ABSTRACT

OLOKA, BONNY MICHAEL. Genetic Linkage Map Construction and QTL Analysis of Important Pest and Agronomic Traits in Two Bi-parental Sweetpotato SNP Mapping Populations. (Under the direction of Dr. G. Craig Yencho).

Genetic improvement of sweetpotato, *Ipomoea batatas* (L.) Lam., for important traits has progressed slowly over the years in sub-Saharan Africa (SSA) where the crop is an important staple food. This is largely due to the crop’s genetic complexity. Sweetpotato has a relatively large genome (ca. 1.5 Mbp) and it is predominantly an autopolyploid hexaploid (2n = 6x = 90), exhibiting some preferential pairing on some chromosomes. It is also an outcrossing species, subject to significant in-breeding depression if heterozygosity is reduced during breeding and it is encumbered with significant self- and cross-incompatibilities, which makes breeding very difficult when one is seeking to improve the crop for the wide array of biotic and abiotic stresses to which it is exposed to in SSA.

This research was conducted to improve our understanding of the inheritance of resistance to three pests of sweetpotato: root-knot nematode (RKN), *Meloidogyne incognita* race 3, the sweetpotato weevils (SPW), *Cylas puncticollis* (Boh.) and *Cylas brunneus* (Fab), and sweetpotato virus disease (SPVD); and three quality traits, storage root β-carotene, starch and dry matter content.

Two bi-parental genetic mapping populations were used to study the inheritance of these traits. The ‘New Kawogo’ x ‘Beauregard’ (NKB) population consisting of 287 individuals was developed in Uganda at the National Crops Resources Research Institute (NaCRI), and phenotyped in three locations, namely; NaCRI, National Semi-Arid Resources Research Institute (NaSARRI) and Ngeta Zonal Agricultural Research and Development Institute (NgeZARDI), in Uganda for two seasons for SPVD, SPW, storage root
dry matter content, and β-carotene content. DNA was extracted from NaCRRRI and shipped to NCSU for genotyping. The second mapping population consisted of 240 individuals of the ‘Tanzania’ x ‘Beauregard’ (TB) population. The TB mapping population was developed at NC State University, and phenotyped for resistance to RKN in a series of greenhouse trials at NCSU.

A modified genotyping by sequencing (GBS) protocol optimized for sweetpotato (GBSpoly) was used to generate single- and multiple-dose single nucleotide polymorphism (SNP) markers in both populations. Novel genetic linkage mapping and QTL analysis algorithms, (MAPpoly and QTLPoly respectively) developed specifically for complex polyploids like sweetpotato were used to facilitate the analysis of both mapping populations.

In the NKB population, we identified 13,096 SNP markers and developed a 5,055-cM genetic map comprising 15 linkage groups (LG), associated with the basic chromosome number for sweetpotato. We identified 2 QTL for SPVD resistance on linkage groups 1 and 2 explaining a total variability of 36%. For SPW resistance, we identified 1 QTL on linkage group 14 at the NaCRRRI site, and another QTL on linkage group 3 at the NgeZARDI location. The QTL associated with resistance to SPVD at NaCRRRI explained a total variability of 13.2% while a different QTL at NgeZARDI explained 11.7%. We further identified 2 QTL for storage root dry-matter content and 3 QTL for β-carotene content explaining a total variation of 36.7% and 29.6% respectively. Additive effects of the QTL alleles from both parents were estimated and important alleles that directly impact the mean of the traits were identified.

In the TB population, we identified 14,813 SNP markers and developed a 2,120 cM genetic linkage map. We identified 2 QTL on linkage groups 3 and 7, explaining a total variation of 66% of resistance to the root knot nematode, *M. incognita* race 3. A BLAST search
of the sequence regions within the QTL revealed putative candidate genes, with potential for validation and future use for molecular breeding in sweetpotato.

The work presented here represents a significant step forward in our understanding of the genetic architecture of important traits in cultivated sweetpotato and sets the stage for future utilization of genomics-assisted breeding in applied sweetpotato breeding programs. This research is one aspect of a larger effort focused on the development and efficient application of new genomic, statistical and bioinformatic tools for global sweetpotato improvement under the Genomic Tools for Sweetpotato Improvement project (GT4SP) supported by the Bill and Melinda Gates Foundation.
Genetic Linkage Map Construction and QTL Analysis of Important Pest and Agronomic Traits in Two Bi-parental Sweetpotato SNP Mapping Populations

by
Bonny Michael Oloka

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Horticultural Science

Raleigh, North Carolina

2019

APPROVED BY:

_______________________________  _______________________________
G. Craig Yencho             Zhao-Bang Zeng
Committee Chair

_______________________________  _______________________________
Christian Maltecca            Bode A. Olukolu
DEDICATION

“To my beloved parents Mr. Emmanuel Aripa and Mrs. Grace Aripa, my siblings Martin, Faith and Joshua, my wife and best friend Daisy, my daughter Maari, and to the entire sweetpotato community.”
BIOGRAPHY

Bonny Michael Oloka was born in Tororo, Uganda on July 7, 1984. He was raised with his two brothers and one sister from Kampala, by their parents Mr. Emmanuel Aripa-Ketta and Mrs. Grace Aripa. He attended primary school in Makerere Primary School, then went to Old Kampala Senior Secondary School and Caltec Academy Makerere for his secondary school education. He obtained his Bachelor of Science (Biochemistry, Chemistry) in January 2009 and Master of Science (Plant Breeding and Seed Systems) in January 2016 from Makerere University.

In 2008, Bonny started his career as a Research Technician in the National Banana Research Program, Kawanda, where he worked for four years on banana biotechnology, breeding and tissue culture. In 2012, he was awarded a scholarship for masters training by the National Agricultural Research Organization (NARO) under the Eastern African Agricultural Productivity Project (EAAPP) to pursue masters training in rice breeding and seed systems. In 2015, Bonny was awarded a scholarship for PhD studies at North Carolina State University (NCSU) through a collaborative project between NCSU and NARO titled Genomic Tools for Sweetpotato Improvement project (GT4SP) funded by the Bill and Melinda Gates Foundation.

After completion of his PhD training, Bonny will return to Uganda and use his acquired knowledge and skills to work on improving sweetpotato varieties for nutrition, health and food security using modern genomic tools in sub-Saharan Africa.
ACKNOWLEDGMENTS

I am forever grateful to my major Professor, Dr. G. Craig Yencho for his mentorship, guidance, support and training throughout my doctoral studies. I am truly grateful for his excellent supervision and continuous encouragement in the pursuit of my academic goals, without which I would not have completed my training.

My deepest gratitude goes to my committee members; Dr. Bode Olukolu, who optimized the protocol for genotyping hexaploid sweetpotato and developed ‘GBSpoly’ which gave us high quality sequence data, Dr. Zhao-Bang Zeng, who oversaw the development of the statistical tools and pipeline for SNP calling, linkage mapping and QTL analysis, and Dr. Christian Maltecca who introduced me to practical quantitative genetics and analysis of next-generation sequence data using shell scripting and command-based tools. I would like to thank Dr. Guilherme da Silva Pereira and Dr. Marcelo Mollinari, both in Dr. Zeng’s group, - who worked tirelessly to develop and optimize bioinformatics software for hexaploid sweetpotato in collaboration with the GT4SP project, and patiently trained me how to efficiently use them in my study.

My sincere gratitude goes out to the entire team at National Crops Resources Research Institute, Namulonge, led by Dr. Benard Yada who developed and phenotyped the NKB population which we used in this study. I’m very grateful to Dr. Milton Otema Anyanga, Doreen Chelangat, Paul Musana, Agnes Alajo, Joan Adero, Jenniffer Asiimwe and Justine Nanteza for all the support they offered.

Special thanks to Victor Amankwaah who always found time to help me with my RKN work in the greenhouse amidst his busy schedule as a fellow graduate student. Stella Nhanala,
Adam Canal, Samuel Acheampong and Jonathan Kinczyk, - my colleagues and graduate students in the program who always provided me with a helping hand whenever requested. We had a way of helping each other out and that made all the difference! I sincerely thank Ken Pecota, Ana Pecota, Sharon Williamson, Meri Reeber, Amber Harmon, Ragy Ibrahim and Dominic Gillooly who provided me with all the resources that I needed for my work in the greenhouse and lab. Finally, heartfelt thanks to Wendy Koch, Bryan Ellerbrock, Xiaofei Zhang, Chris Heim, Jeremy Machacek, Mark Clough and his team, and the entire staff of the Sweetpotato and Potato Breeding and Genetics Programs at NC State University for their training and support throughout my studies. Thank you so much for giving me a home away from home, and for making my training enjoyable! God bless you all.
TABLE OF CONTENTS

LIST OF TABLES .......................................................................................................................... viii
LIST OF FIGURES ..................................................................................................................... x

CHAPTER 1 ................................................................................................................................. 1
Introduction ................................................................................................................................. 1
Literature Review ....................................................................................................................... 8
  Origin, Genetics and Improvement ......................................................................................... 8
  Molecular markers in sweetpotato improvement .................................................................. 9
  Genetic Linkage Map Construction in Sweetpotato .............................................................. 10
  QTL analysis in sweetpotato ................................................................................................. 11
  Storage root yield, dry matter, starch and β-carotene content ............................................. 13
  Sweetpotato virus disease and its management .................................................................... 14
  The Sweetpotato weevil ....................................................................................................... 16
  Root-knot nematode in sweetpotato ..................................................................................... 19
  REFERENCES ......................................................................................................................... 23

CHAPTER 2 ................................................................................................................................ 36
Linkage mapping and QTL analysis of sweetpotato virus disease resistance in hexaploid sweetpotato ................................................................. 36
  Abstract ................................................................................................................................. 38
  Introduction ........................................................................................................................... 39
  Materials and Methods ......................................................................................................... 45
  Results .................................................................................................................................. 52
  Discussion ............................................................................................................................ 55
  REFERENCES ......................................................................................................................... 61

CHAPTER 3 ................................................................................................................................ 88
Identification of QTL for sweetpotato weevil resistance in a biparental mapping population ......................................................................................................................... 88
  Abstract ................................................................................................................................. 90
  Introduction ........................................................................................................................... 91
  Materials and Methods ......................................................................................................... 96
  Results .................................................................................................................................. 101
  Discussion ............................................................................................................................ 103
  REFERENCES ......................................................................................................................... 106
CHAPTER 4
Genetic mapping of QTL for storage root dry matter and β-carotene content in sweetpotato

Abstract
Introduction
Materials and Methods
Results
Discussion
REFERENCES

120
120
122
123
126
130
132
135

CHAPTER 5
QTL analysis for root-knot nematode resistance in cultivated sweetpotato

Abstract
Introduction
Materials and Methods
Results
Discussion
REFERENCES

148
148
150
151
156
162
166
172

CHAPTER 6
What are the Next Steps for Sweetpotato Genomics Research in SSA?

APPENDICES
Appendix A
Appendix B

191
192
196
197
198
LIST OF TABLES

CHAPTER 2

Table 1. Summary statistics of integrated linkage map for ‘New Kawogo’ x ‘Beauregard’ (NKB) population created using MAPpoly ......................................................71

Table 2. Number of markers per dosage in both parents. NK. dosage is the marker dosage in New Kawogo. B. dosage is the marker dosage in Beauregard. The dosage groups 0, 1, 2, 3 and 4 correspond to nulliplex, simplex, duplex, triplex and quadruplex marker groups respectively ...........................................72

Table 3. Summary statistics of QTL for SPVD resistance on the NKB map across sites and environments. $\sigma_{QTL}^2$ is the variance associated with the QTL. $H^2_{QTL}$ is the heritability of the QTL ............................................................................................................73

Table 4. Summary statistics of QTL for SPVD resistance in NgeZARDI. $\sigma_{QTL}^2$ is the variance associated with the QTL. $H^2_{QTL}$ is the heritability of the QTL ............................................................................................................73

CHAPTER 3

Table 1. Experimental sites [Source: (Anyanga , et al., 2017)]........................................111

Table 2. Summary statistics of integrated linkage map for ‘New Kawogo’ x ‘Beauregard’ (NKB) population created using MAPpoly [Source: Oloka, et al. (in prep.)] ........................................................................................................111

Table 3. Summary statistics of QTL for SPW resistance in NaCRRI and NgeZARDI. $\sigma_{QTL}$ is the variance associated with the QTL. $H^2$ is the heritability / amount of variability explained by the QTL ........................................................................................................112

Table 4. Summary statistics of QTL for HCA esters in the NKB population. $\sigma_{QTL}$ is the variance associated with the QTL. $H^2$ is the heritability / amount of variability explained by the QTL ........................................................................................................112
CHAPTER 4

Table 1. Experimental sites [Source: (Anyanga, et al., 2017)] ........................................141

Table 2. Summary statistics of QTL for storage root β-carotene content on the NKB map across sites and environments. \( \sigma^2_{QTL} \) is the variance associated with the QTL. \( h^2_{QTL} \) is the heritability of the QTL. The QTL range is the 95% support interval of the QTL ................................................................. 142

Table 3. Summary statistics of QTL for storage root dry matter content on the NKB map across sites and environments. \( \sigma^2_{QTL} \) is the variance associated with the QTL. \( h^2_{QTL} \) is the heritability of the QTL. The QTL range is the 95% support interval of the QTL ............................................................................. 142

CHAPTER 5

Table 1. Summary of individual chromosomes for the integrated TB linkage map created using MAPpoly ................................................................................................. 179

Table 2. Number of markers per dosage in both parents. T (Tanzania) dosage is the marker dosage in Tanzania. B (Beauregard) dosage is the marker dosage in Beauregard. The dosage groups 0, 1, 2, 3 and 4 correspond to nulliplex, simplex, duplex, triplex and quadruplex marker groups respectively ................................................................. 180

Table 3. Summary statistics of QTL for RKN resistance on the TB map. \( \sigma^2_{QTL} \) is the variance associated with the QTL. \( h^2_{QTL} \) is the heritability of the QTL. QTL limit is the support interval for the QTL on the linkage group ........................................... 181
LIST OF FIGURES

CHAPTER 2

Figure 1. Distribution of the adjusted means in all locations combined (joint adjusted means) and in each location (marginal means); NaCRRI, NaSARRI and NgeZARDI. Numbers on the x-axis show field SPVD scores on a 1 – 9 scale where 1 is resistant and 9 is highly susceptible...........................................74

Figure 2. Boxplot showing the distribution of marker dosages in the ‘New Kawogo’ X ‘Beauregard’ (NKB) mapping population. New Kawogo 0-1 Beauregard means nulliplex by simplex for New Kawogo X Beauregard, 0-2 is nulliplex by duplex, 2-1 is duplex by simplex, and so on...........................................75

Figure 3. Multipoint “De novo” genetic linkage map of NKB population. The integrated map represents all the 15 chromosomes of hexaploid sweetpotato. Numbers 1 to 15 are the linkage groups, while the numbers below them are the number of SNP markers for each linkage group..................................................................................76

Figure 4. Plot of QTL analysis of the entire genome for SPVD resistance in the NKB mapping population obtained with the joint model that averages out the separate effects of QTLs in all environments. QTL are identified on linkage group 1 and 2. LOP is the log of p-values .............................................................................77

Figure 5. Logarithm of p-value profile (LOP) for SPVD resistance in NaCRRI in the NKB population. The triangle on linkage group 1 represents the location of mapped QTL on the linkage group.........................................................77

Figure 6. Breakdown of genotypic values showing additive effects from each allele for QTL 1 in the population. Alleles a, b, c, d, e, f are from New Kawogo whereas alleles g, h, i, j, k, l are from Beauregard. Numbers on the y-axis show the genetic value (deviation from population mean) of the allele. The zero (0) point is the population mean (sum of allele frequency x value) of the trait from a normal distribution.................................................................78

Figure S1. Integrated SNP-based genetic linkage map of New Kawogo x Beauregard. Each of the six horizontal bars on each parent represents a homologous chromosome. The red blocks represent markers identified in that region ...........79

CHAPTER 3

Figure 1. Distribution of the adjusted means for SPW in all locations combined (joint adjusted means) and in each location (marginal means); NaCRRI, NaSARRI and NgeZARDI. The vertical axes of the histograms represent the frequency of clones in a given class represented by vertical bars ..........................113
Figure 2. Distribution of mean number of feeding holes in storage roots of the progeny and parents of the ‘New Kawogo’ (NK) x ‘Beauregard’ (B) mapping population collected from BuZARDI, RwebitaZARDI and AbiZARDI. Vertical bars represent the number of clones in each class of mean number of feeding holes averaged over seasons ..................................................114

Figure 3. Log of p-value (LOP) profile showing mapped QTL for SPW in NaCRRI. The triangle on linkage group 14 shows the position of identified QTL on the linkage group ..................................................................................115

Figure 4. Log of p-value (LOP) profile showing mapped QTL for SPW in NgeZARDI. The triangle on linkage group 3 shows the position of identified QTL on the linkage group ..................................................................................116

Figure 5. Logarithm of p-value (LOP) profile showing mapped QTL for HCA esters in storage roots of the NKB population analyzed jointly from 3 sites: NaCRRI, NaSARRI and NgeZARDI. The triangle on linkage group 1 shows the position of identified QTL on the linkage group ..................................................................................117

Figure 6. Breakdown of genotypic values showing additive effects from each allele for QTL 1 in the population when planted in NaCRRI. Alleles a, b, c, d, e, f are from New Kawogo whereas alleles g, h, i, j, k, l are from Beauregard. Numbers on the y-axis show the genetic value (deviation from population mean) of the allele. The zero (0) point is the population mean (sum of allele frequency x value) of the trait from a normal distribution .................118

Figure 7. Breakdown of genotypic values for HCA concentration showing additive effects from each allele of the parents. Alleles a, b, c, d, e, f are from New Kawogo whereas alleles g, h, i, j, k, l are from Beauregard ..................................................................................119

CHAPTER 4

Figure 1. Distribution of the adjusted means across all environments for parents and progeny of the NKB population. Numbers on the x-axis are the predicted means. Vertical bars represent the number (frequency) of genotypes within a given bin ..................................................................................143

Figure 2. Plot of QTL analysis for storage root beta-carotene content (BC) in the NKB mapping population. QTL are identified on linkage group 2, 3 and 12 .................144

Figure 3. Plot of QTL analysis for storage root dry matter content (BC) in the NKB mapping population. QTL are identified on linkage group 4 and 10 .................145
Figure 4. Plot of mapped QTL for storage root beta-carotene content (BC), starch content (STCH) and dry matter content (DM) in the NKB mapping population. Minor QTL for STCH and the QTL for BC mapped very close to each other on linkage groups 3 and 12 ................................................................. 146

Figure 5. Breakdown of genotypic values of β-carotene showing additive effects from each allele of the parents. Alleles a, b, c, d, e, f are from New Kawogo whereas alleles g, h, i, j, k, l are from Beauregard. Numbers on the y-axis show the genetic value (deviation from population mean) of the allele. The zero (0) point is the population mean (sum of allele frequency x value) of the trait from a normal distribution................................................. 147

CHAPTER 5

Figure 1. Genetic correlation of the distribution of resistance reactions in the progeny and parents of TB mapping population among different replications separated by time. Raw1 to Raw5 are the raw data for the number of RKN egg masses counted per root system from Rep1 to Rep5. Joint is the joint adjusted mean. Pearson correlation (p*** < 0.001) calculated among the adjusted means of 5 replications ................................................................. 182

Figure 2. Histogram and boxplot of the distribution of resistance reactions in the TB population in terms of the number of galls of RKN egg masses per root system. Most of the progeny (70%) reacted like Tanzania, the resistant parent while 2% were more susceptible than Beauregard. The red bar on the boxplot shows where half of the population (50%) is located .......... 183

Figure 3. Histogram of the distribution of a) average dry vine weight and b) average dry root weight in the parents and progeny of the TB population. Most of the genotypes segregated towards the female parent, Tanzania. Average dry vine weight was normally distributed whereas average dry root weight was right skewed ............................................................................. 184

Figure 4. Boxplot showing the distribution of marker dosages in the TB mapping population. Tanzania 0-1 Beauregard means nulliplex by simplex for Tanzania x Beauregard, 0-2 is nulliplex by duplex, 2-1 is duplex by simplex, and so on........................................................................................................... 185

Figure 5. Integrated SNP-based genetic linkage map of Tanzania x Beauregard mapping population ........................................................................................................................................ 186
Figure 6. QTL profile for the TB mapping population showing the raw data for no. of egg masses (RKN galls) per root system in reps 1 to 5 plotted for QTL analysis. Raw1 is the data for Rep 1, Raw 2 for Rep 2 and so on. Mean is the adjusted mean for all the five reps. LOP is the logarithm of probability values .................................................................................................................187

Figure 7. QTL profile of the entire genome for number of egg masses (RKN galls) per root system in the TB population obtained with the joint model with weighted averages of the five replications (Mean). QTL were identified on linkage groups 2 and 7. LOP is the logarithm of probability values .................................................................................................................................188

Figure 8. Confidence interval (95%) for the mapped QTL for resistance to RKN in the TB mapping population .................................................................................................................................189

Figure 9. Breakdown of genotypic values showing additive effects from each allele at the QTL using their decomposed BLUPs in the population. Alleles a, b, c, d, e, f are from Beauregard whereas alleles g, h, i, j, k, l are from Tanzania. Numbers on the y-axis show the genetic value (deviation from population mean) of the allele. The zero (0) point is the population mean (sum of allele frequency x value) of the trait from a normal distribution .................................................................................................................................190
CHAPTER 1

Introduction

Sweetpotato, *Ipomoea batatas* [L.] Lam., is one of the ten most important food crops in the world with a total estimated global production of 106 million metric tons in 2014 (FAOSTAT, 2016). Sub-Saharan Africa (SSA) and Asia, produce most of the global sweetpotato production with 85.4% in China, 11.6% in Africa, 2.3% in the Americas and 0.1% in Europe. In east Africa, production is mainly concentrated in the areas around Lake Victoria, with Uganda being the largest producer in the region averaging 2.35 million metric tons annually and ranking third globally after China and Nigeria (FAOSTAT, 2016).

