# QTL mapping for pest and disease host plant resistance in cassava cultivars Kiroba and AR37-80 and coincidence of QTL with introgression regions from *Manihot glaziovii*

by

Inosters Wambua Nzuki

Submitted partial fulfillment of the requirement for the degree

Philosophiae Doctor

In the Faculty of Natural and Agricultural Sciences

Department of Genetics

University of Pretoria

Pretoria

January 2019

Under the supervision of Prof A A Myburg and

Co-supervision of Dr. M Ferguson and Prof DK Berger

# Declaration

I, **Inosters Wambua Nzuki** declare that, I understand the University of Pretoria policy concerning plagiarism and I hereby state that this thesis is my own original work and where other people's work has been used (either from a printed source, Internet or any other source), this has been properly acknowledged and referenced in accordance with departmental requirements. I further declare that this work has not been previously submitted by me to the University of Pretoria or any other institution of examination.

Inosters Wambua Nzuki

**January 2019**

# Thesis Summary

# QTL mapping for pest and disease resistance in cassava cultivars Kiroba and AR37-80 and coincidence of QTL with introgression regions from *Manihot glaziovii*

**Inosters Wambua Nzuki**

**Supervised by Prof AA Myburg**

Co-supervision of Dr M Ferguson and Prof DK Berger

Submitted partial fulfillment of the requirement for the degree

Philosophiae Doctor

Department of Genetics

University of Pretoria

## Summary

Cassava (*Manihot esulenta* Crantz.) is a staple food crop for more than 800 million people worldwide. It is drought tolerant and offers a flexible harvesting regime since the roots can remain in the soil and be harvested when needed. It is a food security crop when cereal crops fail. Biotic and abiotic stresses including pest and diseases negate this potential. Its heterozygous nature, long growing cycle and low seed yield per pollination poses challenges in breeding. In addition, it is highly outcrossing making it difficult to develop an adequately sized $F_2$ population hence limiting genetic studies to $F_1$ progenies.

Cassava brown streak disease (CBSD) has emerged to be the great threat to cassava production reducing useable roots or leading to total crop loss, and if not checked it can impact more than 200 million people in Africa who depend on the crop for their food and income generation.

The plausible approach to combat CBSD is to combine breeding for host plant resistance with sanitation measures and the planting of virus-free stakes. The breeding for host plant resistance should be performed as quickly and efficiently as possible, taking advantage of genomic, transformation and molecular marker technology. Conventional breeding, which does not use data generated from molecular tools, takes up to 10 years to deliver a new cultivar since the plants have to be grown for 12 months before selection can be made. If molecular markers were found to be associated with field resistance, then $F_1$ progeny generated in a breeding program could be screened and selected at the seedling stage, thereby drastically reducing the breeding cycle and providing an accurate way of efficiently pyramiding resistance from different sources.

In this project, quantitative trait loci (QTL) mapping for resistance to CBSD, cassava Mosaic disease (CMD), and cassava green mites (CGM) was performed using an $F_1$ mapping population developed between CBSD resistant Tanzanian landrace, Kiroba, and a susceptible breeding clone, AR37-80. This aimed to construct a SNP based linkage map using the segregating population and provide a tool to identify QTL. The study investigated the presence of genomic regions in Kiroba derived from *M. glaziovii* and their contribution to the field resistance observed in Kiroba. The introgression regions in Kiroba were compared to those from Namikonga, the well-known CBSD resistant cultivar and other genotypes of African origin to understand the source of their resistance.

The results show that only two QTL are linked to CBSD root necrosis and are located on chromosomes V and XII, while seven are associated with CBSD foliar symptoms only and are located on chromosomes IV, VI, XVII, and XVIII. The QTL on chromosomes XI and XV are linked with both CBSD foliar and root necrosis symptoms. Two QTL found on chromosome XII and XIV are linked to CMD, while two QTL located on chromosomes V and X are linked to CGM resistance.

The analysis of introgression regions in Kiroba revealed the existence of large *Manihot glaziovii* like regions on chromosome I, XVII, and XVIII. The introgression segments on chromosomes XVII and XVIII overlap with QTL associated with CBSD foliar symptoms. These regions contain domains associated with host plant disease resistance. The introgression region on chromosome I in Kiroba is of a different haplotype to the characteristic “Amani

haplotype" found in the landrace Namikonga and other genotypes analyzed in this study. Kiroba also does not have a large introgression block on chromosome IV found in other genotypes. Kiroba is closely related to a sampled Tanzanian "tree cassava." This supports the observation that some of the QTL associated with CBSD resistance in Kiroba are different to those observed in Namikonga.

This study provides an understanding of the genetic basis of the field resistance observed in the local cassava landraces, the genomic regions contributing to the resistance and the source of the resistance. This information is valuable in pyramiding QTL for host plant disease resistance.

# Preface

In recent times, the world has been experiencing climate change which has led to epidemiological changes in disease distribution and spread. Cassava brown-streak disease, a viral disease of cassava, has been previously reported to be restricted by altitude, but recent research findings show a changing spread to regions that were thought to be free. In this regard, pre-emptive breeding strategies are urgently needed to safeguard the livelihoods of millions of people who depend on cassava for their daily uptake. The most efficient and sustainable way to control viral diseases in cassava is the development of resistant cultivars that possess natural molecular barriers that hinder the invasion and spread of the virus in the plants. Complex defence mechanisms have been discovered in plants that involve R-genes and AGONATE proteins. The main defence mechanism against potyviruses has been discovered to be translation initiation factors eIF4E and its isoforms.

Today, the field of genomics and bioinformatics coupled with reduced cost of sequencing and genotyping, provides an opportunity for breeders to utilize molecular tools to accelerate the breeding of cultivars with traits preferred by farmers in a cost effective way. Cassava has a long breeding cycle and to shorten it and reduce population sizes breeders have to embrace modern advances in biotechnology including Marker assisted selection (MAS). Development of molecular markers such as Single nucleotide polymorphisms (SNPs) is a key step in the discovery of quantitative trait loci (QTL) associated with major pests and diseases of cassava.

There is hope for cassava due to the recent discovery of some landraces in East Africa that are infected with CBSD and CMD viruses which show disease symptoms but the yield is not affected. Some of these landraces include Kiroba, Namikonga and Nachinyaya. Though, the exact source of this field resistance is not well understood, breeders postulate that it is derived from wild relatives of cassava due to historical breeding records from Amani breeding station. The Amani breeding station for cassava was based in Tanzania in the 1930s and utilized wild relatives of cassava to breed cultivars resistant against viral diseases such as cassava mosaic virus and cassava brown streak virus. The program ceased operation in 1958 and some of the material were transferred to other research centers in Africa and it is postulated that some found

their way into farmer's fields. Thus, breeders suggest some of the landraces that have been observed to show field resistance against viral diseases today may be part of the interspecific hybrids developed at Amani station but grown by farmers under different names and without any known pedigree. In this regard, research to unravel the source of resistance observed in these landraces is a top priority for cassava breeders. It was against this backdrop that this PhD was conceptualized; (1) to develop a genetic linkage map and identify QTL associated with pest and disease resistance in bi-parental cassava cross between Kiroba and AR37-80 (2) Sequence some of the local landraces noted to have field resistance alongside *M.glaziovii* and undertake a comparative genomic study. The comparative analysis was postulated to discover introgression blocks and whether they coincide with major QTL and see whether they contribute to the observed field resistance or not. Genome wide relatedness and the kinship relationship portion of this study was envisaged to reveal the possible sources of resistance in African cassava germplasm. It was expected that this PhD study would provide insight into the genetic basis of the field resistance in the local cassava landraces and avail baseline data for future breeding strategies in cassava.

**Chapter 1** of this thesis provides a comprehensive review of cassava, its production, constraints and progress made so far in cassava breeding for host plant resistance. The biotic and abiotic stresses limiting cassava production including diseases and pests such as CBSD, CMD and CGM are reviewed. The threat posed by CBSD and its eminent threat in zones considered free from the disease is discussed together with the challenges encountered during breeding in cassava. Approaches to combat CBSD such as those implemented by ETH and IITA are comprehensively covered. The sources of resistance in African germplasm, including the Amani breeding program, are reviewed. The local landraces exhibiting field resistance and suspected to be interspecific hybrids from the Amani breeding program are discussed. The chapter concludes by discussing the various technologies used in plants to deploy new cultivars such as MAS and QTL mapping.

**Chapter 2** is a research chapter with detailed methodologies used to develop a mapping population from a cross between Kiroba and AR37-80 and testing its integrity to identify true crosses. The Genotyping by sequencing (GBS) method used to generate SNP data from the mapping population is discussed. The development of high-density and framework linkage maps based on the SNP data are presented and discussed. The results show that the mapping

population contained some genotypes that were either off types/self's and were not suitable for further study and hence excluded from subsequent analysis. The final genetic linkage map is composed from106 $F_1$ progeny and 1,974 SNP markers and spans 18 chromosomes covering a distance of 1,698 cM.

**Chapter 3** is the second research chapter in this study and discusses how the phenotypic trials were established, statistical analysis of the phenotypic data and the detection of quantitative trait loci (QTL). The results of this chapter show that fifteen significant QTL were detected; two are associated with CBSD root necrosis only, and were detected on chromosomes V and XII, while seven were associated with CBSD foliar symptoms only and were detected on chromosomes IV, VI, XVII, and XVIII. QTL on chromosomes XI and XV were associated with both CBSD foliar and root necrosis symptoms. Two QTL were found to be associated with CMD and were detected on chromosomes XII and XIV, while two were associated with CGM and were identified on chromosomes V and X.

**Chapter 4** is the third research chapter that deals with the resequencing and comparative genomic analysis of the Kiroba genome and other cassava landraces. The methods used to identify introgression regions in Kiroba derived from *M. glaziovii* are discussed. The QTL regions found in the introgression regions are investigated together with the genes present and their relation with host plant disease resistance. The genome wide relatedness of Kiroba and other genotypes of African origin are investigated and presented. The results of this chapter reveal the presence of a major introgression region in Kiroba derived from *M. glaziovii* and co-location of some QTL with the introgression regions. Further, the chapter unravels the existence of a different source of resistance in the African germplasm, other than that found in Namikonga.

**Chapter 5** is found at the end of the thesis and consists of the concluding remarks that discuss the findings of the study and how they fit into the context of the research that has already been performed. The future prospects of the findings in terms of improvements, limitations, and impact are discussed.

The results of this study have been published in the journal Frontiers in Plant Sciences under the tittle 'QTL mapping for pest and disease resistance in cassava and coincidence of some QTL with introgression regions derived from *Manihot glaziovii*". The mapping population developed in this study is part of the data used to construct the consensus map for cassava in the paper "High-Resolution Linkage Map and Chromosome-Scale Genome Assembly for Cassava (*Manihot esculenta Crantz*) from 10 Populations" by the International Cassava Genetic Map Consortium (ICGMC). The results of this PhD study have also been presented in various international conferences listed below.

1. **Inosters Nzuki**, Zander Myburg, Ferguson M. and Appolinaire Djikeng (**2012**) Understanding the genetic basis of field resistance to Cassava Brown streak disease (CBSD). In: *Global cassava Partnership for the 21st century: Second Scientific Conference GCP21-II*, 18-22 June, 2012, Kampala, Uganda
2. **Inosters Nzuki**, Zander Myburg, Manpreet S Katari, Ferguson M., Geoffrey Mkamilo, Esther Masumba and Kapinga F. (**2016**) QTL mapping for cassava brown-streak disease and the Introgression regions in Kiroba derived from *Manihot glaziovii*. In: *World Congress on root and tuber crops: Third Scientific Conference of the Global Cassava Partnership for the 21st Century and 17th Symposium of the International Society for Tropical Root Crops*, Nanning Guangxi, China Jan 18-22, 2016.

3. Morag E. Ferguson G.S.M., Esther A. Masumba, Fortunus A. Kapinga, Kasele Salum, **Inosters Nzuki**, Bernadetha Kimatha, Heneriko P. Kulembeka, Caroline Sichalwe, Ana Luis Garcia Oliveira, Edward Kanju, Simon Jeremiah and Steve Rounsley (**2016**) QTL Mapping for Virus Resistance in Cassava. In: *Plant and Animal Genome conference XXIV*, 08 - 13 Jan 2016, San Diego, CA.

4. Morag Ferguson E.M., Fortunus Kapinga , **Inosters Nzuki** , Bernadetha Kimatha , Kasele Salum , Heneriko Kulembeka , Geoffrey Mkamilo , Jessica Lyons , Jessen Bredeson,Trushar Shah, Manpreet Katari, Edward Kanju, Alexander Myburg , Albe van der Merwe , Daniel Rokhsar , Steve Rounsley (**2017**) QTL Associated with Field Resistance to Cassava Brown Streak Disease in the Context of the Genetic Relationship Among Genotypes. In: *Plant and Animal Genome XX*, 13-18 Jan, 2017 San Diego, CA.

# Acknowlegements

I acknowledge the Almighty God for his mercy and grace through the Lord Jesus Christ by whose power and protection, I have been able to accomplish this work. I express my sincere gratitude to my advisor Prof. Zander Myburg for the continuous support of my PhD study and related research, for his patience, motivation, and immense knowledge. His guidance helped me throughout the research and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD study.Besides my advisor, I would like to thank the rest of my thesis committee: Dr Albe van der Merwe and Dr. Morag Ferguson, for their insightful comments and encouragement. I am grateful to Dr Morag Ferguson for her research support, obtaining of the funds for the project and editing the thesis.

My sincere thanks also go to Dr. Manpreet Katari of the University of New York who provided training and assistance in bioinformatics analysis, without his precious support it would not have been possible to do the analysis. I thank the technical staff from Agricultural Research Institute (ARI) Tanzania for their kind assistance in the establishment of the mapping populations and managing the field trials for the last four years. Am particularly grateful to Esther Masumba, Fortunus Kapinga and Geoffrey Mkamilo for their immense support for the field work. My special thanks goes to Gambo, Kasele Salum and Bernadetha Kimata and the rest of the technical staff from ARI Naliendele who were instrumental in managing the phenotyping trials

Last but not the least, I would like to thank my family: my wife, Catherine Wambua, my sons Enock Mumo and Joshua Mwendwa and daughter Faith Mutanu for supporting me spiritually throughout my study.

**TABLE OF CONTENTS**

# List of Figures

# List of Tables

# List of Supplementary Data

**Appendix 1** Supplementary Figures

1. Supplementary Figure S1: Scale of 1-5 used for scoring CBSD root necrosis and foliar symptoms
2. Supplementary Figure S2: LOD profile for stable QTL associated with CBSD root necrosis
3. Supplementary Figure S3: LOD profile for stable QTL associated with CBSD foliar symptoms
4. Supplementary Figure S4: LOD profile for QTL associated with CBSD foliar symptoms detected on chromosome 15, 11 and 18
5. Supplementary Figure S5: LOD profile for stable QTL associated with CGM detected on chromosome 5, and 10
6. Supplementary Figure S6: Major and minor introgression regions in Kiroba

**Appendix 2** Supplementary Table S1: Main QTL in version 5 and 6 of cassava genome assembly

**Appendix 3** Published paper "QTL Mapping for Pest and Disease Resistance in Cassava and Coincidence of some QTL with Introgression Regions derived from *M. glaziovii* (doi: 10.3389/fpls.2017.01168).

# List of Electronic Supplementary Files

**Electronic supplementary File SS1** Candidate genes found within the QTL regions

1. **Electronic Supplementary File SS1_1** Significant terms found within the QTL regions
2. **Electronic Supplementary File SS1_2** Significant terms found within the QTL region and related to disease resistance

**Electronic Supplementary File SS2** Supplementary information for introgression analysis

1. **Electronic Supplementary File Table SS2_1** Genes found in the QTL region overlapping introgression segments in Kiroba and related to diseases resistance in plants
2. **Electronic Supplementary File SS2_2** Enrichment analysis showing terms related to disease resistance in plants found in the QTL regions overlapping the introgression segments

*3.* **Electronic Supplementary File SS2_3** 133 significant fragments about 1000bp in length found to have SNPs that are more like M.glaziovii in Kiroba but different from *M. esculenta*

*4.* **Electronic Supplementary File SS2_4** Genes found in the most significant introgression segments introgressed into Kiroba from *M. glaziovii*

5. **Electronic Supplementary File SS2_5** Annotation terms found in the introgression regions of Kiroba, genes associated with them and their orthologues in Arabidopsis

**Electronic Supplementary File SS3** Genotypic data for linkage mapping

**Electronic Supplementary File SS4** Data for QTL analysis using GACD software.

# Symbols and Abbreviations

| | |
|---|---|
| AFLP | Amplified fragment length polymorphism |
| AGO | Argonaute proteins |
| ARI | Agricultural research institute |
| ANOVA | Analysis of variance |
| ASARECA | Association for Strengthening Agricultural Research in Eastern and Central Africa |
| bp | Base pairs |
| CBSD | Cassava brown streak disease |
| CBSV | Cassava brown streak virus |
| CGM | Cassava green mite |
| CMD | Cassava mosaic disease |
| CIAT | International Centre for Tropical Agriculture |
| CIM | Composite interval mapping |
| °C | Degrees celcius |
| cM | CentiMorgan |
| CP | Cross-pollinated (out-breeder full-sib family) |
| DH | Double haploid |
| DNA | Deoxyribonucleic acid |
| dNTP | Deoxynucleotide triphosphate |
| EST | Expressed sequence tag |
| F1 | Filial generation 1 |
| FAO | Food and Agriculture Organization |
| g | Gram |
| GACD | Genetic Analysis for Clonal F1 and Double cross |
| GBS | Genotyping by sequencing |
| GW | Genome-wide |
| GSS | Genome survey sequences |
| GS | Genomic selection |

| | |
|---|---|
| Ha | Hectare |
| HTGS | High-throughput genomic sequencing |
| IITA | International Institute of Tropical Agriculture |
| ILRI | International Livestock Research Institute |
| IBS | Identity by states |
| IM | Interval mapping |
| LOD | Logarithm (base 10) of odds |
| LRR | Leucine rich repeats |
| MAP | Months after planting |
| MAS | Marker-assisted selection |
| masl | Meters above sea level |
| Mb | Megabase pairs |
| NBS | Nucleotide binding sites |
| NEPAD | New Partnership for African Development |
| ng | Nanogram |
| PCR | Polymerase Chain Reaction |
| PVE | Phenotypic variance explained |
| QTL | Quantitative trait loci |
| RAPD | Random amplified polymorphic DNA |
| RFLP | Restriction fragment length polymorphism |
| RGA | Resistant gene analogue |
| RNA | Ribose nucleic acid |
| SE | Standard error |
| SNP | Single nucleotide polymorphism |
| SSR | Simple sequence repeats |
| US$ | United States dollars |

# Chapter 1

## Literature Review:

Progress, Challenges and Opportunities for breeding Cassava against Pests and Diseases in Africa

# 1

## 1.1 Introduction

Cassava (*Manihot esculenta* Crantz) is the most important staple crop in Africa based on production (FAOSTATS, 2017). Its total production of 178 million metric tonnes (MT) is twice that of maize at 84 million MT. It is consumed daily by more than 800 million people, mainly from sub-Saharan Africa (Lebot, 2009). It is adapted to poor soils and harsh climatic conditions and is regarded as a food security crop, providing a yield when cereals and other food crops fail (Raphael, 2008). It is drought tolerant and offers a flexible harvesting regime since the roots can remain in the soil and be harvested when needed, thus forming an important safeguard against unexpected food shortages (Nzuki et al., 2017). Due to this, the New Partnership for African Development (NEPAD) has cited cassava as a powerful poverty fighter (Nweke et al., 2002). Apart from being used as human food, cassava is also used as animal feed and for industrial purposes (Nzuki et al., 2017).

Annually, approximately 297 million MT of cassava are produced worldwide with Africa contributing more than 178 million MT (59.9%) (FAOSTAT, 2017), which is greater than for any other crop in Africa. Though the yields are low in Africa, overall production of 59.9% of the global total is high. Statistics indicate the average yield of cassava in Africa to be 10 (MT)/ha, yet its potential yield is estimated to be 100 MT/ha (FAOSTAT, 2017h). In East Africa, cassava yield is very low with 8.1 MT/ha in Democratic Republic of Congo, 9.8 MT/ha in Tanzania, 10.6 MT/ha in Kenya and 12 MT/ha in Uganda. However, yields in some Asian countries are much higher, such as in China with 16.3 MT/ha, Indonesia 16.2 MT/ha, Thailand 22.9 MT/ha, and India 31.4 MT/ha (FAOSTAT, 2015).

Cassava yields remain low in Africa owing to the growing of cultivars with low potential and pest and disease susceptibility, declining soil fertility coupled with lack of farm inputs, and poor farming systems. In Africa, the population is expected to double in the next 50 years and with cassava being such an important food security crop, it is imperative that yields are increased to levels at least similar to that in Asia (United Nations, 2017).

## 1.2 Major constraints to production in cassava

In addition to farmers growing inherently low yielding, unimproved cultivars, and poor agronomic practices, pests and diseases are a major constraint to cassava production (IITA, 2003). Cassava green mite (CGM), cassava mealybug, and the variegated grasshopper are the major cassava pests while cassava mosaic disease (CMD), cassava brown streak disease (CBSD), and cassava bacterial blight are some of the common diseases (Campo et al., 2011). Viral diseases are the most threatening and yield-limiting due to their current epidemiological and distribution trends (Legg et al., 2011).

CBSD, originally considered to be localized mainly in the coastal regions of East Africa and limited by altitude below 1,000 m a.s.l. (Alicai et al., 2007), is now rapidly spreading to previously unaffected areas at alarming rate (Campo et al., 2011). It is thought that this is being fueled by changing climatic conditions, and the diversity and population dynamics of whitefly in the region (Legg et al., 2008). Conventional breeding is already underway in CBSD affected areas, but pre-emptive breeding is urgently needed in unaffected areas of Central and West Africa, the major cassava producing regions in Africa, to mitigate the threat posed by CBSD (IITA, 2003).

Cassava mosaic disease (CMD) is caused by cassava mosaic geminiviruses and just like CBSD, it is also spread by by a white fly vector, *Bemisia tabaci* (Legg and Hillocks, 2003). Currently, CMD is managed through phytosanitary practices and the use of conventional resistance breeding. Though early research on breeding of cassava for resistance to CMD began at Amani in the 1930's (Jennings, 1994), the disease has continued to affect crops without a concrete solution towards its eradication. Three genomic regions associated with CMD resistance are currently known (Wolfe et al., 2016). CMD1 is a quantitative source of resistance identified in the Amani derived interspecific cultivar TMS 30572 (Fregene et al., 2000; Mohan et al., 2013), CMD2 is a region of large effect which has been widely used in the IITA breeding program and has been mapped in several studies (Akano et al., 2002; Lokko et al., 2005; Okogbenin et al., 2007; Rabbi et al., 2014), and CMD3 is on the same chromosome as CMD2 (Okogbenin et al., 2012). Recently a genome wide association study revealed the presence of an interacting locus close to CMD2, which suggests either epistatic interactions or multi-allelic effects (Wolfe et al., 2016).

Cassava green mite (CGM) was discovered in Uganda in 1971 and since then, it has spread to other countries in Africa, causing damage to cassava fields (Gutierrez et al., 1988). It is a serious pest of cassava that impacts growth and yield (Skovgård et al., 1993). It causes 60% chlorophyll depletion and about 50% leaf area reduction, leading to heavy yield reductions and sometimes total crop failure (Hui, 2012). Most studies on CGM have focused on conventional breeding (Chipeta et al., 2013; Chalwe et al., 2015), with limited published work on molecular breeding (Macea Choperena et al., 2012). Two SSR markers, namely NS1099 (chromosome XIV) and NS346 (chromosome XVIII), showed high association with resistant families (Ceballos et al., 2010). Two green mite predators of cassava particularly in Africa, namely *Typhlodromalus manihoti* and *Euseius fusti*, have previously been used for control of CGM (Onzo et al., 2005 Neuenschwander, 2003).

### 1.3 Cassava Brown Streak Disease

CBSD is a fast spreading and devastating viral disease that is threatening food security in sub-Saharan Africa due to its adverse effect on yield and in some cases total crop loss (ASARECA, 2009). CBSD causes a brown dry necrotic rot in the storage root tissues, leading either to complete spoilage or significant reductions in quality (Hillocks, 1997). In addition, farmers tend to harvest early to avoid root necrosis leading to yield losses (Hillocks and Jennings, 2003).

CBSD also causes a foliar chlorosis that may impact yield, although there are no clear data to support this observation at the moment (Alicai et al., 2016). The foliar chlorosis initially follows the veins and symptoms are most often evident on the lower leaves (ASARECA, 2009). On the stems, the disease appears as brown streaks on the upper green portions (Nichols, 1950). Dieback and root deformation is also evident in the advanced stages of the disease (Hillocks, 1997). The typical response of cassava plants is to show symptoms on the leaves, stems and roots, but symptoms can be variable and unpredictable, and appear to be influenced by environmental conditions and/or geographical location (Hillocks and Thresh, 2000). Some cassava cultivars only show symptoms on the leaves and not on the roots, and inversely for others (Kaweesi et al., 2014). Some cultivars do not show symptoms on either the leaves or roots and are known as field resistant. The symptoms may appear in young plants, then disappear but reappear at a later stage (ASARECA, 2009).

CBSD is caused by two viruses, namely cassava brown streak virus (CBSV) and Ugandan cassava brown streak virus (UgCBSV) (Mbanzibwa et al., 2009). More than 70 coat protein sequences and two complete sequences of CBSV from Uganda and from Tanzania (Lake Victoria area) are now available, and the data reveal a clear distinction between the two species. Winter et al (2010) initially reported some geographical localization of the different species, but new evidence suggests their distributions overlap substantially (Abarshi et al., 2012). These viruses are known to be transmitted by whitefly (*Bemesia tabaci*) (Maruthi et al., 2005) and through the planting of infected materials. Additionally, the movement of materials from infected areas to non-infected areas is a major avenue for spread of the viruses (ASARECA, 2009).

Cassava brown streak viruses (CBSVs), are monopartite +ssRNA viruses encapsulated into flexuous filamentous particles and have been classified into the genus *Ipomovirus* (Family *Potyviridae*) (Mbanzibwa et al., 2009). A cassava brown streak virus was demonstrated as the causative agent of CBSD through sap transmission (mechanical transmission) and transmission to herbaceous hosts in glasshouse studies (Lister, 1959). This has also been supported by graft transmission (Thresh, 2003).

The most plausible approach to combat CBSD is to combine breeding for host plant resistance with sanitation measures and the planting of virus-free stakes. It is important that breeding for host plant resistance is performed as quickly and efficiently as possible, and should take advantage of genomics tools, including both transformation and molecular marker technology. Conventional breeding alone takes up to 10 years to deliver a new cultivar (Nweke et al., 2002). This is partly due to the fact that plants need to be grown for 12 months before selection for CBSD resistance is made. If molecular markers were found to be associated with field resistance, then F1 generated in a breeding program could be screened and selected at the seedling stage, thereby drastically reducing the breeding cycle and providing an accurate way of efficiently pyramiding resistance genes from different sources and hence breeding for durable resistance.

## 1.4 CBSD spread and economic losses in ECA countries

CBSD was first reported in coastal Tanzania in the 1930s at the Amani breeding research station (Storey, 1936) and was thought to be restricted by altitude to coastal regions of southern and East Africa and neighboring regions such as Malawi and Zambia. Since 2004, it spread rapidly to other regions of high altitude where it was originally thought to be absent (Muhanna and Mtunda, 2003). It is now causing significant yield and root losses in the Lake Victoria areas of Tanzania, Kenya and Uganda, and has been reported in Rwanda and the Democratic Republic of Congo (DRC) (Mulimbi et al., 2012). It is currently thought to be the greatest threat to food security in the Great Lakes region (Alicai et al., 2007). Recent disease modeling studies predict CBSD hotspots in West Africa and Latin America (Campo et al., 2011).

The brown corky root necrosis caused by CBSD results in significant loss of quality, which translates into yield loss (Hillocks et al., 2001). This is in addition to deformation of the roots and loss of photosynthetic capacity through leaf chlorosis and partial or total plant death through dieback (Gondwe, 2011). In Tanzania yield losses have been estimated at 34% amounting to about 350,000 MT in the coastal regions alone (Muhanna and Mtunda, 2003). This translates to about USD 16.5 million losses due to CBSD in Tanzania (Kanju et al., 2002). In Malawi, 40% incidences have been recorded in the coastal regions mostly along the shores of Lake Malawi (Legg and Hillocks, 2003). The mean loss in production was estimated to range from between 20-25% of the harvest. In Mozambique, the incidence of CBSD is mainly in Nampula and Zambezi provinces with a mean incidence of 30% and 40% respectively (Thresh, 2003). In Rwanda, a surveillance analysis conducted by the National Agricultural Research Institute in 2010 showed 15.7% infection of local cultivars and 36.9% in improved cultivars (Munganyinka et al., 2018). It is now considered the greatest threat to food security in Uganda (Legg et al., 2011).

## 1.5 Challenges in breeding for CBSD resistance

Cassava is a challenging crop to breed (Ceballos et al., 2004). It is an out-crossing clonally propagated species that harbours high levels of heterozygosity (Bredeson et al., 2016). For this reason it suffers from rapid inbreeding depression (Kawuki, 2010). In addition, flowering is

variable and highly influenced by the environment, and seed set and germination rates are low (Jennings, 1963). It has a relatively long generation time from seven or eight months for early bulking cultivars to one year in normal cultivars (Jennings, 1957). Although cassava cultivars that show resistance to both CMD and CBSD have been released through conventional breeding (Muhanna and Mtunda, 2003), the purely field-based selection process is rather ineffective and limiting especially where more than one trait has to be selected simultaneously. In order to determine the response to CBSVs, $F_1$ breeding progeny have to be grown to maturity (12 months) before they can be evaluated (Ferguson personal comm.). In addition, as the virus accumulates in plants over time, evaluation needs to be done over several years (Amuge et al., 2017). The field screening also relies on the natural transmission of whiteflies, which is not 100% effective in a single season (Maruthi et al., 2017). Large populations with expensive field trials hinder the breeding of new cassava cultivars with desired level of resistance to CBSD within a reasonable time frame (Kaweesi et al., 2014). Due to these challenges, it takes about eight to twelve years to release a new cassava cultivar through conventional breeding (Blair et al., 2007).

### 1.6 Approaches to combat CBSD

Current approaches to combat CBSD include breeding for host plant resistance and sanitation (Patil et al., 2015). Sanitation includes distribution of clean planting material to farmers, removal of infected plants selection of healthy stems for replanting and disinfection of tools used for cutting the planting stakes. This method has been advocated by researchers and has been accredited to reducing the spread of CBSD in farmer fields (Legg et al., 2017). This is particularly effective since CBSD transmission is semi-persistent, and the pattern of symptom expression is variable (Legg et al., 2011). However, this solution is short lived and a more permanent control of CBSD is urgently needed to counter the predicted CBSD spread in cassava growing regions (Campo et al., 2011). Furthermore the symptoms are confusing and farmers are unable to notice them early rendering the phytosanitary methods ineffective (Legg et al., 2017).

The most sustainable and feasible control strategy to reduce CBSD losses is the deployment of resistant cultivars (Hillocks and Jennings, 2003). These cultivars are envisaged to possess molecular barriers to prevent or hinder the invasion or spread of the virus in the plant by

employing complex molecular defence mechanisms to antagonise virus activity. These mechanisms could involve small interfering RNA (siRNA's) that play a role in plant defence systems (Harvey et al., 2011). R genes that encode proteins with nucleotide binding site (NBS) and leucine-rich repeat (LRR) domains have also been reported in plants (Ameline-Torregrosa et al., 2008). RNA interference (RNAi) has been used induce crop protection against insects (Price and Gatehouse, 2008). It has been used to produce male sterility in tobacco and enhance nutrition content in tomato, wheat and potato (Muhammad Shah Nawaz-ul-Rehman, 2007).

Cassava breeders in the East Africa region have identified a number of germplasm clones with high levels of tolerance to CBSD which are being used in breeding programmes in Tanzania (Hillocks and Jennings, 2003). These cultivars include Namikonga, Kiroba, Nachinyaya, AR40-6 and NDL06/132. These cultivars may be infected by the virus and may show leaf symptoms, but the onset of root necrosis is delayed and limited until 12 to 18 months after planting or later, and the root fill is not impaired and the full yield potential is realized (Hillocks and Jennings, 2003). The resistance in these cultivars is sometimes called 'field resistance' by breeders (Hillocks and Jennings, 2003). These 'field resistant' cultivars offer a promising breeding window to combat CBSD.

Breeding for CBSD resistance must be combined with CMD resistance, since the two diseases co-exist and pose a double challenge to farmers. The challenge is to efficiently introgress sustainable CMD resistance with durable CBSD resistance, and other farmer and market-preferred traits, all within a reasonable time frame and cost (Vanderschuren et al., 2012). In addition, it has been noted that many CBSD resistant cultivars are also CMD susceptible (Masumba et al., 2017). For example, Kiroba is being utilized in cassava breeding programs to introgress tolerance to farmer preferred cultivars which are currently susceptible to CBSD. It is largely male sterile, gets infected by cassava brown streak viruses (CBSVs), shows mild CBSD leaf symptoms with almost no root necrosis even under very high disease pressure and is regarded as CBSD tolerant by breeders (Kaweesi et al., 2014). In contrast, Kiroba is very susceptible to CMD. It is high yielding (20 MT/ha) with a high dry matter content of over 30% (Muhanna and Mtunda, 2003). AR37-80 is a CIAT improved cultivar being a cross between a CMD resistant line (C33) from IITA and CW259-42 which is a backcross of MTAI 8 (Rayong 60) and an interspecific cross between *M. flabellifollia* and CM 2766-5. It was developed through MAS, being positively selected for markers for the CMD2 resistance locus and

markers for CGM resistance. It is resistant to CMD and CGM but susceptible to CBSD (Blair et al., 2007; Okogbenin et al., 2012).

