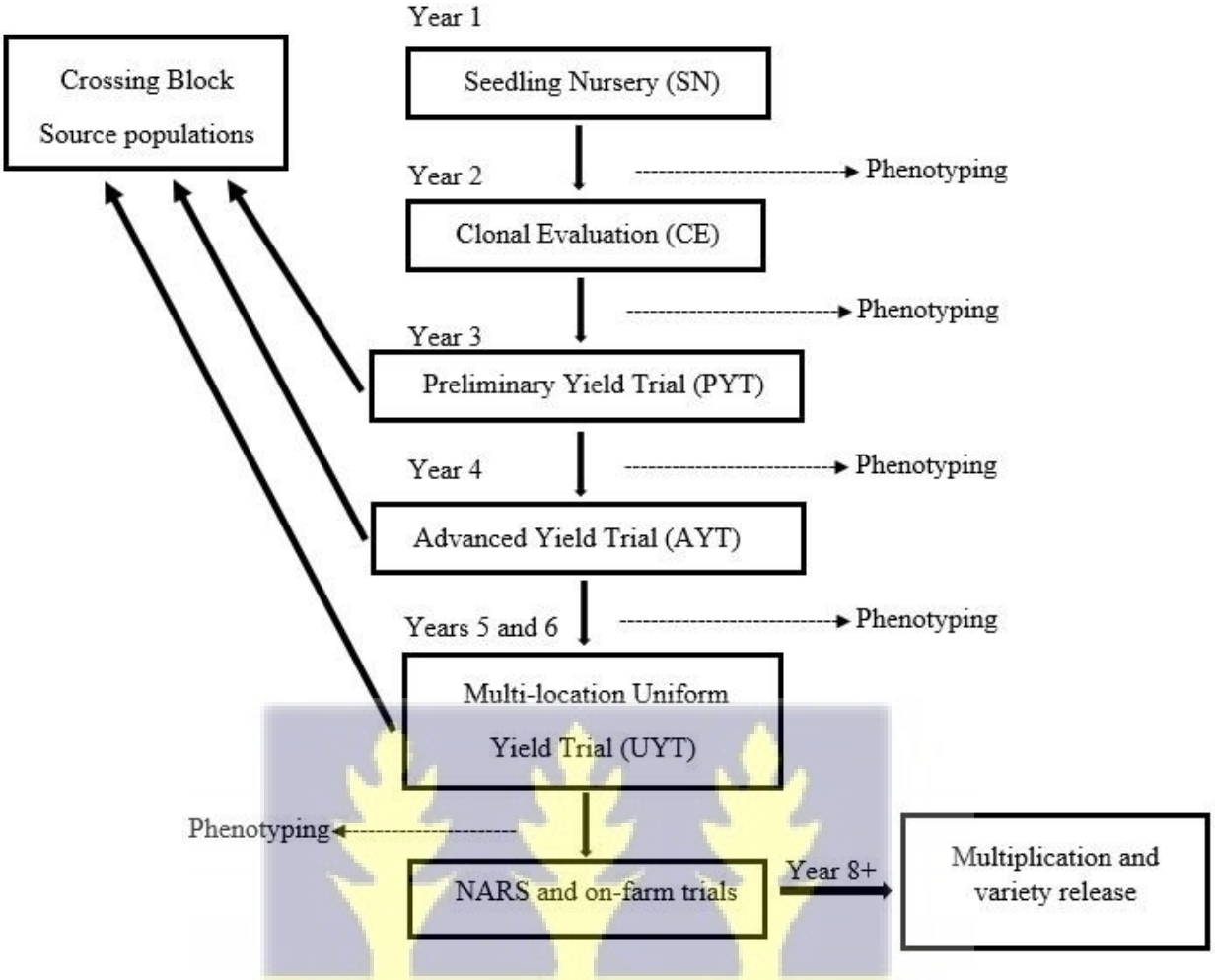


Figure 1: Schematic illustrating the selection stages in a variety development process, cyclic improvement scheme and the need for phenotyping at each stage. Modified from (Wolfe *et al*., 2017).

2.5 Marker Assisted Selection (MAS)

In this approach, molecular markers linked to genes controlling traits of interest through linkage disequilibrium are used to make selections, thus limiting reliance on phenotypic based selection which is lengthy and tedious (Ben-Ari and Lavi, 2012). MAS was initially thought to give plant breeders an advantage of selecting for traits that are genetically complex, by relying on the linked markers for early screening of genotypes, thereby reducing cycle time and minimizing costs (Xu and Crouch, 2008).

However, MAS has not been efficient in selection of such genetically complex traits which often have a large environmental influence (Hasan *et al*., 2021) reduced the length of the breeding cycle as had been anticipated since the selected genotypes still have to be evaluated on field and phenotyped to confirm expression of genes (Ceballos, *et al*., 2015).

It was also found that the quality of markers used in MAS has an impact on its efficiency for selection, such that certain marker types like single nucleotide polymorphisms (SNP) are more robust and widely distributed in the genome than restriction fragment length polymorphisms (RFLPs). The numbers of markers used also affect the reliability of the association with the targeted gene, in a way that large numbers are likely to identify a marker in close linkage to a gene of interest (Ceballos, *et al*., 2015).

In cassava, MAS has been used for selecting traits with relatively simple inheritance and controlled by few genes but not so with complex traits (Ceballos *et al*., 2015). Accordingly, progress has been made in using MAS to introgress cassava mosaic disease (CMD2) and cassava green mite (CGM) resistance genes into some lines in Latin America (Ferguson *et al*., 2012; Carmo *et al*., 2015) and the identification of progeny from self-pollinating cassava populations that have high inbreeding values for development of partial inbred lines (Oliveira *et al.*, 2018).

A few quality traits have also been selected through MAS including carotenes, dry matter content and cyanogenic glucosides (Guimarães *et al*., 2007). However, for some quantitatively inherited traits, MAS has been unable to clearly resolve quantitative trait loci (QTL) markers and the trait of interest due to the limited number of markers used (Ferguson *et al*., 2012). MAS tends to rely on few markers with large effects to draw association with traits of

interest and this limits its use for quantitative traits that are controlled by many genes with small effects (Robertsen *et al*., 2019). For these reasons, it has not been widely adopted in most breeding programs.

Also, false negative or false positive identification of markers especially when dealing with complex traits have further deterred its widescale adoption (Brumlop *et al*., 2010; Ben-Ari and Lavi, 2012). These may arise when biparental mapping populations that are not representative of the allelic diversity in a breeding program are adopted for quantitative trait locus (QTL) discovery. Moreover, such populations are often very costly to generate which may lead to compromise on population size, thereby causing loss of power. This would cascade into the necessity for validating discovered QTLs, which requires more effort (Jannink *et al*., 2010).

2.6 Genomic selection (GS)

As compared to MAS, marker effects across the entire genome are used to calculate the genetic estimated breeding value (GEBV) for a trait in an untested population (Meuwissen *et al*., 2013; Ceballos *et al*., 2015; Vivek et *al*., 2016). Typically, GS utilizes a training population (TR) of individuals that phenotyped and genotyped to model a prediction that converts genotype data from test population (TE) into GEBVs (Jannink *et al*., 2010). Therefore, GS is quite efficient for estimating polygenic effects on quantitative traits, thereby increasing selection efficiency.

In comparison to MAS, genomic selection does not rely on mapping populations, but rather marker effects are detected from training populations that are similar to breeding populations, and used to make associations with desired phenotypes (Wang et *al*., 2018). GS also

facilitates simultaneous improvement of several traits through a selection index, thus enabling accelerated genetic gains. This is particularly vital when demand led breeding is benchmarked to meet end-user preferences which might have different genetic complexities.

In cassava breeding, GS has been tested for predicting cassava brown streak disease (CBSD) resistance in an untested population resulting in prediction accuracies ranging 0.27 – 0.42 (Kayondo *et al*., 2018). Optimization of training population size has also been found to increase genomic prediction accuracy for foliar CBSD compared to random training populations, though the overall prediction accuracy remained low (Ozimati *et al*., 2018). It has also been suggested that seedling data may be used to predict clonal performance for CBSD root necrosis if the genetic correlation for CBSD incidence at seedling and clonal stages is high (Ozimati *et al*., 2019).

In another study, GS using 56 of the most informative SNPs showed high predictions for dry matter in a population of 358 cassava genotypes resulting in halving of the selection cycle (de Oliveira *et al*., 2012). Prediction accuracies for traits like dry matter and other more complex traits could also be improved by increasing training population (TR) size (Wolfe *et al*., 2017).

This has been re-echoed in a study on potato where predictions for starch content in potato were low (0.30 – 0.31) and a subsequent increase in TR size increased prediction accuracy (Sverrisdóttir *et al*., 2017). This is likely a result of the tandem increase in genetic diversity as the TR size increases, which enables accurate estimation of all genotypic effects that account for all phenotypic variation in a TR, leading to improved genomic prediction accuracy (Berro *et al*., 2019).

Apart from size, the composition of the training population has also been found to influence prediction accuracies since GS relies on linkage disequilibrium (LD; Pierre *et al*., 2023). The use of highly diverse training sets (TR) that are distinct from testing sets (TE) often leads to lower prediction accuracies whereas closely related TR and TE sets increase prediction accuracy (Habier *et al*., 2007; Berro *et al*., 2019). Therefore, TRs should be systematically constituted to maximize marker variance, reduce collinearity among markers and to capture genetic variability within a breeding population (Jannink *et al*., 2010).

2.6.1 Constituting a training population.

Generally, training population (TR) design is intended to minimize costs of phenotyping associated with the smaller selected numbers of individuals and to maximize prediction accuracy for the test set (TE). Composition of a TR may be influenced by the intended purpose of genomic selection, whether it is for rapid recurrent selection within a closed population, selection within a biparental population or selection within a germplasm collection (Lorenz and Nice, 2017).

Training population from segregating progenies of the same cross. In this TR, a portion of progenies from a biparental cross is selected, genotyped, phenotyped and used to train a model for predicting other progenies (TE) in the family. This population maximizes the linkage disequilibrium (LD) structure created in the initial hybridization to create an accurate model. This model may be effective in predicting future selection cycles generated by intermating of individuals within the family. Such within-family predictions are often accurate but very limited when across family predictions are made (Lorenz and Nice, 2017).

Training population including related and unrelated genotypes. In this case, the TR is constituted by collating multiple related and /or unrelated families. This may at times lower prediction accuracies if multiple unrelated families are collated leading to a poor relationship between the TR and TE. However, prediction accuracy could be improved by pooling families that share one common parent or by using high marker densities to improve sharing of information among unrelated families (Lorenz and Nice, 2017).

Training population from a diverse germplasm collection. In this scenario, the TR is selected from a diverse set of genotypes comprising a germplasm collection in such a way that it captures the entire diversity available in the collection. The rest of the genotypes can then form the TE. Like with other cases, the extent to which the TR is similar to the TE will influence prediction accuracy.

Several statistical criteria and optimizing algorithms may be used to select informative training populations to maximize prediction accuracy including stratified sampling, mean of the coefficient of determination (CD mean), mean of predictor error variance (PEV mean), stratified CD mean and random sampling (Isidro *et al*., 2015; Lorenz and Nice, 2017).

2.6.2 Phenotyping training populations

Well optimized and informative TRs reduce the phenotyping burden and costs thereof by retaining a smaller portion of the population for predictions. Even then, the quality of phenotyping done on a TR has a major influence on the outcome of prediction accuracy. The use of more replications and locations during evaluation of the TR often reduces phenotypic variance but increases heritability of traits, leading to better estimates of marker effects and better genomic selection outcomes (Lorenz and Nice, 2017). However, the number of

replications and locations to use should be optimized to minimize resource allocation towards phenotyping activities (Jannink *et al*., 2010).

2.6.3 Genotyping training and test populations

The ever-decreasing costs of genotyping have facilitated high density genotyping of large germplasm collections, thereby providing more information for determining the genetic merit for selecting superior genotypes (Wang *et al.*, 2020). Use of high-density marker sets generally increases coverage of the genome with markers, that facilitate capture of more genetic variability associated with a trait, thereby increasing reliability of GEBV estimates (DoVale *et al*., 2021).

2.6.4 Genomic prediction models

Apart from developing efficient genotyping and phenotyping methods, statistical methods also play an important role in genomic prediction accuracy. These are often grouped as either relationship based or marker-effect based models (Zhang *et al*., 2019). Relationship-based models estimate breeding values using only relationship information but do not utilize marker effects.

For instance, genomic best linear unbiased prediction of genomes (GBLUP) model uses a genomic information-based relationship matrix with more accurate relationship coefficients among genotypes to estimate the genetic merit of individuals (VanRaden, 2008). GBLUP has been used to predict cassava brown streak disease (CBSD) in cassava, giving considerable accuracy (r = 0.44) owing to its robustness and computational efficiency (Ozimati *et al*., 2018).

Marker effect-based methods on the other hand work by estimating the genetic marker effects and then accumulate those effects to generate breeding values for individuals (Zhang *et al*., 2019). For instance, Ridge Regression BLUP (RRBLUP) assumes uniform variance from all genetic markers while modelling their association with a phenotype.

However, broadly other studies group these models as parametric, semi-parametric and non-parametric models (Sahebalam *et al*., 2014; Crossa *et al*., 2014). Parametric models generally assume linearity between predictors (SNP markers) and response variables (phenotypes), which favors them for predicting traits majorly controlled by additive gene action. Semi-parametric models on the other hand are able to account for additive as well as non-additive genetic effects (Gianola *et al*., 2008) whereas non-parametric models generally account for non-additive genetic effects (González-Recio *et al*., 2008).

Parametric models including GBLUP and RRBLUP have been used severally to predict agronomic traits in cassava (Ozimati *et al*., 2018; Andrade *et al*., 2019) giving moderate accuracy, whereas no records show use of Super BLUP (SBLUP) and compressed BLUP (CBLUP) in cassava. Parametric methods of Bayesian nature (A, B, C, LASSO) have also showed promising prediction accuracy, albeit lower than GBLUP, for cassava agronomic traits with underlying assumptions of unequal contribution of each marker to total genetic variance (Esuma, *et al*., 2021; Ozimati *et al*., 2022; Phumichai *et al*., 2022).

Further, use of Bayesian methods for genomic high-density polygenic traits having diagnostic markers may increase the model's accuracy, with the latter being modeled as fixed effects (Stich and Van Inghelandt, 2018). Semi-parametric models such as Reducing Kernel Hilbert Spaces (RKHS) have also been used to predict agronomic qualities (yield, vigor, shoot

weight, CMD) in cassava with considerable accuracy (Wolfe *et al*., 2017; de Andrade *et al*., 2022), showing the importance of accounting for both additive and non-additive genetic variation during genomic predictions.

Non-parametric models on the other hand account for only non-additive genetic variation (dominance and epistasis) while estimating marker effects. Models such as random forests (RF) have also been used to predict cassava agronomic qualities with varying prediction accuracies (Kayondo *et al*., 2018; Phumichai *et al*., 2022).

2.6.5 Challenges associated with genomic selection.

Limited cost-effective phenotyping options. The difficulty and expense of measuring several economically important productive and adaptive traits remains a major limiter to the effective use of genomic selection to improve cassava (Ceballos *et al*., 2020; Mbanjo *et al*., 2021). To get accurate phenotype estimates, it is imperative that training populations (TR) get evaluated with replications and in multiple locations in the field. This however, requires a significant resource allocation which might not be accessible to most breeding programs in the developing world (Abincha *et al*., 2021). Cassava breeders also need to adopt high throughput phenotyping tools for assessing carefully selected and optimized TR sizes (Wolfe *et al*., 2017) for intrinsic root qualities that have end-user appeal.

Training population design. Cassava breeders still face a daunting task of collating optimal training populations (TR) that maximize prediction accuracy while minimizing costs of genotyping and phenotyping (Wolfe *et al*., 2017). Whereas large training TRs have been linked to increased genomic prediction accuracy, (Isidro *et al*., 2015; Ozimati *et al*., 2018),

they inadvertently accrue substantial costs which resource constrained breeding programs cannot afford (Lorenz and Nice, 2017).

Relatively high genotyping costs. Whereas significant progress has been achieved in reducing genotyping costs for cassava (Ceballos *et al*., 2020; Mbanjo *et al*., 2021), resource constrained breeding programs are often restricted to use of low-density marker sets which are affordable. Low-density markers partly cover the cassava genome, resulting in insufficient capture of genetic variance. Ultimately, this may translate into low genomic prediction accuracies for key traits in cassava. However, a number of studies show that low density marker sets can be optimized for a particular TR to achieve maximum prediction accuracies similar to those associated with high density marker panels (Chang *et al*., 2018).

With this background, it is evident that GS has not been tested for predicting end-user-linked traits like texture and cooking time in cassava, possibly due to the difficulty involved in phenotyping such traits, as well as scanty genomic association data available for them. Therefore, it is imperative to adopt high throughput phenotyping approaches to complement genomic selection approaches to accelerate genetic gains in cassava for end-user traits.

2.7 Genome wide association studies (GWAS)

Advances in genotyping coupled with lower costs thereof are increasing the efficiency of association studies on a broad range of genetic variants to map QTLS controlling economically important traits in crops (Mbanjo *et al*., 2021). Key to this approach is proper definition and measurement of the phenotypic segregation across diverse genotype sets to increase chances of precisely associating SNPs to a trait. Therefore, defining the allelic basis for such

variability has potential to provide ground for improving several quantitative traits in crops through marker assisted selection (MAS) and genomic selection (GS).

For instance, following a GWAS on a diverse carotene panel constituting 665 genotypes, (Esuma *et al*., 2016) found 4 SNPS highly associated with a genomic region on chromosome 1 of cassava, that was in proximity to the *Manes.01G124200.1* gene which is associated with carotene accumulation in cassava. Similarly, a major locus associated with dry matter content in cassava on chromosome 1 was found to be in physical linkage with the carotene locus in a diverse panel trough association of phenotypes and SNP markers (Rabbi *et al*., 2017).

Elsewhere, GWAS revealed SNPs significantly linked with CBSD foliar severity in cassava on chromosomes 4 and 11 and chromosomes 5, 11 and 18 for root severity, thus providing insight into the polygenic nature of CBSD resistance (Kayondo *et al*., 2018). However, most efforts have focused on uncovering the genetic basis of agronomic traits such as yield components (Zhang *et al*., 2018) as well as disease resistance which are directly linked to the economic value of the crop. Minimal effort has targeted the unravelling of the genetics underpinning several root quality traits that influence end user preference.

Moreover, demand for crops with good texture, taste, flavor, starch content and composition stems from consumer (Iragaba *et al*., 2020) and industry preferences for peculiar products with unique properties (Gebhardt *et al*., 2014). Consequently, as seen with other tuber crops, there has been minimal genetic gain for such traits in cassava, in part due to their phenotypic and genetic complexity (Ceballos, *et al*., 2015). It is therefore no wonder that improved cassava varieties in Uganda have at times been found deficient of end-user preferred traits (Nakabonge *et al*., 2018; Tsairidou *et al*., 2020).

Therefore, to augment the future selection of genotypes with improved tuber quality traits using genomic tools, there is need to adopt rapid but accurate phenotyping methods for complex traits: additionally, knowledge of gene number, chromosomal location and identity, and the magnitude of allelic effects on the traits of interest is required.

2.8 Root quality traits of consumer and industrial importance

Cassava genotypes can be differentiated based on color of their roots, physical properties, functional and physicochemical attributes of their flour and starches. Such differences are normally exploited by consumers and industries (Ayetigbo *et al*., 2018) for a variety of foods and environmentally benign products such as bioplastics, biopharmaceuticals, bioethanol, textiles and cosmetics among others (Vasconcelos *et al*., 2017; Yazid *et al*., 2018).

Additionally, the heterozygous nature of cassava often maintains high genetic variation across genotypes, thus providing opportunities for exploitation and selection. For instance, high amylose starch (> 30%), amylose-free starch and small granule starch have been found in several genotypes, showing the wide diversity for this trait and potential for utilization in industry (Rolland-Sabaté *et al*., 2012).

Further, the physico-chemical attributes in cassava may play a role in the expression of preferred traits including texture (Ahmed, 2023), taste and cooking time, which have been found to influence consumer's choice for specific varieties (Nakabonge *et al*., 2018; Iragaba *et al*., 2020).

However, accurate phenotyping of such proxy traits remains a challenge due to their inherent complexity that arises from their association with other root properties and lack of high throughput methods to characterize large breeding populations. Consequently, this has further presented an impediment towards integration of root quality proxy traits in routine selection activities.

2.8.1 Texture

Texture denotes the mechanical and surface characteristics of a product that are perceptible by senses of touch and feel (Costell and Durán, 2002). It is one of the key traits that influence acceptability of cassava products. For instance, cassava pastes and doughs are perceived to be more desirable when they are soft, sticky and elastic (Bechoff *et al*., 2018). Besides, texture also defines use of cassava in various industrial applications (Maieves *et al*., 2012).

Majority of cassava in East Africa is consumed in boiled state, and therefore farmers tend to cultivate varieties that meet their standards for texture (Franck *et al*., 2011). The boiling process of cassava roots is thought to cause hydration, gelatinization and cell wall modifications, which change the structure of the root, thus influencing end-product characteristics (Beléia *et al*, 2005).

Textural modifications may also result from varying storage conditions of the roots such as deep freezing (Bokanga, 1999). Also, the nature of genotype, maturity periods, environment in which cassava is grown, also influence the texture of the roots (Sajeev *et al*., 2010).

Further, cassava texture depends on cell arrangement and intercellular spaces that constitute the plant's tissues and their general structural organization. The parenchymatous tissue is an

example of this structural organization. It is rich in reserve substances such as cellulose and varies in cell form. These cells are generally polyhedral and have only primary cell walls (Damião, 1993).

More recently, texture of plant tissues has been attributed mainly to cell wall characteristics such as the nature and content of several pectin compounds, which are thought to play a role in cell separation when the root is cooked, thus affecting the softness of the roots (Maieves *et al*., 2012).

Softening of cassava roots is ideal for the development of industrial products (Ngea *et al.*, 2016) such as ethanol in which case roots are subjected to enzymatic hydrolysis prior to fermentation. Root dry matter content also plays a role in the final textural properties during processing (Franck *et al*., 2011). Current methods of screening cultivars for texture involve use of either sensory analysis or texture analyzer whose throughput is low, thus impractical to use when evaluating many clones, that are typical in early-stage breeding trials.

For rapid phenotyping, Iragaba *et al*. (2019) recently proposed the use of a Fruit Hardness Tester (penetrometer) to measure the force of penetration in boiled roots. The use of a penetrometer increased throughput for screening boiled roots for softness and was highly correlated ($r^2 = 0.91$) to ordinal scores from consumer testing.

However, sample preparation prior to score taking is still a hindrance to screening of large populations and the penetrometer is prone to human error accruing from long time use. This necessitates development of more efficient phenotyping approaches for this trait.

2.9 Plant phenotyping

Broadly, plant phenotyping entails comprehensive evaluation of plant characters such as leaf and root features, and the measurement of individual characters that form the basis of more complex traits (Li *et al*., 2014), as the plant interacts with its environment. Globally, advances in genotyping have far outpaced phenotyping capabilities, thus creating a serious bottleneck to genetic analyses for crop improvement (Shi *et al*., 2018). Attempts to curb this have among other technologies introduced robotic controlled imaging platforms (drones) that focus on capturing data above the ground, since the foliar are well exposed.

Whereas these platforms have largely been successful for routine agronomic monitoring such as growth, response to stress and plant architecture (Li *et al*., 2014), thus favoring crops whose productivity is measured above the ground, root and tuber crop quality has been adequately taken care of. Therefore, it is imperative that phenotyping is done across the different plant parts in order to augment crop improvement with data across the plant spectrum (Fahlgren *et al*., 2015).

Also, cost efficiency of the tool used should be considered to be relevant to plant breeders, since the majority of costs in phenotyping are associated with labor (Chawade *et al*., 2019). Equally important is the need to develop large data analysis and curation platforms that can extract vital information from the large volumes of raw data in timely manner to enable farmers and breeders to make timely decisions (Minervini *et al*., 2015).

2.10 Near Infrared Spectroscopy (NIRS)

Advances in material science and plant phenomics have found NIRS to offer high throughput analytics that can be employed for both laboratory and field set ups for plant phenotyping. In

comparison to lab wet chemistry analysis methods which are often costly, time consuming and destructive (Guo and Baianu, 2011), NIRS offers semi to non-destructive and environmentally benign means of evaluating plant characteristics. This tool is gaining wide recognition in analysis of materials in diverse domains including agriculture and pharmaceuticals (López *et al*., 2013; Sarin *et al*., 2019).

Near infrared essentially denotes a region of the electromagnetic spectrum covering wavelengths 780 – 2500 nm and has ability to interact with matter, causing differential vibration of molecules depending on their specific frequencies, thus creating unique spectral bands that contain information about the sample. This interaction follows Beer-Lambert's law, making it possible to relate band intensity to composition in the sample (López *et al*., 2013).

Most of the bands observed in NIR are majorly overtones and they vary in intensity depending on the composition of the material. The most intense overtones are usually associated with hydrogen atom binding to carbon, nitrogen and oxygen, commonly found in food constituents such as protein, lipids, carbohydrates and fiber.

For accurate data acquisition, NIRS must be calibrated for a number of components or elements in a sample, a process that normally involves spectra data acquisition followed by wet chemistry analysis of the same sample and subsequent calibration model development using regression methods such as partial least squares regression (PLS) and modified partial least square regression (MPLS; Alamu *et al*., 2020).

In plant physiology, NIRS has shown potential to track changes in a plant due to varying stresses such as disease intensity and drought, thus could provide rapid means of screening

stress response in crops. For instance, (Bajwa *et al*., 2017) used the visible (red, green and blue) and part of the NIR region to identify band combinations that showed plant disease condition.

A two-class discriminant model developed was able to detect 97% of the healthy plants but could not discriminate accurately between Soybean cyst nematode (SCN) and sudden death syndrome (SDS), two common foliar diseases in soy bean. Soil properties such as water holding capacity have been captured by visible region of NIR thus highlighting the versatility of the tool for both plant and soil sciences (Blaschek *et al*., 2019).

In cassava breeding, NIRS has potential to revolutionize the acquisition of root quality data through increased speed of analysis, flexibility for on-field data collection, minimization of error, ease of use, reduction of drudgery and reducing cost of phenotyping. For instance, (Ikeogu *et al*., 2017) demonstrated that a portable vis/NIRS could predict carotenoids (R^2_c = 96%, R^2_{cv} = 90%, RPD = 3.6 and SECV = 0.63) and dry matter content (R^2_c = 99%, R^2_{cv} = 95%, RPD = 4.5 and SECV = 0.)9 from fresh cassava root samples with relatively good precision using modified partial least square (MPLS) regression model.

However, it was observed that prediction accuracies for both traits increased when the root was homogenized than in whole state, though dry matter prediction was still comparable to the oven method for both sample types. However, extra sample processing such as homogenization increases time and cost of using the tool, which reduces throughput.

Emphasis should therefore be put on developing NIR spectra acquisition protocols that minimize sample preparation (Alamu *et al*., 2020). For instance, NIRs phenotyping throughput

would be increased tremendously in cassava if spectra were collected on whole fresh roots rather than homogenized samples (Ikeogu *et al*., 2017).

The cyanogenic potential of cassava is another key trait that influences varietal preferences. Consumers of boiled cassava in Uganda prefer cassava roots with low hydrogen cyanide content (HCN; not bitter) (Iragaba, *et al*., 2020b). Measuring HCN early in seedling and clonal populations usually presents a challenge due to the slow wet chemistry methods.

However, in recent studies, NIR quantitative and classification models showed high prediction accuracy ($R^2_{pq} \leq 0.86$; $R^2_{pc} \leq 0.99$) for distinguishing genotypes with high or low cyanogenic potential, which could be applied for early screening at clonal or seedling population level (Sánchez *et al*., 2014; Kanaabi *et al*., 2023). NIRs was also been used to predict multiple traits (dry matter, carotenoids and cyanide) with a single scan thus saving time, compared to single analyses-based phenotyping by wet chemistry methods.

NIRS has also accurately predicted starch, sugars (Lebot *et al*, 2013), fiber and protein in cassava and moderately predicted minerals (Lebot *et al*., 2009). For texture analysis, there is no evidence of use of NIRS for measuring this trait in cassava. Elsewhere, NIRS has been explored for texture analysis in sugar beet, though performance of predictions was low for both whole and sliced beets (r^2 values from 0.31 to 0.62).

In contrast, higher prediction accuracy (r^2_p of 0.85 - 0.97) for texture has been observed in taro deep fried chips using PLSR, with the maximum force for break of the chips taken as the reference value (Su and Sun, 2019).

For steam boiled potatoes, (Boeriu *et al*., 1998) showed that moist, waxy, firm characters were best predicted using a quantitative model based on PLS by relating NIR spectra data to sensory scores. Consequently, regression coefficients (R^2_p) ranged between 0.89 and 0.94. This study also noted the high resolving capacity of PLS regression in extracting information from a broad- spectrum range rather than relying on peak height or area and relating it to composition of the sample. This would result in development of reproducible and accurate calibration models.

Likewise, (van Dijk *et al*., 2002) also demonstrated the prediction of potato texture based on a number of texture predictors such as mealiness, firmness and crumbliness. It was established that dry matter was highly correlated with these descriptors: henceforth a PLSR model was developed based on dry matter to predict the texture. Also, relatively strong correlations for moisture - M ($R^2 = 0.85$), mealy ($R^2 = 0.79$), crumbly ($R^2 = 0.79$), waxy - ($R^2 = 0.77$), grainy- ($R^2 = 0.73$), mashable - ($R^2 = 0.70$), and firm- ($R^2 = 0.68$) were found between the NIR spectra and the descriptors.

The study concluded that the dry matter content was a key determinant for texture in steam boiled potatoes. This approach could be extended to cassava to gain understanding of the different physicochemical traits in the root that are linked to texture in cassava, for possible adoption as proxies for NIR based phenotyping.

The literature discussed underlines the progress made in cassava breeding, the limitations to improved variety adoption and the tools that breeders need to access to address such limitations. Overall, for breeders to select for end-user preferences as part of routine selection processes, they need to be equipped with substantial genetic diversity onto which selection

can be imposed, accurate high throughput phenotyping tools to quicken the selection process and knowledge of genetic mechanisms that regulate these traits to inform selection strategy. A combination of these tools would facilitate full exploitation of available genetic diversity to develop relevant cassava varieties.

CHAPTER THREE

Assessing the variability of texture and associated traits in boiled cassava

3.1 Introduction

Genetic gains are among other factors incumbent on availability of high genetic variability of traits of interest (Dumont *et al*., 2019; Iragaba *et al*., 2020) in breeding populations since breeders often must make phenotypic based selections. Therefore, morphological and molecular characterization of breeding populations should be routinely done to classify genotypes and to understand their taxonomic status (Lyimo *et al*., 2012; Adjei *et al*., 2023). This would inform decisions such as choice of progenitors for hybridization to create desirable variation while minimizing genetic drag (Hübner & Kantar, 2021).

Moreover, despite the importance of end-user qualities in cassava variety adoption in Uganda, minimal efforts have been dedicated to characterizing local germplasm morphologically or molecularly (Iragaba *et al*., 2020). This could partly explain the low selection response for cassava genotypes with improved cooking traits in the cassava breeding program, leading to low improved variety adoption.

With the advent and popularization of single nucleotide polymorphic (SNP) markers, efforts to characterize global cassava breeding populations has confirmed existence of sufficient genetic diversity (Ferguson *et al*., 2019). However, quality characterization particularly for consumer traits has not been done and this could limit efforts to exploit heterosis during hybridization. In essence, the combination of molecular and morphological characters is critical for selecting progenitors to generate superior genotypes (Ceballos *et al*., 2012; Njoku *et al*., 2015).

Morphological characteristics of cassava are routinely assessed in breeding programs (Peprah *et al*., 2020; Manze *et al*., 2021) possibly due to ease of phenotyping and particular demand placed by variety release committees for traits such as yield, disease resistance and plant architecture.

However, efforts by the Root, Tubers and Bananas (RTBs) project, a global initiative championing the development of mid and high throughput phenotyping tools for end-user traits, is making progress by leveraging Near infrared spectroscopy (NIRs), hyperspectral imaging (Meghar *et al*., 2023; Alamu *et al*., 2020) and other moderate throughput methods.

Improvements in phenotyping of cooking traits will enable cassava breeders to identify new genetic variation from which progenitors can be selected for population improvement and also improve selection intensity and accuracy at various stage gates (Araus *et al*., 2018). Overall, availability of genetic diversity for end-user linked cooking traits in local populations would determine the magnitude of improvement that can be achieved through selection (Daemo *et al*., 2023), but also anchor future demand led breeding initiatives.

Overall, there is need to conduct thorough morphological profiling of existing genetic stocks for root texture and other cooking qualities to inform selection of progenitors for the recurrent selection schemes in the Ugandan cassava breeding program. Therefore, this study was conducted to; (i) assess the variation in cooking traits within a cycle 2 (C2) genomic selection panel, (ii) evaluate the relationships among texture and water absorption qualities of boiled cassava and (iii) to estimate trait heritabilities.

3.2 Materials and Methods

Population

Genetic stocks from a cycle 2 genomic selection panel in the Ugandan cassava breeding program were constituted into a diverse panel of 434 accessions (Figure 2). The C2 population was generated from successive cycles of selection and hybridization of clones selected based on genomic estimated breeding values (GEBVs) from cycle zero (C0) and cycle one (C1) populations (Ozimati *et al*., 2019; Nandudu *et al*., 2024). The first cycle (C0) of genomic selection was conducted in 2013 when three hundred and ninety-five (395) diverse accessions generated by crossing 52 breeding lines sourced from IITA - Nigeria, CIAT-Colombia and within the East African region, were evaluated at Namulonge.

In 2014, 100 best accessions were selected and intercrossed to generate cycle 1 (C1) botanical seed. A total of 4874 seedlings were raised in a nursery bed and transplanted for evaluation in 2015. Accessions that had poor plant architecture and/or expressed severe symptoms of cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) were culled at the end of the 2015.

The rest of the accessions were advanced for clonal evaluation in 2016 and afterwards, 110 promising lines were identified for population improvement (C2). The C2 parents were crossed in 2017 and seedlings evaluated in 2018. Selection was made and a total of 434 accession were constituted into a diversity panel for conducting genetic studies on cooking qualities in cassava.

These materials were majorly segregating for cassava brown streak disease (CBSD) resistance alleles, which is a major priority for the breeding program. The study also included

three checks such as NAROCASS 1, a popular commercial variety known for its superior yield, Mkumba, a CBSD tolerant variety imported from Tanzania and TME204 an introduced variety from Nigeria.

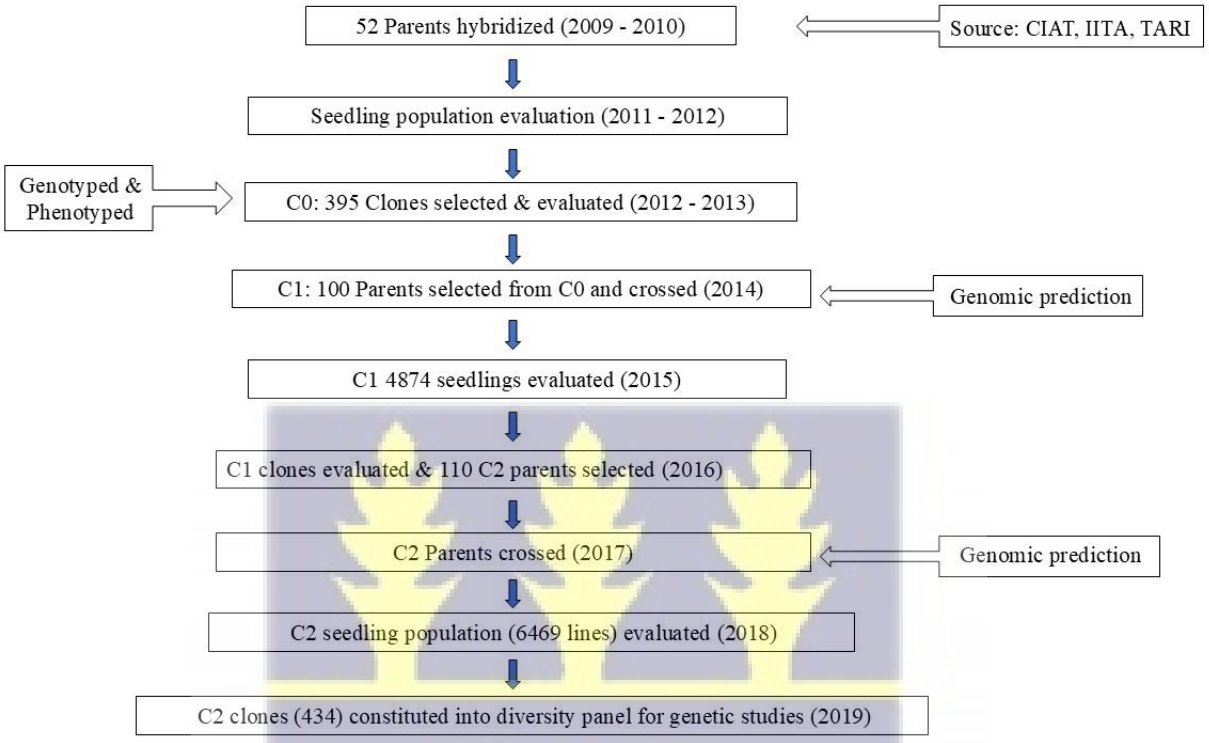


Figure 2: Showing pedigree and recombination cycles of the C2 population used in this study. Initial (C0) progenitors sourced from CIAT; International Center for Tropical Agriculture, IITA; International Institute of Tropical Agriculture and East Africa.

3.2.1 Phenotypic evaluation

Selected accessions (434) were evaluated at two locations, the National Crops Resources Research Institute (NaCRRI), Namulonge and the National Semi-Arid Resources Research Institute (NaSARRI), Serere for two growing seasons (2019 - 2020 and 2020 – 2021). Geographically, NaCRRI is located at latitude 0°5′ N, longitude 32°61′ E, and has an elevation

of 1,120 meters above sea level (masl) plus an average rainfall of 1,719 mm annually (Babaousmail and Ojara, 2025).

NaSARRI is located in the Eastern part of Uganda at latitude 1.52°N and longitude 33.45°E (Onyutha *et al*., 2021). The experiments were laid out in augmented incomplete block design with each block measuring 26 × 9 m (24 accessions and 3 checks) and plot measuring 1 × 9 meters. For each plot, ten (10) plants were established in a single row with interplant distance of 1 m.

Additionally, plots were separated from each other using an interrow (between plots) distance of 1 m. Checks were randomized within each block and adjacent blocks were separated by a 2 m alley. The experiments were rainfed and unfertilized, but routine weed control was done with hand hoes.

3.2.2 Data collection

After 12 months of field evaluation, all plants were harvested by uprooting and the roots were collated in a heap. Four relatively uniformly sized roots free of CBSD root necrosis were selected form each plot and placed in labelled (Block, Plot and Accession) gunny bags. A total of 250 root samples from different plots were collected. Roots were taken to the laboratory and immediately washed with clean tap water to remove soil and other debris. The roots were padded with a dry cloth to remove excess moisture and taken for sampling.

In the lab, each of the four clean roots from a plot were peeled and then sliced cross sectionally with a clean stainless-steel knife to generate 3 cylindrical root portions of approximately 6 cm length according to (Tran and Escobar, 2019). The portions were each further

cut into 2 half cylinders to generate a total of 6 half cylinders per root and 24 half cylinders for a plot. These portions were then analyzed for water absorption, softness, toughness and stiffness.

Assessment of Water absorption

This was determined according to (Tran *et al*., 2020) using a gravimetric assay. Briefly, the 24 pieces from different roots of a plot were mixed and then randomly assigned to four pre-weighed (Wo) perforated aluminum sample holders (Carlisle, Korea), making 6 half cylinders per sample holder (Supp. Fig 3). The weights of the sample holders plus root pieces were also recorded (W1) and the sample holders were placed into a stainless-steel pan containing clean boiling water over a gas flame.

The pan contained a large volume of water (4L per 400 – 600 g root) to minimize significant reductions in water temperature upon introduction of fresh roots. The pan was covered with an aluminum top, and water absorption was measured after 30 minutes cooking. This cooking time has been found to give moderate heritability estimates for water absorption in an earlier study (Namakula *et al*., 2023).

After 30 minutes of cooking, sample holders were quickly removed from the pan, padded with a dry cloth to remove excess water, and immediately weighed (W_{30}) using a 0.1 g precision digital scale (Metler Toledo, NewClassic MF, MS6001S, Switzerland). Water absorption (WAB) was determined as the percent increment in weight of boiled root pieces using the formulae.

$$\text{WAB} = \frac{W30 - W1}{W1 - W0} \times 100$$

Where WAB is the percentage water absorption measured after boiling cassava roots, W_{30} is the weight (g) of the sample holders and cassava roots cooked for 30 minutes, W1 is the weight (g) of the sample holders and the uncooked cassava roots and W0 is the weight of the empty sample holders.

Assessment of boiled root softness

This was done using a penetrometer proposed by Iragaba *et al*. (2019) and a texture analyzer by Tran & Escobar (2019). The root sections from WAB30 were randomly divided into two sets, each having 12 root sections. The root sections were wrapped in aluminum foil to keep them hot and immediately; one set was taken for softness assessment with penetrometer while the second set was assessed using a texture analyzer.

Using a penetrometer (Model number: FHT-1122, Vetus Industrial Company Limited, Hefei, China), the still hot 12 root sections were scored for softness (Soft_p) by puncturing the root. The penetrometer was set to maximum force of 11.1 kgf/cm2 and fitted with a 7.9 mm probe. The probe was then pressed into the boiled cassava piece to a depth of 1 cm and a reading corresponding the force applied was taken. This procedure was repeated for other eleven pieces to generate a total of 12 softness scores for each plot.

Softness assessment by texture analyzer (TA XT PLUSC, UK; 50 kg load cell) was done using the extrusion method proposed by Tran & Escobar (2019) with minor adjustments. This method uses both a compression probe (P/75) and a modified 5 blade grid (CIAT) for extrusion. Prior to testing the samples, the texture analyzer was set to a return-to-start mode, with pre-test speed of 2 mm/sec, test seed of 1 mm/sec, post-test speed of 10 mm/sec, compression distance of 20 mm and a trigger force of 500g.

Each of the hot root sections was loaded onto the extrusion grid with the sample fibers perpendicular to the blades and the test was started by allowing the compression probe to press the cassava sample through the extrusion grid to the set distance (20 mm). This was repeated for the remaining root sections and at least 8 test profiles were generated for each plot. The resulting texture profile curve for each sample was processed using Exponent Connect Software (Connect 7.0.6.0) to generate data on peak force (softness), area under the curve (toughness) and gradient (stiffness).

3.2.3 Data analysis

Descriptive statistics for each parameter in different seasons were determined using *stat.desc* function of *pastecs* package in R (Grosjean *et al*., 2004) to inspect variations among locations and seasons. Least significant differences (LSD) test was used to ascertain differences in performance of the different locations and seasons. Analysis of variance (ANOVA) was done to determine the major sources of variation for texture and water absorption in the population. This was done by fitting a fixed effects model for each trait using the *lm* function,

$$Y_{ijkm} = \mu + G_i + B_j + L_k + S_m + (G \times L)_{ik} + e_{ijkm}$$

Where Y_{ijk} is the phenotypic value of accession i in block j at location k in season m, G_i, B_j, L_k and S_m the fixed effects of accession, block, location and season respectively and G × L the accession by location interaction effect. The output of the analysis was generated using the *anova* function of *stats* package in R studio.

Two mixed effects models were fitted on the single location and combined data (location and season) using *lmer* function of the *lme4* package (Bates *et al*., 2015) in R to estimate components of variance;

$Y_{ijk} = \mu + G_i + B_j + S_k + (G \times S)_{ik} + e_{ijk}$ (Reduced model for single location across seasons)

$Y_{ijk} = \mu + G_i + \delta_j + \gamma_k + e_{ijk}$ (Full model across locations and seasons)

Where Y_{ijk} is the raw phenotypic value of a clone; μ the population mean; G_i is the random effect of the genotype with $G_i \sim N(0, \sigma^2_i)$, B_j the random effect of the incomplete block with distribution $N(0, \sigma^2_B)$, S_k is the random effect of a season, $G \times S$ is the random effect of the accession by season interaction with distribution $N(0, \sigma^2_{GS})$, δ_j the random effect of the block-location-season assumed to be distributed $N(0, \sigma^2_\delta)$, γ_k is the random effect of the accession by location interaction with distribution $N(0, \sigma^2_\gamma)$ and e_{ijk} the residual variance assumed to be random and with distribution $N(0, \sigma^2_e)$. Variance from the models were used to calculate broad sense heritability for each environment (H^2_s) and across environments (H^2_a);

$$H^2_s = \frac{\sigma^2 G}{\sigma^2 G + \frac{\sigma^2 G \times S}{nS} + \frac{\sigma^2 R}{nS}}$$ ..... (Holland *et al*., 2003)

Where H^2_s is the broad sense heritability at location level, σ^2G is the genetic variance, σ^2GxS is the variance due to genotype by season interaction and σ^2R the residual variance.

$$H^2_a = \frac{\sigma^2 G}{\sigma^2 G + \frac{\sigma^2 G \times E}{nE} + \frac{\sigma^2 R}{nE}}$$ ..... (Piepho *et al*., 2008)

Where σ^2G is the genetic variance, σ^2GxE is the variance due to the genotype by environment interaction, *nE* is the number of environments and σ^2R is the residual variance.