In the developing world, sweetpotato is a staple food crop providing important calories, fiber, potassium, protein and energy requirements (Woolfe, 1992; Low , et al., 2007) particularly in SSA and parts of Asia (CIP, 2015). Local landraces are grown mainly in SSA in small plots, and these are cream or white-fleshed sweetpotato that have high dry-matter content (28-30%) and low sugars (Mwanga, et al., 2007). In the United States of America (USA), sweet, low dry matter (<20%), orange-fleshed types are preferred (Cervantes-Flores , et al., 2011). Orange-fleshed types have higher nutritive value due to their increased β-carotene and vitamin C content. In addition to sweetpotato being consumed as tablestock in the U.S., a significant amount of the crop is processed into French fries and baby foods, and there are several value-added products in the growing market for chips, beers, spirits, and flours.

The ability of sweetpotato to grow under different agro-ecological environments, relatively good performance in poor soils with low inputs, short growing season and its ability for storage in the ground prior to harvest, has enabled the crop to earn its place as a famine reserve food security crop in SSA with good potential for health and poverty alleviation
(Mukasa, 2004). The full potential of sweetpotato has not yet been recognized despite its widely recognized global importance. This is due to many factors including: the crop’s complex hexaploid genetics ($2n = 6x = 90$); it’s large genome size of $4.8-5.3 \text{ pg/2C nucleus}$ (Ozias-Akins & Jarret, 1994; Yada, et al., 2017d; Yada, et al., 2017d; Yada, et al., 2017d; Yada, et al., 2017d); it’s out-crossing nature with significant self- and cross-incompatibilities, high heterozygosity; and the fact that there are limited genomic resources available for the crop.

Constraints to sweetpotato production in SSA include drought, sweetpotato virus disease, sweetpotato weevils ($Cylas$ $puncticollis$ and $Cylas$ $brunneus$), $Alternaria$ stem blight, nematodes ($Meloidogyne$ $spp$, $Pratylenchus$ $brachyurus$ and $Helicotylenchus$ $sp$), absence of clean planting material due to weak seed systems, poor yielding varieties that have low nutritive value, significant storage losses and marketing problems, and a limited range of processing options (Woolfe, 1992; Mwanga, 2001).

Aphid and white fly transmitted sweetpotato virus disease (SPVD) is one of the most damaging constraints in many production regions in SSA, causing up to 90% yield loss in susceptible genotypes. SPVD results from a combined infection of $Sweet$ $potato$ $feathery$ $mottle$ $virus$ (SPFMV) and $Sweet$ $potato$ $chlorotic$ $stunt$ $virus$ (SPCSV) (Adikini, et al., 2015). The sweetpotato weevils, $Cylas$ $spp$, are the most serious insect pests of sweetpotato worldwide (Rees, et al., 2003). $Cylas$ $puncticollis$ (Boheman) and $Cylas$ $brunneus$ (Fabricius) are uniquely African species (Downham, et al., 2001) causing crop losses of up to 98% in areas with low rainfall and long dry spells (Smit, 1997). Plant parasitic nematodes are reported to affect root and tuber crops in Uganda including sweetpotato (Namaganda , et al., 1993; Makumbi-Kidza, et al., 2000), cassava and banana (Coyne, et al., 2003; Coyne, et al., 2006).
New and improved varieties of sweetpotato would significantly contribute to reducing many of these production problems. Most traits of economic importance in sweetpotato exhibit quantitative inheritance or appear to do so due to the polyploid nature of the crop, thus presenting enormous challenges for breeders using conventional breeding techniques (Jones, 1986; Collins, et al., 1999).

The development and use of genomic and bioinformatics tools to facilitate sweetpotato molecular breeding has lagged behind that of other globally important crops such as maize, rice and cassava. Until recently, the global community of sweetpotato breeders and researchers lacked basic genomic resources such as a reference genome sequence, molecular markers (especially single nucleotide polymorphisms), and sequence information for identification of genes of economic significance (Varshney, et al., 2010). They also lack a comprehensive detailed molecular genetic linkage map to facilitate quantitative trait loci (QTL) identification, marker-assisted breeding, gene cloning, and comparative genomic breeding research (Monden & Tahara, 2017). In light of all these challenges, sweetpotato is widely considered as an “orphan crop” with virtually nonexistent private investment in the development of sweetpotato genomics infrastructure, leaving all these efforts in the hands of the public sector and non-governmental organizations (NGO’s) (Yada, 2014). The development and use of modern genomic tools in sweetpotato breeding will facilitate more efficient introgression of important traits and subsequent faster genetic gain.

Some progress has been made towards the development of genomic resources to facilitate molecular breeding in sweetpotato over the last two decades. However, due to the species’ complex genomic architecture, the construction of a genetic linkage map for sweetpotato has been challenging given its outcrossing, highly heterozygous nature, and the
fact that it is a complex polyploid \((2n = 6x = 90)\) with a large number of possible genotypes expected. Until recently, the most commonly used approach for linkage map construction in an outcrossing heterozygous species was based on the use of single-dose restriction fragments (SDF) (Wu, et al., 1992) combined with other breeding and mapping strategies such as the use of testcross and pseudo-testcross populations to generate segregating populations (Grattapaglia & Sederoff, 1994). In this approach, linkage mapping is conducted for each parent separately with homologous chromosomes nonaligned, resulting in the development of two parental maps.

Single dose markers (simplex markers) are those that are present in one parent as a single copy. Simplex markers segregate in the F1 generation in a 1:1 fashion and regardless of the ploidy of the genome, half of the progeny will carry this marker and the other half will not. In hexaploid sweetpotato, only simplex, duplex and triplex marker dosages can produce polymorphisms in the progeny whereas alleles in higher dosages are not polymorphic. This is based on the four hypothesized expected segregation ratios for the inheritance of a dominant marker, predicted from cytological characteristics (autohexaploid, tetradiaploid, allohexaploid) (Jones, 1965; Kriegner, et al., 2003; Cervantes-Flores, et al., 2008a).

A handful of studies have constructed genetic linkage maps and conducted QTL analyses in hexaploid sweetpotato (Ukoskit, et al., 1997; Kriegner, et al., 2003; Cervantes-Flores, et al., 2008a; Chang, et al., 2009; Cervantes-Flores, et al., 2011; Zhao, et al., 2013a). Genetic linkage maps are the starting point for localization of single genes and/or quantitative trait loci (QTL) associated with traits of agricultural importance (Geest, et al., 2017). These genetic strategies provide valuable information for understanding the biological basis and genomic architecture of complex traits (Lee, 1995).
Few QTL have been identified for cultivated sweetpotato. To date, QTL for storage root dry matter, starch, β-carotene, and root-knot nematode resistance have been mapped (Cervantes-Flores et al., 2011; Zhao et al., 2013a; Yu et al., 2014). Using the “Tanzania” x “Beauregard” map, Cervantes-Flores et al. (2008b) identified 9 QTL associated with resistance to RKN. Using the same map and the same mapping population, this group further identified 13 QTL for dry matter content, 12 QTL for starch content, and 8 QTL for β-carotene content (Cervantes-Flores et al., 2011). Li et al. (2010) identified one QTL for storage root starch content in the map of “Zhengshu 20” and Zhao et al. (2013) detected a total of 27 QTL for storage root dry matter content.

Due to the complex nature of the sweetpotato genome, genetic linkage mapping and QTL analysis has been more challenging in sweetpotato compared to other crop species. Though they have improved our basic understanding of the genome organization and polyploid nature of sweetpotato, the developed maps have been based on random amplified polymorphic DNA (RAPD), sequence-related amplified polymorphism (SRAP) and amplified fragment length polymorphism (AFLP) markers that have limited utility in breeding programs. To the best of our knowledge, there are currently no high density genetic linkage maps based on the more frequently utilized SNP markers to facilitate rapid and precise QTL detection. On the other hand, until recently (Fei et al., in preparation http://sweetpotato.plantbiology.msu.edu/) sweetpotato also lacks a whole genome sequence or gene annotation platform. This has hindered the validation of mapped QTL regions and/or isolation of target genes for cloning, gene pyramiding, marker assisted breeding, or any other relevant purposes.

Higher density linkage maps can be constructed by applying next-generation sequencing-based genotyping methods such as genotyping-by-sequencing (GBS), restriction
site-associated DNA sequencing (RAD-seq) and double digest RAD-seq to facilitate rapid discovery of thousands of single nucleotide polymorphism (SNP) markers in mapping populations through the sequencing of DNA fragments generated via restriction enzyme digestion (Baird, et al., 2008; Elshire, et al., 2011).

In this study, we phenotyped and genotyped two bi-parental mapping populations previously developed by Cervantes-Flores et al. (Cervantes-Flores, 2006) at North Carolina State University (NCSU) and Yada et al. (Yada, 2014) at the National Crops Resources Research Institute (NaCRRI). We used a GBS approach to develop SNP markers and develop high density genetic linkage maps for both populations. ‘Tanzania’ is a widely grown sweetpotato released landrace in sub-Saharan Africa. It is cream-fleshed, has high dry-matter content (32%) and is resistant to RKN and SPVD (Mwanga, et al., 2001). Beauregard is a popular U.S variety that is orange-fleshed, has low dry-matter content (18%), and is susceptible to sweetpotato weevil (SPW), RKN and SPVD (Rolston , et al., 1987). ‘New Kawogo’ is a white-fleshed, high dry-matter (32%), SPW and SPVD resistant released landrace cultivar popular in Uganda (Mwanga, et al., 2001). The ‘Tanzania’ by ‘Beauregard’ (TB) mapping population from NCSU consists of 240 progenies that are segregating for RKN resistance, among other quality traits (Cervantes-Flores, 2006). The ‘New Kawogo’ x ‘Beauregard’ mapping population consists of 287 progenies and is segregating for storage root yield, dry-matter, starch and β-carotene content, SPVD and SPW resistance (Yada, 2014).

The long-term objective of this study is to contribute to the development of breeder-friendly genomic tools that will facilitate sweetpotato improvement programs in SSA and to develop improved varieties more efficiently for faster genetic gains of key traits of economic
importance. This research, though it is focused primarily on SSA, also has significant spill-over potentials for the global sweetpotato community.

This dissertation is organized into four specific objectives:

1. Linkage mapping and QTL analysis of sweetpotato virus disease resistance in hexaploid sweetpotato
2. Identification of QTL for sweetpotato weevil resistance in a biparental mapping population.
3. Genetic mapping of QTL for storage root dry matter and β-carotene content in sweetpotato.
4. QTL analysis for root-knot nematode resistance in cultivated sweetpotato.
Literature Review

Origin, Genetics and Improvement

Sweetpotato, *Ipomoea batatas* (L.) Lam., is a clonally propagated perennial dicot belonging to the Convolvulaceae (morning glory) family (Austin, 1988). It is a hexaploid with 90 somatic chromosomes (2n=6X=90), has a large genome size of 1.5 Mbp (Wu et al., 2018), and is highly heterozygous, thus complicating studies on its genetics and cytology (Jones, 1986). Furthermore, cultivated sweetpotato is self- and cross-incompatible and each successful cross usually results in the production of less than 3 botanical seeds (Cervantes-Flores, et al., 2011), making its improvement through conventional breeding very challenging.

A number of hypotheses have been examined and studies conducted in an effort to explain the origin of sweetpotato. Research suggests that sweetpotato is an autopolyploid (Ukoskit, et al., 1997), exhibiting both autopolyploid and auto-allopolyploid types of inheritance (Ukoskit & Thompson, 1997; Kriegner, et al., 2003; Cervantes-Flores, et al., 2008a). An autopolyploid is a polyploid that originated by chromosomal doubling of a single species, resulting in several copies of the same genome. The analogous chromosomes can therefore randomly pair between each other (Cervantes-Flores, 2006; Cervantes-Flores, et al., 2008). However, an allopolyploid is a polyploid that originated from the abnormal hybridization between two different species, resulting in two different genomes in a single individual. In this scenario, chromosomes only pair to their homologs from the same genome, and to a lesser extent on the homeologous chromosome (Ramsey & Schemske, 2002). Shiotani and Kawase (1989) hypothesized the genome constitution of sweetpotato as $B_1B_1B_2B_2B_2B_2$, and suggested additional homology between the $B_1$ and $B_2$ genomes basing on the frequent formation of tetravalents and hexavalents. The degree of this homology, however, cannot be
accurately estimated because the genomic components of sweetpotato are still poorly understood (Yang, et al., 2017), though this is rapidly changing.

Among the approximately 50 genera and over 700 species of sweetpotato in the morning glory family, *Ipomea batatas* is the only species of major economic importance for its storage roots and vines (Woolfe, 1992). However, *I. aquatica* is also used in South East Asia as a raw salad, a cooked green vegetable, or animal fodder. The wild relatives of sweetpotato are mainly used for genetic studies due to their relatively simpler genomes (Nakayama, et al., 2010). Numerous studies on *I. trifida* have led to the conclusion that it is the closest relative and putative progenitor of cultivated sweetpotato (Hirakawa, et al., 2015).

**Molecular markers in sweetpotato improvement**

The use of molecular markers in sweetpotato improvement is still at its infancy despite the widely recognized global importance of the crop. This is due to the complex nature of the genetics of sweetpotato (Chang, et al., 2009) and also limited funding. Until recently, most of the studies have used dominant RAPD and AFLP markers which are relatively easy to score, or SSR and ISSR markers that have low abundance in the genome (Yada, et al., 2017).

More recently, advances in DNA sequencing methods through next-generation sequencing technology and bioinformatics resources have enabled large numbers of SNP markers to be developed and utilized (Mwanga, et al., 2017, Shirasawa, et al., 2017). SNPs are the most abundant DNA polymorphisms in the genome and can therefore be more readily utilized in a cost-effective next-generation genotyping platform.
Next-generation sequencing technologies have mostly been applied in SNP discovery, whose downstream application is linkage map construction, genetic diversity analysis, association mapping, QTL discovery and genomic assisted breeding.

**Genetic Linkage Map Construction in Sweetpotato**

Several studies have developed genetic linkage maps in hexaploid sweetpotato. To the best of our knowledge, only five genetic linkage maps have hitherto been developed for cultivated sweetpotato. The first linkage map of sweetpotato was reported by Ukoskit and Thompson (1997), who used 196 randomly amplified polymorphic DNA (RAPD) markers in a mapping population consisting of 76 progenies derived from a cross between “Vardaman” (early root yield, drought tolerant, susceptible to RKN) and “Regal” (resistant to RKN). The second map was developed by Kriegner et al. (2003) using a total of 632 and 435 amplified fragment length polymorphism (AFLP) markers in a segregating population consisting of 64 randomly selected clones derived from a cross between “Tanzania” (resistant to SPVD) and “Bikilamaliya” (susceptible to SPVD) respectively. The third map was developed by Cervantes-Flores et al. (2008a) using a total of 1166 and 960 AFLP markers for the “Tanzania” (resistant to RKN) and “Beauregard” (susceptible to RKN) maps respectively, derived from a mapping population consisting of 240 progenies. The fourth map was constructed by Li et al. (2010) using 473 and 328 sequence-related amplified polymorphism (SRAP) markers in a mapping population consisting of 240 individuals derived from a cross between “Luoxushu 8” (high starch content) and “Zhengshu” (low starch content). The fifth map was developed by Zhao et al. (2013a) using AFLP and SSR markers in a segregating F1 population consisting of 202 individuals derived from a cross between “Xushu 18” (high yield, moderate dry matter)
and “Xu 781” (low yield, high dry matter). However, due to the type of marker systems used, these maps consisted of many sweetpotato linkage groups containing relatively few DNA markers. These maps were also developed separately for each of the parents and these studies always ended up having two separate maps for the population.

SNP markers, which are codominant, overcome most of these challenges due to their abundance and distribution throughout the genome. SNPs can be used to increase the number of segregating loci, and they provide several advantages over previous methodologies. More recently, Shirasawa et al. (2017) developed a high-density SNP genetic map in sweetpotato, which was the first integrated genetic linkage map in a polysomic hexaploid (Shirasawa, et al., 2017). They used a double-digest RAD-seq strategy (Shirasawa, et al., 2016) to genotype an S1 population derived from selfing a single sweetpotato cultivar, “Xushu 18”. They mapped 28,037 double-simplex SNPs onto 96 linkage groups, covering a total distance of 33,020.4 cM. Their use of an S1 mapping population derived from self-pollination of a single cultivar greatly simplified their analysis of the genetic mode of inheritance (Shirasawa, et al., 2017). But their large map size (33,020.4 cM) means that it was substantially inflated, which typically occurs when genotyping errors are present. This map cannot therefore be used in its current form for QTL analysis or studies of the genetic architecture of important traits.

**QTL analysis in sweetpotato**

Using the developed genetic linkage maps, several studies have been conducted to map and identify genomic regions associated with traits of importance such as RKN resistance, dry-matter content, starch content, β-carotene content, and storage root yield. Cervantes-Flores et al. (2008a) conducted QTL mapping of RKN resistance (Cervantes-Flores, et al., 2008b).
identified nine chromosome regions that were associated with RKN resistance of which seven, \textit{rkn1} to \textit{rkn7}, were identified in “Tanzania”, and two, \textit{rkn8} and \textit{rkn9}, were in “Beauregard”. The QTL that were identified in “Tanzania” together with their interactions explained \~40\% of the variation in severity of infection. Nakayama \textit{et al.} (2012) conducted QTL mapping analysis and identified genomic regions with a large effect on resistance to southern root-knot nematode (SRKN) using sequence-specific PCR markers associated with QTL for resistance to multiple races of SRKN in sweetpotato.

Using a mapping population of 240 individuals, Cervantes-Flores \textit{et al.} (2011) conducted QTL mapping analysis for dry-matter, starch and \(\beta\)-carotene content. They observed transgressive segregation in all three traits and suggested that the traits were quantitative. In this study, they identified 13 QTL for dry-matter content, 12 for starch content and 8 for \(\beta\)-carotene. They further noted that dry-matter and starch content were highly correlated \((r > 0.8)\), while \(\beta\)-carotene content was negatively correlated with both starch content \((r = -0.6)\) and dry-matter content \((r = -0.3)\). It was therefore a qualified assumption that most of the QTL regions that significantly affected starch content also affected dry-matter content.

The first genetic map for sweetpotato that included 90 complete linkage groups was developed by Zhao \textit{et al.} (2013a). Using this map, they were able to subsequently conduct analyses and identify QTL for dry-matter content (Zhao \textit{et al.}, 2013a), starch content (Xiao-xia \textit{et al.}, 2014), and yield (Li \textit{et al.}, 2014). The linkage map that they used was based on AFLP and SSR markers, and a mapping population of 202 clones derived from a cross between “Xushu 18” (most popular cultivar in China, has high yield, moderate starch and dry-matter content), and “Xu 781” (has low yield, high starch and dry-matter content). They also observed
transgressive segregation in all three traits, albeit in a different population, and identified 27 QTL for dry-matter content, 8 QTL for starch content, and 9 QTL for storage root yield.

In addition to sweetpotato genome complexity, which poses a significant challenge to researchers in the community, the crop also lacks a whole genome sequence with gene annotation. This has hindered the validation of mapped QTL regions and the identification of target genes for further studies or targeted incorporation in breeding lines. Thus, hypothesis driven, genomic research on sweetpotato has been more difficult compared to other crops possessing more abundant genomic resources (Monden & Tahara, 2017).