Pyramiding of an array of favourable resistance genes for both CMD and CBSD using conventional methods becomes extremely difficult, requiring very large numbers of progeny, and vast expensive field experiments (Kunkeaw et al., 2010). To shorten the breeding cycle and reduce population sizes, cassava breeders must embrace modern advances in biotechnology, including Marker Assisted Selection (MAS) (Blair et al., 2007). Molecular markers associated with CBSD and CMD resistance/tolerance could be used in marker-assisted breeding programs to increase the precision of selection by allowing breeders to make decisions on the basis of genetic make-up of individuals at the seedling stage (Nzuki et al., 2017). For effective marker assisted selection, markers must be very closely linked to the trait of interest to reduce the possibility of recombination between the trait and the marker, thus providing greater reliability. Also, the marker should occur at a low frequency within recipient cassava clones, making it applicable across an array of recipient backgrounds (Ferguson personal communication). Markers associated with one source of field tolerance to CBSD have been identified and are currently being validated (Rabbi et al., 2012). MAS has also been applied for CMD (Okogbenin et al., 2012).

To speed up the breeding process, breeders need to take advantage of genomics tools, including transformation, gene editing and molecular marker technology. Transformation technology is being used at the Donald Danforth Plant Sciences Centre (DDPSC), International Institute of Tropical Agriculture (IITA), ETH (Zurich), Agriculture Research Institute (ARI)-Tanzania, and the National Agriculture Research Organisation (NARO), Uganda (Vanderschuren et al., 2012). Initially cassava was recalcitrant to transformation approaches (Li et al., 1996), but significant progress has been made and now transformation approaches with good transformation efficiencies, and tissue culture systems that allow selection and regeneration of transgenic cassava plants, are available (González et al., 1998).

### 1.7 Sources of CBSD resistance and the Amani Breeding Program

Cassava breeding for host plant resistance, particularly CMD, began at the Amani breeding station in Tanzania in the 1930's (Hillocks and Jennings, 2003). After disappointing results

from screening over 100 cultivars sourced globally for CMD and CBSD resistance, the program focused on crossing the most resistant cultivars and on inter specific hybridization (Nichols, 1947). Derivatives from *Manihot melanobasis* (now recognized as *Manihot ssp. flabellifolia*) were found to be very good candidates showing strong resistance with the ability to localize the virus at the base of the stem (Jennings, 1957). Crosses with *Manihot glaziovii* backcrossed three times and intercrossed with resistant hybrids produced interspecific hybrids that were rated over 80% resistant to CMD and moderately resistant to CBSD (Jennings, 1957). Some of the best known intercrosses at Amani included cultivars 5318/34, 46106/27, and 5543/156 (Jennings, 1994). The Amani breeding program was closed in 1958 and the materials that had been developed were distributed to many research centers in Africa. The Amani hybrids also became important for the International Institute of Tropical Agriculture's (IITA's) cassava breeding program in the 1970s at Moor Plantation in Nigeria (Beck, 1982). When the Amani program ceased, it is thought that some of the inter specific crosses found their way into farmers' fields in Tanzania and have been incorporated as farmer cultivars (Kanju et al., 2002). The clones may have lost their identities and are being grown by farmers under different local names. Cassava breeders have identified some of these interspecific hybrids, and they show strong field resistance to CBSD. They are infected by the CBSD causative viruses; some show leaf symptoms, but the onset of root necrosis is delayed and limited allowing full yield potential (Hillocks and Jennings, 2003). These cultivars are a rich genetic stock for cassava breeding, however, the genetic basis and mechanism of resistance are not clearly understood.

Namikonga, Kiroba and Nachinyaya are local landraces (farmer cultivars) that are thought to be former Amani hybrids that have lost their identity (Nzuki et al., 2017). In Kenya, Namikonga is referred to as Kaleso, which was presumably accession number 46106/27 and was a third backcross derivative from *M. esculenta* x *M. glaziovii* (Hillocks and Jennings, 2003). Kiroba is generally confined to the coastal region in Tanzania, whereas Nachinyaya is confined to the southern coastal areas. It is also popular in Cabo Delgado in Mozambique with the same name. Nachinyaya stems give the impression that the clone could be derived from *M. glaziovii* because they resemble the wild species. Nachinyaya was identified in 1995 from southern Tanzania (Thresh, 2003) with delayed onset of root necrosis (Hillocks and Jennings, 2003). Kiroba does not have any visible wild characteristics, shows mild CBSD leaf symptoms in Tanzania, but does not show any leaf symptoms in Uganda (Kaweesi et al., 2014), with almost no root necrosis even under very high disease pressure. A diversity assessment using SNP markers shows Kiroba to be quite closely related to Namikonga, the known Amani derivative,

(Ferguson et al., 2012). Namikonga is known to confer good tolerance to CBSD and putative QTL for CBSD from this source of resistance have already been identified (Masumba et al., 2017). It is hypothesized that Kiroba may also be a derivative from a *M. glaziovii* × *M. esculenta* interspecific cross.

Genetic diversity analysis in East, Central and Southern Africa regions show a moderate genetic variation among the cultivars within the national breeding programs (Kawuki et al., 2013), though there exist some variation among cultivars and landraces in the region. Among the countries studied, Tanzania has diverse landraces some of which have been noted to be tolerant against viral diseases (Hillocks, 2003). Recent diversity assessment using SNP markers revealed a close but complex relationship among the Tanzanian landraces (Ferguson et al., 2012; Wolfe et al., 2016; Nzuki et al., 2017). The landraces do not exist within a formal germplasm conservation initiative and are found in farmers' fields and National Agricultural Research Systems (Kawuki et al., 2013. The narrow genetic base among cultivars present a bottleneck in breeding against viral diseases. To broaden the diversity and improve on the farmer preferred cultivars, initiatives for crossing the local cultivars with genotypes from other regions such as West Africa or South America, utilizing the local landraces or tapping into the wild relatives are urgently needed.

### 1.8 Research approaches to gene discovery for resistance to viral diseases.

The cassava draft genome (v4.1) was initially released in 2009 (Prochnik et al., 2012) based on a partially inbred line, namely AM560-2. It spanned 533 Mb with 12,977 scaffolds. A consensus genetic linkage map, based on 10 bi-parental populations, was constructed (ICGMC, 2015) and used to anchor further sequence information into a physical map (v5.1). The recent assembly (v6) spans 582 Mb (Bredeson et al., 2016) with the annotation showing that 518 of the scaffolds contain 96.6% of the genes. Whole genome re-sequencing of 15 popular cassava cultivars including Namikonga and *M. glaziovii* (Bredeson et al., 2016) provides an opportunity for comparative analysis with the reference genome. Moreover, research efforts towards tapping into the rich reservoirs of wild relatives of cassava for important genes for viral disease resistance including CMD and CBSD have been initiated (Olsen and Schaal, 2001).

Different disease-resistance mechanisms to detect pathogens and to induce defence responses, including the use of resistance (R) genes have been discovered in plants (Kim et al., 2012). The largest class of these R genes encode proteins with nucleotide binding site (NBS) and leucine-rich repeat (LRR) domains (Ameline-Torregrosa et al., 2008). It has been understood that plants have developed resistance or even immunity by expressing R genes as receptors that recognize an invading pathogen and mediate defence signaling to prevent pathogen spread (Kim et al., 2012). Screening the R-genes from various plants and analyzing their diversity is ongoing to understand the R-gene architecture and their functions (Hammond‐Kosack and Kanyuka, 2007). *In silico* analyses of genome sequences to identify and characterize NBS-LRR-encoding resistance genes in several plant species have been done (Gedil et al., 2012). A combination of computational and experimental analyses revealed the presence of large numbers of NBS-LRR-encoding genes in *Arabidopsis* (Meyers et al., 2003), rice (Zhou et al., 2003), poplar (Tuskan et al., 2006), grapevine (Moroldo et al., 2008), papaya (Porter et al., 2009), and *Medicago truncatula* (Ameline-Torregrosa et al., 2008). In cassava, conserved NBS-LRR sequences mapped to the draft cassava genome assembly with high sequence similarity, thus identifying resistance gene analogues (RGA) (Gedil et al., 2012). Further characterization and experimental validation of these predicted R genes is required to develop functional gene-targeted markers to be used in molecular resistance breeding to combat CBSD and CMD.

RNA silencing is an alternative natural antiviral defence mechanism that has been explored in plants. It involves the Argonaute (AGO) proteins that use bound small interfering (si)RNAs to target cleavage or translational suppression of complementary RNA (Harvey et al., 2011). Plant viruses encode suppressor proteins of siRNA as counter-defence mechanisms that influence the accumulation and spread of viruses in infected plants (Roth et al., 2004). There are many classes of small RNAs, including sense-antisense small interfering RNAs (siRNAs), microRNAs (miRNAs), heterochromatic siRNAs (hc-siRNAs), piwi-interacting RNAs (piRNAs) and trans-acting siRNAs (ta-siRNAs) (Allen et al., 2005) that have been researched in plants. MiRNAs are involved in plant responses to the environment. Several miRNAs are up-regulated or down-regulated by abiotic stress, including high salinity, drought and low temperatures (Sunkar and Zhu, 2004). MiRNAs are also induced by pathogens, suggesting their involvement in plant-microorganism interactions (Katiyar-Agarwal et al., 2007). Thus, miRNAs can provide a novel platform to better understand plant development and resistance to biotic and abiotic stresses. Next generation high throughput sequencing, particularly short-

read Illumina technologies, have been used to develope protocols for investigating small RNAs (Jarvie and Harkins, 2007). Comparative genomics has advanced to a level whereby it is now possible to perform BLAST searches using known miRNAs against Genebank databases, including genome survey sequences (GSS), high throughput genomic sequences (HTGS), expressed sequence tags (ESTs), and non-redundant (NR) nucleotides easily.

## 1.9 Development of genetic marker technologies

In the past, plant breeders have relied on phenotypic selection to advance breeding programs by assessing various agronomic and quality traits in crops (Khan et al., 2015). These early breeding schemes depended on morphological characters or markers that were largely influenced by the environment (Khan et al., 2015). In the early 1980s, biochemical markers were developed and applied in crops (Tanksley, 1988). Allozyme markers were the first to be developed and easily gained acceptance by complimenting morphological markers (Tanksley, 1993). Allozymes are based on protein polymorphism and gained acceptance because the allelic forms of the enzymes could easily be separated by starch gel electrophoresis and detected by staining the gels (Collard et al., 2005). The advantages of allozyme markers included the low cost, technical simplicity and the fact that the markers are co-dominant, so that homozygous and heterozygous genotypes could be distinguished (Weising et al., 2005)). Disadvantages of this method include a limited number of suitable allozyme loci in the genome, and the requirement for fresh tissue of the appropriate developmental stage (Weising et al., 2005).

Later in the 1990s, DNA markers were developed and allowed genetic variation to be observed at the DNA level (Williams et al., 1990; Vos et al., 1995; Powell et al., 1996). Restriction fragment length polymorphisms (RFLP) were the first DNA based genetic markers to be developed and relied on the use of restriction enzymes and resolution of the fragments by gel electrophoresis using radiolabeled probes (Botstein et al., 1980). The development of the polymerase chain reaction (PCR) in the 1980s (Mullis and Faloona, 1987), led to the development of more PCR based marker systems such as randomly amplified polymorphic DNA (RAPD) (Williams et al., 1990), amplified fragment length polymorphism (AFLPs) (Vos et al., 1995), and microsatellites (simple sequence repeats, SSRs) (Powell et al., 1996). SSRs were the preferred markers for linkage mapping due to their co-dominance and high levels of

polymorphism (Collard et al., 2005). However, significant effort is required since SSR markers must be developed specifically for each species and are not transferable between species (Goldstein et al., 1999). AFLP markers are easy to use and reveal large sets of genetic loci, but are dominant (heterozygotes cannot be discerned) and their transferability between detection platforms (for instance, polyacrylamide gel electrophoresis, gel-based sequencers, and capillary sequencers) can be difficult and tedious (Papa et al., 2007).

Recently, and as a result of advances in genomics and declining costs of sequencing, SNP markers have gained more acceptance in various crops (Collard et al., 2005). This is particularly so because SNP markers are based on sequence differences at a single base pair position and are generally bi-allelic. They are the most abundant class of polymorphism in genomes and their sheer abundance compensates for the lack of allelic polymorphism (Jehan and Lakhanpaul, 2006). Estimates of SNP density in cassava indicate a frequency of one SNP for every 121 base pairs (bp) (Kawuki et al., 2009). SNPs offer greater marker density than SSR markers (Vignal et al., 2002), increasing the probability of obtaining a marker closely linked to a target functional gene. This means that linkage defined by SNPs can be more stable than other markers, since the linkage can be lost when those markers are applied to other populations with different recombination patterns (Xu and Crouch, 2008). SNPs also provide gene-based markers that may prove useful in identifying candidate genes of interest associated with QTL (Rafalski, 2002). SNP-based markers are now being used to generate dense genetic linkage maps (Rabbi et al., 2012).

Technologies to visualize SNPs in a cost-effective manner have been developed, including genotyping by sequencing (GBS) (Elshire et al., 2011) and KASpar technology by KBiosciences (Smith, 2013). During GBS, the complexity of the genome is reduced through the use of restriction enzymes. High-throughput next generation sequencing (NGS) is used to sequence from restriction enzyme cut sites. Alignment of the resulting reads and variant calling delivers a large number of SNPs. It is simple, quick, specific and highly reproducible (Poland and Rife, 2012). In cassava, rice and sorghum, GBS-generated SNP markers have been used to construct dense genetic linkage maps (Deschamps et al., 2012). Today, it is being applied in many crops including polyploids (Poland and Rife, 2012). The KASPar technique is based on competitive allele specific PCR which enables the detection of SNPs and indels (Smith, 2013). Instead of genome wide coverage, this technology amplifies a few targeted SNPs. In cassava, 1,190 SNP markers were validated and converted to KASP genotyping assays (Ferguson et al.,

2012). KASP genotyping will gain more use in various crops in the near future due to its technical simplicity and application.

## 1.10 Genetic linkage mapping in cassava

The first genetic linkage map of cassava was based on SSR markers and was derived from two non-inbred elite lines, namely TMS 30572 and CM2177-2 from IITA and CIAT, respectively (Okogbenin et al., 2008). The map was based on 268 F1 progeny, with 100 markers spanning 1236.7 cM distributed on 22 linkage groups, with an average marker distance of 17.92 cM. A second genetic linkage map for cassava was generated from 58 $F_1$ progeny from a cross between Rayong 90 (female) and Rayong 5 (male) and had 33 linkage groups that were comprised of 119 AFLP markers and 18 SSR markers, and spanned 1095 cM with an average marker distance of 7.99 cM (Kunkeaw et al., 2010). More recently, a map comprising of 568 markers (434 SNPs and 134 SSRs) distributed across 19 linkage groups and spanning 1,837 cM was published (Rabbi et al., 2012). The average marker distance was 3.4 cM and 94.2% of the mapped SNPs and SSRs were localized on scaffolds of the draft cassava genome sequence. Recently, a consensus genetic linkage map was published for cassava, based on 10 mapping populations, which included the population described in the current thesis (ICGMC, 2015). The map had 22,400 markers distributed across 18 linkage groups, corresponding to the number of chromosomes for this species. This linkage map enabled scaffolds from the draft genome sequence of cassava to be anchored into a physical map (Bredeson et al., 2016).

## 1.11 Quantitative trait loci (QTL) mapping

QTL mapping has been reported in many crops for various traits including yield, quality, disease resistance and abiotic stress (Edwards et al., 1987). QTL mapping has been used to position several blast resistance genes in a molecular map of rice (Wang et al., 1994) and in maize, QTL mapping for grey leaf spot has been reported (Bubeck et al., 1993). QTL mapping for resistance to stripe virus disease in rice has also been done (Wang et al., 2011). In cassava, QTL mapping for cassava greenmite, cassava mosaic and cassava brown-streak diseases is underway (Masumba et al., 2017; Nzuki et al., 2017).

During QTL mapping, a suitable mapping population has to be developed using parental genotypes that are highly contrasting phenotypically for the target trait (eg. highly resistant and susceptible lines) (Semagn et al., 2006). In addition, the parental lines should be genetically divergent to enhance the possibility of identifying a large set of polymorphic markers that are well distributed across the genome (Mackay et al., 2009). This requires a molecular polymorphism survey across a set of potentially useful lines so as to identify the most suitable ones for generation of a mapping population (Prasanna, 2005).

QTL analysis has been based on different population types such as F2-derived populations, recombinant inbred lines (RILs), near-isogenic lines (NILs), double haploids, and backcross populations (Kearsey and Farquhar, 1998). The principle behind this is the occurrence of recombination events between genetic loci during meiosis, when gametes are formed and transmitted from parents to offspring (Wang et al., 2011). In cassava, $F_1$ mapping populations are normally used due to the difficulty of crossing, the low seed set, heterozygous nature of the crop and the long breeding cycle (Okogbenin and Fregene, 2002). A pseudo-$F_2$ population has also been used for QTL mapping in cassava (Okogbenin et al., 2008).

After establishing a reliable mapping population, the second step is to perform phenotypic characterization with respect to the trait of interest (Semagn et al., 2006). Good quantitative data is important to get the discrimination needed for accurate QTL analysis (Mackay et al., 2009). Finally, an association between genotypes and phenotypes in the mapping population is established (Prasanna, 2005). If the parents carry different alleles for the QTL controlling a given trait, the trait values in the segregating mapping population will be associated with the alleles of the markers that are closely linked to the QTL (Wang et al., 2011). By scanning the markers on a linkage map for association with trait values, likely map positions for QTL can be detected, which is an important step towards understanding the inheritance and genetic basis of the traits (Semagn et al., 2006).

A number of factors limit the accuracy of locating QTL (Prasanna, 2005; Semagn et al., 2006). Experimental design is important and key factors such as small sample size may influence the accuracy of QTL detection (Xu, 2003). The power to detect true QTL diminishes when a small sample size is used, leading to bias in the estimation of the proportion of the genetic variance explained by the mapped QTL (Beavis, 1998). Consequently, only the major or large QTL are detected, while the small and intermediate QTL are discarded. The true QTL that are not detected in small sample sizes have effects that are detected at the regions that are declared as

having QTL. This results in the overestimation of QTL effects and is known as "The Beavis effect'' (Xu, 2003). This can be avoided by use of a large population size to increase the number of recombinant events obtained, thereby greatly increasing the chances of detecting all the QTL (Prasanna, 2005; Semagn et al., 2006). However, QTL mapping of very complex traits controlled by many genes with small effects remains a challenge and may produce unreliable results (Schön et al., 2004). An $F_2$ population can thus be more powerful to detect QTL of additive effects (Semagn et al., 2006). Environmental effects also influence the expression detection of QTL (Mackay et al., 2009). This can be resolved through replication and multi-locational trials (Prasanna, 2005).

Genome Wide Association Studies (GWAS) is an alternative approach to QTL mapping, and relies on a panel of diverse cultivars with minimal population structure (Wolfe et al., 2016). Associations between genotype and phenotype are determined within the panel hence a mapping population does not have to be developed (Kayondo et al., 2018).The results do not indicate which cultivar actually harbors the QTL of interest (Kawuki et al., 2016).With a recently sequenced genome (Prochnik et al., 2012) and SNP-based genetic linkage maps for cassava (ICGMC, 2015; Masumba et al., 2017; Nzuki et al., 2017), it is now possible to study the genetic architecture of key traits using modern GWAS approaches and to improve those traits with GS (Nakaya and Isobe, 2012). Already in cassava, GWAS and GS (Wolfe et al., 2016; Kayondo et al., 2018) are being applied to validate QTL associated with CMD and CBSD that have been identified (Masumba et al., 2017; Nzuki et al., 2017). The QTL on chromosome 4 and chromosome 11 associated with root necrosis detected in this study have been confirmed through this approach (Kayondo et al., 2018).

## 1.12 Marker-assisted breeding

Molecular markers are now being applied in a number of ways to enhance the efficiency of conventional breeding (Ribeiro et al., 2012). In cassava these methods include marker- assisted selection (MAS), genomic selection (GS) and marker assisted recurrent selection (MARS) (Okogbenin et al., 2007; Wolfe et al., 2016; Kayondo et al., 2018). GS is a method that uses genotypic information to predict breeding values for particular phenotypes without specific knowledge of the individual genes contributing to that trait (Hayes and Goddard, 2001). GS requires a training or reference population and uses a panel of markers that are dense enough

to cover all QTL (Hayes et al., 2013). With MAS only limited proportions of the total genetic variance is captured by markers (Nakaya and Isobe, 2012). By using GS, the entire genome is divided into chromosome segments by omitting significance testing and then estimating the effects of genes. Genomic selection has been applied across a whole range of crop species (Hayes et al., 2013). Uganda and Nigeria are implementing genomic selection for cassava, though it is still in an experimental or validation phase (Kawuki et al., 2016; Wolfe et al., 2016; Kayondo et al., 2018). GS and MAS are now used to complement each other. Often one round of MAS will be performed before GS (Rana et al., 2019).

MAS relies on the principle of finding markers linked to the gene or genes influencing a trait of interest. To do this, a segregating mapping population is developed between a resistant and susceptible parent (Semagn et al., 2006). Markers are then used to construct a genetic linkage map based on recombination frequency relative to other markers in a chromosome or linkage group (Okogbenin et al., 2008). Statistics are then used to identify which markers are most frequently inherited together with the quantitative trait of interest from phenotypic data, where there is minimum recombination between marker and phenotypic state (Mackay et al., 2009). If a marker is closely linked to a gene or genes that influence a quantitative trait they are usually carried over together in a breeding cross, and will both be present in the progeny (Semagn et al., 2006). Thus, in MAS a marker is selected for, usually at the seedling stage, as a proxy for the gene or genes that influence the trait (Prasanna, 2005; Semagn et al., 2006). Marker assisted selection uses DNA markers that are tightly linked to the target loci as a substitute for or to assist phenotypic screening (Okogbenin et al., 2008). Markers are used to determine whether a breeding cross has transferred the desired trait (Okogbenin et al., 2007). If the marker gene is present, it is highly probable that the desired trait will be present in the progeny (Ribaut and Hoisington, 1998).

Marker assisted breeding (MAB) in crops is being applied in two ways, namely marker assisted selection (MAS) and marker assisted recurrent selection (MARS) (Xu and Crouch, 2008). MARS is simply MAS performed for several generations, and uses background selection (Ribaut and Hoisington, 1998). Background selection focuses on eliminating the unwanted contributions of genetic material, typically from a donor parent during trait introgression (Namuth-Covert, 2011). This has been applied in maize, reducing the number of generations needed for recovery of the recurrent genome from eight to three generations (Frisch et al., 1999). In maize breeding, it has been used to transfer quality maize phenotype (QMP), controlled by a mutant allele of the *opaque2* gene from one elite maize inbred line to another

elite maize inbred line (Heisey and Edmeades, 1999). On the other hand, MAS is foreground selection whereby molecular markers are used to assist progress with forward selection, thus breeders select for a desired trait or QTL (Xu and Crouch, 2008).

The use of molecular markers increases the effectiveness of selection and heritability by shortening the development time of cultivars (Ribaut and Hoisington, 1998). MAS in cassava uses markers linked to specific traits of interest to quickly and effectively select plants carrying those traits in segregating populations (Ribeiro et al., 2012). This is done for a single trait or for several traits simultaneously (Semagn et al., 2006). Traits that are difficult to phenotype can be quickly and accurately genotyped using MAS (Blair et al., 2007). SSR markers have been used to screen for CMD resistance in cassava (Mohan et al., 2013; Rabbi et al., 2014). More recently, SNPs linked to CMD and CBSD have been discovered (Masumba et al., 2017; Nzuki et al., 2017 (this study)). The chromosome regions bearing these SNPs have been confirmed (Wolfe et al., 2016). Thus, since the markers are now available and linked to CMD and CBSD resistance in cassava, selection will be possible at the seedling stage and this will reduce breeding time and cultivar development. These markers will also be available for 'background' selection, *i.e.* through repeated cycles of backcrossing to farmer-cultivars and applying marker assisted backcrossing. An end product which closely resembles the original farmer cultivar, but fortified with resistance genes, is envisaged to be recovered.

Conventional breeding methods are tedious and time consuming and require all the $F_1$ seeds to be grown to maturity (Xu and Crouch, 2008). Disease screening using conventional methods takes a long time, and owing to the long breeding cycle of cassava, it may take up to 10-12 years to deliver a new cultivar using conventional methods alone (Blair et al., 2007; Kawuki et al., 2016). Complimenting these efforts with molecular markers that have been identified will drastically reduce the breeding cycle and provide an accurate way of efficiently pyramiding resistance from different sources. This strategy is expected to significantly reduce population size, screening time and ultimately the cost of breeding. Thus, breeders will be able to work on a larger number of populations or seedlings per population. It is envisaged that breeders will make crosses and select at the seedling stage for a number of traits simultaneously significantly reducing the number of seedlings to advance for phenotypic trials. Researchers are optimistic that the molecular markers so identified will allow selection in the absence of disease pressure, so that pre-emptive breeding can take place in unaffected areas like West Africa and other parts of the world where CBSD is predicted to spread.

## 1.13 Conclusions

Cassava brown steak disease is the most significant biotic constraint to cassava production in Eastern, Central and Southern Africa. Significant losses have been reported in East and Central Africa and the disease spread is now evident in higher altitudes as opposed to the lower altitudes below 1000 m.a.s.l. as previously thought (Alicai et al., 2007). The disease is spreading rapidly towards West Africa where it will most likely disrupt the livelihood of millions of poor people who depend on cassava for their daily food intake (Campo et al., 2011). Farmers have been harvesting early as a strategy to avoid root necrosis thereby incurring yield losses (Kaweesi et al., 2014). Phytosanitary approaches have been used to control CBSD but the most feasible and sustainable way to combat the disease is the deployment of resistant cultivars. These cultivars will slow the spread of the disease and dramatically reduce the impact of CBSD with regard to yield loss. Resistant cultivars are likely to possess physical, biochemical, or molecular characteristics that reduce or prevent the entry and/or multiplication of the virus within the plant. A dominant gene with large effect, known as CMD2, has been widely and successfully deployed against CMD (Okogbenin et al., 2012). QTL linked to CMD and CBSD have recently been published and breeders can now take advantage of the validated markers and use them for MAS. Furthermore, cassava land landraces with promising levels of field resistance to CMD and CBSD have been identified. Different sources of CBSD field resistance in these landraces have been proposed and are being investigated (Wolfe et al., 2016; Nzuki et al., 2017).

It is important that good levels of resistance to biotic stresses are discovered, or if not available, generated through transgenic or gene editing approaches. To this end, it is important that a wide array of global germplasm is screened for biotic resistance, including germplasm from West Africa and Latin America, and including wild relatives. Once identified, desired germplasm needs to be incorporated into breeding programs, and the target resistance incorporated either through genomic selection approaches using all marker data across a genome to aid prediction, or through target approaches such as MAS.

The future of cassava breeding will heavily depend on a combination of conventional strategies and modern molecular techniques to precisely identify, introgress and pyramid QTL associated with various agronomic and farmer/consumer preferred traits. The QTL dispersed in different parental genotypes will have to be recombined and tracked using molecular tools to deliver products that meet end-user preferences.

## 1.14 References


Abarshi, M., Mohammed, I., Jeremiah, S., Legg, J., Kumar, P.L., Hillocks, R., et al. (2012). Multiplex RT-PCR assays for the simultaneous detection of both RNA and DNA viruses infecting cassava and the common occurrence of mixed infections by two cassava brown streak viruses in East Africa. *J. Virol. methods* 179**,** 176-184.

Akano, A., Dixon, A., Mba, C., Barrera, E., and Fregene, M. (2002). Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theor. Appl. Genet.* 105**,** 521-525.

Alicai, T., Ndunguru, J., Sseruwagi, P., Tairo, F., Okao-Okuja, G., Nanvubya, R., et al. (2016). Cassava brown streak virus has a rapidly evolving genome: Implications for virus speciation, variability, diagnosis and host resistance. *Scientific Reports* 6**,** 36164.

Alicai, T., Omongo, C., Maruthi, M., Hillocks, R., Baguma, Y., Kawuki, R., et al. (2007). Re-emergence of cassava brown streak disease in Uganda. *Plant. Dis.* 91**,** 24-29.

Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* 121**,** 207-221.

Ameline-Torregrosa, C., Wang, B.-B., O'Bleness, M.S., Deshpande, S., Zhu, H., Roe, B., et al. (2008). Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *J. Plant Physiol* 146**,** 5-21.

Amuge, T., Berger, D.K., Katari, M.S., Myburg, A.A., Goldman, S.L., and 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)**,** 9747.

ASARECA (2009). "Act now to stop the spread of the cassava brown streak disease", in: *The New Vision.* (Uganda: ASARECA).

Beavis, W.D. (1998). "QTL analyses: power, precision, and accuracy," in *Molecular dissection of complex traits,* ed. A.H. Paterson. (Boca Raton: CRC Press), 145-162.

Beck, B. (1982). "Historical perspectives of cassava breeding in Africa", in: *Workshop on Root Crops in Eastern Africa.* (ed.) S.K. Hahn. (Kigali,Rwanda: IDRC).

Blair, M.W., Fregene, M.A., Beebe, S.E., and Ceballos, H. (2007). "Marker-assisted selection in common beans and cassava," in *Marker-assisted selection: Current status and future perspectives in crops, livestock, forestry and fish,* ed. J.R. Guimaraes, B.D. Scherf, A. Sonnino, J.D. Dargie. (Rome: FAO), 81-115.

Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32**,** 314.

Bredeson, J.V., Lyons, J.B., Prochnik, S.E., Wu, G.A., Ha, C.M., Edsinger-Gonzales, E., et al. (2016). Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat. Biotechnol.* 34**,** 562-570.

Bubeck, D., Goodman, M., Beavis, W., and Grant, D. (1993). Quantitative trait loci controlling resistance to gray leaf spot in maize. *Crop. Sci.* 33**,** 838-847.

Campo, B.V.H., Hyman, G., and Bellotti, A. (2011). Threats to cassava production: known and potential geographic distribution of four key biotic constraints. *Food. Secur.* 3**,** 329-345.

Ceballos, H., Iglesias, C.A., Perez, J.C., and Dixon, A.G. (2004). Cassava breeding: opportunities and challenges. *Plant Mol Biol* 56(4)**,** 503-516

Ceballos, H., Okogbenin, E., Pérez, J.C., López-Valle, L.A.B., and Debouck, D. (2010). "Cassava," in *Handbook of Plant Breeding: Root and Tuber Crops,* ed. J.E. Bradshaw. Springer), 53-96.

Chalwe, A., Melis, R., Shanahan, P., and Chiona, M. (2015). Inheritance of resistance to cassava green mite and other useful agronomic traits in cassava grown in Zambia. *Euphytica.* 205**,** 103-119.

Chipeta, M.M., Bokosi, J.M., Saka, V.W., and Benesi, I.R. (2013). Combining ability and mode of gene action in cassava for resistance to cassava green mite and cassava mealy bug in Malawi. *Plant. Breed. Crop Sci.* 5**,** 195-202.

Collard, B., Jahufer, M., Brouwer, J., and Pang, E. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica.* 142**,** 169-196.

Deschamps, S., Llaca, V., and May, G.D. (2012). Genotyping-by-sequencing in plants. *J. Biol.* 1**,** 460-483.

Edwards, M., Stuber, C., and Wendel, J. (1987). Molecular-marker-facilitated investigations of quantitative-trait loci in maize. Numbers, genomic distribution and types of gene action. *Genet.* 116**,** 113-125.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6**,** e19379.

FAOSTAT (2015). "Statistical database of the Food and Agriculture Organization of the United Nations". (http://faostat.fao.org).

FAOSTAT (2017). "Statistical database of the Food and agriculture organization of the United Nations". (http://faostat.fao.org).

Ferguson, M.E., Hearne, S.J., Close, T.J., Wanamaker, S., Moskal, W.A., Town, C.D., et al. (2012). Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. *Theor. Appl. Genet.* 124**,** 685-695.

Fregene, M., Bernal, A., Duque, M., Dixon, A., and Tohme, J. (2000). AFLP analysis of African cassava (Manihot esculenta Crantz) germplasm resistant to the cassava mosaic disease (CMD). *Theor. Appl. Genet.* 100(5)**,** 678-685. doi: 10.1007/s001220051339.

Frisch, M., Bohn, M., and Melchinger, A.E. (1999). Comparison of selection strategies for marker-assisted backcrossing of a gene. *Crop. Sci.* 39**,** 1295-1301.

Gedil, M., Kumar, M., and Igwe, D. (2012). Isolation and characterization of resistant gene analogs in cassava, wild *Manihot* species, and castor bean (*Ricinus communis*). *Afri. J. Biotech.* 11**,** 15111-15123.

Goldstein, R.E., Jiménez, G., Cook, O., Gur, D., and Paroush, Z.e. (1999). Huckebein repressor activity in Drosophila terminal patterning is mediated by Groucho. *Arthropod. Struct. Dev.* 126**,** 3747-3755.

Gondwe, F. (2011). "Economic losses experienced by small-scale famers in Malawi due to cassava brown streak virus disease", (ed.) E.D. Store. (Malawi).

González, A., Schöpke, C., Taylor, N., Beachy, R., and Fauquet, C. (1998). Regeneration of transgenic cassava plants (*Manihot esculenta Crantz*) through Agrobacterium-mediated transformation of embryogenic suspension cultures. *Plant Cell Rep.* 17**,** 827-831.

Gutierrez, A.P., Yaninek, J.S., Wermelinger, B., Herren, H.R., and Ellis, C.K. (1988). Analysis of Biological Control of Cassava Pests in Africa. III. Cassava Green Mite Mononychellus tanajoa. *J. Appl. Ecol.* 25(3)**,** 941-950

Hammond-Kosack, K.E., and Kanyuka, K. (2007). *Resistance genes (R genes) in plants.* John Wiley & Sons, Ltd.