Phenotypic correlations among cooking traits were determined with Peason's correlation coefficient using the *cor* function in *stats* package of R and the correlation matrix was visualized using *complexheatmap* package in r (Gu, 2022). Measures of variability including phenotypic variability, genotypic variability, phenotypic coefficient of variation (PCV), genotypic coefficient of variation (GCV), genetic advance (GA) and genetic advance as percent of mean (GAM) were estimated according to Tiwari *et al*. (2019);

Phenotypic variance (σ^2P) = $\sigma^2G + \sigma^2E + \sigma^2GxE$

Genotypic coefficient of variation (GCV) = $\frac{\sqrt{\sigma^2 G}}{\mu}$ X 100

Where μ is the mean of a cooking trait in the population.

Phenotypic coefficient of variation (PCV) = $\frac{\sqrt{\sigma^2 P}}{\mu}$ x 100

Where σ^2P is the phenotypic variance and μ is the mean of a cooking trait in the population. Genetic advance was estimated as;

Genetic Advance (GA) = $I \times \sigma_p \times H^2$,

Where I is the selection differential which varies with selection intensity (Selection intensity of 5% was considered in this work, which gave I of 2.06), σ_p is the phenotypic standard deviation and H^2 the broad sense heritability.

Genetic advance as a percentage of mean (GAM %) was determined as;

$$GAM = \frac{GA}{\mu} \times 100,$$

Where GA is the genetic advance and μ the population mean.

Principal component analysis

This was done to establish the contribution of each of the traits to the total variation in the population (Enyew *et al*., 2019). Prior to the analysis, the data was standardized to remove scale differences between the variables (Forkman *et al*., 2019). The analysis was conducted using *prcomp* function of stats package in R and the first two principal components (PCs) were visualized using *factoextra* package in R (Adjei *et al*., 2022).

Cluster analysis

This was done to ascertain grouping tendency and morphological relatedness in the population (Sinha and Mishra, 2013). The strength of clustering structure within the population was assessed using an agglomerative coefficient (ac) that ranges from 0 to 1, whereby coefficients close to one show very strong clustering structure (Maechler *et al*., 2023).

This was implemented using the *agnes* function of *cluster* package in R. Further, a standardized data matrix was used to estimate pairwise genetic similarity values between genotypes (Adjei *et al*., 2022) using Euclidean distance. This was implemented using the *stats* package in R. Genetic distances calculated were used to construct a *ward.D* clustering algorithm hierarchical dendrogram and visualized using *fviz_nbclust* function of *factoextra* package in R (Kassambara *et al*., 2016).

3.3 Results

3.3.1 Trait distribution

All traits were positively skewed in the population but the degree of skewness varied among traits and locations (Figure 3). Gradient was the most skewed trait among all traits and in both locations whereas softness assessed by texture analyzer (Soft_T) was the least skewed across locations. Area under the curve was most skewed in Namulonge but in Serere, it was nearly normally distributed (Supp. Fig. 4). Apart from WAB, all traits were more skewed in Namulonge than Serere locations.

Soft_p had a low mean (1.67 kgfcm^{-2}) in Namulonge and it ranged from 0.31 – 6.59 kgfcm^{-2} (Table 1) whereas in Serere, Soft_p had a higher mean (3.94) and wider range (0.53 – 8.62). Across seasons, Soft_p still remained consistently lower in Namulonge than Serere (Figure 4A). Estimates of broad sense heritability were generally low for Soft_p, though higher in Namulonge (H^2_{Sofp} = 0.25) than Serere (H^2_{Sofp} = 0.02). In the combined analysis, heritability for Soft_p was slightly higher (H^2_{sofp} = 0.38) than in either location (Table 1).

Similarly, Soft_T had a lower mean in Namulonge (10015.74 gF) and narrow range (2618.91 - 23066.85) compared to Serere with mean 18073.42 and range 4337.69 to 35836.89 gramforce (gF). Across seasons, Namulonge generally had lower softness measurements than Serere. Broad sense heritability for Soft_T was moderate and generally higher in Namulonge ($H^2_{Soft\_T}$ = 0.57) than in Serere ($H^2_{Soft\_T}$ = 0.44) and the combined estimate across locations was $H^2_{Soft\_T}$ = 0.55.

WAB had a higher mean in Namulonge (10.11 %) than Serere (2.93 %) but with a narrow range (-1.74 - 38.16 %) than Serere (-5.06 – 49.99 %). Likewise, across seasons, Namulonge

had higher WAB estimates than Serere (Figure 5C). Broad sense heritability was generally low in individual locations though estimates at Namulonge (H^2_{WAB} = 0.38) were higher than at Serere (H^2_{WAB} = 0.04). The combined heritability estimate across locations and seasons was moderate (H^2_{WAB} = 0.52).

The mean toughness (Area under the curve) was lower for Namulonge (105945.2 g.mm) than Serere (197183.8 g.mm) and the latter also had a wider range for the trait. Across seasons, Namulonge generally had lower estimates of toughness than Serere (Figure 5D). Heritability estimates in the two locations were moderate and similar ($H^2_{Tough\_Nam}$ = 0.48, $H^2_{Tough\_Sere}$ = 0.49) but slightly higher in the combined analysis ($H^2_{Tough\_comb}$ = 0.51).

The gradient (stiffness) was also generally lower in Namulonge with a mean of 568.73 g/mm and higher in Serere (mean = 1220.45). This trend was also consistent across seasons with Serere having higher stiffness than Namulonge (Figure 5E). Heritability was low in Namulonge (H^2_{Stiff} = 0.29) and moderate at Serere (H^2_{Stiff} = 0.66). Likewise, the combined estimate of heritability across location was moderate (H^2_{Stiff} = 0.62).

Soft_T (softness by texture analyzer) had the lowest CV ranging from 32.86% at Serere to 45.13% in the combined location and season data, whereas WAB30 had the highest CVs ranging from 75.52 at Namulonge to 196.04 % at Serere (Table 1).

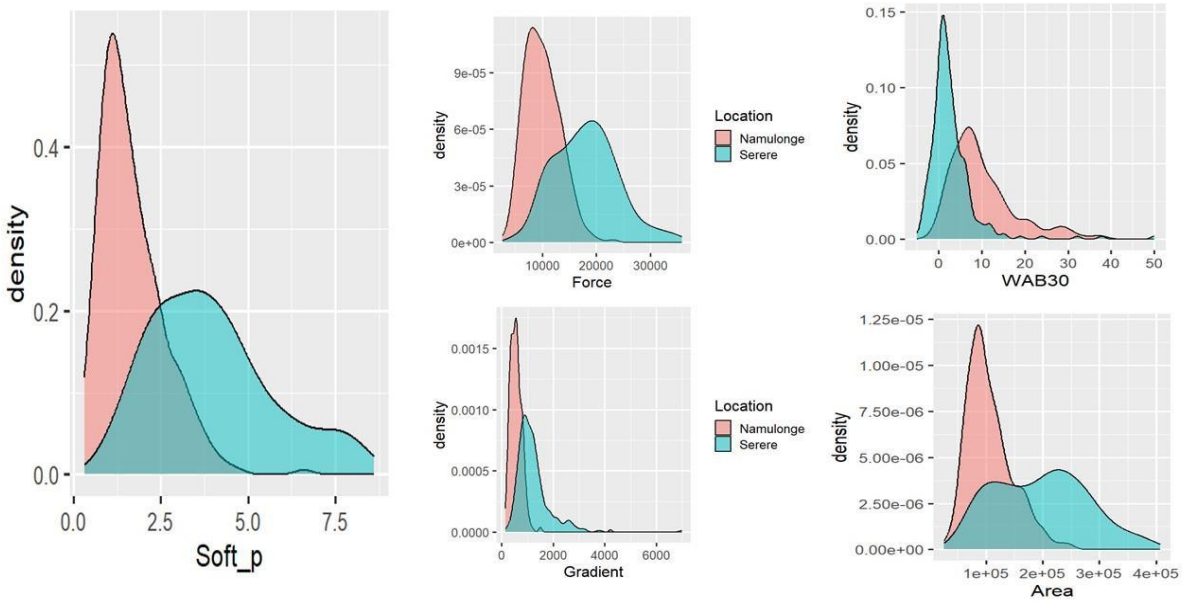


Figure 3: Distribution of traits by location in the population. Soft_p: Softness assessed by penetrometer; Force: Peak force representing softness measured by texture analyzer (soft_T); WAB30: Water absorption determined after 30 minutes of cooking; Gradient: stiffness of the cassava sample and Area: equivalent to the toughness of a sample.

Overall, broad sense heritability for cooking traits was mostly moderate and stiffness had the highest heritability among all cooking traits. Also, heritability estimates for softness using the texture analyzer (Soft_T) was higher than those using the penetrometer (Soft_p). Comparison of cooking traits across locations showed that Soft_p, Soft_T, Toughness and Stiffness were more variable in Serere than Namulonge. Likewise, these traits were also more variable in season 2 than season 1 (Supp. Fig. 4).

Table 1: Summary statistics of individual and combined location data across two seasons

Trait	Mean ± SEM	Min	Max	% CV	H^2
		NaCRRI			
Soft_p	1.67 ± 0.05	0.31	6.59	54.16	0.25
Soft_T	10015.74 ± 193.72	2618.91	23066.85	32.94	0.57
WAB	10.11 ± 0.45	-1.74	38.16	75.52	0.38
Toughness	105945.2 ± 2352.18	26269.92	247362.7	37.81	0.48
Stiffness	568.73 ± 17.96	139.28	4216.56	53.77	0.29
		Serere			
Soft_p	3.94 ± 0.12	0.53	8.62	45.09	0.02
Soft_T	18073.42 ± 385.02	4337.69	35836.89	32.86	0.44
WAB	2.93 ± 0.37	-5.06	49.99	196.04	0.04
Toughness	197183.8 ± 5299.17	45019.55	406892.3	41.46	0.49
Stiffness	1220.45 ± 44.21	208.3	7000.99	55.88	0.66
		Combined NaCRRI and Serere			
Soft_p	2.69 ± 0.08	0.31	8.62	65.84	0.38
Soft_T	13647 ± 268.06	2618.91	35836.89	45.13	0.55
WAB	6.87 ± 0.34	-5.06	49.99	112.31	0.52
Toughness	147071.70 ± 3357.01	26269.92	406892.3	52.45	0.51
Stiffness	862.5 ± 26.32	139.29	7000.99	70.12	0.62

Soft_p: Softness assessed by penetrometer; Soft_T: softness assessed by texture analyzer (Peak Force); WAB: Water absorption assessed after 30 minutes of boiling; Toughness: of boiled root; Gradient: the gradient of the texture profile curve which is equivalent to the stiffness of a sample.

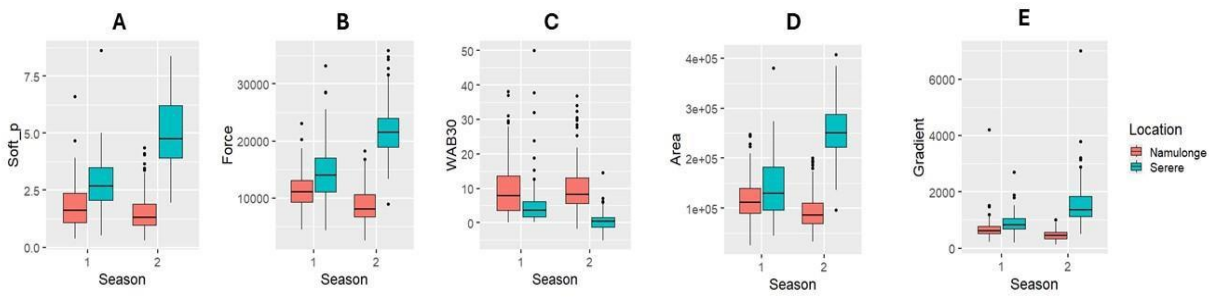


Figure 4: Boxplots showing differences in performance of the C2 cassava population for cooking traits across locations.

3.3.2 Analysis of variance

The combined analysis of variance showed that there were significant differences among the genotypes (Table 2) for Soft_p, Soft_T, WAB30, toughness and stiffness ($p<0.001$). The analysis also revealed that block and environment effects also significantly influenced all cooking traits ($p<0.001$). The genotype by environment (GxE) effects were only significant for WAB ($p<0.001$).

For Soft_p, environment effects were the most influential accounting for up to 93.08% of the total variation. Likewise for Soft_T, most of the variation was attributed to environment effects (93.83%). Similarly, variation for WAB30 in the population was majorly explained by environment effects (85.28%). Environment effects also remained the biggest contributors to the variation observed in toughness and stiffness texture traits.

Table 2: Analysis of variance associated with each of the cooking traits

SOV	Df	Soft_p	Soft_T	WAB	Toughness	Stiffness
Geno (G)	249	2.41***	32893737.00***	69.26***	4825500000.00***	72.30***
Block	19	9.27***	99469848.00***	112.40***	14086000000.00***	158.80***
Env't (E)	3	189.77***	2353949060.00***	1319.04***	403360000000.00***	4119.2***
G x E	207	1.252	11063913.00	36.82***	1751700000.00	21.40
Residuals	49	1.176	11454485.00	9.16	1917600000.00	22.30

Geno: Genotype effect; **Env't**: Environment effects; **GxE**: Genotype by environment interaction effects; **Soft_p**: Softness assessed by penetrometer; **Soft_T**: softness assessed by texture analyzer (Force); **WAB30**: Water absorption assessed after 30 minutes of cooking. Significance: *p< 0.05; ** p< 0.01; ***p< 0.001.

Mean separation of treatments using Least Significant Differences (LSD) at 5% probability level showed that location Serere had statistically significantly ($p<0.05$) higher mean Soft_p, Soft_T, toughness and stiffness measurements than Namulonge. In contrast, Namulonge had significantly higher ($p<0.05$) mean WAB30 than Serere across all seasons (Supp. Fig. 5).

3.3.3 Phenotypic correlations among cooking traits

Phenotypic correlations among the cooking traits were all statistically significant ($p<0.001$). The highest correlation was positive between Soft_T and Toughness ($r = 0.96$), while the lowest correlation was negative between WAB30 and stiffness (Figure 5). Correlation between Soft_T and Soft_p which are measures of softness was positive and high ($r = 0.79$). Contrastingly, WAB30 was negatively correlated with all other traits and the highest negative

correlation was between WAB30 and Soft_p (-0.85). Also, all traits determined by the texture analyzer (Soft_T, Area and gradient) were highly positively correlated.

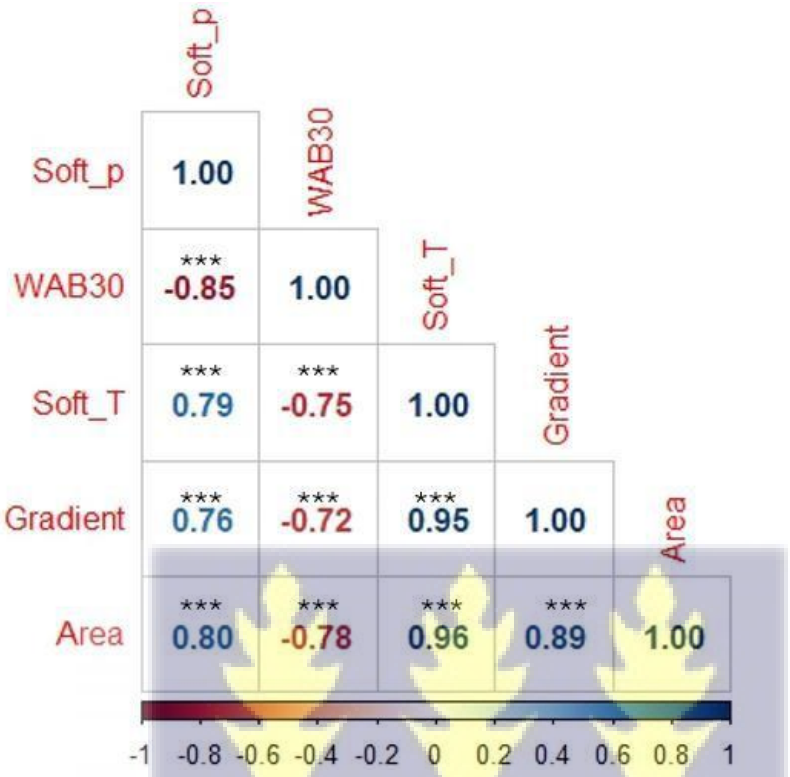


Figure 5: Phenotypic correlations among cooking traits of boiled cassava

3.3.4 Genetic variability of cooking traits in the population

Genotypic variance was lower than environment and genotype by environment variances for Soft_p and WAB30. For Soft_T, Toughness and Stiffness, the genotypic variance was lower than environment variance but higher than genotype by environment variance. All traits had a larger magnitude of phenotypic coefficient of variation (PCV) than genotypic coefficient of variation (GCV). WAB30 had the highest PCV (96.83 %) and GCV (45.01 %) while Soft_p had the lowest PCV and GCV (Table 3). Generally, GCV across traits ranged from 7.43 for Soft_p to 45.01% for WAB30 and PCV from 19.32 to 50.36%

respectively. Soft_T had a higher GCV and PCV than Soft_p. GAM ranged from 21.56% for Soft_p to 113.36% for WAB30.

Table 3: Variance components and coefficients of variation of cooking traits assessed

Trait	σ^2g	σ^2p	GCV	PCV	GA	GAM
Softp	0.04	0.55	7.43	19.32	0.58	21.56
Soft_T	3013731	2.82×10^7	12.72	26.44	6017.36	44.09
WAB30	9.56	52.86	45.01	96.83	7.79	113.36
Toughness	4.08×10^8	4.47×10^9	13.73	30.22	67477.47	45.88
Stiffness	53856	263041	26.91	50.36	655.04	75.95

σ^2g: genotypic variance, **σ^2p:** phenotypic variance, **GCV**: genotypic coefficient of variation, **PCV**: phenotypic coefficient of variation, **GA**: Genetic advance and GAM: genetic advance as percentage of mean.

3.3.5 Principal Component Analysis

To further understand the pattern of variation among the cooking traits and the contribution of each trait to the observed variation, principal component analysis was done. Most of the variability in the population was explained by principal component 1 (PC1) accounting for up to 75.3% of total variation (Table 4) and the first two PCs cumulatively explained up to 89.3% of total variation. Also, PC1 had the highest eigen value (3.77) whereas other PCs had eigen values less than 1. Traits that were highly correlated with PC1 ($r > 0.8$) were Soft_p, Soft_T, Toughness and Stiffness whereas WAB30 was highly correlated ($r = 0.67$) with PC2. Since all cooking traits contributed substantially to the most informative PCs, all traits were found relevant for discriminating the cassava genotypes in the population.

The first two PCs were further plotted on a biplot for visual inspection of the spatial variability and extent of grouping within the population (Figure 6). The genotypes were colored based on the optimal number of groups determined through hierarchical cluster analysis. The accessions were mainly scattered around the center and others along the PC axes. There was a noticeable overlap between clusters and 1 and 2 though the general dispersion pattern was able to separate the two clusters based on the cooking traits.

Majority of the traits (Soft_p, Soft_T, Toughness and Stiffness) were associated with dispersion observed in cluster 1 while WAB30 was associated with cluster 2. The plot also showed that traits Soft_p, Soft_T, Toughness and Stiffness were distributed opposite to WAB30 which is indicative of a negative association between the latter and the former. The small acute angles among Soft_p, Soft_T, Toughness and Stiffness confirmed strong positive correlations among these traits. Also, all traits had relatively long vectors which further confirms their relevance in discriminating the accessions.

Table 4: Principal component analysis (PCA) of cooking traits in cassava showing eigen vectors, eigen values, variance for each PC and cumulative variance.

Trait	PC1	PC2	PC3	PC4	PC5
Soft_p	**-0.91**	-0.15	0.01	-0.39	-0.01
WAB30	0.71	**0.67**	-0.16	-0.16	0.00
Force	**-0.95**	0.18	-0.16	0.11	-0.17
Gradient	**-0.81**	0.44	0.38	0.06	0.04
Area	**-0.94**	0.09	-0.30	0.09	0.14
Eigen	3.77	0.69	0.29	0.2	0.05
Variance (%)	75.34	13.97	5.7	4	0.99
Cumulative Variance (%)	75.34	89.31	95.01	99.01	100

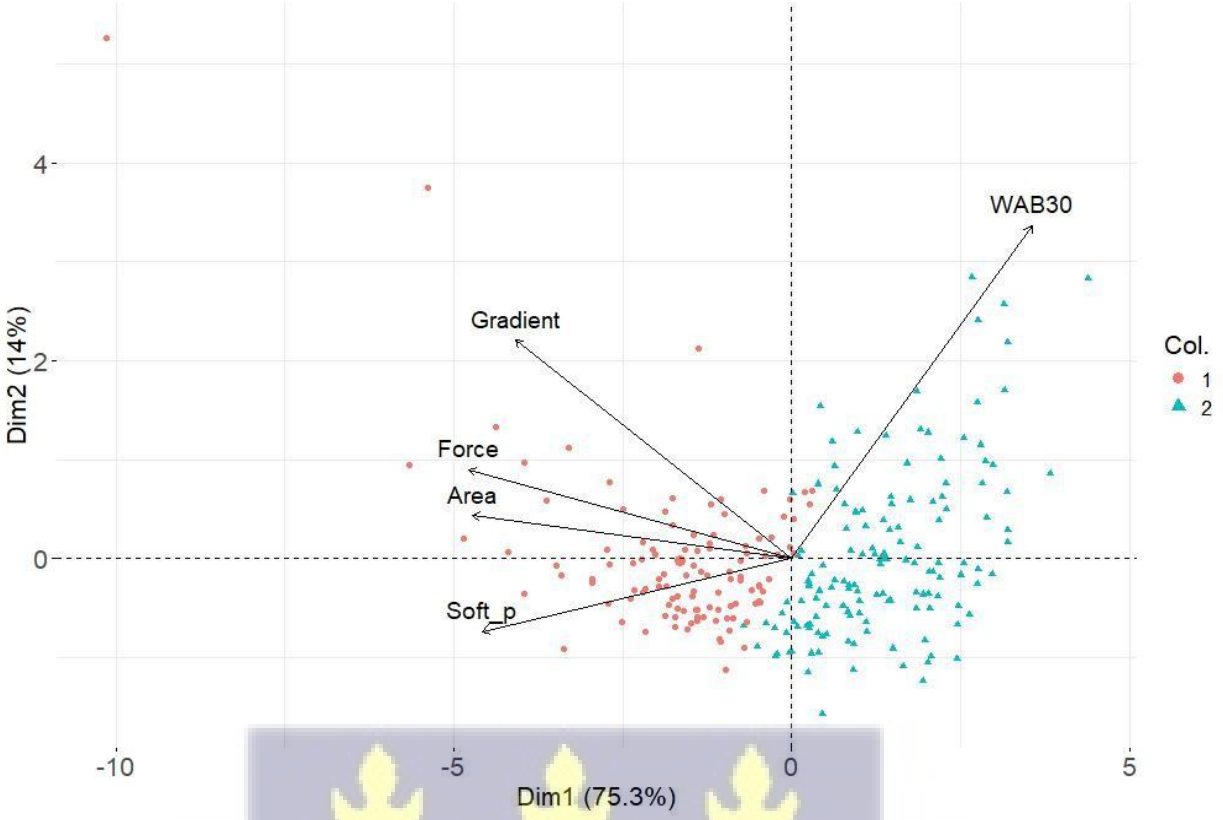


Figure 6: A PCA biplot of the first two principal components of cooking traits.

3.3.6 Cluster Analysis

There was considerable variation among the accessions based on their cooking traits as determined using cluster analysis. The strength of clustering structure within the population gave an agglomerative coefficient of 0.99 indicating very strong structure. Analysis based on *ward.D* hierarchical clustering found two optimal clusters in the population (Figure 7). Cluster one comprised 112 accessions while cluster 2 comprised 137 accessions.

The means of Soft_p, Soft_T, Toughness and Stifness in cluster 1 were statistically significantly higher ($p<0.05$) than those in cluster 2 (Table 5), whereas mean for WAB30 was significantly higher in cluster 2 than 1. Similarly, the ranges for all traits except WAB30 were wider in cluster one than cluster two.

The mean for Soft_p in cluster 1 (3.5 kgfcm^{-2}) was twice that in cluster 2 (1.78 kgfcm^{-2}). The mean of Soft_T in cluster 1 was 16985.79 gF and 9971.44 gF in cluster 2 (Table 6). The mean of WAB30 in cluster two (10.45%) was approximately three times larger than that in cluster 1 (3.51%). All checks including UG110017, Mkumba and TMEB204 were scattered in cluster 1 which was associated with hard textured phenotypes and low WAB30. Means of Stiffness and Toughness in cluster 1 were nearly twice the averages in cluster 2.

Further, based on Soft_p accession UG16F103P002 was the softest in cluster1 (1.75 Kgfcm^{-2}) whereas UG16F312P031 was the hardest (7.42 Kgfcm^{-2}). Based on Soft_T, UG16F290P236 had the softest texture (12537.02 gF) while UG16F312P031 had the hardest texture in cluster 1.

Accession UG16F292P023 had the highest WAB30 while UG16F308P041 had the lowest (Supp. Table 1) in cluster 1. In cluster 2, UG16F042P002 had the lowest Soft_p (0.38 kgfcm^{-2}), the highest WAB30 and UG16F004P004 had the lowest Soft_T (4395.56 gF).

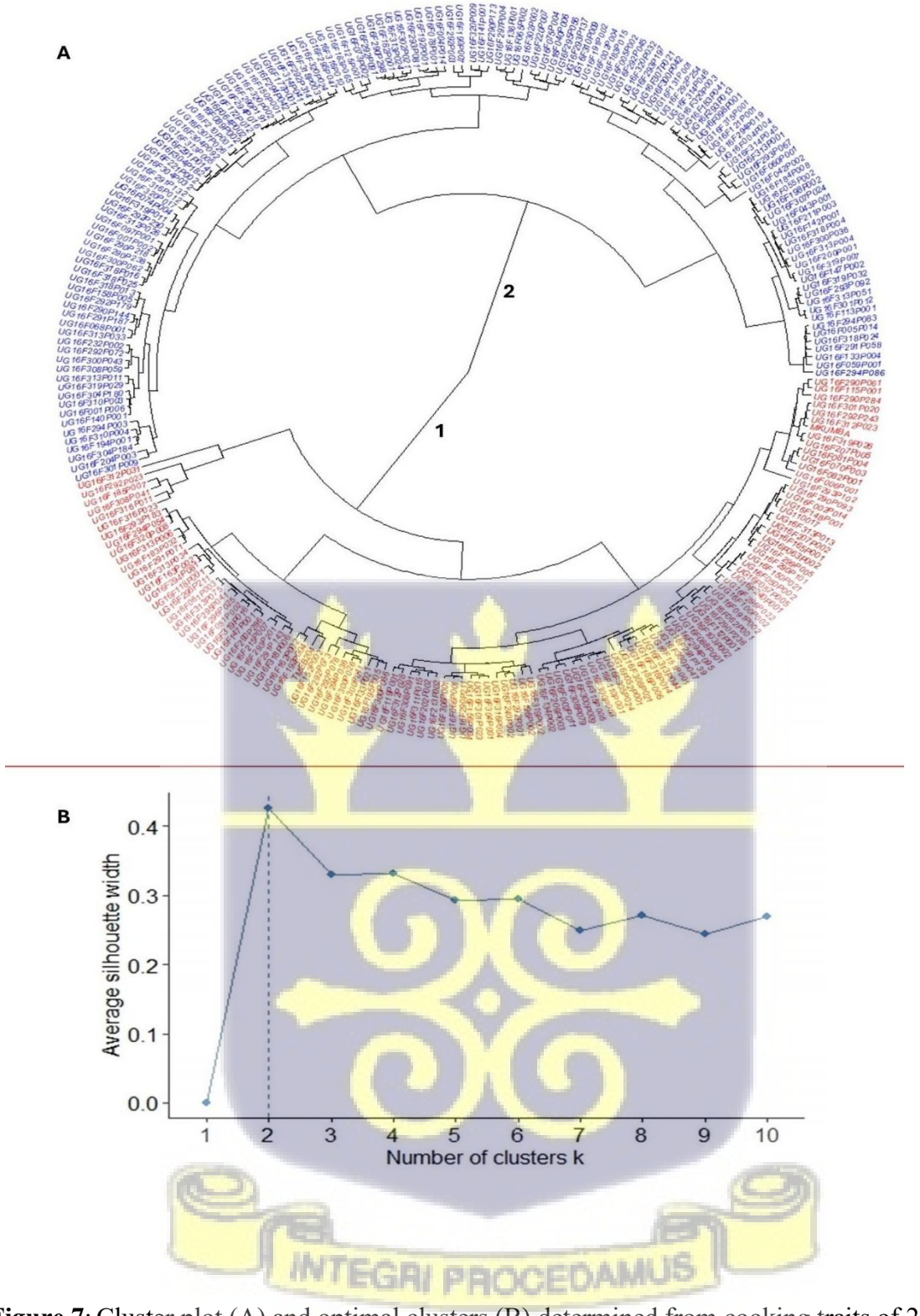


Figure 7: Cluster plot (A) and optimal clusters (B) determined from cooking traits of 250 accessions.

Table 5: Performance statistics of clusters 1 and 2 from cooking traits in cassava

Trait	No. Acc	Cluster	Mean ± SEM	Min	Max	CV %
Soft_p	112	1	$3.5^{a} \pm 0.09$	1.75	7.42	27.37
	137	2	$1.78^{b} \pm 0.07$	0.38	4.47	70.35
Soft_T	112	1	$16985.79^{a} \pm 320.20$	12537.02	31277.19	19.95
	137	2	$9971.44^{b} \pm 210.79$	4395.56	14878.02	24.74
WAB30	112	1	$3.51^{a} \pm 0.34$	-3.57	14.63	102.46
	137	2	$10.45^{b} \pm 0.59$	0.81	36.86	66.29
Stiffness	112	1	$1153.84^{a} \pm 62.68$	652.93	7000.99	57.49
	137	2	$573.91^{b} \pm 15.72$	230.91	1140.66	32.06
Toughness	112	1	$187238.7^{a} \pm 3667.97$	128207.6	332533.5	20.73
	137	2	$100379.3^{b} \pm 2204.34$	43824.19	158838.7	25.70

Means within a trait that have different subscript letters are statistically significantly different ($p<0.05$.

3.4 Discussion

Morphological characterization of breeding populations is vital for effective utilization of existing variability in a breeding program (Nduwumuremyi *et al.*, 2018) for identifying individual accessions with desirable qualities that can be used for population improvement. For cassava, the need to increase acceptance of cassava varieties through breeding has led to the bench marking of end user preferences during population improvement, selection and variety development (Iragaba *et al.*, 2024). Therefore, an informed understanding of the variability of cooking traits in existing cassava germplasm will be critical for developing varieties with acceptable thresholds of texture and water absorption traits in Uganda and beyond.

The positive skew distribution of Soft_p observed across locations corroborates with findings by Namakula *et al*., (2023) using the same method of evaluation but different populations in Uganda. However, Uchendu *et al*., 2021) found a more normally distributed Soft_p, which indicates that distribution of this trait is dependent on the population used. Further the range observed in this study after 30 minutes of boiling (0.31 – 8.62 $Kgfcm^{-2}$) was wider than that observed by Namakula *et al*. (2023) and Iragaba *et al*. (2019) who used smaller populations.

The range of Soft_T observed in this study (2618.91 - 35836.89 gF) was wider than that reported by Iragaba *et al*., (2024) which also confirms the wide variability of softness in the C2 population based on the texture analyzer. The range for WAB30 (-5.06 – 49.99 %) observed in this study was also higher than that reported by Kouadio *et al*. (2011) and Tran *et al*. (2020), though this could be attributed to the smaller populations used in the previous studies. This further confirms high variability for WAB30 within the population used in this study.

The range of Toughness observed in this work (26269.92 - 406892.3 g/mm) was also much wider than that reported by (Iragaba *et al*., 2024). Likewise, Stiffness also had a wider range in this study than that reported by (Iragaba *et al*., 2024), implying there was also considerable variation for these traits in the population.

The CVs reported in this study varied across traits. The CV reported across locations for Soft_p (65.84%) was much higher than that reported by (Uchendu *et al*., 2022) and (Namakula *et al*., 2023) but similar to that reported by (Wembabazi *et al*., 2022) using similar methods but different populations and sizes.

The coefficient of variation (CV) is often used as a measure of accuracy in an experiment, whereby low CVs (<10) indicate good accuracy in an experiment (Al-Naggar *et al*., 2020). Soft_T on the hand had a low CV compared to Soft_p, which indicates that the texture analyzer is a more reliable tool for measuring boiled cassava root softness than the penetrometer. The CV for Soft_T was also comparable to that reported by Adinsi *et al*. (2023).

WAB30 had the highest CV among traits and this was comparable to that reported by Pierre *et al*., (2023) but higher than that found by Wembabazi *et al*., (2022). This implies there is need to improve on the phenotyping efficiency for this trait. Toughness had a CV comparable to that reported by Adinsi *et al*. (2023).

Broad sense heritability for Soft_p remained low in single and across locations similar to findings by Iragaba *et al*. (2019) and Wembabazi *et al*. (2022) but lower than estimates reported by Namakula *et al*. (2023), much as the latter used single location data, larger plot sizes and different cooking times 30 and 45 minutes). Similar heritability for Soft_p to that in this study has also been reported by Uchendu *et al*. (2022) from a study conducted across three environments and two seasons.

Broad sense heritability for Soft_T was generally moderate and higher than that of Soft_p which further confirms the texture analyzer as a more reliable tool for assessing boiled root softness. For WAB30, heritability was also moderate but much lower than that reported by Tran *et al*. (2020) who used single location estimates with no seasonal effects.

Trait heritability is critical for estimating genetic gains (Ceballos *et al*., 2012), selection response (Pradeepkumar *et al*., 2001) in a breeding program and for increasing prediction

accuracy of target traits in a genomic selection (GS) pipeline (Zhang *et al*., 2017). Therefore, it is critical to optimize sampling and phenotyping procedures to maximize heritability estimates for target traits (Ceballos *et al*., 2012).

Traits with high heritability are amenable for direct selection (Fan *et al*., 2019) and could potentially result in fast progress upon selection for accessions with low softness scores. Heritability estimates could also be increased by using bigger experimental plot sizes, more replications (Ceballos *et al*., 2012) and even increasing the number locations for phenotyping trials (Zhang *et al*., 2017).

Variation in cooking traits was majorly attributed to genotype and environment effects and but the genotype by environment (G×E) effects were only statistically significant for WAB. The highly significant block effects observed for all traits (Table 2) could be due to specific field conditions such as soil heterogeneity which could have differentially influenced the checks (Harouna *et al*., 2020).

Comparably, Uchendu *et al*. (2022) also found significant influences of genotype and environment effects on boiled root softness but the genotype by environment (GxE) effect also had a major influence on cooking traits in their study. Moreover, the latter study was conducted in more environments, which therefore provides a more realistic evaluation of genotype by environment interaction (GEI).

The significance of accession and environment main effects shows that some accessions were stable in specific environments. Also, the highly significant genotype main effects across all

cooking traits informs of considerable genetic variation available in the population. This is no surprise given the recombination history of the C2 population used.

This presents a unique opportunity for exploitation in cassava breeding in Uganda. Such variability could be leveraged for progressing in the genetic improvement of cassava for preferred cooking trait thresholds through hybridization and selection (Esuma *et al*., 2016). The correlations observed among all traits were high and significant. These findings corroborate with those reported by Wembabazi *et al*. (2022) who found strong significant negative correlation between Soft_p and WAB30.

However, Uchendu *et al*. (2022) found low to moderate phenotypic correlations among sensorially determined cooking traits, which could be attributed to low accuracy and repeatability of methods used. In another study, sensorially determined cooking traits such as mealiness was found to have a moderate significant correlation with WAB30 and stiffness, implying there was possibility of using instrumental methods to phenotype mealiness (Iragaba *et al*., 2024).

The strong positive correlation between Soft_p and Soft_T is an important find confirming the agreement between the penetrometer and texture analyzer measurements of boiled root softness. A similar finding was also reported by Iragaba *et al*. (2024), implying that one of the two approaches can be used for phenotyping boiled cassava softness efficiently. However, this study also found that the texture analyzer gave better heritability estimates than the penetrometer (Table 1).

The strongly correlated traits in this study provide opportunity for simultaneous improvement of all cooking traits by including only one in a selection index (Nduwumuremyi *et al*., 2018; Al-Naggar *et al*., 2020). The negative correlation between WAB30 and root softness (Soft_p and Soft_T) implies that the more water a root absorbs during cooking, the softer it gets (Wembabazi *et al*., 2022). Moreover, root softness is often used as a measure of readiness (how well cooked) of boiled cassava for consumption (Iragaba *et al*., 2019; Tran *et al*., 2021).

Patterns of variation in cooking traits

Whereas there was evidence for presence of considerable variation in the C2 population for cooking traits, not all variation was attributed to genetic variance but rather significant variations in environments. Furthermore, the difference between the PCV and GCV for all traits was relatively large, suggesting that environmental factors had a substantial impact on observed variations (Gadissa *et al*., 2021). The downstream consequence of this is that phenotypic based selection may not be reliable when dealing with such traits having large environment effects.

The lower GCV for Soft_p (7.43%) compared to Soft_T (12.72%) confirms that the texture analyzer was a more reliable phenotyping tool than the penetrometer, in as much as both traits had the largest influence of environmental factors. This implies that the use of multiple test locations and years is critical to revealing the true extent of genetic variation associated with a breeding population for softness, while minimizing environmental errors (Gadissa *et al*., 2021).

It was also observed that traits with low heritability (Table 1) also had low GCV and PCV values (Table 3). Contrastingly, Uchendu *et al*. (2022) found a higher and moderate GCV

(13.26%) for Soft_p but a PCV comparable to this study, which could be attributed to differences in the genetic background of the populations used.

The high GCV but even higher PCV recorded for WAB30 have not been previously reported for cassava, but it was also indicative of larger environment effects influencing the trait. High GCV and PCV have been previously reported for water absorption in rice (Chakraborty *et al*., 2009), though the difference between the two coefficients was very minimal, indicating large genetic effects accounted for observed variation. Therefore, more evaluations should be done for WAB30 in cassava breeding populations at multiple locations to gain better understanding of the extent of genetic control of the trait.

High genetic advance (>20%) as percentage of mean (GAM) was observed for all traits implying that selecting the top 5% of the population would result in gains of 21.56 to 113.36 % over the initial population mean. These findings contrast slightly with those reported by Uchendu *et al*. (2022) who found a lower GAM for Soft_p.

High GAM has been previously associated with strong additive gene action (Nduwumuremyi *et al*., 2018) and together with heritability estimates, could improve selection response (Peprah *et al*., 2020). Therefore, phenotypic based selection could still achieve desirable improvements in these traits (Awad- Allah *et al*., 2022).

The lower GAM for Soft_p reported in this study compared to other cooking qualities implies that selection for the former in the Uganda breeding program would progress much slower than other cooking traits. Very high GAM (>80 %) such as observed for WAB30 in

this study has been previously reported in other studies on WAB (Chakraborty *et al*., 2009) and other traits (Shilpashree *et al*., 2021) in different crops.

Cluster analysis was employed to visually assess the optimal number of groups that could form in the population, whereby each group contained individuals with similar qualities (Al-Naggar *et al*., 2020). The C2 population was split into two optimal groups whereby cluster 1 had fewer accessions than cluster 2 (Figure 7). The two clusters represented uncorrelated groups where group 1 had higher means for Soft_p, Soft_T, Area and Gradient than group 2 and WAB30 was higher in group 2 than group 1 (Table 5).

Such across group variation may be exploited to maximize heterosis for lower means of texture which are desirable for most cooking traits except WAB30. This would ensure considerable progress in developing soft textured and high-water absorbing cassava varieties. This notion has been previously corroborated by Al-Naggar *et al*. (2020).

Clustering based on origin of the accessions was not considered since the C2 population had already undergone 2 cycles of recombination which implied that the accessions shared common alleles among them (Figure 2). Therefore, the morphological differences observed in the C2 population are the likely cause of genetic diversity. Nevertheless, it is still important to collaboratively engage other cassava breeding programs for germplasm exchanges in order to boost available diversity with new alleles for cooking traits against stable genetic backgrounds for agronomic traits.

Multivariate analysis of phenotypic variability in the population using principal component analysis (PCA) has also been proven as a reliable means of visualizing and quantifying

morphological diversity in a population and ascertaining the contributing variables to such diversity (Sinha & Mishra, 2013; Al-Naggar *et al*., 2020).

The PCA analysis found that the first two PCs (Figure 6) accounted for majority of the variance (89.3%), which is higher than that reported by Uchendu *et al*. (2022) for similar cooking traits. The PCA biplot (also confirmed that WAB30 was negatively correlated with other cooking traits (Soft_p, Soft_T, Toughness and Stiffness), which has been reported previously (Wembabazi *et al*., 2022; Iragaba *et al*., 2024).

All traits except WAB were highly negatively correlated with PC1 ($r > -0.8$) whereas WAB30 was highly positively correlated with PC2 (Table 4), implying that all traits could be used to discriminate among genotypes (Gadissa *et al*., 2021). From the biplot, it was evident that WAB30 had the longest vector among all traits, implying that there was more variation in the trait than others.

Such discrimination could be vital for identifying progenitors for future strategic hybridizations to maximize improvement of cooking time in the breeding pipeline targeting fresh cassava market segment. Groups identified through cluster analysis were used to group the PCA plot resulting in distinct groups on the plot with minimal admixture.

This distinction among groups may be further used for developing thresholds for classifying cooking traits, leading to optimization of high throughput qualitative phenotyping tools such as NIRS (Sampaio *et al*., 2022) for rapid assessment of cooking traits in boiled cassava.

Implications for breeding consumer preferred cassava

Effective use of genetic resources in a breeding program is incumbent on full understanding of the variability among phenotypic and genetic characters in available germplasm. Short of this critical information, genetic erosion occurs as breeders discard 'unwanted' genotypes during the various selection stage gates. Therefore, by characterizing germplasm, a breeder is informed of the need for conservation for future use and for routine use in identifying potential progenitors for population improvement.

Whereas phenotypic based assessment of genetic diversity is associated with many limitations including low polymorphism, environmental influences and physiological growth stage on phenotypic expression (Behera *et al*., 2012; Al-Naggar *et al*., 2020), phenotypic traits are important preliminary indicators of genetic diversity in a population and they provide realistic information crucial for characterizing genetic resources.

This study found considerable phenotypic variation for cooking traits in a C2 population, which could be benchmarked for selecting for consumer preferences during variety development. Moreover, it is important to evaluate this population in more agroecologies to identify broadly or specifically adopted genotypes for use in variety development or as progenitors for population improvement (Dossou-Aminon *et al*., 2015).

The heritability estimated for different cooking traits can be leveraged by cassava breeders for designing appropriate selection strategies to maximize gains for cooking traits during population improvement and variety development. Nevertheless, care should be taken when interpreting estimates of genetic gains using broad sense heritability since there is a likelihood of overestimating expected gains. This is especially plausible if the non-additive portion

of genetic variance (σ^2_G) is substantial, which cannot be exploited through phenotypic recurrent selection commonly used for cassava improvement (Ceballos *et al*., 2012).

The strong correlations among cooking traits present a unique opportunity for indirectly selecting for all cooking qualities using one trait, thereby contributing to optimized resource usage during phenotyping and selection in a breeding pipeline.

3.5 Conclusion

This study was undertaken to determine the extent of phenotypic variation for cooking qualities that have a high appeal among boiled cassava consumers. There was considerable phenotypic variation in the C2 population used in this study for all texture traits and water absorption (WAB30). All the traits were also important in discriminating the accessions.

There was evidence that variation observed in the cooking qualities among accessions is due to true inherent genetic variation in the population, which can be assessed further through genetic analysis. Such analysis of genetic diversity would be vital for the optimal use of available genetic resources for benchmarking consumer preferred textural traits during varietal improvement.

Knowledge of existing diversity also informs of the need to conserve for future reference to avoid genetic erosion and also the need for bolstering diversity through germplasm exchanges with other cassava breeding programs. There were strong correlations among texture traits and water absorption which can be harnessed for simultaneous improvement of all cooking traits, ensuring resource optimization.

Heritability estimates ranged from low to moderate, implying there was substantial genetic control of these traits. This find shows there is potential for genetic improvement of cooking traits though phenotypic selection. The study also found that choice of phenotyping tool for assessing cooking qualities can influence the variability detected. The texture analyzer generally gave better discrimination among accessions for boiled root softness than the penetrometer. However, more effort should be focused on developing higher throughput methods amenable for routine use in assessing cooking traits in cassava breeding populations.

CHAPTER FOUR

Assessing the NIRS prediction accuracy for cooking qualities in boiled cassava

4.1 Introduction

The advent of end-user focused breeding has raised awareness on the need to make urgent changes in the cassava variety development strategy to increase variety adoption (Ceballos *et al*., 2020). Among the changes required is the need to adopt rapid phenotyping tools to enable breeders rapidly select genotypes with desirable end-user trait thresholds at all stage-gates and to minimize genetic drain (Alamu *et al*., 2020).