**Storage root yield, dry matter, starch and β-carotene content**

There has been difficulty in improving sweetpotato for storage root yield, dry matter, starch and β-carotene content in SSA and Asia (Chang, et al., 2009). This is mainly because these traits are quantitatively inherited or appear to be so because of the polyploid nature of the crop. These traits are also controlled by multiple genes with significant genotype by environment interaction effects that further complicate improvement efforts (Adebola, et al., 2013; Yada, et al., 2011). Additionally, storage root dry matter and β-carotene content are negatively correlated traits, which further complicate selection efforts (Gruneberg, et al., 2005; Cervantes-Flores, et al., 2011).

Improving these traits, like most others in sweetpotato, has been through the use of mass selection breeding strategies (Jones, 1986). In hexaploid sweetpotato, this relatively simple method has not been effective at increasing genetic gain from selection (Yada, 2014). The most common molecular tools that have been utilized (RAPD, AFLP and SSR) can only be analyzed as dominant markers for polyploid species because different doses of
heterozygotes cannot be differentiated. These markers have been used in genetic linkage mapping (Kriegner, et al., 2003; Cervantes-Flores, et al., 2008a), QTL analysis for root-knot nematode resistance, storage root dry matter content, starch content, and β-carotene content (Cervantes-Flores, et al., 2008b; Chang, et al., 2009; Zhao, et al., 2013a). Nonetheless, none of these QTL have been utilized in marker assisted selection schemes for sweetpotato improvement due to the inherent limitations of the marker technology in polyploids (Yada et al., 2017d).

**Sweetpotato virus disease and its management**

There are over 30 viruses that have been reported to infect sweetpotato globally (Clark, et al., 2012). These belong to the crinivirus, potyvirus, ipomovirus, badnavirus, carlavirus, cucumovirus, and begomovirus genera. Only six out of all these have been reported in Uganda and these are *Sweet potato feathery mottle virus* (SPFMV), *Sweet potato chlorotic stunt virus* (SPCSV), *Sweet potato chlorotic flecks virus* (SPCFV), *Sweet potato collusive virus* (SPCV), *Sweet potato mild mottle virus* (SPMMV) and *Sweet potato leaf curl Uganda virus* (SPLCUV) (Gibson, et al., 1998; Aritua, et al., 2007). Sweetpotato virus disease (SPVD) is a severe disease complex resulting from the dual infection by SPCSV, transmitted by the silverleaf whitefly [*Bemisia tabaci* (Gennadius)], and SPFMV, transmitted by the aphid [*Aphis gossypii* (Glover)] (Gibson, et al., 1997; Karyeija, et al., 2000).

SPVD causes up to 98% yield losses in SSA (Gibson, et al., 1998; Karyeija, et al., 2000), approximately 30% yield losses in poorly managed fields in USA (Clark & Hoy, 2006), and over 20% yield losses in China (Feng, et al., 2000). Though many varieties of sweetpotato are resistant to SPFMV, the synergistic, dual infection of SPCSV usually leads to break down
in SPFMV-resistant varieties (Aritua, et al., 1998). Plants infected by SPVD are easily recognizable by farmers due to their severe, clear field symptoms which include stunting, leaf narrowing and distortion, chlorosis, and mosaics (Gibson, et al., 1998).

In SSA where there is high virus pressure, the main approach for management of SPVD has been the removal of infected plants from the field and the use of clean planting material (Aritua, et al., 1998; Byamukama, et al., 2004). This approach has not been very effective due to the absence of a formal sweetpotato seed system in SSA (Tairo, et al., 2005). On the other hand, the limited sources of resistance to SPVD further complicates its management, with only a few demonstrated sources of resistance (or tolerance when the resistant clone is not clean – Robert Mwanga personal communication) in “New Kawogo”, “Tanzania” (Aritua, et al., 1998; Mwanga, 2001), and some wild relatives of sweetpotato (Karyeija, et al., 1998).

Conventional breeding for resistance to SPVD has been difficult and challenging (Yada, et al., 2017b). In SSA, mass selection has been primarily employed to select clones to advance based on observation of SPVD symptoms in the field (Mwanga, et al., 2001). This involves assembly and screening of large populations of sweetpotato genotypes for selection of parents and clones for the next breeding cycle (Yada, et al., 2011). However, this approach is limited by environmental changes (variability) depending on presence or absence of vectors during the season, is slow and generally expensive in the long run.

Few molecular approaches have been exploited in breeding for SPVD resistance. Mwanga et al. (2002) identified two QTL, spfm1 and spcsv1, for SPFMV and SPCSV from a bi-parental cross between “Tanzania” and “Bikilamaliya”. They used AFLP and RAPD markers to map these QTL. In a similar study, Mcharo et al. (2005) and Miano et al. (2008) associated AFLP markers with SPVD resistance using discriminant analysis and logistic
regression. However, the dominant nature of inheritance of AFLP and RAPD markers has limited their utility in breeding programs (Yada, et al., 2017b). In a more recent study, Yada et al. (2017b) used regression analysis and they associated SSR markers with SPVD resistance in the bi-parental “New Kawogo” x “Beauregard” cross.

The Sweetpotato weevil

The sweetpotato weevil, Cylas spp, is the most serious insect pest of sweetpotato worldwide. C. puncticollis (Boheman) and C. brunneus (Fabricius) occur in Uganda, and they are uniquely African species (Downham, et al., 2001). SPW can cause crop losses of up to 98% in areas with low rainfall and long dry spells (Smit, 1997). The larvae feed on storage roots but are not always readily observed until they have caused significant damage. Adults are also difficult to detect given their nocturnal habitat. Damage by SPW is the most important production constraint of sweetpotato in SSA (Stevenson, et al., 2009).

The greatest damage is inflicted on storage roots as the larvae feed (Cockerham, et al., 1954) and secondary pathogen infection, triggering the storage roots to produce sesquiterpenes in response to the damage, which render the once edible roots unfit for consumption (Uritani, et al., 1975). SPWs and their larvae have a concealed feeding behavior as they tunnel through storage roots while feeding. This makes their management difficult (Nottingham & Kays, 2002; Odongo, et al., 2003). Studies have indicated that resistance to SPW is an active process conferred by hydroxycinnamic acid (HCA) esters present in the root latex of resistant cultivars (Stevenson, et al., 2009; Anyanga, et al., 2017). The HCA esters were identified as hexadecylcaffeic acid, hexadecylcoumaric acid, heptadecylcaffeic acid, octadecylcaffeic, octadecylcoumaric acid and 5-0-cafeoylquinic acids (Stevenson, et al., 2009; Anyanga, et al.,
Host plant resistance provides an effective component of any integrated pest management program (IPM) of sweetpotato (Anyanga et al., 2017). However, the development of weevil resistant varieties has not been successful over the years due to the lack of heritable resistance in the material evaluated (Anyanga et al., 2017).

With the aim of identifying the biochemical basis of resistance to SPW, Muyinza et al., (2012) and Anyanga et al. (2013) screened a wide range of sweetpotato genotypes against the highly susceptible cultivar “Tanzania” as a control. They observed considerable variation in the concentration of HCA esters, with “New Kawogo”, an African landrace cultivar from Uganda recording the highest concentration of these compounds. HCA esters located in the storage root latex of resistant cultivars are weakly associated with resistance, with an effect against adult weevil oviposition and feeding (Anyanga et al., 2013). Differences in the concentration of these compounds between varieties might explain differences in resistance in the African sweetpotato gene pool (Anyanga et al., 2017) which is highly heterogeneous in their susceptibility to the pest (Muyinza et al., 2012). Research conducted by Anyanga, et al., (2013) using liquid chromatography-mass spectrometry (LC-MS) on root surface and epidermal extracts (1 min in hexane and 24 h in methanol) and observed significant variation in the concentration of hexadecyl, heptadecyl, octadecyl, and quinic acid esters of caffeic and coumaric acid, with higher concentrations of these compounds correlated with resistance. They proceeded to synthesize all these compounds to enable their positive identification. It was concluded that the selection of sweetpotato varieties with higher levels of HCA’s, particularly in the storage root surface might be used to contribute to the development of resistance to SPW. However, this would entail screening of hundreds of breeding lines using freeze-dried samples of roots and analysis by LC-MS in order to achieve long term population improvement.
Currently, efforts to develop such a chemotyping platform are still too expensive (Yada, et al., 2017c).

In his study on the segregation of HCA esters in sweetpotato, Anyanga, et al (2017) observed significant differences in the total HCA esters among the genotypes of the segregating bi-parental population (287 individuals) of the “New Kawogo” (SPW resistant) X “Beauregard” (SPW susceptible) NKB cross. They attributed this to genetic effects at multiple loci in the sweetpotato genome for this trait. Earlier studies have pointed out that the female parent in this bi-parental cross (New Kawogo) has a high concentration of HCA esters on its root surface, and high field and laboratory resistance to SPW (Stevenson, et al., 2009; Muyinza, et al., 2012; Anyanga, et al., 2013). This population also had progeny that were more resistant than “New Kawogo” and others with higher HCA concentrations. These genotypes were selected as new candidates to be used in a recurrent selection breeding program seeking improvement of this trait. All these studies have shown that resistance to sweetpotato weevils is an active process rather than an escape mechanism by deep rooting genotypes as has been hypothesized by others (Stevenson, et al., 2009; Muyinza, et al., 2012; Anyanga, et al., 2013).

Several recently published studies have contributed much to our understanding of the genetic and biochemical basis of resistance to SPW’s in New Kawogo, the landrace source of resistance to SPW. Yada, et al., (2015) examined the diversity in the NKB population using SSR markers to determine the genetic basis of resistance to weevils and identify SSR markers, and he also conducted studies to determine the genetic basis of resistance and identify SSR markers linked to SPW resistance using logistic regression (Yada, et al., 2017c). Anyanga, et al., (2017), conducted additional SPW resistance studies with the NKB materials which were focused on improving our understanding of biochemical basis of resistance present in New
Kawogo. These key studies were able to report on the first use of co-dominantly inherited SSR markers to identify SPW resistance loci in sweetpotato (Yada, et al., 2017c), the quantification of plant chemicals (HCA esters) that confer resistance to SPW, and the evaluation of levels of insect damage of the NKB mapping populations in the field and laboratory (Anyanga, et al., 2017). Data collected from these studies has shown that these chemical traits are controlled quantitatively, with moderate broad sense heritability ($H^2=0.49$) for weevil resistance. In this research, we have used next-generation sequencing technology to identify genome-wide SNP markers for this same population under the Genomic Tools for Sweetpotato Improvement (GT4SP) project. This has facilitated the development of a linkage map, and the identification of quantitative trait loci (QTL) associated with SPW resistance and HCA concentration for use in marker assisted selection (MAS) and genomic selection (GS).

**Root-knot nematode in sweetpotato**

Plant parasitic nematodes have been reported to affect root and tuber crops in Uganda including sweetpotato (Namaganda, et al., 1993; Makumbi-Kidza, et al., 2000), cassava and banana (Coyne, et al., 2003; Coyne, et al., 2006). Root-knot nematodes (*Melodogyne* spp) (RKN) are worldwide distributed and affect virtually every cultivated crop. On average, the damage inflicted on field crops by plant-parasitic nematodes stands at about 10% worldwide (Whitehead, 1998). It is clear that much of the damage by RKNs goes undetected due to the fact that they feed on the underground portions of plants. In sweetpotato, in addition to feeding on the roots, RKN’s also reduce the market value of the sweetpotato storage roots by directly affecting their quality through the presence of cracks and secondary diseases that infect damaged storage roots (Cervantes-Flores, et al., 2008b).
Parasitic nematodes destroy plants by feeding on root tissue and thereby reducing their ability to take up nutrients, promoting secondary infection from soil microbes and fungi through wound sites, and by serving as vectors for various pathogenic viruses. Of all plant parasitic nematodes, RKNs are the most abundant and most damaging (Sasser, 1980). They form rounded galls (knots) on the root vascular tissue of infected plants (Appendix B) thus disrupting their capacity to uptake nutrients from the soil. On the fleshy storage roots of sweetpotatoes, the infection will appear as cracks although this can be associated with various other environmental factors like soil texture and moisture (Lawrence, et al., 1986). In their research, Lawrence et al. (1986) suggested that RKNs may be predisposing the roots to cracking rather than a causal factor of cracking after observing that the number of cracked roots was not correlated with initial RKN population number. This was after they observed that when rainfall was more uniform, the storage roots did not crack although nematodes were present in them.

Managing RKNs has in the past involved the use of toxic nematicides in combination with cultural practices. However, notwithstanding the obvious health and environmental risk that nematicides pose, their cost is prohibitive to small-scale farmers and growers of the crop due to its relatively low market value (Gasapin, 1984). The safest, most sustainable and economic route is the use of host plant resistance. However, “Beauregard” – the sweetpotato cultivar that was predominant in USA prior to release of “Covington” (Yencho, et al., 2008), is highly susceptible to M. incognita and M. javanica (Cervantes-Flores, et al., 2002).

The genetic mode of resistance of sweetpotato to RKN is not well understood. Ukoskit et al. (1997) hypothesized single gene qualitative resistance whereas multiple gene quantitative resistance has been hypothesized by several other researchers (Cordner, et al., 1954; Giamalva,
et al., 1961; Jones & Dukes, 1980; Cervantes-Flores, et al., 2008b). Jones and Dukes (1980) further suggested that independent sources of resistance to different strains of RKN were responsible for the observed differences in inheritance and that the genes originated from multiple origins (Mcharo, et al., 2005b). Histological studies have shown that RKNs in the juvenile (J2) stage penetrate both susceptible and resistant sweetpotato clones as well as other *Ipomoea* species (Komiyama, et al., 2006). However, they observed that the pathogen only establishes in susceptible genotypes in agreement with observations by earlier researchers (Dropkin, 1969; Paulson & Webster, 1972).

Molecular markers have been widely used in many different crops to identify and map genes associated with resistance to nematodes (Barr, et al., 1998; Wang, et al., 2001; Ynturi, et al., 2006) but very few of such studies have been successfully conducted in sweetpotato (Ukoskit, et al., 1997; Mcharo, et al., 2005b; Cervantes-Flores, et al., 2008b). Ukoskit et al. (1997) used RAPD markers in a population of 71 individuals derived from a cross between “Regal” (resistant) and “Vardaman” (susceptible) and identified the marker OP151500 was weakly associated (P=0.037) with RKN resistance in the cross. Mcharo et al. (2005b) employed two unrelated sweetpotato populations and applied logistic regression and discriminant analysis to study RKN resistance. They report the ability to predict and classify the phenotype with an accuracy of 89% and 88%. However, since their population was small, they could have over-estimated the effect of the markers (Cervantes-Flores, et al., 2008b). Cervantes-Flores et al. (2008b) hypothesized that resistance to RKN is conferred by several genes, basing on molecular and phenotypic data. They detected nine QTL associated with RKN resistance, each of which showed a relatively small genetic effect. More recently, Nakayama et al. (2012) conducted a multi-race analysis of resistance to multiple races of southern root-
knot nematode (SRKN) and suggested that race-specific resistance is more likely conferred by single genes and that the genes for resistance against each race are closely located (Nakayama, et al., 2012). However, just like Ukoskit et al. (1997), their population consisting of 86 F1 progeny plants was small and they therefore could have overestimated the effects of markers they used in the study and also could have missed capturing more alleles arising from different recombination events. The importance of population size to study the inheritance of a trait and to detect QTL has been strongly emphasized by several researchers (Collard, et al., 2005; Doerge, et al., 1997; Cervantes-Flores, et al., 2008a). In a polyploid crop such as sweetpotato, a large population size increases the statistical power and is critical to capture the full spectrum of underlying allelic variation arising from very many possible recombination events (Collard, et al., 2005; Kriegner, et al., 2003; Cervantes-Flores, et al., 2008b).
REFERENCES


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


CHAPTER 2

Linkage mapping and QTL analysis of sweetpotato virus disease resistance in hexaploid sweetpotato

(In a format suitable for submission to Theoretical and Applied Genetics, Springer Berlin Heidelberg)
Linkage mapping and QTL analysis of Sweetpotato Virus Disease Resistance in hexaploid sweetpotato

Bonny Michael Oloka\textsuperscript{1,2}, Benard Yada\textsuperscript{2}, Guilherme da Silva Pereira\textsuperscript{3}, Marcelo Mollinari\textsuperscript{3}, Bode A. Olukolu\textsuperscript{1}, Milton O. Anyanga\textsuperscript{2}, Doreen Chelangat\textsuperscript{2}, Paul Musana\textsuperscript{2}, Agnes Alajo\textsuperscript{2}, Zhao-Bang Zeng\textsuperscript{1}, G. Craig Yencho\textsuperscript{1}

\textsuperscript{1}Department of Horticultural Science, North Carolina State University, 214 Kilgore Hall, Box 7609, Raleigh, NC 27695-7609, USA
\textsuperscript{2}National Crops Resources Research Institute, Namulonge, P.O. Box 7084, Kampala, Uganda.
\textsuperscript{3}Bioinformatics Research Center, Department of Horticultural Science, North Carolina State University, NC 27695-7566

Received for publication _______________2019. North Carolina State University. This paper is a portion of a dissertation submitted by Bonny Michael Oloka. The research reported in this publication was supported by funding provided by the Bill and Melinda Gates Foundation (OPP1052983) through the Genomic Tools for Sweetpotato Improvement Project (GT4SP). We thank Joan Adero and Jennipher Asiimwe for their assistance in DNA extraction, and Hannah Huntley for GBS library prep and sequencing.
Corresponding author email: craig_yencho@ncsu.edu
Abstract

Resistance to sweetpotato virus disease (SPVD), a complex interaction of *Sweet potato chlorotic stunt virus* (SPCSV; *Crinivirus*) and *Sweet potato feathery mottle virus* (SPFMV; *Potyvirus*), has been identified as one of the main biotic stress breeding objectives of sweetpotato improvement programs in sub-Saharan Africa. The complex genetics of sweetpotato has limited understanding of the genomic architecture of the trait and impeded sweetpotato improvement. The objective of this study was to identify quantitative trait loci (QTL) for SPVD in a biparental sweetpotato mapping population consisting of 287 individuals segregating for SPVD resistance. The ‘NKB’ population was derived from a cross between ‘New Kawogo’ (NK), a Ugandan landrace cultivar, and ‘Beauregard’ (B), a major cultivar in the United States. The population was phenotyped using a standardized visual rating scale in three sites in Uganda for two seasons. The broad sense heritability of SPVD was estimated at 0.487. REML variance components analysis of genotype marginal means generated within environments (site and seasons) and adjusted means across all environments were used for QTL analysis using a random effects QTL mapping model implemented using the R-based software ‘*QTLpoly*’, utilizing SNP marker dosage information and the NKB genetic map. We identified 2 QTL for SPVD resistance on linkage groups 1 and 2 explaining a total variability of 36%. Additive effects of the QTL alleles from both parents were estimated and important alleles influencing mean SPVD score were identified. This work represents a significant step forward in our understanding of the genetic architecture of SPVD resistance and sets the stage for future utilization of genomics-assisted breeding in applied sweetpotato breeding programs.
Introduction

Production of sweetpotato, *Ipomoea batatas* (L.) Lam., is severely constrained in eastern Africa and globally by SPVD. The SPVD disease complex, which can be severe and leading to plant death, is one of the main constraints to sweetpotato production in sub-Saharan Africa (SSA), and it is a result of the dual infection by SPCSV, transmitted by the silverleaf whitefly [*Bemisia tabaci* (Gennadius)], and SPFMV, transmitted by the aphid [*Aphis gossypii* (Glover)] (Gibson, et al., 1997; Karyeija, et al., 2000). There are over 30 viruses that have been reported to infect sweetpotato globally (Clark, et al., 2012). However, only six have been reported in Uganda and these are Sweet potato feathery mottle virus (SPFMV), Sweet potato chlorotic stunt virus (SPCSV), Sweet potato chlorotic flecks virus (SPCFV), Sweet potato collusive virus (SPCV), Sweet potato mild mottle virus (SPMMV) and Sweet potato leaf curl Uganda virus (SPLCUV) (Gibson, et al., 1998; Aritua, et al., 2007).

SPVD causes up to 98% yield losses in sub-Saharan Africa (SSA) (Gibson, et al., 1998; Karyeija, et al., 2000) and over 20% yield losses in China (Feng, et al., 2000). Though many varieties of sweetpotato are tolerant to SPFMV, the synergistic, dual infection of SPCSV usually leads to break down in SPFMV-tolerant varieties (Aritua, et al., 1998). Plants infected by SPVD are easily recognizable by farmers due to their severe, clear field symptoms which include stunting, leaf narrowing and distortion, chlorosis, and mosaics (Gibson, et al., 1998; Mwanga, et al., 2002).