Harvey, J.J., Lewsey, M.G., Patel, K., Westwood, J., Heimstädt, S., Carr, J.P., et al. (2011). An antiviral defense role of AGO2 in plants. *PLoS ONE* 6**,** e14639.

Hayes, B., and Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genet.* 157**,** 1819-1829.

Hayes, B.J., Cogan, N.O., Pembleton, L.W., Goddard, M.E., Wang, J., Spangenberg, G.C., et al. (2013). Prospects for genomic selection in forage plant species. *Plant Breed.* 132**,** 133-143.

Heisey, P.W., and Edmeades, G.O. (1999). *CIMMYT 1997/98 world maize facts and trends. Maize production in drought-stressed environments: Technical options and research resource allocation.* Mexico: CIMMYT.

Hillocks, R. (1997). Cassava virus diseases and their control with special reference to southern Tanzania. *Pest. Sci.* 2**,** 125-138.

Hillocks, R., and Jennings, D. (2003). Cassava brown streak disease: a review of present knowledge and research needs. *Pest Sci.* 49**,** 225-234.

Hillocks, R., and Thresh, J. (2000). Cassava mosaic and cassava brown streak virus diseases in Africa: a comparative guide to symptoms and aetiologies. *Roots* 7**,** 1-8.

Hillocks, R.J., Raya, M.D., Mtunda, K., and Kiozia, H. (2001). Effects of brown streak virus disease on yield and quality of cassava in Tanzania. *J. Phyt.* 149(7-8)**,** 389-394.

Hui, L. (2012). Potential geographic distribution of the cassava green mite Mononychellus tanajoa in Hainan, China. *Afri. J.Agri.Res.* 7(7).

ICGMC (2015). High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta* Crantz) from 10 populations. *G3* 5**,** 133-144.

IITA (2003). *Preemptive management of the virulent cassava mosaic disease through an integrated cassava development approach for enhanced rural sector in the south-south and south-east zones of Nigeria.* Ibadan, Nigeria.

Jarvie, T., and Harkins, T. (2007). Small RNA analysis using the Genome Sequencer™ FLX system. *Nat. Biotechnol.* 200**,** 7.

Jehan, T., and Lakhanpaul, S. (2006). Single nucleotide polymorphism (SNP)-methods and applications in plant genetics: a review. *Indian J. Biotechnol.* 5**,** 435.

Jennings, D. (1957). Further studies in breeding cassava for virus resistance. *Afr. J. Agric. Res.* 22**,** 213-219.

Jennings, D. (1994). Breeding for resistance to African cassava mosaic geminivirus in East Africa. *Trop. Sci.* 34**,** 110-122.

Jennings, D.L. (1963). Variation in pollen and ovule fertility in varieties of cassava, and the effect of interspecific crossing on fertility. *Euphytica***,** 69–76.

Kanju, E., Mtunda, K., Muhanna, M., Raya, M., and Mahungu, N. (2002). "Management of Cassava brown streak virus disease in Tanzania", in: *Cassava Brown Streak Virus Disease: Past, Present and Future.* (ed.) J. Hillocks. (Mombasa, Kenya).

Katiyar-Agarwal, S., Gao, S., Vivian-Smith, A., and Jin, H. (2007). A novel class of bacteria-induced small RNAs in Arabidopsis. *Genes Dev.* 21**,** 3123-3134.

Kaweesi, T., Kawuki, R., Kyaligonza, V., Baguma, Y., Tusiime, G., and Ferguson, M.E. (2014). Field evaluation of selected cassava genotypes for cassava brown streak disease based on symptom expression and virus load. *J. Virol.* 11**,** 216.

Kawuki, R., Nuwamanya, E., Labuschagne, M., Herselman, L. Ferguson 2009). "Genetic effects of inbreeding on harvest index and root dry matter content in cassava", in: *Second RUFORUM Biennial Regional Conference on "Building capacity for food security in Africa"*, ed. E. Adipala, Tusiime, G. & Majaliwa, J.G.M: RUFORUM), 377–381.

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

Kawuki, R.S., Kaweesi, T., Esuma, W., Pariyo, A., Kayondo, I.S., Ozimati, A., et al. (2016). Eleven years of breeding efforts to combat cassava brown streak disease. *Breed. Sci.* 66**,** 560-571.

Kayondo, S.I., Pino Del Carpio, D., Lozano, R., Ozimati, A., Wolfe, M., Baguma, Y., et al. (2018). Genome-wide association mapping and genomic prediction for CBSD resistance in Manihot esculenta. *Scientific Reports* 8(1)**,** 1549.

Kearsey, M., and Farquhar, A. (1998). QTL analysis in plants; where are we now? *Heredity* 80**,** 137-142.

Khan, A.S., Ali, S., and Khan, I.A. (2015). Morphological and molecular characterization and evaluation of mango germplasm: An overview. *Scientia Horticulturae* 194**,** 353-366.

Kim, J., Lim, C.J., Lee, B.-W., Choi, J.-P., Oh, S.-K., Ahmad, R., et al. (2012). A genome-wide comparison of NB-LRR type of resistance gene analogs (RGA) in the plant kingdom. *Mol. Cells* 33**,** 385-392.

Kunkeaw, S., Tangphatsornruang, S., Smith, D.R., and Triwitayakorn, K. (2010). Genetic linkage map of cassava (*Manihot esculenta* Crantz) based on AFLP and SSR markers. *Plant. Breed.* 129**,** 112-115. doi: 10.1111/j.1439-0523.2009.01623.x.

Lebot, V. (2009). *Tropical root and tuber crops: cassava, sweet potato, yams and aroids.* Cambridge, USA: CABI.

Legg, J., Ndalahwa, M., Yabeja, J., Ndyetabula, I., Bouwmeester, H., Shirima, R., et al. (2017). Community phytosanitation to manage cassava brown streak disease. *Virus Research* 241**,** 236-253. doi: https://doi.org/10.1016/j.virusres.2017.04.020.

Legg, J.P., Asiimwe, P., Sseruwagi, P., and Brown, J.K. (2008). Changes in the biology of *Bemisia tabaci* on cassava in Africa and their impact on virus disease pandemics. *J. Insect Sci* 8**,** 27-28.

Legg, J.P., and Hillocks, R.J. (2003). "Cassava Brown Streak Virus Disease: Past, Present and Future," in: *Proceedings of an International Workshop.* (ed.) J. Hillocks. (Mombasa, Kenya).

Legg, J.P., Jeremiah, S., Obiero, H., Maruthi, M., Ndyetabula, I., Okao-Okuja, G., et al. (2011). Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Res.* 159**,** 161-170.

Li, H.-Q., Sautter, C., Potrykus, I., and Puonti-Kaerlas, J. (1996). Genetic transformation of cassava (*Manihot esculenta Crantz*). *Nat. Biotechnol.* 14**,** 736-740.

Lister, R. (1959). Mechanical transmission of cassava brown streak virus. *Nature*.

Lokko, Y., Danquah, E., Offei, S., Dixon, A.G., and Gedil, M. (2005). Molecular markers associated with a new source of resistance to the cassava mosaic disease. *Afri. J. Biotechnol.* 4**,** 873-881.

Macea Choperena, E.d.P., Ospina, C., Fregene, M., Montoya-Lerma, J., and Bellotti, A.C. (2012). Microsatellite identification associated with resistance to the mite *Mononychellus tanajoa (Acari: Tetranychidae)*. *Rev. Bras. Entomol.* 38**,** 70-75.

Mackay, T.F.C., Stone, E.A., and Ayroles, J.F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* 10(8)**,** 565-577

Maruthi, M., Hillocks, R., Mtunda, K., Raya, M., Muhanna, M., Kiozia, H., et al. (2005). Transmission of Cassava brown streak virus by *Bemisia tabaci* (*Gennadius*). *J. Phytopathol.* 153**,** 307-312.

Maruthi, M.N., Jeremiah, S.C., Mohammed, I.U., and Legg, J.P. (2017). The role of the whitefly, Bemisia tabaci (Gennadius), and farmer practices in the spread of cassava brown streak ipomoviruses. *Journal of Phytopathology* 165(11-12)**,** 707-717

Masumba, E.A., Kapinga, F., Mkamilo, G., Salum, K., Kulembeka, H., Rounsley, S., et al. (2017). QTL associated with resistance to cassava brown streak and cassava mosaic diseases in a bi-parental cross of two Tanzanian farmer varieties, Namikonga and Albert. *Theor. Appl. Genet.* 130:2069-2090

Mbanzibwa, D., Tian, Y., Tugume, A., Mukasa, S., Tairo, F., Kyamanywa, S., et al. (2009). Genetically distinct strains of Cassava brown streak virus in the Lake Victoria basin and the Indian Ocean coastal area of East Africa. *Arch. Virol.* 154**,** 353-359.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W. (2003). Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. *Plant Cell.* 15**,** 809-834.

Mohan, C., Shanmugasundaram, P., Maheswaran, M., Senthil, N., Raghu, D., and Unnikrishnan, M. (2013). Mapping New Genetic Markers Associated with CMD Resistance in Cassava (*Manihot esculenta Crantz*) Using Simple Sequence Repeat Markers. *J. Agric. Sci.* 5(5)..

Moroldo, M., Paillard, S., Marconi, R., Fabrice, L., Canaguier, A., Cruaud, C., et al. (2008). A physical map of the heterozygous grapevine'*Cabernet Sauvignon*'allows mapping candidate genes for disease resistance. *BMC Plant Biol.* 8**,** 66.

Muhammad Shah Nawaz-ul-Rehman, S.M., Asif Ali Khan,Yusuf Zafar,Rob W. Briddon ( 2007). RNAi-mediated male sterility of tobacco by silencing TA29. *Mol. Biotechnol.***,** 159–165.

Muhanna, M., and Mtunda, K. (2003). "Cassava Brown Streak Virus Disease Transmission Studies in Tanzania", in: *Cassava Brown Streak Virus Disease: Past, Present and Future.* (ed.) J. Hillocks. (Mombasa, Kenya).

Mulimbi, W., Phemba, X., Assumani, B., Kasereka, P., Muyisa, S., Ugentho, H., et al. (2012). First report of Ugandan cassava brown streak virus on cassava in Democratic Republic of Congo. *New Dis. Rep.* 26**,** 11.

Mullis, K.B., and Faloona, F.A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Meth. Enzymol.* 155**,** 335-350.

Munganyinka, E., Ateka, E.M., Kihurani, A.W., Kanyange, M.C., Tairo, F., Sseruwagi, P., et al. (2018). Cassava brown streak disease in Rwanda, the associated viruses and disease phenotypes. *Plant Pathology* 67(2)**,** 377-387

Nakaya, A., and Isobe, S.N. (2012). Will genomic selection be a practical method for plant breeding? *Ann. Bot.***,** 1303-1316.

Namuth-Covert, H.L.M.D. (2011). Introduction to Forward Selection (a.k.a. Foreground Selection. *Mol. Breed.***,** 1-5.

Neuenschwander, P., Borgemeister, C., and Langewald, J. (2003). *Biological Control in IPM Systems in Africa.* CABI.

Nichols, R. (1947). Breeding cassava for virus resistance. *Afr. J. Agric. Res.* 12**,** 184-194.

Nichols, R.F.W. (1950). The Brown Streak Disease of Cassava. *Afr. J. Agric. Res.* 15(3)**,** 154-160

Nweke, F., Nweke, F.I., Lynam, J.K., and Spencer, D.S. (2002). *The cassava transformation: Africa's best-kept secret.* USA: Michigan State University Press.

Nzuki, I., Katari, M.S., Bredeson, J.V., Masumba, E., Kapinga, F., Salum, K., et al. (2017). QTL Mapping for Pest and Disease Resistance in Cassava and Coincidence of Some QTL with Introgression Regions Derived from Manihot glaziovii. *Front Plant Sci.* 8(1168).

Okogbenin, E., Egesi, C., Olasanmi, B., Ogundapo, O., Kahya, S., Hurtado, P., et al. (2012). Molecular marker analysis and validation of resistance to cassava mosaic disease in elite cassava genotypes in Nigeria. *Crop Sci.* 52**,** 2576-2586.

Okogbenin, E., and Fregene, M. (2002). Genetic analysis and QTL mapping of early root bulking in an F1 population of non-inbred parents in cassava (*Manihot esculenta* Crantz). *Theor. Appl. Genet.* 106**,** 58-66.

Okogbenin, E., Marin, J., and Fregene, M. (2008). QTL analysis for early yield in a pseudo F2 population of cassava. *Afr. J. Biotechnol.* 7**,** 2.

Okogbenin, E., Porto, M., Egesi, C., Mba, C., Espinosa, E., Santos, L., et al. (2007). Marker-assisted introgression of resistance to cassava mosaic disease into Latin American germplasm for the genetic improvement of cassava in Africa. *Crop Sci.* 47**,** 1895-1904.

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

Onzo, A., Hanna, R., Negloh, K., Toko, M., and Sabelis, M.W. (2005). Biological control of cassava green mite with exotic and indigenous phytoseiid predators—Effects of intraguild predation and supplementary food. *Biol. Control.* 33(2)**,** 143-152.

Papa, R., Bellucci, E., Rossi, M., Leonardi, S., Rau, D., Gepts, P., et al. (2007). Tagging the signatures of domestication in common bean (*Phaseolus vulgaris*) by means of pooled DNA samples. *Ann. Bot.* 100**,** 1039-1051.

Patil, B.L., Legg, J.P., Kanju, E., and Fauquet, C.M. (2015). Cassava brown streak disease: a threat to food security in Africa. *Journal of General Virology* 96(5)**,** 956-968.

Poland, J.A., and Rife, T.W. (2012). Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome.* 5**,** 92-102.

Porter, B.W., Paidi, M., Ming, R., Alam, M., Nishijima, W.T., and Zhu, Y.J. (2009). Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family. *Mol. Genet. Genomics.* 281**,** 609-626.

Powell, W., Machray, G.C., and Provan, J. (1996). Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.* 1**,** 215-222.

Prasanna, B. (2005). QTL mapping and its applications in crop plants. *Indian Agr. Res. Inst. PUSA, New Delhi* 122.

Price, D.R.G., and Gatehouse, J.A. (2008). RNAi-mediated crop protection against insects. *Trends in Biotechnology* 26(7)**,** 393-400.

Prochnik, S., Marri, P.R., Desany, B., Rabinowicz, P.D., Kodira, C., Mohiuddin, M., et al. (2012). The cassava genome: current progress, future directions. *Trop. Plant Biol.* 5**,** 88-94.

Rabbi, I.Y., Hamblin, M.T., Kumar, P.L., Gedil, M.A., Ikpan, A.S., Jannink, J.-L., et al. (2014). High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. *Virus Res.* 186**,** 87-96.

Rabbi, I.Y., Kulembeka, H.P., Masumba, E., Marri, P.R., and Ferguson, M. (2012). An EST-derived SNP and SSR genetic linkage map of cassava (*Manihot esculenta* Crantz). *Theor. Appl.Genet.* 125**,** 329-342.

Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol.* 5**,** 94-100.

Rana, M., Sood, A., Hussain, W., Kaldate, R., Sharma, T.R., Gill, R.K., et al. (2019). "Chapter 6 - Gene Pyramiding and Multiple Character Breeding," in *Lentils,* ed. M. Singh. Academic Press), 83-124.

Ribaut, J.-M., and Hoisington, D. (1998). Marker-assisted selection: new tools and strategies. *Trends Plant Sci.* 3**,** 236-239.

Ribeiro, P.F., Akromah, R., and Manu-Aduening, J. (2012). Using marker assisted selection to hasten screening of cassava cultivars developed through introgression of Cassava Mosaic Disease (CMD) resistance into cassava landraces in Ghana. *Agri. Sci. Technol.* 2**,** 1.

Roth, B.M., Pruss, G.J., and Vance, V.B. (2004). Plant viral suppressors of RNA silencing. *Virus Res.* 102**,** 97-108.

Schön, C.C., Utz, H.F., Groh, S., Truberg, B., Openshaw, S., and Melchinger, A.E. (2004). Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genet.* 167**,** 485-498.

Semagn, K., Bjørnstad, Å., and Ndjiondjop, M. (2006). Principles, requirements and prospects of genetic mapping in plants. *Afr. J. Biotechnol.* 5**,** 2569-2587.

Skovgård, H., Tomkiewicz, J., Nachman, G., and Münster-Swendsen, M. (1993). The effect of the cassava green mite Mononychellus tanajoa on the growth and yield of cassava Manihot esculenta in a seasonally dry area in Kenya. *Exp. Appl. Acarol.* 17(1)**,** 41-58. doi: 10.1007/bf00156943.

Smith, S.M. (2013). *Application of Genome Reduction, Next Generation Sequencing, and KASPar Genotyping in Development, Characterization, and Linkage Mapping of Single Nucleotide Polymorphisms in the Grain Amaranths and Quinoa.* Msc, Brigham Young University,Provo.

Storey, H. (1936). Virus diseases of East African plants. VI-A progress report on studies of the disease of cassava. *East Afri. J. Agri.* 2**,** 34-39.

Sunkar, R., and Zhu, J.-K. (2004). Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. *Plant Cell.* 16**,** 2001-2019.

Tanksley, S.D. (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature.* 335**,** 6170.

Tanksley, S.D. (1993). Mapping polygenes. *Annu. Rev. Genet.* 27**,** 205-233.

Thresh, J. (2003). "Brief History of Cassava Brown Streak Virus Disease", in: *Cassava Brown Streak Virus Disease: Past, Present, and Future.* (ed.) J. Hillocks. (Mombasa, Kenya).

Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *J. Sci. Food Agric.* 313**,** 1596-1604.

United Nations, D.o.E.a.S.A., Population Division (2017). World Population Prospects: The 2017 Revision, Key Findings and Advance Tables. Working Paper No. ESA/P/WP/248.

Vanderschuren, H., Moreno, I., Anjanappa, R.B., Zainuddin, I.M., and Gruissem, W. (2012). Exploiting the combination of natural and genetically engineered resistance to cassava mosaic and cassava brown streak viruses impacting cassava production in Africa. *PLoS ONE* 7**,** e45277.

Vignal, A., Milan, D., SanCristobal, M., and Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.* 34**,** 1.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23**,** 4407-4414.

Wang, B., Jiang, L., Zhang, Y., Zhang, W., Wang, M., Cheng, X., et al. (2011). QTL mapping for resistance to strip virus disease in rice. *Plant. Breed.* 130**,** 321-327.

Wang, G.-L., Mackill, D.J., Bonman, J.M., McCouch, S.R., Champoux, M.C., and Nelson, R.J. (1994). RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. *Genet.* 136**,** 1421-1434.

Weising, K., Nybom, H., Pfenninger, M., Wolff, K., and Kahl, G. (2005). *DNA fingerprinting in plants: principles, methods, and applications.* CRC press.

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

Wolfe, M.D., Rabbi, I.Y., Egesi, C., Hamblin, M., Kawuki, R., Kulakow, P., et al. (2016). Genome-Wide association and prediction reveals genetic architecture of cassava mosaic disease resistance and prospects for rapid genetic improvement. *Plant Genome.* 9.

Xu, S. (2003). Theoretical basis of the Beavis effect. *Genet.* 165**,** 2259-2268.

Xu, Y., and Crouch, J.H. (2008). Marker-assisted selection in plant breeding: from publications to practice. *Crop. Sci.* 48**,** 391-407.

Zhou, W.C., Kolb, F., Bai, G.H., Domier, L., Boze, L., and Smith, N. (2003). Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat. *Plant Breed.* 122**,** 40-46.

# Chapter 2

# Genetic linkage mapping in a bi-parental cassava cross between Kiroba and AR37-80


**Authors**

Inosters Nzuki[1], Jessen V. Bredeson[2], Jessica B. Lyons[2], Daniel S. Rokhsar[2], Geoffrey S. Mkamilo[3], Fortunus Kapinga[3], Esther Masumba[4], Kasele Salum[5,] Heneliko Kulembeka[5], Alexander A. Myburg[1], Nicholas van der Merwe[1] and Morag E. Ferguson*[6]

**Affiliations**

[1]Department of Genetics, University of Pretoria, Private bag X20, Hatfield 0028, South Africa

[2]Molecular and Cell Biology Department, University of California, Berkeley, California, USA

[3]Naliendele Agricultural Research Institute, P. O. Box 509, Mtwara, Tanzania.

[4]Sugarcane Research Institute, P. O. Box 30031, Kibaha, Tanzania

[5]Ukiriguru Agricultural Research Institute, P. O. Box 1433, Mwanza, Tanzania

[6]International Institute of Tropical Agriculture (IITA), P.O. Box 30709-00100, Nairobi, Kenya

* Corresponding author: m.ferguson@cgiar.org, (+254) 733524685

## Author contributions

The data obtained in this chapter are part of the data used to construct the consensus map for cassava published in the paper "High-Resolution Linkage Map and Chromosome-Scale Genome Assembly for Cassava (*Manihot esculenta Crantz*) from 10 Populations" by the International Cassava Genetic Map Consortium (ICGMC) published in 2014 (DO1 10.1524/g3.114.05008 ). The genetic linkage map developed in this chapter has been published in the journal Frontiers in Plant Sciences under the title 'QTL mapping for pest and disease resistance in cassava and coincidence of some QTL with introgression regions derived from *Manihot glaziovii*" published online in July 2017 (Front. Plant Sci. 8:1168. doi: 10.3389/fpls.2017.01). Inosters Nzuki conducted the study analyzed the data and wrote the manuscript. Morag Ferguson conceived the project and sourced funding for the study. Geoffrey Mkamilo facilitated logistical support for the implementation of the project in Tanzania. Esther Masumba, Benendetta Kimata, Fortunus Kapinga, Kasele salum and Heneriko Kulembeka provided support in the development of the mapping population, establishment of the phenotypic trials and data collection. Jessica Lyons provided laboratory support for GBS, Jensen Bredeson called the SNPs for GBS together with Dan Rokhsar. Prof. Alexander A. Myburg, Dr. Morag Ferguson and Dr. Nicolas van der Merwe provided scientific guidance, advice and editorial support for the manuscript. All authors read and approved the manuscript

**2**

## 2.1 Abstract

An $F_1$ bi-parental mapping population was developed by crossing two distantly related cassava cultivars Kiroba as the female parent and AR37-80 as the male parent. The integrity of the mapping population was assessed initially using SSR genotyping and subsequently using SNPs generated through genotyping-by-sequencing (GBS). Unexpected alleles or allelic combinations from SSR analysis identified 14 admixture $F_1$, presumably having received pollen from a source other than the donor parent during crossing, and one self. A total of 265 progeny withthe expected allelic composition were classified as true. Further identity-by-descent (IBD) analysis on SNP data revealed 90 admixture $F_1$. A genetic linkage map was derived from 175 $F_1$ progeny and 1,974 SNP markers and spanned 18 chromosomes covering a distance of 1,698 cM with a marker density of 0.86 per cM. The number of markers ranged between 31 to 201 per linkage group and the length of the groups ranged from 12.4 to 141.86 cM. A total of 549 markers showed segregation distortion, accounting for 12% of the original 4422 markers and 1548 (35%) sharing segregation pattern with at least one other marker. Some linkage groups contained substantial gaps (maximum of 1,264,551bp (21.4 cM) on chromosome XIV) that required additional markers to close the gaps and further saturate the map. The biparental map obtained in this study contributed to the development of a high resolution linkage map of cassava that helped to anchor chromosome-genome sequence of v4.1 cassava genome.

**Key words:** Genetic linkage map, cassava-brown streak disease, Kiroba, biparental mapping

## 2.2 Introduction

A genetic linkage map gives the order and relative position of genetic markers in the genome of a species or experimental population. Genetic linkage maps have been used for a cultivar of purposes and are fundamental to identifying chromosomal regions that contain genes controlling quantitative traits through QTL analysis. Linkage maps also provide a scaffold for assembling physical maps, facilitate positional cloning of important genes and are important tools for functional genomic studies (Wang et al., 2011). Genetic linkage maps have been constructed for almost all the important crop species. A genetic linkage map of tomato was the first to be developed using RFLPs (Bernatzky and Tanksley, 1986). Such a map was also the first to be used to resolve quantitative traits into discrete Mendelian factors (Tanksley, 1988). With the advent of PCR-based genetic markers many more maps were published, for a larger number of species, and at increasingly high density, for example a genetic linkage map of potato consisting of 10,000 markers was published (Van Os et al., 2006). In cassava, the first genetic linkage map derived from an F1 cross was mainly based on RFLP markers and comprised 132 RFLPs, 30 RAPDs, three microsatellites and three isoenzymes (Fregene et al., 1997). The map comprised of 20 linkage groups spanning 931.6cM with an average marker density of one per 7.9cM. An SSR based genetic linkage map of cassava was later developed from an F2 population (Okogbenin et al., 2012). The map has 100 SSR markers spanning 1236.7cM distributed across 22 linkage groups with an average marker distance of 17.92cM. This was the first SSR based linkage map of cassava. More recently, and with the advance in genomics, a SNP based genetic linkage map in cassava anchored in SSRs was published (Rabbi et al., 2012). The map comprised of 568 markers (434 SNPs and 134 SSRs) spanning 1837cM, distributed across 19 linkage groups and with a marker distance of 3.4cM. Recently, a high resolution consensus linkage map of cassava was published (ICGMC, 2015). The consensus map is derived from ten mapping populations (including the population described here) and has 22,400 SNPs anchored in chromosomes and scaffolds of cassava v4.1 and spans a distance of 2412cM. It is composed of 18 linkage groups consistent with the haploid chromosome number of cassava (2n=36) (Nasser, 2000).

SNP markers are now the most popular marker class for creating dense genetic linkage maps, diversity and genome wide association studies overtaking microsatellites as the markers of choice (Kawuki et al., 2013). Although microsatellites are co-dominant, relatively abundant in genomes with high levels of polymorphism, SNPs are more abundant in genomes and can be

genotyped cost effectively. Low cost per data point was made possible by the advent of next generation sequencing platforms and associated plummeting cost of sequencing. The GoldenGate genotyping assay utilizing the power of allele specific oligos for each target SNP and binding of activated DNA to magnetic particles was among the first to provide a robust and a flexible platform for SNP genotyping (Bibikova and Fan, 2009). KASPar from LGC (www.lgcgroup.com) is a low density SNP genotyping platform based on competitive allele specific PCR (Semagn et al., 2014). It uses two FRET cassettes with HEX or FAM fluorophores which are conjugated to a primer and quenched by resonance transfer. DNA amplification is done using allele specific primers and the dyes separate with the quencher when the FRET cassette primer is hybridized with the DNA. The GoldenGate system has been used in cassava and some markers validated and converted to KASpar assay (Rabbi et al., 2012). The SNaPshot® Multiplex System from life technologies is a primer extension-based method developed for the analysis of SNPs (Applied Biosystems). It is a SNP genotyping by fragment analysis method with a capability of multiplexing up to 10 SNPs in a single reaction. It has the ability to use unlabeled user-defined primers to incorporate SNPS of interest cost effectively. Genotyping by sequencing (GBS) is the most recent SNP genotyping technology and is based on reducing the complexity of the genome with restriction enzymes (Elshire et al., 2011). The GBS method is based on sequencing from restriction sites, an approach similar to Restriction site Associated DNA Sequencing (RAD-Seq) (Baird et al., 2008). In GBS, the DNA is digested using ApeK1(though it can be a range of different enzymes with optimization steps in the beginning to see which are the best enzymes to use and they differ between species) and Y-shaped adaptors are ligated onto the restriction sites with specially designed barcodes to allow multiplexing and reduction of reagent costs. It is a simple and robust approach that is gaining momentum in many genomics and molecular breeding applications, including for diversity studies and linkage mapping. At present GBS is the best approach for SNP based saturated linkage map construction and subsequent QTL mapping of important agronomic traits in crops.

Cassava is a challenging crop to breed owing to its heterozygous nature, long growing cycle and low seed yield per pollination limiting production of adequately sized $F_2$ populations (Kunkeaw et al., 2010). Moreover, it is technically challenging to construct a linkage map of cassava from an $F_2$ population since separate analysis for gametes segregating in male and female parents is required. This generally limits genetic linkage mapping in cassava to an $F_1$ population.

The objective of this study was to develop an $F_1$ mapping population derived from a cross between Kiroba (female parent) and AR37-80 (male parent), to genotype the $F_1$ progeny by GBS and construct a SNP based genetic linkage map for the cross.

## 2.3 Materials and Methods

### 2.3.1 Establishment of crossing blocks and development of mapping population

The mapping population was formed by crossing two parents with contrasting responses to CBSD; 'Kiroba' (female) (Mahungu, 2004) and AR37-80 (male) (Figure2.1). 'Kiroba' is a local landrace from Tanzania which is largely male sterile, gets infected by cassava brown streak viruses (CBSVs) shows mild CBSD leaf symptoms with almost no root necrosis even under very high disease pressure and is regarded as CBSD tolerant by breeders (Kaweesi et al., 2014). In contrast, 'Kiroba is very susceptible to CMD. It is high yielding with high dry matter content of over 30% (Muhanna and Mtunda, 2003). AR37-80 is a CIAT improved cultivar being a cross between a CMD resistant line (C33) from IITA and CW259-42 which is a backcross of MTAI 8 (Rayon 60) and an interspecific cross between *M. flabellifollia* and CM 2766-5. It was developed through MAS, being positively selected for markers for the CMD2 resistance locus and markers for CGM resistance. It is resistant to CMD and CGM but susceptible to CBSD (Blair et al., 2007; Okogbenin et al., 2012).

Pairwise crossing blocks for 'Kiroba' and AR37-80 parents were established at the Sugarcane Research Institute, Kibaha, Tanzania. At flowering (approximately 6 months after planting (MAP)) the female flowers from Kiroba were covered with pollination bags early in the morning (8-10am). The male flowers from AR37-80 with mature pollen were tagged at around 11am, collected and hand pollination done around midday. The pollinated flowers were tagged and after four weeks, the fertilized fruits were covered to avoid shattering. After 8 weeks, the seeds in the bags containing the tags were collected and stored for one month to break dormancy. The seeds were planted in trays and kept in a screen house. After one month the germinated seedlings were transplanted and maintained in a low CMD/CBSD pressure area of Makutopora (5°58'36.87"S, 35°46'00.00"E) in Tanzania in 2011. Stem cuttings were used to establish phenotyping trials in two coastal sites in Tanzania where CBSD disease pressure is high; Chambezi (6°33'19"S 38°44'51"E) and Naliendele (0°23'00.60"S 40°09'50.58''E). The

mapping population was maintained and used as planting material for the establishment of phenotyping trials.


**Figure 2.1** Parents used in the development of the mapping population (a) 'Kiroba', female parent and (b) AR37-80, male parent

### 2.3.2 Genomic DNA isolation and validation of the mapping population using SSR markers

Genomic DNA was extracted from fresh young leaf tissues harvested from the $F_1$ plants following a modified method of Dellaporta (Dellaporta et al., 1983). The leaf tissues were freeze dried in liquid nitrogen and then ground using a Genogrinder beadmill at 150 strokes/min for 2 minutes. The $F_1$ progenies were first screened using 12 SSR markers (Kawuki et al., 2013) that were identified as being polymorphic between the two parents to detect true crosses, admixtures and self's. The admixtures and self's were removed prior to GBS analysis. The PCR reaction conditions used were: 50 ng DNA; 0.8 pmole (F and R) primer; 0.375U Taq polymerase; 1.5 mM $MgCl_2$; 0.2 mM dNTPs, 1x PCR buffer (50 mM KCl; 10 mM Tris-HCl (pH 8.3)). The reaction was done in a volume of 10 µl and the PCR profile consisted of 5 min initial template denaturation at 95°C, 35 cycles of 30 seconds at 95°C, 2 min at the primer annealing temperature and 30 seconds at 72°C. This was followed by a final primer extension

of 72°C for 30 minutes. The PCR fragments were electrophoresed on a capillary sequencer, the ABI3730 (Applied Biosystems). Data capture was done using the Genscan®software (Applied Biosystems) and the resulting fragments analyzed and scored using the Genemapper®software ver 4.1 (Applied Biosystems).

### 2.3.3 Genotyping by sequencing (GBS)

GBS (Elshire et al., 2011) was performed on the confirmed progeny at the University of California, Berkeley (ICGMC, 2015). ApeK1 restriction enzyme (New England Biolabs) was used to digest the DNA as a way to reduce the complexity of the genome. The barcode sequences included in the Elshire et al (2012) manuscript were used together with Y-shaped Illumina adapters. Pair-end sequencing was performed using the Hiseq 2500 (Illumina Inc.). The SNPs were called using GATK (DePristo et al., 2011) against v5.1 of the cassava genome assembly and filtered using custom scripts. SNPs were named according to chromosome number (Roman for v5.1) and the base pair. Loci that deviated from the expected Mendelian segregation ratios based on goodness of fit ($P < 0.05$) were excluded from the analysis. Clustering and principle analysis was performed in R using mclust. The samples were grouped based on identity by state (IBS) calculated using Bayesian Information Criterion (BIC). The clustering, density estimation and discriminant analysis was performed using package (mlcustD).

### 2.3.4 Construction of genetic linkage maps

A genetic linkage map was calculated based on 1,974 SNP markers using the CP option of JoinMap® ver 4.1 (Van Ooijen, 2006) which is appropriate for outcrossing species. A one-step map approach is the most common strategy for linkage mapping in cassava (Kunkeaw et al., 2010; Sraphet et al., 2011; Chen et al., 2012) and was used for this study. Goodness of fit between observed and expected segregation ratios was tested by Chi square. Loci showing segregation distortion ($\chi^2 > 6.0$) were excluded together with identical (redundant) loci before calculating the linkage groups. Markers were grouped using regression method at a minimum LOD threshold of 5. The recombination frequencies were converted into map distances (centimorgan) using the Kosambi mapping function. The position of the markers in each

linkage group was obtained by considering their contribution to the average goodness of fit (mean Chi square) and the nearest neighbour fit (N.N Fit) value. Based on these criterion, markers were removed or added to each linkage group and calculations redone until the best fit and order was obtained. Initially a high density genetic linkage map was calculated then pruned to obtain a framework map with markers spaced at every 5-10cM with an order consistent with v 5.1 of the cassava genome sequence (Goodstein et al., 2012). The level of confidence of the position of markers on the framework map is higher than that of the high-density map.