To this end, a number of moderate and high throughput phenotyping, methods have been developed primarily targeting proximate composition of cassava roots including starch, dry matter and carotenoids (Abincha *et al*., 2021). Majority of these methods have leveraged the robustness of infrared spectroscopy (mid and near infrared; Garces *et al*., 2021) to recognize specific molecular vibrations strongly associated with specific chemical compounds in a sample matrix through absorption or scattering of infrared light (Burns & Ciurczak, 2007).

Near infrared spectroscopy (NIR) is projected to reduce over dependence on laborious and costly analytical techniques such as high-performance liquid chromatography (HPLC) and spectrophotometry across breeding programs that have adopted it (Mbanjo *et al*., 2022), (Hershberger *et al*., 2022).

NIRS relies on the absorption of part of the incident infrared light (IL) by compounds in sample and reflectance of the remaining IL (Ozaki *et al*., 2006). The intensity of reflected IL which corresponds to concentration of a compound in a sample, is then measured by detectors and plotted into an infrared signature.

However, much as NIR has showed great potential for compositional analyses in cassava, little effort has been directed towards sensorial traits which have high end-user demand. For instance, boiled cassava is a major product consumed in Uganda and consumers prefer varieties with a soft texture (Iragaba *et al*., 2020).

Cooking time is another trait which has implications on energy use during preparation of cassava. Phenotyping for texture is often a laborious, slow and costly process either involving use of a texture analyzer or sensory panel. The former is accurate but very slow, while the latter is very subjective and at times producing inconsistent data.

Cooking time has recently been evaluated using water absorption (Tran *et al*., 2020), but this too is a slow and tedious process. The problems associated with these methods have altogether handicapped cassava breeders from including these traits in selection indices particularly in early-stage gates (Iragaba *et al*., 2019), leading to the development of varieties deficient in end user traits.

Recently, Iragaba *et al*. (2019) demonstrated that a portable penetrometer device could provide mid-throughput capacity for assessing softness of boiled roots with moderate repeatability. However, use of this tool was still laborious given that samples had to be cooked prior to assessment.

Therefore, it is imperative that NIR be assessed for its prediction accuracy on cooking traits in cassava. The penetrometer and texture analyzer would be useful in collecting reference data for developing NIR predictions for boiled root texture properties. Predictions for NIRS are developed by regressing infrared spectra data from samples that are varying for a trait, with their respective phenotype data (Munawar *et al*., 2020).

Collection of good quality spectra is hinged on sample homogeneity though achieving this often requires extra processing steps such as sample size reduction which increase drudgery and cost of phenotyping. For instance, Ikeogu *et al*. (2017) demonstrated that cassava root homogenization increased NIR prediction accuracy for root dry matter (RDMC) and total carotenoid content (TCC), though intact roots-based models were also sufficient for screening purposes. Therefore, developing robust and accurate NIR predictions with minimal sample processing would be ideal for mainstreaming this technology for cooking quality assessment during selection processes.

In this regard, this study set out to assess the prediction accuracy of a portable ASD QualitySpec NIR for assessing cassava root softness and water absorption. Specifically, the study purposed to i) assess prediction accuracy of several models on softness and WAB in unprocessed

fresh cassava roots using both linear and non-linear regression approaches and ii) to assess the impact of different spectra preprocessing techniques on NIR predictions.

4.2 Materials and Methods

4.2.1 Materials

A population of 212 clones including 3 checks (main panel) that was part of the C2 population (Chapter 3) planted at two locations at Namulonge (Central Uganda) and Serere (Eastern Uganda) during the 2019/2020 season was used in this study. To maximize variation from the different environments, clones from the different environments were considered unique samples and phenotypes and spectra from both environments were not averaged, resulting in a total of 424 samples.

4.2.2 Sample preparation

Twelve months after planting, all roots in the plot were harvested and collated. Four healthy (CBSD root necrosis free) roots of similar size from each plot were selected. The roots were washed with clean water, dried with a cloth and taken for evaluation. Each of the roots from a plot was sliced transversely with a stainless-steel knife to remove the proximal and distal ends and straight away taken for spectra acquisition (Ikeogu *et al*., 2017).

4.2.3 Spectra collection

Using a portable NIR (ASD Quality SpecTrek, USA) device, spectra were collected from the samples. The proximal side of the sliced root was pressed against the infrared emitting window of the ASD NIR and one spectrum was collected from that root surface. Then the distal

end was also scanned to generate an extra spectrum (Hershberger *et al*., 2022) and a total of two spectra were collected per root and eight spectra per plot. Spectra were collected using a 1 nm resolution and each spectrum generated was an average of 50 sub-spectra collected over a period of 30 seconds.

4.2.4 Cooking quality evaluation

Water absorption (WAB)

After spectra acquisition, water absorption of roots was measured to infer cooking time as previously described in chapter three. The four roots (section 4.2.3) were sectioned into six half cylinders each of length 6 cm resulting in 18 pieces from each genotype. The pieces from different roots were mixed up and then randomly assigned to four pre-weighed (*Wo*) aluminum sample holders (Carlisle, Korea). The weights of the sample holders plus root pieces were recorded (*W1*) and samples were then placed in a pan containing clean boiling water over a gas flame.

The pan contained a large volume of water (4L per 400 – 600 g root) to minimize significant reductions in water temperature upon introduction of fresh roots. The pan was covered with an aluminum top, and water absorption was measured at 30-minutes cooking time. At 30 minutes of cooking, the sample holders were quickly removed from the pan, padded with a dry cloth to remove excess water and immediately weighed (W_{30}) with a 0.1 g precision digital scale (Metler Toledo, NewClassic MF, MS6001S, Switzerland). Water absorption (WAB) was determined as the percent increment in weight after cooking.

Softness and toughness measurements with penetrometer and texture analyzer

Following WAB measurements, root sections were randomly divided into two parts and nine pieces were used for softness measurement using a penetrometer (Model number: FHT-1122, Vetus Industrial Company Limited, Hefei, China) according to (Iragaba *et al*., 2019) and the other nine were assessed using a texture analyzer (TA. XT PlusC with 50kg load cell, UK) as previously described (Chapter 3).

4.2.5 Data analysis

4.2.5.1 Spectra preprocessing

Prior to model development, spectra were pre-treated according to (Ikeogu *et al*., 2017). Spectra were first transformed to log (1/R) using ViewSpec Pro software (ASD 2008), and the full Vis/NIRS wavelength range (350 –2500nm) was subjected to outlier filtration using Mahalanobis distances according to (Hershberger *et al*., 2022).

Spectra pre-treatments were then applied to correct for interferences on three segments of the wavelengths (350nm - 1000nm, 1001nm - 1800nm and 1801nm - 2500nm). The effect of two light-scatter correction methods – Standard Normal Variate (SNV) and Multiplicative Scatter Correction (MSC) were tested on several derivative and smoothing options.

Derivatives were generated using gap segment derivatization where options were given as (m, w, s, delta): where *m* indicated the order the derivative (1 indicated the first derivative, and so on), *w* indicated the gap size (the spacing of data points over which derivation is computed), *s*

indicated the segment size (the range over which points are averaged) and delta indicated the sampling interval (band spacing).

Baseline correction was also applied to correct for any shifts according to (Rinnan *et al*., 2009). These treatments were tested either singly or in combination on spectra and their effects on NIR predictions were compared to no treatment in each calibration set for softness, WAB and Area (toughness). Details of the working of each treatment are given;

Standard Normal Variate (SNV)*:* This correction minimizes scattering when the effective path length and baseline vary among samples of a data set and for granular or powdery samples or when the particle sizes vary among samples. SNV is usually applied first to correct the effects of the multiplicative interferences of scatter and particle size differences by removing the mean and

scaling to unit variance. This correction is given by:

$Si = \frac{So - Sv}{Sd}$, where *Si* = corrected spectrum, *So* = original individual spectrum measured by the NIR device, *Sv* = average value of the sample spectrum to be corrected and *Sd* = standard deviation of the sample spectrum. SNV correction was done using the *standardNormalVariate* function of *prospectr* package in RStudio (Stevens *et al*., 2015).

Multiplicative Scatter Correction (MSC): This is similar to SNV and it attempts to correct spectral imperfections due to particle size differences by linearizing each spectrum to an ideal or reference sample spectrum which in most cases is the average spectrum obtained from all the

data in the training set. The slope and offset of the sample spectra are adjusted to the ideal average spectra to give the MSC corrected spectrum. The process of MSC correction, considering the average spectrum as the reference, was done according to (Ikeogu *et al.*, 2017) as follows;

a. Reference spectrum calculation: $\dot{s}_j = \sum_{i=1}^{n}(Si, j)/n$

b. Using spectral responses in each spectrum to calculate a linear regression against the corresponding points in the reference spectrum: $S_i = a_i\dot{s} + b_i$

c. Subtracting the slope from the regression on the original spectrum and dividing with the offset values to obtain MSC corrected spectrum: $S_{i(MSC)} = (S_j - b_i)/a_i$, where S = spectral responses for all the wavelengths; $\dot{s}$ = average responses of all the training set spectra at each wavelength; Si = responses for a single spectrum in the training set; n = number of training spectra; ai and bi = slope and offset coefficients of the linear regression of the mean spectrum vector versus Sj spectrum. This was done using the *msc* function of *prospectr* package in RStudio (Stevens *et al.*, 2015).

Derivatives and Smoothing: Derivatization attempts to remove both additive and multiplicative effects in form of vertical offsets and linearly sloping baselines from spectra (Rinnan *et al.*, 2009). The basic method of derivation is finite difference where: the first-order derivation takes the difference between two values with a given gap size while second order derivative is then estimated by calculating the difference between two successive points of the first-order derivative spectra. In place of the basic derivative which is usually not feasible for most real

measurements due to noise inflation, the modified smoothing and derivative of the Norris- Williams approach is usually the preferred option:

a. Smooth the spectra. Average over a given number of points.

$$\text{X smooth, i} = \frac{\sum_j^m = -mX\text{org, i} + \text{j}}{2m + 1}$$

where *m* is the radius of the smoothing window centered on the current measurement point *i*.

b. Derive at each wavelength. For the first derivative take the difference between two smoothed values at a given gap distance and for the second-order derivative, take twice the smoothed value at point i and the smoothed value at a gap distance on either side:

$$Xi` = Xsmooth,i + gap – Xsmooth,i – gap$$

$$Xi`` = Xsmooth,i – gap – 2.Xsmooth,i + Xsmooth,i + gap$$

Importantly, derivatization should be done cautiously since it has a propensity for increasing noise in the transformed spectra and falsely improving the correlation between spectra and reference data during calibration development (Faculty *et al*., 2010). Derivatization was done using the *gapDer* function of *prospectr* package in R statistical software (R-studio, 2022).

SavitzkyGolay smoothing

This was done to lower the likely noise inflation accruing from the previous derivatization (Basile *et al*., 2021). This was implemented using the *savitzkyGolay* function of *prospectr* package in R. The magnitude of the polynomial order, frame size, and the derivative order were

altered to find a right combination that did not inflate noise while retaining as much spectral variance as possible.

4.2.5.2 Principal component Analysis (PCA)

Pre-treated spectral data were examined for outliers using PCA. Owing to the large size of the spectral data, PCA served to reduce the dimensionality of the data by tracing fewer linear combinations of variables that contribute more to making samples different from each other (Shao *et al*., 2009). PCA was done using scaled data with *prcomp* function in R.

Principal components explaining atleast 99% of the total variance were retained and k-means clustering was applied to the principal component scores (Ikeogu *et al*., 2017); the number of clusters was equal to the desired number of calibration individuals. The calibration set was constituted by picking individuals from the center of each cluster and leaving the rest for validation. This systematic approach of sampling ensured that the calibration set was truly representative of the population.

4.2.5.3 Spectra filtering with Mahalanobis distances

This was calculated based on a principal components matrix derived from a principal component analysis (PCA) of the spectral matrix. It is a powerful tool for defining sample boundaries and similarity indices between spectra. Mahalanobis distance (Global H value, GH) was used as a spectrum outlier tool to detect instrumental error, sample contamination and differences in sample handling (Basile *et al*., 2021). In essence, GH measures how far a given spectrum lies from

the multivariate mean of the calibration population in multidimensional space. GH was calculated as;

$$Hi = \sqrt{x_i - \mu)^T S^{-1}(x_i - \mu)}$$

Where x_i is a spectral observation from the mean vector μ and S is the covariance matrix of the spectral data. GH values were calculated for all spectra collected and any spectrum that presented a GH > 3 was considered a spectral outlier (Ikeogu *et al*., 2017). Subsequently, outlier spectra were removed from the spectral data set to minimize model distortion in downstream analyses.

Sample set division

Following pretreatment, the remaining spectra and corresponding reference data were divided into calibration (train set) and validation (test set) sets using an 80:20 ratio such that the calibration set contained 80% of the original observations and the validation set comprised 20% of the original observations. This was achieved using the Kennard-Stone sampling algorithm (*ken-Stone*) in *propectr* package in RStudio. To partition a data set, this method selects a subset of samples (selection set) from the original, which is representative of the range and distribution of the original set and includes samples on the boundary of the original.

The method starts by tracing two samples from the original set which are farthest apart using geometric distance. To add another sample to the selection set, the algorithm selects from the remaining samples another individual who has the greatest separation distance from the selected samples. The separation distance of a candidate sample from the selected set is the distance from the candidate to its closest selected sample. This most separated sample is then added to the

selection set and the process is repeated until the required number of samples, k, have been added to the selection set (Kennard & Stone, 1969).

4.2.5.4 Classification of phenotype data

Cluster analysis was done on each of the parameters (cooking traits) to group each into two classes and corresponding spectra for each sample in a particular class was added. This data was then used to train classification models.

4.2.5.5 Developing calibrations

Calibrations were developed using both linear and non-linear regression approaches. Linear methods such as partial least squares (PLS) assume a linear relationship of the modelled sample characteristic (e.g. proximate composition) as a function of infrared spectra data variations (Munawar *et al*., 2020).

PLS. This regression approach utilizes latent variables also called score vector to reduce the dimensionality of the spectra data and model the relationship between input and response variables (Bai *et al*., 2022). In this work, the regression first generated latent variables from spectra data set and then used them as new predictor variables for the various responses (traits). Prior to training the PLS model, tuning hyper parameters were set to a maximum of tune length of 20 principal components and a 10-fold cross validation scheme (partition the training set into ten parts and use 9 parts to train the model and the tenth for testing).

The cross validation done during training of PLS was important for internal assessment of model performance using statistics such as the coefficient of determination (R^2_p), the root mean square

error of prediction ($RMSE_p$) and the mean absolute error (MAE). The PLS model was fit for each trait using the *train* function of *caret* package in Rstudio (Kuhn *et al*., 2023).

Non-linear regression and classification approaches on the other hand assume a non-linear relationship between predictors and response. Non-linearity may increase due to changes to the measuring instrument such as repairs. Also changes to the sample such as particle size variation, deterioration and high concentration of the target constituent could increase non-linearity (Sexton *et al*., 2018). In this study, support vector machine (SVM) was tested for predicting cooking qualities of boiled cassava.

Support vector machine (SVM). is a supervised statistical learning algorithm, used as a nonlinear classification technique. It works with supervised learning models that analyze data used for classification and regression analysis, producing linear boundaries between object groups in a transformed space of the x-variables.

SVM has advantages in dealing with small samples, non- linear and high dimensional data. Also, model performance depends of the selection of kernel function in SVM models, and the commonly used Radial Bias Function (RBF) is used as kernel function (Sampaio & Brites, 2021).

Performance of non-linear models was assessed following cross validation and statistics of model performance were generated including the coefficient of determination (R^2_p) and the kappa coefficient of calibration (K_p).

4.2.5.6 Validation of models

This was done to assess the robustness of trained NIR models and their relevance with respect to the standard assays for texture and water absorption. The optimal trained models were used to predict the test set spectra and the coefficient of cross validation and (R^2_{cv}), root mean square error of cross validation ($RMSE_{cv}$) and ratio of performance to deviation (RPD) were used to assess reliability of quantitative predictions. The R^2_{cv} was determined as the correlation coefficient between the NIR predicted and actual measurements of a trait.

The RPD is a measure of the predictive ability of a model whereby values greater than 3 show that a model is excellent for making accurate and robust predictions but when less than 1.5, the model is deemed unusable (Li *et al*., 2022). It was measured as the standard deviation of predicted samples divided by the root mean square error of prediction ($RMSE_p$). On the other hand, cross validation accuracy and kappa coefficient were used to interpret model reliability for qualitative models (Classification).

4.2.5.7 Effect of feature selection on quantitative model accuracy

Highly multivariate data such as spectra tends to experience multicollinearity which may cause overestimation of model coefficients, and reduce reliability of model predictions (Chan *et al*., 2022). Several methods have been reported to handle multicollinearity and improve model reliability (Gómez *et al*., 2019). In this work, we investigated the effect of interval and variable selection based partial least squares (iPLS), on prediction model performance for cooking traits. Interval selection through PLS relies on local prediction models built from equidistant sub- intervals of the full spectrum region. This provides an overall picture of important information in

different sub-regions, provides focus on such regions while eliminating interfering regions. Interval selection was done using the *ipls* function of *mdatools* package in Rstudio.

4.3 Results

4.3.1 Trait Distribution in train and test sets

Peak force (Soft_T) ranged from 4180 to 28467.28-gram force (gF) in the full and train sets and 6776.13 to 18783.99 in the test set. The mean peak force was highest in the train set (12027.61 gF) and lowest in the test set (11223.29 gF) (Table 6). The coefficient of variation was 35.89, 37.77 and 25.25% in the full, train and test sets respectively. Peak force had a moderately positively skewed distribution in both train and test sets (Figure 8A) with majority of genotypes having force less than 15,000 gF.

Soft_p ranged from 0.38 to 6.59 kgfcm^{-2} in the full set, 0.41 to 6.59 in the train set and 0.38 to 3.5 kgfcm^{-2} in the test set. Mean softness was highest in the train set (2.35 kgfcm^{-2}) and lowest in the test set (1.73 kgfcm^{-2}). The coefficient of variation for softness was 50.04 in the full set, 48.38 in the train set and 49.27 % in the test set. Both train and test sets were highly positively skewed with most genotypes having softness scores less than 2 kgfcm^{-2} (Figure 8B).

Water absorption (WAB) ranged from 0.11 to 33.77 % in the full and train sets and 0.34 to 25.07 % in the test set (Table 6). The mean WAB was highest in the train set (8.04%) and lowest in the test set (6.27%). The coefficient of variation for WAB was 91.08 % in the full set, 91.56 % in the train set and 81.65 % in the test set. WAB was highly positively skewed in both train and test sets (Figure 8C) with majority of the genotypes scoring less 10% WAB.

Toughness (Area under the curve) ranged from 26269.9 to 283,016.1 in the full set and train sets and 56939.6 to 223,624.40 gF/s (gram force per second) in the test set. The mean toughness was highest in the train set 120408.9 (122897.30 gF/s) and lowest in the test set (110,555.9 gF/s). The CV was highest in the train set (42.87 %) and lowest in the test set (30.39 %). Toughness was also moderately positively skewed (Figure 8D) in both train and test sets, with majority of the genotypes having less than 100,000 gF/s.

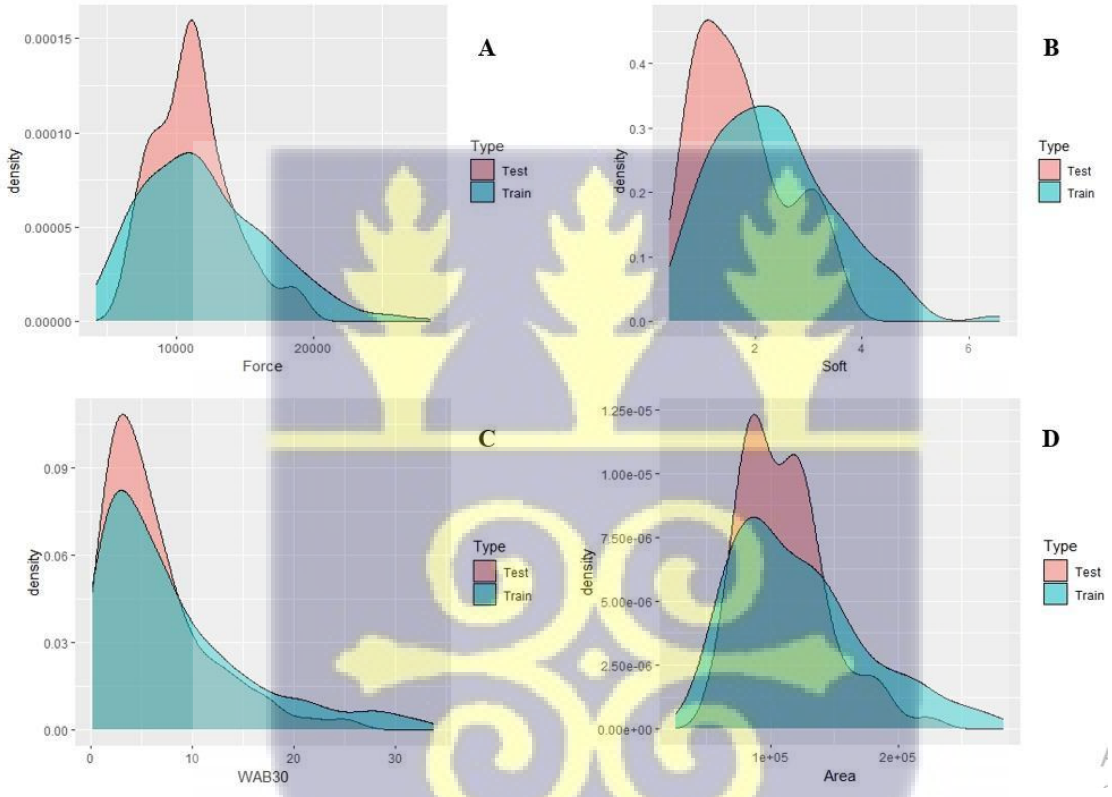


Figure 8: Distribution of train and test sets for Soft_T (A), Soft_p (B), water absorption (C) and Toughness (D). Force = softness assessed by texture analyzer (Soft_T); Soft = softness assessed by penetrometer (Soft_p); WAB30 = water absorption assessed after 30 minutes of boiling (WAB); Area = Area under the texture extrusion curve that is equivalent to sample toughness; Train = Training set used to calibrate the NIRS model; Test = test set used to cross validate the calibrated model.

Table 6: Summary statistics for full, training and validation sets for water absorption (WAB30), Softness by texture analyzer (Soft_T), softness by penetrometer (Soft_p) and Toughness (Area under the curve)

Trait	Set	N	Min	Max	Mean	SEM	CV%
Soft_T	Full	367	4180.01	28467.28	11863.24	222.31	35.89
WAB	Full	363	0.11	33.77	7.69	0.37	91.08
Soft_p	Full	366	0.38	6.59	2.22	0.06	50.04
Toughness	Full	367	26269.9	283016.10	120408.9	2591.44	41.23
Soft_T	Train	292	4180.01	28467.28	12027.61	265.85	37.77
WAB	Train	290	0.11	33.77	8.04	0.43	91.56
Softp	Train	292	0.41	6.59	2.35	0.07	48.38
Toughness	Train	293	26269.92	283016.10	122897.30	3078.08	42.87
Soft_T	Test	74	6776.13	18783.99	11223.29	327.24	25.25
WAB	Test	73	0.34	25.07	6.27	0.59	81.65
Soft_p	Test	74	0.38	3.5	1.73	0.09	49.27
Toughness	Test	74	56939.68	223624.40	110555.9	3906.65	30.39

4.3.2 Principal component analysis of spectra data

Principal component analysis showed variation among genotypes and moderate differentiation between trials (Fig 9A). Also, there were potential outliers detected in the spectra, necessitating their removal prior to downstream model building. The first two principal components (PCs)

accounted for a total of 92.7% of spectra variation among the genotypes. Generally, the variation exhibited in the two locations was promising for linear calibration.

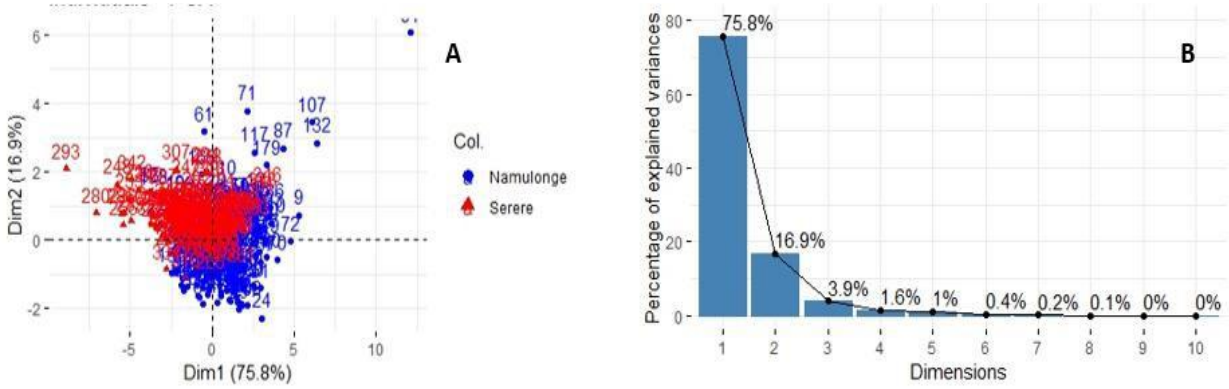


Figure 9: Principal component analysis plot (A) and scree plot showing top 10 PCs and variance explained by the first 10 PCs (B).

4.3.3 Cluster analysis

Cluster analysis of spectra data based on k-means further confirmed the existence of two clusters in the data (Figure 10). The first cluster had a total of 201 accessions with an average silhouette width (si) of 0.35 whereas the second cluster consisted of 164 accessions with an average si of 0.39. Cluster one comprised of 155 individuals from Namulonge and 46 from Serere whereas cluster 2 comprised of 74 individuals from Namulonge and 90 from Serere. However, performance of the clusters for the different traits did not differ.

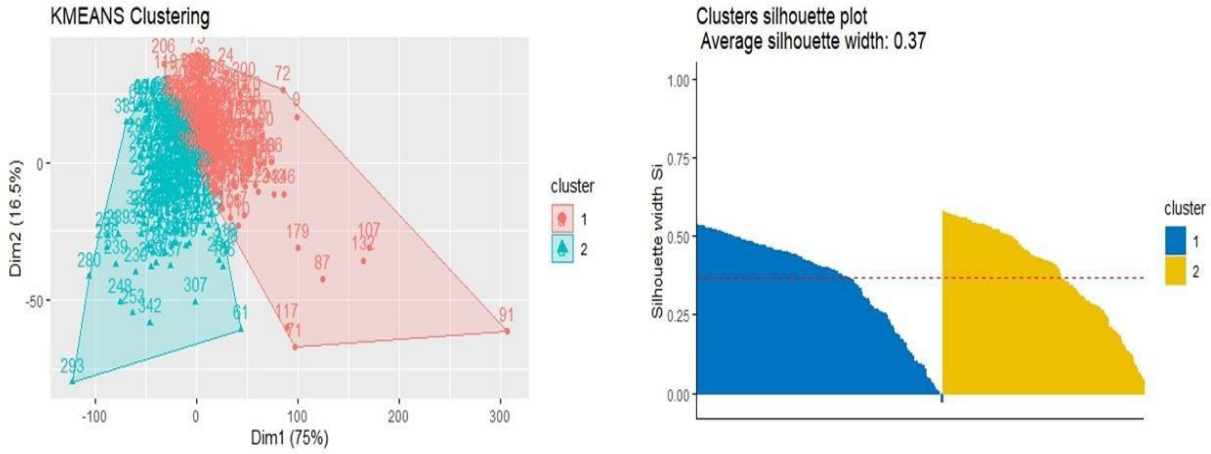


Figure 10: Cluster plot of spectra data showing 2 clusters and a silhouette plot validating the cluster analysis.

Outlier removal

After filtering spectra using mahalanobis distances, a total of 59 samples were removed from the data and 365 were retained for further model calibration. Retained spectra were further processed using a series of mathematical treatments. Further outlier removal was done on measured phenotypes using Grubb's test (Urvoy & Autrusseau, 2014) and 367 samples were retained for Soft-T and Area whereas 363 and 366 samples were retained for WAB30 and Soft_p.

Effect of pretreatments on spectra quality

In order to select an appropriate pretreatment for cleaning the spectra prior to their use, treated and untreated spectra were plotted for visual inspection. A lot of emphasis has been put on developing effective spectra pretreatment methods to reduce spectral interferences and simultaneously increase signal. This has been shown to increase efficiency of multivariate models (Watanabe *et al*., 2018). To understand the effect of the pretreatments on spectra quality, each

treatment application was proceeded by a plot of the spectra to identify possible signal enhancement or loss.

Raw and replicated spectra (Figure 11A) were first averaged to get one spectrum per sample (Figure 11B), then *mahalanobis* distance-based spectra filtration was done to eliminate outliers. The raw averaged spectra (Figure 11B) showed prevalence of water associated overtones (1000 – 2000 nm region) in the roots, typical of fresh cassava. Pretreatments were then applied to the filtered spectra either singly or in combination with others.

Application of the standard normal variate (SNV) resulted in minimization of scattering effects (Figure 12A) across the spectrum while the combination of SNV with the first derivative noticeably resolved peaks in the 700 – 2200nm spectra region but blurred the signal at both tail ends of the spectrum (Figure 12B).

A combination of gap derivatization (m=1, w=11, s=3) smoothing with SNV gave a similar scatter correction to that of SNV (Figure 12C) whereas gap derivatization implemented with a second derivative (m=2), a gap size of 9 (w=9), a segment size of 5 (s=5) and in combination with SNV resulted in higher peak resolution across the spectrum (Figure 12D).

Application of baseline correction and in combination with Savitzky-Golay (SVG) filter and SNV also resolved peaks across the spectrum but also increased signal level (Supp. Fig. 6B). Use of multiple scatter correction (MSC) minimized scattering effects (Supp. Fig. 6C) and when

combined with SVG, peaks were resolved clearly though signal intensity was reduced (Supp. Fig. 6D).

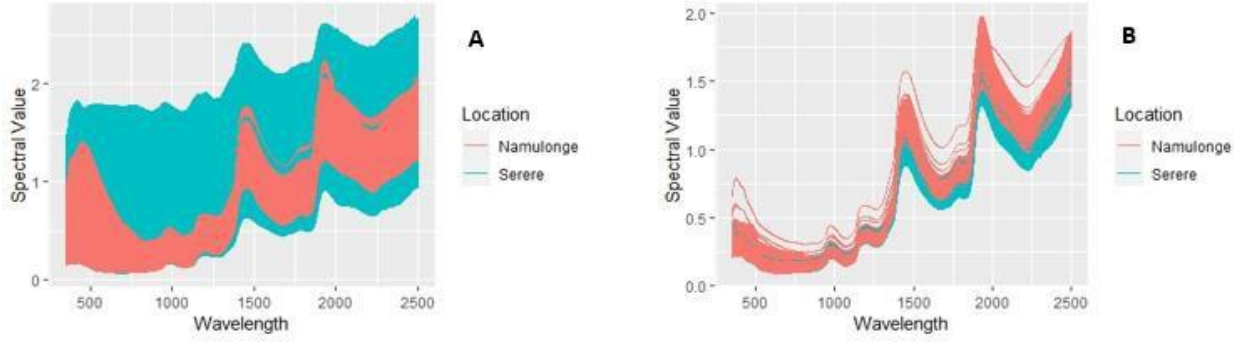


Figure 11: Raw and replicated spectra (A) and average spectra (B) for both Namulonge and Serere.

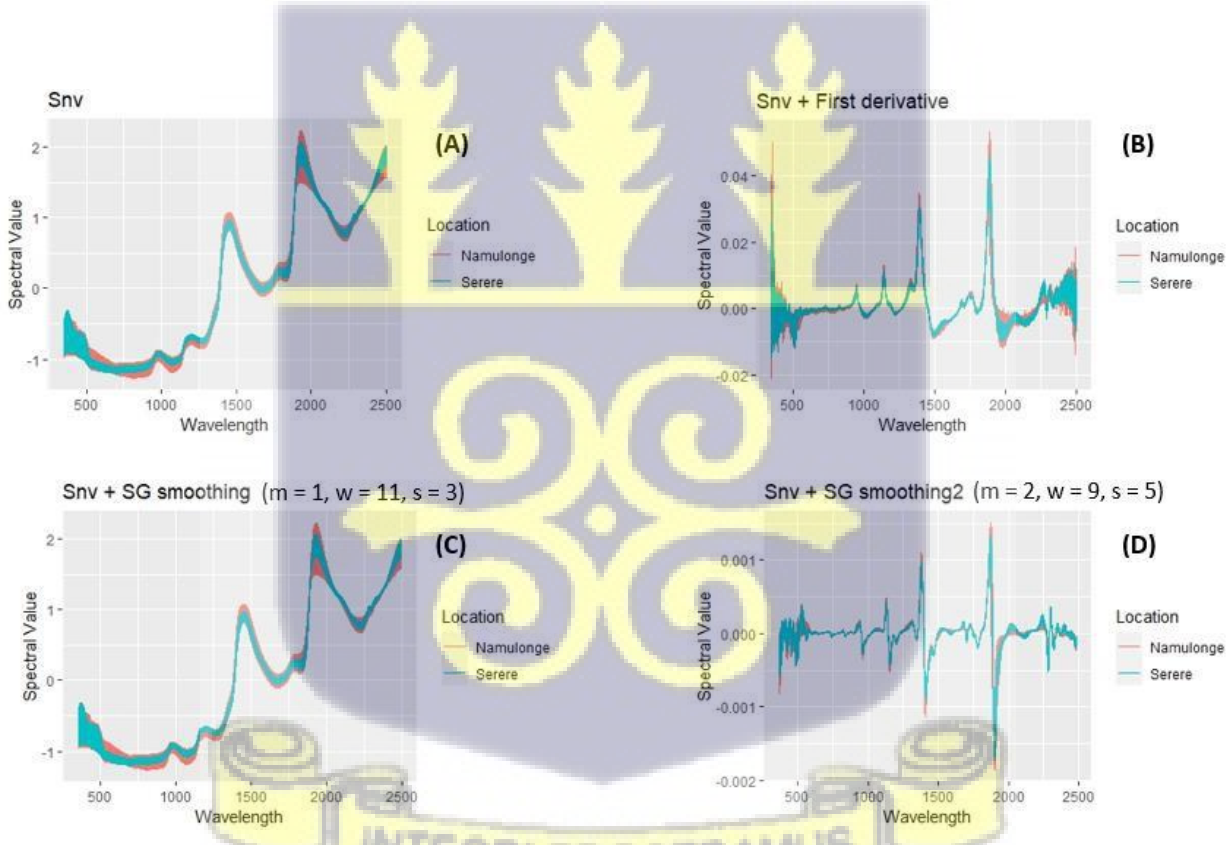


Figure 12: Effect of different Spectra preprocessing techniques including standard normal variate (A), snv + first derivatization (B), snv + savitzky-Golay smoothing 1 (C) and 2 (D) on spectra quality.

Effect of pretreatments on cumulative variance of principal components

This was used as a confirmative step in selecting the best pretreatment based on the cumulative variance retained by the treated spectra in comparison with the untreated. Principal component analysis of spectra data is important for tracking changes in variability within the population and factors contributing to such changes.

The first two principal components from the raw untreated spectra explained up to 91.51% of the cumulative spectral variance among samples whereas application of standard normal variate (SNV) reduced the cumulative variance to 76.96% (Figure 13). This implies that multiplicative interferences due to light scatter and particle size differences accounted for majority (14.55%) of the initial 'noisome' spectral variation.

The combination of baseline correction and SNV further reduced the cumulative variance (2 PCs) to 75.32% implying that baseline shift accounted for 1.64% of initial spectral variation. The combination of baseline correction, gap segment derivatization (derivative =2, gap size = 19, segment size = 9, sampling interval = 7) and SNV also lowered the cumulative variation to 69.18%.

Application of SavitzkyGolay (SG) smoothing treatment (polynomial order = 1, window size = 21, differentiation order = 1, sampling interval = 3) singly on spectra also lowered the cumulative variance (65.4%) of the spectra. When SG was combined with SNV, the cumulative variance of 2 PCs slightly increased to 67.96% and when baseline, SG and SNV were combined, the cumulative variance increased to 71.33%.

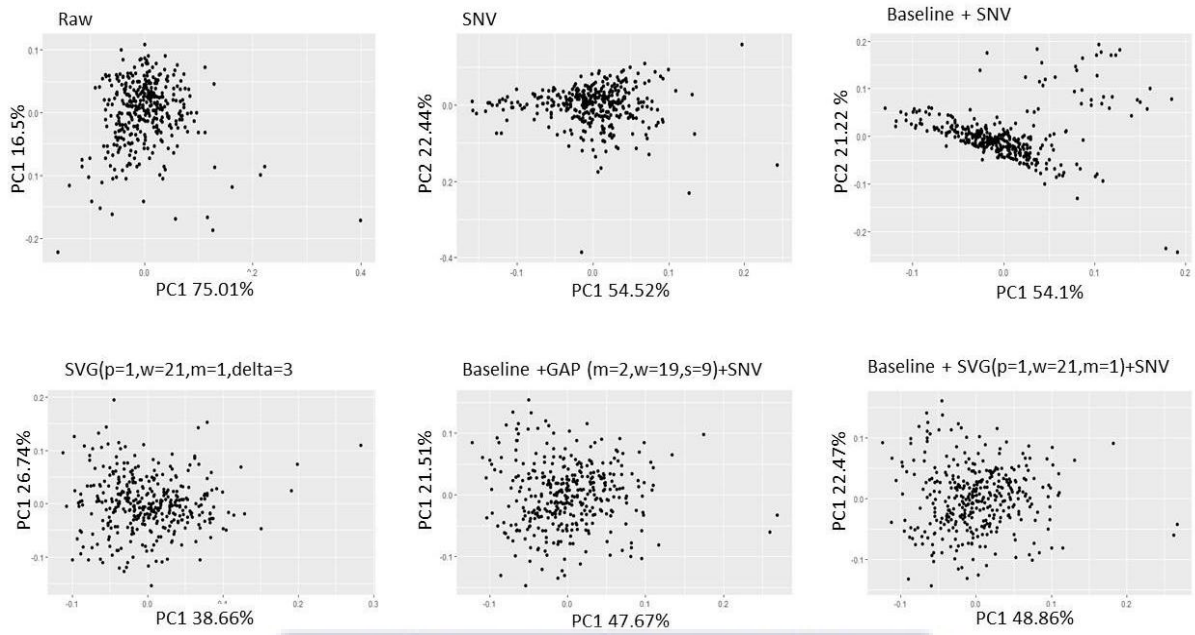


Figure 13: Effect of spectra pretreatments on cumulative variance explained. Raw = untreated spectra; SNV = spectra treatment with standard normal variate; Baseline + SNV = baseline correction in combination with SNV; SVG = SavitzkyGolay smoothing; Baseline + GAP = Baseline correction in combination with gap segment derivatization; Baseline + SVG = Baseline correction in combination with SavitzkyGolay smoothing.

4.3.4 Effect of spectra pretreatments on Linear calibration model statistics

Linear prediction accuracies generally remained low across all traits even after applying several spectra pretreatments. However, effect of spectra pretreatment on calibration accuracy varied with each trait (Table 7).

For Soft_p, the combination of baseline correction (BC), gap segment derivatization (GAP; m=1, w=5, s=3, delta.wav =3) combined with standard normal variate (SNV) gave the highest calibration accuracies for Soft_p (R^2_p = 0.39; R^2_{cv} = 0.39) (Figure 14). The same treatment gave

the highest prediction accuracy for softness measured by texture analyzer (Soft_T; $R^2_p = 0.31$; $R^2_{cv} = 0.35$).

However cross validation accuracy for Soft_T was highest for the treatment combining baseline correction, SG, GAP and SNV ($R^2_{cv} = 0.39$). The combination of BC and SNV gave the highest prediction accuracy for water absorption (WAB30) ($R^2_p = 0.25$; $R^2_{cv} = 0.27$) while BC in combination with SNV and GAP had the lowest prediction accuracy ($R^2_p = 0.28$; $R^2_{cv} = 0.14$).

Likewise, BC combined SNV gave the highest R^2_p for area under the curve ($R^2_p = 0.26$; $R^2_{cv} = 0.18$) while BC+SVG+SNV had the lowest accuracy. Prediction accuracy for gradient was highest under treatments of BC combined with SNV ($R^2_p = 0.13$; $R^2_{cv} = 0.19$). Overall, calibration and cross validation accuracies were highest for Soft_p and lowest for gradient.

Table 7: Effect of spectra pre-treatments on NIR calibration and cross validation statistics of partial least squares regression (PLSR) for cooking traits in cassava

Calibration					Cross validation						
Trait	**Treat**	**R^2p**	**RMSE**	**MAE**	**RMSE**	**R^2_{cv}**	**RPD**	**RPIQ**	**CCC**	**Bias**	**SE**
WAB	1	0.24	6.98	5.13	7.43	0.24	1.14	1.21	0.4	-1.01	7.49
Soft_p	1	0.33	0.95	0.74	0.93	0.36	1.25	1.78	0.54	0.03	0.94
Soft_T	1	0.33	3679.89	2857.84	4258.14	0.12	1.02	1.43	0.32	335.46	4288.02
Toughness	1	0.25	44798.44	36227.21	51695.02	0.09	1.02	1.15	0.27	4392.06	52100
Stiffness	1	0.31	350.94	221.27	910.05	0.02	1.02	0.39	0.05	-81.94	916.44
WAB	2	0.23	7.01	5.18	7.07	0.27	1.17	1.17	0.41	0.63	7.12
Soft_p	2	0.38	0.88	0.69	1.13	0.36	1.22	1.63	0.46	-0.24	1.14
Soft_T	2	0.27	3612.78	2891.28	4338.98	0.39	1.23	1.3	0.46	-975.55	4369.43
Toughness	2	0.19	44418.01	35687.63	51342.88	0.3	1.17	1.25	0.39	-10528.2	51700
Stiffness	2	0.26	350.27	227.65	915.29	0.08	1.03	0.39	0.09	-117.39	921.72
WAB	3	0.25	7.15	5.23	6.32	0.27	1.18	1.03	0.42	-0.29	6.36
Soft_p	3	0.35	0.94	0.73	0.98	0.16	1.08	1.54	0.34	0.8	0.99
Soft_T	3	0.29	3865.84	3006.51	3723.75	0.21	1.19	1.67	0.39	-190.79	3749.89
Toughness	3	0.26	45725.24	35836.36	45750.95	0.18	1.09	1.4	0.35	-6212.89	46072.01
Stiffness	3	0.13	461.68	272.72	275.56	0.19	1.1	1.24	0.24	29.66	277.49

WAB	4	0.28	6.57	4.88	9	0.14	1.08	0.94	0.29	-0.35	9.07
Soft_p	4	0.37	0.95	0.75	0.91	0.26	1.14	1.71	0.49	-0.01	0.92
Soft_T	4	0.32	3597.05	2899.52	4203.63	0.29	1.18	1.23	0.45	-576.29	233.13
Toughness	4	0.23	4476284	36882.6	50066.89	0.16	1.09	1.11	0.31	-4502.86	50418.25
Stiffness	4	0.18	410.85	252.09	536.55	0.05	1.03	0.63	0.11	-8.39	540.32
WAB	5	0.21	6.86	5.13	7.84	0.29	1.17	0.96	0.41	-1.46	7.89
Soft_p	5	0.36	0.95	0.74	0.92	0.28	1.15	1.78	0.5	0.14	0.93
Soft_T	5	0.33	3732.28	2945.93	3773.77	0.19	1.07	1.38	0.41	486.27	3800.25
Toughness	5	0.25	46469.11	36493.66	45767.09	0.13	1.03	1.53	0.31	6632.32	46100
Stiffness	5	0.24	372.07	239.85	892.87	0.03	1.02	0.39	0.06	-20.94	899.14
WAB	6	0.29	6.69	5.03	8.12	0.16	1.08	0.99	0.33	0.21	8.18
Soft_p	6	0.39	0.88	0.7	0.97	0.39	1.28	1.59	0.57	0.02	0.98
Soft_T	6	0.31	3617.6	2846.21	4043.66	0.35	1.24	1.37	0.49	-288.94	4072.03
Toughness	6	0.18	46020.09	37452.82	51137.36	0.16	1.1	1.22	0.27	104.15	51500
Stiffness	6	0.19	412.96	246.04	540.46	0.04	1.03	0.65	0.09	-35.46	544.25

Note: Soft_p: softness measured by penetrometer; WAB: Water absorption measured at 30 minutes after boiling; Soft_T: softness measured by texture analyzer; Treat: spectra pretreatment method; 1 = baseline correction + Savitzky-Golay (SVG) smoothing (p = 1, w = 21, m = 1, delta.wav = 3) + standard normal variate (SNV); 2 = Base + SVG(p = 1, w = 21, m = 1, delta.wav = 3) + GAP(m=1,w=11,s=9,delta.wav=3) + SNV; 3 = Baseline correction (Base) + SNV; 4 = Base + SNV + GAP(p = 1,w=7,s = 9, delt =3); 5 = Base + SVG(p=2,w=39,m=1, delta=2) + SNV; 6 = Base + GAP(m=2,w=21,s=15,delt.w=7) + SNV.