In regions where there is high virus pressure, the main approach for management of SPVD has been the removal of infected plants from the field and the use of clean planting material (Aritua, et al., 1998; Byamukama, et al., 2004). This approach has not been very effective due to the absence of a formal sweetpotato seed system in SSA (Tairo, et al., 2005).
The limited sources of resistance to SPVD further complicates its management, with only a few demonstrated sources of resistance present in the released land races ‘New Kawogo’, ‘Tanzania’ (Aritua, et al., 1998; Mwanga, 2001), and some wild relatives of sweetpotato (Karyeija, et al., 1998). SPVD management has also been approached using biotechnology by transforming some eastern African varieties with the coat protein gene of russet crack strain of SPFMV, but the transgenic plants succumbed to SPCSV in the field (Okada, et al., 2002; Tairo, et al., 2005).

Conventional breeding for resistance to SPVD has been both difficult and challenging (Yada, et al., 2017). Mass selection has been primarily employed to select resistant clones to advance based on observation of SPVD symptoms in the field (Mwanga, et al., 2001). This involves assembly and screening of large populations of sweetpotato genotypes for selection of parents and clones for the next breeding cycle (Yada, et al., 2011). This approach towards population improvement and product development is, however, limited by environmental variability, and it is slow and generally expensive in the long run.

Only a few molecular genetic studies have been conducted in breeding for SPVD resistance. Mwanga et al. (2002) identified two recessive genes, \textit{spfm1} and \textit{spcsv1}, for SPFMV and SPCSV from a bi-parental cross between ‘Tanzania’ and ‘Bikilamaliya’. They used amplified fragment length polymorphism (AFLP) and random amplified polymorphic DNA (RAPD) markers to identify markers linked to these hypothesized genes. In a similar study, Mcharo et al. (2005) and Miano et al. (2008) associated AFLP markers with SPVD resistance using discriminant analysis and logistic regression. However, the dominant nature of inheritance of AFLP and RAPD markers has limited their utility in breeding programs (Yada, et al., 2017b). In a more recent study, Yada et al. (2017) used regression analysis and they
associated simple sequence repeat (SSR) markers with SPVD resistance in a bi-parental population derived from a cross between New Kawogo and Beauregard.

All of the marker-assisted breeding studies of SPVD resistance have relied on marker data that are not fully informative because they have utilized loci with low allelic doses. For a complex polyploid like sweetpotato, it is preferable to utilize higher dose markers because each marker locus needs to be positioned in a homology group (Garcia, et al., 2013) to capture the full spectrum of recombination events. In a segregating sweetpotato population derived from two phenotypically distinct cultivars, different numbers of segregating alleles exist across the genome. In order to guarantee reasonable coverage of its genome for genetic characterization and analysis, a relatively large number of single and multiple dose markers are required (Oloka, et al., 2018). With the advent of high throughput next generation genotyping, it is now possible to identify and utilize single nucleotide polymorphism (SNP) markers for genetic studies in sweetpotato. One of the major advantages for the utilization of SNPs for genetic analysis in polyploids is their allowance for the direct observation of all genotypes at a given locus (Garcia, et al., 2013). Consider one example when segregating individuals in an autohexaploid sweetpotato population are being evaluated for a given locus: AAAAAA, AAAAAa, AAAAAa, AAAaaa, AAAaaa, Aaaaaa and aaaaaa; using the A allele as reference. The number of copies of the reference allele is the dosage of that allele. When a marker system such as SSRs or AFLPs, is used to evaluate such individuals, they are scored as 1 (allele present) for individuals possessing the following genotypes: AAAAAA, AAAAAa, AAAaaa, AAAaaa, AAAAAa, and Aaaaaa; or 0 (allele absent) for aaaaaa (Garcia, et al., 2013) resulting in a loss of a considerable amount of genetic information. However, good quality SNPs can provide allele dosage information contained within a segregating locus, and thus is preferred in genetic
studies. The development of modern genotyping protocols like genotyping-by-sequencing (Elshire, et al., 2011) or RAD seq (Etter, et al., 2011) and their modification and optimization for highly heterozygous and hexaploid sweetpotato (Olukolu et al, in preparation) together with the recent development of two high quality reference genomes of *Ipomoea trifida* (V.3.0) and *Ipomoea triloba* (V.3.0), two of the wild diploid progenitors of cultivated sweetpotato (Wu et al, 2018) has enabled the evaluation of SNP markers throughout the genome of hexaploid sweetpotato.

When SNPs are identified in a crop, they are used for a number of purposes including, but not limited to; quantitative trait loci (QTL) identification, candidate gene identification and comparative genomic breeding research. These are all made possible with a detailed molecular genetic linkage map and a reference genome (Monden & Tahara, 2017). To date, the most commonly used approach for linkage map construction in outcrossing heterozygous species has been based on the use of single-dose restriction fragments (SDF) (Wu, et al., 1992) combined with other mapping strategies such as testcross and pseudo-testcross (Grattapaglia & Sederoff, 1994). In this approach, linkage mapping is conducted for each parent separately using single dose markers, resulting in the development of two parental maps. Simplex markers segregate in the F1 generation in a 1:1 fashion and regardless of the ploidy of the genome, thus half of the progeny will carry this marker and the other half will not. In this scenario, the working hypothesis is that only simplex, duplex and triplex marker dosages produce polymorphisms in the progeny. This is based on the four hypothesized expected segregation ratios for the inheritance of a dominant marker, predicted from cytological characteristics (autohexaploid, tetradiploid, allohexaploid) (Jones, 1965; Kriegner, et al., 2003; Cervantes-Flores, et al., 2008a).
A theoretical model for the prediction of marker genotypes in autotetraploid species was earlier developed by Hackett, et al (1998), which was later adopted in the development of several linkage maps in potato (Meyer, et al., 1998; Schumann, et al., 2017; Luo, et al., 2001) by analyzing marker phenotypes from segregating data of parents and progeny. This model is utilized in the software TetraploidMap (Hackett & Luo, 2003) and has been used to develop linkage maps in blackberry (Castro, et al., 2013), rose (Gar, et al., 2011), potato (Bradshaw, et al., 2008) and alfalfa (Julier, et al., 2003). However, TetraploidMap is not applicable for ploidy levels beyond tetraploids. To simplify the methodology for identifying SNP markers while building a genetic linkage map in hexaploid sweetpotato, Shirasawa et al (2017) used an S1 population derived from self-pollination of a single parent and the OneMap program (Margarido, et al., 2007) to build a map. However, they only used SDF markers (Wu, et al., 1992), thus they were not able to capture much of the information from all segregating loci in the cross.

The work we present here is part of a larger effort under the Genomic Tools for Sweetpotato Improvement Project (GT4SP), with the overarching aim of generating core genomic, genetic and bioinformatics resources for sweetpotato improvement. One of the key genomic resources developed under this project was a high density ‘Beauregard’ x ‘Tanzania’ (BT) integrated genetic linkage map (Mollinari et al, in prep). To the best of our knowledge, the BT map is the highest density, correctly ordered and phased SNP-based genetic linkage map of hexaploid sweetpotato to date. Here, we present a high-density SNP-based genetic linkage map of hexaploid sweetpotato developed from the ‘New Kawogo’ x “Beauregard” (NKB) mapping population. We used the recently described genome sequence of I. trifida and I. triloba as a reference to detect SNPs and to identify homology groups. Our NKB mapping
population consists of 287 individuals segregating for various traits of importance in the sweetpotato breeding community (Yada, 2014). The NKB map, SNP markers and their genotypes were used to identify QTL for SPVD in sweetpotato and we discuss its potential application in sweetpotato breeding programs.
Materials and Methods

Mapping population

The NKB biparental mapping population was previously described by Yada et al., (2015) and was developed in Uganda at the National Crops Resources Research Institute (NaCRRI). The population contains 287 individuals and was generated in 2010 by crossing the high storage root dry-matter, high starch and low β-carotene content Ugandan landrace cultivar, New Kawogo (female) (Mwanga, et al., 2001), with the low storage root dry-matter, low starch and high β-carotene content US cultivar, Beauregard (male) (Rolston, et al., 1987). Phenotyping trials were conducted in three sites for two seasons in 2012. All trials were arranged in a randomized complete block design with 3 replications per site.

Field trials and screening for SPVD resistance

SPVD phenotyping of the NKB population was conducted at three locations: NaCRRI (0°32’ N, 33°35’ E, 1,160 masl); National Semi-Arid Resources Research Institute (NaSARRI) (1°32’ N, 33°27’ E, 1,085 masl); and Ngetta Zonal Agricultural Research and Development Institute (NgeZARDI) (2°202’ N, 33°62’ E, 1,080 masl). The first season trial (2012A) was planted in June 2012 and harvested in November 2012 while the second season trial (2012B) was planted in November 2012 and harvested in May 2013. The delayed harvesting (5 to 6 months after planting) was intended to allow ample exposure to virus pressure. SPVD resistance was measured by disease severity scores of symptoms on the leaves and stems at harvest. A standard disease score rating of 1 (very resistant) to 9 (very susceptible) was used (Gruneberg, et al., 2010).
DNA extraction and genotyping

Genomic DNA was extracted at the NaCRRI Biosciences Laboratory from young leaves (ca. 100g) of each of the progeny and parents using a modified C-TAB method (Doyle & Doyle, 1990) and quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, Delaware). The DNA samples were subsequently shipped to the Department of Horticultural Science at North Carolina State University for genotyping and further analysis.

Library preparation and sequencing was performed at the Genomic Science Laboratory (GSL), North Carolina State University. An optimized genotyping-by-sequencing library preparation protocol (GBSpoly) for highly heterozygous polyploids (Olukolu et al, in preparation) was used to genotype the NKB population. Briefly, the GBSpoly libraries were constructed with two combinations of restriction enzymes, CviAII and Tsel, that were added sequentially. Following restriction digestion, the DNA fragments were purified using AMPure XP magnetic beads (ThermoFisher), cleaned and quantified on a plate reader using the picogreen assay to normalize the concentration of DNA in the samples to between 5 to 10 ng/μl. A set of 96 barcoded adaptors with complementary sequences to Tsel and CviAII overhang sequence were prepared and ligated to the DNA. The samples were pooled by pipetting 5 μl of each ligated DNA sample to make 64 and 96 samples per pool. After washing the pooled libraries, a size selection step using PippinPrep (Sage Science) was included to ensure uniform amplification of fragments (i.e. avoiding PCR bias between large and small fragments). The libraries were comprised of fragments (adapter and genomic insert) within a range of 300-600 bp. Small peaks outside the set range were eliminated by additional Megabead purification.
For sequencing, a 1:25 dilution of the 60 μl amplified library was prepared, quantified, and the values of the amplicons were calculated and converted to nmol/l. The stock amplified library was diluted to 10 nmol/l into 20 μl for sequencing on an Illumina HiSeq 2500 (Illumina, Inc. San Diego, CA). The NKB population was run in single-end sequencing mode, with each of the parents replicated 10 times. DNA for 96 genotypes that had low quality and consequently resulted in low sequence coverage were run on an additional 6 lanes for a total of 12 lanes. The entire NKB population was run on 36 Illumina lanes total.

**Sequence data processing**

The 8-bp buffer designed into the adapter sequence was trimmed off from the raw FASTQ sequence reads using fastx_trimmer in FASTX-Toolkit (version 0.0.13) (http://hannonlab.cshl.edu/fastx_toolkit). The filtered reads (64-bp-long sequence tags) were mapped onto the *I. trifida* (v.3 assembly JBrowse) and the *I. triloba* (v.3 assembly JBrowse) genome sequences (http://sweetpotato.plantbiology.msu.edu/download) as a reference using Bowtie 2 (version 2.2.3.1) (https://sourceforge.net/projects/bowtie-) (Langmead, et al., 2009), with default parameters, which has a very sensitive-local argument. The resulting sequence alignment/map (SAM) files were converted to binary sequence alignment/map (BAM) files and subjected to SNP calling using a modified TASSEL-GBS pipeline (Glaubitz, et al., 2014) to obtain variant call format (VCF) files that contain read counts.

**SNP mining and genotype calling**

The filtering process consisted of both the *I. trifida* and *I. triloba* reference genomes. High-quality variants and genotypes were called using SuperMASSA software (Serang, et al., 2012) with an average read depth filtering criteria requirement of 150 reads per SNP locus. SNPs with less than 20 reads on average were removed. In the next round of quality control
and genotype calling, we removed monomorphic SNPs together with those classified by SuperMASSA as diploids and tetraploids. This was important because classifications other than hexaploid amounts to SNPs with questionable quality. Those SNPs that had more than 25% missing data were also removed. For a data point to be classified as missing, its highest posterior probability was set to be smaller than 0.8. A chi-square test for Mendelian segregation was performed on the remaining SNPs using MAPpoly software (https://github.com/mmollina/MAPpoly). Based on the segregation distortion under polysomic inheritance and considering the amount of missing data, this test was set at \( P < 5 \times 10^{-4} \). SNPs that were inherited from genomic regions with high LD were removed because they provide redundant information. This SNP mining and genotype calling process has been described and explained in detail in Mollinari et al (in preparation).

**Linkage mapping, marker phasing and ordering**

Here, we briefly mention what has been described in detail by Mollinari et al (in preparation). Using MAPpoly software, a two-point recombination fraction analysis was performed on all the filtered polymorphic markers in the population of 287 individuals to estimate all pairwise recombination fractions in the data, and to form a two-point “object”. This two-point “object” was then converted into recombination fraction and LOD score (SNP combinations/linkage phase) matrices. LOD score thresholds were imposed for the most likely linkage phase configuration and recombination fraction for all points on the recombination fraction matrix using maximum likelihood. The recombination fraction matrix, containing phased haplotypes (obtained using a built-in algorithm in MAPpoly) from both parents, was estimated for a given order of markers using multi-dimensional scaling [MDS (Preedy and
Hackett 2016)) to order and reconstruct the map. The final map was ordered and re-estimated using the Hidden-Markov model (HMM) (Mollinari et al., in prep).

**Genetic map and conditional probabilities**

Dosage calling was performed using SuperMASSA software (Serang, et al., 2012) and the 1,371-cM genetic map containing 7,184 filtered SNPs was obtained using MAPpoly. The NKB map is a high density, integrated SNP-based genetic linkage map containing all 15 linkage groups of hexaploid sweetpotato with a marker density averaging 5.2 SNPs/cM. Genotype probabilities were estimated given the dosage at each marker. Assuming no double reduction, there are 400 possible genotype configurations coming from independent inheritance of two chromosomal segments from each hexaploid parent. Having obtained the complete phase information for each parental homolog, the most likely genotype for each individual progeny was determined given the dosage at each marker. To infer the most likely combination of parental homologs in each individual offspring for each position along the chromosome, we used a Hidden Markov Model (Rabiner, 1989; Hackett, et al., 2013; Rak, et al., 2017). To do this, the genotype conditional probabilities were calculated every 1 cM for the whole genome, and for each genotyped individual, we obtained conditional probabilities for all positions in the genome.

**Phenotypic data and QTL analysis**

Restricted maximum likelihood (REML) variance components analysis of the phenotypic data was performed both within sites and seasons and across all environments to generate genotype predicted means for QTL analysis using the following mixed model:

\[ y_{ijmn} = \mu + l_m + s_n + b_{j(mn)} + g_{imn} + e_{ijmn} \]
Where $y$ is the observed value of genotype $i$, $\mu$ is the population mean of the trait, $l_m$ is the fixed effect of location $m$, $s_n$ is the fixed effect of season $n$, $b_{j(mn)}$ is the random effect of block $j$ nested within location $m$ and season $n$ with $b_{j(mn)} \sim N(0, \sigma_b^2)$, $g_{imn}$ is the random effect of genotype $i$ in block $j$ within location $m$ and season $n$ with $g_{imn} \sim N(0, \sigma_g^2)$, and $e_{ijmn}$ is the residual error with $e_{ijmn} \sim N(0, \sigma_e^2)$. To generate variance components that were used for calculating the heritability of SPVD, we used the following model;

$$y_{ijk} = \mu + env_k + b_{j(k)} + g_i + g_{ik} + e_{ijk}$$

Where $env_k$ is the fixed effect of environment $k$ (for a given site and season), $b_{j(k)}$ is the random effect of block $j$ nested within environment $k$ with $b_{j(k)} \sim N(0, \sigma_b^2)$, $g_i$ is the random effect of genotype $i$ with $g_i \sim N(0, \sigma_g^2)$, $g_{ik}$ is the random effect of genotype $i$ nested within environment $k$ with $g_{ik} \sim N(0, \sigma_{ik}^2)$, and $e_{ijk}$ is the residual error with $e_{ijk} \sim N(0, \sigma_{eijk}^2)$. The phenotypic broad sense heritability of resistance to SPVD was calculated using the formula;

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{ge}^2/MN + \sigma_e^2/\text{JMN}}$$

Where $H^2$ is the broad sense heritability, $\sigma_g^2$ is the variance of the genotype, $\sigma_{ge}^2$ is the variance of the genotype nested within the environment, $\sigma_e^2$ is the error variance, $M$ is the number of sites with $m = 1,2,3$, $N$ is the number of seasons with $m = 1,2$, and $J$ is the number of blocks with $j = 1,2$. For QTL mapping, we used a random effects model implemented in the R package QTLpoly [available at https://github.com/guilherme-pereira/qtlpoly]. The software adopted a multiple interval mapping model described by Kao and Zeng (1997) to a random-effect model framework referred to as random-effect multiple interval mapping [REMIM
(Pereira, et al., in prep)] for QTL detection. The variance components associated with the QTL were estimated using the restricted maximum likelihood method. To identify QTL, the QTLpoly scripts performs the following functions; 1) a multiple QTL search using score-based statistics to assess significance and model optimization using forward and backward searches; 2) a QTL search that estimates QTL variance components using REML; 3) estimates of QTL allelic effects that includes the prediction of individual breeding values of a given QTL from the average effects of alleles; and 4) plots of QTL profiles for a given trait under evaluation. The whole genome is evaluated in a 1cM step size and the software fits QTL in positions that explain a significant part of the phenotypic variance (Pereira et al., in prep). QTL detection was followed by characterization. The joint conditional probabilities of markers were used for detecting and characterizing multiple QTL.

QTL peaks were identified from the most significant p-value profiles and plotted along the linkage groups. Previously detected peaks were used to build the model using the R package ‘sommer’ (Covarrubias-Pazaran, 2016). A final analysis included the computation of QTL heritability ($h^2_{QTL}$) as a proportion of phenotypic variance explained by their associated genetic effects. This was calculated as the ratio of the genetic variances over the total phenotypic variance. The REMIM functions and routines that plot p-value profiles used for the identification and characterization of QTL peaks while building full and multiple QTL models, are implemented in the R package “QTLpoly” (Pereira, et al., in press).
Results

Phenotypic data

Resistance reactions to SPVD in the NKB mapping population was highly significant ($p < 0.001$) when analyzed by ANOVA (Yada, et al., 2017). The frequency distribution of means varied widely over all three locations. Plots of the marginal means for the trait in NaSARRI and NgeZARDI showed that most of the progenies were scored as resistant in these sites, while they were all susceptible in NaCRRI (Figure 1). Transgressive segregation was observed [i.e., a few progenies exhibiting lower levels of SPVD resistance than the susceptible parent, and higher resistance levels than the resistant parent]. Most of the genotypes succumbed to SPVD pressure in NaCRRI, which is a hotspot for the disease. A scatter plot and histogram of residuals generated by the SAS MIXED procedure (SAS Institute, Cary, NC, USA) (data not shown) indicated that the SPVD severity scores were heteroscedastic as residual plots showed that susceptible genotypes showed much more variation across replications, whereas resistant genotypes were more consistently resistant. Broad-sense heritability across all environments was estimated at 48.7%.

Linkage map and SNP markers

We identified 13,096 high quality SNP markers after filtering with SuperMASSA and imputation using the *I. trifida* and *I. trilobal* reference genomes. Since the intention of using these markers was for linkage mapping and QTL analysis, we selected markers that were segregating in our population and were free from segregation distortion, missing data and redundancy using this characteristic as an added filtering criterion in MAPpoly software. After this additional filtering, 9,185 markers (70%) were kept for linkage mapping whereas 3,911 (30%) were dropped. These markers had unique dosage scores across the progeny and were
non-redundant. Sixty five percent of these were single dose markers (SDMs) while 35% were multi-dose markers (MDMs). There were more simplex markers identified in Beauregard (5,643 markers) compared to New Kawogo (4,506) (Table 2). The nullipex by simplex (0-1) parental order of SNP genotype was the highest, followed by simplex by nulliplex (1-0), simplex by simplex (1-1), nulliplex by duplex (0-2), duplex by simplex (2-1), simplex by duplex (1-2), duplex by duplex (2-2), triplex by duplex (3-2), duplex by triplex (2-3), nulliplex by triplex (0-3), and so on (summarized in Figure 2). The 9,185 markers were used to develop a 5,055 cM-long genetic map containing 1.8 SNPs per cM (summarized in Table 1). The average length of the linkage groups was 337 cM with the longest being linkage group 4 at 446.1 cM, and the shortest was linkage group 8 at 212.4 cM (Figure 3). All of these ordered markers were phased to the expected number of homologues based on parental dosages, resulting in a fully phased integrated linkage map (Supplementary Figure S1).