## 2.4 Results

### 2.4.1 Development and validation of the mapping population

Controlled crossing through 1,116 hand pollinations between Kiroba and AR37-80 resulted in 2,676 seeds. Twelve hundred and sixty seeds were sown and 445 of them germinated in a screen house, a 35% germination rate. Three hundred $F_1$ seedlings were transplanted in a CMD/CBSD free area in Makutupora, Tanzania and only 280 of them survived in the field due to the hot and dry conditions. The mapping population was validated by SSR screening to identify selfs, off-types and true crosses (Figure 2.2). The $F_1$ with two identical alleles from one parent were regarded as selfs and those with unexpected segregation or alleles, presumably having received pollen from elsewhere were regarded as off-types. Those with expected segregation were classified as true crosses. Fourteen off-types, one self and 265 true crosses were identified.

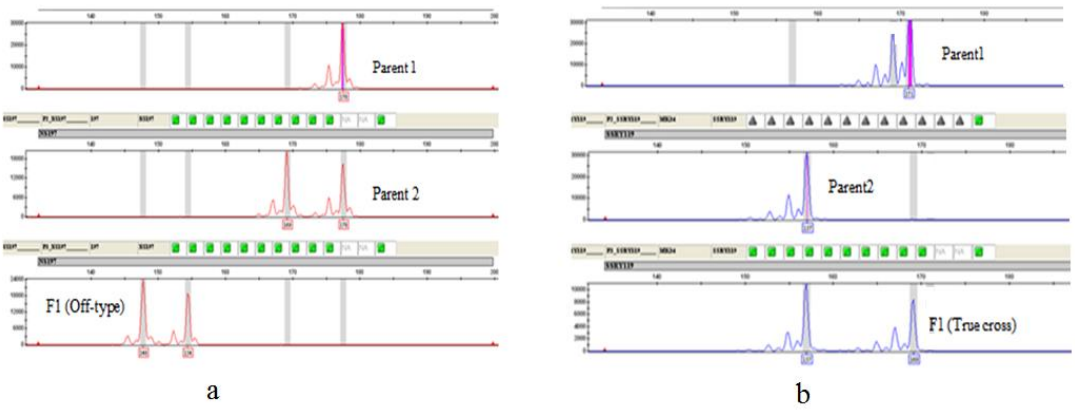



**Figure 2.2** Identification of off-types (a) and true crosses (b) in the mapping population by use of SSR markers. The off-types contain alleles different from those from the parents, or in different segregation patterns to those expected from the parents. True crosses contain the expected alleles and segregation patterns from the parents.

The integrity of the mapping population was further tested by cluster analysis based on the SNP data by splitting it into principal components P1 and P2 (Figure 2.3). This revealed the presence of patterns of relatedness, unrelatedness and some uncertainties of some of the $F_1$. Some $F_1$ are assigned to different populations as they are probably admixtures, having received pollen from another source rather than from the donor parent during crossing which is possible in an outcrossing crop like cassava.

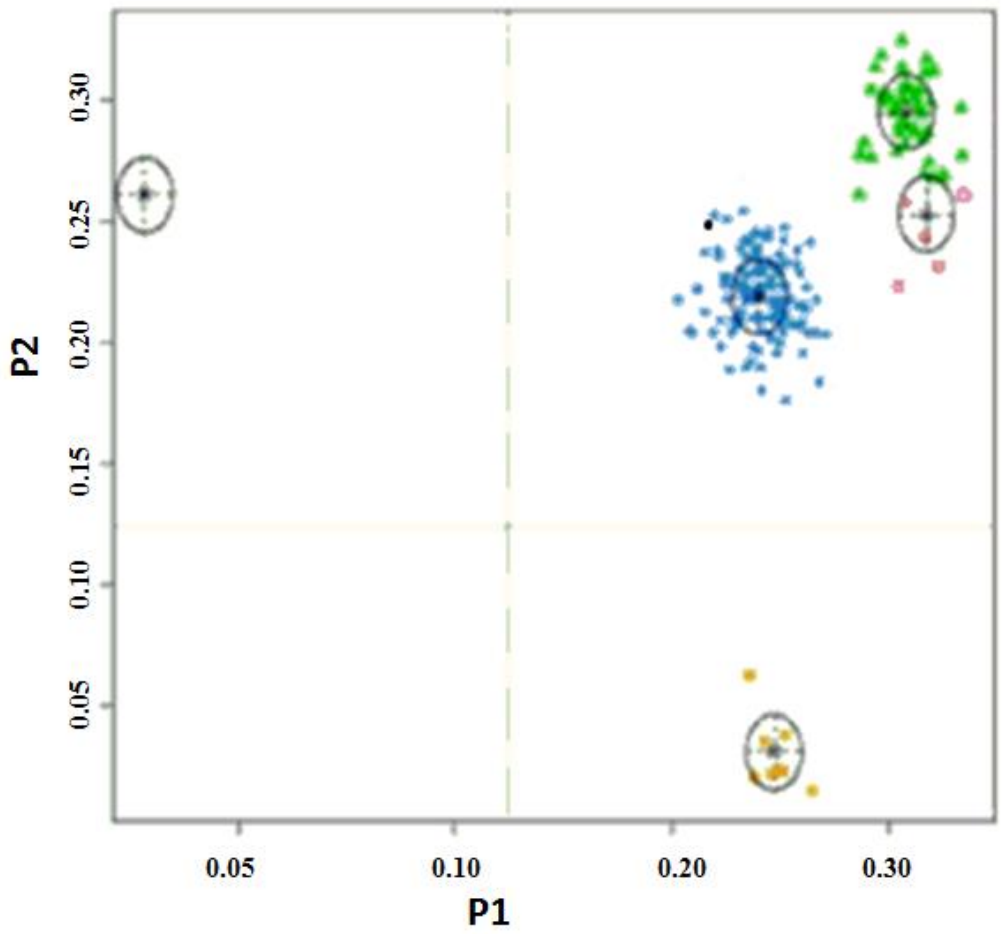



**Figure 2.3** Characterization of the mapping population showing the clustering of the individuals in the mapping population. The blue cluster represents the true $F_1$, the cluster in yellow near X, Y =00, 0.25 represents parental genotypes while the rest of the groupings in green and pink colour represents off-types/selfs. P1 and P2 represent principle component 1 and 2.

Identity by state plot (IBS) done using Mclust is shown in Figure 2.4. It shows relatedness (or unrelatedness) by how and where the points (each representing a pair-wise comparison between two individuals in the dataset) cluster on the plot. All comparisons are color-coded by their comparison type: Blue = parent-parent comparison, red = parent-offspring comparison and black = nominal offspring-offspring comparison. On the X-axis is the measure of the first-degree relatedness by considering cases where an individual shares a single allele with another individual (IBS1 and IBS2*), and on the Y-axis is the measure of 0th-degree relatedness or identity (two shared alleles) between individuals (IBS2). Though a het-het comparison is really

a first-degree relationship, we are more confident that individuals sharing IBS2* are related than individuals sharing IBS1 genotypes.

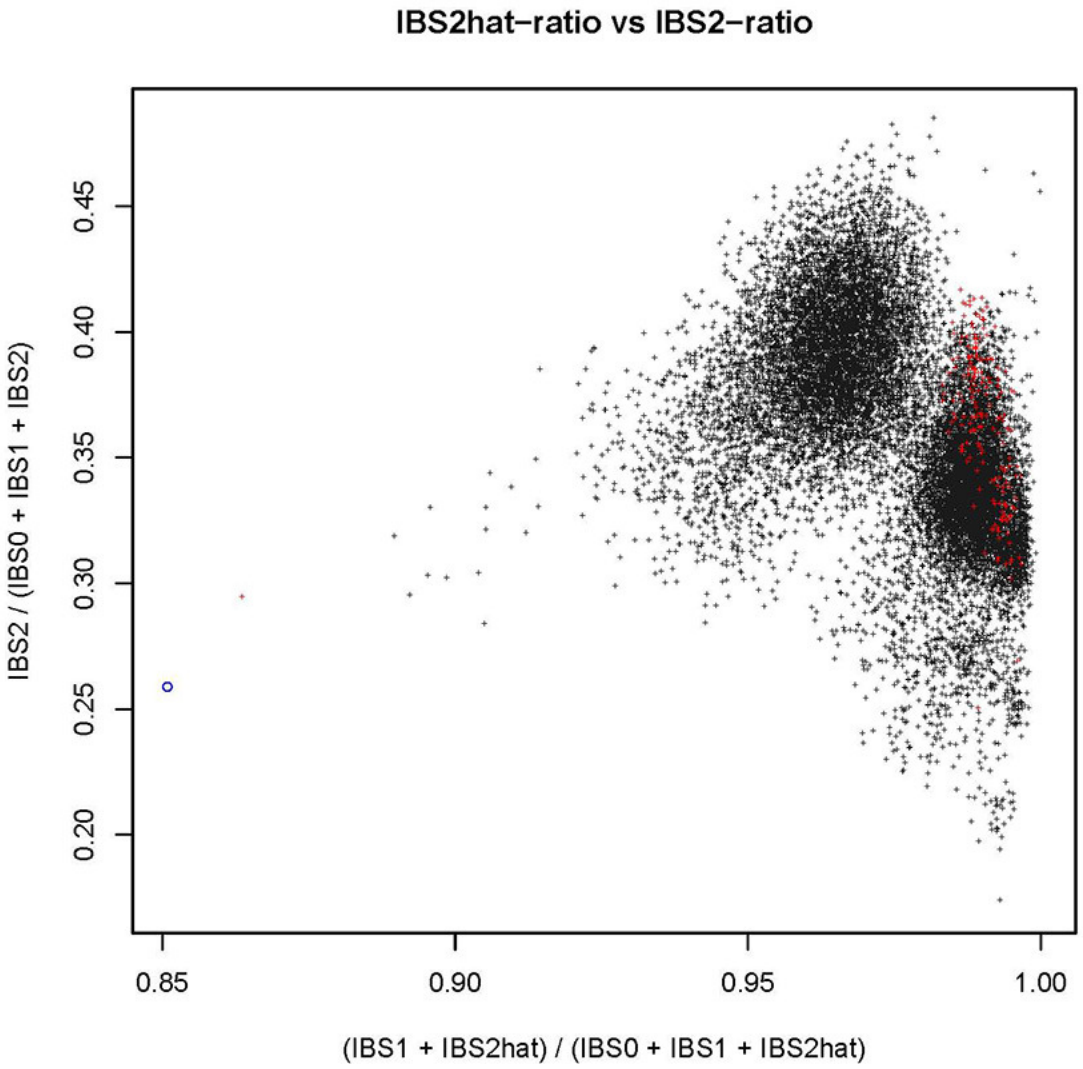



**Figure 2.4** Identity by state (IBS) plot for the mapping population constructed using GBS data. In the figure, Parent-parent comparisons are represented by blue circle, parent-offspring comparisons by red dots and offspring-offspring comparisons by black dots.

The expectations of where the relatedness types cluster on the plot are: (1) clones of either parent will localize to the top left corner of the plot when compared with its known parent having a high proportion of identical (IBS2) genotypes and very few (if any) first-degree or

unshared (IBS1 or IBS0) genotypes (the population appears not to have any of these); (2) full-sibs/$F_1$ of the cross of interest should cluster unto themselves toward the center of the plot with high first-degree (IBS1 and IBS2*) genotypes; (3) selfs will cluster somewhere near X, Y = 1.00, 0.25 (with greater variance on the Y-axis depending on heterozygosity) when compared to its known parent, and will tend to also group more closely with the parent-offspring (red) cluster. Historically, the selfs cluster will smear upward toward the offspring-offspring cluster, as there is still a substantial fraction of relatedness between the two types. Based on these interpretations, the population contains 27 individuals consistent with being selfs (the light cluster at 1.00, 0.25 below the more dense one). The denser cluster above (at X, Y = 1.00, 0.35) consisting of 48 individuals was suspected to be selfs or admixtures. When these 48 individuals were removed from the population many markers became monomorphic in the remaining individuals, indicating that, for whatever reason, these individuals were not desirable for downstream analysis; (4), unrelated individuals and half-sibs will cluster toward the bottom left of the plot, having many unshared alleles (many IBS0 genotypes) and a range of single-allele shared (IBS1) genotypes (depending on the heterozygosity of the source population). If the parents of the cross of interest are outbred and truly unrelated, the parent-parent comparison (the blue point on the plot) will cluster with the unrelated individuals. Historically, half-sib vs full-sib comparisons smear out of the full-sib cluster toward the bottom left of the plot, and this is evident in the plot as the scattered individuals at X,Y = 0.90,0.30 and light scatter of points at the bottom left of the offspring-offspring cluster.

### 2.4.2 High-density genetic linkage map

A total of 4422 SNP markers were found to be segregating in the mapping population (Figure 2.5) with 3873 markers conforming to the expected Mendelian segregation ratios and 549 markers (12%) showing moderate segregation distortion with $0.05 \leq P \leq 0.1$. Some 1,548 (35%) markers were identical and excluded from the study leaving 2,874 markers (65%) for linkage analysis. The <nnxnp> and <llxlm> segregation classes for the markers were the most abundant while triallelic SNPs, <efxeg> class, was less frequent. The final linkage map (Figure 2.6) is composed of 175 $F_1$ individuals with 1,974 SNP markers (Table 2.1) distributed across 21 linkage groups and spanning 1698cM. The average marker density was 0.86. The number of markers ranged from between 31 to 201 in each linkage group and the length of the groups ranged from 12.4 to 141.86 cM. The 21 linkage groups represented the 18 chromosomes of

cassava with chromosome VII, XII, and XIV each represented by two linkage groups. Some linkage groups had substantial gaps between markers (maximum of 1,264,551bp (21.4 cM) in chromosome XIV). This indicates that either there is a recombination 'hotspot' at this location with high rates of crossing over, tending to increase the length of the map, or, there is a lack of molecular markers in that region.

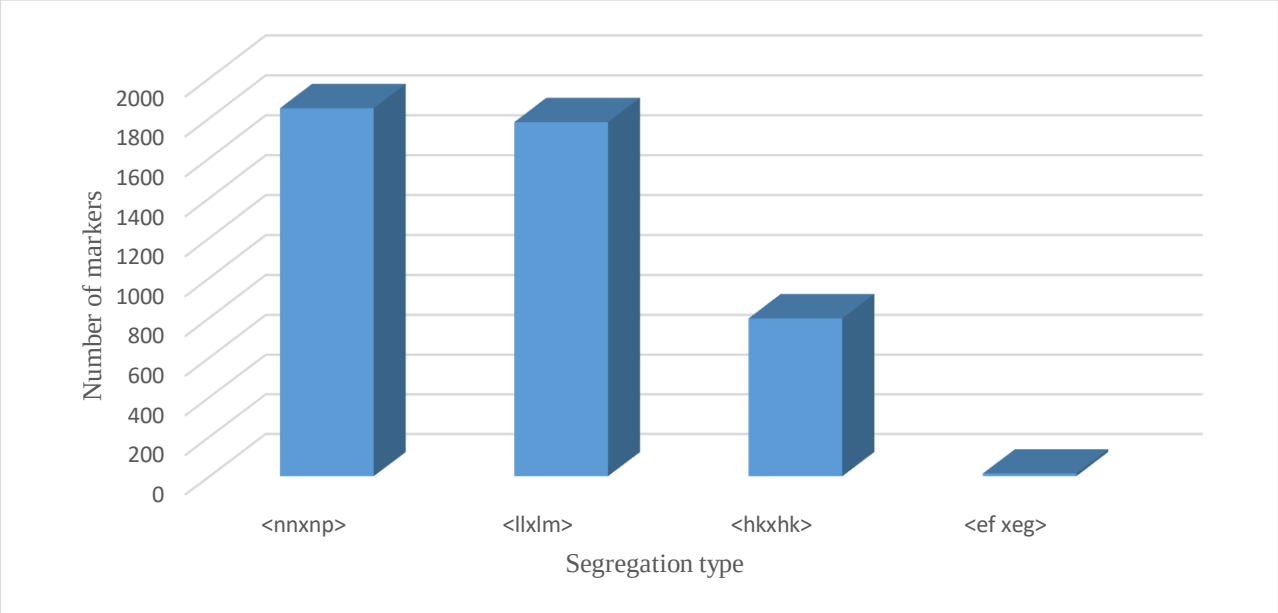



**Figure 2.5** Number of SNP markers and segregation types

**Table 2.1 Characteristics of high density genetic linkage map**

| Linkage group | Chromosome number | No. of markers | distance (cM) | Marker density (cM/marker) |
|---|---|---|---|---|
| 1 | I | 201 | 98.25 | 0.49 |
| 2 | II | 75 | 141.86 | 1.89 |
| 3 | III | 81 | 116.03 | 1.43 |
| 4 | IV | 92 | 97.12 | 1.06 |
| 5 | V | 95 | 96.87 | 1.02 |
| 6 | VI | 157 | 97.16 | 0.62 |
| 7 | VII | 115 | 85.39 | 0.74 |
| 8 | VIII | 88 | 89.04 | 1.01 |
| 9 | IX | 155 | 104.37 | 0.67 |
| 9 | X | 104 | 69.52 | 0.67 |
| 11 | XI | 108 | 96.31 | 0.89 |
| 12 | XII | 57 | 58.42 | 1.02 |
| 12b | XII | 35 | 13.67 | 0.39 |
| 13 | XIII | 76 | 104.93 | 1.38 |
| 14 | XIV | 72 | 26.95 | 0.37 |
| 14b | XIV | 47 | 57.22 | 1.22 |
| 15 | XV | 108 | 106.98 | 0.99 |
| 16 | XVI | 99 | 95.29 | 0.96 |
| 17 | XVII | 108 | 63.61 | 0.59 |
| 18 | XVIII | 70 | 66.90 | 0.96 |
| | **Total** | **1974** | **1698.26** | **0.86** |

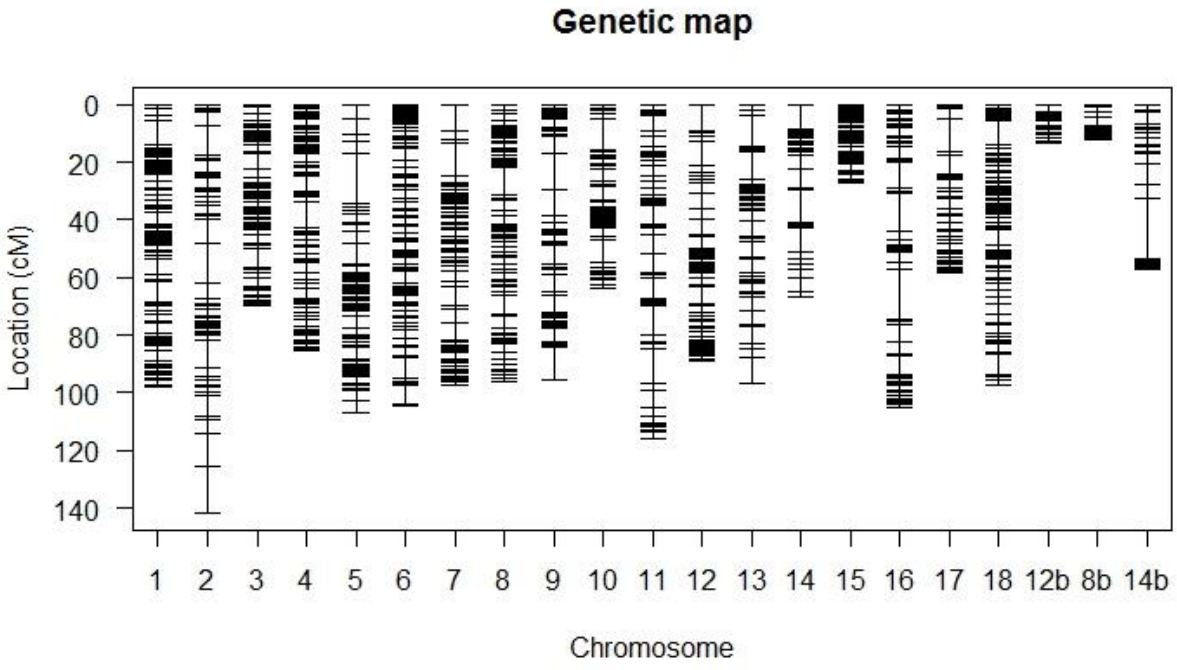



**Figure 2.6** High density genetic linkage map based on SNP markers. Linkage groups 1-18 represent chromosomes I to XVIII of cassava v5.1 assembly. Chromosomes VIII, XII, and XIV are represented by two linkage groups each and the second group is denoted by b.

### 2.4.3 Framework map

The framework map was composed of 230 SNP markers spanning a distance of 1454.9 cM (Table 2.2) and had a marker spacing of at least 5-10 cM (Figure 2.7). It had an average of 6 SNPs per centimorgan (cM).

**Table 2.2 Characteristics of framework genetic linkage map**

| Linkage group | Chromosome | No. of markers | distance (cM) | Marker density (cM/markers) |
|---|---|---|---|---|
| 1 | I | 10 | 55.85 | 5.59 |
| 2 | II | 10 | 145.76 | 14.58 |
| 3 | III | 13 | 80.07 | 6.16 |
| 4 | IV | 13 | 90.54 | 6.96 |
| 5 | V | 12 | 88.75 | 7.40 |
| 6 | VI | 15 | 91.01 | 6.07 |
| 7 | VII | 15 | 82.05 | 5.47 |
| 8 | VIII | 10 | 76.84 | 7.68 |
| 8b | VIII | 4 | 12.70 | 3.17 |
| 9 | IX | 15 | 80.39 | 5.36 |
| 10 | X | 11 | 73.36 | 6.67 |
| 11 | XI | 19 | 86.44 | 4.55 |
| 12 | XII | 13 | 61.47 | 4.73 |
| 12b | XII | 4 | 10.51 | 2.63 |
| 13 | XIII | 11 | 93.35 | 8.49 |
| 14 | XIV | 6 | 18.62 | 3.10 |
| 14b | XIV | 6 | 47.41 | 7.90 |
| 15 | XV | 14 | 84.05 | 6.00 |
| 16 | XVI | 11 | 66.54 | 6.05 |
| 17 | XVII | 8 | 62.75 | 7.84 |
| 18 | XVIII | 10 | 46.48 | 4.65 |
| | Total | 230 | 1454.92 | 6.33 |

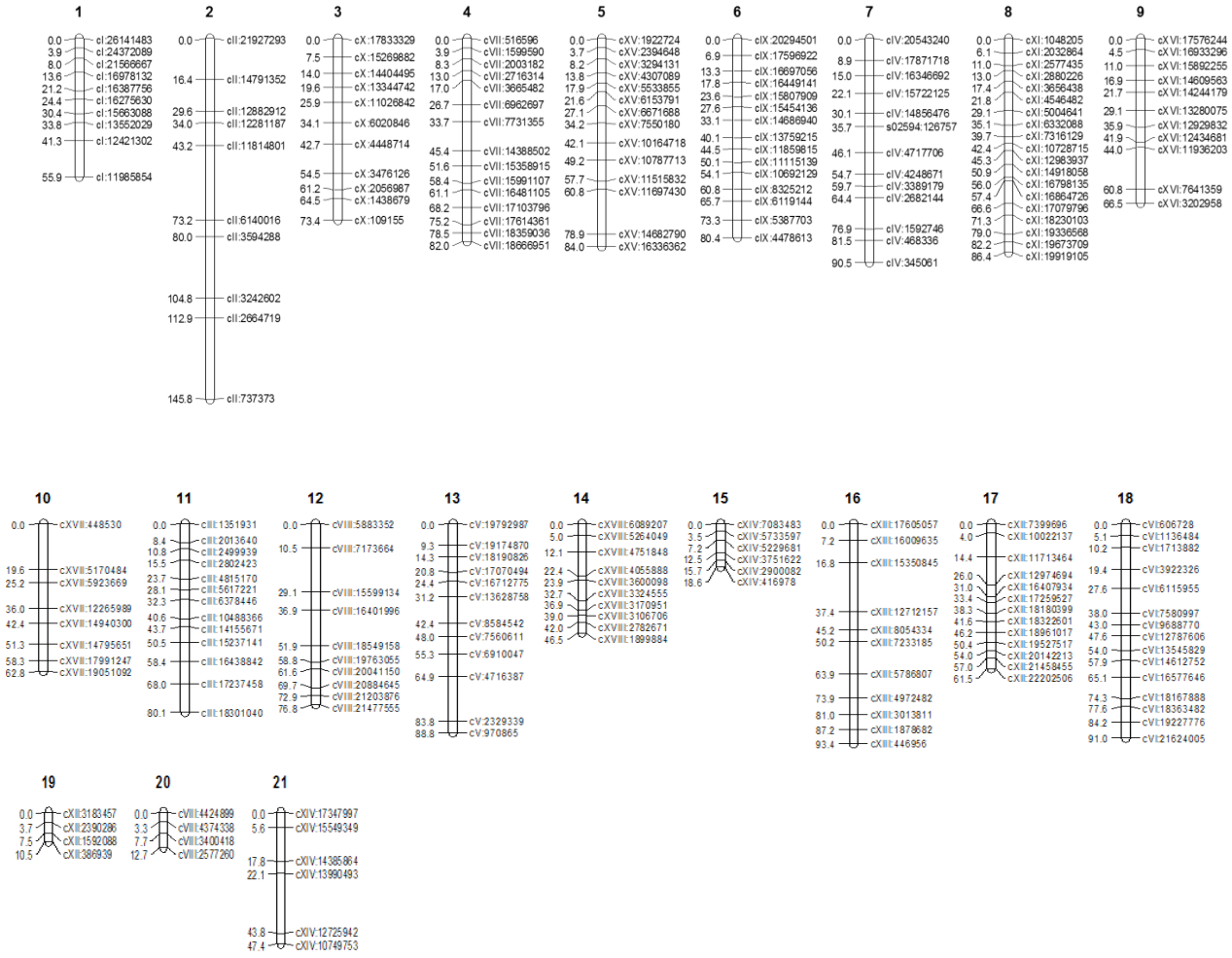



**Figure 2.7** Framework genetic linkage map based on SNP markers. Linkage groups 1-18 represent chromosome I to XVIII of cassava v5.1 assembly. Chromosomes VII, XII, and XIV are represented by two linkage groups each and the second group is denoted as 20,19 and 21 respectively

## 2.5 Discussion

### 2.5.1 Development and validation of the mapping population

. Breeding in cassava is technically challenging due to its heterozygous nature, long growing cycle and low seed yield per pollination. It is highly outcrossing and difficult to develop an adequate sized $F_2$ population usually limiting genetic studies to $F_1$ progeny (Kunkeaw et al., 2010). In this study, an adequate number of seeds were obtained (2676), albeit at low efficiency levels (1116 crosses). The germination rate was also very low at 35%. This is likely to be due to the fact that seeds were germinated in pots on benches, exposing the soil to high diurnal temperature ranges. Placing the seed trays on the ground, thereby reducing diurnal temperature ranges that the soil and seed were exposed to, increased germination rates.

SSR analysis and IBS analysis of SNP data revealed patterns of relatedness. Based on this analysis, some of the $F_1$ progeny were found to be admixtures, most likely having received pollen from a source other than from the donor parent during crossing. The number of progeny with unexpected alleles that were regarded as admixtures was surprisingly high (90). This drastically reduced the number of genotypes available for linkage and QTL mapping in this study. Recent observations indicate female flowers may remain receptive for some time even after pollination and introduction of pollen from elsewhere may have been possible by other pollinating agents such as bees. The present assumption of non-receptivity of the flowers after pollination needs to change and the crossing technique modified to include bagging of the flowers even after pollination.

### 2.5.2 Genetic linkage mapping

During linkage analysis many markers were found to have identical segregation patterns and thus mapped to the same position. Only one marker of identically segregating markers was retained for analysis, thus drastically reducing the number of markers in the final map. Identical markers are common in high density maps and it is prudent to exclude all but one identical marker since they would map in the same position leading to a crowded map. A small number of markers showed segregation distortion which is a common phenomenon in wide crosses (Ren et al., 2014). Segregation distortion can occur for a wide range of reasons, including some

bias during the generation of the population. It is possible that the poor germination posed a bias. Chromosome 1 and 2 initially mapped together and had to be split into two linkage groups thereby increasing the number of linkage groups in the final map. Markers for chromosome 4, 8, and 12 mapped into two linkage groups each further increasing the number of linkage groups to 21 more than the expected number of 18 consistent with the haploid chromosome number in cassava (2n=36) (Nassar, 2000).

The genetic linkage map was based on one-step map approach (Rabbi et al., 2012) due to the lack of enough bridge markers (hk x hk) to integrate the female and male maps. Bridge markers are needed in each linkage group for the maps to be integrated. The present map (1698.2 cM) is shorter than the consensus map (2412.3 cM) (ICGMC, 2015). The marker density of 0.86 markers per cM obtained in this study is higher than the average marker density 1.96 in the consensus map.

The map is fairly saturated with 1,974 SNP makers (approximately one SNP per cM) and a maximum interval size of 21.4cM on chromosome XIV. The mapping population of this study contributed to the development of the consensus map (ICGMC, 2015). The consensus map is based on ten populations and contains 22,400 SNPs anchored in chromosomes and scaffolds of cassava v4.1 assembly. The map obtained here is based on v5.1 genome assembly of cassava in which chromosome numbers are given in Roman numerals. In v6.1 of the genome assembly, Arabic numerals are used to designate chromosome numbers (Bredeson et al., 2016).

### 2.5.3 High-density genetic linkage map

Two maps were generated, one a high-density map using as many SNP markers as possible, and a framework map, with fewer markers evenly distributed across linkage groups, of possibly more accurate position. The level of confidence of the position of markers on the framework map is higher than that of the high-density map. The high-density map obtained for this study is composed of 1974 SNP markers distributed across 21 linkage groups covering a distance of 1698.26cM across the 18 chromosomes of cassava with an average number of markers of 0.86 per cM.

### 2.5.4 Framework map for QTL mapping

QTL mapping requires only a few markers evenly distributed within the linkage groups and this helps to speed up the mapping process. The chi square goodness of fit and marker order are the basic criteria suitable for pruning the high-density map. Markers are further removed from the map giving at least a spacing of 5-10cM, a fairly less dense distribution of markers in the linkage groups leading to a framework map suitable for QTL mapping. The framework map obtained for this study is composed of 230 SNP markers fairly distributed in 21 linkage groups covering a distance of 1454.9cM across the 18 chromosomes of cassava with an average number of markers of 6 per cM. The map distance of the framework map is reduced compared to the 1698.262cM obtained in the high-density map, indicating less error.

## 2.6 Conclusion

The high density genetic linkage map obtained in this study spans the 18 chromosomes of cassava predicted through karyotyping of *M. esculenta* (De Carvalho and Guerra, 2002). It is similar to the consensus map (ICGMC, 2015) and more dense compared to previous maps (Okogbenin et al., 2006; Kunkeaw et al., 2010; Sraphet et al., 2011; Rabbi et al., 2012). The mapping population developed in this study is among the 10 populations that contributed to the construction of the consensus map and is a valuable resource for improvement of the genome assembly of cassava.

## 2.7 References


Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3**,** e3376.

Bernatzky, R., and Tanksley, S.D. (1986). Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. *Genet.* 112**,** 887-898.

Bibikova, M., and Fan, J.-B. (2009). GoldenGate® assay for DNA methylation profiling. *Methods Mol. Biol.***,** 149-163.

Blair, M.W., Fregene, M.A., Beebe, S.E., and Ceballos, H. (2007). "Marker-assisted selection in common beans and cassava," in *Marker-assisted selection: Current status and future perspectives in crops, livestock, forestry and fish,* ed. J.R. Guimaraes, B.D. Scherf, A. Sonnino, J.D. Dargie. (Rome: FAO), 81-115.

Bredeson, J.V., Lyons, J.B., Prochnik, S.E., Wu, G.A., Ha, C.M., Edsinger-Gonzales, E., et al. (2016). Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat. Biotechnol.* 34**,** 562-570.

Chen, X., Fu, Y., Xia, Z., Jie, L., Wang, H., Lu, C., et al. (2012). Analysis of QTL for yield-related traits in cassava using an F1 population from non-inbred parents. *Euphytica.* 187**,** 227-234.

De Carvalho, R., and Guerra, M. (2002). Cytogenetics of *Manihot esculenta Crantz* (cassava) and eight related species. *Hereditas* 136**,** 159-168.

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

DePristo, M.A., Banks, E., Poplin, R.E., Garimella, K.V., Maguire, J.R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43**,** 491-498.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6**,** e19379.

Fregene, M., Angel, F., Gómez, R., Rodríguez, F., Chavarriaga, P., Roca, W., et al. (1997). A molecular genetic map of cassava (*Manihot esculenta Crantz*). *Theor. Appl. Genet.* 95**,** 431-441.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic. Acids Res.* 40**,** 1178-1186.

ICGMC (2015). High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta* Crantz) from 10 populations. *G3* 5**,** 133-144.

Kaweesi, T., Kawuki, R., Kyaligonza, V., Baguma, Y., Tusiime, G., and Ferguson, M.E. (2014). Field evaluation of selected cassava genotypes for cassava brown streak disease based on symptom expression and virus load. *J. Virol.* 11**,** 216.

Kawuki, R., Herselman, L., Labuschagne, M., Nzuki, I., Ralimanana, I., Bidiaka, M., et al. (2013). Genetic diversity of cassava (*Manihot esculenta* Crantz) landraces and cultivars from Southern, Eastern and Central Africa. *Plant Genet. Resour.* 11**,** 170-181.

Kunkeaw, S., Tangphatsornruang, S., Smith, D.R., and Triwitayakorn, K. (2010). Genetic linkage map of cassava (*Manihot esculenta* Crantz) based on AFLP and SSR markers. *Plant. Breed.* 129**,** 112-115.