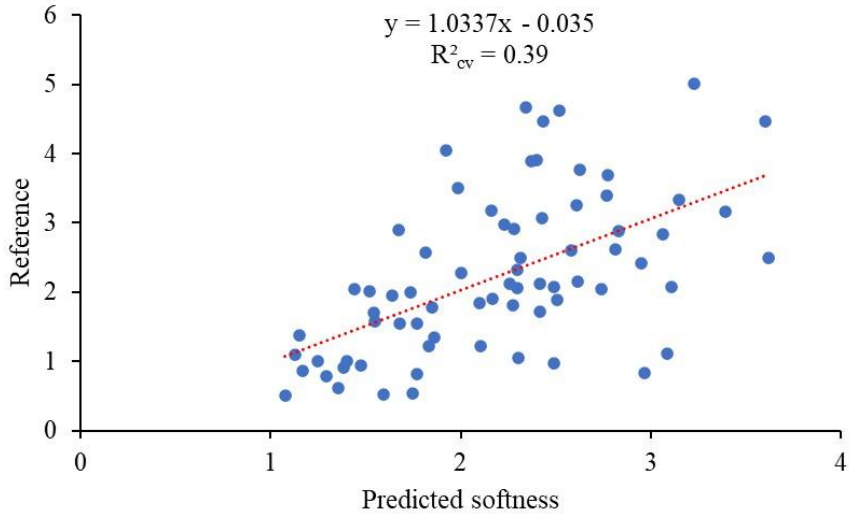


Figure 14: Correlation between predicted and reference values for softness

Effect of feature selection on partial least squares (PLS) model performance of linear models

There was slight increment in overall prediction accuracy for softness (penetrometer, R^2_{cv} = 0.43) when 4 intervals (571 variables) were selected for regression using iPLS (Figure 15). However, model performance did not change for WAB, Soft_T and Toughness using the same method and cross validation prediction accuracies remained low (R^2cv < 0.2) for these traits.

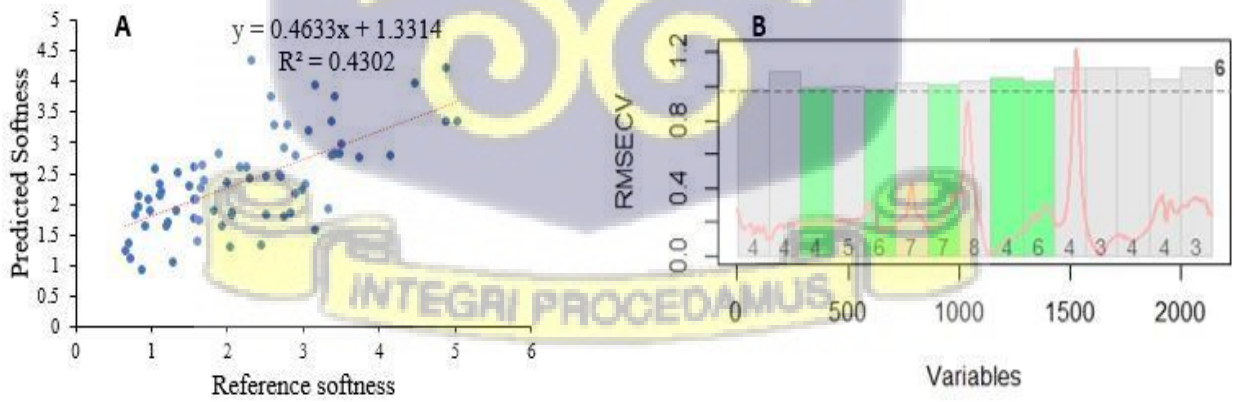


Figure 15: Cross validation prediction accuracy for softness after variable selection iPLS (A) and the intervals (in green) that were most explanatory for softness (B).

4.3.5 Effect of spectra pretreatments on calibration model statistics for cooking traits in cassava using support vector machines (SVM).

Support vector machines (SVM) were used to establish the non-linear relationship between spectra and the respective cooking trait binary classes. Overall, there was improvement in prediction accuracy across all traits when SVM was used for binary classification (Table 8).

Prediction accuracy for Soft_p was highest (R^2_p = 0.78, k_p = 0.39; R^2_{cv} = 0.69, k_{cv} = 0.19) when a combination of GAP, SVG and SNV treatments were used. Using the same treatment, similar prediction accuracy was achieved for Soft_T (R^2_p = 0.77, k_p = 0.38; R^2_{cv} = 0.69, k_{cv} = 0.19). However, in both instances, the cross-validation kappa coefficient (k) remained low.

For Toughness, the highest prediction accuracy was attained using a combination of SNV and GAP (R^2_p = 0.69, k_p = 0.19; R^2_{cv} = 0.74, k_{cv} = 0.28). High prediction accuracy was attained for WAB when SNV was combined with GAP (R^2_p = 0.7, k_p = 0.26; R^2_{cv} = 0.78, k_{cv} = 0.46. This treatment also gave the highest kappa coefficient among all traits, thus increasing reliability of the predictions made. Overall, the reliability associated with WAB was higher than that for other traits.

Table 8: Effect of spectra treatments on classification accuracy using support vector machines

Trait	Treat	R^2_p	Kappa_c	R^2_{cv}	Kappa_cv
Soft_p	GAP (m=1, w=19, s=1, d=3) + SVG (p=1, w=21, m=1) + SNV	0.78	0.39	0.69	0.19
Soft_T	GAP (m=1, w=19, s=1, d=3) + SVG (p=1, w=21, m=1) + SNV	0.77	0.38	0.69	0.19
Toughness	GAP (m=1, w=19, s=1, d=3) + SVG (p=1, w=21, m=1) + SNV	0.68	0.20	0.72	0.12
WAB	GAP (m=1, w=19, s=1, d=3) + SVG (p=1, w=21, m=1) + SNV	0.67	0.19	0.64	0.18
Soft_p	SVG (m=1, w=11, s=3) + SNV	0.75	0.29	0.71	0.14
Soft_T	SVG (m=1, w=11, s=3) + SNV	0.70	0.26	0.72	0.19
Toughness	SVG (m=1, w=11, s=3) + SNV	0.65	0.11	0.72	0.17
WAB	SVG (m=1, w=11, s=3) + SNV	0.64	0.10	0.75	0.39
Soft_p	SNV+GAP	0.76	0.34	0.69	0.19
Soft_T	SNV+GAP	0.71	0.30	0.67	0.11
Toughness	SNV+GAP	0.63	0.04	0.67	0.15
WAB	SNV+GAP	0.70	0.26	0.78	0.46
Soft_p	SNV+GAP (m=2, w=19, s=9, d=7)	0.69	0.13	0.71	0.19
Soft_T	SNV+GAP (m=2, w=19, s=9, d=7)	0.71	0.28	0.65	0.09
Toughness	SNV+GAP (m=2, w=19, s=9, d=7)	0.69	0.19	0.74	0.28
WAB	SNV+GAP (m=2, w=19, s=9, d=7)	0.67	0.33	0.63	0.23

Note: GAP - Gap segment derivatization; SVG – Savitzky-Golay filter; SNV – Standard Normal Variate

4.4 Discussion

Near Infrared Spectroscopy (NIRS) is widely projected to revolutionize crop and product phenotyping while minimizing costs associated. This would be particularly important for crop breeding programs that seek to increase rate of genetic gains for end-user preferred traits by optimizing selection.

NIRs has the potential to increase selection efficiency by providing means by which crop breeders can speedily and accurately phenotype large breeding populations for specific root quality traits. The tool has been tested on cassava root proximate composition with promising outcomes, but no comprehensive study had been done to test its efficiency in predicting cooking traits.

Therefore, this study was conducted to evaluate the accuracy of NIRS in predicting cooking qualities of boiled cassava from fresh roots. The use of fresh unhomogenized roots was intended to reduce sample preparation time, which could easily limit the use of NIR in field conditions. The success of NIR predictions is highly dependent on the accuracy and range of traits to be predicted as well as the quality of infrared signal (Beć *et al*., 2021; Hershberger *et al*., 2022).

The range observed for soft_T (softness by texture analyzer) was wider than that reported by Iragaba *et al*. (2024) implying there was sufficient variability for training the models. The range of Soft_p (softness by penetrometer) in this work was higher than that reported by Uchendu *et al*. (2021) and the range of WAB (water absorption) was higher than that reported by Tran *et al*. (2020).

Importantly, the ranges of all traits in the training set were larger than those in the test set, which is important for building reliable models (Xia *et al*., 2019). All traits were positively skewed but WAB and Soft_p were more positively skewed than soft-T and Area. Reference data distribution may negatively impact performance of PLS regression models especially in cases of high skewness, but less with moderately distributed data (Xia *et al*., 2019).

Principal component analysis (PCA) revealed minimal stratification with respect to location (Figure 9A). This implies that based on NIR spectra, grouping of samples may not reflect location influence on performance of genotypes. In contrast, when PCA clearly stratifies samples based on their spectral and chemical characteristics, then classification of phenotypes using NIRS improves (Munck et *al*., 2005). This notion was proved in a study where PCA was fundamental in the classification of soft and hard wood tree species based on differences in infrared signal corresponding to variation in chemical composition (Toscano *et al*., 2017).

The large spectral variance explained by the first two principal components in this study (Figure 9B) was expected to complement quantitative NIR models. However, large spectral variance may not be associated with variation of the measured phenotype (Bobelyn *et al*., 2010), which could have limited the extent of association between spectra and actual cooking traits during calibration building.

The tendency of spectra data to cluster may be rather important for building NIRs-based classification models by determining the optimal number of clusters forming from a spectral data set (Crase *et al*., 2021). This is underpinned by a fundamental principle that objects with more similarity and less distance in a variable space are likely to form a group

(Amirvaresi & Parastar, 2023). When cluster analysis of spectra data was done, two relatively distinct groups were formed (Figure 10) better than those observed in the PCA.

However, the groups formed were heterogeneous with each containing individual samples from the two locations. When the phenotype statistics from the two groups were evaluated, there was no difference in performance of the two groups (Supp. Table 1). This indicated that the observed marginal spectral variation was not associated with differences in phenotypes of the samples. This finding hinted on the likelihood of not achieving effective differentiation of cooking traits into binary classes using NIRs.

Prior to their use in model development, the spectra were further pretreated with the purpose of removing unwanted variation that could interfere with model performance (Xia *et al*., 2019). Such noisome variation may accrue due to non-linearity caused by light scatter (Rinnan *et al*., 2009) and also instrumental artefacts such as random noise, changes in lamp intensity, and detector response (Mokari *et al*., 2023).

Light scatter could be additive in which case it is reflected as baseline offset in spectra or multiplicative in which it is depicted as scaling of the entire spectrum by a given factor (Huang *et al*., 2019). Among all treatments used in this study, the standard normal variate (SNV) normalization of spectra resulted in the least distortion of the original spectrum (Figure 12A) and retention of most spectral variance (Figure 13).

This is consistent with findings by (Wajizah *et al*., 2020) who also found SNV and MSC to retain most of the spectral variance after treatment. Savitzky-Golay smoothing retained the lowest cumulative variance which is possibly due to erosion of some signal along with

noise. However, none of the treatments tested resulted in clear stratification of the samples and thus not promising for high classification efficiency NIR model building.

To clearly understand the implications of spectra pretreatment on calibration model performance, select pretreatments were tested on both linear and non-linear models. For WAB, no clear differences were found among the linear models generated with different spectra pretreatments with coefficients of calibration (r^2_c) ranging from 0.21 – 0.29 and coefficient of cross validation (r^2_{cv}) ranging from 0.14 – 0.29. These findings are similar to those reported by Davrieux *et al*. (2022) who concluded that linear regressions were not capable of relating sample spectra to actual WAB measurements.

Moreover, the major difference between the two studies is that Davrieux *et al*. (2022) used fresh ground cassava (homogenized) for collecting spectra whereas this study used whole root for spectra acquisition. Sample homogenization has previously been shown to play a role in minimizing noise due to light scattering and differences in effective path length, thereby increasing NIR model performance (Ikeogu *et al*., 2017; Alamu *et al*., 2020; Abincha *et al*., 2021). However, extra sample preprocessing steps could negatively impact the practical use of NIRs in out-of-lab scenarios and also lower throughput of NIRs protocols (Prieto *et al*., 2017).

Linear predictions for Soft_p (softness by penetrometer) were also low but generally higher than those for WAB. Using different spectra treatments, coefficients of calibration (R^2_c) ranged from 0.33 – 0.39 and coefficient of cross validation (R^2_{cv}) ranging from 0.16 – 0.39. Similar findings were reported by (Fatumah Namakula *et al*., 2023) who used homogenized root samples (mashed) for spectra data collection. The similarity of the findings from

the two studies using different sample types (whole vs mashed) suggests that homogenization may not improve predictions for boiled cassava root softness but rather increase sample processing time.

When feature selection was used to isolate the most informative spectral ranges associated with Soft_p, cross validation accuracy increased slightly (R^2_{cv} = 0.43; Figure 15). This indicates that there was increased linearity between selected spectra variables and Soft_p. However, feature selection did not improve predictions for other cooking traits. In other studies, feature selection has generally been found to reduce multicollinearity, which is common in NIR spectra, thereby improving NIR predictions (Gómez *et al*., 2019; Sun *et al*., 2021).

Similar to Soft_p, Soft_T (softness assessment by texture analyzer) linear predictions were low across spectra pretreatments (R^2_c = 0.27 – 0.33; R^2_{cv} = 0.12 – 0.39). This result was consistent with that reported by (Mehgar *et al*., 2019) who used a smaller population (17 varieties) and whole root portions for spectra acquisition.

Predictions for Toughness remained low across all spectral pretreatments (R^2_c = 0.18 – 0.26; R^2_{cv} = 0.09 – 0.3). Prior to this work, no study had reported NIRS predictions on this parameter, which is also linked to end-user preferences for boiled cassava. Overall, the poor linear predictions observed across all cooking qualities indicate lack of linear relationship between sample spectra and the cooking traits (Davrieux *et al*., 2022). This is resonated across the various models with different spectra treatments which failed to achieve an RPD greater than 1.5 (Table 7).

When classification-based prediction using support vector machines (SVM) was applied, calibration accuracy for WAB increased across spectra treatments but the agreement between the model and the classes used to train it remained low. Cross validation accuracy also increased across spectra treatments but the agreement between predicted and actual classes was generally low. These findings differ from those reported by Davrieux *et al*. 2022) who concluded that SVM was unable to classify water absorption. This study showed that a combination of SNV and GAP treatments achieved a moderately reliable classifier ($kappa_{cv}$ = 0.46).

However, the reliability of classifications for other cooking traits (Soft_p, Soft_T, Toughness) remained low ($kappa_{cv}$ = 0.09 - 0.28) which implies the observed classification accuracy was majorly due to chance and the models were still unreliable. Overall, classification-based models were more promising than quantitative linear models for predicting cooking traits in cassava. These findings also showed that there was a possibility of improving quantitative models through feature selection.

Further, this study also found that spectral pretreatments had varying effects on predictions for different traits. Whereas SNV combined with GAP gave the highest classification accuracy and reliability (R^2_{cv}) for WAB, predictions for other traits remained relatively similar even when different combinations of treatments were applied. Similar findings are corroborated by other studies (Hershberger *et al*., 2022; Mbanjo *et al*., 2022) proving that no single pretreatment is effective for different traits.

For further improvements in predictions, quality of reference data and sub sampling for cooking traits could be improved to increase likelihood of finding strong associations with

spectra data. This is because poor quality references and subsampling greatly reduce the linear association between spectra and references (Fystro, 2002; Cécillon *et al*., 2009). Additionally, more populations with variations in cooking qualities should be explored to increase training and validate set sizes, since prediction models have been shown to respond to population size in other studies.

4.5 Conclusion

The phenotyping bottle neck around cooking qualities in cassava puts a limit on their inclusion in selection indices particularly in early-stage gates. This prevailing challenge often cascades into low genetic gains for this end-user linked traits and ultimately leads to development of varieties with low consumer appeal. This study sought to test the prediction accuracy of NIRs for cooking qualities in boiled cassava as a cost-effective means of rapidly assessing these traits in cassava breeding programs that cannot afford to more costly methods.

Both quantitative and qualitative models were tested and the findings presented herein suggest that quantitative NIR predictions cannot yet be used for phenotyping cooking traits whereas qualitative classification models could be adopted with improvements in performance. For WAB, the classification accuracy and reliability attained for the support vector classifier could be useful in early selection stages for separating slow and fast water absorbing cassava genotypes in a breeding population.

Moreover, the minimal sample preparation used in this work (whole roots) is promising for out-of-lab application of the hand-held ASD Quality SpecTrek NIRs in phenotyping multi-location experiments which are often remote and devoid of equipped laboratories. In addition, use of whole roots in the NIR protocol would significantly reduce time taken in

phenotyping and increase the turnover of genotypes phenotyped in a day. Overall, NIR is a promising tool for phenotyping cooking traits, but several modifications should be made to increase reliability of the predictions made.

CHAPTER FIVE

Genome wide association analysis and genomic predictions for texture and cooking time in cassava roots

5.1 Introduction

Genetic loci and markers linked to important traits in cassava have been discovered through genome wide association studies (GWAS), a linkage disequilibrium dependent analysis that associates unrelated alleles (Burghardt *et al*., 2017). In GWAS, genetically diverse populations (assembled or developed through hybridization) are phenotyped and genotyped to provide data sets for linking phenotypic variation to genetic variation (Xiao *et al*. 2022) using linear or non-linear regression models. Several factors affect the power of GWAS (i.e. association study) including number of markers used, heritability of the trait, size of the population and diversity therein (Burghardt *et al*., 2017).

Also, association analysis depends on the quality of genotyping and phenotyping (Amyotte *et al*., 2017; Colonges *et al*., 2022). Sensory based phenotyping for quality traits is often associated with low repeatability (Gasura *et al*., 2020) which could result in low resolution of SNP-trait associations whereas machine automated phenotyping generally improves association mapping (Xiao *et al*., 2022).

The discovery of genetic loci linked to cassava cooking qualities in the past could therefore have been limited by lack of efficient phenotyping means for these traits. Cooking qualities in cassava are particularly cumbersome to measure in large breeding populations (Iragaba *et al*., 2019). For instance, phenotyping for cassava softness, a textural attribute that is preferred may be done using either a sensory panel which is low throughput and at times

subjective or instrumental methods involving a texture analyzer (Bechoff *et al*., 2018) or penetrometer (Iragaba *et al*., 2019).

However, Uchendu *et al.* (2021) recently used a sensory panel to phenotype a diverse population used for association mapping and eventually uncovered significant loci linked to cassava root mealiness, firmness and other cooking qualities in boiled cassava. Automation of texture analysis in cassava using a texture analyzer (Tran & Escobar, 2019) and penetrometer (Iragaba *et al*., 2019) have also been demonstrated to improve repeatability of measures and could augment discovery of more associated loci.

Cooking time on the other hand is often conventionally assayed using a sensory fork test, which is also subjective and slow. However, Tran *et al*. (2020) improved the reliability of the assay by using root water absorption (WAB) as a proxy for cooking time. Therefore, the combination of improved phenotyping efficiency, genotyping with high marker density and appropriate modelling is likely to increase SNP-trait associations (Burghardt *et al*., 2017), leading to a better understanding of the genetic mechanisms underlying texture and cooking time in cassava.

Such discoveries would enable cassava breeders increase rate of genetic gains for cooking qualities through development and adoption of genomic breeding strategies (Ozimati *et al*., 2018). Also, discovered genomic regions could be targets for transformation and gene editing interventions, that are postulated to deliver even quicker genetic gains (Smith *et al*., 2018).

Applications of genomic selection (GS) in cassava have so far been restricted to easy-to-phenotype agronomic traits including CMD, CBSD (Ozimati *et al*., 2018; Kayondo *et*

al., 2018), yield (Andrade *et al*., 2019) and some phenotypically complex root quality traits including dry matter (Wolfe *et al*., 2016), starch (Phumichai *et al*., 2022) and provitamin A (Esuma *et al*., 2021) with promising prediction accuracies. This evidence suggests that GS could also be employed in improving more specific end-user cassava qualities such as texture and cooking time that influence variety adoption.

With this background, the National cassava breeding program in Uganda recently re-aligned its breeding pipeline to develop improved clones and/or populations that combine agronomic superiority with end-user preferred traits such as texture and cooking time. This feat is dauting if conventional recurrent selection is employed, but possible if genomic breeding is adopted. The objective of this study was therefore to (i) determine the genetic loci linked to texture and cooking time in boiled cassava and (ii) assess genomic prediction accuracies for these traits.

5.2 Materials and Methods

5.2.1 Test germplasm and field establishment.

A total of 196 clones including two commercial checks were sourced from the National Crops Resources Research Institute (NaCRRI) in Uganda. These materials were selected from the cycle 2 (C2) seedling population previously established. The clones were established in two growing seasons (2019 – 2020 and 2020 – 2021) across two agroecologies: Namulonge and Serere District (previously described in chapter three).

Experiments were set up in the field at the onset of rains in augmented design with the two checks randomized within each incomplete block. Clones were established in single row plots of 10 plants with an interplant and interplot distance of 1m. Incomplete blocks were separated

by 2m alleys. Throughout the crop cycle, no fertilizer or herbicide was applied and weed control was done using hand weeding.

5.2.2 Assessment of cooking quality

The clones were assayed for texture attributes namely softness, area (toughness) and gradient (stiffness) using a texture analyzer according to Tran and Escobar (2019) and penetrometer (Iragaba *et al*., 2019; Uchendu *et al*., 2021) (Table 9) after boiling the root pieces. Water absorption was assayed using a gravimetric assay according to Tran *et al*. (2020) with minor modifications (chapter 3).

Table 9: Description of cooking traits evaluated for boiled cassava.

Name of trait	Description	Method for phenotyping
Softness	The force (in grams or kgfcm^{-2}) required to compress or break into sample	Texture analyzer and penetrometer
Toughness	The total amount of work required in deforming or cutting a sample	Texture analyzer
Stiffness	The modulus of elasticity of a sample	Texture analyzer
Water absorption	The amount of water absorbed by a sample during cooking	Gravimetric assay

5.2.3 Instrumental texture analysis

After boiling, the root pieces were separated into two portions and one of the portions (8 pieces) was selected for assessment with a Texture analyzer (TA XT PLUSC, UK) loaded with a 50kg loadcell (Chapter 3). The TA was set to a "return-to-start" mode with pre-test speed of 2 mm/sec, test seed of 1 mm/sec, post-test speed of 10 mm/sec,

compression distance of 20 mm and a trigger force of 500g. Using a combination of a compression probe and 5 blade extrusion grid, each boiled root piece was extruded to generate data on softness, toughness and stiffness.

The other half of the cooked root portions were assayed using a penetrometer (FHT 1122, China) to determine root softness. The roots were assayed while still hot by pressing the penetrometer probe 1 cm into the cooked piece and recording the maximum force achieved in $kgfcm^{-2}$ (Iragaba *et al*., 2019).

5.2.4 Genotyping

Healthy and young green leaf samples from three months old plants were harvested from each plot, punched into wells of a genotyping plate that was kept on ice to maintain tissue integrity. The plate was then placed in a desiccator with silica gel to dry the samples and then shipped to DArT Private Limited in Canberra, Australia (www.diversityarrays.com) for genotyping.

Extraction of DNA, Genotyping-by-Sequencing and SNP calling were done for each sample using DArTseq genotyping platform protocols (https://www.diversityarrays.com/technology-and-resources/dartreseq/). A total of 32,250 SNPs were called and further filtered for downstream analyses.

The resulting SNP markers were filtered with a maximum threshold of 95% reproducibility, 95% call rate for markers, and 20% missing values over samples using R package *SNPReady* (Kang *et al*., 2020). The SNPs were also filtered for minor allele frequency (MAF) < 0.05% in order to eliminate the rare allele variants, which were not the focus for

this study. After the filtering and data quality control process, a total of 26,850 SNPs distributed across the 18 cassava chromosomes were retained. The retained markers were used for genomic prediction and GWAS analysis.

5.2.5 Statistical Analyses

Phenotype data analysis

A mixed effects model was fitted on the combined location and seasonal data for analysis of variance using the *lme4* package in R software (Uchendu *et al*., 2021), R core Team (2014), using the model:

$$Y_{ijk} = \mu + \beta_i + G_k + R_{jmn} + (\beta_i \times G_k) + e_{ijkm},$$

where Y_{ijk} is the phenotypic value of a clone; μ the population mean; β_i is the fixed effect of environment i; Gk the random effect of genotype k; Rjnn is the random effect of the incomplete block j in environment *m* at season *n*; $(\beta_i \times G_k)$ is the random effect of the genotype × environment interaction effect associated with environment i and genotype k and e_{ijkmn} is the residual. Best linear unbiased predictors (BLUPS) for each trait were estimated from the model using *ranef* function in R and subsequently used as phenotypes in subsequent analyses.

Population structure and genetic relatedness analysis

Population structure and relatedness among individuals sharing common ancestry (kinship) are the most important factors for confounding in genetic association studies, often leading to spurious associations (Astle & Balding, 2009). We assessed the extent of

stratification in the population using principal component analysis (PCA) with the *prcomp* function in R studio. Kinship was also estimated inform of a pairwise SNP matrix using the kinship function of *statgenGWAS* package in R studio. The kinship matrix was generated using the VanRaden method (VanRaden, 2008) and implemented using the *kinship* function of *statgenGWAS* package (Rossum *et al*., 2023) in R.

Linkage disequilibrium

Linkage disequilibrium refers to the non-random association of alleles at different loci, and it is informative of historical recombination events in a population and the search window within which causal genes may be found (Bush & Moore, 2012). This was estimated using the correlation coefficients (r^2) between the pairs of loci (26,850 SNPs) in each chromosome (Brito *et al*., 2017) using the *LD.Measures* function of *LDcorSV* package (Desrousseaux *et al*., 2017), and the distribution pattern in the whole genome visualized using *GAPIT3* package (R software; R Core Team 2015).

The LD decay graph was plotted as the r^2 between pairs of SNPs against their pairwise physical distance and showing the average pairwise distances at which LD decayed at r^2 = 0.1 (Figure 19).

Genome-wide association mapping

A mixed linear association analysis was done using the filtered SNPs and BLUPs for all traits as phenotypes. This was implemented using the compressed multi locus mixed model (CMLM) with *GAPIT 3* in R studio (Wang *et al*., 2022). The CMLM is similar to the convention mixed linear model (MLM) which tests association between a phenotype and

marker one at a time to determine if the variance effects at a particular locus significantly deviate from zero.

However, the CLMM differs from MLM in that it tends to shrink the size of random effects from an individual level to a group level, and the compression level, the average number of individuals within one group, is optimized by clustering individuals into groups (Zhang *et al*., 2010).

Population structure represented by number of PCs was added as a fixed effect in the model and kinship that accounts for unequal relatedness among individual genotypes as an additive polygenic effect to correct for spurious associations (Badji *et al*., 2020). Five PCs were fitted into the model as fixed effects.

The optimal control of false positive and false negative rate was achieved through visual inspection of the *Q-Q* plots of observed against the predicted negative log10 *p*-values of each of the SNPs. Statistically significant loci were identified as genomic regions whose SNPs passed the genome-wide *Bonferroni* significance thresholds ($\alpha = 0.01/26{,}850$) (Kuo, 2017).

Putative gene identification

SNPs that were significant at the *Bonferoni* threshold but also located within exonic and intronic regions of cassava genes were considered for annotation. Genes located within 10,000 – 100,000 base pairs (10 – 100 kb) of the significant SNPs were recorded as candidate genes through annotation to the *Manihot esculenta* reference genome (v7.1) on phytozome v.21.1 (https://phyto zome.jgi.doe.gov/pz/porta l.html) (Phumichai *et al*., 2021;

Rabbi *et al*., 2022). Functionality of the genes closest to the significant SNPs was done by blast searching the SNP sequences for proteins related to texture and water absorption using the National Center for Biological Information (NCBI) database.

Genomic predictions

BLUPs from all traits were first de-regressed using the formula proposed by Garrick *et al*. (2009) to minimize effects of double shrinkage of genotype values;

$$deregressed\ BLUP = \frac{BLUP}{1 - \frac{PEV}{\sigma^2 c}},$$

where PEV was the prediction error variance of the BLUP and $\sigma^2 c$ the variance of the genotypes.

Deregressed BLUPS from traits were used in a 5-fold cross-validation scheme with 10 replications to evaluate prediction ability (PA) for softness, water absorption (WAB), toughness and stiffness using a single-step genomic best linear unbiased predictor (G-BLUP) model (Ozimati *et al*. 2018).

The combined phenotype-genotype data was randomly partitioned into five subsets and 4 of the sets were assigned to the training set intended to calibrate the prediction model and the remaining set used as the validation set whose phenotypes were assumed to be unknown. This cross-validation scheme was repeated 10 times such that samples were drawn with replacement from the reference set. Pearson's correlation between predicted GEBVs and the observed de-regressed BLUPS was used to determine the accuracy of the predicted phenotypes.

Genomic prediction accuracy was further evaluated using parametric models such as Bayesian models including Bayes A (BA), Bayes ridge regression (BRR) and semi- parametric models such as the reproducing Kernel Hilbert space (RKHS), all under scenarios of (i) optimized and random training sets and (ii) different training population sizes.

This was done because different models have unique assumptions of marker effects, which tend to influence final predictive ability (Nsibi *et al*., 2020). Selection of optimal subpopulations was done using the *TrainSel* function of *TrainSel* package in R. Bayesian models assume that each marker effect comes from a prior distribution, allowing variance across loci to vary (Meuwissen *et al.*, 2001; Nsibi *et al*., 2020). In contrast, GBLUP assumes equal variances for all markers resulting in homogeneous shrinkage of effects toward zero (Gota & Gianola, 2014), which may not be realistic.

Bayesian ridge regression (BRR) on the other hand assumes nonzero and normally distributed marker effects (Nsibi *et al*., 2020) and equal marker variances whereas the Restricted kernel Hilbert Spaces (RKHS) models pairwise distances between samples by a Gaussian kernel, thereby enabling the capture of nonlinear relationships such as dominance and epistasis (Farooq *et al*., 2023). Tukey's honest significance test (R Core Team 2022) was used to test for significant differences among the predictive accuracies of the four models.

5.3 Results

5.3.1 Variability and BLUP correlations among Root cooking qualities

From the combined phenotypes across four contrasting environments, all traits had a skewed distribution towards the tails (Supp. Fig. 3). Raw BLUP values were generated for each trait and used in the association analysis downstream. Broad sense heritability was generally

moderate across all traits ranging from 0.41 for softness (penetrometer) to 0.67 for gradient (texture analyzer). However, SNP based heritability which represents a measure of additive genetic effects was generally low ranging from 0.03 for softness (penetrometer) to 0.06 for toughness. The traits exhibited varying coefficient of variation (CV) with Force having the lowest and WAB the highest (Table 10).

Table 10: Estimates of heritability, mean and CV for cooking traits.

Trait	Method	SNP h^2	H^2	Mean	CV (%)
Soft_p	Penetrometer	0.03	0.41	2.67	66.87
Soft_T	Texture analyzer	0.05	0.59	13588.09	46.53
WAB	Gravimetric	0.05	0.61	7.20	116.72
Stiffness	Texture analyzer	0.03	0.67	885.83	84.73
Toughness	Texture analyzer	0.06	0.51	145472.3	53.74

h^2 = narrow sense heritability; H^2 = broad sense heritability; CV =coefficient of variation

BLUP correlations (Fig 16) showed a moderate correlation (r = 0.44 – 0.57) between Soft_p and other texture analyzer determined parameters. Soft_p was negatively correlated with WAB30 (r = -0.71) but Soft_T, Stiffness and Toughness were moderately negatively correlated with WAB30. In contrast, Soft_T, Toughness and Stiffness were highly positively correlated.

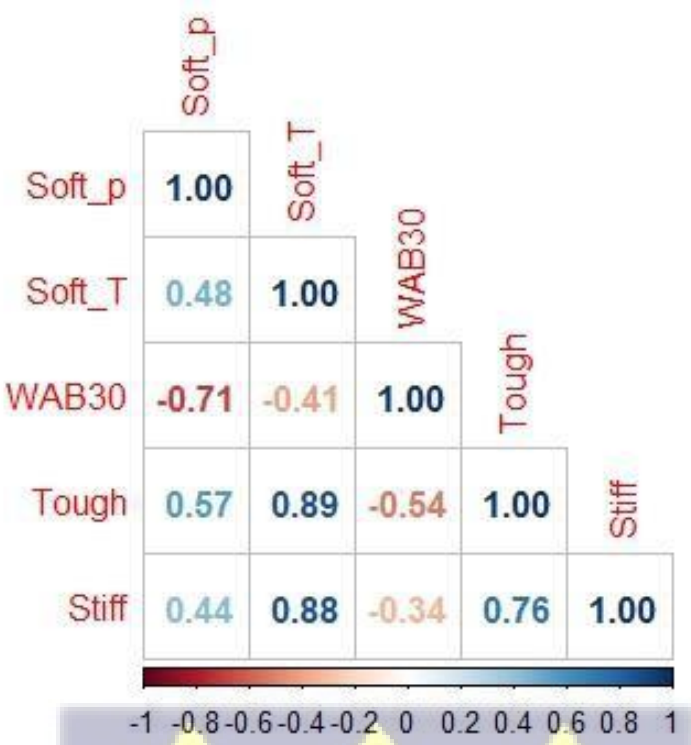


Figure 16: Heat map of BLUP correlations among cooking traits. Soft_p = softness assessment by penetrometer; Soft_T = softness assessment by texture analyzer; WAB30 = water absorption after 30 minutes of boiling roots; Tough = Toughness; Stiff = Stiffness of roots.

5.3.2 SNP distribution and genetic diversity

A total of 26,850 SNPS were generated for each of the 196 genotypes. These SNPs wer not evenly distributed on all chromosomes, with chromosome 16 having the lowest number of SNPs (932) and chromosome one having the highest number (2874 SNPS) (Figure 17). The average number of SNPS per chromosome was 1492.

The SNP set had a polymorphic information content (PIC) ranging from 0.09 to 0.38 with an average PIC of 0.28. Minor allele frequency (MAF) ranged from 0.05 to 0.5 with an average of 0.26. Observed heterozygosity ranged from 0.02 to 0.08 with an average of

0.05 (Table 11). The genetic diversity (GD) score ranged from 0.1 to 0.5 with an average of 0.35.

Table 11: Genetic statistics of the 26,850 SNP markers

Parameter	mean	lower	Upper
Nei's Genetic diversity (GD)	0.35	0.1	0.5
Polymorphic information content (PIC)	0.28	0.09	0.38
Minor allele frequency (MAF)	0.26	0.05	0.5
Observed heterozygosity (Ho)	0.05	0.02	0.08
Inbreeding coefficient (F)	0.86	0.78	0.95

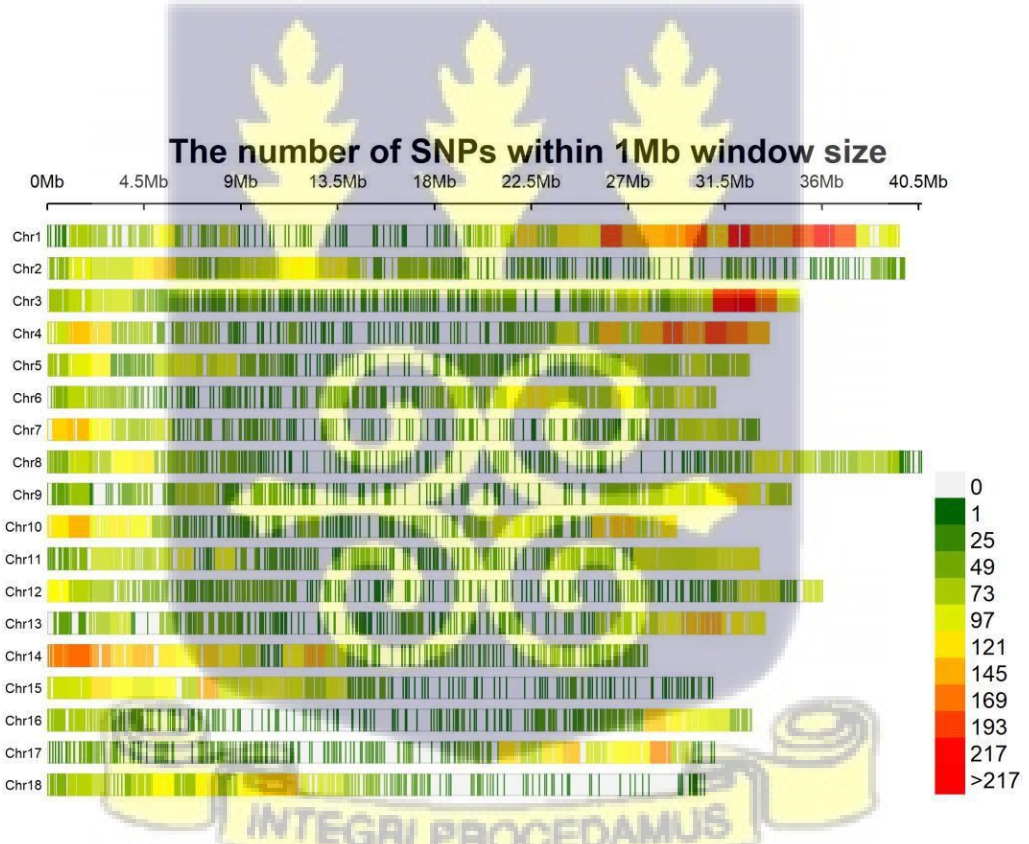


Figure 17: SNP distribution on different chromosomes; chromosome 1 had the highest number of SNPs while chromosome 16 had the lowest.

5.3.3 Linkage disequilibrium and decay

Average LD generally remained low across all chromosomes, but chromosomes 1,13 and16 had higher LD averages than other chromosomes (Supp. Fig. 8). The genome wide LD average peaked at r^2 of 0.27. Linkage disequilibrium (LD) decay was generally fast across all chromosomes, with some observed variations.

Chromosomes 1,3,13,14 and 18 had relatively slower LD decay rates compared to the rest of the chromosomes (Supp. Fig. 8). The genome wide LD average decreased rapidly with physical distance from 0.27 to below $r^2 = 0.2$ at 100 kb. It further dropped below $r^2 < 0.1$ at an approximate distance of 1000kb (Figure 18).

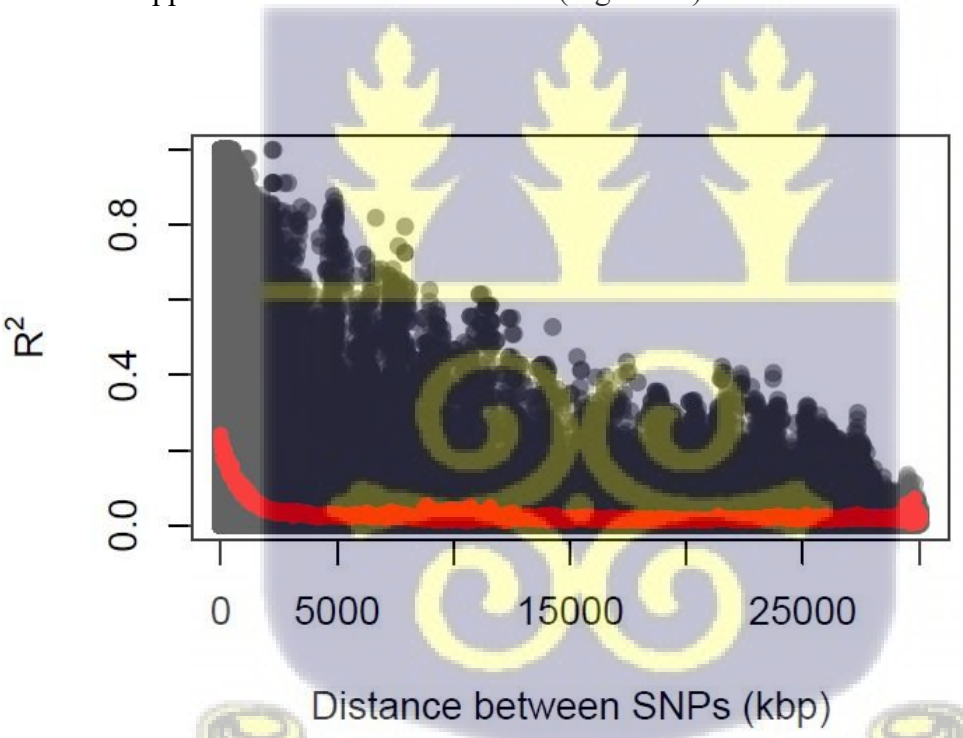


Figure 18: Population wide linkage decay plot showing rapid decay of linkage disequilibrium.

Population structure and relationship

There was minimal stratification among the genotypes as shown by the principal component analysis (PCA) (Figure 19A), and the first 5 principal components (PCs) explained 63.21 % of total genetic variance while the first 10 PCs cumulatively accounted for 77.43 % of total variance.

Kinship estimated using the *vanRaden* method revealed several blocks of closely related individuals along the diagonal (Figure 19B). Areas in red generally showed points of highest correlation between pairs of individuals and those in yellow show low correlation among individuals.

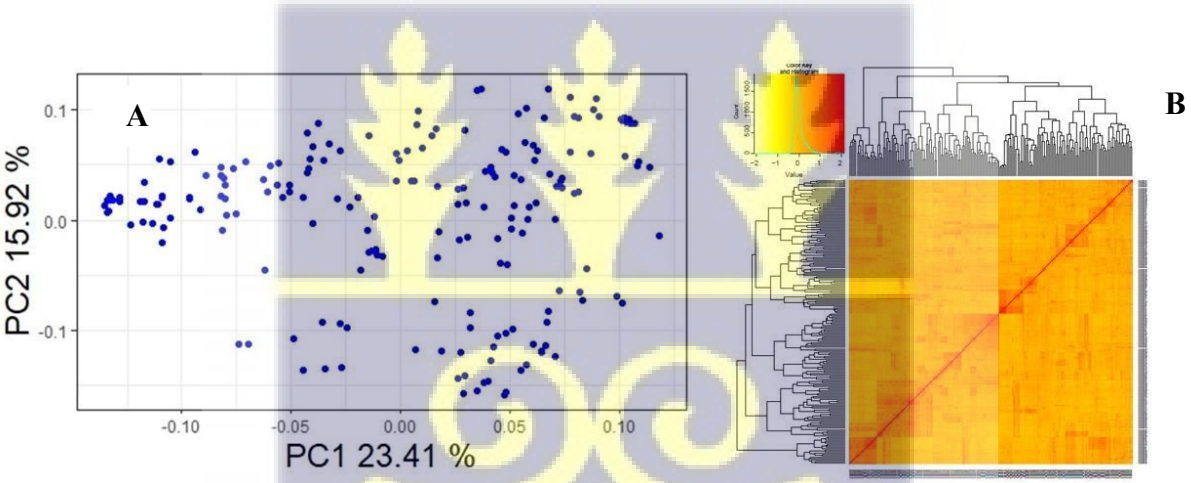


Figure 19: Population structure (A) and kinship plot (B) constructed using 26,850 SNPs, showing moderate stratification in the population.

5.3.4 Genome-wide association analysis

Using the CMLM, the study identified 111 genome wide significant SNP-trait associations across most of the chromosomes (Supp. Table 2). Association analysis was only done on Soft_p, Soft_T, WAB30, Toughness and Stiffness traits because of their relative importance towards end- user preferences for boiled cassava.

Softness

Using a texture analyzer, genetic variation for softness was found associated with major loci on chromosomes 4,5,8,11,17 and18 (Figure 20B). These were tagged majorly by SNP markers on chromosome 4 including S4L3153318 which had a positive effect on softness (Table 12). SNPs S5L9644039 (chr 5), S8L5043186 (chr 8), S11L2768424 (chr 11), S17L23789456 (chr 17) and S18L4974443 (chr18) also had significant effects on softness.

With a penetrometer, variation for softness was highly associated with loci on chromosomes 1,2,4,10,14,17 and 18 (Figure 20A). The most significant locus occurred on chromosome 18 and was tagged by 6 significant markers. It was observed that, SNPs on chromosomes 4, 17 and 18 were consistent for softness assessed by texture analyzer and penetrometer.

Water absorption (WAB)

Genetic variation for WAB was associated with loci on chromosomes 5,6,7,8,10,11,13,16 and 18 (Figure 21C). Marker S18L4974443 (chr18) had a negative effect on WAB followed by S7L29635941 (chr7) with a positive effect. Other markers associated with WAB were found on chromosomes 6 and 11.

Toughness (Area under the curve)

Toughness was significantly associated with loci on six chromosomes 2,4,8,11,17 and 18 (Figure 20D). Markers S2L16500325 (chr2) and S11L146943 (chr11) had a negative effect on toughness whereas S17L23789456 (chr17) and S18L4974443 (chr18) had a positive effect.

Stiffness (Gradient of the curve)

Loci linked to stiffness were found on four chromosomes 1,4, 5 and 8 and tagged by 4 SNP markers. The most significant locus occurred on chromosome 4, tagged by SNPs S4L32681922. (Table 12). On chromosome 8, SNP S8L5961512 was found to be associated with WAB.