QTL analysis

By using score-based statistics (Qu, et al., 2013) and a random effects multiple interval mapping model implemented in the R package QTLpoly, we detected 2 QTL for resistance to SPVD on linkage groups 1 and 2 (Figure 4) explaining 22.7% and 13.5% of the variation in SPVD scores respectively (Table 3). The QTL on linkage group 1 was the strongest and the most significant ($p < 5.22e^{-05}$). When QTL analysis was performed on single sites, there were no QTL captured in NaCRRI and NaSARRI, whereas 1 QTL was observed in NgeZARDI on LG 1, explaining 12.8% of the variation in SPVD resistance within the environment (Figure 5). Table 4 shows summary statistics of the QTL in NgeZARDI.
We set our QTL significance threshold at -log(0.001) = 3 based on a prior simulation study that our group performed (Pereira, et al., in preparation) and observed that this threshold has high detection power and low false-positive discovery rate.

Allele effects from the decomposed best linear unbiased predictions (BLUPs) for QTL 1 showed that marker-assisted selection (MAS) for increasing SPVD resistance would be effective when the focus is to select alleles d, f, g and l from the respective parents. These particular alleles are responsible for decreasing the population mean of the trait as shown in Figure 6. This approach would also select against alleles a, b, i and j in the respective parents because these alleles are responsible for increasing the population mean of SPVD resistance (Figure 6). We noted that both Beauregard and New Kawogo had a combination of both resistance and susceptibility alleles (Figure 5).
Discussion

The NKB mapping population was developed in 2010 and has been studied extensively and is well suited for investigating the architecture of resistance to SPVD for various reasons (Yada, et al., 2015; Yada, et al., 2017). In all screening studies of several sweetpotato genotypes, ‘New Kawogo’ has consistently shown resistance to SPVD (Aritua, et al., 1998; Mwanga, et al., 2002). ‘Beauregard’, on the other hand, is highly susceptible to SPVD (Yada et al., 2017). The F1 mapping population is large enough (287 individuals) to increase statistical power of analysis and also allow for higher numbers of recombinant genotypes that allow better dissection of the trait and QTL analysis.

The nature and applicability of GBS to discover high-throughput SNP markers de novo, in a large number of segregating individuals at a relatively low cost, make it a powerful genotyping platform for various plant species (Chen, et al., 2013). The challenge in generating usable SNP marker data in polyploid crops has prevented breeders, specifically in the sweetpotato community, from incorporating molecular marker data in their breeding and selection schemes. In developing a genetic linkage map for autotetraploid blueberry, McCallum et al (2016) used GBS to genotype their population and identify SNP markers. However, the GBS data they obtained was not sufficient enough in terms of read depth and sequence coverage to call allele dosage and they were restricted to determining presence or absence of alleles only (McCallum, et al., 2016). They treated SNPs like dominant markers thus significantly reducing their versatility and information content. Various researchers have observed that GBS has unique technical challenges like nonuniform distribution of sequence reads coupled with high amounts of missing data (Beissinger, et al., 2013, Tardivel, et al., 2013, Wadl, et al., 2018). We increased the number of usable markers by increasing the mean
sequence coverage of individuals in the mapping population. We noted that a reduction in total sequencing depth was a false economy because it increased our difficulty of downstream analysis. This observation agrees with other researchers seeking to use next-generation sequencing data for similar purposes (Davey, et al., 2011; McCallum, et al., 2016). Non-unique markers were dropped because they map to the same position in the genetic map (Van Geest, et al., 2017) thus providing the same information. Segregation distortion occurs when markers do not fit with Mendelian segregation. This can result from several biological phenomena including gametophytic selection, double reduction, genetic drift, and unknown cytological attributes. Hence, distorted markers were dropped because they can alter estimation of recombination and cause spurious linkage (Carlier, et al., 2004), that can also impact map order or length (Hackett & Broadfoot, 2003).

Our reduced SNP map was still very dense, but the fewer SNPs selected to construct the NKB map made it easier to handle and less computationally intensive for subsequent analyses. We noted that although the Maximum Likelihood criterion orders loci better, it also results in increased map inflation when genotyping errors are present. This was also observed and reported by Shields et al (1991) and by Hackett & Broadfoot (2003). Map inflation results because for each locus, the likelihood approach calculates the probability of each genotype, given the genotypes at the two adjacent loci and on the distances between them (Hackett & Broadfoot, 2003). Whenever a genotyping error is encountered, an inflated distance results between those two loci thus inflating the map. The advantage of the weighted least squares approach is that the distances between markers are calculated from the map distances between all pairs of markers on the chromosome, and therefore the impact of genotyping errors on the distance between adjacent markers is minimized (Hackett & Broadfoot, 2003). However, like
Hackett and Broadfoot (2003) in their simulation study, we observed that the least squares approach gets more orders wrong and therefore we used the map built with the maximum likelihood criterion for our subsequent analyses. Using MAPploy software (Mollinari et al, in preparation), we were able to order all the 9,185 markers in one run, resulting directly into a fully phased integrated genetic map. Map length was considered trivial in our work, due to the inherent similarity of results arising from application of the map in subsequent genetic studies. Our QTL analysis results mapped on the same locations on the linkage groups irrespective of the map length. We observed that genotyping errors may have an effect on map length, but does not have a noticeable effect when the map is used for subsequent QTL analysis.

The broad sense heritability of resistance to SPVD was moderate (H = 48.7%) and this was close to the 51% observed by Yada et al, (2017). Using QTLpoly, we identified two QTL associated with resistance to SPVD that explain a total of 36.2% of the variation of SPVD scores in the mapping population. In hexaploid sweetpotato, this type of analysis presents significant advantages over methods developed for diploids. Polyploid QTL are regulated by multiple alleles and these were detected and their effects modeled using QTLpoly. We set a high threshold for detecting a QTL so as to greatly minimize chances of having significant false positives. The BLUPs for the allele effects showed that additive effects are the most important contributor to SPVD resistance and that ‘New Kawogo’ is the greatest contributor towards resistance, which would be expected. This is probably because the resistance alleles of New Kawogo are able to counteract and overcome effects of the susceptibility alleles in the population.

QTL that have been effectively utilized in breeding programs for cultivar development include QTL for soybean resistance to cyst nematode [Heterodera glycines] (Concibido , et
al., 2004), the \textit{Fhb1} QTL for Fusarium head blight resistance in wheat \textit{Triticum aestivum} (Anderson, et al., 2008), and the \textit{Sub1} QTL for submergence tolerance in rice \textit{Oryza sativa} (Septiningsih, et al., 2009). In each of these cases, the favorable QTL allele had an effect that was large enough and consistent enough to be easily tracked and fixed by standard breeding procedures (Bernardo, 2016). It is difficult to fix a trait in sweetpotato because of the outbreeding nature of sweetpotato, ploidy, the need to maintain high levels of heterozygosity and a high degree of deleterious mutations which are difficult to flush out. Maintaining heterozygosity is paramount to keep these mutations at bay, the ultimate result of this being that a typical backcrossing breeding scheme to incorporate a gene of interest is difficult to achieve in sweetpotato. Mwanga \textit{et al.} (2002) identified two QTL, \textit{spfm1} and \textit{spcsv1}, for SPFMV and SPCSV from a bi-parental cross between \textquote{Tanzania” and “Bikilamaliya”} using AFLP and RAPD markers. However, resistance to SPFMV has been reported to break down when co-infection with SPCSV occurs on the same genotype (Adikini, et al., 2015). Therefore, it is desirable to integrate multiple virus resistance into improved sweetpotato for long term resistance (Mwanga, et al., 2002).

The QTL we have identified here do not account for all sources of variation for SPVD as this was not a traditional heritability study that would account for all components of the genetic variation that would include additive, dominance and epistatic variances. However, our integrated NKB genetic map was developed using single- and multi-dose SNP markers and it represents a significant improvement in comparison to previous SPVD QTL research and we believe it is more reliable in performing QTL analysis than the previous studies (Kriegner, et al., 2003, Cervantes-Flores, et al., 2008, Zhao, et al., 2013). More resistance alleles were captured during our QTL analysis of the NgeZARDI site compared to the other sites. This is
probably because unlike NaCRRI, NgeZARDI is not a known hotspot for SPVD (Yada, et al., 2017) and therefore moderately resistant alleles could not be overcome by the low virus pressure. In this study, we found resistance alleles in the susceptible parent Beauregard. We hypothesize that this means that Beauregard has genes that possibly suppress the expression of its resistant alleles.

Resistance alleles of the resistant parent New Kawogo could overcome the suppressive alleles in the background of Beauregard. Further research will be needed to identify genes that potentially lead to Beauregard’s susceptibility to SPVD. The markers we have identified in this study account for enough of the variation to reduce SPVD severity. Further validation in a different population will be needed to see if these markers are useful enough for future marker assisted breeding in sweetpotato breeding programs especially in SSA where SPVD is a serious problem.

Based on our analyses of the joint adjusted means of SPVD scores and QTL allelic effects in the NKB population using QTLpoly, the top ten genotypes that have favorable alleles whose interactions have a net effect of decreasing the mean of SPVD score and thus contributing to resistance are; NKB 152, NKB 41, NKB 269, NKB 264, NKB 252, NKB 268, NKB 220, NKB 50, NKB 96, and NKB 66. When these lines are used as parents for SPVD population improvement, favorable alleles would recombine to improve the genetic gain for SPVD resistance in SSA.

In summary, the extensive genetic information provided by the integrated NKB genetic linkage map and the SNP genotypes improved our ability to detect QTL for SPVD and to model their allelic effects which is useful for selection. These new tools have enabled us to precisely track a favorable allele and identify full sib individuals containing them irrespective
of whether they have been phenotyped or not, provided they are genotyped. The use of ‘QTLpoly’ software allowed us to utilize SNP dosage information for polymorphism at the nucleotide level throughout all clones genotyped. These robust genomic tools have led to higher quality integrated linkage maps and more robust QTL analysis by analyzing both parents and the mapping population simultaneously. The search of possible candidate genes for mapped QTL will be needed to expound on the observations made in this research as more interdisciplinary studies build a network of information and open new insights and opportunities.
REFERENCES


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


Gar, O., Sargent, D. J., Tsai, C. J., Pleban, T., Shalev, G., Byrne, D. H., Zamir, D., 2011. An autotetraploid linkage map of rose (Rose hybrida) validated using strawberry (Fragaria vesca) genome sequence. PloS ONE, 6(5) e19379.


Rak, K., Bethke, P. C., Palta, J. P., 2017. QTL mapping of potato chip color and tuber traits within an autotetraploid family. Mol Breeding 37(15).


Table 1. Summary statistics of integrated linkage map for ‘New Kawogo’ x ‘Beauregard’ (NKB) population created using MAPpoly.

<table>
<thead>
<tr>
<th>Linkage Group (LG)</th>
<th>cM</th>
<th>No. of SNPS</th>
<th>SNPs/cM</th>
</tr>
</thead>
<tbody>
<tr>
<td>LG1</td>
<td>424.6</td>
<td>759</td>
<td>1.8</td>
</tr>
<tr>
<td>LG2</td>
<td>349.6</td>
<td>576</td>
<td>1.6</td>
</tr>
<tr>
<td>LG3</td>
<td>387.3</td>
<td>858</td>
<td>2.2</td>
</tr>
<tr>
<td>LG4</td>
<td>446.1</td>
<td>851</td>
<td>1.9</td>
</tr>
<tr>
<td>LG5</td>
<td>316.6</td>
<td>565</td>
<td>1.8</td>
</tr>
<tr>
<td>LG6</td>
<td>403.1</td>
<td>648</td>
<td>1.6</td>
</tr>
<tr>
<td>LG7</td>
<td>273.0</td>
<td>568</td>
<td>2.0</td>
</tr>
<tr>
<td>LG8</td>
<td>212.4</td>
<td>348</td>
<td>1.6</td>
</tr>
<tr>
<td>LG9</td>
<td>357.3</td>
<td>783</td>
<td>2.2</td>
</tr>
<tr>
<td>LG10</td>
<td>313.3</td>
<td>607</td>
<td>1.9</td>
</tr>
<tr>
<td>LG11</td>
<td>248.8</td>
<td>515</td>
<td>2.3</td>
</tr>
<tr>
<td>LG12</td>
<td>374.7</td>
<td>796</td>
<td>2.1</td>
</tr>
<tr>
<td>LG13</td>
<td>266.1</td>
<td>540</td>
<td>2.0</td>
</tr>
<tr>
<td>LG14</td>
<td>278.0</td>
<td>575</td>
<td>2.1</td>
</tr>
<tr>
<td>LG15</td>
<td>405.0</td>
<td>806</td>
<td>2.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>5,055.9</strong></td>
<td><strong>9,185</strong></td>
<td><strong>1.8</strong></td>
</tr>
</tbody>
</table>
Table 2. Number of markers per dosage in both parents. NK. dosage is the marker dosage in New Kawogo. B. dosage is the marker dosage in Beauregard. The dosage groups 0, 1, 2, 3 and 4 correspond to nulliplex, simplex, duplex, triplex and quadruplex marker groups respectively.

<table>
<thead>
<tr>
<th>NK. dosage</th>
<th>B. dosage</th>
<th>No. SNPs</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>3164</td>
<td>31.68</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>407</td>
<td>6.23</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>66</td>
<td>0.97</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>7</td>
<td>0.03</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>4295</td>
<td>21.61</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1502</td>
<td>12.03</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>536</td>
<td>4.65</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>179</td>
<td>1.76</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>47</td>
<td>0.31</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>910</td>
<td>2.96</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>550</td>
<td>4.82</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>384</td>
<td>3.58</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>263</td>
<td>2.28</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>63</td>
<td>0.85</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>126</td>
<td>0.28</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>187</td>
<td>1.58</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>230</td>
<td>2.44</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>80</td>
<td>0.84</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>3</td>
<td>0.05</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>41</td>
<td>0.44</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>56</td>
<td>0.56</td>
</tr>
</tbody>
</table>

Total 13,096
Table 3. Summary statistics of QTL for SPVD resistance on the NKB map across sites and environments. $\sigma^2_{QTL}$ is the variance associated with the QTL. $H^2_{QTL}$ is the heritability of the QTL.

<table>
<thead>
<tr>
<th>QTL</th>
<th>LG.</th>
<th>QTL Position</th>
<th>QTL Range (cM)</th>
<th>p-value</th>
<th>$\sigma^2_{QTL}$</th>
<th>$H^2_{QTL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>18.1</td>
<td>10.2</td>
<td>5.22e-05</td>
<td>0.0120</td>
<td>0.227</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>302.9</td>
<td>13.2</td>
<td>5.28e-04</td>
<td>0.0072</td>
<td>0.135</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.362</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Summary statistics of QTL for SPVD resistance in NgeZARDI. $\sigma^2_{QTL}$ is the variance associated with the QTL. $H^2_{QTL}$ is the heritability of the QTL.

<table>
<thead>
<tr>
<th>QTL</th>
<th>LG.</th>
<th>QTL Position</th>
<th>QTL Range</th>
<th>p-value</th>
<th>$\sigma^2_{QTL}$</th>
<th>$H^2_{QTL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>5.39</td>
<td>68</td>
<td>6.64e-04</td>
<td>0.017</td>
<td>0.128</td>
</tr>
</tbody>
</table>
Figure 1. Distribution of the adjusted means in all locations combined (joint adjusted means) and in each location (marginal means); NaCRRI, NaSARRI and NgeZARDI. Numbers on the x-axis show field SPVD scores on a 1 – 9 scale where 1 is resistant and 9 is highly susceptible.
Figure 2. Boxplot showing the distribution of marker dosages in the ‘New Kawogo’ X ‘Beauregard’ mapping population. New Kawogo 0-1 Beauregard means nulliplex by simplex for New Kawogo X Beauregard, 0-2 is nulliplex by duplex, 2-1 is duplex by simplex, and so on.
Figure 3. Multipoint “De novo” genetic linkage map of NKB population. The integrated map represents all the 15 chromosomes of hexaploid sweetpotato. Numbers 1 to 15 are the linkage groups, while the numbers below them are number of SNP markers for each linkage group.
Figure 4. Plot of QTL analysis of the entire genome for SPVD resistance in the NKB mapping population obtained with the joint model that averages out the separate effects of QTLs in all environments. QTL are identified on linkage group 1 and 2. LOP is the log of p-values.

Figure 5. Logarithm of p-value profiles (LOP) for SPVD resistance in NaCRRRI in the NKB population. The triangle on linkage group 1 represent the location of mapped QTL.
Figure 6. Breakdown of genotypic values showing additive effects from each allele for QTL 1 in the population. Alleles a, b, c, d, e, f are from New Kawogo whereas alleles g, h, i, j, k, l are from Beauregard. Numbers on the y-axis show the genetic value (deviation from population mean) of the allele. The zero (0) point is the population mean (sum of allele frequency * value) of the trait from a normal distribution.
Supplementary Figure S1. Integrated SNP-based genetic linkage map of New Kawogo x Beauregard. Each of the six horizontal bars on each parent represents a homologous chromosome. The red blocks represent markers identified in that region.
Linkage Group 3

New Kawogo

Beauregard

Linkage Group 4

New Kawogo

Beauregard
Linkage Group 15

New Kawogo

Beauregard
CHAPTER 3

Identification of QTL for sweetpotato weevil resistance in a biparental mapping population

(In a format suitable for submission to Frontiers in Plant Science, Genetics and Genomics of Polyploid Plants)
Identification of QTL for Sweetpotato Weevil Resistance in a Biparental Mapping Population

Bonny Michael Oloka¹², Benard Yada², Milton O. Anyanga², Guilherme da Silva Pereira¹³, Marcelo Mollinari¹³, Bode A. Olukolu¹, Doreen Chelangat², Paul Musana², Agnes Alajo², Zhao-Bang Zeng¹³, G. Craig Yencho¹

¹ Department of Horticultural Science, North Carolina State University, 214 Kilgore Hall, Box 7609, Raleigh, NC 27695-7609, USA
² National Crops Resources Research Institute, Namulonge, P.O. Box 7084, Kampala, Uganda.
³ Bioinformatics Research Center, Department of Horticultural Science, North Carolina State University, NC 27695-7566

Received for publication _______________2019. North Carolina State University. This paper is a portion of a dissertation submitted by Bonny Michael Oloka. The research reported in this publication was supported by funding provided by the Bill and Melinda Gates Foundation (OPP1052983) through the Genomic Tools for Sweetpotato Improvement Project (GT4SP). We thank Joan Adero and Jennipher Asiimwe for their assistance in DNA extraction, and Hannah Huntley for GBS library prep and sequencing.

Corresponding author email: craig_yencho@ncsu.edu
Abstract

Sweetpotato weevils (Cylas spp.) are the most serious insect pest of cultivated sweetpotato (Ipomoea batatas) in eastern Africa. The complex genetics of sweetpotato as well as a lack of knowledge of the underlying genetics of resistance, which has been associated with the presence of hydroxycinnamic acid esters (HCAs), has impeded the development of cultivars resistant to sweetpotato weevils (SPW). The objective of this study was to determine the genetic basis of resistance to SPW in a biparental sweetpotato mapping population consisting of 287 individuals segregating for resistance to SPW. The population was derived from a cross between ‘New Kawogo’ (NK), a Ugandan landrace cultivar, and ‘Beauregard’ (B), a major cultivar in the United States. Phenotyping was conducted in three sites (NaCRRI, NaSARRI and NgeZARDI) in Uganda over two seasons. The broad sense heritability of SPW resistance, measured at harvest, was estimated at 0.487. REML variance components of predicted genotype means (BLUPs) generated both within sites and seasons, and across all environments were used for QTL analysis using a random effects QTL mapping model using the software ‘MAPpoly’ which utilizes SNP marker dosage information and the NKB genetic map for QTL detection. We identified 1 QTL for SPW resistance in NaCRRI on linkage group 14 and 1 QTL in NgeZARDI on linkage group 3 explaining a variability of 13.2% and 11.7% respectively. For HCA, 1 QTL was identified on linkage group 1, explaining a variability of 16%. Additive effects of the QTL alleles from both parents were estimated and important alleles that directly impact the mean of both traits were identified.
Introduction

Sweetpotato, *Ipomoea batatas* (L.) Lam., is one of the ten most important food crops in the world, with a total global estimated production of 130 million tons per year (FAOSTAT, 2015). Sub-Saharan Africa (SSA) and Asia, account for over 80% of global sweetpotato production (FAOSTAT, 2016) most of which is concentrated in China. In the developing world, sweetpotato is a staple food crop providing important calorie and energy requirements (Yada, 2014) particularly in SSA and parts of Asia (CIP, 2015). Mainly local landraces are grown in SSA, and these are cream- or white-fleshed sweetpotato genotypes that have high dry-matter content (28-32%) and low sugars (Mwanga, et al., 2007). Sweetpotato has the ability to grow under marginal conditions and the high nutritional value of the provitamin A beta-carotene rich orange-fleshed types make them appealing and desirable for resource-poor female farmers and their children who suffer from chronic vitamin A deficiencies (Karyeija, et al., 1998, Low, et al., 2007).