Mahungu, N.M. (2004). Advances in root and tuber crops technologies for sustainable food security, improved nutrition, wealth creation and environmental conservation in Africa: proceedings of the 9th ISTRC-AB symposium. Mombasa, Kenya, 1-5 Nov. 2004. IITA.

Muhanna, M., and Mtunda, K. (2003). "Cassava Brown Streak Virus Disease Transmission Studies in Tanzania", in: *Cassava Brown Streak Virus Disease: Past, Present and Future.* (ed.) J. Hillocks. (Mombasa, Kenya).

Okogbenin, E., Egesi, C., Olasanmi, B., Ogundapo, O., Kahya, S., Hurtado, P., et al. (2012). Molecular marker analysis and validation of resistance to cassava mosaic disease in elite cassava genotypes in Nigeria. *Crop Sci.* 52**,** 2576-2586.

Nassar, N. (2000). Cytogenetics and evolution of cassava (Manihot esculenta Crantz). Genet. Mol. Biol. 23, 1003-1014.

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

Rabbi, I.Y., Kulembeka, H.P., Masumba, E., Marri, P.R., and Ferguson, M. (2012). An EST-derived SNP and SSR genetic linkage map of cassava (*Manihot esculenta* Crantz). *Theor. Appl.Genet.* 125**,** 329-342..

Ren, Y., McGregor, C., Zhang, Y., Gong, G., Zhang, H., Guo, S., et al. (2014). An integrated genetic map based on four mapping populations and quantitative trait loci associated

with economically important traits in watermelon (*Citrullus lanatus*). *BMC Plant Biol.* 14**,** 1.

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single nucleotide polymorphism genotyping using Competitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol. Breed.* 33**,** 1-14.

Sraphet, S., Boonchanawiwat, A., Thanyasiriwat, T., Boonseng, O., Tabata, S., Sasamoto, S., et al. (2011). SSR and EST-SSR-based genetic linkage map of cassava (*Manihot esculenta* Crantz). *Theor. Appl. Genet.* 122**,** 1161-1170.

Tang, N., van der Lee, T., Shahin, A., Holdinga, M., Bijman, P., Caser, M., et al. (2015). Genetic mapping of resistance to Fusarium oxysporum f. sp. tulipae in tulip. *Mol. Breed.* 35**,** 1-17.

Tanksley, S.D. (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature.* 335**,** 6170.

Van Ooijen, J. (2006). JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. *Nature.* 33**,** 1371.

Van Os, H., Andrzejewski, S., Bakker, E., Barrena, I., Bryan, G.J., Caromel, B., et al. (2006). Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map. *Genet.* 173**,** 1075-1087.

Wang, B., Jiang, L., Zhang, Y., Zhang, W., Wang, M., Cheng, X., et al. (2011). QTL mapping for resistance to strip virus disease in rice. *Plant. Breed.* 130**,** 321-327.

# Chapter 3

# Quantitative Trait Loci Mapping for Resistance to Cassava brown-streak disease, Cassava Mosaic disease and cassava Green mite


**Authors**

Inosters Nzuki[1], Geoffrey S. Mkamilo[2], Benendeta Kimata[2] Fortunus Kapinga[2], Esther Masumba[3], Kasele Salum[4,] Heneliko Kulembeka[4], Alexander A. Myburg[1], Nicholas van der Merwe[1] and Morag E. Ferguson*[5]

**Affiliations**

[1]Department of Genetics, University of Pretoria, Private bag X20, Hatfield 0028, South Africa

[2]Naliendele Agricultural Research Institute, P. O. Box 509, Mtwara, Tanzania.

[3]Sugarcane Research Institute, P. O. Box 30031, Kibaha, Tanzania

[4]Ukiriguru Agricultural Research Institute, P. O. Box 1433, Mwanza, Tanzania

[5]International Institute of Tropical Agriculture (IITA), P.O. Box 30709-00100, Nairobi, Kenya

* Corresponding author: m.ferguson@cgiar.org , (+254) 733524685

## Author contributions

This chapter has been published in the journal Frontiers in Plant Sciences under the tittle 'QTL mapping for pest and disease resistance in cassava and coincidence of some QTL with introgression regions derived from *Manihot glaziovii*" published online in July 2017 (Front. Plant Sci. 8:1168. doi: 10.3389/fpls.2017.01). Inosters Nzuki conducted the study analyzed the data and wrote the manuscript. Morag Ferguson conceived the project and sourced for funding of the study. Geoffrey Mkamilo facilitated logistical support for the phenotypic trials established in Tanzania. Esther Masumba, Benendetta Kimata Fortunus Kapinga, Kasele salum and Heneriko Kulembeka provided technical support for establishment of the phenotypic trials and data collection. Prof. Alexander A. Myburg, Dr. Morag Ferguson and Dr. Nicolas van der Merwe provided scientific guidance, advice and editorial support for the manuscript. All authors read and approved the manuscript.

# 3

## 3.1 Abstract

Genetic mapping of QTL for resistance to CBSD, CMD and CGM was performed using Genetic Analysis for Clonal $F_1$ and Double cross populations (GACD) software based on an $F_1$ cross developed between Kiroba and AR37-80. Disease scoring on a scale of 1-5 for CBSD foliar symptoms and root necrosis was done in two consecutive years in two sites in Tanzania. Fifteen significant QTL were identified; two were associated with CBSD root necrosis only, and were detected on chromosomes V and XII, while seven were associated with CBSD foliar symptoms only and were detected on chromosomes IV, VI, XVII, and XVIII. QTL on chromosomes XI and XV were associated with both CBSD foliar and root necrosis symptoms. Two QTL were found to be associated with CMD and were detected on chromosomes XII and XIV, while two were associated with CGM and were identified on chromosomes V and X. The QTL associated with CMD resistance detected on chromosome XII (qCMDc12K) is thought to be the CMD2 locus identified previously.

**Keywords;** Cassava brown streak disease, CMD2 locus, Genetic analysis for clonal F1 and double cross populations (GACD), Quantitative trait loci. .

## 3.2 Introduction

Cassava (*Manihot esculenta* Crantz ) (2n=36) is an important food crop in sub Saharan Africa providing daily calories for more than 800 million resource poor people in rural and peri-urban areas (Lebot, 2009). Cassava is a versatile crop, able to adapt to harsh environmental conditions with minimal rainfall and other farm inputs. It is regarded as a food security crop when cereals fail due to its inherent drought tolerance and the fact that the roots can remain in the soil and be harvested when needed. Cassava is a commercial crop providing income to women and the youth in the rural areas where it is sold raw or processed into various products. It is also used as animal feed. Approximately 229.5 million MT of cassava have been produced worldwide with Africa alone contributing a total production of more than 118 million tones (FAO, 2013), which is greater than for any other crop in Africa. FAO statistics indicate the average yields for cassava in Africa to be 8 tones/ha yet the crop has a potential to yield 80 MT/ha (FAO, 2015). The actual and potential yield gap is huge and bridging the gap is a challenge to research. Genomic tools including molecular markers are valuable for crop improvement in terms of disease resistance, yield and other quality traits, and when used in conjunction with conventional breeding can speed up cultivar development and delivery.

The benefit of cassava in improving food security in Africa is being eroded by biotic and abiotic factors including pests and diseases. Cassava green mite (CGM), cassava mealybug and the variegated grasshopper are the major cassava pests while cassava mosaic disease (CMD), cassava brown streak disease (CBSD), cassava bacterial blight, cassava anthracnose disease are some of the common diseases of cassava (Campo et al., 2011). Viral diseases affecting cassava are the most threatening and limiting factors to cassava production, due to their current epidemiological and distribution trends (Legg et al., 2011). CBSD, originally considered to be localized in the coastal regions and limited by altitude of 1000m below sea level (Alicai et al., 2007) is now rapidly spreading to areas that were considered safe (Campo et al., 2011). This is compounded by changing climatic conditions and distribution trends of whitefly populations in the region. Pre-emptive breeding measures are urgently needed in Central and West Africa; the major cassava producing regions in Africa (www.fao.org) to mitigate against the threat posed by CBSD. These measures include the development of cultivars with broad genetic diversity of resistance genes that can offer molecular barriers to the spread of the disease and use of appropriate genomic tools. The local landraces found in the region are a good source of such diversity.

The most feasible and sustainable control of viral diseases of cassava is through planting of resistant cultivars that can offer molecular barriers and thus limit the spread. Diversity assessments have shown a broad genetic base in the local landraces found in the region (Kawuki et al., 2013) and this is a useful genetic stock that can be utilized to develop resistant cultivars suited for the region and also can act as pre-emptive stock for other regions as yet unaffected. Cassava breeders have identified some local landraces that show strong field resistance towards CBSD, they get infected with the viruses and show leaf symptoms but root necrosis is limited or delayed until harvest (Hillocks and Jennings, 2003). Kiroba is one of the local cultivars grown by farmers in Tanzania that has been identified with strong field resistance towards CBSD. It shows almost no root necrosis even at twelve month after planting despite showing leaf symptoms. It is thought to have been developed at the Amani breeding station in Tanzania but lost its identity when the breeding program ceased in 1958. It is thought to have dispersed into farmers' fields and adopted by farmers under a local name. Genetic diversity assessment shows Kiroba to be closely related to Namikonga (Ferguson et al., 2012), an interspecific hybrid from the Amani breeding program. QTL for CBSD resistance in Namikonga have recently been published (Masumba et al. 2017). These landraces despite having promising field resistance towards CBSD are susceptible to CMD and are attacked by cassava green mites that have also been identified as great threats to cassava production in the region. Pyramiding of QTL for CMD and cassava green mite resistance into these local CBSD resistant landraces is important to counter the emerging threats posed by a combination of these diseases and pests. AR37-80 is a CIAT developed cultivar that has been introduced into Tanzania and shows resistance against CMD and CGM but succumbs to CBSD. The rich genetic stock of these cultivars can be introgressed in the already identified elite local landraces by use of appropriate genomic tools including MAS.

A consensus genetic linkage map of cassava has been published (ICGMC, 2015), and this is a valuable molecular tool that opens opportunities for further study of quantitative traits associated with viral diseases and pests of cassava. Identification of quantitative trait loci (QTL) is an important step in the development of new cultivars with resistance to diseases and pests. Pyramiding of several QTL associated with various traits into farmer preferred cultivars is an important strategy. If molecular markers are found tightly linked to QTL, hence heritable then selection would be done at the seedling stage. This would reduce the time taken for cultivar development and release and ameliorate the otherwise challenging biology of cassava in terms

of being heterozygous and outcrossing coupled with a long breeding cycle and low seed yield per pollination (Kunkeaw et al., 2010).

The objective of this study was to identify QTL associated with CBSD, CMD and CGM using a mapping population developed between Kiroba (female) and AR37-80 (male) and identify markers linked to these traits for possible deployment in MAB.

## 3.3 Materials and methods

### 3.3.1 Phenotyping trials and field evaluation

Stem cuttings of the $F_1$ mapping population described in Chapter 3 were used to establish phenotyping trials in two coastal sites in Tanzania where CBSD disease pressure is high; Chambezi (6°33'19"S 38°44'51"E) and Naliendele (0°23'00.60"S 40°09'50.58"E ) (Figure 3.1).


**Figure 3.1** Map of Tanzania showing phenotyping sites along the coast, in Naliendele and Chambezi. A yellow star indicates the experimental sites

The climate of both Chambezi and Naliendele is tropical and classified as Aw (Tropical wet and dry or savanna climate; with the driest month having precipitation less than 60 mm (2.4 in) and less than 4% of the total annual precipitation) by the Köppen-Geiger system (Rubel and Kottek, 2010). At Chambezi, rainfall occurs between March to May, and October to December while between June and August there is very little rainfall (Figure 3.2) with the average annual of 1015mm and temperature of 26.6 °C. At Naliendele, summers are much rainier than the winters with an annual average rainfall of 1024mm and temperature of 26.3 °C (Figure 3.3). The main soil type in Mtwara is reddish sands and loamy sands, while Chambezi has sandy-loam and clay soils (Mponda et al., 2001).

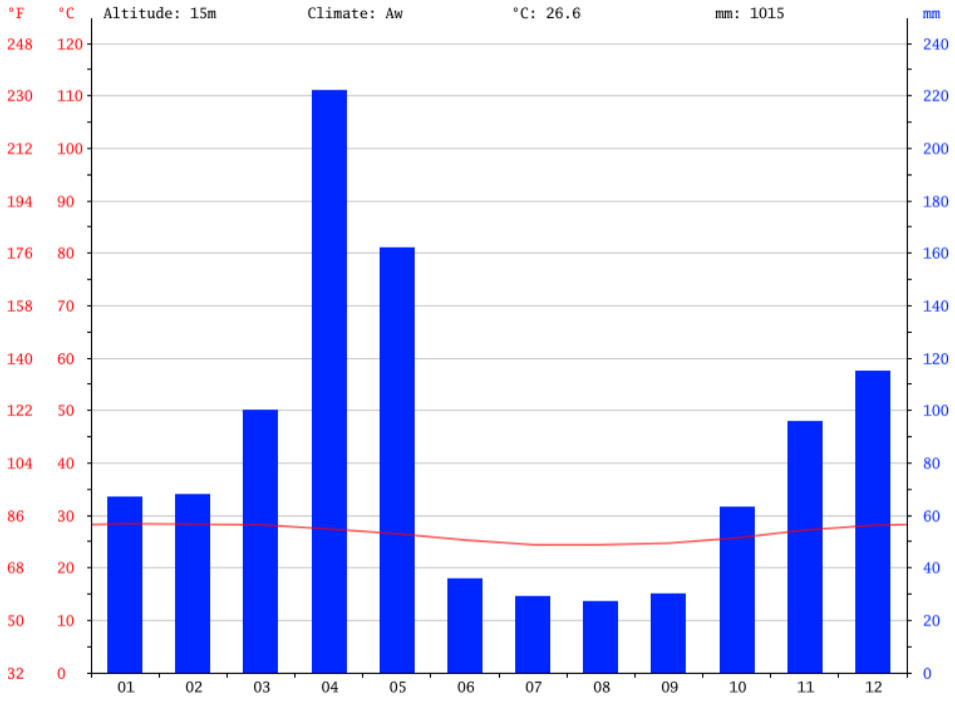



**Figure 3.2** Climate graph of Chambezi, Bagamoyo region, Tanzania. Blue bars are rainfall (mm), red line is temperature.

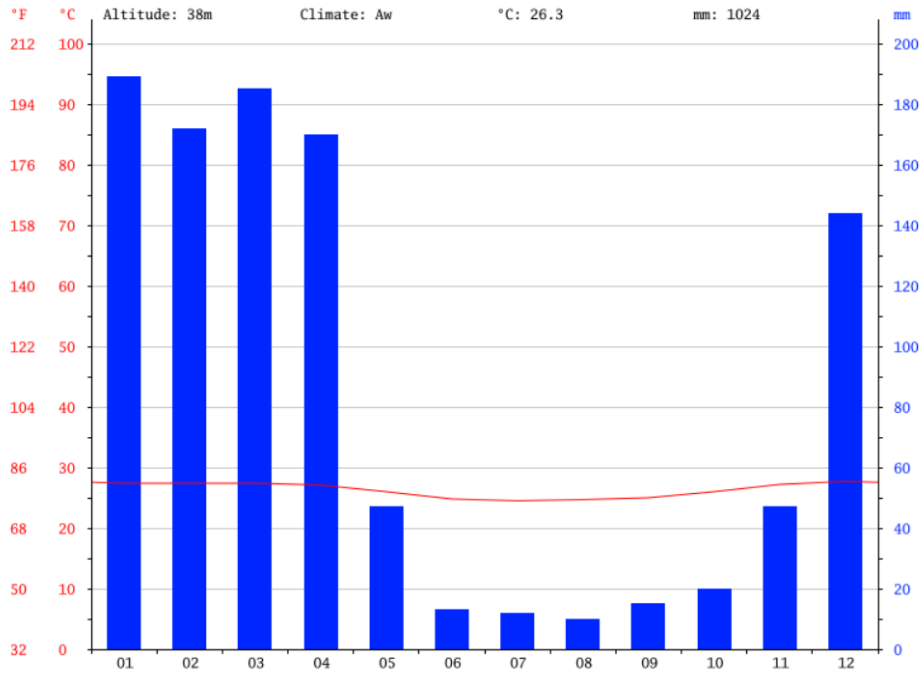
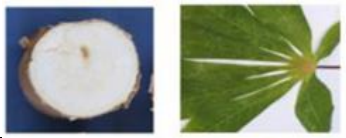



**Figure 3.3** Climate graph of Naliendele, Mtwara region, Tanzania. Blue bars are rainfall (mm), red line is temperature

Phenotyping trials were conducted over two successive growing seasons, 2013-2014 and 2014-2015 using an alpha lattice design with incomplete blocks replicated twice. The $F_1$ progeny were evaluated for CBSD foliar symptoms, CMD and CGM at 3, 6 and 9 MAP using a scale of 1-5 (Alicai et al., 2007). The roots were harvested 12 MAP and chopped into equal slices (5cm) using a fabricated machine cutter. CBSD root necrosis was evaluated on a scale of 1-5 (Hillocks and Thresh, 2000) (Table 3.1). Seven roots per plant and a maximum of seven slices per root were chosen at random for disease scoring. All data were captured electronically on a tablet using the Field Book App (Poland and Rife, 2012).

**Table 3.1 Classification of CBSD foliar and root symptoms on a scale of 1-5**

| Class | CBSD Foliar symptoms | Root symptoms | Pictoral representation |
|---|---|---|---|
| **1** | no visible symptoms | No visible symptoms |  |
| **2** | mild vein yellowing or chlorotic blotches on some leaves | Less than 25% root necrosis |  |
| **3** | pronounced/extensive vein yellowing or chlorotic blotches on leaves, but no lesions or streaks on stems | 25-50% root necrosis |  |
| **4** | pronounced/extensive vein yellowing or chlorotic blotches on leaves and mild lesions or streaks on stems, | 50-75% root necrosis |  |
| **5** | pronounced/extensive vein yellowing or chlorotic blotches on leaves and severe lesions or streaks on stems, defoliation and dieback | 75-100% root necrosis |  |

### 3.3.2 Phenotypic trait analysis and QTL mapping

The statistics for the phenotypic data were computed using GenStat (Ripatti et al., 2009). The mean, skewness, kurtosis, and Shapiro-Wilk normality test were used to infer the distribution and normality of the data. Box plots and normal plot (Q-Q plots) were used to inspect the quality of the data and identify outliers. The mean of each genotype across the replicates in each year and site were calculated and used for QTL mapping. Inclusive interval mapping (ICIM) for QTL detection was done using the genetic Analysis of Clonal $F_1$ and Double cross populations (GACD) software (Zhang et al., 2015). A genome-wide LOD threshold with P value of 0.05 was obtained by permutation test (1000 replications) to identify significant QTL (Manichaikul et al., 2007). Additive and dominant effects were calculated based on the method of Muchero (Muchero et al., 2013);

a= {mu(ac)-mu(bd)}/2;

d= {mu (ad) +mu (bc)}/2-{mu (ac) +mu (bd)}/2

where mu (ac) and mu (bd) are the phenotypic means for heterozygous loci carrying alleles derived from the same species and mu (ad) and mu (bc) are heterozygous loci carrying alleles derived from both species as computed by GACD software.

QTL mode of action was calculated as a ratio of dominance over additivity, d/ a. According to Muchero's paper (Muchero et al., 2013) d/ a ratios of <1 are regarded as reflecting under-dominance or partial dominance while ratios >1 reflect over- dominance.

ICIM used in this study is an improved algorithm of composite interval mapping suitable for biparental crosses (Zhang et al., 2015). It has increased detection power, a reduced false detection rate, and less biased estimates of QTL effects. This approach minimizes the bias due to Beavis effect (Xu, 2003) associated with QTL analysis using a small population size and was suitable for our population that comprised of 106 individuals. QTL were named with q (for QTL), the name of the trait (e.g. CBSDRN for cassava brown streak disease root necrosis or CBSDRNF for root necrosis and foliar symptoms ) followed by "c" for chromosome and the number of the chromosome on which the QTL lies, followed by "K" for Kiroba or Ar for AR37-80 e.g., CBSDRNc11K. In cases where there was more than one QTL on a chromosome, an "L" or "R" was given as a suffix to indicate the left or right arm of the chromosome. When more than one QTL was present per chromosome arm, "a" and "b" were used to discriminate them.

## 3.4 Results

### 3.4.1 Phenotypic trait data quality and frequency distribution

The phenotypic trait data were of good quality without extreme outliers as shown by the Q-Q plot (Figure 3.4).

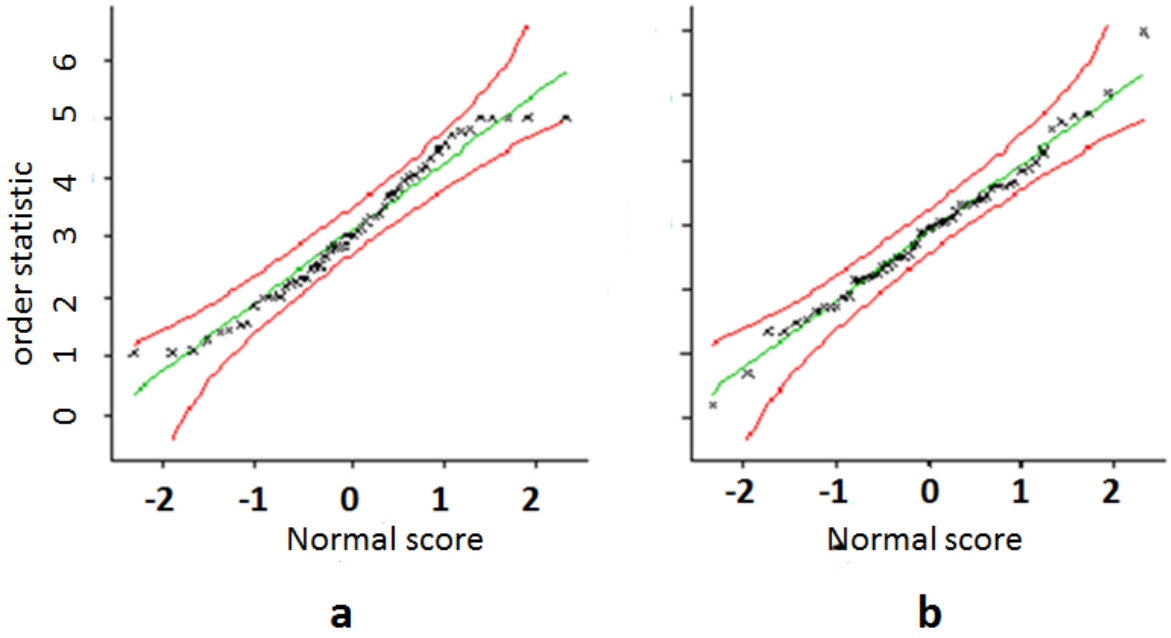



**Figure 3.4** Q-Q plot showing phenotypic data for CBSD root necrosis (a) and (b) CBSD foliar symptoms in Chambezi (2013-2014).

### 3.4.2 Phenotypic trait evaluation and distribution

The summary statistics for the phenotypic data obtained for two years of phenotyping 2013-2014 (year1) and 2014-2015 (year 2) in two locations, Chambezi (C1 and C2) and Naliendele (N1 and N2) are presented in Table 3.2. The highest mean (3.073 of a maximum of 5) was obtained for root necrosis at C1and the lowest was for CBSD foliar symptoms (1.032) in N1. The standard error of the mean (SE) ranged from 0.011 for CBSD foliar symptoms at N1 to 0.163 for root necrosis at C1. CBSD foliar symptoms were positively skewed towards class 1

in N1 and N2 (skewness 3.327 and 4.342). Root necrosis was also positively skewed in N1 and N2 (skewness 2.047and 3.569). This is also reflected in Figure 3.5 showing the frequency distribution of the phenotypic data. CGM was also positively skewed at C2. The highest variance was obtained for root necrosis in C1 (1.439) while the lowest was obtained for CBSD foliar symptoms in N1 (0.008). The SE followed a similar trend as the variance with the highest value was reported for necrosis at C1 and lowest 0.011 reported for CBSD foliar symptoms at N1. Kurtosis was highest in N2 for both CBSD foliar symptoms and root necrosis (20.891 and 20.619 respectively). The distribution of these data for each trait was confirmed by frequency distribution graph (Figure 3.5 and Figure 3.6) for each trait and season. The disease distribution across sites for the same year and across seasons for the same site was significantly different for all traits measured $P<0.05$ (Table 3.3).

**Table 3.2 Descriptive statistics for phenotypic data for two years 2013-2014(1) and 2014 - 2015(2) for two sites, Chambezi (C) and Naliendele (N) in Tanzania**

| Trait* | Mean | Std. error | Std. deviation | Variance | Kurtosis | Skewness | Min | Max | *P*-value |
|---|---|---|---|---|---|---|---|---|---|
| *CGM_N1* | 1.581 | 0.033 | 0.284 | 0.081 | 0.448 | 0.202 | 1.000 | 2.458 | 0.065 |
| *CGM_N2* | 1.978 | 0.038 | 0.315 | 0.099 | 1.536 | 0.597 | 1.200 | 3.000 | 0.075 |
| *CGM_C1* | 1.422 | 0.033 | 0.253 | 0.064 | 0.087 | 0.378 | 1.000 | 2.146 | 0.066 |
| *CGM_C2* | 1.103 | 0.025 | 0.206 | 0.043 | 6.560 | 2.543 | 1.000 | 2.000 | 0.051 |
| *CMD_N1* | 1.807 | 0.085 | 0.743 | 0.552 | -0.193 | 0.811 | 1.000 | 3.778 | 0.170 |
| *CMD_N2* | 2.160 | 0.102 | 0.847 | 0.718 | -1.139 | 0.146 | 1.000 | 3.900 | 0.204 |
| *CMD_C1* | 2.452 | 0.079 | 0.609 | 0.371 | -0.837 | -0.387 | 1.222 | 3.542 | 0.159 |
| *CMD_C2* | 2.413 | 0.087 | 0.709 | 0.503 | 0.130 | 0.184 | 1.000 | 4.300 | 0.174 |
| *CBSD_Foliar _N1* | 1.032 | 0.011 | 0.091 | 0.008 | 11.405 | 3.327 | 1.000 | 1.476 | 0.021 |
| *CBSD_Foliar_N2* | 1.151 | 0.050 | 0.416 | 0.173 | 20.891 | 4.342 | 1.000 | 3.600 | 0.099 |
| *CBSD_Foliar_C1* | 2.443 | 0.071 | 0.542 | 0.294 | 0.511 | 0.229 | 1.111 | 4.000 | 0.141 |
| *CBSD_Folair_C2* | 1.510 | 0.044 | 0.358 | 0.128 | -0.781 | 0.413 | 1.000 | 2.333 | 0.088 |
| *CBSD_Necrosis_N1* | 1.249 | 0.034 | 0.295 | 0.087 | 4.841 | 2.047 | 1.000 | 2.412 | 0.068 |
| *CBSD_Necrosis_N2* | 1.320 | 0.039 | 0.292 | 0.085 | 20.619 | 3.569 | 1.000 | 3.026 | 0.077 |
| *CBSD_Necrosis_C1* | 3.073 | 0.163 | 1.200 | 1.439 | -1.068 | 0.066 | 1.042 | 5.000 | 0.327 |
| *CBSD_Necrosis_C2* | 1.916 | 0.100 | 0.788 | 0.620 | -0.263 | 0.703 | 1.000 | 4.109 | 0.200 |

*C1, Phenotyping in Chambezi year1; C2, phenotyping in Chambezi year 2; N1, phenotyping in Naliendele year1; and N2 phenotyping in Naliendele year 2. All traits are scored on a scale of 1-5 with 1 being no symptoms and 5 being the maximum symptoms. P-value was calculated based on 95% confidence limit

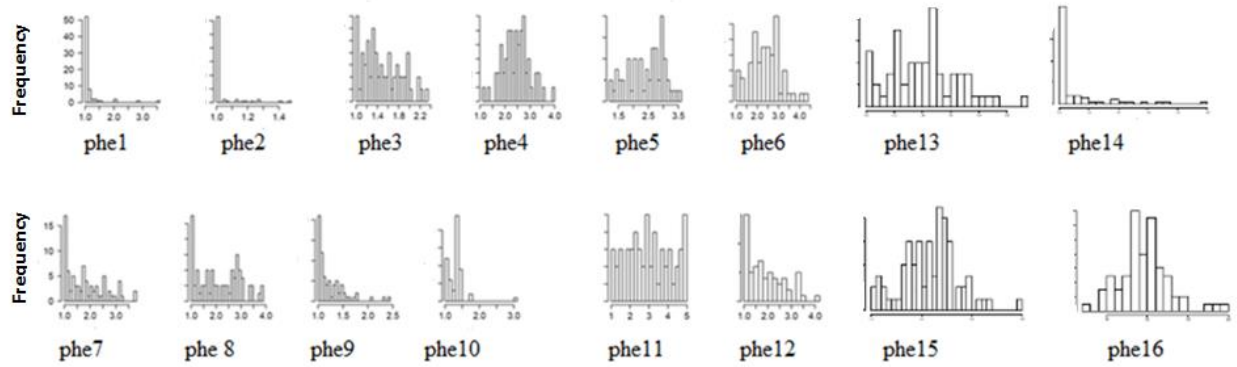



**Figure 3.5** Frequency distribution of phenotypic data for cassava green mite (CGM), cassava mosaic disease (CMD), cassava brown streak disease (CBSD) root necrosis, and CBSD foliar symptoms for two phenotyping years at two locations, Naliendele (N)and Chambezi (C). The labels on the graphs are as follows; phe1=CBSD foliar symptoms at N2, phe2=CBSD foliar symptoms at N1, phe3=CBSD foliar symptoms at C2, phe4=CBSD foliar symptoms at C1, phe5=CMD at C1, phe6=CMD at C2, phe7=CMD at N1, phe8=CMD at N2, phe9=CBSD root necrosis at N1, phe10=CBSD root necrosis at N2, phe11=CBSD root necrosis at C1, phe12=CBSD root necrosis at C2, phe13=CGM at C1 phe14= C2, phe15=CGM N1 and phe16=CGM N2

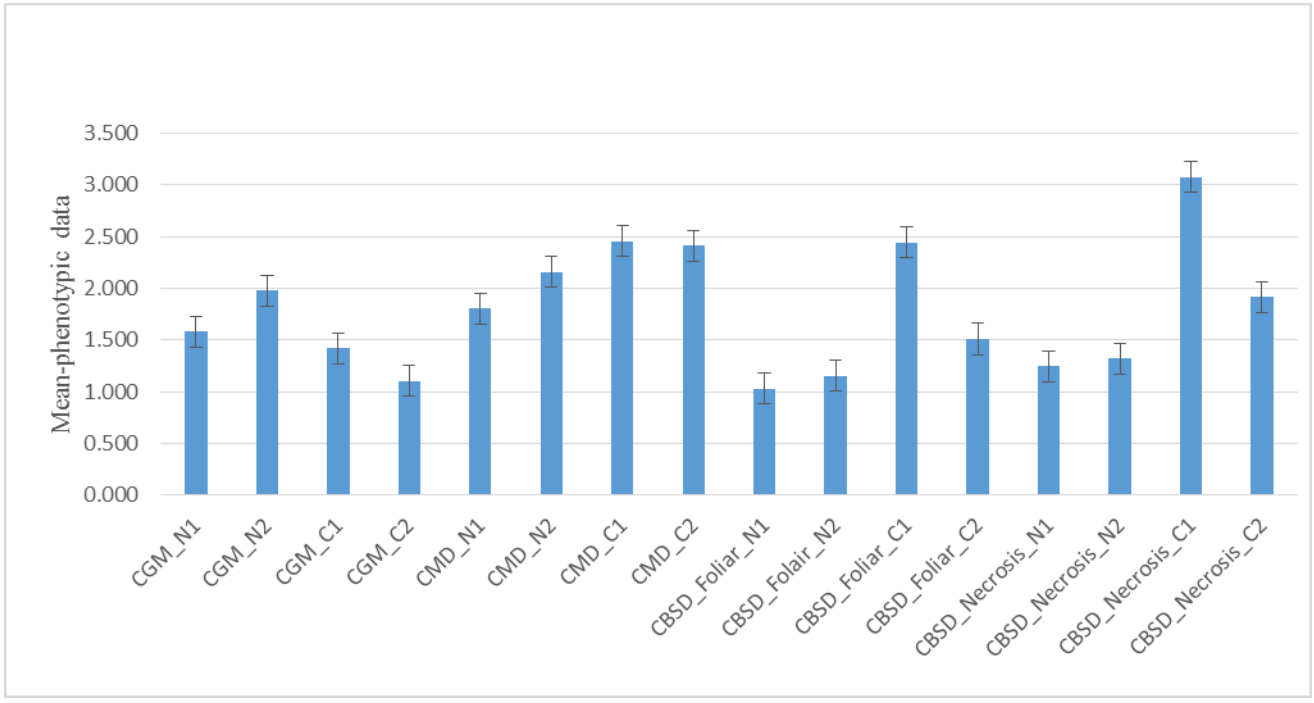



**Figure 3.6** Distribution of the phenotypic trait means across sites and years

**Table 3.3 Analysis of variance (ANOVA) across sites and years**

| | *Source of Variation** | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
|---|---|---|---|---|---|---|---|
| CBSD_Necrosis | N1_N2 | 1.3977286 | 1 | 1.3977286 | 3.194277255 | 0.0753386 | 3.886121 |
| | C1_C2 | 10.487264 | 1 | 10.487264 | 4.800214211 | 0.0295564 | 3.886121 |
| | C1_N1 | 25.4795 | 1 | 25.4795 | 14.55290527 | 0.0001792 | 3.886121 |
| | C2_N2 | 8.9495113 | 1 | 8.9495113 | 10.26905208 | 0.0015636 | 3.886121 |
| CBSD_Foliar | C1_C2 | 27.134009 | 1 | 27.134009 | 131.5793325 | 3.692E-21 | 3.918178 |
| | N1_N2 | 0.5083999 | 1 | 0.5083999 | 5.792936116 | 0.0173668 | 3.907312 |
| | C2_N2 | 4.3844347 | 1 | 4.3844347 | 29.02398243 | 3.118E-07 | 3.911795 |
| | C1_N1 | 65.748541 | 1 | 65.748541 | 491.0939605 | 2.565E-46 | 3.912875 |
| CMD | C1_C2 | 0.047749 | 1 | 0.047749 | 0.108293249 | 0.7426553 | 3.918178 |
| | N1_N2 | 4.5239247 | 1 | 4.5239247 | 7.173528011 | 0.0082653 | 3.907312 |
| | C2_N2 | 2.1493916 | 1 | 2.1493916 | 3.506911615 | 0.0633087 | 3.912331 |
| | C1_N1 | 13.82904 | 1 | 13.82904 | 29.24131893 | 2.871E-07 | 3.912331 |
| CGM | N1_N2 | 5.6987661 | 1 | 5.6987661 | 63.59393325 | 4.451E-13 | 3.907312 |
| | C1_C2 | 3.1694071 | 1 | 3.1694071 | 60.03240176 | 2.985E-12 | 3.918178 |
| | C1_N1 | 0.8384246 | 1 | 0.8384246 | 11.41468356 | 0.0009581 | 3.912875 |
| | C2_N2 | 26.010994 | 1 | 26.010994 | 362.5913146 | 6.149E-40 | 3.911795 |

*C ,Chambezi, N, Naliendele phenotypic sites; SS sum of squares, df degrees of freedom, MS mean square, F (F ratio); *P*-value , significant level and F critical value set by experiment. CMD denotes cassava mosaic disease and CGM cassava green mite.