Apart from WAB, all other traits had at least one significant SNP on chromosome 4 (Table 12), while chromosome 18 was common among four traits namely Soft_T (softness by texture analyzer), Soft_p (softness by penetrometer), WAB, and Toughness (Area) traits. Also, chromosome 11 was common to Soft_T, WAB and Toughness.

Table 12: Summary statistics of top significant SNPs detected across five traits of boiled cassava.

Trait	SNP	Chrom	Position	P.value	maf	R^2 Mod&SNP	Effect
Soft_p	S1L2151431	1	2151431	5.09E-68	0.50	0.98	0.19
Soft_p	S2L16500325	2	16500325	5.75E-10	0.07	0.66	-0.07
Soft_p	S4L3881302	4	3881302	1.59E-06	0.29	0.98	-0.04
Soft_p	S4L31359043	4	31359043	1.06E-06	0.13	0.82	0.05
Soft_p	S10L20288297	10	20288297	2.12E-45	0.48	0.99	-0.10
Soft_p	S10L20815786	10	20815786	6.63E-65	0.42	0.98	-0.14
Soft_p	S14L24139366	14	24139366	1.15E-06	0.14	0.85	-0.04
Soft_p	S17L23789456	17	23789456	6.31E-08	0.16	0.81	0.05
Soft_p	S18L4974443	18	4974443	2.2E-103	0.10	0.97	0.44
Soft_T	S4L31110972	4	31110972	6.08E-20	0.14	1.00	0.11
Soft_T	S4L31200835	4	31200835	1.78E-18	0.15	1.00	0.10
Soft_T	S4L31249019	4	31249019	1.78E-18	0.15	1.00	0.10
Soft_T	S4L31533184	4	31533184	9.29E-08	0.14	0.99	0.06
Soft_T	S4L31674609	4	31674609	1.73E-10	0.15	1.00	0.07
Soft_T	S5L9644039	5	9644039	6.58E-90	0.28	0.99	-0.27
Soft_T	S8L5043186	8	5043186	6.68E-64	0.47	0.99	0.17
Soft_T	S11L2768424	11	2768424	9.8E-75	0.08	0.95	-0.33
Soft_T	S11L31356213	11	31356213	8.2E-112	0.27	0.97	-0.45
Soft_T	S17L23789456	17	23789456	1.19E-06	0.16	0.97	0.05
Soft_T	S17L23808459	17	23808459	8.63E-49	0.15	1.00	0.22
Soft_T	S18L4974443	18	4974443	1.43E-34	0.10	1.00	0.17
WAB	S5L18056419	5	18056419	1.17E-07	0.25	0.79	-0.69
WAB	S6L21272537	6	21272537	3.94E-10	0.25	1.00	-0.81
WAB	S6L23860699	6	23860699	2.1E-129	0.15	0.84	8.97
WAB	S6L27073265	6	27073265	1.49E-09	0.36	0.93	-0.68
WAB	S7L29635941	7	29635941	1.25E-35	0.18	0.94	2.93

WAB	S8L6899246	8	6899246	5.95E-25	0.07	1.00	2.45
WAB	S10L2115716	10	2115716	1.5E-153	0.48	0.88	9.54
WAB	S11L2973212	11	2973212	3.13E-12	0.06	0.64	1.50
WAB	S11L8566741	11	8566741	1.64E-56	0.12	1.00	-3.54
WAB	S13L28732449	13	28732449	4.41E-07	0.37	0.75	0.52
WAB	S13L29379434	13	29379434	5.25E-10	0.24	0.81	0.73
WAB	S16L27770001	16	27770001	4.97E-11	0.44	0.85	0.70
WAB	S18L4974443	18	4974443	1.76E-12	0.10	0.84	-1.36
Toughness	S2L16500325	2	16500325	1.49E-21	0.07	0.75	-0.21
Toughness	S4L2144235	4	2144235	2.7E-08	0.29	0.96	-0.08
Toughness	S4L31042027	4	31042027	2.22E-45	0.15	1.00	0.30
Toughness	S4L31158648	4	31158648	8.39E-41	0.14	1.00	0.29
Toughness	S8L6327008	8	6327008	1.67E-16	0.47	1.00	0.09
Toughness	S11L146943	11	146943	4.02E-09	0.12	0.58	-0.10
Toughness	S17L23789456	17	23789456	8.39E-54	0.16	1.00	0.34
Toughness	S18L4974443	18	4974443	8.98E-52	0.10	1.00	0.39
Stiffness	S1L6072007	1	6072007	2.96E-91	0.064	0.88	-0.13
Stiffness	S4L32681922	4	32681922	2.45E-73	0.077	0.94	-0.09
Stiffness	S5L9644039	5	9644039	1.08E-63	0.281	0.99	-0.05
Stiffness	S8L5961512	8	5961512	1.18E-09	0.444	1.00	0.01

Soft_p = Softness assessed by penetrometer; Soft_T = Peak force equivalent to softness assessed by texture analyzer; WAB = Water absorption assessed after 30 minutes of boiling; Toughness = Area under the texture profile curve equivalent to toughness; Stiffness = stiffness of boiled cassava; Chrom = chromosome; maf = minor allele frequency.

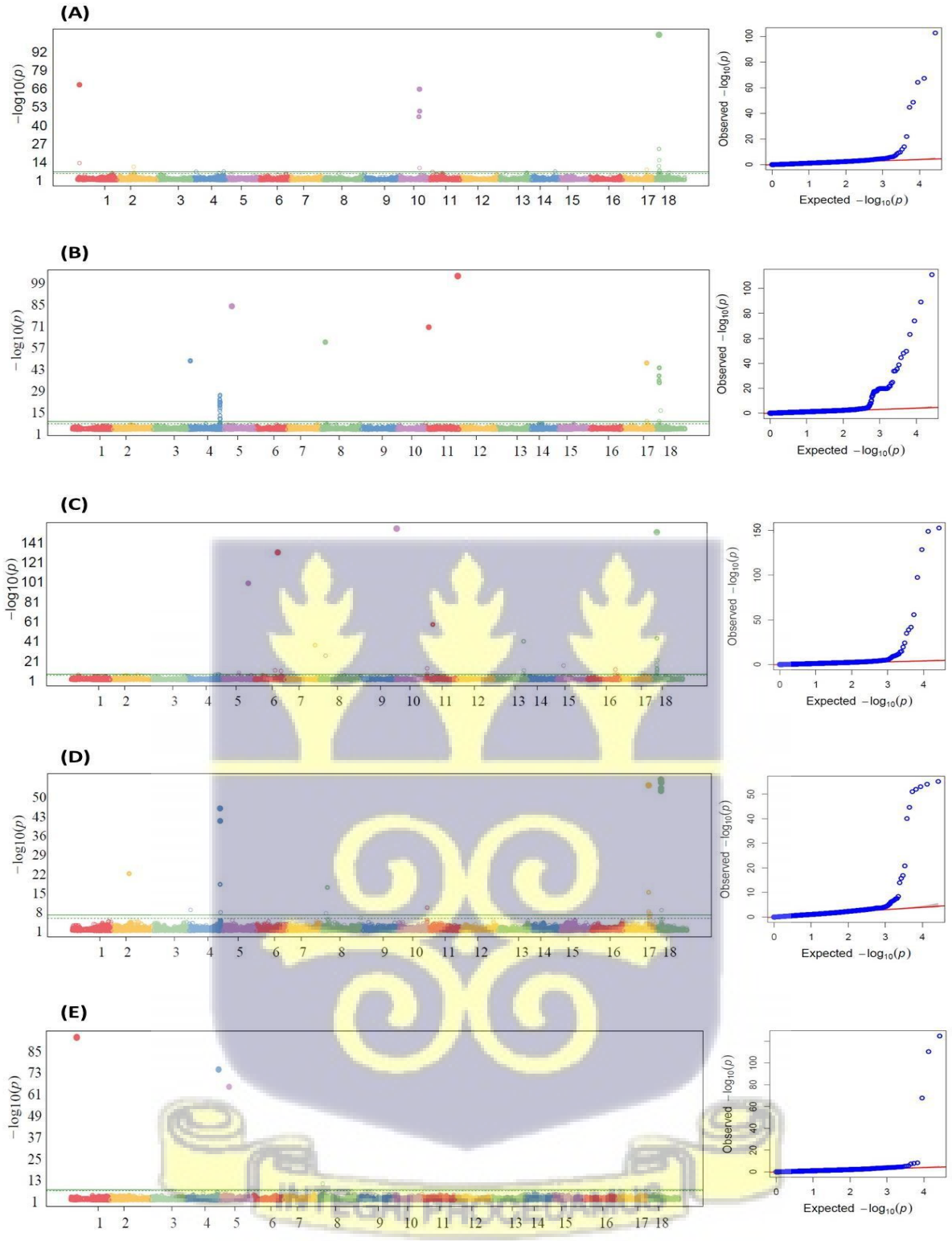


Figure 20: Manhattan and Q-Q plots of significant associations for **(A) –** softness by penetrometer; **(B)** – softness by texture analyzer; **(C)** – water absorption (WAB30); **(D)** - toughness and **(E)-** stiffness of boiled cassava.

5.3.5 Candidate gene identification and annotation

The genome wide significant SNPs from the CMLM analysis were investigated further for presence of potential genes in their vicinity using linkage disequilibrium and gene annotation. Candidate genes were searched up and downstream of significant SNPs on phytozome online database (https://phytozomenext.jgi.doe.gov/jbrowse/index.html?data=genomes%2FMesculenta_v7_1) and identified from the cassava reference genome (*Manihot esculenta v7.1*) (Table 13).

A search within 100kb of significant SNPS revealed Soft_p (penetrometer) to be associated with 14 putative genes on chromosomes 1,2,4,10,14,17 and 18 including Manes.04G136700 coding for *glycosyltransferase* (chr4), Manes.10G089100 (chr10) linked to *squalene epoxidase* (Table 13), Manes.14G162600 (chr14) linked to *calcium ion transporting ATPase*, Manes.17G069200 (chr17) coding for *mannosyl-oligosaccharide 12-alpha-mannosidase*, Manes.18G044200 (chr18) coding for *arginine-N-methyltransferase* and Manes.18G044401 coding for a *glycine rich cell wall structural protein.*

A total of 15 candidate genes of which the majority were found on chromosome (chr) 4 were identified to be associated with softness assessed by texture analyzer (Soft_T). Notably, genes Manes.04G139200 associated with Beta-*galactosidase 1 (*S4_L31533184), Manes.04G141800 associated with *acetyltransferase*, Manes.04G141600 linked to *glycylpeptide N-tetradecanoyltransferase* all on chromosome four and Manes.05G103501 (chr5) associated with *expansin A1* were identified.

On chromosome 8, Manes.08G047500 was associated with *poly-alpha-galacturonosidase* while Manes.17G069200 linked to *alpha mannosidase* was located on chromosome 17.

On chromosome 18, the most notable SNP was associated with Manes.18G044200 linked to *Arginine N-methyltransferase*, which was previously identified for Soft_p (penetrometer).

For WAB, the study identified 10 putative genes collocated with the significant SNPs on chromosomes 5,6,7,13 and 18. These were Manes.05G122800 (chr5) associated with *polygalacturonase*, Manes.06G100100 linked to *Mannose-1-phosphate guanylyltransferase* and Manes.07G106500 associated with *Ribosomal lysine N-methyltransferase 4*. On chromosome 13, the study found Manes.13G114500 coding for *alpha-beta hydrolase family of enzymes*, Manes.13G114400 coding for *esterase* and Manes.13G120400 coding for *ethylene- responsive transcription factor*.

The most notable SNP on chromosome 18, S18L4974443 was associated with genes Manes.18G044401 coding for *a glycine rich cell wall structural protein*, Manes.18G044200 associated with *arginine N-methyltransferase* and Manes.18G044300 linked to a *subtilisin-like protease*.

Significant SNPs associated with toughness of boiled cassava were collocated with genes Manes.04G134700 (chr4) linked to *hexokinase*, Manes.11G000700 (chr11) associated with *aspartic protease* and Manes.17G069200 (chr17) coding for *alpha mannosidase*, Manes.17G069500 (chr17) linked to *P-N-acetylmuramoyl-pentapeptide- transferase*, and on chromosome 18 Manes.18G044401, Manes.18G044200 and Manes.18G044300 previously identified for WAB.

Stiffness (gradient) was associated with genes Manes.01G033900 (chr1) coding for *arabinase*, Manes.04G154800 (chr4) coding *xyloglucan fucosyltransferase*, Manes.05G103401 (chr5) coding for a *MADs box protein* and on chromosome 8, Manes.08G053400 linked to *folylpolyglutamate synthase* and Manes.08G053200 associated with *pectinesterase inhibitor 1*.

These findings highlight the quantitative nature of traits under study and the major role of genes on chromosomes 4 and 18 in modulation of texture properties and water absorption since majority of the associated genes originated from these chromosomes. Also, Manes.18G044401, Manes.18G044200 and Manes.18G044300 all on chromosome 18 were commonly linked with softness (penetrometer and texture analyzer), WAB and toughness of boiled cassava.

Table 13: Genome wide significant SNPs and associated putative genes and proteins

Trait	SNP	Chr	Pos	Putative gene	Protein
Soft_p	S1L2151431	1	2151431	Manes.01G009750	*Stress response protein NST1-like*
Soft_p	S1L2151431	1	2151431	Manes.01G009900	*Inositol-tetrakisphosphate 1-kinase*
Soft_p	S2L16500325	2	16500325	Manes.02G178800	*Exocyst complex component*
Soft_p	S4L31359043	4	31359043	Manes.04G137500	*Glutamate dehydrogenase*
Soft_p	S4L31359043	4	31359043	Manes.04G137700	*Ubiquitin carboxyl-terminal hydrolase*
Soft_p	S4L31359043	4	31359043	Manes.04G136800	*Iaa-amino acid hydrolase*
Soft_p	S4L31359043	4	31359043	Manes.04G136700	*Glycosyltransferase*
Soft_p	S10L20288297	10	20288297	Manes.10G089100	*Squalene epoxidase*
Soft_p	S14L24139366	14	24139366	Manes.14G162600	*Ca2+-transporting atpase*
Soft_p	S17L23789456	17	23789456	Manes.17G069200	*Alpha-mannosidase*
Soft_p	S17L23789456	17	23789456	Manes.17G069500	*Phospho-N-acetylmuramoyl-pentapeptide-transferase*
Soft_p	S18L4974443	18	4974443	Manes.18G044200	*Arginine-N-methyltransferase*
Soft_p	S18L4974443	18	4974443	Manes.18G044300	*Subtilisin like protease*
Soft_p	S18L4974443	18	4974443	Manes.18G044401	*Glycine rich cell wall structural protein*
Soft_T	S4L31200835	4	31200835	Manes.04G134900	*Calmodulin-binding family protein*
Soft_T	S4L31200835	4	31200835	Manes.04G13470	*Hexokinase*
Soft_T	S4L31533184	4	31533184	Manes.04G139200	*Beta-galactosidase 1*
Soft_T	S4L31533184	4	31533184	Manes.04G139800	*Serine/threonine protein kinase*
Soft_T	S4L31533184	4	31533184	Manes.04G139400	*Transmembrane 9 superfamily protein*
Soft_T	S4L31674609	4	31674609	Manes.04G141800	*Dihydrolipoyllysine-residue Acetyltransferase*
Soft_T	S4L31674609	4	31674609	Manes.04G141600	*G-N-tetradecanoyltransferase*
Soft_T	S5L9644039	5	9644039	Manes.05G103501	*Expansin_A1*
Soft_T	S8L5043186	8	5043186	Manes.08G047900	*Arabinogalactan peptide 23*
Soft_T	S8L5043186	8	5043186	Manes.08G048200	*Serine/threonine-protein kinase*
Soft_T	S8L5043186	8	5043186	Manes.08G047500	*Poly-alpha-galacturonosidase*
Soft_T	S11L2768424	11	2768424	Manes.11G027600	*Phospholipase d*

Soft_T	S17L23789456	17	23789456	Manes.17G069200	*Alpha-mannosidase*
Soft_T	S17L23789456	17	23789456	Manes.17G069500	*P-N-acetylmuramoyl-pentapeptide-transferase*
Soft_T	S18L4974443	18	4974443	Manes.18G044200	*Arginine N-methyltransferase*
WAB	S5L18056419	5	18056419	Manes.05G122800	*Polygalacturonase*
WAB	S6L21272537	6	21272537	Manes.06G072700	*Aspartyl proteases*
WAB	S6L23860699	6	23860699	Manes.06G100100	*Mannose-1-phosphate guanylyltransferase*
WAB	S7L29635941	7	29635941	Manes.07G106500	*Ribosomal lysine N-methyltransferase 4*
WAB	S13L28732449	13	28732449	Manes.13G114500	*Esterase*
WAB	S13L28732449	13	28732449	Manes.13G114400	*Esterase D14L*
WAB	S13L29379434	13	29379434	Manes.13G120400	*Ethylene-responsive transcription factor*
WAB	S18L4974443	18	4974443	Manes.18G044401	*Glycine rich cell wall structural protein*
WAB	S18L4974443	18	4974443	Manes.18G044200	*Arginine N-methyltransferase*
WAB	S18L4974443	18	4974443	Manes.18G044300	*Subtilisin like protease*
Toughness	S4L31158648	4	31158648	Manes.04G134700	*Hexokinase*
Toughness	S11L146943	11	146943	Manes.11G000700	*Aspartic protease*
Toughness	S17L23789456	17	23789456	Manes.17G069200	*alpha-mannosidase*
Toughness	S17L23789456	17	23789456	Manes.17G069500	*P-N-acetylmuramoyl-pentapeptide-transferase*
Toughness	S18L4974443	18	4974443	Manes.18G044200	*Arginine-N-methyltransferase*
Toughness	S18L4974443	18	4974443	Manes.18G044300	*Subtilisin like protease*
Toughness	S18L4974443	18	4974443	Manes.18G044401	*Glycine rich cell wall structural protein*
Stiffness	S1L6072007	1	6072007	Manes.01G033900	*Arabinanase*
Stiffness	S4L32681922	4	32681922	Manes.04G154800	*Xyloglucan fucosyltransferase*
Stiffness	S5L9644039	5	9644039	Manes.05G103401	*MADS box protein*
Stiffness	S8L5961512	8	5961512	Manes.08G053400	*Folylpolyglutamate synthase*
Stiffness	S8L5961512	8	5961512	Manes.08G053400	*glycosyltransferase*
Stiffness	S8L5961512	8	5961512	Manes.08G053200	*Pectinesterase inhibitor 1*

Soft_p = Softness assessed by penetrometer; Force = Softness assessed by texture analyzer; WAB = water absorption; Area = Toughness; Gradient = Stiffness; Chr = Chromosome; Pos = SNP position.

5.3.6 Estimation of genomic predictive ability through cross validation

Impact of prediction models on accuracy

Overall mean prediction accuracies remained low across traits and models using 100% of the sub-samples (196 genotypes). However, prediction accuracies remained relatively higher for WAB than other traits in the cross validation (Figure 21). The restricted kernel Hilbert space (RKHS) model had the highest mean prediction accuracy for WAB ($r^2_{cv} = 0.27$), while Bayes A had the lowest accuracy ($r^2_{cv} = 0.21$) (Table 14). For Soft_p, prediction accuracy ranged from 0.14 (GBLUP) to 0.21 (Bayes A) whereas force had prediction accuracies ranging from 0.003 (GBLUP) to 0.09 (BA).

Prediction accuracies for gradient ranged from -0.13 (GBLUP) to 0.05 (BRR) and those for area ranged from -0.08 (GBLUP) to 0.04 (BA). The GBLUP model generally gave lower prediction accuracies across all traits, while Bayes A gave better predictions overall (Table 14). The Tukey HSD test revealed statistically significant differences ($p<0.05$) between prediction accuracies given by GBLUP and BA.

Also, comparison between accuracies given by GBLUP and BRR and RKHS and GBLUP were significantly different across traits (Figure 22). Overall, GBLUP performed poorly across models and prediction accuracies by BA, BRR and RKHS were similar across all traits.

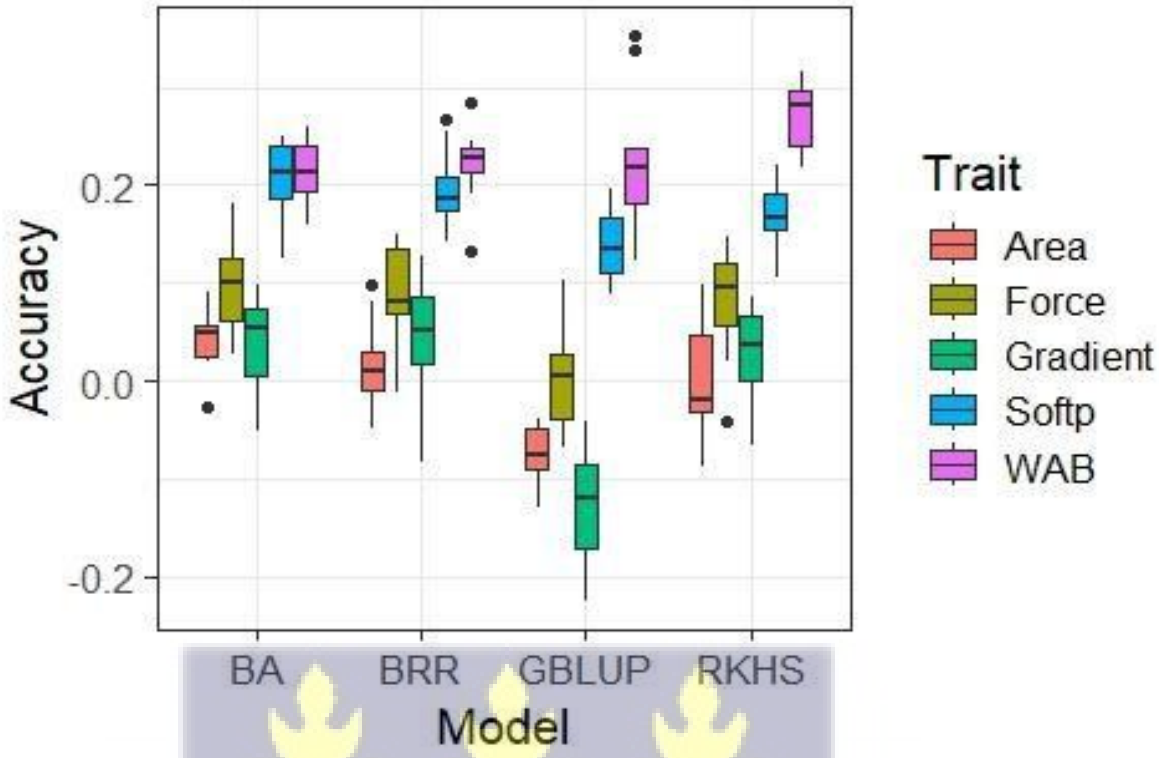


Figure 21: Genomic prediction accuracy through cross validation using the original training panel of 196 genotypes showing low prediction accuracy across traits (Training set = 157, Test set = 39). BA = Bayes A; BRR = Bayes Ridge Regression; GBLUP = Genomic Best Linear Unbiased prediction; RKHS = Reproducing Kernel Hilbert Spaces. Area = Toughness; Force = softness by texture analyzer; Gradient = Stiffness; Soft_p = softness by penetrometer; WAB = water absorption assessed after 30 minutes of boiling.

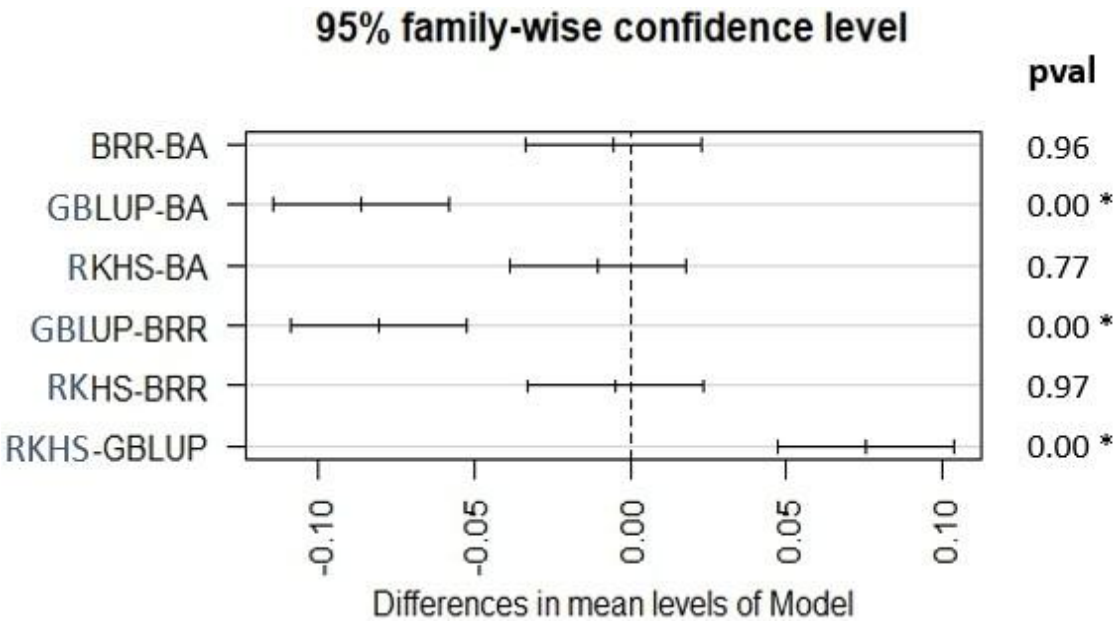


Figure 22: Confidence intervals for model comparison and associated p-values. *Indicates model pairwise comparisons that are significant at $p<0.05$.

Table 14: Average prediction accuracy for each model across cooking traits in cassava

Model	Train	Test	r²_Soft_p	r²_Soft_T	r²_WAB30	r²_Tough	r²_Stiff
GBLUP	157	39	0.14	0.003	0.23	-0.08	-0.13
BRR	157	39	0.19	0.09	0.22	0.02	0.05
RKHS	157	39	0.17	0.08	0.27	-0.002	0.02
BA	157	39	0.21	0.09	0.21	0.04	0.03

r²_Soft_p = average prediction accuracy for softness by penetrometer; **r²_Soft_T** = prediction accuracy for softness by texture analyzer; **r²_WAB30** = prediction accuracy for water absorption assessed after 30 minutes of boiling; **r²_Tough** = prediction accuracy for toughness (Area); **r²_Stiff** = prediction accuracy for stiffness (Gradient).

Cross validation with optimized set of 156 genotypes

Using the optimized training panel (156 genotypes) that constituted 80% of the original population, there was a general reduction in prediction accuracies (Figure 23) across most traits using different models in comparison to the original panel (196 genotypes). Prediction accuracy for WAB remained higher than other traits and the BRR model gave the highest mean accuracy (0.25), while GBLUP gave the lowest prediction accuracy (0.1).

Similarly, BRR gave the highest prediction accuracy for Soft_p (0.17) and GBLUP gave the lowest accuracy (0.07). The highest mean prediction accuracy for force was attained with Bayes A (0.08) while GBLUP had the lowest accuracy (-0.08). Bayes A also had the highest prediction accuracy for Area (0.02) and GBLUP the lowest (-0.12). Likewise, Bayes A had the highest prediction accuracy for gradient and GBLUP the lowest (-0.13). The GBLUP model still performed poorly across traits and Bayes A overall gave the higher prediction accuracies across traits.

Cross validation with optimized set of 176 genotypes

Prediction accuracies were slightly improved for Soft_p and WAB when an optimized training panel of 176 genotypes was used (Figure 23), representing 90% of the original panel. Prediction accuracy for force, area and gradient dropped slightly in comparison to the optimal set of 156. Bayes A and RKHS both gave a prediction accuracy of 0.27 like that achieved with the full set.

Prediction accuracy for softness ranged from 0.18 to 0.19 with Bayes A and BRR giving the highest accuracy. For force, BRR gave the highest prediction accuracy at 0.04 while GBLUP gave the lowest accuracy (-0.06). Similarly, BRR gave the highest prediction accuracy for area and GBLUP the lowest. Bayes A gave the highest prediction accuracy for gradient and GBLUP the lowest. Overall, using large training populations gave better prediction accuracies across traits and BA, BRR and RKHS had similar prediction accuracies across traits and populations.

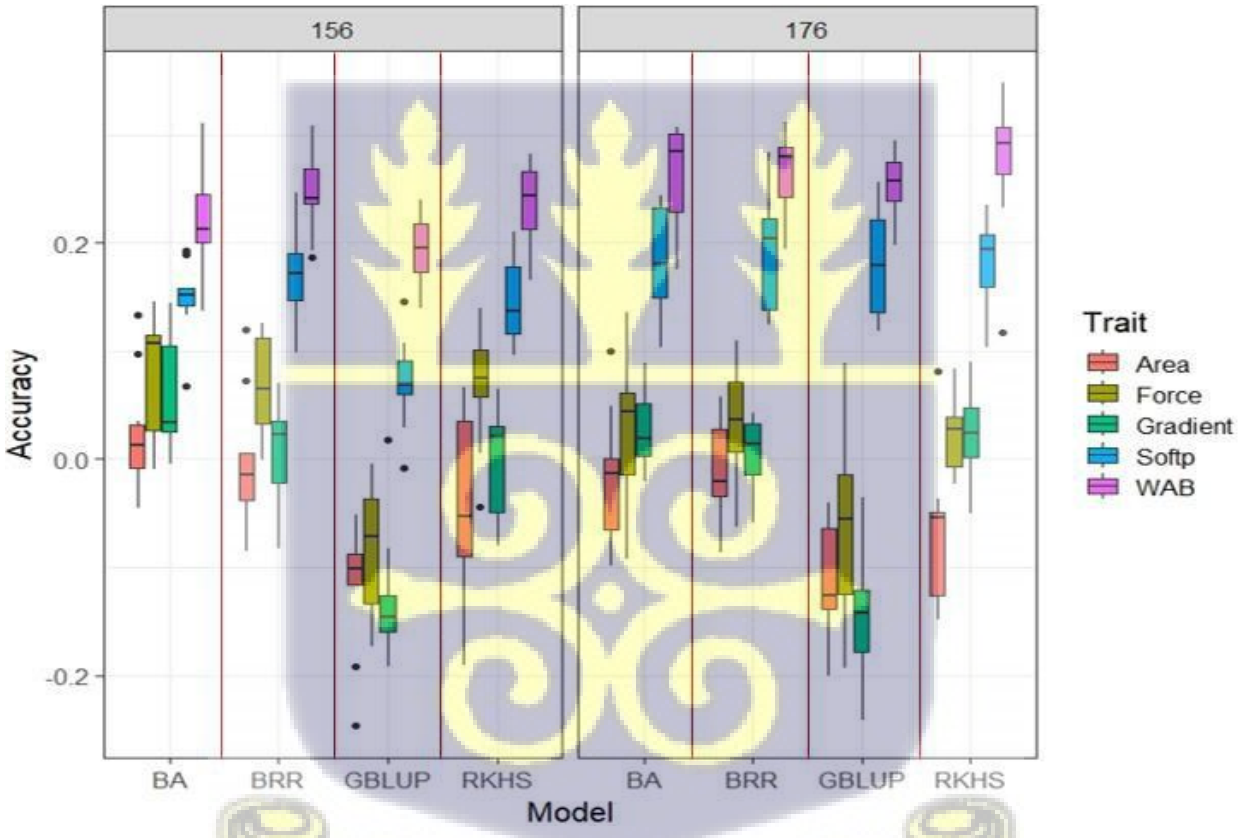


Figure 23: Cross validation prediction accuracy for cooking qualities in cassava using GBLUP, BRR, BA and RKHS models across two optimized training populations of size 156 and 176.

5.4 Discussion

Knowledge of the genetic architecture of traits linked to boiled cassava quality traits is a leap forward towards breeding for relevant cassava varieties. This work complements and builds on previous GWAS efforts on cassava cooking traits by uncovering more associated loci. This is achieved by employing automated and semi-automated phenotyping using a texture analyzer and penetrometer respectively on a cycle 2 genomic selection panel (196 genotypes) to improve data quality for associating with genotype data. It also assesses genomic prediction accuracies to determine its potential use in selecting for end-user preferred boiled cassava traits.

There was moderate genetic diversity (GD) with a score of 0.35 (Table 11) which was higher than that previously reported in a similar study by Uchendu *et al*. (2021), indicating the usefulness of the marker set used in this study for association analysis. The PIC (0.28) also was higher than that reported by (Uchendu *et al*., 2021), indicating more informativeness of the SNPs used in this study. The moderate broad sense heritability recorded in this study for Soft_p (H^2 = 0.41) was lower than those for other cooking traits, but still higher than that that previously reported by Iragaba *et al*., (2019; H^2 = 0.37) and Uchendu *et al*. (2022; H^2 = 0.36).

This indicates there is minimal contribution of genetic factors to Soft_p but higher contribution of environment to the expression of the trait. Conversely, the moderate heritability associated with using a texture analyzer Soft_T; (H^2 = 0.59) is indicative of high contribution of genetic factors and better repeatability compared to the penetrometer for assessing boiled root softness.

The moderate heritability of 0.61 reported for water absorption (WAB) in this study is lower than that found by Tran *et al*. (2020) much as the latter only used single location estimates. The large residuals possibly attributed to low method repeatability, could have affected heritability estimates for WAB. Heritability estimates reported for toughness (H^2 = 0.51) and stiffness (H^2 = 0.67) in this study have not been reported elsewhere. SNP based heritability was generally low across all traits (h^2 < 0.1), implying there were minimal additive genetic effects influencing cooking qualities for boiled cassava.

Low narrow sense heritability has also been reported by Phumichai *et al*. (2022) for cassava fresh root weight (h^2 = 0.1), starch content (h^2 = 0.25) and high estimates for pasting properties (h^2 = 0.58 – 0.85). Additionally, Rabbi *et al*. (2022) found low h^2 estimates for cassava green mite (CGM) severity (h^2 = 0.17), harvest index (h^2 = 0.31) and plant type (h^2 = 0.38) but moderate heritabilities for plant architectural traits, dry matter and total carotenoids (h^2 = 0.43 – 0.72). However, no studies in cassava have previously reported narrow sense heritability for cooking traits.

The moderate BLUP correlations between the penetrometer and texture analyzer for softness further confirms that the two tools are comparable in measuring boiled root softness. The strong positive correlations among texture analyzer traits (Soft_T, Toughness and stiffness) also indicate that selection for one of these traits would simultaneously improve the rest. Likewise, the negative correlation between WAB30 and other textural parameters could be exploited to develop fast cooking cassava varieties with soft texture simultaneously.

Genetic correlations among traits could arise due to linkage disequilibrium (LD) or pleiotropy (one gene controlling more than one unrelated biological processes) (Rabbi *et al*., 2022). However, correlations due to LD are often temporary and unreliable since LD tends to decay in subsequent cycles of recombination in a population. In contrast, pleiotropy-based correlations are more reliable due to control of genetic loci and could be beneficial for breeding especially if a particular locus causes desirable changes in two or more traits (Chen *et al*., 2010).

Moreover, inferences made on genetic (BLUP) correlations are more reliable than phenotypic ones (chapter 3) since the latter could be subject to environmental correlations (Souza *et al*.,1998). This could explain the drop in magnitude of correlation coefficients from phenotypic-based (Chapter 3) to BLUP-based correlations (Figure 16).

Marker density, population structure and linkage disequilibrium

The study used a moderate density marker set of 35,521 SNPs, which were further trimmed to 26,850 SNPs. The density of markers used in an association panel greatly influences the resolution of the associations made (Muqaddasi *et al*., 2019). High marker densities (i.e. > 100,000 SNPs generally tend to increase mapping resolution by facilitating discovery of additional markers explaining more genotypic variance, though this could result in high costs of genotyping. Effect of marker density is majorly determined by linkage disequilibrium (LD), where low LD necessitates increasing number of markers (Bejarano *et al*., 2018; dos Santos *et al*., 2022).

The LD decay observed in this study (i.e., $r^2 = 0.27$) was low, but fast and similar to that reported by Rabbi *et al*. (2022). However, this was lower than that reported by Uchendu *et al*. (2021). This is not surprising given that the C2 test population used in this work had already undergone three rounds of recombination (C0 – C2) that could have reduced LD bock sizes with each round.

Further, cassava is innately an outcrossing crop that often shows fast LD decay, due to high number of recombination events, that tend to increase rate of LD decay (Vos *et al*., 2017). Moreover, by increasing the recombination events in a population, its resolution for association mapping tends to improve, thus favoring discovery of SNP-trait relationships (Flint-Garcia *et al*., 20).

Also, evaluation of population structure and relatedness among individuals in an association panel is recommended for correction of spurious associations (Souza *et al*., 1998; Rabbi *et al*. 2022). Based on principal component analysis (PCA), there was minimal structure in the population (Figure 19). This could be attributed to the systematic intercrossing history of the C2 population, which is typical of recurrent selection programs, resulting in genetic similarity among genotypes.

Based on the observed minimal stratification and elaborate kinship, we proceeded to add 5 principal components (PCs) as a fixed effects and kinship as an additive polygenic effect in the CMLM model since such a model better corrected for false positives, as deduced from the *Q-Q* plots.

Significant SNPs associated with cooking qualities

From this study, a total of 111 SNPs were found associated with water absorption (WAB), Soft_T (texture analyzer), Soft_p (penetrometer), Toughness and Stiffness of boiled cassava. Once validated, these markers could be adopted for marker assisted selection for improving cooking qualities in breeding programs.

Most of the genes linked to Soft_p were involved in cell wall modification pathways (Table 13). Moreover, changes in cell wall structure and composition are often linked to noticeable alterations in texture of fruits (Fuentes *et al*., 2019). For instance, on chromosome 2, the Manes.02G178800 linked to the *exocyst complex component* has been implicated in mediating cell wall secretion (Mitsuda *et al*., 2005) leading to cell wall assembly in A. *thaliana* (Marković *et al*., 2020).

Another notable gene, Manes.04G136700 (S4L31359043) associated with softness was located on chromosome 4. This protein expressed *glycosyltransferase* (GT) known for its role in catalyzing the formation of cell wall matrix polysaccharide linkages (Amos & Mohnen, 2019), through coordination of pectin synthesis (Harholt, 2006) in plants. This could lead to cell wall strengthening which would affect final texture in boiled cassava.

On chromosome 14, Manes.14G162600 was found coding for *calcium-transporting ATPase* which primarily pumps excess calcium ions (Ca^{2+}) from the cytosol into organelles for storage and reallocation to cell walls where it affects structural integrity (Pooviah 1985; Robertson,

2013). In tandem with cell wall depolymerizing enzymes, calcium tends to cross link with polysaccharides having a low degree of methylation to form stiffer cell walls (Temple *et al*., 2021), which could eventually confer a hard texture on boiled cassava.

Alpha-mannosidase (Manes.17G069200) on chromosome 17 has also been found in tomato to be responsible for *N- glycan* processing, which is linked short shelf life and softening of tomato (Irfan *et al*., 2016). This study also found a *subtilisin like protease* (*subtilase*) (Manes.18G044300) on chromosome 18 that is linked to regulation of cell wall properties (Schaller *et al*., 2018). A study on tomato found *subtilases* to affect *pectin methylestestarase* (PME) activity, a major enzyme associated with homogalacturonan (major pectin component) demethylation (Meyer *et al*., 2016).

In the vicinity, a glycine rich cell wall structural protein (GRP) coded by Manes.18G044401 (chr18) has been linked to structural and defense roles in plant cell walls (Ringli *et al*., 2001). Some GRPs may even act as scaffoldings for callose deposition leading to cell wall fortification (Mangeon *et al*., 2010), a process that could alter final texture in cassava. Softness measured by texture analyzer (Soft_T) was also linked to genes involved in cell wall modulation, highlighting the critical role played by cell walls in determining softness of final products.

For instance, the calmodulin-*binding family protein* coded by Manes.04G134900 (chr4), has been linked to secondary cell wall biosynthesis and strengthening (Badmi *et al*., 2018). Calmodulin is a calcium ion (Ca^{2+}) sensor that modulates activities of several proteins in plants

that in turn control a myriad of processes, including cell wall regeneration and fortification (Snedden *et al*., 2001).

The most notable gene on chromosome 4 was Manes.04G139200 (S4L31533184) coding for *β-galactosidase 1*, a cell wall modifying enzyme involved in disassembly of the cell wall leading to softening in fruits (Yang *et al*., 2018; Pan *et al*., 2022). In potato, *galactosidase* was found to shorten galactose side chains of pectins in tubers, resulting in lowering of cell wall materials (CWM) (Huang *et al*., 2016). In contrast, (Uchendu *et al*., 2021) found *β-galactosidase 3* to be associated with taste and color of boiled cassava.

Cell wall polysaccharides are also documented among the factors that influence water absorption in foods and high amounts of non-cellulosic polysaccharides in cell walls have been shown to increase hydration properties of the food (Ramasamy, 2014). In this study, water absorption (WAB) was majorly linked to Manes.07G106500 (S7L29635941) coding for *ribosomal lysine N-methyltransferase 4* and Manes.18G044300 (Subtilisin-like *protease*), Manes.18G044401 (glycine-rich cell wall structural protein) and Manes.18G044200 coding for *Arginine-N-methyl-transferase* (S18L4974443), that were previously linked to Soft_T (Table 13).

This implies that there is a strong pleiotropic role played by these genes for the two traits. Ethylene responsive transcription factor (ERTf) coded by Manes.13G120400 was also found linked to water absorption. This gene has been found to regulate several cell wall modifying genes in apple (Zhai *et al*., 2022) and potato (Dobránszki *et al*., 2021).

Toughness was associated with Manes.04G134700 (S4L31158648) coding for *hexokinase* previously found associated with softness (texture analyzer), Manes.11G000700 (S11L146943) coding for *aspartic proteases* whose role in texture expression is not clearly understood. On chromosome 17, Manes.17G069200 and Manes.17G069500 previously associated with Soft_p (penetrometer) and on chromosome 18, Manes.18G044200, Manes.18G044300 and Manes.18G044401 were also associated with Toughness. This is also indicative of pleiotropic effects of these genes in regulating softness and toughness, which were also moderately positively correlated (Figure 16).

Stiffness was associated with Manes.01G033900 (S1L6072007) coding for *arabinanase* which digests arabinan units that form part of the side chains of pectic rhamnogalacturonan I in the cell wall, potentially affecting the structural integrity of cell walls (Skjøt *et al*., 2002). This notion is confirmed in a study on apple where loss of highly branched (1/5)-a-L-arabinans preceded the loss of firm texture in apples during storage (Peña and Carpita, 2004).

On chromosome 4, Manes.04G154800 (S4L32681992) coding for *xyloglucan fucosyltransferase was* also associated with root stiffness. This enzyme is associated with addition of L-fucose units to cell wall polysaccharides like glycans and xyloglucans which contribute to cell wall strength (Soto *et al*., 2019). It also regulates cellular adhesion (Verger *et al*., 2016), which in turn affects stiffness of plant tissues (Paniagua *et al*., 2014). The putative Manes.05G103401 (S5L9644039) coded for MADS box protein, a regulator of locular gel formation was found to increase tomato firmness when fully expressed (Huang *et al*., 2016). The effect of this protein on cassava texture needs to be studied further.

Folylpolyglutamate synthase coded by Manes.08G053400 (chr8) was also associated with stiffness in this study. A loss of function of this gene in A *thaliana* was linked to reduced lignification in cell walls and consequently less recalcitrance (Srivastava *et al*., 2015). In the vicinity, the study also identified *glycosyltransferase* (Manes.08G053400), that is associated with cell wall polysaccharide synthesis such as xylose, pectin and xyloglucan, thus directly contributing to cell wall integrity (Amos & Mohnen, 2019). It catalyzes formation of glycosidic bonds by attaching sugar moieties to appropriate acceptor polysaccharides in the cell wall, and therefore influence mechanical properties of plant tissues (Scheible & Pauly 2004).

Reduction in wall integrity and cellular adhesion in potato has been linked to low acetylation and caused a stiff texture of tubers (Amos & Mohnen, 2019). This study also found *Manes.08G053200* (chr8) associated with a pectin esterase inhibitor, which regulates the degree of pectin demethylation thus maintaining wall structural integrity (Wormit & Usadel 2018). These findings provide foundation for subsequent studies to further dissect, the genetics of boiled cassava texture. Moreover, the pleiotropy observed for certain genes could be exploited for improving several cooking traits simultaneously.

Genomic prediction for cooking qualities

The emphasis on genomic predictions for cooking qualities arises from the growing need to deliver faster genetic progress on end user preferred traits that have been found to drive variety adoption in cassava. In this context, this work represents the first evaluation of genomic prediction on cassava cooking root attributes. The findings showed that prediction accuracies remained low across traits and with different models ranging from -0.002 to 0.27 (Table 14).

However, there was indication that prediction accuracies for Soft_p and WAB could be raised with increase in training population size (Figure 23).

Impact of prediction models on prediction accuracy

The performance of a genomic prediction model is majorly dependent on the inherent genetic makeup of a trait under study and the assumptions of the model when treating marker effects (Desta *et al*., 2014). Prediction abilities for BA, BRR and RKHS remained superior to GBLUP across most traits (Soft_p, Soft_T and Stiffness. This could be a result of the relaxed assumptions of GBLUP that each of the markers explains equal proportion of total additive genetic variance (Kayondo *et al*., 2018) while assuming absence of epistasis (Lorenzana & Bernardo, 2009), all of which may not hold true for some cooking traits.