The sweetpotato weevil, *Cylas* spp, is the most serious insect pest of sweetpotato worldwide. *C. puncticollis* (Boheman) and *C. brunneus* (Fabricius) occur in Uganda and are uniquely African species (Downham, et al., 2001). SPW can cause crop losses of up to 98% in areas with low rainfall and long dry spells (Smit, 1997). The larvae feed on storage roots but are not always readily observed until they have caused significant damage. Adults are also difficult to detect given their nocturnal habitat. Damage by SPW is one of the most important production constraints of sweetpotato in SSA (Stevenson, et al., 2009).

The greatest damage is inflicted on storage roots as the larvae feed (Cockerham, et al., 1954), triggering the storage roots to produce sesquiterpenes in response to the damage as well as the secondary pathogen infection, which renders edible roots unfit for consumption (Uritani
SPWs and their larvae have a concealed feeding behavior as they tunnel through storage roots while feeding. This makes their management difficult (Nottingham & Kays, 2002; Odongo, et al., 2003). Studies have indicated that resistance to SPW is partially mediated by hydroxycinnamic acid (HCA) esters present in the root latex of resistant cultivars (Stevenson, et al., 2009; Anyanga, et al., 2017). The HCA esters associated with resistance have been identified as hexadecylcaffeic acid, hexadecylcoumaric acid, heptadecylcaffeic acid, octadecylcaffeic, octadecylcoumaric acid and 5-O-caffeoylquinic acids (Stevenson, et al., 2009; Anyanga, et al., 2013; Yada, et al., 2017c). Host plant resistance provides an effective component of any integrated pest management program (IPM) of sweetpotato (Anyanga, et al., 2017). However, the development of weevil resistant varieties has not been successful over the years due to the perceived lack of heritable resistance in existing sweet potato germplasm (Anyanga, et al., 2017).

With the aim of identifying the biochemical basis of resistance to SPW, Muyinza et al., (2012) and Anyanga et al (2013) screened a wide range of sweetpotato genotypes using the highly susceptible cultivated landrace “Tanzania” as a control. They observed considerable variation in the concentration of HCA esters, with “New Kawogo”, an African landrace cultivar from Uganda recording the highest concentration of these compounds and highest levels of resistance to SPW. HCA esters located in the root latex of resistant cultivars have been associated with resistance, though only moderately, with an effect against adult weevil oviposition and feeding (Anyanga, et al., 2013). Differences in the concentration of these compounds between varieties might explain differences in resistance in the African sweetpotato gene pool (Anyanga, et al., 2017) which is highly heterogeneous in their susceptibility to the pest (Muyinza, et al., 2012). Research conducted by Anyanga, et al.,
(2013) using liquid chromatography-mass spectrometry (LC-MS) of root surface and epidermal extracts (1 min in hexane and 24 h in methanol) demonstrated that there was significant variation in the concentration of hexadecyl, heptadecyl, octadecyl, and quinic acid esters of caffeic and coumaric acid, with higher concentrations of these compounds correlated with resistance. They synthesized these compounds to enable their positive identification. It was proposed that the selection of sweetpotato varieties with higher levels of these HCA’s, particularly in the storage root surface might be used to develop germplasm with resistance to SPW. However, this would entail screening of hundreds of breeding lines using freeze-dried samples of roots and analysis by LC-MS in order to achieve long term population improvement. Currently, efforts to develop such a chemotyping platform are still too expensive (Yada, et al., 2017).

In their study on the segregation of HCA esters in sweetpotato, Anyanga, et al (2017) observed significant differences in the total HCA esters among the genotypes of the segregating bi-parental population (287 individuals) of the “New Kawogo” (resistant) x “Beauregard” (susceptible) (NKB cross). They attributed this to genetic effects at multiple loci in sweetpotato. Earlier studies have pointed out that the female parent in this bi-parental cross (New Kawogo) has a high concentration of HCA esters on its root surface, and high field and laboratory resistance to SPW (Stevenson, et al., 2009; Muyinza, et al., 2012; Anyanga, et al., 2013). This population also had genotypes that were more resistant and clones with higher HCA concentrations than “New Kawogo”. These genotypes were selected as new candidates to be used in a recurrent selection breeding program seeking improvement of this trait in order to achieve long term genetic gain. All these studies have suggested that resistance to sweetpotato weevils is mediated by expression of the HCA allelochemicals in sweetpotato in
addition to an escape mechanism by deep rooting genotypes as has been hypothesized by others (Stevenson, et al., 2009; Muyinza, et al., 2012; Anyanga, et al., 2013).

Several recently published studies have contributed much to our understanding of the genetic and biochemical basis of resistance to SPW’s in New Kawogo, the landrace source of resistance to SPW. Yada, et al., (2015), examined the diversity in the NKB population using SSR markers, and they also conducted studies to determine the genetic basis of resistance and identify SSR markers linked to SPW resistance using logistic regression (Yada, et al., 2017c). Anyanga, et al., (2017), conducted additional SPW resistance studies with the NKB materials, that were focused on improving our understanding of biochemical basis of resistance present in New Kawogo. These key studies were able to report on the first use of co-dominantly inherited SSR markers to identify SPW resistance loci in sweetpotato (Yada, et al., 2017c), the quantification of plant chemicals (HCA esters) that confer resistance to SPW, and the evaluation of levels of insect damage of the NKB mapping populations in the field and laboratory (Anyanga, et al., 2017). Data collected from these studies has shown that these chemical traits are controlled quantitatively.

With the advent of high throughput next generation genotyping, it is now possible to identify and utilize single nucleotide polymorphism (SNP) markers for genetic studies in sweetpotato. One of the major advantages for the utilization of SNPs for genetic analysis in polyploids is their genome-wide abundance and allowance for the direct observation of all genotypes at a given locus (Garcia, et al., 2013). The development of modern genotyping protocols like genotyping-by-sequencing (GBS) (Elshire, et al., 2011) or restriction amplified polymorphic DNA sequencing (RAD seq) (Etter, et al., 2011) and their modification and optimization for highly heterozygous and hexaploid sweetpotato using GBSpoly (Olukolu et
al, in prep) together with two complete reference genomes of *Ipomoea trifida* (V.3.0) and *Ipomoea triloba* (V.3.0), two of the wild diploid progenitors of cultivated sweetpotato (Wu et al., 2018) has enabled the evaluation of SNP markers throughout the genome of hexaploid sweetpotato.

In this study, we used SNP markers, our previously constructed integrated genetic linkage map of the NKB mapping population (Oloka, et al., in prep) and we identified QTL for SPW and HCA esters in sweetpotato. The NKB population was developed and previously mapped with 133 SSR markers by Yada et al., (2015) and it consisted of a segregating population of 287 individuals derived from a cross between ‘New Kawogo’, an African landrace cultivar with demonstrated resistance to SPW, and ‘Beauregard’, a major US cultivar that is highly susceptible to SPW (Yada, et al., 2015). Here we discuss the potential application of these identified minor QTL in breeding sweetpotato for improved resistance to SPW.
Materials and Methods

Mapping population phenotyping

The NKB biparental mapping population was previously described by Yada et al., (2015) and was developed in Uganda at the National Crops Resources Research Institute (NaCRRI) (0°32′N, 32°35′E, 1,150 m.a.s.l). The population consisted of 287 individuals and was generated in 2010 by crossing the SPW resistant Ugandan landrace cultivar, New Kawogo (Mwanga, et al., 2001; Anyanga , et al., 2017), with the SPW susceptible USA cultivar, Beauregard (Rolston , et al., 1987; Anyanga , et al., 2017). The population development and phenotyping for SPW has been previously described (Yada, et al., 2017; Anyanga , et al., 2017). Phenotyping trials were conducted in three sites for two seasons in 2012 (Table 1). All trials were arranged in a randomized complete block design (RCBD) with 3 blocks per site. Each plot consisted of 5 plants with 30 cm spacing on ridges 1.0 m apart. Plots were inoculated artificially with laboratory-reared weevils as described by Muyinza et al, (2012). Harvesting was done 5 months after planting and the harvested storage roots were scored for SPW severity using a 1-9 scale where 1 = no weevil damage on any root and 9 = severe damage symptoms on all roots in the plot (Gruneberg, et al., 2010).

Profiling HCA esters in storage root samples

To profile hydroxycinnamic acid esters (HCAs), storage roots were sampled from each plot and processed at NaCRRI. For HCA analysis, only samples from NaCRRI and NgeZARDI were used. The procedure for profiling HCA esters in the population has been described in detail by Yada et al., (2017) and Anyanga et al., (2017). The concentration of HCA esters in the storage root extracts was analyzed by liquid chromatography-mass spectrometry (LC-MS) at the Natural Resources Institute (NRI), University of Greenwich, UK, as described by
Anyanga et al., (Anyanga , et al., 2013). Briefly, storage roots were harvested from trial sites in NaCRRI and NgeZARDI, washed and cleaned. The clean roots were cut transversely into 2-3 small root disks, weighed, packed into labelled bags and freeze-dried using a vacuum freeze dryer (True-Ten Industrial Co., Taichung City, Taiwan) for 72 hours prior to blending into a powder. The powder (50 mg) was methanol extracted and the extract (1 ml) was used for HCA profiling LC-MS.

**No-Choice bioassay to evaluate weevil feeding and oviposition**

No-choice SPW feeding and oviposition was evaluated in the laboratory as described by Anyanga et al, (2017). Observations were made on storage roots of individual NKB progenies inoculated with 10-week-old gravid female *C. puncticollis* adults for 24 hours to feed and lay eggs. The inoculated roots were placed in 2 L plastic jars with the tops covered with muslin cloth for aeration. After 24 hours, the weevils were removed and the number of feeding holes that were formed on the roots were counted and recorded for each clone. The eggs that were laid in the roots during the time of weevil exposure were left to incubate at room temperature (24°C) under a 12:12 light dark regime until emergence. The number of adults emerging was recorded in 5-week intervals from day 25 to day 50. This same method was followed for the *C. brunneus* assay.

**Genetic map and conditional genotype probabilities of SNP markers**

Because sweetpotato is an autopolyploid crop, the dosage of a SNP marker is important for building an integrated map. To construct the integrated map for the NKB population, marker dosage estimation was first performed using SuperMASSA software, a graphical Bayesian model for SNP genotyping in polyploids (Serang, et al., 2012). We then used MAPpoly software [available at https://github.com/mmollina/MAPpoly (Mollinari et al., in
to construct a linkage map of NKB, which consisted of a 5,055-cM integrated genetic map containing 9,185 filtered SNPs. MAPpoly was used because it is able to compute millions of pairwise recombination fractions including all kinds of simplex and multiplex SNP configurations. The resulting NKB map consisted of a high density, integrated SNP-based genetic linkage map containing all 15 linkage groups of hexaploid sweetpotato with a marker density of 1.8 SNPs/cM. Genotype probabilities were estimated given the dosage at each marker locus. This method has been described by Mollinari et al., (in prep). In short, the steps followed by MAPpoly included: 1) SNP filtering; 2) pairwise recombination fraction estimation; 3) linkage group formation; 4) de novo ordering using a multi-dimensional scaling (MDS) algorithm; and 5) multipoint map reconstruction including phasing. Having obtained the complete phase information for each parental homolog, the most likely genotype for each individual progeny was determined given the dosage at each marker. To infer the most likely combination of parental homologs in each individual offspring for each position along the chromosome, a Hidden Markov Model was used (Rabiner, 1989; Hackett, et al., 2013; Rak, et al., 2017). The genotype conditional probabilities were calculated every 1 cM for the entire genome (Mollinari et al., in prep). For each genotyped individual, we obtained conditional probabilities for all positions in the genome.

**Phenotypic data and QTL analysis**

Restricted maximum likelihoods (REML) variance components analysis of the phenotypic data was performed both within sites and seasons (within environments) and across all environments to generate genotype predicted means for QTL analysis using the following mixed model:

\[ y_{ijmn} = \mu + l_m + s_n + b_{j(mn)} + g_{imn} + e_{ijmn} \]
Where \( y \) is observed value of genotype \( i \), \( \mu \) is the population mean of the trait, \( l_m \) is the fixed effect of location \( m \), \( s_n \) is the fixed effect of season \( n \), \( b_{j(mn)} \) is the random effect of block \( j \) nested within location \( m \) and season \( n \) with \( b_{j(mn)} \sim N(0, \sigma_b^2) \) \( \{b_{j(mn)} \) is normally distributed with mean 0 and variance \( \sigma_b^2 \}, g_{imn} \) is the random effect of genotype \( i \) in block \( j \) within location \( m \) and season \( n \) with \( g_{imn} \sim N(0, \sigma_g^2) \), and \( e_{ijmn} \) is the residual error with \( e_{ijmn} \sim N(0, \sigma_e^2) \). Marginal means of each location were obtained and adjusted phenotypic means across all three environments were also obtained. These phenotypes were then merged with the genotype probabilities and searched for associations between phenotypic means and molecular markers, given the genetic map.

To generate variance components that were used for calculating the heritability of SPW, we used the following model:

\[
y_{ijk} = \mu + env_k + b_{j(k)} + g_i + g_{ik} + e_{ijk}
\]

Where \( env_k \) is the fixed effect of environment \( k \) (for a given site and season), \( b_{j(k)} \) is the random effect of block \( j \) nested within environment \( k \) with \( b_{j(k)} \sim N(0, \sigma_b^2) \), \( g_i \) is the random effect of genotype \( i \) with \( g_i \sim N(0, \sigma_g^2) \), \( g_{ik} \) is the random effect of genotype \( i \) nested within environment \( k \) with \( g_{ik} \sim N(0, \sigma_{ik}^2) \), and \( e_{ijk} \) is the residual error with \( e_{ijk} \sim N(0, \sigma_e^2) \). The phenotypic broad sense heritability of resistance to SPW was calculated using the formula:

\[
H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{JMN} + \frac{\sigma_e^2}{JMN}}
\]

Where \( H^2 \) is the broad sense heritability, \( \sigma_g^2 \) is the variance of the genotype, \( \sigma_{ge}^2 \) is the genotype-environment interaction variance, \( \sigma_e^2 \) is the error variance, \( M \) is the number of sites with \( m = 1, 2, 3 \), \( N \) is the number of seasons with \( m = 1, 2 \), and \( J \) is the number of blocks with \( j \).
For QTL mapping, we used a random effects model implemented in the R package QTLpoly [available at https://github.com/guilherme-pereira/qtlpoly]. The software adopted a multiple interval mapping model described by Kao and Zeng (1997) to a random-effect model framework referred to as random-effect multiple interval mapping [REMIM (Pereira, et al., in prep)] for QTL detection. The variance components associated with the QTL are estimated using restricted maximum likelihood method. This REMIM method which is employed by “QTLpoly” to detect and characterize QTL has been described in detail in Pereira, et al. (in prep).

QTL peaks were identified from the most significant \( \hat{p} \) -value profiles and plotted along the linkage groups. A final analysis included the computation of QTL heritabilities \( (H^2) \) as a proportion of phenotypic variance explained by their associated genetic effects. This was simply calculated as the ratio of the genetic variances over the total phenotypic variance. Finally, the allele contribution to the population mean of the trait was got by averaging the genotypic values containing them.
Results

Field resistance reaction to SPW

Differences in SPW resistance of individual genotypes in the NKB mapping population were highly significant ($p < 0.001$) when analyzed with ANOVA (Yada, et al., 2017). Plots of the joint adjusted and marginal means for the trait in individual locations (sites) showed that most of the progenies were resistant (Figure 1), averaging between scores of 2 and 4 on a scale of 1 to 9. The frequency distribution of predicted genotype means varied widely over all three locations. Broad-sense heritability across all environments was estimated at 48.7%.

Feeding and oviposition SPW assay

Sweetpotato weevil feeding and oviposition varied widely for all the genotypes in the NKB population (Figure 2). Averaged across all locations, the mean number of feeding holes per storage root in individual genotypes of the NKB population ranged between 20 and 30 holes. The histograms for BuZARDI, RwebitaZARDI, and AbiZARDI were right skewed towards ‘New Kawogo’, the resistant parent. Forty genotypes collected from RwebitaZARDI had mean number of feeding holes less than 10, which is a similar level of resistance to SPW as New Kawogo. Genotype means of the number of feeding holes in inoculated storage roots across all sites ranged from 0 to 90 holes per SPW inoculated storage root.

Linkage map and SNP dosage

The completed NKB linkage map (summarized in Table 1) consisted of 15 linkage groups containing all 90 homologous chromosomes mapped using 9,185 SNP markers, totaling 5,055 cM in length, with an average marker density of 1.8 SNPs/cM. This map included both single dose and multiple dose markers evenly distributed on all 15 chromosomes and has been described in detail in Oloka, et al. (in prep).
QTL analysis

Using score-based statistics (Qu, et al., 2013) and the REMIM model implemented in the R package QTLpoly, we detected one QTL for resistance to SPW in NaCRRI on linkage groups 14 (Figure 3) explaining 13.2% of the variation, and 1 QTL in NgeZARDI on linkage group 3 (Figure 4) explaining 11.7% of the variation. There were no QTL identified on the joint model for QTL analysis when all three sites were combined. Table 3 shows the summary statistics of these identified QTL.

QTL analysis for HCA ester concentration on storage roots revealed 1 QTL on linkage group 1 explaining 16% of the HCA ester concentration (Figure 5). Table 4 shows the summary statistics for QTL for HCA esters. We noted that these QTL mapped to different linkage groups. The individual allele contribution obtained from averaging the genotypic values in NaCRRI (Figure 6) and NgeZARDI (figure not shown) showed that New Kawogo provided more resistant alleles compared to the susceptible parent Beauregard (Figure 6). For HCA concentration in storage roots, both parents had a combination of resistant and susceptibility alleles (Figure 7).

QTL analysis for the no-choice bioassay data did not detect any QTL. This data had very high environmental variability coming from storage roots from different agroecological environments without replications or controls. This resulted in a significant amount of noise in the data, giving negative genetic variances on analysis and we therefore could not estimate the error.
Discussion

We report here the first QTL mapping study for resistance to SPW and HCA ester concentration in a biparental mapping population of cultivated sweetpotato. Recent studies have concentrated on understanding the mechanism of resistance and suggested that the presence of HCA esters on the storage roots of resistant cultivars may be a major contributing factor towards resistance (Stevenson, et al., 2009; Muyinza, et al., 2012; Anyanga, et al., 2013). This study dug deeper into unraveling the genetic basis of resistance to SPW utilizing all available genomics and bioinformatics technologies at the time of the study.

We were not able to identify any QTL for resistance to SPW in the joint multiple QTL mapping model that combined data from all three environments in a single run. This could be due to a number of factors, including but not limited to; insufficient population size, highly heterogeneous weevil population in the different environments, significant genotype by environment effects, among others. Simulations to determine the optimum population size and marker density for QTL analysis in polyploid populations indicated that population size is critical, and to have adequate power to detect QTL, panels of 400 or more genotypes are recommended (Rak, et al., 2017, Rosyara 2015).

The NKB mapping population was developed in 2010 and has been studied extensively and is suited for investigating the architecture of resistance to SPW for various reasons, even though it might not contain the optimal number of genotypes (Yada, et al., 2015; Yada, et al., 2017; Anyanga, et al., 2017). In all screening studies of several sweetpotato genotypes, ‘New Kawogo’ has consistently shown better resistance to SPW compared to a wide range of genotypes (Mwanga et al., 2001). ‘Beauregard’, on the other hand, is highly susceptible to SPW (Yada et al., 2017). The F1 mapping population is large (287 individuals) enough to
increase statistical power of analysis and also allow for higher expression of recombinant
genotypes that allow better dissection of the trait and QTL analysis for both SPW resistance
and HCA esters.

The information provided by the integrated NKB genetic linkage map and the SNP
genotypes was extensive enough to improve the power to detect QTL for SPW and to model
their effects. QTLpoly software allowed us to utilize SNP dosage information for detection of
polymorphism at the nucleotide level throughout all clones genotyped. These robust genomic
tools led to higher quality linkage maps and more robust QTL analysis by analyzing both
parents and the mapping population simultaneously. The REMIM method, which considers an
integrated genetic map, was proposed by Gazaffi et al., (2014), in which QTL mapping is
performed based on the advantages of composite interval mapping (CIM) approach (Zeng,
1994), where cofactors are included in the model to remove the effect of QTL located outside
the mapping region, thus significantly increasing statistical power. We used a mapping model
with three genetic effects, assuming that a QTL may also segregate in different patterns in
progeny as a function of its genetic effects and of the linkage phase between markers and QTL
alleles (Pereira, et al., in prep).