### 3.4.3 Test of normality

The Shapiro-Wilk normality test results are presented in Table 3.4. CBSD foliar symptoms at Chambezi in 2013-2014 ($P$=0.928> 0.05), CGM in both Chambezi and Naliendele in 2013-2014 ($P$=0.363 and 0.385) and CMD at Chambezi in 2014-2015 phenotyping year ($P$=0.533) were normally distributed. The rest of the traits deviated from a normal distribution ($P$<0.05).

**Table 3.4 Shapiro-Wilk test for normality**

| Data variate: | Test statistic W: | Probability: |
|---|---|---|
| CBSD_Folair _C1 | 0.9905 | 0.928 |
| CBSD_Foliar_N1 | 0.4117 | <0.001 |
| CBSD_Foliar_C2 | 0.9546 | 0.017 |
| CBSD_Foliar_N2 | 0.4078 | <0.001 |
| CGM_C1 | 0.9781 | 0.363 |
| CGM_N1 | 0.9824 | 0.385 |
| CGM_C2 | 0.5773 | <0.001 |
| CGM_N2 | 0.966 | 0.054 |
| CMD_C1 | 0.9548 | 0.028 |
| CMD_N2 | 0.9026 | <0.001 |
| CMD_C2 | 0.9836 | 0.533 |
| CMD_N2 | 0.9383 | 0.002 |
| CBSD_Necrosis_C1 | 0.9554 | 0.043 |
| CBSD_Necrosis_N1 | 0.771 | <0.001 |
| CBSD_Necrosis_C2 | 0.9233 | <0.001 |
| CBSD_Necrosis_N2 | 0.6512 | <0.001 |

### 3.4.4 Significant QTL associated with CBSD, CMD, and CGM detected across environments

Significant QTL were declared based on LOD threshold of above 2.5 and their stability across at least two environments (Table 3.5). To be declared 'real' QTL had to have a LOD above 2.5 in two environments (sites or seasons) and have exactly the same flanking markers. If the QTL was consistent across root necrosis and foliar symptoms in one or two environments only, it was declared a putative QTL, prefixed by "p." Other putative QTL had closely located, but not identical, flanking markers. Based on these criteria, three QTL associated with CBSD root necrosis, namely qCBSDRNc5K, qCBSDRNFc11K, and qCBSDRNc12K, were detected on chromosomes V, XI, and XII, with maximum LOD-values of 6.20, 13.45, and 11.05, respectively, explaining up to 10.1% of the phenotypic variation (PVE%) (Figure 3.7, 3.8 and 3.9 respectively). Seven QTL associated with CBSD foliar symptoms only, namely qCBSDFc4KL, qCBSDFc4R, qCBSDFc6KRa and b, qCBSDFc17K, qCBSDFc18Ka and b, were detected on chromosomes IV, VI, XVII and XVIII, respectively. They have maximum LODs of 2.78, 60.67, 54.75, 20.92, 27.01, 23.72 and 23.08, respectively, and explained up to 8.45% of the variation (Table 3.5). Two QTL were associated with both root necrosis and foliar symptoms, namely qCBSDRNFc11KR (Figure 3.10) and qCBSDFc15K, although the latter

was only associated with root necrosis in one environment (Naliendele 2014 -2015). Two QTL, namely qCMDc12Ar and qCMDc14Ar, associated with CMD resistance were detected on chromosomes XII and XIV (Table 3.5), with maximum LOD scores of 13.20 and 4.41, and maximum PVE% of 13.01 and 13.36%. QTL associated with CGM, namely qCGMc5Ar and qCGMc10Ar, were detected on chromosomes V and X with maximum LODs of 20.19 and 24.03, and maximum PVE% of 10.56% and 10.08%. Seven putative QTL were also identified, five of these were associated with both CBSD root necrosis and foliar symptoms on chromosomes IV, VI and XI (Table 3.6). pqCBSDRNFc4KLa and b were closely located to one another on the genome and each occurred in one environment (N2 and C2 respectively) but for foliar symptoms and root necrosis. One QTL on the left arm of chromosome VI (pqCBSDRNFc6KL) had inconsistent flanking markers for different traits, and on the right arm of the same chromosome, close to a consistent QTL, qCBSDFc6KRb, the putative QTL pqCBSDRNFc6KR, occurred in two environments (N1 and N2) but for one root necrosis and one foliar symptom trait only.

A putative QTL region on the left arm of chromosome XI (pqQTLRNFc11KL) had inconsistent flanking markers stretching from cXI: 2970283 to cXI: 7719727 but was present for root necrosis in three environments, and foliar symptoms in a single environment (Naliendele 2014 -2015) (Table 3.6). The last putative CBSD foliar resistant QTL occurred on chromosome XI at around 19Mb (Table 3.6). One putative QTL for CGM occurred on chromosome X but with inconsistent flanking markers.

**Table 3.5 Significant QTL detected in Kiroba x AR37-80 mapping population**

| QTL name | Trait | Location | Chr. | Left Marker (v5.1) | Right Marker (v5.1) | LOD | PVE (%) | a | d | d/a |
|---|---|---|---|---|---|---|---|---|---|---|
| qCBSDFc4KL | CBSD_3 | C1 | 4 | cIV:2397127 | cIV:3389179 | 2.51 | 8.45 | 0.01 | 0.14 | 27 |
| | CBSD_3 | C2 | 4 | cIV:2397127 | cIV:3389179 | 2.78 | 4.19 | -0.1 | -0.22 | 2.15 |
| qCBSDFc4KR | CBSD_3 | N1 | 4 | cIV:12722062 | cIV:15281535 | 60.67 | 6.09 | 0 | 0.13 | - |
| | CBSD_6 | N1 | 4 | cIV:12722062 | cIV:15281535 | 20.77 | 7.38 | 0.01 | 0.42 | 72.95 |
| | CBSD_6 | N2 | 4 | cIV:12722062 | cIV:15281535 | 10.16 | 6.66 | 0.01 | 0.9 | 79.77 |
| qCBSDRNc5K | Root necrosis | C2 | 5 | cV:8584542 | cV:9172040 | 6.2 | 10.18 | -1.18 | -1.14 | 0.96 |
| | Root necrosis | N1 | 5 | cV:8297662 | cV:8525472 | 2.87 | 3.53 | -0.34 | -0.23 | 0.68 |
| qCBSDFc6KRa | CBSD_3 | N1 | 6 | cVI:12787606 | cVI:13554612 | 54.75 | 6.09 | 0 | 0.13 | - |
| | CBSD_6 | N1 | 6 | cVI:12787606 | cVI:13554612 | 24.12 | 6.49 | -0.01 | 0.55 | - 60.99 |
| | CBSD_6 | N2 | 6 | cVI:12787606 | cVI:13554612 | 13.19 | 6.36 | 0 | 0.83 | - |
| qCBSDFc6KRb | CBSD_6 | N1 | 6 | cVI:15579060 | cVI:16110806 | 20.92 | 7.26 | 0.46 | -0.49 | -1.06 |
| | CBSD_6 | N2 | 6 | cV1:15579060 | cVI:16110806 | 3.04 | 2.64 | 0.32 | -0.36 | -1.13 |
| qCBSDRNFc11K | Root necrosis | N1 | 11 | cXI:15686140 | cXI:15799548 | 3.26 | 7.81 | 0.06 | 0.77 | 13.97 |
| | Root necrosis | N2 | 11 | cXI:15686140 | cXI:15799548 | 13.45 | 3.12 | -0.02 | 0.78 | - 34.69 |
| | CBSD_6 | C1 | 11 | cXI:15686140 | cXI:15799548 | 3.06 | 12.28 | 0.05 | 0.43 | 9.17 |
| | CBSD_6 | N2 | 11 | cXI:15686140 | cXI:15799548 | 13.54 | 6.4 | -0.02 | 0.83 | - 36.39 |
| qCBSDRNc12K | Root necrosis | C1 | 12 | cXII:16811592 | cXII:17259527 | 3.31 | 8.97 | -0.15 | 0.44 | -3.06 |
| | Root necrosis | N1 | 12 | cXII:16407934 | cXII:16811592 | 7.42 | 8.02 | 0.03 | 0.78 | 27.57 |
| | Root necrosis | N2 | 12 | cXII:16407934 | cXII:16811592 | 11.05 | 3.11 | 0.02 | 0.78 | 33.71 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| qCBSDFc15K | CBSD_3 | N1 | 15 | cXV:4187935 | cXV:4668905 | 54.68 | 6.09 | 0 | 0.13 | - |
| | CBSD_3 | N2 | 15 | cXV:4187935 | cXV:4668905 | 32.68 | 4.49 | -0.01 | 0.49 | -55.14 |
| | CBSD_6 | N2 | 15 | cXV:4187935 | cXV:4668905 | 17.51 | 6.53 | -0.05 | 0.85 | -18.77 |
| | Root necrosis | N2 | 15 | cXV: 4187935 | cXV:4668905 | 7.96 | 2.62 | 0.01 | 0.89 | 66.98 |
| qCBSDFc17K | CBSD_6 | C2 | 17 | cXVII:18990126 | cXVII:19117961 | 2.73 | 6.09 | -0.68 | -0.84 | 1.23 |
| | CBSD_6 | N1 | 17 | cXVII:18990126 | cXVII:19117961 | 27.01 | 7.39 | -0.45 | -0.51 | 1.14 |
| qCBSDFc18Ka | CBSD_6 | C2 | 18 | cXVIII:5764853 | cXVIII:6089207 | 2.79 | 8.02 | 0 | 0.73 | - |
| | CBSD_6 | N1 | 18 | cXVIII:5764853 | cXVIII:6089207 | 23.72 | 6.59 | 0 | 0.58 | - |
| qCBSDFc18Kb | CBSD_6 | C2 | 18 | cXVIII:5328493 | cXVIII:5467691 | 2.96 | 7.17 | 0 | 0.75 | - |
| | CBSD_6 | N1 | 18 | cXVIII:5328493 | cXVIII:5467691 | 23.08 | 6.59 | 0 | 0.58 | - |
| qCGMc5Ar | CGM_3 | C1 | 5 | cV:8584542 | cV:9172040 | 20.19 | 6.48 | -0.4 | -0.44 | 1.09 |
| | CGM_3 | N2 | 5 | cV:8584542 | cV:9172040 | 4.27 | 10.56 | -0.17 | -0.2 | 1.16 |
| qCGMc10ArR | CGM_6 | C2 | 10 | cX:14135599 | cX:15319838 | 24.03 | 4.11 | -0.75 | -0.79 | 1.05 |
| | CGM_3 | N1 | 10 | cX:14135599 | cX:15319838 | 3.83 | 10.08 | -0.26 | -0.33 | 1.27 |
| qCMDc12Ar | CMD_3 | C2 | 12 | cXII:7399696 | cXII:9335575 | 13.2 | 3.09 | 0.03 | -0.72 | -27.31 |
| | CMD_6 | C1 | 12 | cXII:7399696 | cXII:9335575 | 3.36 | 13.01 | -0.15 | -0.22 | 1.44 |
| | CMD_3 | N2 | 12 | cXII:7399696 | cXII:12974694 | 13.2 | 3.09 | 0.03 | -0.72 | -27.31 |
| | CMD_6 | N2 | 12 | cXII:7399696 | cXII:12974694 | 3.36 | 13.1 | -0.15 | -0.22 | 1.44 |
| qCMDc14Ar | CMD_6 | N1 | 14 | cXIV:15032980 | cXIV:16101978 | 4.41 | 13.36 | 0.82 | -0.66 | -0.8 |
| | CMD_3 | C2 | 14 | cXIV:15032980 | cXIV:16101978 | 3.54 | 1.22 | -0.88 | 0.83 | -0.94 |

a = additive effects, d = dominance effects, d/a = QTL mode of action, all calculated according to Muchero et al., 2013.

**Table 3.6 Main putative QTL in the bi-parental mapping population between Kiroba and AR37-80**

| QTL name | Trait | Location | Chromosome | Left Marker (v5.1) | Right Marker (v5.1) | LOD | PVE (%) |
|---|---|---|---|---|---|---|---|
| pqCBSDRNFc4KLa | Root necrosis | N2 | 4 | cIV:4826000 | cIV:8748959 | 17.57 | 3.12 |
| | CBSD_6 | N2 | 4 | cIV:4826000 | cIV:8748959 | 12.51 | 6.57 |
| pqCBSDRNFc4KLb | Root necrosis | C2 | 4 | cIV:8833268 | cIV:8899624 | 2.64 | 8.46 |
| | CBSD_3 | C2 | 4 | cIV:8833268 | cIV:8899624 | 2.60 | 5.47 |
| pqCBSDRNFc6KL | Root necrosis | N2 | 6 | cVI:1126402 | cVI:4229496 | 12.67 | 3.10 |
| | Root necrosis | N1 | 6 | cVI:7580997 | cVI:8850841 | 6.22 | 6.83 |
| | CBSD_3 | N1 | 6 | cVI:5913911 | cVI:7196510 | 53.26 | 6.09 |
| pqCBSDRNFc6KR | Root necrosis | N2 | 6 | cVI:16482426 | cVI:16676368 | 12.83 | 3.10 |
| | CBSD_6 | N1 | 6 | cVI:16482426 | cVI:16676368 | 12.15 | 7.38 |
| pqCBSDRNc11KL | Root Necrosis | N1 | 11 | cXI:7483911 | cXI:7719727 | 3.94 | 2.81 |
| | Root Necrosis | N2 | 11 | cXI:4904507 | cXI:5653088 | 16.90 | 3.15 |
| | Root Necrosis | C1 | 11 | cXI:2970283 | cXI:4502176 | 2.69 | 17.75 |
| | CBSD_6 | N2 | 11 | cXI:4156929 | cXI:4502564 | 3.06 | 12.28 |
| pqCBSDRNFc11KR | CBSD_3 | C2 | 11 | cXI:19060880 | cXI:19576509 | 3.75 | 8.64 |
| pqCGMc10ArL | CGM_3 | N1 | 10 | cX:5019671 | cX:5019989 | 4.25 | 13.16 |
| | CGM_3 | C1 | 10 | cX:5019989 | cX:5188325 | 2.84 | 0.97 |

C1,Chambezi field experiment 2013 -2014, C2, Chambezi field experiment 2014 -2015, N1, Naliendele field experiment 2013 -2014, N2, Naliendele field experiment 2014 - 2015;PVE, phenotypic variation explained

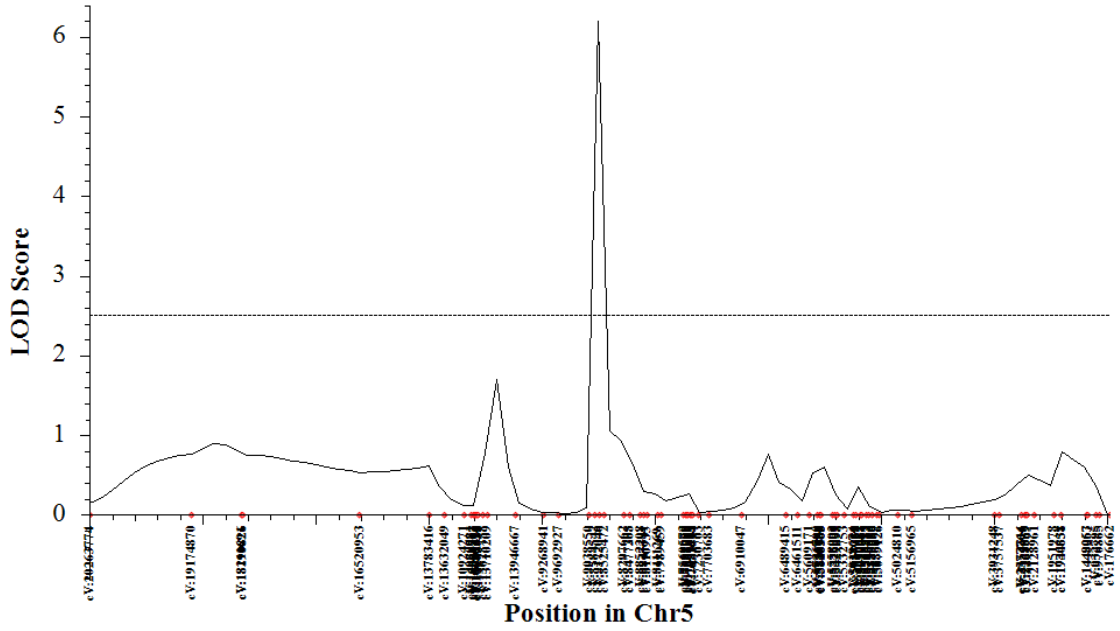



**Figure 3.7** LOD profile for QTL associated with CBSD root necrosis (qCBSDRNc5K) detected on chromosome V at Chambezi (C1). Broken horizontal line represents the LOD significance threshold of 2.5.

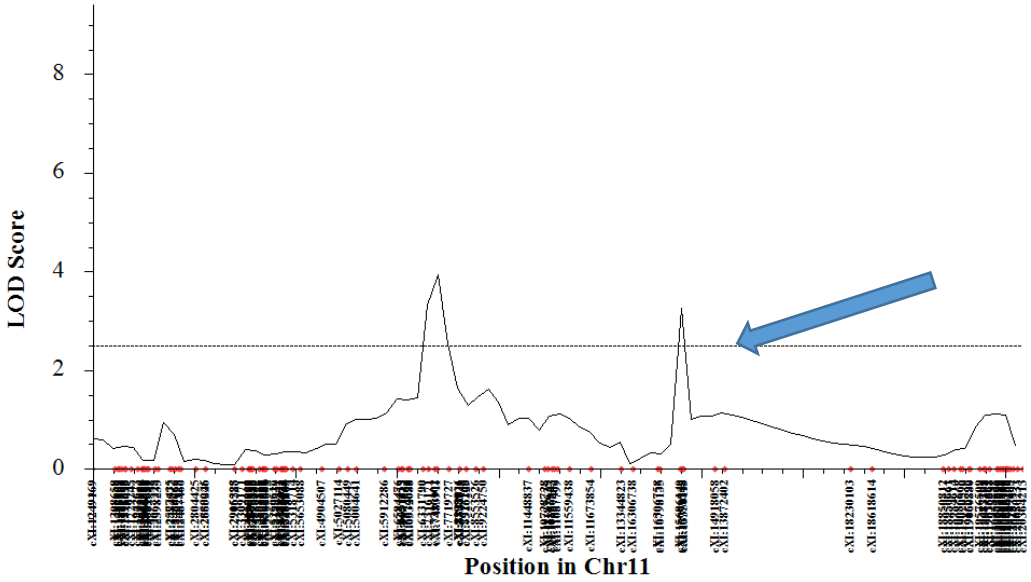



**Figure 3.8** LOD profile for QTL associated with CBSD root necrosis (q(CBSDRNc11KR) detected in chromosome 11 at Chambezi (C1). Broken horizontal line represents the LOD significance threshold of 2.5.

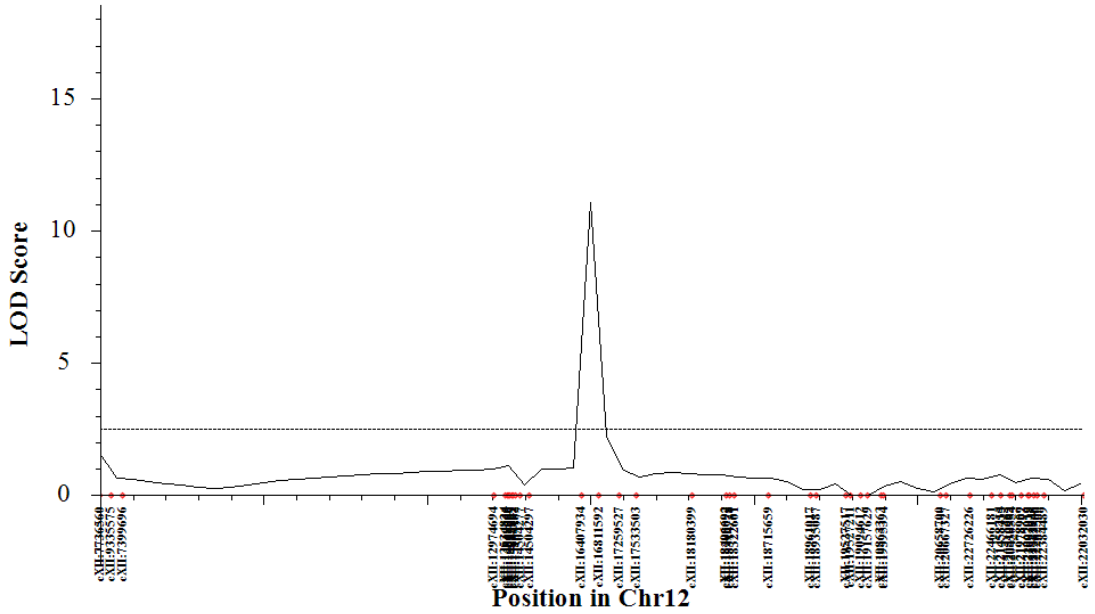



**Figure 3.9** LOD profile for QTL associated with CBSD root necrosis (qCBSDRNc12K) detected on chromosome XII at Chambezi (C1). Broken horizontal line represents the LOD significance threshold of 2.5.

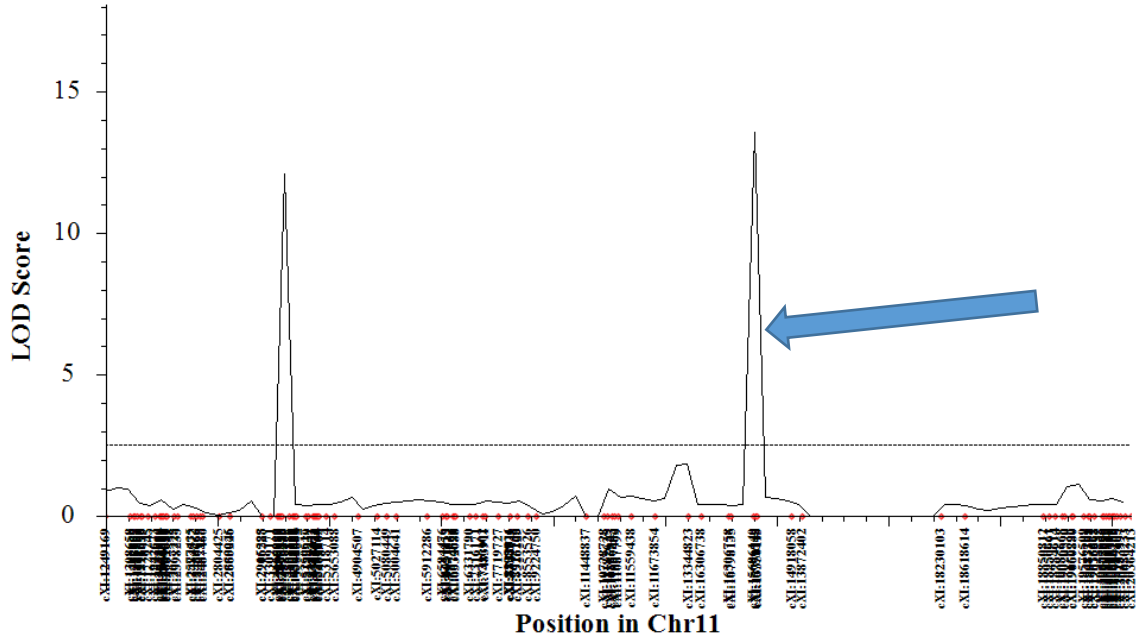



**Figure 3.10** LOD profile for QTL associated with CBSD foliar symptoms (qCBSDRNFc11KR) detected on chromosome XI at Naliendele (N1). The same QTL is associated with CBSD root necrosis and was detected at Naliendele and Chambezi across years and seasons.

### 3.4.5 Additive and dominant effects

Five QTL, namely qCBSDFc17K, qCBSDFc4KL and –R, qCGMc10ArR and qCGMc5Ar exhibited consistent evidence of over-dominance across sites whereas three QTL, qCBSDFc18Ka, -b and, qCMDc14Ar exhibited under-dominance across (Table 3.5). qCBSDRNc5K showed evidence of partial-dominance. The rest of the QTL were not consistent in their mode of action.

## 3.5 Discussion

### 3.5.1 Phenotypic analysis

CBSD disease distribution varied with site and season. CBSD foliar symptoms were higher in Chambezi for both phenotyping years; 2013–2014 and 2014–2015 (mean score 2.443 and 1.510) as compared to Naliendele (mean score 1.032 and 1.151). CBSD root necrosis was also higher in Chambezi for the two years (mean score 3.073 and 1.916) as compared to Naliendele (mean score 1.249 and 1.320). This shows that the CBSD disease pressure was higher in Chambezi as compared to Naliendele. All classes of scores (1–5) were represented in Chambezi for both leaf and root symptoms, but only classes 1–3 in Naliendele. The presence of different strains of virus in Chambezi or an increased population of whitefly may have contributed to this observation (Ndunguru et al., 2015). The phenotypic data in this study generally do not follow a perfect normal distribution and this is not surprising as it is a common phenomenon observed in many mapping populations (Zou, 2009). CBSD foliar symptoms were positively skewed toward class1 in both the years in Naliendele (skewness 3.327, 4.342) unlike in Chambezi where the distribution tended to be normally distributed (skewness 0.0229, 0.413). This is supported by the kurtosis analysis showing Naliendele being positively skewed (kurtosis 11.405 and 20.891) and Chambezi minimally skewed indicating tendency for normality (kurtosis <1 for all years). CBSD root necrosis follows a similar trend; Naliendele positively skewed (skewness, 2.047 and 3.569 kurtosis, 4.841 and 20.6190) and Chambezi minimally skewed with a tendency to be normally distributed (skewness and kurtosis <1). This shows that disease pressure and distribution was high in Chambezi in all the years but much lower in Naliendele

### 3.5.2 QTL associated with CBSD resistance

Eleven QTL associated with CBSD resistance were identified using ICIM; two were associated with CBSD root necrosis and seven with CBSD foliar symptoms, and two with both foliar symptoms and root necrosis. These QTL were significant by LOD score (>2.5) and were detected in at least two environments. Given their relative stability, the QTL derived in this mapping population offer valuable targets for further characterization and utilization through genomic based approaches including marker-assisted breeding and genomic selection for improving cassava productivity. The QTL associated with root necrosis were found on chromosome V (qCBSDRNc5K located between 8.2–9.1 Mbp), XI (qCBSDRNFc11KR located between 15.6–15.7 Mbp), XII (qCBSDRNc12K located between 16.4–17.2 Mbp) and XV (qCBSDRNFc15K located between 4.1 and 4.6 Mbp) (v5.1) with the highest LOD across years and environments in each case being 6.2, 13.45, 11.05 and 7.96, respectively. Interestingly the QTL associated with CBSD root necrosis, namely qCBSDRNFc11KR, at N1 and N2, mapped to the same interval and was flanked by the same markers as the one identified for CBSD foliar symptoms on chromosome XI detected at C1 and N2. This seems to suggest that CBSD root necrosis and CBSD foliar symptoms are influenced to some extent, but not exclusively, by the same gene(s), or by closely linked genes at this locus. This QTL region ranges from 21756682 to 21873545 bp on v6.1 of the cassava reference genome and lies within a CBSD resistance region (19872319–23751929 bp) (v6.1) on chromosome XI identified through genome wide association studies (GWAS) of populations in Uganda, partly derived from seed of Tanzanian origin (Kawuki et al., 2016). In those GWAS studies, four SNPs in this region were found to be associated with root severity and/or a disease index, and one SNP associated with maximum root severity. The fact that this QTL and the QTL observed here in Kiroba are located in the same region indicates that Kiroba or close relative(s) sharing the same QTL may be the main source of resistance to CBSD root necrosis used in the Ugandan germplasm.

Interestingly qCBSDRNFc11KR is in a different position (15.6 to 15.7 Mbp) from the main QTL associated with CBSD root necrosis found in another Tanzanian farmer cultivar, Namikonga (4.5 to 4.7Mbp) (Masumba et al., 2017). In this study however, we did identify a region of putative QTL for root necrosis (N1, N2 and C1) and foliar symptoms (N2) with inconsistent flanking markers ranging from cXI: 2970283 to cXI:7719727 (Table 4.6).

Five out of the 15 QTL identified were location specific, occurring in both years at one location (qCBSDFc4KL in Chambezi only; qCBSDFc4KR qCBSDFc6KRa and -b and qCBSDRNFc15K in Naliendele only). This could indicate genotype x environment interaction since our data show the disease distribution was not consistent across the sites. This may have also been influenced by the presence of different CBSVstrains in Chambezi and Naliendele experimental sites (Ndunguru et al., 2015). It is interesting to note that qCBSDRNFc11K was only found in Naliendele for root necrosis in both years, but was significant for foliar symptoms across locations.

### 3.5.3 QTL associated with CMD and CGM resistance

QTL associated with CMD, namely qCMDc12Ar and qCMDc14Ar, were detected on chromosomes XII and XIV. qCMDc12Ar [7399696–9335575 bp (v5.1) equivalent to 7929400 8645361 bp (v6.1)] coincides with the major resistance locus (3.56–11.38 Mb) (v6.1) identified on chromosome XII (Wolfe et al., 2016) and confirmed as the QTL linked to the CMD2 locus (Akano et al., 2002). In that study a major peak was found at 7926132 bp and is similar to the one reported for Namikonga (Masumba et al., 2017). This is a major resistance locus that has been previously cited on scaffold 06906 and 05214 (Rabbi et al., 2014). The markers SSRY28 and NS158 that are linked to the CMD2 locus (Akano et al., 2002; Lokko et al., 2005; Fregene et al., 2006; Okogbenin et al., 2007; Alabi et al., 2011) lie within qCMDc12Ar reported in this study. A CMD3 locus has also been reported on the same chromosome (Okogbenin et al., 2012) and is flanked by the marker NS198 which is approximately 36 cM away from the CMD2 locus. The presence of an additional interactive locus close to CMD2 has led to the postulation of multiple epistatic loci for CMD resistance or multiple alleles (Wolfe et al., 2016).The male parent of the mapping population studied here, AR37-80, was selected for CMD2 at CIAT using molecular markers (Blair et al., 2007; Okogbenin et al., 2007).

Most studies on CGM have focused on conventional breeding (Chipeta et al., 2013; Chalwe et al., 2015), with limited published work on molecular breeding (Macea Choperena et al., 2012). Two SSR markers, NS1099 (chromosome XIV) and NS346 (chromosome XVIII), showed high association with resistant families after screening five families each comprising 30

individual genotypes with 500 markers using bulk segregant analysis (BSA) (Ceballos et al., 2010; Macea Choperena et al., 2012). In this study we report two QTL linked to CGM resistance, namely qCGMc5Ar and qCGMc10Ar, on chromosomes V and X, with maximum LOD of 20.19 at C1 (with PVE 6.48) and 24.03 at C2 (with PVE of 4.11). These results, if validated and translated into marker-assisted breeding strategies, will complement conventional breeding approaches to improve cassava cultivars for CGM resistance.

## 3.6 Conclusions

This study has identified four QTL linked to resistance to CBSD root necrosis on chromosomes V, XI, XII, and XV, and although there appears to be a region on the left arm of chromosome 11 that confers resistance to root necrosis in both biparental populations derived from the resistant parents Namikonga and Kiroba, all other root necrosis QTL in Kiroba do not appear in Namikonga (Masumba et al., 2017). Once validated, this would provide opportunities for pyramiding of QTL for root necrosis for enhanced resistance and greater durability. Nine QTL linked to CBSD foliar symptoms have also been identified on chromosomes IV, VI, XVII, and XVIII, with QTL on chromosomes XI and XV being consistent between root necrosis and foliar symptoms. This observation suggests that resistance to CBSD foliar symptoms may be controlled by some of the same loci as for root necrosis, but additional loci are trait specific. A QTL associated with CMD resistance is consistent with the CMD2 locus on chromosome XII. All the QTL were detected using a small phenotyping population of 106 individuals and need to be validated on an expanded population size, to get greater accuracy. Once validated and applied through MAS or genomic selection, this study will aid in breeding cassava for multiple host plant disease resistance, contributing to food security and enhanced economic growth in Africa.

## 3.7 References


Akano, A., Dixon, A., Mba, C., Barrera, E., and Fregene, M. (2002). Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theor. Appl. Genet.* 105**,** 521-525.