These findings contrast with prediction accuracy reported for fruit firmness, a proxy correlated to softness in tomato where GBLUP had moderate cross validation prediction accuracy of 0.61 using a training set of 122 lines and test set of 41 genotypes (Duangjit *et al*., 2016). Elsewhere, prediction ability for apple firmness was found to be moderate using GBLUP and Bayes models (Jung *et al*., 2022). Prior to this study, no attempt had been made to understand genomic prediction ability for cassava textural traits.

The RKHS model gave the highest prediction ability for WAB (r = 0.27) while BA had the lowest accuracy. This may be indicative of the role played by non-additive genetic effects such a dominance and epistasis in the expression of WAB in cassava. In contrast, a study on cooking qualities in beans (1,325 lines comprising four populations) showed Bayes models,

GBLUP and RKHS had varying prediction accuracies (r = 0.04 – 0.69) for water absorption capacity (WAC) across several populations (Diaz *et al*., 2021).

The negative prediction accuracies observed for toughness and stiffness using GBLUP, and RKHS models could be a result of the genotype by environment (G×E) interaction experienced in the C2 population (Ozimati *et al*., 2018), much as the variance for G×E for these traits was lower in magnitude than the genotypic variance. This occurrence has been observed in other genomic prediction studies on cassava (Wolfe *et al*., 2017; Yonis *et al*., 2020; Esuma *et al*., 2021) but no other plausible explanation is given.

Implication of findings to breeding for cooking qualities

These findings unanimously point at the vast role played by plant cell walls in conferring different cooking qualities in cassava, which is in tandem with previous works (Menoli & Beleia, 2007; Uchendu *et al*., 2021). Cell wall content and composition could therefore be a vital breeding target for improvement of cooking qualities in cassava, albeit the need for effective phenotyping methods.

The putative genes discovered in this study present opportunity for further understanding of their expression patterns and how they affect cassava cooking qualities. The genes could also be targets for transgenic and gene editing approaches for accelerated cooking quality improvement in cassava. Further, markers identified should be validated to be certain of their consistency (Rabbi *et al*. 2022) prior to development into Kompetitive allele specific PCR (KASP) markers for large scale deployment.

The study has also provided a catalogue of loci and putative genes that generally improve our understanding of the genetic architecture of cooking qualities in boiled cassava. These findings also show that phenotyping approach used on an association panel influences discovery of significant associations.

Using the penetrometer to measure softness resulted in unearthing of slightly more significant SNPs compared to the texture analyzer for the same trait. However, the texture analyzer was associated with unique significant SNPs, some of which were previously discovered in a similar study by Uchendu *et al*. (2021).

The pleiotropic effects of some loci over several traits could be exploited further for simultaneous improvement of cooking qualities especially if the traits are positively correlated (i.e., Soft_p, Soft_T, Toughness and Stiffness). Also, more significant associations could be unearthed by improving the phenotyping methods for cooking qualities to minimize residuals and by exploring non-linear regression approaches in order to capture non-linear genetic (non-additive) associations (Romagnoni *et al*., 2019; Enoma *et al*., 2022).

Genomic predictions for cooking traits could be enhanced by increasing and optimizing training population size and by including significantly associated markers as covariates in the prediction model (Ozimati *et al*., 2018). This would provide cassava breeders impetus to adopt genomic selection for simultaneous improvement of agronomic as well as end-user preferred cooking traits, leading to development of relevant varieties.

5.5 Conclusion

This study sought to understand the genetic basis of cooking qualities in cassava that are linked to consumer preferences for boiled cassava and genomic prediction ability for the same. To increase chances of discovering significant associations, phenotyping was semi-automated using a penetrometer, texture analyzer and gravimetric assay. The study identified twenty-four SNPs that explained substantial phenotypic variation of the cooking traits. These were found occupying majority of the chromosomes, with many of the significant polymorphisms falling on chromosomes 4 and 18.

The study also identified several candidate genes associated with cell wall modification, and this confirms the vital role cell walls play in modulating texture and cooking time in cassava. Some of the candidate genes had a pleiotropic effect on several traits, which could be useful for synchronized improvement of corelated traits.

The candidate genes identified are also potential targets for transgenic and genome editing approaches of developing desired texture phenotypes. The cell wall modifying proteins could further provide selectable breeding targets for developing varieties that are fast cooking and with a soft texture in genetic backgrounds of important agronomic traits. Genomic prediction accuracy for all traits remained low but could be improved by increasing training population size and diversity. Overall, this work improves our understanding of the genetic complexity around cooking traits in cassava and provides basis for explorative work on genomic based breeding and targets for transcriptomic assessment.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The motivation for this study was to unravel the genetic architecture of boiled cassava texture and its associated traits as well as to test the possibility of using near infrared spectroscopy (NIRS) for phenotyping texture in fresh cassava roots, all geared at equipping cassava breeders with relevant tools to fast-track progress for these traits. Specifically, this study sought to i) determine the phenotypic variability for texture traits in a C2 genomic selection population, ii) develop and validate NIRs predictions for texture and iii) identify genomic regions linked to natural variation for texture through GWAS and to assess genomic prediction accuracies for texture traits.

The variability study found substantial genetic variations among accessions for textural traits. This implies that there is room for achieving genetic progress for textural traits through phenotypic selection. However, environment effects also significantly influenced these traits, implying that it is necessary to evaluate accessions in multiple agroecological conditions prior to selection.

The study identified genotypes UG16F042P002, UG16F307P021, UG16F292P254, UG16F004P004, UG16F293P013, UG16F098P001, UG16F320P032, UG16F060P001, UG16F313P001 and UG16F314P046 as promising candidate parents combining both low softness values (< 2 kgF/cm^2) and high-water absorption (WAB) capacity (> 15%), which are desirable for end users.

The moderate broad sense heritability (Table 1) observed for all traits ($H^2 > 0.4$) shows that there is substantial genetic control of these traits, which could be harnessed through introgression of desirable textural traits from parents to offspring. Additionally, this work showed that improvements in phenotyping resulted in better heritability estimates. The texture analyzer (TA) gave better heritability estimates than the penetrometer for boiled root softness, indicating that the TA was a more reliable tool for assessing texture, in as much as it is amorphous, slow and laborious to use.

The high positive correlations among textural traits found in this study imply that selecting for reduced levels of one texture trait would indirectly reduce levels of other traits, which is desirable for resource optimization during selection. For instance, selecting for low softness thresholds would indirectly lead to simultaneous selection for low toughness and stiffness thresholds. The negative correlation between WAB and other textural traits is also desirable for selecting fast cooking accessions with a soft texture.

Importantly, this study found that the C2 population clustered into two phenotypically dissimilar groups, where the first cluster comprised of accessions that combined high texture thresholds and low WAB whereas the second cluster had accessions that combined low texture thresholds and high WAB.

The accessions in cluster two constitute an invaluable genetic resource for the National cassava breeding program, for identification of promising accessions with desirable cooking trait thresholds that can either be conserved for future use or be fast tracked through advanced selection stages for eventual deployment to end-users in Uganda.

From the near infrared spectroscopy (NIRS) study, quantitative predictions based on partial least squares regression (PLSR) were unable to accurately predict all cooking traits (texture and WAB) indicating there is minimal linearity between cooking traits and spectra ($r^2_{cv} < 0.4$), even under different spectra pre-processing regimes. This has been previously corroborated in other findings by the Roots, tubers and Banana (RTB) foods project that is championing global efforts to develop high throughput phenotyping methods for end-user preferred traits across RTBs.

Using classification-based prediction models, there was a remarkable improvement in NIR prediction accuracy for cooking traits. Using a support vector binary classifier with a linear kernel, WAB the highest accuracy and reliability ($r^2_{cv} = 0.78$; $k_{cv} = 0.46$), rendering the model suitable for early-stage gate screening of clonal or seedling populations for WAB30. Classifiers for other cooking traits suffered low reliability challenges much as the accuracy was high.

Through GWAS, 111 SNPs were found associated with cooking traits (Supp. Table 2) of which 24 were notable. These SNPs were widely distributed across most chromosomes but SNPs on chromosomes 4, 17 and 18 were common to Soft_p, Soft_T and Toughness. Some SNPs on chromosome 18 were common to WAB, Toughness, Soft_p and Soft_T.

This shows there is a pleiotropic effect of some loci over several cooking traits, which can be exploited for simultaneous improvement of affected traits, especially when those traits are positively correlated as observed in this study. For Soft_T and Stiffness, SNPs on chromosome 4 explained most of the phenotypic variation in the traits which underlines the importance of this chromosome in regulating these two traits.

For Soft_p and WAB, SNPs on chromosome 18 explained most of the phenotypic variation. Majority of the SNPs were found tagging genes that have been previously linked with cell wall modifications. This finding could be leveraged for broader interventions on indirect texture modification based on cell walls using genetic engineering and gene editing which are known to deliver fast gains.

Genomic predictions for all cooking traits remained low ($r < 0.4$) but comparable to previous predictions on CBSD and CMD in other studies. This could be attributed to the small population used in this study which could have had insufficient phenotypic variability that could not capture all the SNP diversity.

This study complemented earlier efforts that seek to provide relevant tools and resources for cassava breeders to fast-track gains for end-user traits as well as agronomic superiority to meet market demands in Uganda. Whereas previous improvement strategies have often started with germplasm acquisition, this study sought to understand the available diversity and use it to train high throughput NIR prediction models amenable for rapid phenotyping of cooking traits in cassava. Further, with a comprehensive understanding of the genetic architecture and markers linked to cooking traits, genomic based selection could become the major tool of cassava improvement in Uganda.

Moreover, as Uganda grapples with the prevailing CBSD scourge on cassava, it is critical that all improvement strategies account first, for disease resistance of varieties to ensure food security. Having a stable genetic background for CBSD resistance would make breeding for end-user

preferences more realistic and progress would be achieved. A combination of these two qualities would increase cassava adoption among farmers but also in various food, feed and industrial sectors in Uganda.

RECOMMENDATIONS

Given the substantial influence of environment factors on cooking traits, there is need to test breeding lines in many environments to gauge the true performance and stability of individuals prior to imposing selection.

Similarly, there is need to conduct genotype by environment (GxE) studies for cooking traits over many environments and several seasons in Uganda as basis for identifying environments in which breeding lines should be screened, to guide optimal resource usage during variety development.

To understand the molecular basis of genetic diversity, SNP markers should be used to assess the extent of population stratification within existing breeding lines. This would inform the hybridization strategies to use to maximize heterosis or identify need for boosting genetic diversity through germplasm exchange, mutagenesis or genetic transformation approaches.

There is need to undertake detailed physicochemical profiling of breeding lines along with sensory and instrumental analyses to find proxies linked to texture and WAB traits. Such proxies could be easier to phenotype in early-stage gates thereby providing breeders with means of selection and help minimize genetic drain.

Physicochemical proxies linked to texture and WAB should be used to train NIR linear predictions as a possible means of improving model accuracy. This would improve the quantitative prediction of cooking traits.

To improve NIR classification models, trait thresholds should be determined for each consumer trait and then used to classify phenotype data. This would improve discrimination of one class from another, thereby improving NIR qualitative prediction of cooking traits.

There is need to validate SNP markers found in association with cooking traits to assess their robustness, distribution and reliability. This would inform choice of markers for developing Kompetitive allele specific PCR (KASP) markers that can be deployed in marker assisted breeding strategies for accelerating gains for cooking traits.

Additionally, there is need to conduct gene expression studies to confirm effect of putative genes on cooking traits. The most promising genes with substantial effect on cooking traits would be adopted as candidates for manipulation using more precise breeding tools such as genome editing to enhance cooking traits during variety development.

To enhance genomic prediction accuracy for cooking traits, there is need to train models using larger training populations. This would enhance reliability of models to predict breeding values, leading to the mainstreaming of genomic selection in accelerating genetic gains for cooking traits in cassava.

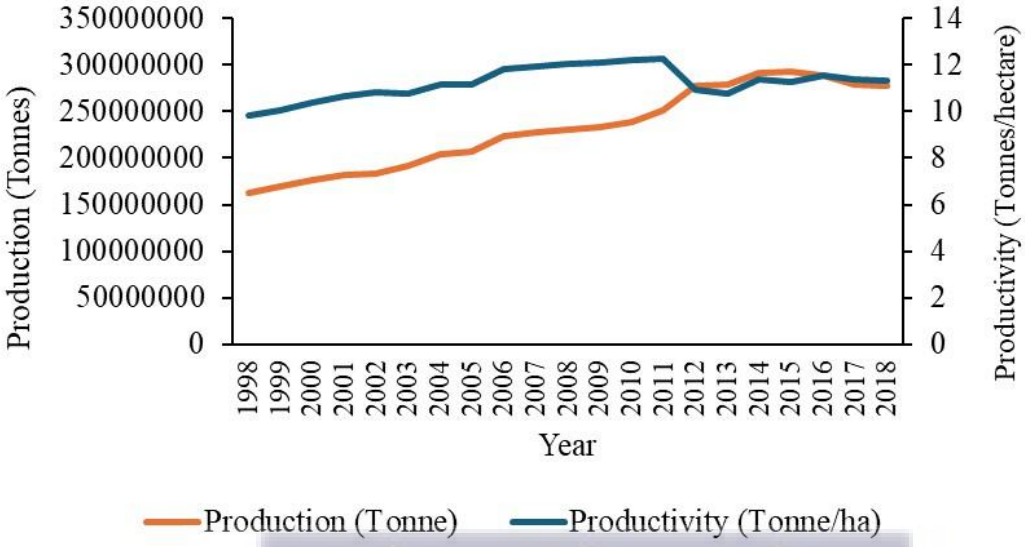
Additionally, the significant SNPs associated with cooking traits should be modeled as covariates to improve genomic prediction accuracy. Further, there is need to explore non-linear models that capture non-additive genetic variance, through machine learning approaches to test

prediction accuracies for cooking traits. This is meant to improve performance of genomic prediction models to enhance their reliability in predicting breeding values for selection.

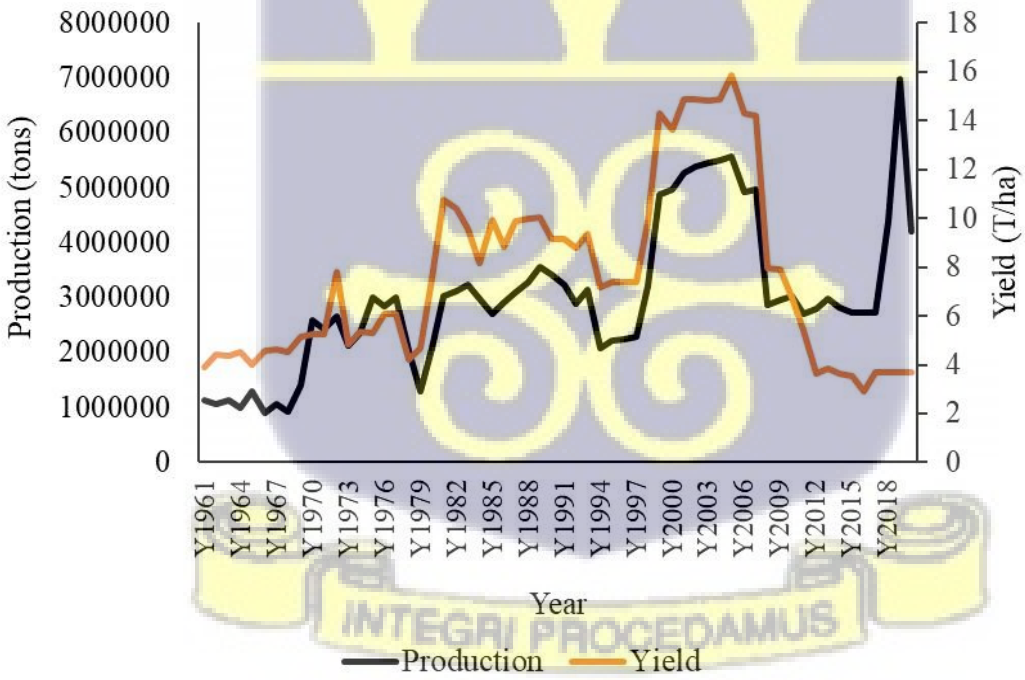
Genotypes that ranked highest with good texture and WAB levels (Supp. Table 1) should be adopted as progenitors for population improvement or evaluated further for variety development.

To enhance association mapping outputs, larger diversity panels than what was used in this study, should be explored to increase chances of finding more significant polymorphisms using more rigorous statistical approaches that account for both additive and non-additive genetic effects.

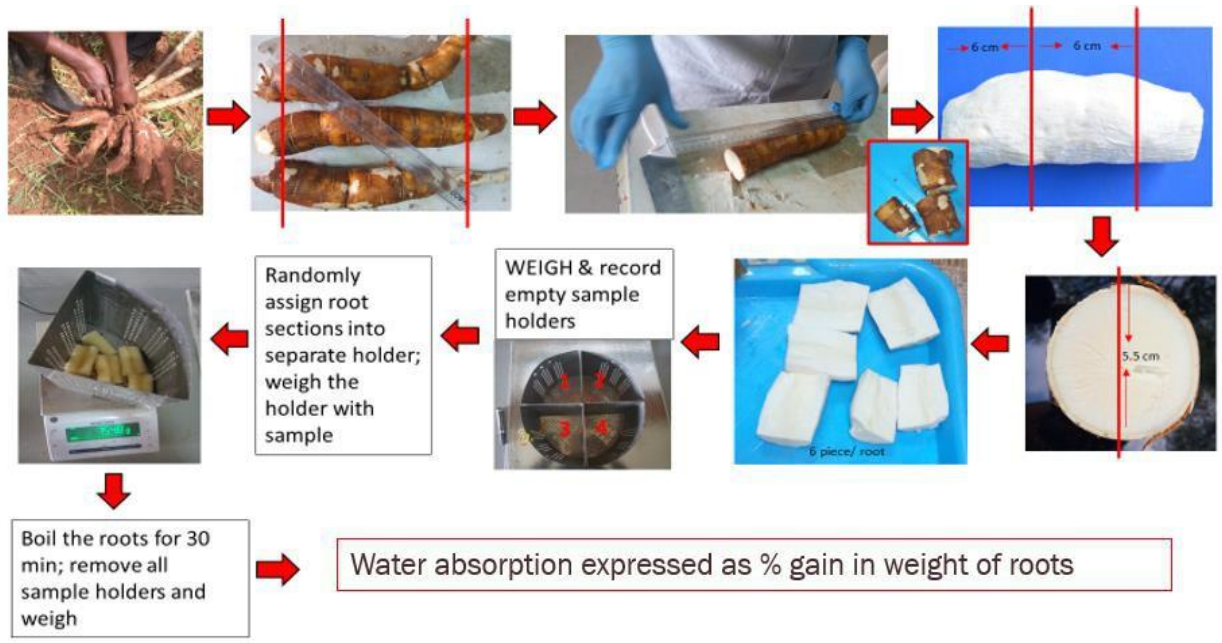
APPENDICES



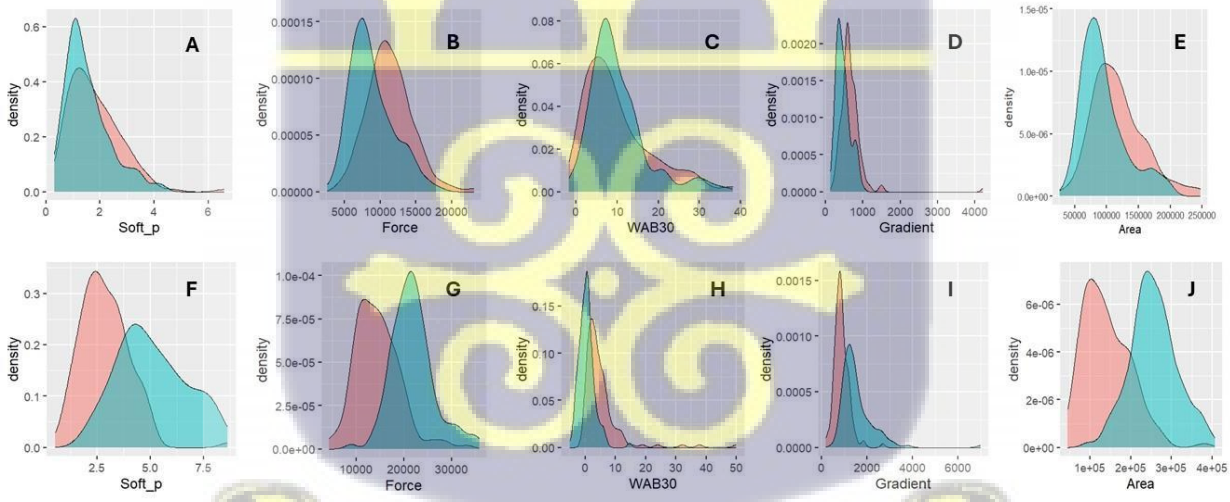
Supplementary Figure 1: Global production and productivity trend of cassava.



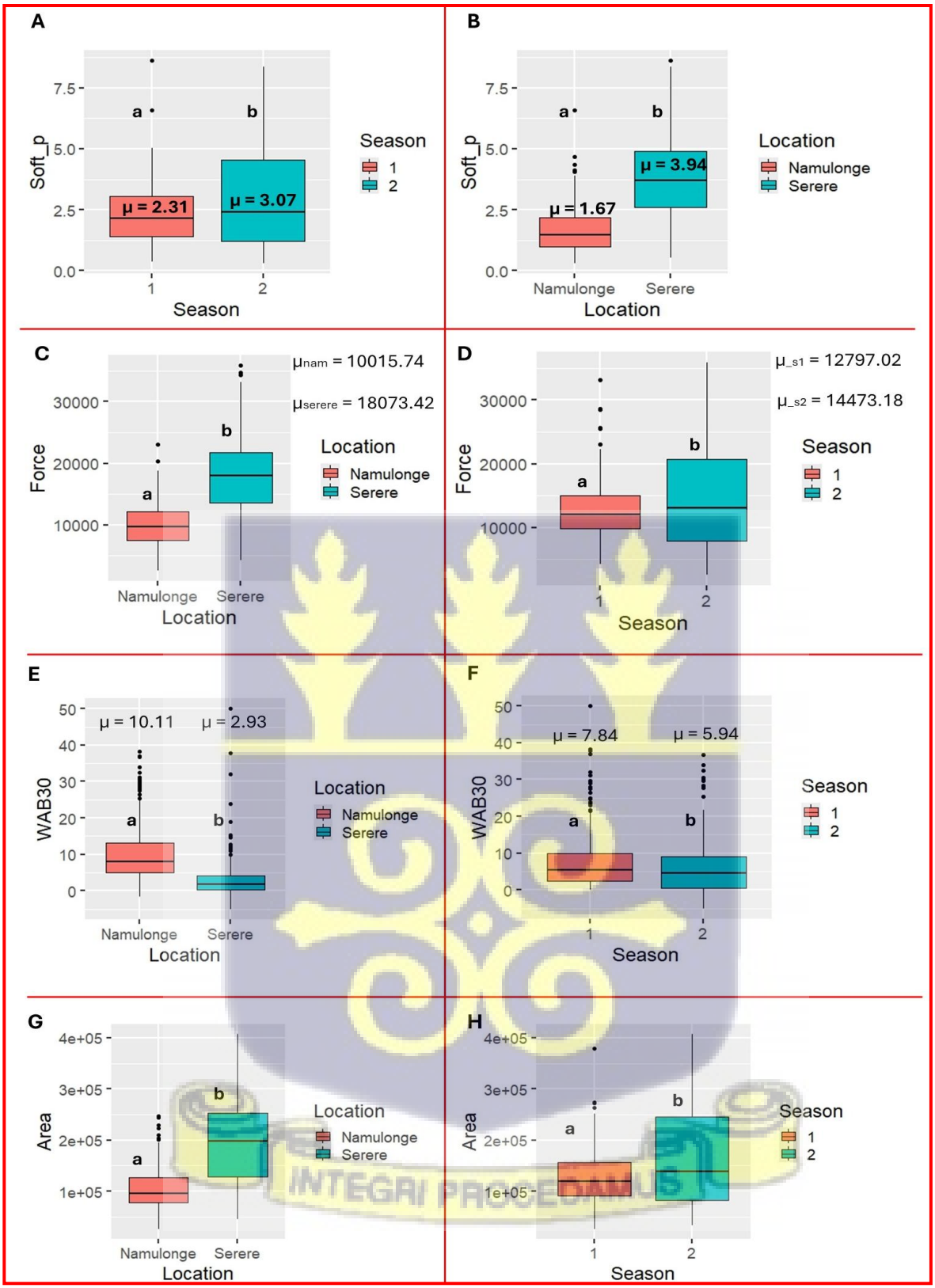
Supplementary Figure 2: Cassava production and productivity in Uganda.



Supplementary Figure 3: Step by step illustration of the assessment of water absorption in cassava roots.



Supplementary Figure 4: Distribution of traits in individual locations across seasons. A – E: Trait distributions for Namulonge environment; F – J: Trait distributions in Serere environment



Supplementary Figure 5: Mean separation for location and season across all cooking traits

Supplementary Table 1: Mean performance of accessions used in this study and their corresponding clusters.

Cluster	Accession	Soft_p	WAB	Soft_T	Stiffness	Toughness
1	MKUMBA	3.32	4.30	16900.22	1127.11	194697.7
1	TMEB204	2.47	1.50	20913.14	1495.66	236903.9
1	UG110017	3.15	4.58	13635.02	774.49	149151.2
1	UG16F002P011	1.98	2.65	16231.89	799.94	167175.1
1	UG16F002P014	2.60	5.14	13907.46	836.81	152439.1
1	UG16F006P001	2.27	10.09	15528.76	978.30	174085.5
1	UG16F050P002	2.15	6.01	14918.83	885.77	133474.7
1	UG16F051P002	3.87	2.76	17928.36	1024.74	200057.9
1	UG16F052P001	3.02	-1.52	13282.09	1563.57	165057.5
1	UG16F057P005	2.35	5.75	14949.41	921.46	152328.4
1	UG16F061P003	4.13	0.65	17819.61	1093.80	218823.5
1	UG16F062P001	2.17	11.50	14516.40	891.49	138639.7
1	UG16F063P002	3.26	7.10	14525.28	802.65	178348.9
1	UG16F066P001	3.78	0.71	15075.76	988.08	176719.4
1	UG16F070P003	2.71	8.49	17647.79	1011.08	188808.9
1	UG16F081P004	3.10	8.49	17884.22	1027.63	182717.7
1	UG16F082P001	3.53	7.63	15970.81	957.09	184665.9
1	UG16F088P001	2.05	9.13	13839.74	884.83	156010.6
1	UG16F091P002	2.85	11.82	14177.37	880.14	141484.6
1	UG16F100P002	2.99	4.33	15839.10	890.50	150307.3
1	UG16F103P002	1.75	6.40	13380.50	831.63	162986.6
1	UG16F110P001	2.53	1.64	14688.55	929.94	171739.4
1	UG16F112P001	2.20	12.61	13548.49	789.27	151558.6
1	UG16F115P001	2.51	4.01	17668.87	1102.24	179209.7
1	UG16F118P001	4.01	0.32	17112.88	1267.12	196582
1	UG16F119P001	4.34	1.99	15415.69	946.63	186775.5
1	UG16F133P021	5.34	-2.66	19357.32	1032.54	245981.7
1	UG16F134P002	2.48	3.26	13620.43	782.99	169264.5
1	UG16F147P001	4.25	-0.02	17876.99	877.41	203388.6
1	UG16F148P001	3.26	5.72	13724.15	744.21	147915.4
1	UG16F150P021	2.19	6.86	14889.22	888.98	128784.9
1	UG16F151P001	3.15	1.17	14452.61	892.91	202459.4
1	UG16F152P002	3.74	-2.46	20078.45	1554.16	273019.4
1	UG16F156P004	3.66	-0.06	16797.42	926.16	189219
1	UG16F162P002	4.27	6.56	22091.91	1694.77	250368.7
1	UG16F163P002	3.42	5.96	18014.39	1249.83	208867.8
1	UG16F165P001	3.09	8.51	14794.56	858.76	156390.3
1	UG16F183P032	3.63	2.53	18271.16	1145.32	150114.3
1	UG16F185P007	4.57	2.09	21947.23	2157.96	245457.3

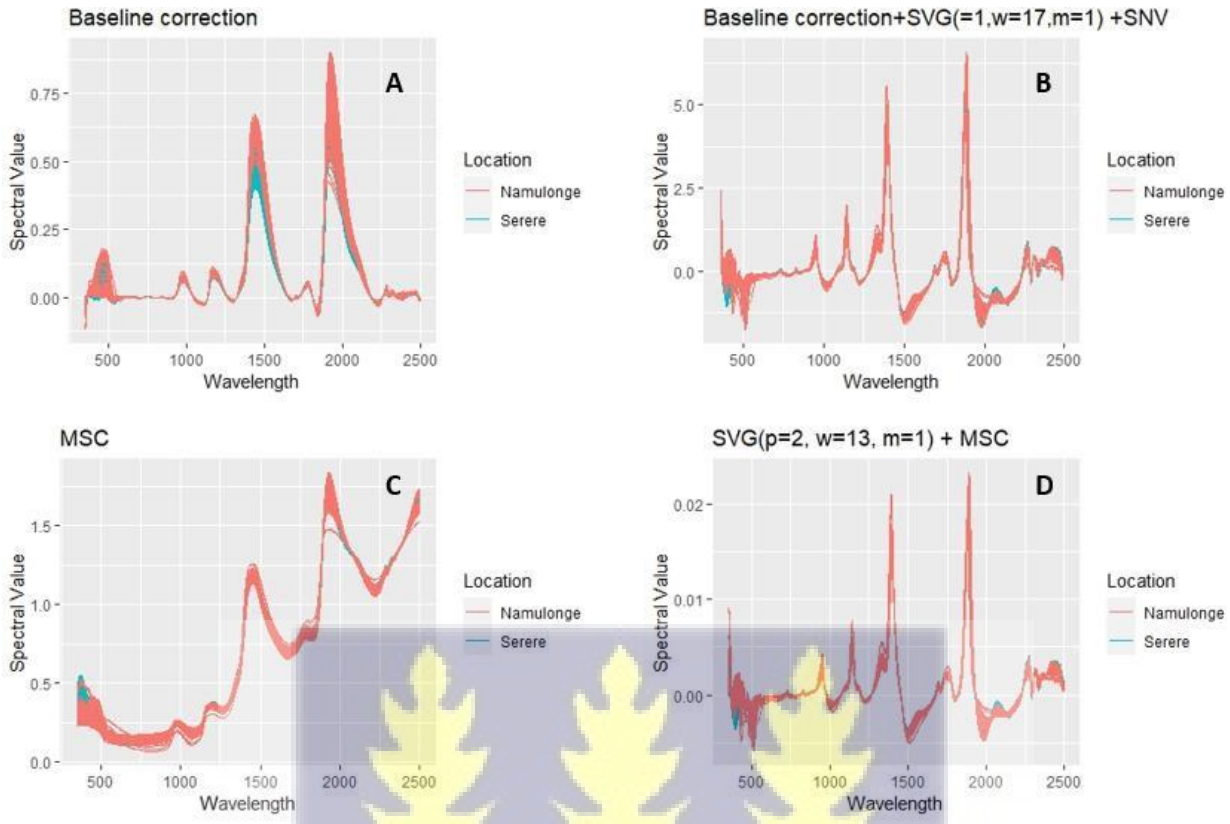
1	UG16F186P006	2.90	0.47	14539.92	1090.27	164888.8
1	UG16F192P004	4.26	2.57	23258.92	1558.09	259248.7
1	UG16F195P005	4.39	1.02	19721.00	1151.94	245376.5
1	UG16F196P002	4.20	6.20	13991.60	1038.65	136783.8
1	UG16F202P002	4.58	3.32	15387.05	836.42	157799.3
1	UG16F207P002	3.28	6.75	14017.05	999.71	154199.7
1	UG16F207P005	3.34	4.99	15415.22	1176.51	170788.3
1	UG16F209P003	2.85	2.08	13331.10	793.66	170836.5
1	UG16F213P002	4.53	1.65	14292.72	977.75	165844.9
1	UG16F215P007	3.08	1.56	18547.72	918.92	216808.9
1	UG16F240P001	2.06	6.07	15938.40	1008.87	138607.1
1	UG16F240P002	3.70	1.57	14084.34	765.80	128207.6
1	UG16F290P061	2.52	6.66	18089.71	1081.78	167197.7
1	UG16F290P093	2.85	4.51	13910.04	889.14	148838.3
1	UG16F290P101	2.82	4.75	14736.61	852.16	175424.8
1	UG16F290P211	3.53	1.47	15351.87	996.20	177594.7
1	UG16F290P236	2.60	12.68	12537.02	737.98	155523.9
1	UG16F290P284	2.69	5.24	21369.23	1270.18	181786.9
1	UG16F290P286	3.90	1.09	17976.30	875.99	192102.5
1	UG16F291P071	3.77	0.82	18783.99	1097.60	154990.4
1	UG16F291P139	2.76	7.64	13007.38	837.93	157846.6
1	UG16F291P143	3.93	0.66	18735.58	1356.39	229354.3
1	UG16F292P023	4.38	14.63	29805.30	2895.84	332533.5
1	UG16F292P073	5.92	-0.71	23608.06	1282.42	224938.4
1	UG16F292P078	5.33	0.55	23876.64	1376.49	273064.9
1	UG16F292P243	3.21	3.44	16034.46	1057.75	183190.8
1	UG16F292P301	3.50	1.48	15305.40	832.69	140803.4
1	UG16F293P102	3.52	14.24	15809.55	2174.60	147277.5
1	UG16F293P183	4.29	2.23	19412.98	1295.39	181103.8
1	UG16F293P201	3.27	0.49	18916.28	916.89	193223.6
1	UG16F294P036	5.13	4.28	18673.91	1270.45	213454.4
1	UG16F294P054	4.46	4.55	19893.63	1240.94	171460.2
1	UG16F294P081	3.44	5.52	18172.31	1175.94	198473
1	UG16F294P095	2.69	10.44	13191.55	855.38	163291.8
1	UG16F295P041	4.47	2.35	18739.38	999.05	206742.8
1	UG16F296P022	2.50	3.49	15502.52	900.54	138024.5
1	UG16F296P025	4.33	-0.95	18521.23	1057.86	209590.2
1	UG16F296P211	3.72	3.09	18261.63	1274.60	194770.1
1	UG16F299P005	3.09	4.32	15040.67	921.29	168282.7
1	UG16F300P009	3.00	1.71	15400.25	812.76	167436.3
1	UG16F300P026	4.87	0.47	18271.30	1159.98	212687.1
1	UG16F300P041	4.81	0.89	19589.19	1240.68	222022.5
1	UG16F300P047	2.67	0.34	14127.36	900.26	200007.6
1	UG16F300P053	2.67	1.99	15962.54	938.60	179569.6

1	UG16F300P066	4.99	0.31	18850.42	1377.17	259024.6
1	UG16F301P020	3.17	4.51	17326.49	1072.16	182061.8
1	UG16F301P023	3.72	1.21	15265.43	1019.92	199495.5
1	UG16F304P001	3.67	1.57	15916.44	919.45	194823
1	UG16F305P024	2.39	0.89	15439.87	945.54	169807.2
1	UG16F306P001	3.81	1.35	15415.30	987.10	165501.3
1	UG16F308P021	2.98	0.21	15895.31	1023.18	187559.1
1	UG16F308P041	5.75	-3.57	22704.68	2212.85	245357.4
1	UG16F308P079	2.30	2.53	15307.67	774.39	159641.3
1	UG16F308P091	4.39	6.49	14884.60	887.22	170620.3
1	UG16F311P015	4.29	4.32	14170.88	846.72	168642.5
1	UG16F312P023	3.04	2.82	16997.04	1112.71	184341.5
1	UG16F312P028	4.49	3.43	17098.06	1023.83	189034.2
1	UG16F312P031	7.42	-1.34	31277.19	7001.00	267858.2
1	UG16F313P006	3.54	1.38	18077.18	1017.24	174306.9
1	UG16F313P013	3.01	6.40	13764.23	1074.53	162163.6
1	UG16F313P034	4.09	4.79	18465.95	1783.08	213945.4
1	UG16F313P044	2.63	-1.74	14143.44	652.93	194020.5
1	UG16F313P075	3.95	1.57	17301.92	1085.30	215577.2
1	UG16F315P005	2.74	1.14	15904.17	835.11	158025.5
1	UG16F316P004	3.89	1.04	13460.90	911.07	150170.5
1	UG16F316P011	5.44	-0.71	27853.93	2200.31	307065.3
1	UG16F316P023	3.79	2.41	25143.23	2126.79	291807.9
1	UG16F318P006	3.09	1.01	18844.04	1258.21	222308.7
1	UG16F318P014	2.98	-0.54	14068.24	920.76	173413.7
1	UG16F319P002	3.93	2.23	13884.67	1245.79	167016.6
1	UG16F319P016	3.57	3.06	18118.55	977.29	208950.4
1	UG16F319P026	3.32	4.87	15441.74	1298.67	186659
1	UG16F320P008	4.16	3.59	19025.74	1239.51	167559.4
2	UG16F001P002	1.89	7.59	12571.13	659.39	139728.8
2	UG16F001P006	2.75	4.38	11624.72	659.38	116097.5
2	UG16F002P004	1.24	15.61	9016.79	597.10	66278.05
2	UG16F004P004	0.50	21.75	4395.56	230.91	43824.19
2	UG16F004P014	1.41	13.81	11026.95	592.60	108232.5
2	UG16F005P002	1.70	16.94	7687.08	400.05	81280.93
2	UG16F005P014	1.57	6.55	6976.29	360.46	70833.11
2	UG16F040P006	2.52	18.79	13065.95	967.35	153355.9
2	UG16F042P002	0.38	36.86	5851.73	270.46	58902.55
2	UG16F043P001	1.29	11.69	7555.57	447.12	77308.53
2	UG16F055P002	1.52	11.11	8882.58	476.54	85433.58
2	UG16F059P001	1.88	5.44	8057.61	369.02	90049.72
2	UG16F060P001	0.71	29.01	10024.17	575.67	102461
2	UG16F064P002	1.80	8.12	9279.09	464.94	90776.74
2	UG16F065P002	2.27	13.19	12974.15	685.84	114540.7

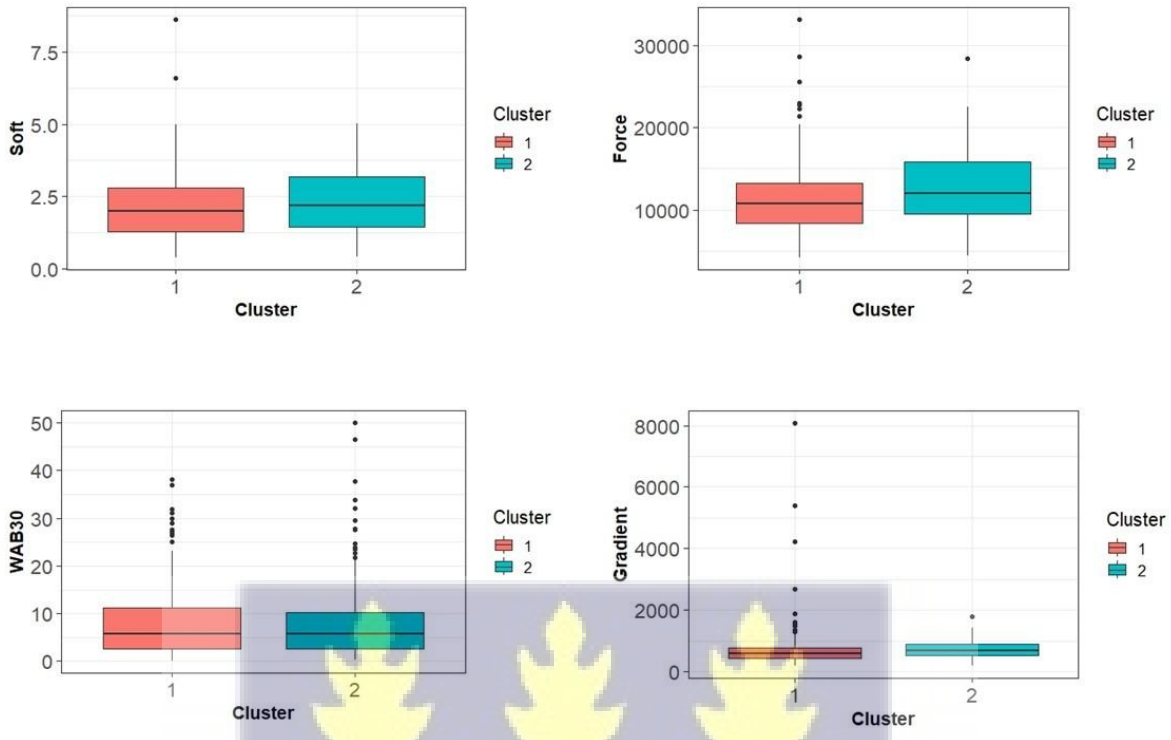
2	UG16F068P001	3.47	3.20	13919.65	711.76	154874.9
2	UG16F073P002	1.38	17.16	12274.02	835.13	103977.4
2	UG16F074P004	2.51	6.05	11283.21	797.76	130437.9
2	UG16F075P004	2.62	13.75	13202.73	833.42	131983.9
2	UG16F078P001	0.84	11.14	11235.52	590.18	103035
2	UG16F089P002	1.92	8.21	9028.54	472.23	92115.14
2	UG16F097P001	1.85	6.67	14878.02	752.56	140783.1
2	UG16F098P001	0.66	15.91	5548.97	329.20	49457.55
2	UG16F113P001	0.97	11.55	6977.33	337.63	69071.27
2	UG16F121P001	1.31	12.67	5473.99	245.51	60427.97
2	UG16F125P001	0.98	15.62	11673.96	728.25	93240.1
2	UG16F131P001	0.84	18.50	7152.40	387.99	75164.27
2	UG16F133P004	1.18	5.41	6433.43	278.65	82982.49
2	UG16F136P001	2.03	12.63	11581.12	555.46	114514.3
2	UG16F140P001	2.80	4.14	10828.98	710.90	122620.3
2	UG16F141P001	1.40	11.95	12307.45	661.87	125034.7
2	UG16F142P001	1.17	9.33	7598.10	376.66	74262.59
2	UG16F147P002	0.92	10.13	7858.72	407.15	80789.95
2	UG16F150P001	1.93	10.19	10457.25	595.84	118510.7
2	UG16F158P005	2.51	3.48	12934.44	666.07	131048.1
2	UG16F172P012	1.93	8.48	9254.86	503.87	96799.05
2	UG16F177P001	1.59	9.56	11100.29	694.08	97288.88
2	UG16F182P001	1.13	13.67	10396.94	677.93	89054.17
2	UG16F183P015	1.56	18.10	8095.14	482.58	83596.75
2	UG16F183P041	1.03	19.42	5745.22	318.27	55222.68
2	UG16F183P065	0.78	18.83	11395.88	629.20	110172
2	UG16F184P008	1.27	11.68	9324.57	516.05	85061.89
2	UG16F190P007	1.56	8.62	14455.55	1126.52	147380.2
2	UG16F191P002	0.96	14.90	9274.64	528.31	92428.2
2	UG16F192P001	1.17	8.48	10611.05	597.97	109583.2
2	UG16F194P001	2.73	7.12	10682.35	529.34	128808.4
2	UG16F198P002	1.41	10.07	8892.79	617.04	85807.18
2	UG16F200P001	1.13	7.68	8192.11	384.48	94274.65
2	UG16F204P003	4.11	2.80	7545.84	336.77	103923.2
2	UG16F210P001	1.70	7.42	10167.76	539.97	96099.98
2	UG16F211P003	1.20	8.12	7175.64	410.88	81405.82
2	UG16F220P007	2.66	15.40	13296.21	859.94	109558.7
2	UG16F221P002	2.45	5.88	9211.45	525.51	83952.47
2	UG16F232P002	3.43	2.27	11818.65	774.89	92235.65
2	UG16F290P144	3.09	3.23	13415.03	737.98	138443.3
2	UG16F290P173	0.94	9.66	12808.13	830.75	121110.1
2	UG16F290P191	1.86	7.09	11142.53	616.93	115614.8
2	UG16F290P218	2.11	6.15	13153.28	656.13	137016
2	UG16F290P238	1.86	5.25	13141.23	765.55	134479