In order to have a relatively high QTL detection power (>80% on average) and at the
same time controlling the false discovery rate and keeping it below 20%, our p-value thresholds
were based on 1,000 simulated quantitative traits with three QTLs each (Pereira, et al., in prep).
From those simulations utilizing different threshold combinations, we noted that using 0.01
and $10^{-3}$ as the forward and backward thresholds respectively resulted in having lesser QTLs
detected but with very high confidence.
The broad sense heritability of resistance to SPW was moderate (H = 49.1%) and this is agreement to that observed by Yada et al. (2017). We identified four resistance QTL explaining a total of 37.7% of the variation for SPW and 15.9% of the variation for storage-root HCA ester concentration in the mapping population. The BLUPs for the allele effects showed that additive effects are the most important contributor to SPW resistance and HCA concentration and that ‘New Kawogo’ is the greatest contributor towards resistance.

In this study, we found resistance alleles in the susceptible parent Beauregard and susceptible alleles in the resistant parent New Kawogo. This is possible when the genetic background of Beauregard has genes that suppress the expression of its resistant alleles, and the genetic background of New Kawogo has genes that mask the expression of susceptible alleles. We are not sure what these genes are at this moment. The QTL we have identified here do not account for all sources of variation for SPW and HCA as this was not a traditional heritability study that would account for all components of the genetic variation that would include additive, dominance and epistatic variances. Further research will be needed to identify genetic factors that potentially account for New Kawogo’s resistance and Beauregard’s susceptibility to SPW. The markers we have identified in this study account for enough of the variation to reduce SPW severity. Further validation in a different population will be needed to see if these markers are useful enough for future marker assisted breeding in sweetpotato breeding programs especially in sab-Saharan Africa where SPW is a serious problem.
REFERENCES


Downham, M. C., Smit, N.E J.M; Laboke, P O; Hall, D R; Odongo, B 2001. Reduction of Pre-harvest Infestations of African Sweetpotato Weevils *Cylas brunneus* and *C.*


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


Rak, K., Bethke, P. C., Palta, J. P., 2017. QTL mapping of potato chip color and tuber traits within an autotetraploid family. Mol Breeding, 37(15).


## Tables

**Table 1. Experimental sites** [Source: (Anyanga, et al., 2017)]

<table>
<thead>
<tr>
<th>Location</th>
<th>GPS coordinates</th>
<th>Agro-ecological zone</th>
<th>Average daily temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCRRI, Namulonge</td>
<td>0°32’N, 32°35’E, 1,150 m.a.s.l*</td>
<td>Moist, tall grassland</td>
<td>27°C</td>
</tr>
<tr>
<td>NaSARRI, Serere</td>
<td>1°32’N, 3°27’E, 1,085 m.a.s.l</td>
<td>Dry, short grassland</td>
<td>31°C</td>
</tr>
<tr>
<td>NgeZARDI, Ngetta</td>
<td>2°202’N, 33°62’E, 1,080 m.a.s.l</td>
<td>Dry, short grassland</td>
<td>30°C</td>
</tr>
</tbody>
</table>

* m.a.s.l is meters above sea level

**Table 2. Summary statistics of integrated linkage map for ‘New Kawogo’ x ‘Beauregard’ (NKB) population created using MAPpoly** [Source: Oloka, et al. (in prep.).]

<table>
<thead>
<tr>
<th>Linkage Group (LG)</th>
<th>cM</th>
<th>No. of SNPS</th>
<th>SNPs/cM</th>
</tr>
</thead>
<tbody>
<tr>
<td>LG1</td>
<td>424.6</td>
<td>759</td>
<td>1.8</td>
</tr>
<tr>
<td>LG2</td>
<td>349.6</td>
<td>576</td>
<td>1.6</td>
</tr>
<tr>
<td>LG3</td>
<td>387.3</td>
<td>858</td>
<td>2.2</td>
</tr>
<tr>
<td>LG4</td>
<td>446.1</td>
<td>851</td>
<td>1.9</td>
</tr>
<tr>
<td>LG5</td>
<td>316.6</td>
<td>565</td>
<td>1.8</td>
</tr>
<tr>
<td>LG6</td>
<td>403.1</td>
<td>648</td>
<td>1.6</td>
</tr>
<tr>
<td>LG7</td>
<td>273.0</td>
<td>568</td>
<td>2.0</td>
</tr>
<tr>
<td>LG8</td>
<td>212.4</td>
<td>348</td>
<td>1.6</td>
</tr>
<tr>
<td>LG9</td>
<td>357.3</td>
<td>783</td>
<td>2.2</td>
</tr>
<tr>
<td>LG10</td>
<td>313.3</td>
<td>607</td>
<td>1.9</td>
</tr>
<tr>
<td>LG11</td>
<td>248.8</td>
<td>515</td>
<td>2.3</td>
</tr>
<tr>
<td>LG12</td>
<td>374.7</td>
<td>796</td>
<td>2.1</td>
</tr>
<tr>
<td>LG13</td>
<td>266.1</td>
<td>540</td>
<td>2.0</td>
</tr>
<tr>
<td>LG14</td>
<td>278.0</td>
<td>575</td>
<td>2.1</td>
</tr>
<tr>
<td>LG15</td>
<td>405.0</td>
<td>806</td>
<td>2.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>5,055.9</strong></td>
<td><strong>9,185</strong></td>
<td><strong>1.8</strong></td>
</tr>
</tbody>
</table>
Table 3. Summary statistics of QTL for SPW resistance in NaCRRI and NgeZARDI. $\sigma_{QTL}$ is the variance associated with the QTL. $H^2$ is the heritability / amount of variability explained by the QTL.

<table>
<thead>
<tr>
<th>QTL location</th>
<th>LG.</th>
<th>QTL Position</th>
<th>QTL Range (cM)</th>
<th>p-value</th>
<th>$\sigma_{QTL}$</th>
<th>$H^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCRRI</td>
<td>14</td>
<td>258.9</td>
<td>127.8</td>
<td>3.36e-03</td>
<td>0.0105</td>
<td>0.132</td>
</tr>
<tr>
<td>NgeZARDI</td>
<td>3</td>
<td>125.5</td>
<td>26.0</td>
<td>2.02e-03</td>
<td>0.0127</td>
<td>0.117</td>
</tr>
</tbody>
</table>

Table 4. Summary statistics of QTL for HCA esters in the NKB population. $\sigma_{QTL}$ is the variance associated with the QTL. $H^2$ is the heritability / amount of variability explained by the QTL.

<table>
<thead>
<tr>
<th>QTL</th>
<th>LG.</th>
<th>QTL Position</th>
<th>QTL Range (cM)</th>
<th>p-value</th>
<th>$\sigma_{QTL}$</th>
<th>$H^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>150.2</td>
<td>64.7</td>
<td>7.81e-04</td>
<td>1944</td>
<td>0.160</td>
</tr>
</tbody>
</table>
Figure 1. Distribution of the adjusted means for SPW in all locations combined (joint) and in each location; NaCRRI, NaSARRI and NgeZARDI (marginal). The vertical axes of the histograms represent the frequency of clones in a given class represented by vertical bars.
Figure 2. Distribution of mean number of feeding holes in storage roots of the progeny and parents of the ‘New Kawogo’ (NK) x ‘Beauregard’ (B) mapping population collected from BuZARDI, RwebitaZARDI, and AbiZARDI. Vertical bars represent the number of clones in each class of mean number of feeding holes averaged over seasons.
Figure 3. Log of p-value (LOP) profile showing mapped QTL for SPW in NaCRRI. The triangle on linkage group 14 shows the position on the identified QTL on the linkage group.
Figure 4. Log of p-value (LOP) profile showing mapped QTL for SPW in NgeZARDI. The triangle on linkage group 3 shows the position on the identified QTL on the linkage group.
Figure 5. Log of p-value (LOP) profile showing mapped QTL for HCA esters on the storage roots of the NKB population analyzed jointly from three sites: NaCRRI, NaSARRI and NgeZARDI. The triangle on linkage group 1 shows the position on the identified QTL on the linkage group.
Figure 6. Breakdown of genotypic values showing additive effects from each allele at the QTL in the population when planted in NaCRRI. Alleles a, b, c, d, e, f are from New Kawogo whereas alleles g, h, i, j, k, l are from Beauregard. Numbers on the y-axis show the genetic value (deviation from population mean) of the allele. Zero (0) is the population mean (sum of allele frequency * value) of the trait from a normal distribution. Letters on the x-axis represent allelic combinations.
Figure 7. Breakdown of genotypic values for HCA concentration showing additive effects from each allele of the parents. Alleles a, b, c, d, e, f are from New Kawogo whereas alleles g, h, i, j, k, l are from Beauregard.
CHAPTER 4

Genetic mapping of QTL for storage root dry matter and β-carotene content in sweetpotato

(In a format suitable for submission to Molecular Breeding)
Genetic mapping of QTL for storage root dry matter and β-carotene content in sweetpotato

Bonny Michael Oloka¹,², Benard Yada², Guilherme da Silva Pereira¹,³, Bode A. Olukolu¹, Milton O. Anyanga², Doreen Chelangat², Paul Musana², Agnes Alajo², Marcelo Mollinari¹,³, Zhao-Bang Zeng³, G. Craig Yencho¹

¹ National Agricultural Research Organization (NARO), National Crops Resources Research Institute (NaCRRRI), Namulonge, P.O. Box 7084, Kampala, Uganda
² Bioinformatics Research Center, Department of Statistics, North Carolina State University
³ Department of Horticultural Science, North Carolina State University, 214 Kilgore Hall, Box 7609, Raleigh, NC 27695-7609, USA

Received for publication ____________2019. North Carolina State University. This paper is a portion of a dissertation submitted by Bonny Michael Oloka. The research reported in this publication was supported by funding provided by the Bill and Melinda Gates Foundation (OPP1052983) through the Genomic Tools for Sweetpotato Improvement Project (GT4SP). We thank Joanne Adero and Jenniffer Asiimwe for their assistance in DNA extraction, and Hannah Huntley for GBS library prep and sequencing.

Corresponding author email: craig_yencho@ncsu.edu
Abstract

The development and adoption of orange-fleshed sweetpotato (OFSP) in sub-Saharan Africa is desired for the improvement of food security and nutritional requirements of the growing population. However, the genetic complexity of sweetpotato [Ipomoea batatas (L.) Lam] pose significant challenges in breeding, and marker assisted selection tools are needed for population improvement and product development. We identified 2 quantitative trait loci (QTL) for storage root dry-matter and 3 QTL for β-carotene content in a 287-individuals hexaploid sweetpotato mapping population (NKB) derived from a cross between ‘New Kawogo’ (NK), a white-fleshed, high dry-matter African landrace cultivar, and ‘Beauregard’ (B), an orange-fleshed, low dry-matter sweetpotato cultivar popular in the USA. We used a previously described integrated genetic linkage map of NKB, and a random-effect multiple interval mapping model (REMIM) implemented in the R package QTLpoly and SNP’s together with their genotypes to identify the QTLs. Additive effects of the QTL were characterized using best linear unbiased predictions (BLUPs) to identify alleles contributing toward the phenotypic mean for parent and progeny selection.
Introduction

Sweetpotato, *Ipomoea batatas* (L.) Lam., is the 7th most important crop in the world, with a total global estimated production of 130 million tons per year (FAOSTAT, 2015). Sub-Saharan Africa (SSA) and Asia, produce over 80% of global sweetpotato production (FAOSTAT, 2015) most of which is concentrated in China. In most of SSA, sweetpotato is a staple food crop providing important calorie and energy requirements because they are easy to propagate, maintain, and yield well under marginal conditions (Karyeija, et al., 1998; Cervantes-Flores, et al., 2011). Mainly local landraces are grown in SSA, and these are typically cream- or white-fleshed sweetpotatoes that have high dry-matter content (28-30%) and low sugars (Mwanga, et al., 2007) which is a sharp contrast with the predominant varieties in USA that are sweet, orange-fleshed types with low dry matter. The caloric and nutritional value of orange-fleshed types, which are particularly rich in β-carotene make them appealing and desirable for resource-poor farmers (Karyeija, et al., 1998). The OFSP types, in addition to being high in β-carotene, are also high in vitamin C, fiber, iron, potassium and protein (Woolfe, 1992; Low, et al., 2007) thus making them an ideal food source to combat malnutrition, hunger and vitamin A deficiency mainly among pregnant women and children under the age of 5. Thus, a significant amount of sweetpotato breeding work in SSA has focused on developing higher dry-matter, semi-sweet OFSP to address the vitamin A deficiency (VAD) needs of women and children.

There has been a marked difficulty in improving sweetpotato for storage root dry matter, starch and β-carotene content in SSA and Asia (Chang, et al., 2009). This is mainly because of the generally self- and cross- incompatible, highly heterozygous and hexaploid (2n = 6x = 90) nature of the crop, which imposes significant breeding challenges. Many of the
nutritional traits in sweetpotato are also quantitatively inherited or appear to be so because of the polyploid nature of the crop and have significant genotype-by-environment interaction effects (Adebola, et al., 2013; Yada, et al., 2011). Additionally, storage root dry matter and β-carotene content are negatively correlated traits, and this further complicates selection efforts (Gruneberg, et al., 2005). Cross incompatibilities are common with each successful cross typically resulting in production of one or two botanical seed, thus further constraining improvement efforts.

Improving these traits, just like all others in sweetpotato, has been by mass selection (Jones, 1986). In hexaploid sweetpotato, this relatively simple method has not been very effective (Yada, 2014). It has become necessary to develop and utilize new genomic tools and breeding approaches to facilitate the genetic improvement of these traits. To date, the most common molecular tools that have been utilized are random amplified polymorphic DNA (RAPD) markers, amplified fragment length polymorphism (AFLP) markers and simple sequence repeat (SSR) markers (Yada, et al., 2017). These markers can only be analyzed as dominant markers for polyploid species because different doses of heterozygotes cannot be differentiated. These markers have been used in genetic linkage mapping (Kriegner, et al., 2003; Cervantes-Flores, et al., 2008a), QTL analysis for root-knot nematode resistance, storage root dry matter content, starch content, and β-carotene content (Cervantes-Flores, et al., 2008b; Chang, et al., 2009; Zhao, et al., 2013a). Nonetheless, none of these QTL have been utilized in marker assisted-selection schemes for sweetpotato improvement due to the inherent limitations of the marker technology in polyploids (Yada et al., 2017). The sampling of the genome in single doses, which is the norm when using linkage maps constructed using single dose markers, requires that the QTL also segregate as single dose markers which is not a
practical assumption (Pereira et al., in prep). Previous techniques have relied on marker data that are not fully informative because they only utilize loci with low allelic doses. However, for polyploid sweetpotato, it is desirable to utilize higher dose markers because each marker locus can then be localized in a homology group, potentially providing much more information for QTL effect modeling (Garcia, et al., 2013). In a segregating population of hexaploid sweetpotato, different numbers of segregating alleles exist. In order to obtain reasonable coverage of its genome for genetic characterization and analysis, a relatively large number of markers and progeny are required to capture the full spectrum of recombination events in segregating progeny, together with an integrated linkage map (Oloka, et al., 2018).

In this research, we used single and multiple dose SNP markers from the ‘New Kawogo’ by ‘Beauregard’ (NKB) genetic linkage map (Oloka et al., in prep) to identify QTL for dry-matter and β-carotene content in sweetpotato. The NKB map is a SNP-based integrated genetic linkage map developed from a segregating mapping population consisting of 287 individuals derived from a cross between ‘New Kawogo’, an African landrace cultivar, and ‘Beauregard’, a major US cultivar (Yada, 2014). Here, we describe the use of a random-effect multiple interval mapping model (REMIM) implemented in the R package QTLpoly for the detection and characterization of QTL associated with dry-matter and β-carotene content and discuss their potential application in breeding sweetpotato.
Materials and Methods

Mapping population

The NKB biparental mapping population was developed at National Crops Resources Research Institute [NaCRRI (0°32’N, 32°35’E, 1,150 m.a.s.l)] by Yada et al., (2015). This population of 287 individuals was generated in 2010 by crossing of the high storage root dry-matter, high starch and low β-carotene Ugandan landrace cultivar, New Kawogo, and low storage root dry-matter content, low starch content and high β-carotene content USA cultivar, Beauregard (Rolston, et al., 1987). This population development has been described in detail by Yada et al., (2015). Phenotyping trials were conducted in three sights for two seasons in 2012 (Table 1). This population was used for QTL analysis in this study, and is currently being maintained vegetatively in the greenhouse at NaCRRI.

Determination of storage root dry-matter, starch and β-carotene content

The phenotyping and analysis of storage root dry-matter, starch and β-carotene content of the NKB population has been described in detail by Yada et al., (2017). Briefly, marketable storage roots from each plot were randomly selected from each field, brought to NaCRRI, washed and processed. For determination of dry-matter content, 100 g-fresh weight samples were dried in a hot air oven at 65°C for 72 hours and the final dry weight was calculated as a percentage of the fresh weight. For storage root starch and β-carotene content, another 100 g-fresh weight samples were dried in a vacuum freeze dryer (True-Ten Industrial Co., Taichung City, Taiwan), weighed and milled (mesh size 0.425 mm) using a Thomas Wiley Mini-Mill (Thomas Scientific, Swedesboro, New Jersey), then shipped to International Potato Center (CIP), Ghana, for root chemistry profiling using near-infrared reflectance spectroscopy (NIRS). The NIRS calibration equations for β-carotene and starch content were developed at...
CIP, Lima, Peru (Zum Fielde, et al., 2009) and the dried and weighed (2.0 g) samples were scanned using a FOSS NIRS Rapid Content Analyzer instrument (FOSS NIRSystem, Hoganas, Sweden). The measurement of diffuse reflectance from the samples in the near-infrared region of the electromagnetic spectrum using a monochromator (XDS 1000, Hoganas, Sweden) gave us spectral data. These results were recorded as log (1/R) at 2 nm increments and averaged. Processing of the results was done using ISIscan, version 4.2.0 (FOSS NIRSystems, Hoganas, Sweden) using calibration equations developed by Zum Felde et al., (2009). The values for the calibration coefficients of determination ($R^2_c$) were 0.98 and 0.97 for starch and β-carotene respectively. High coefficients of determination ($R^2_{cv}$) were also yielded for the cross-validation process with 0.8 for starch and 0.96 for β-carotene (Yada et al., 2017).

**Genetic map and conditional probabilities**

The development of the NKB genetic map has been described in Oloka et al., (in prep). Dosage call was performed using SuperMASSA software (Serang, et al., 2012) and the 5,055-cM genetic map containing 9,185 filtered SNPs was obtained using MAPpoly (Mollinari et al., in prep). The map is a high density, integrated SNP-based genetic linkage map containing all 15 linkage groups of hexaploid sweetpotato and marker density of 1.8 SNPs/cM. Genotype probabilities were estimated given the dosage at each marker. This method has been comprehensively described by Mollinari et al., (in prep). In short, assuming no double reduction, there are 400 possible genotype configurations coming from independent inheritance of two chromosomal segments from each parent. Having obtained the complete phase information for each parental homolog, the most likely genotype for each individual progeny was determined given the dosage at each marker. To infer the most likely combination of parental homologs in each individual offspring for each position along the chromosome, we
used a Hidden Markov Model (Rabiner, 1989; Hackett, et al., 2013; Rak, et al., 2017). The genotype conditional probabilities were calculated every 1 cM for the whole genome (Mollinari et al., in prep). For each genotyped individual, we obtained conditional probabilities for all positions in the genome.