Alabi, O., Rayapati, N., and Kumar, L. (2011). *Cassava mosaic disease: A curse to food security in Sub-Saharan Africa.*

Alicai, T., Omongo, C., Maruthi, M., Hillocks, R., Baguma, Y., Kawuki, R., et al. (2007). Re-emergence of cassava brown streak disease in Uganda. *Plant. Dis.* 91**,** 24-29.

Barkan, A., and Small, I. (2014). Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* 65**,** 415-442.

Blair, M.W., Fregene, M.A., Beebe, S.E., and Ceballos, H. (2007). "Marker-assisted selection in common beans and cassava," in *Marker-assisted selection: Current status and future perspectives in crops, livestock, forestry and fish,* ed. J.R. Guimaraes, B.D. Scherf, A. Sonnino, J.D. Dargie. (Rome: FAO), 81-115.

Campo, B.V.H., Hyman, G., and Bellotti, A. (2011). Threats to cassava production: known and potential geographic distribution of four key biotic constraints. *Food. Secur.* 3**,** 329-345.

Ceballos, H., Okogbenin, E., Pérez, J.C., López-Valle, L.A.B., and Debouck, D. (2010). "Cassava," in *Handbook of Plant Breeding: Root and Tuber Crops,* ed. J.E. Bradshaw. Springer), 53-96.

Chalwe, A., Melis, R., Shanahan, P., and Chiona, M. (2015). Inheritance of resistance to cassava green mite and other useful agronomic traits in cassava grown in Zambia. *Euphytica.* 205**,** 103-119.

Chipeta, M.M., Bokosi, J.M., Saka, V.W., and Benesi, I.R. (2013). Combining ability and mode of gene action in cassava for resistance to cassava green mite and cassava mealy bug in Malawi. *Plant. Breed. Crop Sci.* 5**,** 195-202.

Du, L., and Chen, Z. (2000). Identification of genes encoding receptor-like protein kinases as possible targets of pathogen-and salicylic acid-induced WRKY DNA-binding proteins in Arabidopsis. *Plant. J.* 24**,** 837-847.

Eulgem, T., Rushton, P.J., Robatzek, S., and Somssich, I.E. (2000). The WRKY superfamily of plant transcription factors. *Trends. Plant. Sci.* 5**,** 199-206.

FAO (2013). "Statistical database of the Food and Agriculture Organization of the United Nations". ( http://faostat.fao.org).

FAO (2015). "Statistical database of the Food and agriculture organization of the United Nations. ". (http://faostat.fao.org).

Ferguson, M.E., Hearne, S.J., Close, T.J., Wanamaker, S., Moskal, W.A., Town, C.D., et al. (2012). Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. *Theor. Appl. Genet.* 124**,** 685-695.

Fregene, M.A., Morante, N., Sánchez, T., Marín, J.A., Ospina, C., Barrera Sabogal, E., et al. (2006). "Molecular markers for introgression of useful traits from wild *Manihot* relatives of cassava, marker-assisted selection (MAS) of disease and root quality traits". (CGSpace: CIAT).

Hillocks, R., and Jennings, D. (2003). Cassava brown streak disease: a review of present knowledge and research needs. *Pest Sci.* 49**,** 225-234.

Hillocks, R., and Thresh, J. (2000). Cassava mosaic and cassava brown streak virus diseases in Africa: a comparative guide to symptoms and aetiologies. *Roots* 7**,** 1-8.

ICGMC (2015). High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta* Crantz) from 10 populations. *G3* 5**,** 133-144.

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., et al. (2002). bZIP transcription factors in Arabidopsis. *Trends Plant Sci.* 7**,** 106-111.

Kawuki, R., Herselman, L., Labuschagne, M., Nzuki, I., Ralimanana, I., Bidiaka, M., et al. (2013). Genetic diversity of cassava (*Manihot esculenta* Crantz) landraces and cultivars from Southern, Eastern and Central Africa. *Plant. Genet. Resour.* 11**,** 170-181.

Kunkeaw, S., Tangphatsornruang, S., Smith, D.R., and Triwitayakorn, K. (2010). Genetic linkage map of cassava (*Manihot esculenta* Crantz) based on AFLP and SSR markers. *Plant. Breed.* 129**,** 112-115.

Kuroda, H., Takahashi, N., Shimada, H., Seki, M., Shinozaki, K., and Matsui, M. (2002). Classification and expression analysis of Arabidopsis F-box-containing protein genes. *Plant. Cell Physiol.* 43**,** 1073-1085.

Lebot, V. (2009). *Tropical root and tuber crops: cassava, sweet potato, yams and aroids.* Cambridge, USA: CABI.

Legg, J.P., Jeremiah, S., Obiero, H., Maruthi, M., Ndyetabula, I., Okao-Okuja, G., et al. (2011). Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Res.* 159**,** 161-170.

Lokko, Y., Danquah, E., Offei, S., Dixon, A.G., and Gedil, M. (2005). Molecular markers associated with a new source of resistance to the cassava mosaic disease. *Afri. J. Biotechnol.* 4**,** 873-881.

Lozano, R., Hamblin, M., Prochnik, S., and Jannink, J.-L. (2015). Identification and distribution of the NBS-LRR gene family in the Cassava genome. *BMC Genomics* 16**,** 360.

Macea Choperena, E.d.P., Ospina, C., Fregene, M., Montoya-Lerma, J., and Bellotti, A.C. (2012). Microsatellite identification associated with resistance to the mite *Mononychellus tanajoa (Acari: Tetranychidae)*. *Rev. Bras. Entomol.* 38**,** 70-75.

Manichaikul, A., Palmer, A.A., Sen, Ś., and Broman, K.W. (2007). Significance thresholds for quantitative trait locus mapping under selective genotyping. *Genet.* 177**,** 1963-1966.

Masumba, E.A., Kapinga, F., Mkamilo, G., Salum, K., Kulembeka, H., Rounsley, S., et al. (2017). QTL associated with resistance to cassava brown streak and cassava mosaic diseases in a bi-parental cross of two Tanzanian farmer varieties, Namikonga and Albert. *Theor. Appl. Genet.* 130:2069-2090

Muchero, W., Sewell, M.M., Ranjan, P., Gunter, L.E., Tschaplinski, T.J., Yin, T., et al. (2013). Genome anchored QTLs for biomass productivity in hybrid Populus grown under contrasting environments. *PLoS ONE* 8**,** e54468.

Mponda, E.M Kafiriti, Wembah-Rashid, M.G. Kajimbwa, E. Ndedya, L., Kinyunyu, H., et al. (2001). "Farmer seed experts and their knowledge on seeds, forest food products andmedicinal plants", (ed.) D.R.D. Ministry of agriculture and Foodsecurity. (Tanzania: Naliendele Agricultural Research Institute ).

Ndunguru, J., Sseruwagi, P., Tairo, F., Stomeo, F., Maina, S., Djinkeng, A., et al. (2015). Analyses of twelve new whole genome sequences of cassava brown streak viruses and Ugandan cassava brown streak viruses from East Africa: Diversity, supercomputing and evidence for further speciation. *PLoS ONE* 10**,** e0139321.

Okogbenin, E., Egesi, C., Olasanmi, B., Ogundapo, O., Kahya, S., Hurtado, P., et al. (2012). Molecular marker analysis and validation of resistance to cassava mosaic disease in elite cassava genotypes in Nigeria. *Crop Sci.* 52**,** 2576-2586.

Okogbenin, E., Porto, M., Egesi, C., Mba, C., Espinosa, E., Santos, L., et al. (2007). Marker-assisted introgression of resistance to cassava mosaic disease into Latin American germplasm for the genetic improvement of cassava in Africa. *Crop Sci.* 47**,** 1895-1904.

Poland, J.A., and Rife, T.W. (2012). Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome.* 5**,** 92-102.

Prikryl, J., Rojas, M., Schuster, G., and Barkan, A. (2011). Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc. Natl. Acad. Sci.* 108**,** 415-420.

Rabbi, I.Y., Hamblin, M.T., Kumar, P.L., Gedil, M.A., Ikpan, A.S., Jannink, J.-L., et al. (2014). High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. *Virus Res.* 186**,** 87-96.

Rabbi, I.Y., Kulembeka, H.P., Masumba, E., Marri, P.R., and Ferguson, M. (2012). An EST-derived SNP and SSR genetic linkage map of cassava (*Manihot esculenta* Crantz). *Theor. Appl.Genet.* 125**,** 329-342.

Ripatti, S., Becker, T., Bickeböller, H., Dominicus, A., Fischer, C., Humphreys, K., et al. (2009). Genestat: an information portal for design and analysis of genetic association studies. *E. J. Hum. Genet.* 17**,** 533-536.

Rubel, F., and Kottek, M. (2010). Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorologische Zeitschrift* 19(2)**,** 135-141.

Van den Burg, H.A., Tsitsigiannis, D.I., Rowland, O., Lo, J., Rallapalli, G., MacLean, D., et al. (2008). The F-box protein ACRE189/ACIF1 regulates cell death and defense responses activated during pathogen recognition in tobacco and tomato. *Plant Cell* 20**,** 697-719.

Wolfe, M.D., Rabbi, I.Y., Egesi, C., Hamblin, M., Kawuki, R., Kulakow, P., et al. (2016). Genome-Wide association and prediction reveals genetic architecture of cassava mosaic disease resistance and prospects for rapid genetic improvement. *Plant Genome.* 9.

Xu, S. (2003). Theoretical basis of the Beavis effect. *Genet.* 165**,** 2259-2268.

Zhang, L., Meng, L., Li, H., and Wang, J. (2015). GACD: Integrated Software for Genetic Analysis in Clonal F1 and Double Cross Populations. *J. Hered.* 106**,** 269-283.

Zou, F. (2009). QTL mapping in intercross and backcross populations. *Methods Mol. Biol.* 573**,** 157-173.

# Chapter 4

# Identification of introgression regions in Kiroba derived from *Manihot glaziovii and* comparison with other genotypes of African origin


**Authors**

Inosters Nzuki[1], Manpreet S. Katari[2], Jessen V. Bredeson[3], Trushar Shah[4] , Steve Rounsley[5]Alexander A. Myburg[1], Nicholas van der Merwe[1] and Morag E. Ferguson*[4]

**Affiliations**

[1]Department of Genetics, University of Pretoria, Private bag X20, Hatfield 0028, South Africa

[2]New York University, New York, USA.

[3]Molecular and Cell Biology Department, University of California, Berkeley, California, USA

[4]International Institute of Tropical Agriculture (IITA), P.O. Box 30709-00100, Nairobi, Kenya

[5]Genus plc, DeForest WI USA.

* Corresponding author: m.ferguson@cgiar.org, (+254) 733524685

## Author contributions

This chapter has been published in the journal Frontiers in Plant Sciences under the tittle 'QTL mapping for pest and disease resistance in cassava and coincidence of some QTL with introgression regions derived from *Manihot glaziovii*" published online in July 2017 (Front. Plant Sci. 8:1168. doi: 10.3389/fpls.2017.01). Inosters Nzuki conducted the study, analyzed the data and wrote the manuscript. Morag Ferguson conceived the project and sourced for funding of the study. Jessen Bredesson and Manpreet S. Katari provided training and technical support on introgression analysis. Steve Rounsley provided scientific and bioinformatics support and guidance. Prof. Alexander A. Myburg, Dr. Morag Ferguson and Dr. Nicolas van der Merwe provided scientific guidance. All authors read and approved the manuscript.

**4**

## 4.1 Abstract

The genome of Kiroba was assembled at very high coverage of 91.8 % based on re-sequencing data available from GenBank. It contains genomic segments that are consistently the same as in *M. glaziovii* but alternatively different from *M. esculenta.* These genomic regions are more like *M. glaziovii* and are concluded to have been introgressed into Kiroba. There are large *M. glaziovii* introgression regions in Kiroba on chromosomes I, XVII and XVIII. The introgression regions contain significant annotation terms that have known orthologues in Arabidopsis which function in pathogen response. The introgression segments on chromosomes XVII and XVIII overlap with QTL associated with CBSD foliar symptoms. The QTL region on chromosome XVII and XVIII contains significant genes that have functions related to disease resistance in plants. They encode proteins with similarity to the nucleotide binding site and other domains characteristic of plant defence proteins including leucine-rich repeat receptor like protein kinases (LRR) that have been reported to be responsible for pathogen recognition in plants. The QTL regions lie perfectly within the introgression regions and may be deemed to have been derived from *M. glaziovii* and are responsible for the observed field resistance against pests and diseases in Kiroba. The introgression region on chromosome I is of a different haplotype to the characteristic "Amani haplotype" found in the landrace Namikonga and unlike some other genotypes, Kiroba does not have a large introgression block on chromosome IV. Kiroba is closely related to a sampled Tanzanian "tree cassava". This supports the observation that some of the QTL associated with CBSD resistance in Kiroba are different to those observed in Namikonga.

**Keywords;** Field resistance, genome introgression Kiroba, *M. glaziovii,*

## 4.2 Introduction

Cassava breeders have identified cultivars that show strong field resistance against viral diseases. The source of this resistance is assumed to be from the Amani breeding program of the 1930's (Hillocks and Jennings, 2003;Jennings, 1994) whose breeding materials may have found their way into farmer's fields after closure the program and are being grown under different names (Kanju et al., 2002). Kiroba is a farmer cultivar grown in coastal Tanzania (Muhanna and Mtunda, 2003) and is thought to be one of the former Amani hybrid that may have lost its identity. Kiroba does not have any visible wild characteristics, shows mild CBSD leaf symptoms in Tanzania, but does not show any leaf symptoms in Uganda (Kaweesi et al., 2014), with almost no root necrosis even under very high disease pressure. A diversity assessment using SNP markers shows Kiroba to be quite closely related to a known Amani derivative, the landrace Namikonga (Ferguson et al., 2012). Namikonga is known to confer good tolerance to CBSD and putative QTL for CBSD have already been identified (Masumba et al., 2017). It is an inter-specific hybrid grown by farmers in Tanzania (Rwegasira and Rey, 2012) and in Kenya, Namikonga is referred to as "Kaleso" which according to Dr Jennings, former Amani breeder, this was accession number 46106/27 (Jennings, 1994) and was a third backcross derivative from *M. esculenta x M. glaziovii*. It is hypothesized that Kiroba may also be a derivative from a *M. glaziovii* × *M. esculenta* interspecific cross. Nachinyaya is another landrace from southern Tanzania also known as Cabo Delgado in Mozambique (Thresh, 2003). It has been noted to have delayed onset of necrosis and the stems resemble wild species (Hillocks and Thresh, 2000), it is related to Amani clones (Bredeson et al., 2016) and possibly is an interspecific hybrid derived from *M. glaziovii* or *M. melanobasis*. These cultivars have the potential to be utilized in cassava breeding programs to introgress CBSD resistance to farmer-preferred cultivars like Albert or promising breeder's lines such as AR37-80. The source and mechanism of their resistance and their pedigrees is not clearly understood. Genetic and genomic approaches are useful in identifying genomic regions that confer resistance in these cultivars, information that is vital for the efficient movement of these traits in breeding, using MAS and GS approaches.

The emergence of new strains of virus (Ndunguru et al., 2015) present a challenge to the already existing resistant cultivars and researchers have to look for alternative sources of resistance including tapping into the rich genetic reservoir of the wild progenitors of cassava and its derivatives to broaden the genetic base of the farmer preferred cultivars. The genetics and

evolution of cassava is associated with its wild progenitors (Olsen and Schaal, 2001). The genes for host plant viral resistance including CMD and CBSD in the cultivars above may have originated through genome hybridization. Genomic characterization of these cultivars to identify introgressed genomic segments and discover resistance (R) genes may unravel the source and genomic location of their resistance allowing molecular breeders to develop functional gene-targeted markers that can be used in resistance breeding to combat CBSD and CMD (Gedil et al., 2012).

The genome assembly of cassava is now available (Bredeson et al., 2016) and scaffolds and contigs have been anchored into chromosomes through the use of the consensus map (ICGMC, 2015). This along with QTL data being made available by researchers will allow breeders to use a targeted approach in cassava breeding as they know which genes or section of genome to target. This will eventually accelerate deployment of resistant cultivars with efficiency and precision using modern approaches that circumvent the long breeding cycle of cassava and its inherent challenges of heterozygosity and outcrossing.

The objective of this study was (1) to determine whether there are genomic regions in 'Kiroba' derived from *M. glaziovii* (2) if so, where does the QTL identified earlier in this study lie in respect to the introgression regions (3) are the genes present within QTL regions overlapping introgression regions associated with diseases resistance?

## 4.3 Materials and Methods

### 4.3.1 Re-sequencing

The cultivars for this study were; *M. glaziovii*, a wild species of *Manihot*, also known as Ceará or India rubber, Albert, a Tanzanian cultivar said to be a pure *M. esculenta*, and Kiroba, one of the parents used in the mapping population reported in this study. DNA extractions, library preparations and sequencing was done as described in Bredeson et al. (2016). SRA BioSample accession numbers are as follows: Kiroba (SAMN02693378), Albert (SAMN04117017), *M. glaziovii* (SAMN02693380).

### 4.3.2 Identification of genomic regions in Kiroba derived from *M. glaziovii*

Sequence quality assessment was done using FastQC (Patel and Jain, 2012). The first 10 bases were trimmed using fastx trimmer, and then de novo assembly performed using abyss-pe (Simpson et al., 2009). Default parameters were used with a k-mer of 64. The purpose of assembling the Kiroba genome was to obtain high quality scaffolds for alignment and SNP analysis. The quality of the Kiroba assembly was assessed by N50 length statistics derived from the abyss-pe output. Based on the results of the assembly, scaffolds and contigs smaller than 200 bp were discarded to avoid using low quality reads in subsequent analysis. Assemblies of *M. glaziovii* and Albert (*M. esculenta*), were downloaded from Phytozome v.10 (Goodstein et al., 2012), and together with the Kiroba assembly, were aligned to v5.1 of the cassava genome assembly, using Bowtie2 (Langmead and Salzberg, 2012) followed by SNP calling using GATK (McKenna et al., 2010).

### 4.3.3 Calculating a score for possible introgression sites

The genotype information from the VCF file (the output of GATK pipeline discussed above) was coded as follows; 0/0 = 0, meaning homozygous to the reference, 0/1 =1, heterozygous to the reference and 1/1 = 2, homozygous alternative. Any SNP with a missing value (−/−) for any genotype was removed. For each SNP, the absolute value for difference between the *M. glaziovii* and Kiroba (GK) and the difference between Albert and Kiroba (AK) scores were calculated. The final score for the SNP was determined using the equation: Diffscore = AK+ (2 − GK). Any SNP Diffscore value >2 was regarded as indicative of introgression. A loop was created to look at 1,000 bp at a time starting at position 1. The tail of the contig <1,000 bp was not analyzed. Poisson test was used to identify enriched regions of the genome, indicative of introgression (scores >2). The *P* values were corrected for multiple hypothesis testing using a false discovery rate (FDR) with a 10% cut off to identify fragments that are significant and at least 1,000 bp.

### 4.3.4 Identification of genes present in the putative introgression regions and within the detected QTL regions

Putative introgression regions were aligned to the cassava reference genome v6 (Bredeson et al., 2016) using BLAST (Altschul et al., 1990). Gene lists containing functions and annotations were obtained from the top BLAST matches. Functional categories such as GO-Terms, PFAM domains, KOG, and PANTHER, were used for gene set enrichment analysis. Hypergeometric test with FDR correction cutoff of 10% (Benjamini and Hochberg, 1995) was performed to determine the significance of the functional term enrichment. The gene list for genes within the QTL regions and introgression segments were analyzed and tested for significance based on *P*-value and FDR.

### *4.3.5* Comparative analysis of Kiroba and Namikonga Haplotype derived from *M. glaziovii*

Chromosomal locations of introgression regions found in Kiroba were compared to those found in other genotypes by incorporating Kiroba into an earlier analysis (Bredeson et al., 2016). Comparative genome analysis was done to investigate whether Kiroba shared the same *M. glaziovii* haplotype as Namikonga and other genotypes including TME 117 thought to be the source of CBSD resistance in Namikonga. Kiroba is thought to be a possible former Amani interspecific hybrid just like Namikonga and is postulated to have found its way into farmer fields and now being grown with unknown identity in Tanzania.

### 4.3.6 Genome-wide relatedness

Genetic relatedness analysis between Kiroba and 40 other accessions (Goodstein et al., 2012; Bredeson et al., 2016) including Namikonga and tree cassava was performed using the kinship coefficient $\hat{\pi}$, and identity by descent (IBD) probabilities computed with PLINK (Purcell et al., 2007). The SNP data for all the accessions was obtained from Phytozome v.10 (Goodstein et al., 2012) and filtered using Vcftools. A network plot showing the first degree relatedness was drawn using Cytoscape (Shannon et al., 2003).

## 4.4 Results

### 4.4.1 Characterization of Kiroba Genome Assembly

Sequences were generated as paired-end reads from Illumina HiSeq 2500 and produced 310 million reads with 40X coverage (2 × 150 bp) of Kiroba, 257 million reads of 40X coverage of *M. glaziovii*, and 328 million reads with 42X coverage of Albert (*M. esculenta*) (Bredeson et al., 2016). The minimum contig size considered for *de novo* assembly into scaffolds was 500 bp (Table 4.1). The N50 scaffold length for the Kiroba assembly was 3,376 bp (Figure 4.1). N50 length is the length of the smallest contig (scaffold) for which the collection of all contigs of that length or longer contains at least half of the sum of the lengths of all contigs. The final assembly had 13.83 million reads that were assembled into scaffolds whose sizes add up to 161.3 million bases. The percent coverage of the Kiroba assembly (using scaffolds >500 bp) used in this analysis was approximately 21.2% (161/760) of the *M. esculenta* genome (~760 million bases). However, Figure 4.1 shows there were contigs as small as 25bp which when taken into consideration, 697 million bases are obtained giving 91.8% coverage of the genome.

**Table 4.1 N50 score and genome coverage**

| | N* | Min contig size (bp) | N50 | Max | Sum |
|---|---|---|---|---|---|
| **Contigs** | 13.89e6 | 500 | 2006 | 38372 | 152.7e6 |
| **Scaffolds** | 13.83e6 | 500 | 3376 | 76353 | 161.3e6 |

*N= Number of contigs and scaffolds

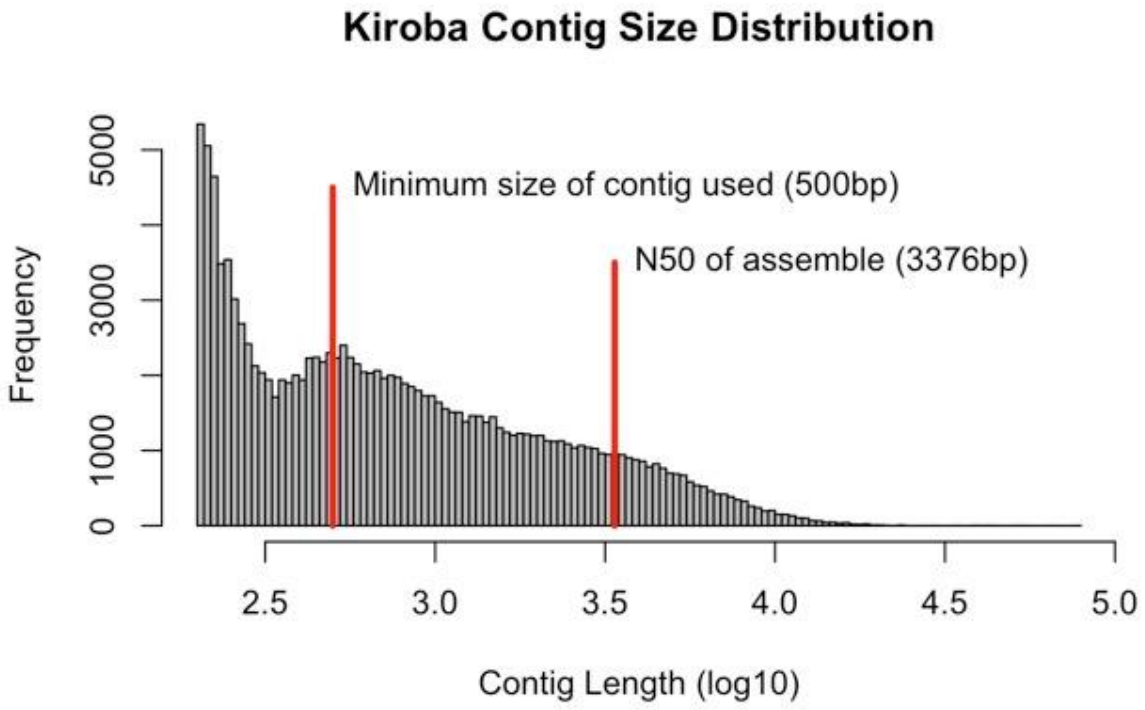



**Figure 4.1** Kiroba genome assembly showing the contig size in log10.

### 4.4.2 Genomic regions introgressed into Kiroba and their co-location with detected QTL regions

Large *M. glaziovii* introgression regions in Kiroba were found on chromosome 1 (24888098 33833841 bp), chromosome 17 (13700269–23436220 bp), and chromosome 18 (6805369–25209274 bp) (v6.1) of cassava genome (Figure 4.2).There are also small introgression segments dispersed within the genome (Figure 4.3) The QTL on chromosome 17, qCBSDFc17K (18990126–19117961 bp) and on chromosome 18, qCBSDFc18Kb (5764853–6089207 bp) associated with CBSD foliar symptoms lie perfectly within the large introgression regions, however qCBSDFc18Ka lies outside the introgression region. The QTL regions associated with CBSD root necrosis (qCBSDRNc5K, qCBSDRNFc11KR, and CBSDRNc12K), CMD (qCMDc12Ar and qCMDc14Ar), and CGM (qCGMc10Ar and qCGMc5Ar) are located on chromosomes 5, 10, 11, 12, and 14 that lack introgression regions derived from *M. glaziovii*. The introgression region in chromosome 18 encodes many protein domains that are associated with defence, including kinases, F-box family proteins which contain leucine-rich repeats (LRR), tetratricopeptide repeat (TPR)-like superfamily protein,

protein kinase superfamily proteins, and pentatricopeptide repeat (PPR) superfamily proteins (Table 4.3).

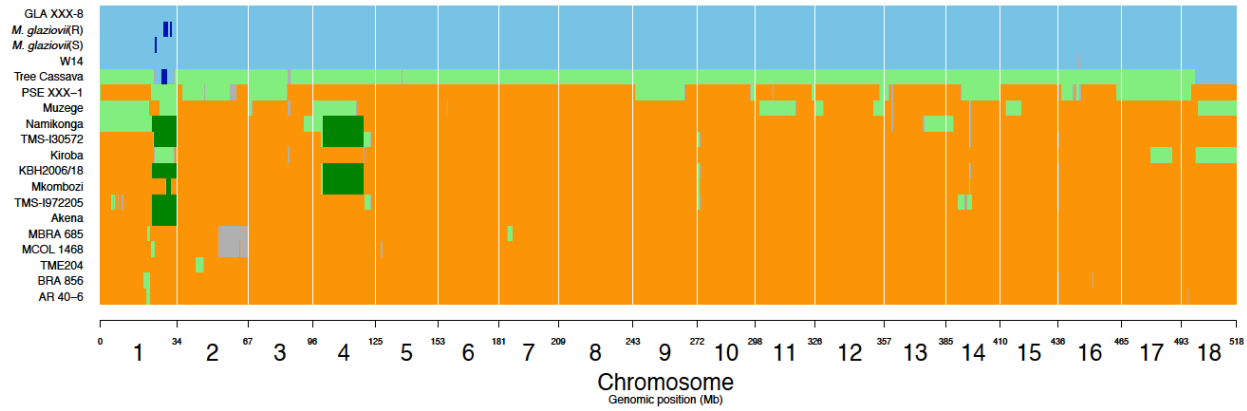



**Figure 4.2** *M. glaziovii* introgression segments in selected cassava genotypes, including Kiroba and Namikonga based on whole genome sequence data. Orange indicates *M. esculenta* genotype; blue indicates *M. glaziovii*, light green represents hybrid *M. esculenta* and *M. glaziovii*, dark green indicates the presence of a shared haplotype proposed to be inherited from the Amani program (modified from Bredeson et al., 2016; Permission for the modification and publication of the adapted figure has been obtained from the copyright holder).

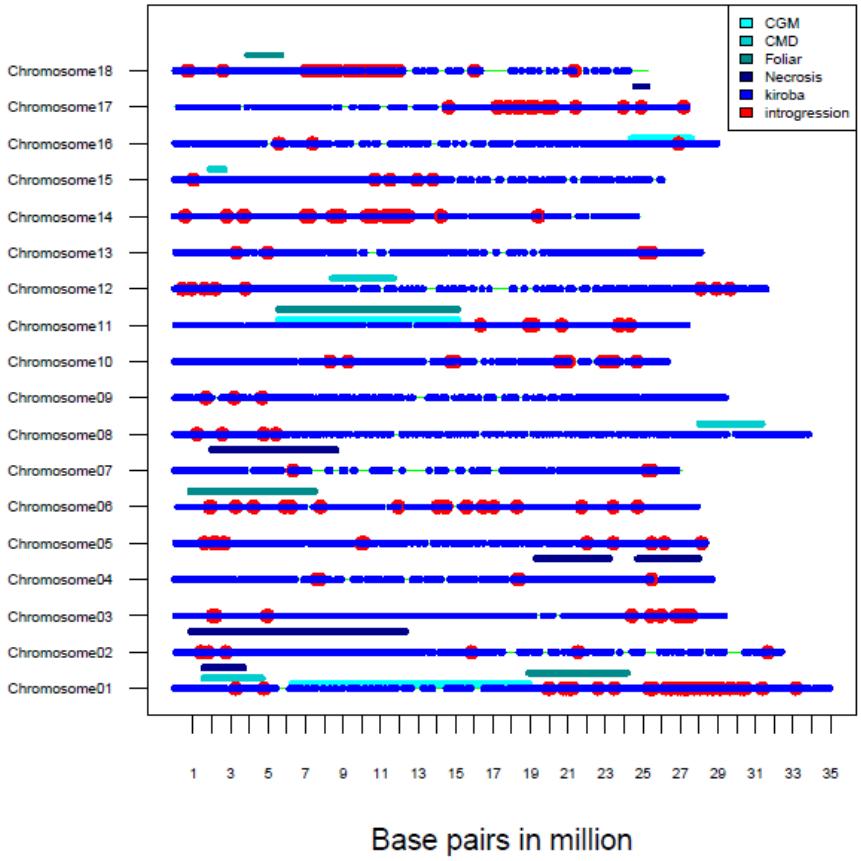



**Figure 4.3** Cassava chromosomes showing additional putative introgression sites and co-location with detected QTL.

### 4.4.3 Identification of *M. glaziovii* haplotype present in Kiroba and Namikonga

Six genotypes, Namikonga, TMS130572, KBH2006/18, Mkombozi, TMS 1972205, and Akena, share a common *M. glaziovii* haplotype on chromosome 1 (Figure 4.2 above, dark green), designated as the "Amani haplotype" (Bredeson et al., 2016). Namikonga, TMS130572, KBH2006/18, and Mkombozi also share a common haplotype on chromosome 4 (Bredeson et al., 2016). The *M. glaziovii* haplotype in Kiroba is different from that of Namikonga and the other five genotypes that share the same "Amani haplotype" on chromosomes 1 and 5.

### 4.4.4 Genome wide relatedness

Individuals sharing putative first-degree relationships were identified as pairs sharing a $\pi$ of 0.45 or greater (Table 4.2). The results extend an earlier study (Bredeson et al., 2016) with the addition of Kiroba. A network of first degree relatedness (Figure 4.3) results shows two "hub"

genotypes, namely TME117 and KBH 2006/18, each with many parent–offspring relationships. TME117 appears to be more related to East African germplasm such as Namikonga (with which it has a parent–offspring relationship), Albert, Ebwanatereka (EBW-A and EBW-2), and Kibaha, whereas KBH 2006/18 is more closely related to West African and South American germplasm such as TME204, TME419, AR40-6, and AR37-80. 'Kiroba', Muzege, and KBH 2006/18 have a parent–offspring relationship with tree cassava, with Muzege providing a link between the TME117 and KBH2006/18 hubs. Interestingly 'Kiroba' appears to be more related to the KBH2006/18 hub than the TME117 hub that contains most of the East African germplasm.