2	UG16F290P298	1.23	10.81	10231.68	611.60	90885.37
2	UG16F291P014	2.56	6.29	10333.84	599.92	89645.62
2	UG16F291P058	1.51	3.36	6748.38	349.82	60680.06
2	UG16F291P132	2.31	1.87	11787.75	625.79	105195
2	UG16F291P187	3.14	3.44	10959.86	746.59	132913.4
2	UG16F292P003	1.45	9.99	13197.54	1140.66	101750.4
2	UG16F292P045	2.00	16.16	7788.04	443.94	70839.18
2	UG16F292P072	3.37	4.49	11280.15	766.35	86580.13
2	UG16F292P091	1.20	10.69	11097.32	705.65	90526.93
2	UG16F292P179	2.72	2.71	13963.45	669.77	136058.5
2	UG16F292P230	1.73	6.17	12409.16	709.66	115105.4
2	UG16F292P254	0.49	19.71	7562.88	375.15	89083.13
2	UG16F292P314	1.70	6.50	10434.84	545.48	115675.1
2	UG16F293P013	0.59	15.08	5301.83	292.92	54174.65
2	UG16F293P067	0.41	31.13	11394.99	616.33	118075.5
2	UG16F293P081	1.34	9.13	11146.99	636.72	105237.5
2	UG16F293P092	0.90	9.23	5776.55	271.79	63994.96
2	UG16F293P137	1.82	15.70	14545.91	841.20	137991.9
2	UG16F293P197	0.65	19.83	9312.90	558.19	81946.69
2	UG16F294P003	3.43	10.15	11807.32	786.37	123630.2
2	UG16F294P016	2.06	7.55	11081.16	630.57	117444.8
2	UG16F294P019	0.91	13.11	5118.09	254.84	58666.59
2	UG16F294P042	1.73	8.32	11958.48	802.13	100146.8
2	UG16F294P083	1.77	5.91	5959.93	287.67	71130.39
2	UG16F294P086	1.68	7.19	4949.98	236.22	51918.08
2	UG16F295P047	1.72	4.83	11164.52	620.84	108638
2	UG16F295P056	1.88	17.74	12413.63	803.02	134697.6
2	UG16F297P004	1.92	13.20	12104.67	561.51	132068.5
2	UG16F300P036	1.51	9.91	7673.65	374.58	78154.73
2	UG16F300P042	0.78	19.90	7179.58	411.97	66810.84
2	UG16F300P043	2.93	4.43	10701.72	673.68	90802.31
2	UG16F300P062	1.93	3.61	13630.42	758.63	127476.2
2	UG16F301P009	4.47	5.14	10248.35	559.51	126147.4
2	UG16F301P012	0.92	12.42	6865.51	347.83	72454.19
2	UG16F302P002	2.02	24.47	10621.72	533.13	116452.8
2	UG16F302P021	1.66	11.79	10889.08	765.03	85780.28
2	UG16F304P032	2.34	4.52	10243.96	502.72	100331.9
2	UG16F304P092	3.03	7.85	8605.19	503.02	81719.16
2	UG16F304P180	2.70	2.20	10803.83	668.68	111827.5
2	UG16F304P182	2.51	6.34	10810.62	655.49	93524.89
2	UG16F304P184	2.86	8.40	10684.55	601.26	116248.6
2	UG16F307P021	0.43	25.37	7991.59	383.60	88666.88
2	UG16F307P024	1.59	11.25	7244.01	462.17	79297.66
2	UG16F307P026	2.11	6.11	10139.18	576.36	96658.44

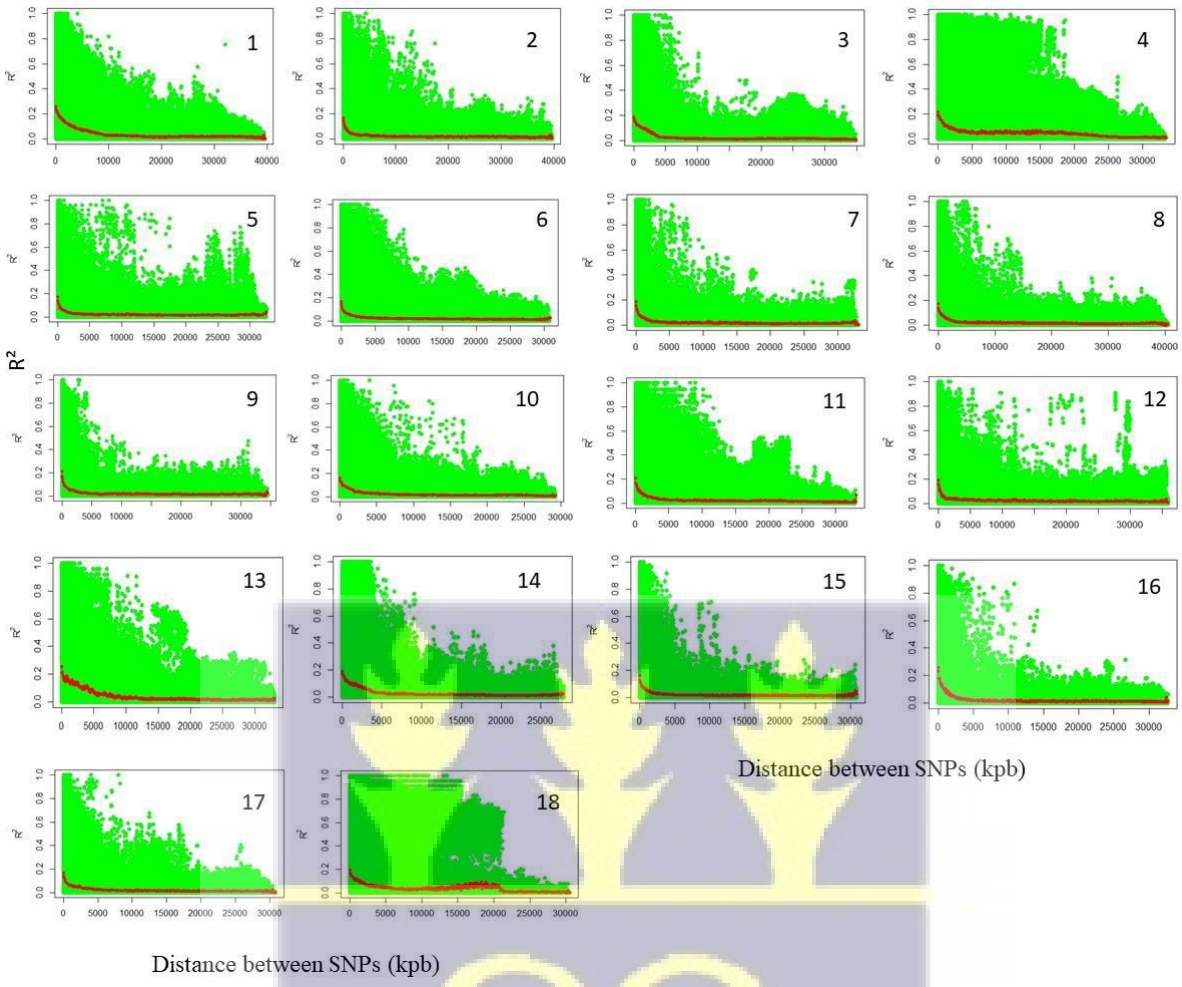
2	UG16F308P059	3.10	1.13	10792.08	708.08	89704.9
2	UG16F310P003	2.85	4.65	10564.76	600.37	117494.7
2	UG16F310P004	3.49	8.78	10655.36	511.23	124999.6
2	UG16F310P009	1.31	18.97	9334.86	518.03	89643.99
2	UG16F312P026	2.31	6.35	12317.22	827.95	120398.8
2	UG16F313P001	1.05	29.71	7161.17	473.45	86156.07
2	UG16F313P004	1.31	8.93	8448.80	427.45	87180.96
2	UG16F313P005	2.49	2.67	7921.96	580.28	96534.44
2	UG16F313P011	2.70	0.98	11599.63	770.42	121492.6
2	UG16F313P024	1.29	12.65	11174.93	829.50	85831.1
2	UG16F313P033	3.27	2.15	12289.45	662.91	158838.7
2	UG16F313P051	1.15	8.26	6096.26	319.48	58631.05
2	UG16F314P041	1.64	6.67	11508.42	574.10	121170.9
2	UG16F314P045	0.95	31.10	8955.76	629.35	70159.54
2	UG16F314P046	1.09	21.66	7594.53	404.04	80922.26
2	UG16F315P001	1.43	13.80	5626.88	276.57	56070.12
2	UG16F316P010	0.95	20.16	9894.50	492.00	128850.9
2	UG16F316P012	2.25	2.20	13674.87	853.43	127997.1
2	UG16F318P003	1.20	5.88	11710.42	571.91	118805.1
2	UG16F318P004	1.40	9.54	7651.34	370.15	90400.25
2	UG16F318P013	2.37	4.30	13827.39	692.73	140117.9
2	UG16F318P016	1.88	2.68	11653.99	594.05	127384.3
2	UG16F318P024	2.21	5.21	6791.55	359.52	72835.97
2	UG16F318P025	1.68	3.22	12826.05	632.06	131993.6
2	UG16F319P007	0.98	10.53	8150.28	396.03	89995.39
2	UG16F319P011	2.34	5.86	9944.29	812.95	134197.1
2	UG16F319P029	2.75	0.81	12116.18	691.01	142920.5
2	UG16F319P032	0.93	9.31	6520.22	308.26	69262.22
2	UG16F320P002	1.93	7.15	10709.67	665.16	104256.9
2	UG16F320P003	0.92	16.66	6406.41	320.45	68672.21
2	UG16F320P009	1.49	9.79	11639.49	769.64	133256.6
2	UG16F320P017	2.26	1.26	13121.96	917.14	97649.82
2	UG16F320P024	1.61	2.96	10347.98	530.38	109196
2	UG16F320P032	0.68	22.84	8959.52	574.84	79138.81



Supplementary Figure 6: Effect of spectra treatments Baseline correction (A), baseline +SVG+SNV (B), multiple scatter correction (MSC) and SVG with MSC (D)



Supplementary Figure 7: Performance of different clusters for Soft_p, Soft_T, WAB30 and Gradient



Supplementary Figure 8: Chromosome specific LD - decay curves showing different rates of decay.

Supplementary Table 2: All genome-wide significant SNPs associated with texture traits and water absorption in cassava.

Trait	SNP	Chrom	Position	P.value	maf	R^2 mod	effect	% PVE
Soft_p[a]	S1L2151431	1	2151431	5.09E-68	0.50	0.98	0.19	0.25
Soft_p	S1L2164300	1	2164300	1.19E-12	0.49	0.95	0.05	0.00
Soft_p	S2L16500325	2	16500325	5.75E-10	0.07	0.66	-0.07	2.04
Soft_p	S4L3881302	4	3881302	1.59E-06	0.29	0.98	-0.04	0.38
Soft_p	S4L31359043	4	31359043	1.06E-06	0.13	0.82	0.05	0.32
Soft_p	S10L20288297	10	20288297	2.12E-45	0.48	0.99	-0.10	0.09
Soft_p	S10L20815786	10	20815786	6.63E-65	0.42	0.98	-0.14	0.00
Soft_p	S10L20903044	10	20903044	3.66E-09	0.48	0.98	-0.03	0.00
Soft_p	S10L20975749	10	20975749	2.26E-49	0.49	0.99	-0.11	0.00
Soft_p	S14L24139366	14	24139366	1.15E-06	0.14	0.85	-0.04	0.64
Soft_p	S17L23789456	17	23789456	6.31E-08	0.16	0.81	0.05	0.78
Soft_p	S18L4890744	18	4890744	1.98E-10	0.09	0.86	0.07	0.00
Soft_p	S18L4974443	18	4974443	2.2E-103	0.10	0.97	0.44	2.31
Soft_p	S18L4992526	18	4992526	1.24E-22	0.09	0.95	0.11	0.00
Soft_p	S18L5206934	18	5206934	1.04E-14	0.11	0.93	0.08	0.00
Soft_p	S18L5254501	18	5254501	8.06E-07	0.10	0.76	0.05	0.00
Soft_p	S18L5469618	18	5469618	1.91E-07	0.11	0.81	0.05	0.00
Soft_T	S4L2144235	4	2144235	2.21E-50	0.29	0.97	-0.18	0.00
Soft_T	S4L30955244	4	30955244	1.8E-20	0.14	1.00	0.11	0.00
Soft_T	S4L31007604	4	31007604	7.48E-15	0.14	1.00	0.09	0.00
Soft_T	S4L31011535	4	31011535	5.15E-18	0.13	1.00	0.10	0.00
Soft_T	S4L31042027	4	31042027	2.18E-22	0.15	1.00	0.11	0.00
Soft_T	S4L31051618	4	31051618	4.21E-18	0.15	1.00	0.10	0.00
Soft_T	S4L31093822	4	31093822	7.89E-17	0.15	1.00	0.09	0.22
Soft_T	S4L31094633	4	31094633	4.21E-18	0.15	1.00	0.10	0.00
Soft_T	S4L31110972	4	31110972	6.08E-20	0.14	1.00	0.11	3.49
Soft_T	S4L31111492	4	31111492	1.15E-08	0.19	1.00	0.05	0.00
Soft_T	S4L31137980	4	31137980	1.78E-18	0.15	1.00	0.10	1.04
Soft_T	S4L31158648	4	31158648	8.42E-21	0.14	1.00	0.11	0.00
Soft_T	S4L31193930	4	31193930	1.8E-20	0.14	1.00	0.11	0.00
Soft_T	S4L31199376	4	31199376	3.75E-20	0.15	1.00	0.10	0.00
Soft_T	S4L31200835	4	31200835	1.78E-18	0.15	1.00	0.10	4.55
Soft_T	S4L31246645	4	31246645	6.31E-11	0.16	1.00	0.07	0.00
Soft_T	S4L31249019	4	31249019	1.78E-18	0.15	1.00	0.10	15.37
Soft_T	S4L31261983	4	31261983	1.18E-24	0.14	1.00	0.13	0.00
Soft_T	S4L31320230	4	31320230	1.8E-20	0.14	1.00	0.11	0.00
Soft_T	S4L31320947	4	31320947	4.21E-18	0.15	1.00	0.10	0.00
Soft_T	S4L31411919	4	31411919	4.79E-22	0.15	1.00	0.11	0.00

Soft_T	S4L31507436	4	31507436	1.48E-25	0.15	1.00	0.13	0.00
Soft_T	S4L31533184	4	31533184	9.29E-08	0.14	0.99	0.06	49.10
Soft_T	S4L31545964	4	31545964	1.8E-20	0.14	1.00	0.11	0.00
Soft_T	S4L31546043	4	31546043	7.34E-14	0.15	1.00	0.08	0.00
Soft_T	S4L31549479	4	31549479	1.8E-20	0.14	1.00	0.11	0.00
Soft_T	S4L31554179	4	31554179	4.21E-18	0.15	1.00	0.10	0.00
Soft_T	S4L31555044	4	31555044	1.8E-20	0.14	1.00	0.11	0.00
Soft_T	S4L31559183	4	31559183	1.8E-20	0.14	1.00	0.11	0.00
Soft_T	S4L31567761	4	31567761	1.8E-20	0.14	1.00	0.11	0.00
Soft_T	S4L31613479	4	31613479	2.98E-19	0.15	1.00	0.10	0.00
Soft_T	S4L31618658	4	31618658	1.8E-20	0.14	1.00	0.11	0.00
Soft_T	S4L31636455	4	31636455	3.15E-20	0.14	1.00	0.11	0.00
Soft_T	S4L31659450	4	31659450	1.8E-20	0.14	1.00	0.11	0.00
Soft_T	S4L31664456	4	31664456	1.8E-20	0.14	1.00	0.11	0.00
Soft_T	S4L31674609	4	31674609	1.73E-10	0.15	1.00	0.07	5.71
Soft_T	S4L31709909	4	31709909	1.93E-08	0.13	1.00	0.06	0.00
Soft_T	S4L31717733	4	31717733	1.71E-08	0.15	0.99	0.06	0.00
Soft_T	S4L31793897	4	31793897	4.32E-16	0.15	1.00	0.09	0.00
Soft_T	S5L9644039	5	9644039	6.58E-90	0.28	0.99	-0.27	2.55
Soft_T	S8L5043186	8	5043186	6.68E-64	0.47	0.99	0.17	0.02
Soft_T	S11L2768424	11	2768424	9.8E-75	0.08	0.95	-0.33	7.02
Soft_T	S11L31356213	11	31356213	8.2E-112	0.27	0.97	-0.45	2.31
Soft_T	S17L23789456	17	23789456	1.19E-06	0.16	0.97	0.05	2.82
Soft_T	S17L23808459	17	23808459	8.63E-49	0.15	1.00	0.22	0.49
Soft_T	S17L23813265	17	23813265	1.37E-06	0.13	0.94	0.06	0.00
Soft_T	S18L4890744	18	4890744	1.61E-39	0.09	1.00	0.20	0.00
Soft_T	S18L4940687	18	4940687	1.2E-06	0.09	0.95	0.06	0.00
Soft_T	S18L4974443	18	4974443	1.43E-34	0.10	1.00	0.17	5.31
Soft_T	S18L4992526	18	4992526	3.73E-36	0.09	1.00	0.19	0.00
Soft_T	S18L5206934	18	5206934	2.73E-45	0.11	1.00	0.21	0.00
Soft_T	S18L5469618	18	5469618	1.82E-34	0.11	1.00	0.17	0.00
Soft_T	S18L6610954	18	6610954	2.32E-14	0.23	0.96	-0.06	0.00
WAB[c]	S5L18056419	5	18056419	1.17E-07	0.25	0.79	-0.69	3.67
WAB	S5L27322223	5	27322223	3.09E-98	0.28	0.98	-4.94	0.00
WAB	S6L21272537	6	21272537	3.94E-10	0.25	1.00	-0.81	10.18
WAB	S6L23860699	6	23860699	2.1E-129	0.15	0.84	8.97	0.18
WAB	S6L27073265	6	27073265	1.49E-09	0.36	0.93	-0.68	2.60
WAB	S7L29635941	7	29635941	1.25E-35	0.18	0.94	2.93	25.97
WAB	S8L6899246	8	6899246	5.95E-25	0.07	1.00	2.45	0.00
WAB	S10L2115716	10	2115716	1.5E-153	0.48	0.88	9.54	2.03
WAB	S11L2973212	11	2973212	3.13E-12	0.06	0.64	1.50	7.71
WAB	S11L2996798	11	2996798	1.59E-06	0.06	0.67	0.95	0.00
WAB	S11L8566741	11	8566741	1.64E-56	0.12	1.00	-3.54	6.78

WAB	S13L28732449	13	28732449	4.41E-07	0.37	0.75	0.52	1.94
WAB	S13L29157709	13	29157709	1.53E-39	0.40	1.00	-1.71	1.48
WAB	S13L29379434	13	29379434	5.25E-10	0.24	0.81	0.73	5.74
WAB	S15L7407184	15	7407184	6.35E-15	0.17	1.00	1.07	0.00
WAB	S16L27770001	16	27770001	4.97E-11	0.44	0.85	0.70	4.50
WAB	S18L4753895	18	4753895	6.32E-08	0.16	0.83	-0.85	0.00
WAB	S18L4787558	18	4787558	1.84E-15	0.11	0.90	-1.48	0.00
WAB	S18L4890744	18	4890744	8.9E-150	0.09	0.95	-15.60	0.00
WAB	S18L4974443	18	4974443	1.76E-12	0.10	0.84	-1.36	27.21
WAB	S18L4992526	18	4992526	1.58E-42	0.09	0.94	-3.38	0.00
WAB	S18L5206934	18	5206934	2.16E-10	0.11	0.84	-1.18	0.00
WAB	S18L5254501	18	5254501	3.27E-11	0.10	0.85	-1.29	0.01
WAB	S18L5277008	18	5277008	1.04E-08	0.11	0.81	-1.06	0.00
WAB	S18L5469618	18	5469618	2.73E-20	0.11	0.90	-1.87	0.00
Toughness[d]	S2L16500325	2	16500325	1.49E-21	0.07	0.75	-0.21	30.37
Toughness	S4L2144235	4	2144235	2.7E-08	0.29	0.96	-0.08	1.73
Toughness	S4L31042027	4	31042027	2.22E-45	0.15	1.00	0.30	0.42
Toughness	S4L31111492	4	31111492	1.24E-17	0.19	1.00	0.14	0.37
Toughness	S4L31158648	4	31158648	8.39E-41	0.14	1.00	0.29	3.84
Toughness	S4L31613479	4	31613479	1.94E-07	0.15	0.98	0.09	0.00
Toughness	S8L5043186	8	5043186	4.49E-07	0.47	0.97	0.06	0.00
Toughness	S8L6327008	8	6327008	1.67E-16	0.47	1.00	0.09	3.20
Toughness	S11L146943	11	146943	4.02E-09	0.12	0.58	-0.10	27.05
Toughness	S17L23789456	17	23789456	8.39E-54	0.16	1.00	0.34	17.66
Toughness	S17L23793678	17	23793678	8.7E-08	0.13	0.93	0.10	0.00
Toughness	S17L23808459	17	23808459	2.88E-07	0.15	0.89	0.09	0.00
Toughness	S17L23813265	17	23813265	1.03E-14	0.13	0.97	0.16	0.00
Toughness	S17L25596778	17	25596778	6.86E-07	0.12	0.96	0.09	0.00
Toughness	S18L4890744	18	4890744	6.5E-56	0.09	1.00	0.45	0.00
Toughness	S18L4974443	18	4974443	8.98E-52	0.10	1.00	0.39	15.37
Toughness	S18L4992526	18	4992526	1.01E-52	0.09	1.00	0.42	0.00
Toughness	S18L5206934	18	5206934	8.27E-55	0.11	1.00	0.41	0.00
Toughness	S18L5469618	18	5469618	3.32E-08	0.11	0.98	0.11	0.00
Stiffness[e]	S1L6072007	1	6072007	1.8E-125	0.06	0.60	-0.21	7.72
Stiffness	S4L30963834	4	30963834	1.39E-08	0.24	1.00	-0.02	19.83
Stiffness	S4L31533120	4	31533120	4.26E-09	0.16	1.00	-0.02	0.00
Stiffness	S4L31674609	4	31674609	4.01E-08	0.15	1.00	-0.02	0.00
Stiffness	S4L32681922	4	32681922	3.7E-111	0.08	0.77	-0.16	15.88
Stiffness	S5L9644039	5	9644039	1.42E-68	0.28	0.99	-0.06	8.08

REFERENCES

Abincha, W., Ikeogu, U. N., Kawuki, R., Egesi, C., Rabbi, I., Parkes, E., Kulakow, P., Edema, R., Gibson, P., Owor, B.-E., & Biancolillo, A. (2021). Portable Spectroscopy Calibration with Inexpensive and Simple Sampling Reference Alternatives for Dry Matter and Total Carotenoid Contents in Cassava Roots. *Mdpi.Com*. https://doi.org/10.3390/app11041714.

Adebayo, G.W. (2023). Cassava production in africa: A panel analysis of the drivers and trends. *Heliyon 9* (2023) e19939, https://doi.org/10.1016/j.heliyon.2023.e19939.

Adinsi, L., Honfozo, F., Hotegni, F., & Moussa, I. D. (2023). Standard Operating Protocol for sample preparation, determination of instrumental texture of steam-cooked cassava. Bio-physical characterization of quality traits*, WP2*. https://doi.org/10.18167/agritrop/00723.

Adjei, E. A., Esuma, W., Alicai, T., Bhattacharjee, R., Dramadri, I. O., Agaba, R., Chamba, E. B., & Odong, T. L. (2022). Phenotypic Diversity within Ugandan Yam (Dioscorea species) Germplasm Collection. *International Journal of Agronomy, Volume 2022, Article ID 5826012, 10 pages*, https://doi.org/10.1155/2022/5826012.

AGRA. (2018). AGRA Uganda Operational Plan. https://agra.org/wpcontent/uploads/2018/01/agra-uganda final.pdf.

Ahmed, M. (2023). Utilization of Pectin in the Food and Drug Industries. https://www.intechopen.com/books/12308.

Al-Naggar, A. M. M., Shafik, M. M., Musa, R. Y. M., Thi, P., & Ha, T. (2020). Genetic diversity based on morphological traits of 19 maize genotypes using principal component analysis and GT biplot. *Annual Research & Review in Biology, 2020*, *35*(2), 68–85. https://doi.org/10.9734/ARRB/2020/v35i230191.

Alamu, E. O., Adesokan, M., Asfaw, A., & Maziya-Dixon, B. (2020a). Effect of Sample Preparation Methods on the Prediction Performances of Near Infrared Reflectance Spectroscopy for Quality Traits of Fresh Yam (Dioscorea spp.). *Applied Sciences Volume 10, Issue 17,* 10.3390/app10176035 https://doi.org/10.3390/app10176035.

Alamu, E. O., Nuwamanya, E., Cornet, D., Meghar, K., Adesokan, M., Tran, T., Belalcazar, J., Desfontaines, L., & Davrieux, F. (2020b). Near-infrared spectroscopy applications for high-throughput phenotyping for cassava and yam: A review. *International Journal of Food Science and Technology*, *56*(3), 1491–1501. https://doi.org/10.1111/ijfs.14773.

Amirvaresi, A., & Parastar, H. (2023). Miniaturized NIR spectroscopy and chemometrics: A smart combination to solve food authentication challenges. *Frontiers in Analytical Science*, *3*. https://doi.org/10.3389/FRANS.2023.1118590/PDF.

Adjei, A.E., Esuma, W., Alicai, T., Bhattacharjee, R., Dramadri, I. O., Edema, R., Chamba, E. B., & Odong, T. L. (2023). Genetic diversity and population structure of Uganda's yam (Dioscorea spp.) genetic resource based on DArTseq. *PLoS ONE*, *18*(2 February), 1–16. https://doi.org/10.1371/journal.pone.0277537.

Amuge, T., Berger, D. K., Katari, M. S., Myburg, A. A., Goldman, S. L., & Ferguson, M. E. (2017). A time series transcriptome analysis of cassava (Manihot esculenta Crantz) varieties challenged with Ugandan cassava brown streak virus. *Scientific Reports*, *7*(1), 1–21. https://doi.org/10.1038/s41598-017-09617-z.

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., & Cairns, J. E. (2018). Translating High-Throughput Phenotyping into Genetic Gain. *Trends in Plant Science*, *23*(5), 451–466. https://doi.org/10.1016/j.tplants.2018.02.001.

Awad-Allah, M.M.A., Shafie, W.W.M., Alsubeie, M.S., Alatawi, A., Safhi, F.A., Salha Mesfer ALshamrani, S.M., Albalawi, D.A., Al-Amrah, H., Alshehri, D., Alshegaihi, R.M., Basahi, M.A and Masrahi, A.S (2022). Utilization of Genetic Resources, Genetic Diversity and Genetic Variability for Selecting New Restorer Lines of Rice. *Genes* 2022, 13(12), 2227; https://doi.org/10.3390/genes13122227.

Ayetigbo, O., Latif, S., Abass, A and Müller, J. (2018). Comparing characteristics of root, flour and starch of biofortified yellow-flesh and white-flesh cassava variants, and sustainability considerations: A review. *Sustainability, Volume 10, Issue 9*, 10.3390/su10093089, *Mdpi.Com*.

Bai, X., Zhang, L., Kang, C., Quan, B., Zheng, Y., Zhang, X., Song, J., Xia, T and Wang, M. (2022). Near-infrared spectroscopy and machine learning-based technique to predict

quality-related parameters in instant tea. *Scientific Reports 2022 12:1*, *12*(1), 1–8. https://doi.org/10.1038/s41598-022-07652-z.

Bajwa, S.G., Rupe, J.C and Mason, J. (2017). Soybean disease monitoring with leaf reflectance. *Remote Sensing, 2017, 9, 127*; doi:10.3390/rs9020127, *Mdpi.Com*.

Balabin, R. M., Safieva, R. Z and Lomakina, E. I. (2007). Comparison of linear and nonlinear calibration models based on near infrared (NIR) spectroscopy data for gasoline properties prediction. *Chemometrics and Intelligent Laboratory Systems*, *88*(2), 183–188. https://doi.org/10.1016/J.CHEMOLAB.2007.04.006.

Barandica, O. J., Pérez, J. C., Lenis, J. I., Calle, F., Morante, N., Pino, L., Hershey, C. H and Ceballos, H. (2016). Cassava breeding II: Phenotypic correlations through the different stages of selection. *Frontiers in Plant Science*, *7*, https://doi.org/10.3389/FPLS.2016.01649/FULL.

Basile, T., Marsico, A. D and Perniola, R. (2021). NIR analysis of intact grape berries: Chemical and physical properties prediction using multivariate analysis. *Foods*, *10*(1). https://doi.org/10.3390/foods10010113.

Bates, D., Mächler, M., Bolker, B. M and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01.

Beć, K. B., Grabska, J and Huck, C. W. (2021). NIR spectroscopy of natural medicines supported by novel instrumentation and methods for data analysis and interpretation. *Journal of Pharmaceutical and Biomedical Analysis*, *193*. https://doi.org/10.1016/j.jpba.2020.113686.

Bechoff, A., Tomlins, K., Fliedel, G., Becerra Lopez-lavalle, L. A., Westby, A., Hershey, C and Dufour, D. (2018). Cassava traits and end-user preference: Relating traits to consumer liking, sensory perception, and genetics. *Critical Reviews in Food Science and Nutrition*, *58*(4), 547–567. https://doi.org/10.1080/10408398.2016.1202888.

Behera, D.T.K., Gaikwad, A.B., Saxena, S., Bharadwaj, C and Munshi, A.D (2012). Morphological and molecular analyses define the genetic diversity of Asian bitter gourd ('Momordica charantia'L.). *Australian Journal of Crop Science,* AJCS 6(2):261-267 (2012), https://search.informit.org/doi/abs/10.3316/INFORMIT.054435151250212.

Beléia, A., Prudencio-Ferreira, S. H., Yamashita, F., Sakamoto, T. M and Ito, L. (2005). Sensory and instrumental texture analysis of cassava (Manihot esculenta, crantz) roots. *Journal of Texture Studies*, *35*(5), 542–553. https://doi.org/10.1111/j.1745-4603.2004.35505.x.

Ben-Ari, G and Lavi, U. (2012). Marker-assisted selection in plant breeding. In *Plant Biotechnology and Agriculture*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-381466-1.00011-0.

Berro, I., Lado, B., Nalin, R. S., Quincke, M., & Gutiérrez, L. (2019). Training Population Optimization for Genomic Selection. *The Plant Genome*, *12*(3). https://doi.org/10.3835/PLANTGENOME2019.04.0028.

Holland, B. (2007). Genetic architecture of complex traits in plants. *Elsevier*, *Current Opinion in Plant Biology*, 10:156–161, 2023, https://doi.org/10.1016/j.pbi.2007.01.003, https://www.sciencedirect.com/science/article/pii/S1369526607000076.

Bobelyn, E., Serbanc, A-S., Nicuc, M and Lammertynb, J (2010). Postharvest quality of apple predicted by NIR-spectroscopy: Study of the effect of biological variability on spectra and model performance. *Postharvest Biology and Technology* 55 (2010) 133–143. doi:10.1016/j.postharvbio.2009.09.006.

Blaschek, M., Roudier, P., Poggio, M and Hedley, C.B (2019). Prediction of soil available water-holding capacity from visible near-infrared reflectance spectra. *Scientific Reports*, (2019) 9:12833, https://doi.org/10.1038/s41598-019-49226-6.

Boeriu, C., Yuksel, D., van der Vuurst de Vries, R., Stolle-Smits, T and van Dijk, C. (1998). Correlation between near infrared spectra and texture profiling of steam cooked potatoes. *Journal of Near Infrared Spectroscopy*, 6, A291 – A297.

Bokanga Mpoko. (1999). *CASSAVA: Post-harvest Operations*., *INPhO - Post harvest compendium*, Food And Agriculture Organinsation (FAO), International Institute of tropical Agriculture (IITA), http://www.cgiar.org/iita/.

Borku, A.W., Tora, T.T and Masha, M (2025). Cassava in focus: A comprehensive literature review, its production, processing landscape, and multi-dimensional benefits to society. *Food Chemistry Advances* 7 (2025) 100945.

Brumlop, S., and Finckh. (2011). Applications and potentials of marker assisted selection (MAS) in plant breeding. Final report of the F+E project "Applications and Potentials of Smart Breeding" (FKZ 350 889 0020). Federal Agency for Nature Conservation.Burns, D and Ciurczak, E. (2007). Handbook of near-infrared analysis. https://books.google.com/books?hl=en&lr=&id=6EEd1a0uka0C&oi=fnd&pg=PP1&dq=Ha ndbook+of+infrared+spectroscopy+Burns+third+edition&ots=L497P2JSsx&sig=Ul5-Pr4AhmmcA2iNqLQOI9Gu_P8.

Bush, W. S and Moore, J. H. (2012). Chapter 11: Genome-Wide Association Studies. *PLoS Computational Biology*, *8*(12). https://doi.org/10.1371/JOURNAL.PCBI.1002822.

Carmo, C. D. do, Silva, M. S. da, Oliveira, G. A. F., Oliveira, E. J. de, Carmo, C. D. do, Silva, M. S. da, Oliveira, G. A. F and Oliveira, E. J. de. (2015). Molecular-assisted selection for resistance to cassava mosaic disease in Manihot esculenta Crantz. *Scientia Agricola*, *72*(6), 520–527. https://doi.org/10.1590/0103-9016-2014-0348.

Ceballos, H., Rojanaridpiched, C., Phumichai, C., Becerra, L.A., Kittipadakul, P., Iglesias, C and Gracen, V.E (2020). Excellence in cassava breeding: perspectives for the future. *Crop Breeding, Genetic and Genomics*. 2020;2(2):e200008. https://doi.org/10.20900/cbgg20200008 *Cbgg.Hapres.Com*.

Ceballos, H., Davrieux, F., Talsma, E., Belalcazar, J., Chavarriaga, P and Andersson, M.S (2017). Carotenoids in cassava roots. *Carotenoids, Intechopen,* Chapter 12, *Books.Google.Com*.

Ceballos, H., Hershey, C and Becerra-López-Lavalle, L. A. (2012). New Approaches to Cassava Breeding. *Plant Breeding Reviews*, *36*, 427–504, https://doi.org/10.1002/9781118358566.ch6.

Ceballos, H., Hershey, C., Iglesias, C and Zhang, X. (2021). Fifty years of a public cassava breeding program: evolution of breeding objectives, methods, and decision-making processes. *Theoretical and Applied Genetics*, *134*(8), 2335–2353. https://doi.org/10.1007/S00122-021-03852-9.

Ceballos, H., Iglesias, C. A., Pérez, J. C and Dixon, A. G. O. (2004). Cassava breeding: Opportunities and challenges. *Plant Molecular Biology*, *56*(4), 503–516. https://doi.org/10.1007/S11103-004-5010-5.

Ceballos, H., Kawuki, R. S., Gracen, V. E., Yencho, G. C and Hershey, C. H. (2015). Conventional breeding, marker-assisted selection, genomic selection and inbreeding in clonally propagated crops: a case study for cassava. In *Theoretical and Applied Genetics*, (Vol. 128, Issue 9). https://doi.org/10.1007/s00122-015-2555-4.

Ceballos, H., Pérez, J. C., Barandica, O. J., Lenis, J. I., Morante, N., Calle, F., Pino, L and Hershey, C. H. (2016). Cassava breeding I: The value of breeding value. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/FPLS.2016.01227/FULL.

Cécillon, L., Barthès, B. G., Gomez, C., Ertlen, D., Genot, V., Hedde, M., Stevens, A and Brun, J. J. (2009). Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (NIRS). *European Journal of Soil Science*, *60*(5), 770–784. https://doi.org/10.1111/J.1365-2389.2009.01178.X.

Chakraborty, R., Chakraborty, S., Dutta, B.K and Paul, S.B (2009). Genetic variability and genetic correlation among nutritional and cooking quality traits in bold grain rice. *ORYZA-An International Journal on Rice, 2009, Indianjournals.Com*. https://www.indianjournals.com/ijor.aspx?target=ijor:oryza&volume=46&issue=1&article=005.

Chan, J., Leow, Y-L., Bea, S. M. H., Cheng, K. T., Phoong, W.K., Hong, S.W and Chen, Z-W. (2022). Mitigating the multicollinearity problem and its machine learning approach: a review. *Mathematics*, 10(8), 1283; https://doi.org/10.3390/math10081283 *Mdpi.Com*.

Chang, L., Toghiani, S., Ling, A., Aggrey, S.E and Rekaya, R. (2018). High density marker panels, SNPs prioritizing and accuracy of genomic selection. *BMC Genetics* (2018) 19:4 DOI 10.1186/s12863-017-0595-2.

Chawade, A., Ham, J.V., Blomquist, H., Bagge, O., Alexandersson, E and Ortiz, R (2019). High- Throughput Field-Phenotyping Tools for Plant Breeding and Precision Agriculture. *Agronomy* 2019, 9, 258; doi:10.3390/agronomy9050258, *Mdpi.Com*.

Chen, Y and Lübberstedt, T (2010). Molecular basis of trait correlations. *Cell.ComY Chen, T Trends in Plant Science, cell.Com*. 1385(10)00097-X, DOI: 10.1016/j.tplants.2010.05.004 https://www.cell.com/trends/plant-science/fulltext/S1360-.

Enyew, M., Dejene, T., Lakew, B and Wored, F (2019). Clustering and principal component analysis of Barley (*Hordeum volugare* L.) Landraces for major morphological traits from North Western Ethiopia. *International Journal of Agricultural Science and Food Technology* 5(1): 058-063. DOI: http://doi.org/10.17352/2455-815X.000043.

Costell, E and Durán, L (2009). Food texture: sensory evaluation. *Food Engineering* - Volume II, https://books.google.co.ug/books/about/Food_Engineering_Volume_II.html?id=JO2yCwAAQBAJ&redir_esc=y.

Crase, S., Hall, B and Thennadil, S. (2021). Cluster Analysis for IR and NIR Spectroscopy: Current Practices to Future Perspectives. Computers, Materials and Continua, 69(2), 1945-1965. https://doi.org/10.32604/cmc.2021.018517.

Crossa, J., Pérez, P., Hickey, J., Burgueño, J., Ornella, L., Ceró N-Rojas, J., Zhang, X., Dreisigacker, S., Babu, R., Li, Y., Bonnett, D and Mathews, K. (2014). Genomic prediction in

CIMMYT maize and wheat breeding programs. *Heredity,* 2014, 112, 48–60. https://doi.org/10.1038/hdy.2013.16, nature.com.

Daemo, B.B., Yohannes, D.B., Beyene, T.M and Abtew, W.B (2023). Genetic variability and divergence analysis in cassava (Manihot esculenta Crantz) accessions for storage root yield and other agronomic characters. *Journal of Agriculture and Food Research* 14, (2023) 100874, *Elsevier,* https://doi.org/10.1016/j.jafr.2023.100874, https://www.sciencedirect.com/science/article/pii/S2666154323003812.

Davrieux, F., Belalcazar, J., Zhang, X., & Tran, T. (2022). NIRS & biophysical analyses: Tentative prediction of cassava cooking properties-*Year 2. High-Throughput Phenotyping Protocols (HTPP), WP3*. https://doi.org/10.18167/agritrop/00715. https://agritrop.cirad.fr/603474/.

Davrieux, F., Dufour, D., Dardenne, P., Belalcazar, J., Pizarro, M., Luna, J., Londoño, L., Jaramillo, A., Sanchez, T., Morante, N., Calle, F., Becerra Lopez-Lavalle, L. A and Ceballos, H. (2016). LOCAL regression algorithm improves near infrared spectroscopy predictions when the target constituent evolves in breeding populations. *Journal of Near Infrared Spectroscopy*, *24*(2), 109–117. https://doi.org/10.1255/jnirs.1213.

Davrieux, F., Zhang, X., Londoño, L., Belalcazar, J and Tran, T (2023). Cassava Cooking Properties Characterization using NIRS on Fresh Ground Cassava; High-Throughput

Phenotyping Protocols (HTPP), WP3. St Pierre, La Réunion, France: *RTBfoods Calibration Report*, 27 p. https://doi.org/10.18167/agritrop/00732.

de Andrade, L. R. B., Sousa, M. B. e., Wolfe, M., Jannink, J. L., de Resende, M. D. V., Azevedo, C. F., & de Oliveira, E. J. (2022). Increasing cassava root yield: Additive-dominant genetic models for selection of parents and clones. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/FPLS.2022.1071156.

Delwiche, S. R and Reeves, J. B. (2010). A Graphical Method to Evaluate Spectral Preprocessing in Multivariate Regression Calibrations: Example with Savitzky – Golay Filters and Partial Least Squares Regression. Applied Spectroscopy Vol. 64, Issue 1, pp. 73-82 (2010). https://opg.optica.org/as/abstract.cfm?URI=as-64-1-73.

de Oliveira, E. J., de Resende, M. D. V., da Silva Santos, V., Ferreira, C. F., Oliveira, G. A. F., da Silva, M. S., de Oliveira, L. A and Aguilar-Vildoso, C. I. (2012). Genome-wide selection in cassava. *Euphytica*, *187*(2), 263–276. https://doi.org/10.1007/s10681-012-0722-0.

de Oliveira, E. J., Morgante, C. V., de Tarso Aidar, S., de Melo Chaves, A. R., Antonio, R. P., Cruz, J. L and Filho, M. A. C. (2017). Evaluation of cassava germplasm for drought tolerance under field conditions. *Euphytica*, *213*(8). https://doi.org/10.1007/S10681-017-1972-7.

Desta, Z. A and Ortiz, R (2014). Genomic selection: genome- wide prediction in plant improvement. *Trends in Plant Science*, Volume 19, Issue 9, p592-601, DOI: 10.1016/j.tplants.2014.05.006 *Elsevier*. https://www.sciencedirect.com/science/article/pii/S1360138514001411.

Dossou-Aminon, I., Loko, L.Y., Adjatin, A., Ewédjè, E.B.K., Dansi, A., Rakshit, S., Cissé, N., Patil, J.V., Agbangla, C., Sanni, A., Akoègninou, A and Akpagana, K (2024). Genetic Divergence in Northern Benin Sorghum (Sorghum bicolor L. Moench) Landraces as Revealed by Agromorphological Traits and Selection of Candidate Genotypes. *The Scientific World Journal*, Volume 2015, http://dx.doi.org/10.1155/2015/916476, https://www.hindawi.com/journals/tswj/2015/916476/abs/.

DoVale, J.C., Carvalho, H.F., Sabadinb, F and Fritsche-Neto, R (2021). Reduction of genotyping marker density for genomic selection is not an affordable approach to long- term breeding in cross-pollinated crops. *BioRxiv,* 2021, https://doi.org/10.1101/2021.03.05.434084.

Dufour, D., Hershey, C., Hamaker, B. R and Lorenzen, J. (2020). Integrating end-user preferences into breeding programmes for roots, tubers and bananas. *International Journal of Food Science and Technology*, *56*(3), 1071–1075. https://doi.org/10.1111/ijfs.14911.

Dumont, T., Thong-Chane, A., Barau, L., Siegmund, B and Hoarau, J. Y. (2019). Genetic

Variabilities and Genetic Gains for Yield Components in Regional Sugarcane Breeding Programmes on Réunion Island. *Sugar Technology*, *21*(6), 868–878. https://doi.org/10.1007/s12355-019-00718-9.

Esuma, W., Herselman, L., Labuschagne, M.T., Ramu, P., Lu, F., Baguma, Y., Buckler, E.S and Kawuki, R.S (2016). Genome-wide association mapping of provitamin A carotenoid content in cassava. *Euphytica*, 212, 97–110 (2016). https://doi.org/10.1007/s10681-016-1772-5.

Esuma, W., Kawuki, R.S, Herselman, L and Labuschagne, M.T (2016). Diallel analysis of provitamin A carotenoid and dry matter content in cassava (Manihot esculenta Crantz). *Breeding Science*, 66: 627–635 (2016), doi:10.1270/jsbbs.15159.

Esuma, W., Kawuki, R.S, Herselman, L and Labuschagne, M.T (2016). Stability and genotype by environment interaction of provitamin A carotenoid and dry matter content in cassava in Uganda. *Breeding Science,* 66: 434–443 (2016) doi:10.1270/jsbbs.16004.

Esuma, W., Nanyonjo, A. R., Miiro, R., Angudubo, S and Kawuki, R. S. (2019). Men and women's perception of yellow-root cassava among rural farmers in eastern Uganda. *Agriculture & Food Security*, *8*(1), 10. https://doi.org/10.1186/s40066-019-0253-1.

Esuma, W., Ozimati, A., Kulakow, P., Gore, M. A., Wolfe, M. D., Nuwamanya, E., Egesi, C and Kawuki, R. S. (2021). Effectiveness of genomic selection for improving provitamin A

carotenoid content and associated traits in cassava. *G3 Genes, Genomes, Genetics*, *11*(9). https://doi.org/10.1093/G3JOURNAL/JKAB160.

Esuma, W., Ozimati, A., Kulakow, P., Gore, M.A., Wolfe, M.D., Nuwamanya, E., Egesi, C and Robert S Kawuki (2021). Effectiveness of genomic selection for improving provitamin A carotenoid content and associated traits in cassava. G3 Genes, Genomes, Genetics, Volume 11, Issue 9, September 2021, https://doi.org/10.1093/g3journal/jkab160.

Fahlgren, N., Gehan, M and Baxter, I (2015). Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Current Opinion in Plant Biology* 2015, 24:93–99, http://dx.doi.org/10.1016/j.pbi.2015.02.006.

Fan, Z., Gao, Y., Guo, L., Cao, Y., Liu, R and Zhang, Q (2019). Phenotypic variations and heritability in hybrid populations of bearded iris. *Hortscience* 54(6):988–992. 2019. https://doi.org/10.21273/HORTSCI13884-19.

FAOSTAT Statistical Database (2024). Rome, Italy: Food and Agriculture Organization of the United Nations. Available at: https://www.fao.org/faostat/en/.

Fatumah Namakula, B., Nuwamanya, E., Kanaabi1, M., Gibson, P., Wembabazi, E., Paula, I., & Kawuki, R. S. (2023). NIRS Predictions, Phenotypic Variability and Optimization of Cooking Time for Evaluation of the Root Softness of Boiled Cassava. *Uganda Journal of Agricultural Sciences, 2021*, *21*(2), 41–53. https://doi.org/10.4314/ujas.v21i2.4.

Ferguson, M. E., Shah, T., Kulakow, P and Ceballos, H. (2019). A global overview of cassava genetic diversity. *PLoS ONE*, *14*(11), 1–16. https://doi.org/10.1371/journal.pone.0224763.

Ferguson, M., Rabbi, I., Kim, D-J., Gedil, M., Lopez-Lavalle, L. A. B and Okogbenin, E. (2012). Molecular Markers and Their Application to Cassava Breeding: Past, Present and Future. *Tropical Plant Biology*, *5*(1), 95–109. https://doi.org/10.1007/s12042-011-9087-0.

Ferris, R. S. B., Muganga, A., Matovu, R., Kolijn, S., Hagenimana, V and Karuni, E. (2002). Marketing opportunities for starch and high-quality flour production from cassava and sweet potato in Uganda. 1–88. http://books.google.com.co/books?id=uR7xAAAAMAAJ.

Forkman, J., Josse, J and Piepho, H-P (2019). Hypothesis tests for principal component analysis when variables are standardized. *Journal of Agricultural, Biological, and Environmental Statistics*, Volume 24, Number 2, Pages 289–308, https://doi.org/10.1007/s13253-019-00355-5.

Forsythe, L., Tufan, H., Bouniol, A., Kleih, U and Fliedel, G. (2020). An interdisciplinary and participatory methodology to improve user acceptability of root, tuber and banana varieties. *International Journal of Food Science and Technology*, 1–9. https://doi.org/10.1111/ijfs.14680.

Franck, H., Christian, M., Noël, A., Brigitte, P., Joseph, H.D., Cornet, D and Mathurin, N.C (2011). Effects of cultivar and harvesting conditions (age, season) on the texture and taste of boiled cassava roots. Food Chemistry, Volume 126, Issue 1, 1 May 2011, Pages 127-133, https://doi.org/10.1016/j.foodchem.2010.10.088.

Fystro, G. The prediction of C and N content and their potential mineralisation in heterogeneous soil samples using Vis–NIR spectroscopy and comparative methods. Plant and Soil 246, 139–149 (2002). https://doi.org/10.1023/A:1020612319014.

Gadissa, F., Abebe, M and Bekele, T (2021). Agro- morphological traits-based genetic diversity assessment in Ethiopian barley (Hordeum vulgare L.) landrace collections from Bale highlands, Southeast Ethiopia. *Agriculture & Food Security*, *10*(1). https://doi.org/10.1186/s40066-021-00335-4.

Garces, V., García-Quintero, A., Lerma, T. A., Palencia, M., Combatt, E. M and Arrieta, Á. A. (2021). Characterization of Cassava Starch and Its Structural Changes Resulting of Thermal Stress by Functionally-Enhanced Derivative Spectroscopy (FEDS). *Polysaccharides 2021, Vol. 2, Pages 866-877*, *2*(4), 866–877. https://doi.org/10.3390/POLYSACCHARIDES2040052.

Gebhardt, C., Urbany, C and Stich, B. (2014). Dissection of Potato Complex Traits by Linkage and Association Genetics as Basis for Developing Molecular Diagnostics in Breeding

Programs. *Genomics of Plant Genetic Resources* (pp. 47–85). Springer Netherlands. https://doi.org/10.1007/978-94-007-7575-6_3.

Gianola, D and van Kaam, J.B.C.H.M (2008). Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics*, Volume 178, Issue 4, 1 April 2008, Pages 2289–2303, https://doi.org/10.1534/genetics.107.084285.

Gómez, D., Salvador, P., Sanz, J and Casanova, J.L (2019). Potato yield prediction using machine learning techniques and sentinel 2 data. *Remote Sensing* 2019, 11(15), 1745; https://doi.org/10.3390/rs11151745, *Mdpi.Com*.

González-Recio, O., Gianola, D., Long, N., Weigel, K.A., Rosa, G.J.M and Avendaño, S (2008). Nonparametric methods for incorporating genomic information into genetic evaluations: an application to mortality in broilers. *Genetics*, Volume 178, Issue 4, 1 April 2008, Pgs 2305–2313, https://doi.org/10.1534/genetics.107.084293, https://academic.oup.com/genetics/article- abstract/178/4/2305/6073912.

Graffham, A., Kleih, U., Jagwe, J., Nabawanuka, J., Wanda, K., Kalunda, P and Ntibarikure, G. (2000). Industrial markets for starch-based products: an assessment of the industrial potential for cassava in Uganda. Natural Resources Institute *(NRI), FOODNET*/International Institute of Tropical Agriculture (IITA), Kawanda Agricultural Research Institute, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.202.4372&rep=rep1&type=pdf.

Grosjean, P., Ibanez, F and Etienne, M (2024). Package for Analysis of Space-Time Ecological Series (Pastecs). *R Cran*, https://eagle.fish.washington.edu/trilobite/notebooks/steven/ea-gle.fish.washington.edu/derm ochelys/R/win-library/3.0/pastecs/doc/pastecs.pdf.

Gu, Z (2022). Complex heatmap visualization. *iMeta Wiley 1*(3). https://doi.org/10.1002/imt2.43.

Guimarães, P. E., Ruane, J., Beate D. Scherf, Andrea Sonnino and James D. Dargie (2007). Marker-assisted selection – Current status and future perspectives in crops, livestock, for-estry and fish. *Food and Agricultural Organization of the United Nations, Rome* 2007, Book.

Guo Jun and Baianu I.C. (2011). NIR Calibrations for Soybean Seeds and Soy Food Composi-tion Analysis: Total Carbohydrates, Oil, Proteins and Water Contents. *Nature Precedings,* https://doi.org/doi:10101/npre.2011.6611.1.

Habier, D., Fernando, R.L and Dekkers, J.C.M (2007). The impact of genetic relationship infor-mation on genome-assisted breeding values. *Genetics*, Volume 177, Issue 4, 1 December 2007, Pages 2389–2397, https://doi.org/10.1534/genetics.107.081190.

Harouna, D. V., Venkataramana, P. B., Matemu, A. O and Ndakidemi, A. P. (2020). Agro-morphological exploration of some unexplored wild vigna legumes for domestication. *Agronomy*, *10*(1). https://doi.org/10.3390/agronomy10010110.

Harry, T., Joel, O., Ademiluyi, F.T and Oduola, K (2016). Performance evaluation of local cassava starches with imported starch for drilling fluid. American Journal of Engineering

Research (AJER), Volume-5, Issue-10, pp-111-120, e-ISSN: 2320-0847 p-ISSN: 2320-0936.

Henry, G., Westby, A and Collinson, C. (1998). European Group on RTB-Global Cassava End- uses & Markets, Phase 1-(FAO-ESCB) Global cassava end-uses and markets: current situation and recommendations for further study. *Phase 2*, 1–58. http://www.hubrural.org/IMG/pdf/global_cassava_end_use_study.pdf.

Hershberger, J., Gaby, E., Mbanjo, N., Peteti, P., Ikpan, A., Ogunpaimo, K., Nafiu, K., Rabbi, I. Y and Gore, M. A. (2022). Low-cost, handheld near-infrared spectroscopy for root dry matter content prediction in cassava. *The Plant Phenome Journal*, *Wiley Online Library*, *5*(1). https://doi.org/10.1002/ppj2.20040.

Holland, J. B., Nyquist, W. E and Cervantes-Martínez, C. T. (2003). Estimating and inter-preting heritability for plant breeding: An update. *Plant Breeding Reviews*, 22, 9–112. https://doi.org/10.1002/9780470650202.ch2.

Hübner, S., & Kantar, M. B. (2021). Tapping Diversity From the Wild: From Sampling to Implementation. *Frontiers in Plant Science*, *12*(January), 1–14. https://doi.org/10.3389/fpls.2021.626565.

Ikeogu, U. N., Davrieux, F., Dufour, D., Ceballos, H., Egesi, C. N and Jannink, J.-L (2017).

Rapid analyses of dry matter content and carotenoids in fresh cassava roots using a portable visible and near infrared spectrometer (Vis/NIRS). *PLoS ONE*, *12*(12). https://doi.org/10.1371/journal.pone.0188918.

Ilukor, D., Mukasa, J. B and Mugisha, S (2022). Fuel from the Farm: An analysis of the profitability and factors driving farmers' decisions to produce bioethanol from cassava in Northern Uganda. *African Journal of Rural Development*, *7*(1), 89–110.

Iragaba, P., Adinsi, L., Delgado, L. F., Nanyonjo, A. R., Nuwamanya, E., Wembabazi, E., Kanaabi, M., Honfozo, L., Hotegni, F., Djibril-Moussa, I., Londoño, L. F., Bugaud, C., Dufour, D., Kawuki, R. S., Akissoé, N and Tran, T (2024). Definition of sensory and instrumental thresholds of acceptability for selection of cassava genotypes with improvedboiling properties. *Journal of the Science of Food and Agriculture*. https://doi.org/10.1002/JSFA.13363.

Iragaba, P., Hamba, S., Nuwamanya, E., Kanaabi, M., Nanyonjo, R.A., Mpamire, D., Muhumuza, N., Khakasa, E., Tufan, H.A and Kawuki, R.S (2021). Identification of cassava quality attributes preferred by Ugandan users along the food chain. *International Journal of Food Science and Technology*, Volume 56, Issue 3, March 2021, Pages 1184–1192, https://doi.org/10.1111/ijfs.14878.

Iragaba, P., Kawuki, R. S., Bauchet, G., Ramu, P., Tufan, H. A., Earle, E. D., Gore, M. A and Wolfe, M. (2020). Genomic characterization of Ugandan smallholder farmer-preferred cassava varieties. *Crop Science*, *60*(3), 1450–1461, https://doi.org/10.1002/csc2.20152.

Iragaba, P., Nuwamanya, E., Wembabazi, E., Baguma, Y., Dufour, D., Earle, E. D., Kerr, R. B., Tufan, H. A., Gore, M. A., & Kawuki, R. S. (2019). Estimates for heritability and consumer-validation of a penetrometer method for phenotyping softness of cooked cassava roots. *African Crop Science Journal*, *27*(2), 147. https://doi.org/10.4314/acsj.v27i2.3.

Isidro, J., Jannink, J.-L., Akdemir, D., Poland, J., Heslot, N and Sorrells, M. E. (2015). Training set optimization under population structure in genomic selection. *Springer*, *128*(1), 145–158. https://doi.org/10.1007/s00122-014-2418-4.

Jannink, J., Lorenz, A.J and Iwata, H (2010). Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics*, Volume 9, Issue 2, March 2010, Pages 166–177, https://doi.org/10.1093/bfgp/elq001.

Kanaabi, M., Namakula, F. B., Nuwamanya, E., Kayondo, I. S., Muhumuza, N., Wembabazi, E., Iragaba, P., Nandudu, L., Ritah Nanyonjo, A., Baguma, J., Esuma, W., Ozimati, A., Settumba, M., Alicai, T., Ibanda, A and Kawuki, R. S (2023). Rapid analysis of hydrogen cyanide in fresh cassava roots using NIRSand machine learning algorithms: Meeting end user demand for low cyanogenic cassava. *The Plant Genome, 2023•Wiley Online Library*. https://doi.org/10.1002/tpg2.20403.

Kassambara, A (2016). Extract and Visualize the Results of Multivariate Data Analyses (facto-extra).” http://www.sthda.com/english/rpkgs/factoextra.

Katono, K., Alicai, T., Baguma, Y., Edema, R., Bua, A and Omongo, C.A. (2015). Influence of host plant resistance and disease pressure on spread of cassava brown streak disease in Uganda. *American Journal of Experimental Agriculture* 7(5): 284-293, 2015, Article no. AJEA.2015.130, DOI: 10.9734/AJEA/2015/15563.

Katono, K., Macfadyen, S., Omongo, C. A., Colvin, J., Karungi, J., & Otim, M. H. (2023). Effect of *Bemisia tabaci* SSA1 host density and cassava genotype on host feeding capacity and parasitism by two Hymenoptera parasitoid species. *Biocontrol Science and Technology*, *33*(1), 19–34. https://doi.org/10.1080/09583157.2022.2151976.

Kawuki, R., Kaweesi, T., Esuma, W., Pariyo, A., Kayondo, I.S., Ozimati, A., Kyaligonza, V., Abaca, A., Orone, J., Tumuhimbise, R., Nuwamanya, E., Abidrabo, P., Amuge, T., Ogwok, E., Okao, G., Wagaba, H., Adiga, G., Alicai, T., Omongo C., Bua, A., Ferguson, M., kanju, E and Baguma, Y (2016). Eleven years of breeding efforts to combat cassava brown streak disease. *Breeding Science* 66: 560–571 (2016) doi:10.1270/jsbbs.16005.

Kawuki, R., Pariyo, A., Amuge, T., Nuwamanya, E., Ssemakula, G., Tumwesigye, S., Bua, A., Baguma, Y., Omongo, C., Alicai, T and Orone, J (2011). A breeding scheme for local adoption of cassava (Manihot esculenta Crantz). *Journal of Plant Breeding and Crop Science* Vol. 3(7). pp. 120-130, July 2011; http://www.academicjournals.org/jpbcs.

Kayondo, S. I., Del Carpio, D. P., Lozano, R., Ozimati, A., Wolfe, M., Baguma, Y., Gracen, V., Offei, S., Ferguson, M., Kawuki, R and Jannink, J. L. (2018). Genome-wide

association mapping and genomic prediction for CBSD resistance in Manihot esculenta. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-19696-1.

Kennard, R. W and Stone, L. A. (1969). Computer Aided Design of Experiments. *Technometrics*, *11*(1), 137–148. https://doi.org/10.1080/00401706.1969.10490666.

Kilimo Trust. (2012). Development of Inclusive Markets in Agriculture and Trade (DIMAT) Value. Value Chain Analysis (VCA) of the of the Cassava Sub-sector in Uganda (Report).

Kleih, U., Phillips, D., Jagwe, J and Kirya, M. (2012). Cassava Market and Value Chain Analysis Uganda Case Study Final Report. *Cassava Adding Value (CAVA).*

Kouadio, O., Nindjin, C., Bonfoh, B., N'dri, D and Amani, G.N. Water absorption as an evaluation method of cooking quality for yam (Dioscorea alata) and cassava (Manihot esculenta crantz). *Procedia Food Science, 2011,* doi:10.1016/j.profoo.2011.09.024., https://www.sciencedirect.com/science/article/pii/S2211601X11000253.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C and Hunt, T (2023). Classification and Regression Training Package (caret). https://orcid.org/0000-0003-2402-136X.

Lebot, V., Champagne, A., Malapa, R., & Shiley, D. (2009). NIR determination of major constituents in tropical root and tuber crop flours. *Journal of Agricultural and Food Chemistry*, *57*(22), 10539–10547. https://doi.org/10.1021/jf902675n.

Lebot, V., Malapa, R and Jung, M. (2013). Use of NIRS for the rapid prediction of total N, minerals, sugars and starch in tropical root and tuber crops. *New Zealand Journal of Crop and Horticultural Science*, *41*(3), 144–153. https://doi.org/10.1080/01140671.2013.798335.

Legg, J., Kumar, P., Makeshkumar, T., Tripathi, L., Ferguson, M., Kanju, E., Ntawuruhunga, P., Cuellar, W (2015). Cassava virus diseases: biology, epidemiology, and management. *Advances in Virus Research Volume* 91, 2015, Pages 85-142, https://doi.org/10.1016/bs.aivir.2014.10.001.

Legg, J., Somado, E. A., Barker, I., Beach, L., Ceballos, H., Cuellar, W., Elkhoury, W., Gerling, D., Helsen, J., Hershey, C., Jarvis, A., Kulakow, P., Kumar, L., Lorenzen, J., Lynam, J., McMahon, M., Maruthi, G., Miano, D., Mtunda, K., Ntawuruhunga, P., Okogbenin, E., Pezo, P., Terry, E., Thiele, G., Thresh, M., Wadsworth, J., Walsh, S., Winter, S., Tohme, J and Fauquet, C. (2014). A global alliance declaring war on cassava viruses in Africa. *Food Security*, *6*(2), 231–248. https://doi.org/10.1007/s12571-014-0340-x.

Li, L., Zhang, Q and Huang, D (2014). A review of imaging techniques for plant phenotyping. *Sensors* 2014, 14(11), 20078-20111; https://doi.org/10.3390/s141120078.

Li, Y., Wang, G., Guo, G., Li, Y., Via, B.K and Pei, Z (2022). Spectral pre-processing and multivariate calibration methods for the prediction of wood density in Chinese white poplar by visible and near infrared spectroscopy. *Forests* 2022, 13(1), 62 https://doi.org/10.3390/f13010062.

López, A., Arazuri, S., García, I., Mangado, J and Jarén, C (2013). A review of the application of near-infrared spectroscopy for the analysis of potatoes. *Journal of Agricultural and Food Chemistry* Vol 61, Issue 23, *ACS Publications*.

Lorenz, A., & Nice, L. (2017). Training population design and resource allocation for genomic selection in plant breeding. *Genomic Selection for Crop Improvement: New Molecular Breeding Strategies for Crop Improvement*, 7–22. https://doi.org/10.1007/978-3-319-63170-7_2.

Lyimo, L. D., Tairo, F. M and Rweyemamu, C. L. (2012). Morphological and molecular characterization of cassava (Manihot esculenta Crantz) landraces using SSR markers. *International Journal of Integrative Biology*, *13*(2), 71–77.

Maechler, M., Rousseeuw, P., Struyf, A and Hubert, M (2013). Finding Groups in Data": Cluster Analysis Extended. Package "cluster." https://svn.r-project.org/R-packages/trunk/cluster/.

Maieves, H.A., Oliveira, D.C.De, Bernardo, C., Müller, C.M.D.O and Amante, E.R (2012). Microscopy and texture of raw and cooked cassava (manihot esculenta crantz) roots. *Journal of Texture Studies*, *43*(2), 164–173. https://doi.org/10.1111/j.1745-4603.2011.00327.x.

Manze, F., Rubaihayo, P., Ozimati, A., Gibson, P., Esuma, W., Bua, A., Alicai, T., Omongo, C., and Kawuki, R. S (2021). Genetic Gains for Yield and Virus Disease Resistance of Cassava Varieties Developed Over the Last Eight Decades in Uganda. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/FPLS.2021.651992/FULL.

Mbanjo, E. G. N., Rabbi, I. Y., Ferguson, M. E., Kayondo, S. I., Eng, N. H., Tripathi, L., Kulakow, P., & Egesi, C. (2021). Technological Innovations for Improving Cassava Production in Sub-Saharan Africa. In *Frontiers in Genetics* (Vol. 11, p. 1829). Frontiers Media S.A. https://doi.org/10.3389/fgene.2020.623736.

Meghar, K., Tran, T., Delgado, L. F., Ospina, M. A., Moreno, J. L., Luna, J., Londoño, L., Dufour, D and Davrieux, F (2023). Hyperspectral imaging for the determination of relevant cooking quality traits of boiled cassava. *Journal of the Science of Food and Agriculture*. https://doi.org/10.1002/JSFA.12654.

Meuwissen, T., Hayes, B., & Goddard, M. (2013). Accelerating Improvement of Livestock with Genomic Selection. *Annual Review of Animal Biosciences*, *1*(1), 221–237. https://doi.org/10.1146/annurev-animal-031412-103705.

Minervini, M., Scharr, H and Tsaftaris, S.A (2015). Image analysis: the new bottleneck in plant phenotyping. *IEEE Signal Processing Magazine* (Volume: 32, Issue: 4, July 2015), DOI: 10.1109/MSP.2015.2405111.

Mokari, A., Guo, S and Bocklitz, T (2023). Exploring the steps of infrared (IR) spectral analysis: Pre-processing, (classical) data modelling, and deep learning. *Molecules* 2023, 28(19), 6886; https://doi.org/10.3390/molecules28196886.

Munawar, A.A., Yunus, Y., Devianti and Satriyo, P (2020). Calibration models database of near infrared spectroscopy to predict agricultural soil fertility properties. *Data in Brief*, *30*. https://doi.org/10.1016/J.DIB.2020.105469.

Munck, L and Møller, B. (2005). Principal component analysis of near infrared spectra as a tool of endosperm mutant characterization and in barley breeding for quality. *Czech Journal of Genetics and Plant Breeding,* Vol. 41, Iss. 3, (2005): 89-95. DOI:10.17221/3666-CJGPB.

Naing, T. W., Nwet Win, N and Ngwe, D. H. (2018). Application of starch and its derivative from cassava (Manihot esculenta Crantz.) as disintegrants in pharmaceutical tablets. *Journal of the Myanmar Academy of Arts and Science*, 2018 Vol. XVI. No. 1A.

Nakabonge, G., Samukoya, C and Baguma, Y. (2018). Local varieties of cassava: conservation, cultivation and use in Uganda. *Environment, Development and Sustainability*, *20*(6), 2427–2445. https://doi.org/10.1007/s10668-017-9997-6.

Nandudu, L., Sheat, S., Winter, S., Ogbonna, A., Kawuki, R and Jannink J-L (2024) Genetic complexity of cassava brown streak disease: insights from qPCR-based viral titer analysis and genome-wide association studies. *Frontiers in Plant Science*. 15:1365132. doi: 10.3389/fpls.2024.1365132.

Nanyonjo, A.R., Angudubo, S., Candiru, W., Hamba, S., Esuma, W., Kawuki, R.S., Kleih, U., Fliedel, G and Forsythe, L (2019). State Of Knowledge for Boiled Cassava in Uganda. *Food Science, Gender & Market*. Kampala, Uganda. RTBfoods State of Knowledge Report, 22 p. https://doi.org/10.18167/agritrop/00695.

Nassar, N., & Ortiz, R. (2010). Breeding cassava to feed the poor. *Scientific American*, *302*(5), 78–84. https://doi.org/10.1038/scientificamerican0510-78.

Nduwumuremyi, A., Melis, R., Shanahan, P., & Theodore, A. (2018). Analysis of phenotypic variability for yield and quality traits within a collection of cassava (Manihot esculenta) genotypes. *Journals.Co.ZaA Theodore, A Nduwumuremyi, P Shanahan, R MelisSouth African Journal of Plant and Soil,* 2018, 35(3), 199–206, https://doi.org/10.1080/02571862.2017.1354406.

Ngea, G., Guillon, F., Ngang, J.J.E., Bonnin, E., Bouchet, B and Saulnier, L (2016). Modification of cell wall polysaccharides during retting of cassava roots. *Food Chemistry* Volume 213, 15 December 2016, Pages 402-409, https://doi.org/10.1016/j.foodchem.2016.06.107.

Njoku, D. N., Ikeogu, U. N., Ewa, F and Egesi, C. (2015). Crossability and germinability potentials of some cassava (Manihot esculenta Crantz) progenitors for selection. *Journal of Plant Breeding and Crop Science*, *7*(3), 61–66. https://doi.org/10.5897/jpbcs2014.0479.

Nkouaya Mbanjo, E.G., Hershberger, J., Peteti, P., Agbona, A., Ikpan, A., Ogunpaimo, K., Kayondo, S.I., Abioye, R.S., Nafiu, K., Alamu, E.O., Adesokan, M., Maziya-Dixon, B., Parkes, E., Kulakow, P., Gore, M.A., Egesi, C and Rabbi, I.Y. (2022). Predicting starch content in cassava fresh roots using near-infrared spectroscopy. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/FPLS.2022.990250/FULL.

Nsubuga, F. W., Olwoch, J. M and Rautenbach, C. J. de W. (2011). Climatic Trends at Namulonge in Uganda: 1947-2009. *Journal of Geography and Geology*, *3*(1), 119–131, https://doi.org/10.5539/jgg.v3n1p119.

Nuwamanya, E., Baguma, Y., Wembabazi, E and Rubaihayo, P. (2013). A comparative study of the physicochemical properties of starches from root, tuber and cereal crops. *African Journal of Biotechnology*, 10(56), 12018–12030. https://doi.org/10.4314/ajb.v10i56.

Nuwamanya, E., Baguma, Y., Emmambux, N., Taylor, J and Patrick, R. (2010). Physicochemical and functional characteristics of cassava starch in Ugandan varieties and their progenies. *Journal of Plant Breeding and Crop Science*, *2*(1), 1–11, http://www.academicjournals.org/jpbcs.

Okareh, O., Ajayeoba, T., Ugbekile, O and Hammed, T.B (2021). The use of cassava leaves as food and medicinal herbs in rural communities and the perceived health risks. Bionature 41(1), 2021: 19 – 38.

Okello, G and Akullu, A (2017). Feasibility of cassava industrialization in Uganda: A strategic study. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Feasibility+and+Market+for+Cassava+Industrialisation+in+Uganda%3A+A+Strategic+Study&btnG=.

Oliveira, P.H.G.A., Barbosa, A.C.O., Diniz, R.P., Oliveira, E.J.De and Ferreira, C.F (2018). Molecular marker assisted selection for increasing inbreeding in S1 populations of cassava. *Annals of the Brazilian Academy of Sciences*, (2018) 90(4): 3853-3869, http://dx.doi.org/10.1590/0001-3765201820180278.

Onyutha, C., Asiimwe, A., Muhwezi, L and Mubialiwo, A (2021). Water availability trends across water management zones in Uganda. *Atmospheric Science Letters*, *22*(10), 1–14.https://doi.org/10.1002/asl.1059.

Ozaki, Y., Christy, A and McClure, W. (2006). Near-infrared spectroscopy in food science and technology. Wiley-Interscience publication.

Ozimati, A. A., Esuma, W., Manze, F., Iragaba, P., Kanaabi, M., Ano, C. U., Egesi, C and Kawuki, R. S. (2022). Utility of Ugandan genomic selection cassava breeding populations

for prediction of cassava viral disease resistance and yield in West African clones. *Frontiers in Plant Science*, *13*, 4541. https://doi.org/10.3389/FPLS.2022.1018156/BIBTEX.

Ozimati, A., Kawuki, R., Esuma, W., Kayondo, I. S., Wolfe, M., Lozano, R., Rabbi, I., Kulakow, P., & Jannink, J. L. (2018). Training population optimization for prediction of cassava brown streak disease resistance in West African Clones. *G3: Genes, Genomes, Genetics*, *8*(12), 3903–3913. https://doi.org/10.1534/g3.118.200710.

Ozimati, A., Kawuki, R., Esuma, W., Kayondo, S. I., Pariyo, A., Wolfe, M., & Jannink, J.-L. (2019). Genetic Variation and Trait Correlations in an East African Cassava Breeding Population for Genomic Selection. *Crop Science*, *59*(2), 460. https://doi.org/10.2135/cropsci2018.01.0060.

Parmar, A., Sturm, B and Hensel, O. (2017). Crops that feed the world: Production and improvement of cassava for food, feed, and industrial uses. *Food Security*, *9*(5), 907–927. https://doi.org/10.1007/S12571-017-0717-8.

Peprah, B.B., Parkes, E., Manu-Aduening, J., Kulakow, P., van Biljon, A and Labuschagne, M. (2020). Genetic variability, stability and heritability for quality and yield characteristics in provitamin A cassava varieties. *Euphytica*, *216*(2). https://doi.org/10.1007/s10681-020-2562-7.

Pérez, J. C., Lenis, J. I., Calle, F., Morante, N., Sánchez, T., Debouck, D and Ceballos, H. (2011).

Genetic variability of root peel thickness and its influence in extractable starch from cassava (Manihot esculenta Crantz) roots. *Plant Breeding*, *130*(6), 688–693. https://doi.org/10.1111/j.1439-0523.2011.01873.x.

Peter Waigumba, S., Nyamutoka, P., Wanda, K., Adebayo, A., Kwagala, I., Menya, G., Acheng, S., Nuwamanya, E., Matovu, M., Kaliisa, R., Muyinza, H., Nyakaisiki, E and Naziri, D. (2016). Market Opportunities and Value Chain Analysis of Fresh Cassava Roots in Uganda Extending the shelf-life of fresh cassava roots for increased incomes and postharvest losses reduction Expanding Utilization of Roots, Tubers and Bananas and Reducing Their Pos. *CGIAR – Technical report*.

Piepho, H.-P., Möhring, J., Melchinger, A. E and Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica*, 161(1–2), 209–228. https://doi.org/10.1007/s10681-007-9449-8.

Phumichai, C., Kongsil, P., Kittipadakul, P., Wannarat, W., Aiemnaka, P., Nathaisong, P., Hunsawattanakul, S., Fungfoo, P., Rojanaridpiched, C., Vichukit, V., Chunwongse, J., Tongyoo, P., Kijkhunasatian, C., Chotineeranat, S., Piyachomkwan, K., Wolfe, M. D., Jannink, J.-L and Sorrells, M. E. (2022). Genome-wide association mapping and genomic prediction of yield-related traits and starch pasting properties in cassava. *Springer*, *135*(1), 145–171, https://doi.org/10.1007/s00122-021-03956-2.

Huang, J., Romero-Torres, S and Moshgbar, M (2010). Practical Considerations in Data Pre-treatment for NIR and Raman Spectroscopy. *American Pharmaceutical Review*, 2010 (Article).

Pradeepkumar, T., Dijee, B., Joy, M and Radhakrishnan, N (2001). Genetic variation in tomato for yield and resistance to bacterial wilt. *Journal of Tropical Agriculture* 39 (2001): 157-158, http://14.139.185.57:8080/jspui/bitstream/123456789/2218/1/39_2_157-158_0971-636X.pdf.

Pradyawong, S., Juneja, A., Bilal Sadiq, M., Noomhorm, A and Singh, V. (2018). Comparison of cassava starch with corn as a feedstock for bioethanol production. *Energies*, *11*(12), 1–11. https://doi.org/10.3390/en11123476.

Prieto, N., Pawluczyk, O., Dugan, M. E. R and Aalhus, J. L. (2017). A review of the principles and applications of near-infrared spectroscopy to characterize meat, fat, and meat products. *Applied Spectroscopy* Volume 71, Issue 7, July 2017, Pages 1403-1426, *71*(7), 1403–1426. https://doi.org/10.1177/0003702817709299.

Rabbi, I., Udoh, L., Wolfe, M., Parkes, E.Y., Gedil, M.A., Dixon, A., Ramu, P., Jannink, J-L and Kulakow, P (2017). Genome-wide association mapping of correlated traits in cassava: dry matter and total carotenoid content. *The Plant Genome* Volume 10, no. 3, doi: 10.3835/plantgenome2016.09.0094.

Rabbi, I. Y., Kayondo, S. I., Bauchet, G., Yusuf, M., Aghogho, C. I., Ogunpaimo, K., Uwugiaren, R., Smith, I. A., Peteti, P., Agbona, A., Parkes, E., Lydia, E., Wolfe, M., Jannink, J. L., Egesi, C., & Kulakow, P. (2022). Genome-wide association analysis reveals new insights into the genetic architecture of defensive, agro-morphological and quality- related traits in cassava. *Plant Molecular Biology*, *109*(3), 195–213. https://doi.org/10.1007/S11103-020-01038-3.

Rinnan, Å., Berg, F.V.D and Engelsen, S.B (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, Vol. 28, No. 10, 2009, doi:10.1016/j.trac.2009.07.007.

Robertsen, C., Hjortshøj, R.L and Janss, L.L (2019). Genomic selection in cereal breeding. *Agronomy* 2019, 9, 95; doi:10.3390/agronomy9020095.

Andrade, L.R.B.d., Sousa, M.B.e., Oliveira, E.J., Resende, M.D.V.d and Azevedo, C.F (2019) Cassava yield traits predicted by genomic selection methods. *PLoS ONE* 14(11): e0224920. https://doi.org/ 10.1371/journal.pone.0224920.

Sahebalam, H., Gholizadeh, M., Hafezian, H and Farhadi, A (2019). Comparison of parametric, semiparametric and nonparametric methods in genomic evaluation. *Journal of Genetics,* 2019, *98*(4), https://doi.org/10.1007/s12041-019-1149-3.

Sajeev, M. S., Sreekumar, J., Unnikrishnan, M., Moorthy, S. N and Shanavas, S (2010). Kinetics of thermal softening of cassava tubers and rheological modeling of the starch. *Journal of Food Science and Technology*, *47*(5), 507–518. https://doi.org/10.1007/s13197-010-0087-0.

Sampaio, P. S., & Brites, C. M. (2021). Near-Infrared Spectroscopy and Machine Learning: Analysis and Classification Methods of Rice. *Integrative Advances in Rice Research*. https://doi.org/10.5772/INTECHOPEN.99017.

Sánchez, T., Ceballos, H., Dufour, D., Ortiz, D., Morante, N., Calle, F., Zum Felde, T., Domínguez, 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. https://doi.org/10.1016/j.foodchem.2013.11.081.

Sarin, J., Torniainen, J., Prakash, M., Rieppo, L., Afara, I.O and Töyräs, J (2019). Dataset on equine cartilage near infrared spectra, composition, and functional properties. *Nature, Scientific Data* (2019) 6:164, https://doi.org/10.1038/s41597-019-0170-y.

VanRaden, P.M (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science* Volume 91, Issue 11, November 2008, Pages 4414-4423. https://doi.org/10.3168/jds.2007-0980.

Sexton, J., Everingham, Y., Donald, D., Staunton, S and White, R (2018). A comparison of non-linear regression methods for improved on-line near infrared spectroscopic analysis of a

sugarcane quality measure. *Journal of Near Infrared Spectroscopy*, Volume 26, Issue 5 https://doi.org/10.1177/0967033518802448.

Shao, Y., He, Y., Bao, Y and Mao, J. (2009). Near-Infrared Spectroscopy for Classification of Oranges and Prediction of the Sugar Content. *International Journal of Food Properties*, 12(3), 644–658. https://doi.org/10.1080/10942910801992991.

Shi, R., Junker, A., Seiler, C and Altmann, T (2018). Phenotyping roots in darkness: disturbance-free root imaging with near infrared illumination. *Functional Plant Biology*, 2018, Vol. 45, No. 4, 400-411.

Shilpashree, N., Devi, S.N., Manjunathagowda, D.C., Muddappa, A., Abdelmohsen, S.A.M., Tamam, N., Elansary, H.O., Zin El-Abedin, TK., Abdelbacki, A. M.M and Janhavi, V. (2021). Morphological Characterization, Variability and Diversity among Vegetable Soy-bean (Glycine max L.) Genotypes. *Plants, 2021*, https://doi.org/10.3390/plants10040671.

Sinha, A. K and Mishra, P.K. (2013). Agro-morphological characterization and morphology based genetic diversity analysis of landraces of rice variety (Oryza sativa L.) of Bankura district of West Bengal. *International Journal of Current Research*, *5*(10), 2764–2769.

Stevens, A., Ramirez-Lopez, L and Stevens, M. (2015). *Package "prospectr."* http://bioconductor.statistik.tu-dortmund.de/cran/web/packages/prospectr/prospectr.pdf.

Stich, B and Van Inghelandt, D. (2018). Prospects and potential uses of genomic prediction of key performance traits in tetraploid potato. *Frontiers in Plant Science*, *9*(March), 1–12. https://doi.org/10.3389/fpls.2018.00159.

Su, W-H and Sun, D-W (2019). Advanced Analysis of Roots and Tubers by Hyperspectral Techniques. *Advances in Food and Nutrition Research*, Volume 87, 2019, Pages 255-303, https://doi.org/10.1016/bs.afnr.2018.07.003.

Sun, R., Zhou, Jy. & Yu, D. Nondestructive prediction model of internal hardness attribute of fig fruit using NIR spectroscopy and RF. Multimedia Tools and Applications 80, 21579–21594 (2021), https://doi.org/10.1007/s11042-021-10777-4.

Sverrisdóttir, E., Byrne, S., Sundmark, E. H. R., Johnsen, H. Ø., Kirk, H. G., Asp, T., Janss, L., and Nielsen, K. L. (2017). Genomic prediction of starch content and chipping quality in tetraploid potato using genotyping-by-sequencing. *Theoretical and Applied Genetics*, *130*(10), 2091–2108. https://doi.org/10.1007/s00122-017-2944-y.

Thiele, G., Dufour, D., Vernier, P., Mwanga, R. O. M., Parker, M. L., Schulte Geldermann, E., Teeken, B., Wossen, T., Gotor, E., Kikulwe, E., Tufan, H., Sinelle, S., Kouakou, A. M., Friedmann, M., Polar, V and Hershey, C. (2020). A review of varietal change in roots, tubers and bananas: consumer preferences and other drivers of adoption and implications for breeding. *International Journal of Food Science and Technology*, 1–17. https://doi.org/10.1111/ijfs.14684.

Tiwari, D., Tripathi, S., Tripathi, M.P., Khatri, N and Bastola, B.R (2019). Genetic variability and correlation coefficients of major traits in early maturing rice under rainfed lowland environments of Nepal. *Advances in Agriculture*, Volume 2019, Article ID 5975901, 9 pageshttps://doi.org/10.1155/2019/5975901.

Toscano, G., Rinnan, Å., Pizzi, A and Mancini, M. (2017). The Use of Near-Infrared (NIR) Spectroscopy and Principal Component Analysis (PCA) to Discriminate Bark and Wood of the Most Common Species of the Pellet Sector. *Energy and Fuels*, *31*(3), 2814–2821. https://doi.org/10.1021/ACS.ENERGYFUELS.6B02421.

Tran, T and Escobar, A. (2019). SOP for characterization of cooking time and texture of boiled cassava: texture-extrusion. Biophysical characterization of quality traits, WP2. https://doi.org/10.18167/agritrop/00594.

Tran, T., Zhang, X., Ceballos, H., Moreno, J. L., Luna, J., Es Escobar, A., Morante, N., Belal-cazar, J., Becerra, L. A and Dufour, D (2020). Correlation of cooking time with water absorption and changes in relative density during boiling of cassava roots. *International Journal of Food Science and Technology*, Volume 56, Issue 3, March 2021, Pages 1193–1205, https://doi.org/10.1111/ijfs.14769.

Tsairidou, S., Hamilton, A., Robledo, D., Bron, J.E and Houston, R.D. (2020). Optimizing low-cost genotyping and imputation strategies for genomic selection in atlantic salmon. *G3: Genes, Genomes, Genetics*, *10*(2), 581–590, https://doi.org/10.1534/g3.119.400800.

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*(4), 403–408. https://doi.org/10.5829/idosi.wjas.2012.8.4.1667.

Uchendu, K., Njoku, D. N., Ikeogu, U. N., Dzidzienyo, D., Tongoona, P., Offei, S and Egesi, C. (2022). Genotype-by-environment interaction and stability of root mealiness and other organoleptic properties of boiled cassava roots. *Scientific Reports*, *12*(1), https://doi.org/10.1038/s41598-022-25172-8.

Uchendu, K., Njoku, D., Paterne, A., Rabbi, I.Y., Dzidzienyo, D., Tongoona, P and Egesi, C (2021). Genome- wide association study of root mealiness and other texture-associated traits in cassava. *Frontiers in Plant Science*, 12:770434, doi: 10.3389/fpls.2021.770434.

Famine Early Warning Systems Network (FEWS NET, 2017). Uganda staple food market fundamentals, Uganda. A report, http://www.fews.net.

Urvoy, M., & Autrusseau, F. (2014). Application of grubbs' test for outliers to the detection of watermarks. *IH and MMSec 2014 - Proceedings of the 2014 ACM Information Hiding and Multimedia Security Workshop*, 49–60. https://doi.org/10.1145/2600918.2600931.

van Dijk, C., Fischer, M., Holm, J., Beekhuizen, J-G., Stolle-Smits, T and Boeriu, C. (2002). Texture of Cooked Potatoes (*Solanum tuberosum*). Relationships between Dry Matter

Content, Sensory-Perceived Texture, and Near-Infrared Spectroscopy. *Journal of Agricultural and Food Chemistry*, *50*(18), 5082–5088. https://doi.org/10.1021/jf011509w.

Vivek, B., Krishna, G.K., Vengadessa, V., Babu, R., Zaidi, P.H., Kha, L.Q., Mandal, S.S., Grudloyma, P., Takalkar, S., Krothapalli, K., Singh, I.S., Ocampo, E.T.M., Xingming, F., Burgueño, J., Azrai, M., Singh, R.P and Crossa, J (2016). Use of genomic estimated breeding values results in rapid genetic gains for drought tolerance in maize. The plant genome, March 2017, vol. 10, no .1; doi: 10.3835/plantgenome2016.07.0070.

Wagaba, H., Beyene, G., Aleu, J., Odipio, J., Okao-Okuja, G., Chauhan, R. D., Munga, T., Obiero, H., Halsey, M. E., Ilyas, M., Raymond, P., Bua, A., Taylor, N. J., Miano, D and Alicai, T. (2017). Field level RNAi-mediated resistance to cassava brown streak disease across multiple cropping cycles and diverse east African agro-ecological locations. *Frontiers in Plant Science*, *7*(January). https://doi.org/10.3389/fpls.2016.02060.

Wajizah, S.S and Munawar, A.A (2020). Near infrared spectroscopy (NIRS) data analysis for a rapid and simultaneous prediction of feed nutritive parameters. *Data in brief 29* (2020) 105211.

Wang, L., Janss, L. L., Henshall, J., Huang, C.-H., Marois, D., Alemu, S., Sørensen, A. C and Jensen, J. (2020). Effect of genomic selection and genotyping strategy on estimation of variance components in animal models using different relationship matrices. *Genetics Selection Evolution,* 2020, 52(1). https://doi.org/10.1186/s12711-020-00550-w.

Wang, X., Xu, Y., Hu, Z., & Xu, C. (2018). Genomic selection methods for crop improvement: Current status and prospects. *The Crop Journal*, 6(4), 330-340, https://doi.org/10.1016/j.cj.2018.03.001.

Watanabe, L.S., Bovolenta, Y.R., Junior, V.R.A., Barbin, D.F., Madeira, T.B and Nixdorf, S.L (2018). Investigation of NIR spectra pre-processing methods combined with multivariate regression for determination of moisture in powdered industrial egg. *Acta Scientiarum*, 2018, vol. 40, January-December 2019, ISSN: 1806-2563 1807-8664, DOI: https://doi.org/10.4025/actascitechnol.v40i1.30133.

Wembabazi, E., Dzidzienyo, D., Ofori, K., Ibanda, A., Ozimati, A.A., Esuma, W., Nuwamanya,E., Iragaba, P., Kanabi, M., Danquah, E.Y and Kawuki, R.S (2022). Consumer preferred boiled cassava cooking qualities in white and yellow fleshed advanced breeding populations in Uganda. *African Journal of Rural Development,* 7(3), 315-324 *2022, repository.Ruforum.Org*.

Wolfe, M. D., Del Carpio, D. P., Alabi, O., Ezenwaka, L. C., Ikeogu, U. N., Kayondo, I. S., ... and Jannink, J. L. (2017). Prospects for genomic selection in cassava breeding. *The Plant Genome*, 10(3), plantgenome2017-03, https://doi.org/10.3835/plantgenome2017.03.0015.

Wolfe, M. D., Rabbi, I. Y., Egesi, C., Hamblin, M., Kawuki, R., Kulakow, P., ... and Jannink, J. L. (2016). Genome-wide association and prediction reveals genetic architecture of cassava

mosaic disease resistance and prospects for rapid genetic improvement. *The Plant Genome*, 9(2), plantgenome2015-11, https://doi.org/10.3835/plantgenome2015.11.0118.

Xia, Y., Huang, W., Fan, S., Li, J., & Chen, L. (2019). Effect of spectral measurement orientation on online prediction of soluble solids content of apple using Vis/NIR diffuse reflectance. *Infrared Physics & Technology*, 97, 467-477, https://doi.org/10.1016/j.infrared.2019.01.012.

Xu, Y and Crouch, J.H. Marker-Assisted Selection in Plant Breeding: From Publications to Practice. Crop Science, vol. 48, doi: 10.2135/cropsci2007.04.0191.

Yonis, B. O., Pino del Carpio, D., Wolfe, M., Jannink, J. L., Kulakow, P and Rabbi, I. (2020). Improving root characterization for genomic prediction in cassava. *Scientific Reports 2020 10:1*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-64963-9.

Zhang, A., Wang, H., Beyene, Y., Semagn, K., Liu, Y., Cao, S., Cui, Z., Ruan, Y., Burgueño, J., Vicente, F. S., Olsen, M., Prasanna, B. M., Crossa, J., Yu, H and Zhang, X (2017). Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. *Frontiers in Plant Science*, *8*, 1916. https://doi.org/10.3389/FPLS.2017.01916/BIBTEX.

Zhang, H., Yin, L., Wang, M., Yuan, X and Liu, X. (2019). Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig

populations. *Frontiers in Genetics*, *10*(MAR). https://doi.org/10.3389/FGENE.2019.00189/FULL.

Zhang, S., Chen, X., Lu, C., Ye, J., Zou, M., Lu, K., Feng, S., Pei, J., Liu, C., Zhou, X., Ma, P., Li, Z., Liu, C., Liao, Q., Xia, Z., & Wang, W. (2018). Genome-Wide Association Studies of 11 Agronomic Traits in Cassava (Manihot esculenta Crantz). *Frontiers in Plant Science*, *9*. https://doi.org/10.3389/fpls.2018.00503.