**Table 4.2 Identicals and first degree relationships**

| **Identicals** | | | | | |
|---|---|---|---|---|---|
| **Individual 1** | **Individual 2** | **IBD0 (Z0)** | **IBD1 (ZI)** | **IBD2 (Z2)** | $\hat{\pi}$ |
| EBW-2 | EBW-A | 0 | 0.0016 | 0.9983 | 0.999 |
| TME14K | TME3 | 0 | 0.0021 | 0.9979 | 0.999 |
| TME204 | TME419 | 0 | 0.002 | 0.9979 | 0.999 |
| MAUS7 | TMS50595 | 0 | 0.0029 | 0.9971 | 0.999 |
| TME3 | TME7 | 0 | 0.0077 | 0.9923 | 0.996 |
| TME14K | TME7 | 0 | 0.0085 | 0.9914 | 0.996 |
| **Likely parent -offspring** | | | | | |
| AR37-80 | KBH2006_18 | 0 | 1 | 0 | 0.5 |
| AR40-6 | KBH2006_18 | 0 | 1 | 0 | 0.5 |
| CM3306-4-5 | KBH2006_18 | 0 | 1 | 0 | 0.5 |
| COL1468-5 | KBH2006_18 | 0 | 1 | 0 | 0.5 |
| KBH2006_18 | Kibandameno | 0 | 1 | 0 | 0.5 |
| KBH2006_18 | MCOL22 | 0 | 1 | 0 | 0.5 |
| KBH2006_18 | Muzege | 0 | 1 | 0 | 0.5 |
| KBH2006_18 | TME204 | 0 | 1 | 0 | 0.5 |
| KBH2006_18 | TME419 | 0 | 1 | 0 | 0.5 |
| KBH2006_18 | TreeCassava | 0 | 1 | 0 | 0.5 |
| Kibandameno | Muzege | 0 | 1 | 0 | 0.5 |
| Kibandameno | Namikonga | 0 | 1 | 0 | 0.5 |
| Kiroba | TreeCassava | 0 | 1 | 0 | 0.5 |
| Muzege | TME117 | 0 | 1 | 0 | 0.5 |
| Muzege | TreeCassava | 0 | 1 | 0 | 0.5 |
| Namikonga | TME117 | 0 | 1 | 0 | 0.5 |
| **Likely full-sibling** | | | | | |
| Kibandameno | Me001Vu | 0.1745 | 0.6624 | 0.1631 | 0.494 |
| Albert | Kibandameno | 0.1731 | 0.6795 | 0.1474 | 0.487 |
| AVOCA | Kibandameno | 0.1508 | 0.7358 | 0.1133 | 0.481 |
| Kibandameno | TME117 | 0.1412 | 0.7742 | 0.0846 | 0.472 |
| Kibandameno | Me004Vu | 0.2212 | 0.6195 | 0.1593 | 0.469 |
| KBH2006_18 | TMS30572 | 0.1001 | 0.8749 | 0.025 | 0.463 |
| Albert | Nachinyaya | 0.0161 | 0.7668 | 0.2171 | 0.601 |
| Me003Vu | TME117 | 0.0027 | 0.8003 | 0.1969 | 0.597 |
| EBW-2 | TME117 | 0.0029 | 0.8283 | 0.1688 | 0.583 |
| EBW-A | TME117 | 0.0028 | 0.8284 | 0.1688 | 0.583 |
| Albert | TME117 | 0.0177 | 0.8027 | 0.1796 | 0.581 |
| CM3306-4-5 | MCOL22 | 0.0058 | 0.8347 | 0.1595 | 0.577 |
| Kibandameno | Me003Vu | 0.0836 | 0.7168 | 0.1996 | 0.558 |
| Mkombozi | TME14K | 0.0023 | 0.8874 | 0.1103 | 0.554 |
| Mkombozi | TME3 | 0.0014 | 0.8924 | 0.1061 | 0.552 |
| Mkombozi | TME7 | 0.0131 | 0.8784 | 0.1085 | 0.548 |
| TME204 | TME3 | 0.211 | 0.5031 | 0.2859 | 0.537 |
| EBW-2 | Kibandameno | 0.1308 | 0.7276 | 0.1416 | 0.505 |
| EBW-A | Kibandameno | 0.1309 | 0.7275 | 0.1415 | 0.505 |
| Kibaha | Kibandameno | 0.0194 | 0.6352 | 0.3455 | 0.663 |
| NDL06_132 | Nachinyaya | 0.0041 | 0.7078 | 0.2881 | 0.642 |
| AVOCA | TMS50595 | 0.004 | 0.741 | 0.2549 | 0.625 |
| AVOCA | MAUS7 | 0.0037 | 0.7428 | 0.2535 | 0.625 |
| Me001Vu | TME117 | 0.0026 | 0.7698 | 0.2276 | 0.613 |
| Me003Vu | TMS50595 | 0.0039 | 0.7824 | 0.2137 | 0.605 |
| MAUS7 | Me003Vu | 0.0035 | 0.7838 | 0.2127 | 0.605 |
| Kibaha | TME117 | 0.0028 | 0.7915 | 0.2057 | 0.602 |

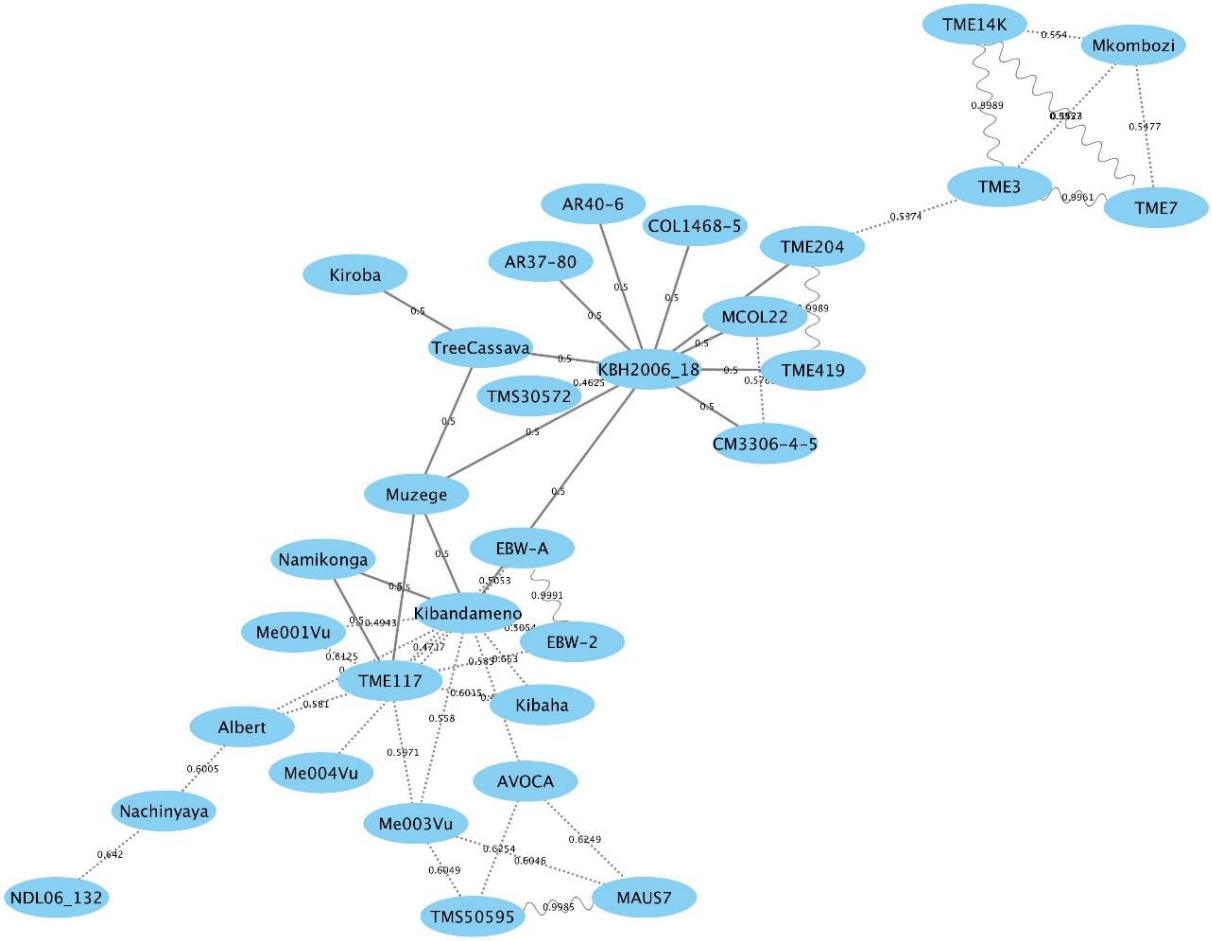



**Figure 4.4** Genome-wide relationship between Kiroba and other accessions. Each line drawa a relationship between accessions. Parent-offspring relationship is represented by a solid line, full-siblings by dotted lines and identicals by sinewave

### 4.4.5 Genes with significant pest and disease resistance functional annotation terms within QTL regions

Significant terms and their associated annotations related to disease resistance within the QTL regions were obtained (Table 4.3). They are significant by *P*-value and FDR correction. A total of 276 significant terms were detected within the QTL regions and 17 (6.2%) of them were related to pest or disease resistance in plants (Table 4.4). 7.8% of the terms detected within the QTL regions for root necrosis are related to disease resistance. 3.3% of terms detected within QTL regions were related to CBSD foliar symptoms, 8% within QTL regions related to CMD

resistance, and 3.3% detected within QTL regions related to CGM are associated with pest or disease resistance in plants. WRKY DNA-binding and leucine rich repeat (LRR) protein domains were found within QTL related to CBSD root necrosis on chromosome V and XII, while tetratricopeptide repeat (TPR) (chromosome VI), LRR (chromosome X), and F box and pentatricopeptide repeat (PPR) (chromosome XVIII) are present within QTL regions associated with CBSD leaf symptoms. WRKYDNA-binding domains are present in the QTL region on chromosome V detected for CGM, while LRR and NBARC domains are present in the QTL regions detected for CMD resistance on chromosome XIV. The significant genes found on chromosome XI were not related to disease resistance in plants. There are only seven genes between the flanking markers of QTL on chromosome XVII, and none are particularly associated with disease resistance.

### 4.4.6 Enrichment analysis in the areas of introgression

Enrichment analysis for functional terms related to disease resistance in plants is shown in Table 4.4. There are six terms that are significant by *P*-value and FDR correction ($P<0.05$). They contain genes with orthologues in Arabidopsis that have functions related to defence response, pathogen response and cell cycle during geminivirus infection.

**Table 4.3 Significant terms related to disease resistance, and candidate genes, found within the QTL**

| QTL | Term annotation | P-value | FDR adj P-value | Genes |
|---|---|---|---|---|
| qCBSDFc6K | K12200 programmed cell death 6-interacting protein | 3.77E-06 | 6.99E-06 | Manes.06G062600 |
| qCBSDFc6K | PTHR23083 Tetratricopeptide repeat protein, tpr | 0.001423 | 0.001748025 | Manes.06G062500 |
| qCBSDFc10K | KOG3017 Defence-related protein containing SCP domain | 7.80E-11 | 3.12E-10 | Manes.10G089800, Manes.10G089900 Manes.10G090000 |
| qCBSDFc10K | K13449 pathogenesis-related protein 1 | 1.49E-06 | 1.99E-06 | Manes.10G089900 |
| qCBSDFc10K | PF08263 Leucine rich repeat N-terminal domain | 0.030015 | 0.034017167 | Manes.10G090500 |
| qCBSDFc10K | GO:0006952 defence response | 0.017246 | 0.032191706 | Manes.10G091500 |
| qCBSDFc18K | PF01566 Natural resistance-associated macrophage protein | 3.39E-07 | 4.41E-06 | Manes.18G080200, Manes.18G080100 |
| qCBSDFc18K | PF00646 F-box domain | 4.10E-06 | 2.13E-05 | Manes.18G080900, Manes.18G080800, Manes.18G080700 Manes.18G080500 |
| qCBSDFc18K | PF01535 PPR repeat | 0.007722 | 0.01056644 | Manes.18G081800, Manes.18G080300 Manes.18G081300 |
| qCBSDFc18K | PTHR22844 F-box and WD40 domain protein | 0.001121 | 0.001466151 | Manes.18G082000 |
| qCBSDRNc12K | PF03106 WRKY DNA -binding domain | 0.000512 | 0.000512491 | Manes.12G107400 |
| qCBSDRNc12K | PF08263 Leucine rich repeat N-terminal domain | 0.017957 | 0.020070087 | Manes.12G108300 |
| qCBSDRNc12K | PF01566 Natural resistance-associated macrophage protein | 1.55E-05 | 8.81E-05 | Manes.12G109500 |
| qCBSDRNc5K | PF03106 WRKY DNA -binding domain | 0.009542 | 0.011276814 | Manes.05G106900 |
| qCGMc5AR | PF03106 WRKY DNA -binding domain | 0.007003 | 0.008210781 | Manes.05G106900, |
| qCMDc14AR | PF00560 Leucine Rich Repeat | 0.065133 | 0.065132504 | Manes.14G161200 Manes.14G162700 |
| qCMDc14AR | GO:0006396 RNA processing | 0.005399 | 0.012675835 | Manes.14G164400 |
| qCMDc14AR | PTHR11017 Leucine-rich repeat-containing protein | 0.003087 | 0.003420248 | Manes.14G165100 |
| qCMDc14AR | PF00931 NB-ARC domain | 0.005761 | 0.006539726 | Manes.14G165100, Manes.14G165300 |
| qCMDc14AR | GO:0006952 defence response | 0.013098 | 0.02720445 | Manes.14G165100, |

**Table 4.4 Summary of terms detected within the QTL regions**

| Trait | Number of significant terms | Number of terms of related to disease resistance | % |
|---|---|---|---|
| CBSD root necrosis | 51 | 4 | 7.8 |
| CBSD foliar symptoms | 149 | 8 | 5.3 |
| CGM | 30 | 1 | 3.3 |
| CMD | 46 | 4 | 8 |
| Total | 276 | 17 | 6.2 |

**Table 4.5 Enrichment analysis for terms related to disease resistance in cassava and their orthologues in Arabidopsis**

| **Term** | ***P*-value (FDR)** | **Arabidopsis orthologue** | **Function in Arabidopsis** | **Cassava genes** |
|---|---|---|---|---|
| KOG3017 | 0.003151045 | AT2G14580.1 | defence response | Manes.07G050200,Manes.07G050700,Manes.07G050300, Manes.06G028000,Manes.07G050400,Manes.06G027000 |
| KOG4579 | 0.019138275 | AT4G22730.1 | LRR/defence responses | Manes.02G033600,Manes.02G100700,Manes.17G105400, Manes.04G152100 |
| PTHR14009 | 0.032163933 | AT3G59820.1 | Pathogen response | Manes.01G018300 |
| KOG2739 | 0.032647798 | AT3G50690.1 | Leucine-rich repeat (LRR) family protein | Manes.02G092300 |
| KOG1859 | 0.032647798 | AT3G17920.1 | Leucine-rich repeat (InterPro:IPR001611) | Manes.06G004700 |
| PF04258 | 0.047620471 | AT1G63690.1 | Pathogen response and cell cycle during gemivirus infection | Manes.01G059500,Manes.02G020000,Manes.02G140400 |

## 4.5 Discussion

### 4.5.1 Characterization of Kiroba genome assembly

The scaffold assembly of Kiroba is reported to have a N50 score of 3,376. Thus half of the genome is present in contigs with at least this size. Most of the contigs of Kiroba were of small sizes of less than 200bp and were filtered during the assembly. In the final assembly, 13.83 million reads were assembled into scaffolds representing 161.3 million base pairs (Mbp). Comparison with cassava reference genome (Goodstein et al., 2012) (760Mbp) shows that the percent coverage of the Kiroba assembly with larger contigs used in this analysis was 21.2%. However, if the contigs and scaffolds of smaller sizes are considered, 697 Mbp are obtained giving a 91.8% coverage of the genome. These results indicate that the Kiroba genome assembly has high coverage as compared to the reference, and the average length of the scaffolds in the assembly are of a large size above 3,376bp, thus the assembly is adequate and is of high quality and useful for comparative analysis.

### *4.5.2* Genomic regions in Kiroba derived from *M. glaziovii*

This study shows that there are large introgression regions on chromosome 1, 17 and 18 that are consistently the same as in *M. glaziovii* and alternatively different from *M. esculenta*. These regions are *M. glaziovii* like and are presumed to have been introgressed into Kiroba through hybridization during the early breeding activities involving wild relatives of cassava. The introgression regions contain significantly enriched terms by FDR correction and have genes with functions related to disease resistance in plants. These genes have orthologs in Arabidopsis with functions related to disease resistance (AT3G17920.1) (Wang et al., 2008), disease response (AT4G22730.1) (Shen and Hanley-Bowdoin, 2006), pathogen response (AT3G59820.1) and cell cycle during geminivirus infection (AT1G63690) (Ascencio-Ibáñez et al., 2008). Two percent of the total number of the genes in the introgression regions relate to disease resistance in plants and this is very significant as compared to the 1% of total number of genes predicted in cassava (Lozano et al., 2015).

### 4.5.3 Genomic introgression into Kiroba and co-location with detected QTL

There are large *M. glazovii* like regions in chromosome 1 (24888098–33833841 bp), 17 (13700269–23436220 bp), and 18 (6805369–25209274 bp) (v6.1) of Kiroba (Figure 4.2) which is an indication of genomic introgression from wild species. The co-location of qCBSDFc17K and qCBSDFc18Kb associated with CBSD foliar symptoms on chromosome 17 and 18 with the introgression segments suggest that resistance to CBSD foliar symptoms may be partly derived from *M. glaziovii*.

Approximately 7% of the Kiroba genome comprises *M. glaziovii/M. esculenta* introgression segments, whereas in Namikonga approximately 14% of the genome has this form (Bredeson et al., 2016). Six genotypes (Akena, TMS-1972205, Mkombozi, KBH2006/18, TMS-I30572 and Namikonga) share a common *M. glaziovii* haplotype on chromosome 1 (Bredeson et al., 2016), and four genotypes (Mkombozi, KBH2006/18, TMS-130572, and Namikonga) also share a common haplotype on chromosome 4. These were designated as the "Amani haplotypes" (Bredeson et al., 2016). Kiroba does not share the so called "Amani" M. glaziovii haplotype on chromosome 1 or 4 that Namikonga shares with other genotypes. Unlike Namikonga, it does not have a direct relationship with TME117, but is more closely related to an interspecific hybrid "tree cassava" from Tanzania (*M. glaziovii* × *M. esculenta*), and the KBH 2006/18 hub (Bredeson et al., 2016). Tree cassava and Kiroba are related as parent–offspring according to the first degree relatedness analysis in this study. It is possible that tree cassava was used as a parent in the interspecific Amani breeding program, from where Kiroba is thought to have originated, as the rate of seed production in pure *M. glaziovii* × *M. esculenta* crosses can be very low (E. Kanju, pers. commun.).

This rather distinct relationship to the Namikonga-TME117- Nachinyaya-Albert-NDL06/132 cluster of germplasm (Figure 5.3) supports the finding of some different QTL associated with CBSD root necrosis and foliar symptoms, compared to those identified in Namikonga (Masumba et al., 2017). This would provide opportunities for pyramiding of QTL associated with CBSD resistance for more durable field resistance.

### 4.5.4 Candidate genes in QTL regions

The QTL regions contain genes encoding several protein domains that have been reported to be involved in disease resistance in plants. The F-box protein domains found on chromosome XVIII are said to contain LRR domains associated with pathogen responses (Kuroda et al., 2002; Van den Burg et al., 2008). The pentatricopeptide repeat (PPR) superfamily protein (Barkan and Small, 2014) present in chromosome XVIII and tetratricopeptide repeat (TPR)-like superfamily protein (Prikryl et al., 2011) found on chromosome VI have been implicated in plant defence mechanisms. WRKY DNA binding domains found within QTL regions of chromosome V (associated with CGM and CMD) and XII (associated with CMD) are transcription factors involved in plant defence responses (Du and Chen, 2000). The F-box (Jakoby et al., 2002) and WRKY DNA binding domains (Eulgem et al., 2000) have orthologs in Arabidopsis with genes that function in plant defence. One percent of the total predicted genes in cassava contain these protein domains (Lozano et al., 2015) and have been shown to have high sequence similarity to proteins from other plant species, thus it is not surprising to find these genes in the QTL regions and within the putative introgression segments.

## 4.6 Conclusion

The genome of Kiroba has been assembled at very high coverage of 91.8 %. It has some genomic segments that are consistently the same as in *M. glaziovii* but alternatively different from *M. esculenta.* These genomic regions are more like *M. glaziovii* and are concluded to have been introgressed into Kiroba through hybridization in the early breeding efforts involving crossing with *M. esculenta*. Major *M. glaziovii* introgression regions are found on chromosomes I, XVII and XVIII, although the introgression region on chromosome I is not the characteristic "Amani haplotype". The large introgression regions on chromosome XVII and XVIII overlap QTL associated with CBSD foliar symptoms indicating that at least some of this resistance may be derived from wild species. The introgression region on chromosome I in Kiroba is a different haplotype from that of Namikonga. Furthermore, Kiroba does not share a common "Amani haplotype" on chromosome 1 and 4 that Namikonga shares with other genotypes including TME

17. Moreover, the QTL identified in Kiroba are different than those identified in Namikonga. This is suggestive of a different source of CBSD resistance in East African germplasm. Kiroba is closely related to tree cassava as parent-offspring and it may have been used as a donor parent in early breeding efforts involving crosses of *M. glaziovii* x *M. esculenta.*

The introgression regions that overlap with the detected QTL contain genes that encode proteins with similarity to the nucleotide binding site and other domains characteristic of plant defence proteins including LRR that have been reported to be responsible for pathogen recognition in plants. The introgressed genomic regions that overlap with the identified QTL regions have the potential of contributing to the observed field resistance against diseases in Kiroba depending on the environment and disease pressure. The QTL regions that perfectly lie within the introgression regions on chromosome XVII and XVIII are deemed to have been derived from *M. glaziovii* and are responsible for the observed CBSD resistance in Kiroba. Furthermore, the introgressed regions are highly significant and consistently much more similar to *M. glaziovii* hence presumed to be derived from *M. glaziovii.*

The findings of this research elicits the need for genomic characterization of other local landraces in the East Africa region to reveal more in terms of genetic architecture and explore the possibility of their utilization and improvement for future varietal development and release. It is worth noting that the *M. glaziovii*-like genomic regions in Namikonga, the best CBSD resistant cultivar identified by breeders seems to be different from those found in Kiroba. This research reveals the need to tap into the rich reservoir of genes present in the wild progenitors of cassava to improve on the farmer preferred cultivars by pyramiding genes conferring resistance against CBSD, CMD and CGM. This is useful and is expected to positively impact on the cassava molecular breeding strategies to combat the threat of the viral diseases that are ravaging the region with huge economic losses estimated to be millions of dollars annually.

## 4.7 References


Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *Mol. Biol.* 215**,** 403-410.

Ascencio-Ibáñez, J.T., Sozzani, R., Lee, T.-J., Chu, T.-M., Wolfinger, R.D., Cella, R., et al. (2008). Global Analysis of Arabidopsis Gene Expression Uncovers a Complex Array of Changes Impacting Pathogen Response and Cell Cycle during Geminivirus Infection. *Plant. Physiol.* 148**,** 436-454.

Barkan, A., and Small, I. (2014). Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* 65**,** 415-442.

Beck, B. (1982). "Historical perspectives of cassava breeding in Africa", in: *Workshop on Root Crops in Eastern Africa.* (ed.) S.K. Hahn. (Kigali,Rwanda: IDRC).

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *R. Stat. Soc.* 57**,** 289-300.

Bredeson, J.V., Lyons, J.B., Prochnik, S.E., Wu, G.A., Ha, C.M., Edsinger-Gonzales, E., et al. (2016). Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat. Biotechnol.* 34**,** 562-570.

Du, L., and Chen, Z. (2000). Identification of genes encoding receptor-like protein kinases as possible targets of pathogen-and salicylic acid-induced WRKY DNA-binding proteins in Arabidopsis. *Plant. J.* 24**,** 837-847.

Eulgem, T., Rushton, P.J., Robatzek, S., and Somssich, I.E. (2000). The WRKY superfamily of plant transcription factors. *Trends. Plant. Sci* 5**,** 199-206.

Ferguson, M.E., Hearne, S.J., Close, T.J., Wanamaker, S., Moskal, W.A., Town, C.D., et al. (2012). Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. *Theor. Appl. Genet.* 124**,** 685-695.

Gedil, M., Kumar, M., and Igwe, D. (2012). Isolation and characterization of resistant gene analogs in cassava, wild *Manihot* species, and castor bean (*Ricinus communis*). *Afri. J. Biotech.* 11**,** 15111-15123.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic. Acids Res.* 40**,** 1178-1186.

Hillocks, R., and Jennings, D. (2003). Cassava brown streak disease: a review of present knowledge and research needs. *Pest Sci.* 49**,** 225-234.

Hillocks, R., and Thresh, J. (2000). Cassava mosaic and cassava brown streak virus diseases in Africa: a comparative guide to symptoms and aetiologies. *Roots* 7**,** 1-8.

ICGMC (2015). High-resolution linkage map and chromosome-scale genome assembly for cassava (*Manihot esculenta* Crantz) from 10 populations. *G3* 5**,** 133-144.

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., et al. (2002). bZIP transcription factors in Arabidopsis. *Trends Plant Sci.* 7**,** 106-111.

Jennings, D. (1957). Further studies in breeding cassava for virus resistance. *Afr. J. Agric. Res.* 22**,** 213-219.

Jennings, D. (1994). Breeding for resistance to African cassava mosaic geminivirus in East Africa. *Trop. Sci.* 34**,** 110-122.

Kanju, E., Mtunda, K., Muhanna, M., Raya, M., and Mahungu, N. (2002). "Management of Cassava brown streak virus disease in Tanzania", in: *Cassava Brown Streak Virus Disease: Past, Present and Future.* (ed.) J. Hillocks. (Mombasa, Kenya).

Kaweesi, T., Kawuki, R., Kyaligonza, V., Baguma, Y., Tusiime, G., and Ferguson, M.E. (2014). Field evaluation of selected cassava genotypes for cassava brown streak disease based on symptom expression and virus load. *J. Virol.* 11**,** 216. doi: 10.1186/s12985-014-0216-x.

Kuroda, H., Takahashi, N., Shimada, H., Seki, M., Shinozaki, K., and Matsui, M. (2002). Classification and expression analysis of Arabidopsis F-box-containing protein genes. *Plant. Cell Physiol.* 43**,** 1073-1085.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* 9**,** 357-359.

Lozano, R., Hamblin, M., Prochnik, S., and Jannink, J.-L. (2015). Identification and distribution of the NBS-LRR gene family in the Cassava genome. *BMC Genomics* 16**,** 360.

Masumba, E.A., Kapinga, F., Mkamilo, G., Salum, K., Kulembeka, H., Rounsley, S., et al. (2017). QTL associated with resistance to cassava brown streak and cassava mosaic diseases in a bi-parental cross of two Tanzanian farmer varieties, Namikonga and Albert. *Theor. Appl. Genet.* 130:2069-2090.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20**,** 1297-1303.

Muhanna, M., and Mtunda, K. (2003). "Cassava Brown Streak Virus Disease Transmission Studies in Tanzania", in: *Cassava Brown Streak Virus Disease: Past, Present and Future.* (ed.) J. Hillocks. (Mombasa, Kenya).

Ndunguru, J., Sseruwagi, P., Tairo, F., Stomeo, F., Maina, S., Djinkeng, A., et al. (2015). Analyses of twelve new whole genome sequences of cassava brown streak viruses and Ugandan cassava brown streak viruses from East Africa: Diversity, supercomputing and evidence for further speciation. *PLoS ONE* 10**,** e0139321.

Nichols, R. (1947). Breeding cassava for virus resistance. *Afr. J. Agric. Res.* 12**,** 184-194.

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

Patel, R.K., and Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE* 7**,** e30619.

Prikryl, J., Rojas, M., Schuster, G., and Barkan, A. (2011). Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc. Natl. Acad. Sci.* 108**,** 415-420.

Prochnik, S., Marri, P.R., Desany, B., Rabinowicz, P.D., Kodira, C., Mohiuddin, M., et al. (2012). The cassava genome: current progress, future directions. *Trop. Plant Biol.* 5**,** 88-94.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81**,** 559-575.

Rwegasira, G.M., and Rey, C.M. (2012). Response of selected cassava varieties to the incidence and severity of cassava brown streak disease in Tanzania. *Agri. Sci.* 4**,** 237.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13**,** 2498-2504.

Shen, W., and Hanley-Bowdoin, L. (2006). Geminivirus infection up-regulates the expression of two Arabidopsis protein kinases related to yeast SNF1-and mammalian AMPK-activating kinases. *Plant Physiol.* 142**,** 1642-1655.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., and Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Res.* 19**,** 1117-1123.

Thresh, J. (2003). "Brief History of Cassava Brown Streak Virus Disease", in: *Cassava Brown Streak Virus Disease: Past, Present, and Future.* (ed.) J. Hillocks. (Mombasa, Kenya).

Van den Burg, H.A., Tsitsigiannis, D.I., Rowland, O., Lo, J., Rallapalli, G., MacLean, D., et al. (2008). The F-box protein ACRE189/ACIF1 regulates cell death and defense responses activated during pathogen recognition in tobacco and tomato. *Plant Cell* 20**,** 697-719.

Wang, Y., Zhang, W.-Z., Song, L.F., Zou, J.J., Su, Z., and Wu, W.H. (2008). Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in Arabidopsis. *Plant. Physiol.* 148**,** 1201-1211.

# Chapter 5

# Concluding Remarks

# 5

## 5.1 Concluding Remarks

This study has identified four significant QTL linked to resistance to CBSD root necrosis on chromosomes V, XI, XII, and XV, and these appear to be different from those found in the cultivar Namikonga (Masumba et al., 2017) and may be another possible source of CBSD resistance. Nine significant QTL linked to CBSD foliar symptoms have also been identified on chromosomes IV, VI, XVII, and XVIII, with QTL on chromosomes XI and XV being consistent between root necrosis and foliar symptoms. This observation suggests that resistance to CBSD foliar symptoms may be controlled by some of the same loci to root necrosis, but additional loci are trait specific. A QTL associated with CMD resistance is consistent with the CMD2 locus on chromosome XII.

The genome of Kiroba was assembled at very high coverage of 91.8 % (697 million bases), however, the scaffolds >500bp used in this analysis was 21.2% (760 million bases The genome contain large *M. glazovii* like regions in chromosome 1(24888098–33833841 bp), 17 (13700269–23436220 bp), and18 (6805369–25209274 bp) (v6.1 of cassava assembly) which is an indication of genomic introgression from this wild species. The co-location of qCBSDFc17K and qCBSDFc18Kb associated with CBSD foliar symptoms on chromosome 17 and 18 with the introgression segments suggests that resistance to CBSD foliar symptoms may be partly derived from *M. glaziovii*.

The introgression region on chromosome I of Kiroba is not the characteristic "Amani haplotype" found in Namikonga, Akena, TMS-1972205, Mkombozi, KBH2006/18, and TMS-I30572. Four genotypes (Mkombozi, KBH2006/18, TMS-130572, and Namikonga) also share a common haplotype on chromosome 4. Thus, it can be concluded that Kiroba does not share the so called "Amani" *M. glaziovii* haplotype on chromosome1 or 4 that Namikonga shares with other genotypes. Approximately 7% of the Kiroba genome comprises an *M. glaziovii*/*M. esculenta* introgression segment, whereas in Namikonga approximately 14% of the genome has this form (Bredeson et al., 2016). Unlike Namikonga, it does not have a direct relationship with TME117, but is more closely related to an interspecific hybrid "tree cassava" from Tanzania (*M. glaziovii* × *M. esculenta*), and the KBH 2006/18 hub (Bredeson et al., 2016). Tree cassava and Kiroba are

related as parent–offspring according to the first degree relatedness analysis in this study. It is possible that tree cassava was used as a parent in the interspecific Amani breeding program, from where Kiroba is thought to have originated, as the rate of seed production in pure *M. glaziovii* × *M. esculenta* crosses can be very low (E. Kanju, pers. comm.). This rather distinct relationship to the Namikonga-TME117-Nachinyaya-Albert-NDL06/132 cluster of germplasm supports the finding of some different QTL associated with CBSD root necrosis and foliar symptoms in Kiroba and Namikonga (Masumba et al., 2017). This signifies different sources of resistance and may provide opportunities for pyramiding QTL associated with CBSD resistance for more durable field resistance.

## 5.2 Future perspectives

The results from this study presents strong evidence of the presence of genomic segments in Kiroba putatively derived from *M. glaziovii*. The results show 'Kiroba' to be a possible interspecific hybrid closely related to tree cassava but having a different haplotype than the so called "Amani haplotype" that has been identified in Namikonga. These results are not surprising because interspecific hybridization in cassava with its wild relatives was utilized in the early cassava breeding programs (Nassar, 2000) and many derivatives of the crosses are now not traceable using morphological characters alone. Furthermore, several of the interspecific hybrids with known pedigrees are being grown by farmers under different names (Kaweesi et al., 2014) and show strong field resistance towards viral diseases (Thresh, 2003). Kiroba shows strong field resistance against CBSD similar to that observed in Namikonga (Rwegasira and Rey, 2012). The pedigree of Namikonga is thought to be known unlike that of Kiroba. It may have been lost when the Amani breeding program closed and the breeding materials transferred to other breeding stations across Africa (Legg and Hillocks, 2003).

Due to the increased challenges of viral diseases of cassava such as CBSD and CMD, tapping into the wild relatives of cassava and their derivatives as reservoir and source of resistance genes is key for cassava improvement in East Africa. Understanding their genomic architecture is important for future breeding of cassava. Knowledge about the genetic basis of observed resistance and its source

hereby identified in Kiroba will greatly contribute to the improvement of levels of resistance by enabling the pyramiding of QTL controlling various diseases and pests.

## 5.3 References


Bredeson, J.V., Lyons, J.B., Prochnik, S.E., Wu, G.A., Ha, C.M., Edsinger-Gonzales, E., et al. (2016). Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat. Biotechnol.* 34**,** 562-570.

Kaweesi, T., Kawuki, R., Kyaligonza, V., Baguma, Y., Tusiime, G., and Ferguson, M.E. (2014). Field evaluation of selected cassava genotypes for cassava brown streak disease based on symptom expression and virus load. *J. Virol.* 11**,** 216.

Legg, J.P., and Hillocks, R.J. (2003). "Cassava Brown Streak Virus Disease: Past, Present and Future," in: *Proceedings of an International Workshop.* (ed.) J. Hillocks. (Mombasa, Kenya).

Masumba, E.A., Kapinga, F., Mkamilo, G., Salum, K., Kulembeka, H., Rounsley, S., et al. (2017). QTL associated with resistance to cassava brown streak and cassava mosaic diseases in a bi-parental cross of two Tanzanian farmer varieties, Namikonga and Albert. *Theor. Appl. Genet.* 130:2069-2090

Nassar, N. (2000). Cytogenetics and evolution of cassava (*Manihot esculenta Crantz*). *Genet. Mol. Biol.* 23**,** 1003-1014.

Rwegasira, G.M., and Rey, C.M. (2012). Response of selected cassava varieties to the incidence and severity of cassava brown streak disease in Tanzania. *Agri. Sci.* 4**,** 237.

Thresh, J. (2003). "Brief History of Cassava Brown Streak Virus Disease", in: *Cassava Brown Streak Virus Disease: Past, Present, and Future.* (ed.) J. Hillocks. (Mombasa, Kenya).