**Phenotypic data and QTL analysis**

Restricted maximum likelihoods (REML) variance components analysis of the phenotypic data was performed both within sites and seasons and across all environments to generate genotype adjusted means for QTL analysis using the following mixed model:

\[
y_{ijmn} = \mu + l_m + s_n + b_{j(mn)} + g_{imn} + e_{ijmn}
\]

Where \(y_{ijmn}\) is the observed phenotypic value of genotype \(i\), \(\mu\) is the population mean of the trait, \(l_m\) is the fixed effect of location \(m\), \(s_n\) is the fixed effect of season \(n\), \(b_{j(mn)}\) is the random effect of block \(j\) nested within location \(m\) and season \(n\) with \(b_{j(mn)} \sim N(0, \sigma_b^2)\) [i.e. \(b_{j(mn)}\) follows a normal distribution with mean 0 and variance \(\sigma_b^2\)], \(g_{imn}\) is the random effect of genotype \(i\) in block \(j\) within location \(m\) and season \(n\) with \(g_{imn} \sim N(0, \sigma_g^2)\), and \(e_{ijmn}\) is the residual error with \(e_{ijmn} \sim N(0, \sigma_e^2)\). Marginal means of each location were obtained and adjusted phenotypic means (predicted means) across all three environments were also obtained. These phenotypes were then merged with the genotype probabilities into a single “object”. This “object” was searched for associations between phenotypic means and SNP markers, given the genetic map for QTL analysis. To generate variance components that were used for calculating the heritability of the traits, we used the following mixed model:

\[
y_{ijk} = \mu + env_k + b_{j(k)} + g_i + g_{ik} + e_{ijk}
\]
Where $env_k$ is the fixed effect of environment $k$ (for a given site and season), $b_{j(k)}$ is the random effect of block $j$ nested within environment $k$ with $b_{j(k)} \sim N(0, \sigma^2_b)$. $g_i$ is the random effect of genotype $i$ with $g_i \sim N(0, \sigma^2_g)$, $g_{ik}$ is the random effect of genotype $i$ nested within environment $k$ with $g_{ik} \sim N(0, \sigma^2_{ik})$, and $e_{ijk}$ is the residual error with $e_{ijk} \sim N(0, \sigma^2_{eijkl})$. The phenotypic broad sense heritability of the traits was calculated using the formula:

$$H^2 = \frac{\sigma^2_g}{\sigma^2_g + \frac{\sigma^2_{ge}}{MN} + \frac{\sigma^2_e}{JMN}}$$

Where $H^2$ is the broad sense heritability, $\sigma^2_g$ is the variance of the genotype, $\sigma^2_{ge}$ is the genotype-environment interaction variance, $\sigma^2_e$ is the error variance, $M$ is the number of sites with $m = 1, 2, 3$, $N$ is the number of seasons with $m = 1, 2$, and $J$ is the number of blocks with $j = 1, 2$. The QTL mapping procedure has been described in detail in our earlier papers and chapters (Oloka, et al., in preparation, Pereira, at al., in preparation).
**Results**

**Phenotypic data**

When analyzed with ANOVA, the means of storage root dry matter, starch and β-carotene content in the NKB mapping population were highly significant \((p < 0.05)\) (Yada, et al., 2017). Plots of the joint adjusted means for the traits showed that the frequency distribution of joint adjusted means varied widely (Figure 1). The distribution of means was continuous for storage root dry matter content, while it was skewed to the right for β-carotene content. Broad-sense heritability across all environments was estimated at 80% and 93% respectively for storage root dry matter content and β-carotene content. These plots revealed that storage root β-carotene content was right-skewed with most of the genotypes in the population having low β-carotene levels like the female parent, ‘New Kawogo’. Histogram for storage root dry-matter content approximated a normal distribution with its mean (29.1%) and mode (28.8%) very close as represented by the boxplot on top of the figure (Figure 1).

**Linkage map and SNP dosage**

The completed NKB linkage map consisted of 15 linkage groups containing all 90 homologous chromosomes mapped using 9,185 SNP markers, totaling 5,055 cM in length, with an average marker density of 1.8 SNP/cM (Oloka et al., in prep). This map included both single dose and multiple dose markers evenly distributed on all 15 chromosomes.

**QTL analysis**

By using score-based statistics (Qu, et al., 2013) and a multiple QTL mapping model implemented in the R package QTLpoly, we detected 3 QTL for storage root β-carotene content on linkage groups 2, 3 and 12 (Figure 2) explaining 14.2%, 11.5% and 10.8% of the variation, respectively and accounting for a total variation of 36.7% (Table 2), whereas 2 QTL
for storage root dry-matter content were identified on linkage groups 4 and 10 (Figure 3) explaining 14.9% and 14.6% of the variation, respectively and accounting for a total variation of 29.6%. A QTL plot of β-carotene, starch and dry matter content (Figure 5) showed that the Minor QTL for STCH (QTL that was not strong enough to meet the set significance threshold) and the QTL for BC mapped very close to each other on linkage groups 3 and 12.

The individual allelic contributions obtained from averaging the genotypic values across sites showed that both parents had a combination of alleles that either increased or decreased the phenotypic mean of the trait in the population (Figure 6).
**Discussion**

Sweetpotato is a one of the leading food crops in SSA with significant potential for food security and poverty alleviation. However, efforts to develop nutrient-rich, value-added products with increased levels of pro-vitamin A β-carotene in SSA have been met with significant adoption challenges. This is mainly because the preferred white- or cream-fleshed types have desirable attributes like high dry matter content (30% – 34%) that is mostly associated with starch content, that are also tolerant to key local biotic and abiotic stresses. In this research, we report on the identification of QTL for storage root β-carotene content, starch and dry matter content to aid sweetpotato breeding programs in improving these traits simultaneously in their breeding populations.

The distribution of adjusted means for dry-matter content was continuous, while that for starch and dry matter content were slightly skewed. Transgressive segregations was observed for all the traits, which was in agreement with the observation of Cervantes-Flores et al., (2011) in the TB mapping population. Estimates for the broad sense heritability of these traits (0.93 for BC, 0.91 for STCH, 0.81 for DM) are high, meaning that they can be effectively improved through conventional breeding approaches. These estimates are close to what was observed by Yada et al., (2017) in the same population, but using different analysis approaches. The BLUPs for the allele effects showed that additive effects are the most important contributor to the traits and that ‘New Kawogo’ was the greatest contributor for STCH and DM, while Beauregard was the greatest contributor for BC.

Cervantes-Flores et al (2011) was the first to map QTL for dry matter, starch and β-carotene content in sweetpotato in a 240-individual bi-parental mapping population developed from a cross between ‘Tanzania’ and ‘Beauregard’ (Cervantes-Flores, et al., 2011). They used
AFLP markers on two parental maps developed from their population and reported the identification of 13 QTL (5 in Tanzania and 8 in Beauregard) for storage root dry matter content explaining a total variation of 15-24%, 12 QTL (5 from Tanzania and 7 in Beauregard) for starch content explaining a total variation 17-30%, and 8 QTL (4 from Tanzania and 4 from Beauregard) for β-carotene content explaining a total variation of 17-35%. Using the SSR and AFLP linkage map developed by Zhao et al., (2013), Xiao-xia et al., (2014) identified 28 QTL for dry matter content and 8 QTL for starch content (Xiao-xia, et al., 2014) in a population of 202 individuals derived from a bi-parental cross between two important Chinese sweetpotato cultivars Xushu 18 and Xu 781.

It is important to note that in all these previous studies, a pseudo test-cross QTL mapping strategy was used, and QTL were placed on separate parental maps that could not be integrated. Therefore, it is difficult to directly compare our results for mapped QTL with the previous due to the differences in technology and methodology. The integrated NKB genetic map developed using single- and multi-dose SNP markers presents significant improvements compared to the previous work, and its reliability in performing QTL studies have been discussed in detail in Oloka et al., (Chapter 2). Furthermore, the QTL described here may be more reliable because we have taken into account the BLUPs of all the environments and also their phenotypic joint adjusted means to map QTL.

In this study, we found a combination of additive alleles from both parents that either increase the population mean of the trait or decrease it. For New Kawogo, we observed that additive alleles for storage root dry matter content are able to increase the population mean irrespective of the negative alleles. The same applies to Beauregard with alleles for β-carotene content.
The extensive genetic information provided by the integrated NKB genetic linkage map and the SNP genotypes improved our power to detect QTL for storage root β-carotene, starch and dry-matter content and to model their effects in cultivated sweetpotato. The use of MAPpoly software allowed us to utilize SNP dosage information for polymorphism at the nucleotide level throughout all clones genotyped. The QTLpoly linkage mapping approach has enabled more effective utilization of the high-quality integrated linkage map for QTL analysis by analyzing the allelic contributions of both parents of the NKB mapping population simultaneously. The search of possible candidate genes for mapped QTL will be our next step to expound on the observations made here as more interdisciplinary studies build a network of information and open new challenges, insights and opportunities.
REFERENCES


Zhao, N. et al., 2013. A genetic linkage map based on AFLP and SSR markers and mapping of QTL for dry-matter content in sweetpotato. Mol Breed, Volume 32: 807-820.

### Tables

Table 1. Experimental sites [Source: (Anyanga, et al., 2017)]

<table>
<thead>
<tr>
<th>Location</th>
<th>GPS coordinates</th>
<th>Agro-ecological zone</th>
<th>Average daily temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCRRI, Namulonge</td>
<td>0°32’N, 32°35’E, 1,150 m.a.s.l*</td>
<td>Moist, tall grassland</td>
<td>27°C</td>
</tr>
<tr>
<td>NaSARRI, Serere</td>
<td>1°32’N, 3°27’E, 1,085 m.a.s.l</td>
<td>Dry, short grassland</td>
<td>31°C</td>
</tr>
<tr>
<td>NgeZARDI, Ngetta</td>
<td>2°202’N, 33°62’E, 1,080 m.a.s.l</td>
<td>Dry, short grassland</td>
<td>30°C</td>
</tr>
</tbody>
</table>

*m.a.s.l is meters above sea level
Table 2. Summary statistics of QTL for storage root β-carotene content on the NKB map across sites and environments. $\sigma^2_{QTL}$ is the variance associated with the QTL. $h^2_{QTL}$ is the heritability of the QTL. The QTL range is the 95% support interval of the QTL.

<table>
<thead>
<tr>
<th>QTL</th>
<th>LG.</th>
<th>QTL Position</th>
<th>QTL Range (cM)</th>
<th>p-value</th>
<th>$\sigma^2_{QTL}$</th>
<th>$h^2_{QTL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>43.6</td>
<td>282.2</td>
<td>7.39e-04</td>
<td>2.8</td>
<td>0.142</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>125.5</td>
<td>90.6</td>
<td>2.08e-04</td>
<td>2.3</td>
<td>0.116</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>308.5</td>
<td>94.4</td>
<td>6.72e-04</td>
<td>2.2</td>
<td>0.108</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.367</td>
</tr>
</tbody>
</table>

Table 3. Summary statistics of QTL for storage root dry matter content on the NKB map across sites and environments. $\sigma^2_{QTL}$ is the variance associated with the QTL. $h^2_{QTL}$ is the heritability of the QTL. The QTL range is the 95% support interval of the QTL.

<table>
<thead>
<tr>
<th>QTL</th>
<th>LG.</th>
<th>QTL Position</th>
<th>QTL Range (cM)</th>
<th>p-value</th>
<th>$\sigma^2_{QTL}$</th>
<th>$h^2_{QTL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>401.3</td>
<td>39.7</td>
<td>4.14e-04</td>
<td>0.207</td>
<td>0.149</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>72.4</td>
<td>36.5</td>
<td>7.74e-04</td>
<td>0.201</td>
<td>0.146</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.296</td>
</tr>
</tbody>
</table>
Figure 1. Distribution of the joint adjusted means across all environments for parents and progeny of the NKB population. Numbers on the x-axis are the predicted means. Vertical bars represent the number (frequency) of genotypes within a given bin.
Figure 2. Plot of QTL analysis for storage root β-carotene content (BC) in the NKB mapping population. QTL were located on linkage groups 2, 3 and 12.
Figure 3. Plot of QTL analysis for storage root dry matter content (DM) in the NKB mapping population. QTL were identified on linkage groups 4 and 10.
Figure 4. Plot of mapped QTL for storage root β-carotene content (BC), starch content (STCH) and dry matter content (DM) in the NKB mapping population. Minor QTL for STCH and the QTL for BC mapped very close to each other on linkage groups 3 and 12.
Figure 5. Breakdown of genotypic values of β-carotene showing additive effects from each allele of the parents. Alleles a, b, c, d, e, f are from New Kawogo whereas alleles g, h, i, j, k, l are from Beauregard. Letters on the x-axis show alleles and their combinations. Numbers on the y-axis show the genetic value (deviation from population mean) of the allele. Zero (0) is the population mean (sum of allele frequency * value) of the trait from a normal distribution.
CHAPTER 5

QTL analysis for root-knot nematode resistance in cultivated sweetpotato

(In a format suitable for submission to HortScience)
QTL analysis for root-knot nematode resistance in cultivated sweetpotato

Bonny Michael Oloka, Victor A. Amankwaah, Kenneth V. Pecota, Bode A. Olukolu

Department of Horticultural Science, North Carolina State University, 214 Kilgore Hall, Box 7609, Raleigh, NC 27695-7609, USA

Guilherme Da Silva Pereira, Marcelo Mollinari, Zhao-Bang Zeng

Bioinformatics Research Center, Department of Horticultural Science, North Carolina State University, NC 27695-7566

G. Craig Yencho

Department of Horticultural Science, North Carolina State University, 214 Kilgore Hall, Box 7609, Raleigh, NC 27695-7609, USA

Received for publication _______________2019. North Carolina State University. This paper is a portion of a dissertation submitted by Bonny Michael Oloka. The research reported in this publication was supported by funding provided by the Bill and Melinda Gates Foundation (OP1052983) through the Genomic Tools for Sweetpotato Improvement Project (GT4SP). We thank Anna Pecota, Meri Reeber, Amber Dawn Harmon and Dominic Gillooly for their assistance in the greenhouse.

Corresponding author email: craig_yencho@ncsu.edu
Abstract

The root knot nematode [Meloidogyne incognita (Kofoid & White) Chitwood] (RKN) causes significant storage root quality reduction and yield losses in cultivated sweetpotato [Ipomoea batatas (L.) Lam.]. In this study, resistance to RKN was examined in a mapping population consisting of 240 progenies derived from a cross (TB) between ‘Tanzania’, a predominant African landrace cultivar, and ‘Beauregard’, a major cultivar in the United States. We performed quantitative trait (QTL) analysis using a random-effect QTL mapping model to locate QTL on the TB genetic map associated with RKN resistance. The integrated TB linkage map contained 14,813 well distributed SNPs and totaled 2,120 cM. An RKN bioassay incorporating potted cuttings of each genotype was conducted in the greenhouse and replicated five times over a period of 10 weeks. For each experiment, each genotype was inoculated with ca. 20,000 RKN eggs. Plant resistance reactions were assessed by counting the number of RKN egg masses (galls) that formed on the entire root system of a given clone. Resistance to RKN in the progeny was highly skewed towards the resistant parent, exhibiting medium to high levels of resistance. We identified two QTL: one on linkage group 3, and one on linkage group 7. Genetic effects were modeled for each putative QTL and models of effects for RKN egg masses were fit for all reps. In the joint model, the QTL were shown to have significant effects for RKN and explained 66% of the variation for RKN egg masses. This work represents a significant step forward in our understanding of the genetic architecture of RKN, and sets the stage for future utilization of genomics assisted breeding in applied sweetpotato breeding programs.
Introduction

Plant parasitic nematodes are major pathogens of many cultivated crops (Agrios, 1997). The root knot nematode (RKN) species, *Meloidogyne incognita* is the most important of the plant parasitic nematodes and is responsible for billions of dollars in crop losses annually (Sasser & Freckman, 1987; Cervantes-Flores, et al., 2008b). Root-knot nematodes (*Meloidogyne* spp) are distributed worldwide and the damage they inflict on field crops is estimated at about 10% worldwide (Whitehead, 1998). In eastern Africa, RKNs are reported to affect sweetpotato (Namaganda, et al., 1993; Makumbi-Kidza, et al., 2000; Karuri, et al., 2017), cassava, and banana (Coyne, et al., 2003; Coyne, et al., 2006). However, much of the damage by RKNs goes undetected due to their associations with fungi and bacteria in disease complexes (Cervantes-Flores, et al., 2008b). The cracking and secondary infections resulting from this reduce the market value of the sweetpotato storage roots by directly affecting their quality.

RKNs are obligatory sedentary endoparasites with a life cycle of 1 to 2 months. Embryos develop in eggs produced by adult females, and hatch as second-stage larvae (L2) that move through the soil and invade the plant root (Sasser, 1980). Upon infection, the worm establishes a feeding site within the root cortex and undergoes three additional molts to become an adult. *M. incognita* is a mitotic parthenogenetic species that produces a continuous infection chain once established. The feeding sites appear as rounded galls (knots) on the root vascular tissue of infected plants thus disrupting their capacity to uptake nutrients from the soil. On fleshy storage roots, the infection can appear as cracks although this is also associated with various other environmental factors like soil texture and moisture (Lawrence, et al., 1986). In their research on RKNs, Lawrence et al. (1986) suggested that the pathogens may be
predisposing the roots to cracking, rather than a causal factor of cracking after observing that the number of cracked roots was not correlated with initial RKN population number. This was after they observed that when rainfall was more uniform, the storage roots did not crack although nematodes were present in them (Lawrence, et al., 1986). There is also a genetic component to this as some genotypes crack more often than others when grown in warm, wet sandy soils.

Managing root-knot nematodes has often involved the use of neurotoxic nematicides in combination with cultural practices. However, notwithstanding the obvious health and environmental risk that nematicides pose, their cost is prohibitive to small scale farmers and growers of sweetpotato due to its relatively low market value (Gasapin, 1984). The safest, most sustainable and economic route is the use of host plant resistance. However, ‘Beauregard’ – a popular sweetpotato cultivar grown in USA and other parts of the world, is highly susceptible to *M. incognita* and *M. javanica* (Cervantes-Flores, et al., 2002).

The mode of resistance of sweetpotato to RKN is not well understood. Ukoskit *et al.* (1997) hypothesized single gene qualitative resistance, whereas multiple gene quantitative resistance has been hypothesized by several other researchers (Cordner, et al., 1954; Giamalva, et al., 1961; Jones & Dukes, 1980; Cervantes-Flores, et al., 2008b). Jones and Dukes (1980) suggested that independent sources of resistance to different strains of RKN were responsible for the observed differences in inheritance and that the genes originated from multiple origins (Mcharo, et al., 2005b). Histological studies have shown that RKNs in the juvenile (J2) stage penetrate both susceptible and resistant sweetpotato clones as well as other *Ipomoea* species (Komiyama, et al., 2006). However, Komiyama *et al* (2006) observed that localized necrotic reactions prevent further pathogen development in resistant clones whereas in susceptible
genotypes, the pathogen is able to establish itself, in agreement with observations by earlier researchers (Dropkin, 1969; Paulson & Webster, 1972). Nematodes that fail to establish feeding have been observed to either die or leave the roots (Koyimana et al., 2006). It has also been noted that resistance to *M. incognita* occurs via a hypersensitive reaction in sweetpotato (Dean & Struble, 1953; Gentile, et al., 1962; Martin & Birchfield, 1973; Jones & Dukes, 1980) as well as in other crops (Okamoto & Mitsui, 1974).

Molecular markers have been widely used in many different crops to identify and map genes associated with resistance to nematodes (Barr, et al., 1998; Wang, et al., 2001; Ynturi, et al., 2006) but very few of such studies have been successfully conducted in sweetpotato (Ukoskit, et al., 1997; Mcharo, et al., 2005b; Cervantes-Flores, et al., 2008b; Nakayama, et al., 2012). Ukoskit et al. (1997) used RAPD markers in a population of 71 individuals derived from a cross between ‘Regal’ (resistant) and ‘Vardaman’ (susceptible) and the marker OP151500 was weakly associated (P=0.037) with RKN resistance in the cross. Mcharo et al. (2005b) employed two unrelated sweetpotato populations and applied logistic regression and discriminant analysis to study RKN resistance. They report the ability to predict and classify the phenotype with an accuracy of 88.75% and 88.04%. However, since their population size was small, they could have over-estimated the effect of the markers (Cervantes-Flores, et al., 2008b). Cervantes-Flores et al. (2008b) hypothesized that resistance to RKN is conferred by several genes, based on molecular and phenotypic data. They detected nine QTL associated with RKN resistance, each of which showed a relatively small genetic effect. They used single dose AFLP markers on two separate maps, one for Tanzania and the other for Beauregard. They also identified 3 unmapped duplex markers that explained most of the variation of resistance (~45%), but they were not able to place those particular markers into the framework.
map, and thus, they could not locate the potential QTL. They concluded that the higher dose markers are more informative when placed on the genetic map, and that they were potentially associated with 1 or 2 major genes (Cervantes-Flores, et al., 2008b). More recently, Nakayama et al. (2012) conducted a multi-race analysis of resistance to multiple races of southern root-knot nematode (SRKN) and suggested that race-specific resistance is more likely conferred by single genes and that the genes for resistance against each race are closely located (Nakayama, et al., 2012). However, just like Ukoskit et al. (1997), their population was small and they therefore could have overestimated the effects of markers they used in the study and also could have missed capturing more alleles arising from different recombination events. The importance of population size to study the inheritance of a trait and to detect QTL has been strongly emphasized by several researchers (Collard, et al., 2005; Doerge, et al., 1997; Cervantes-Flores, et al., 2008a). In a polyploid crop such as sweetpotato, a large population size increases the statistical power which is critical to capture the full spectrum of underlying allelic variation arising from very many possible recombination events (Collard, et al., 2005; Kriegner, et al., 2003; Cervantes-Flores, et al., 2008b). These previous techniques have relied on marker data that are not fully informative because they utilized loci with low allelic doses. For a complex polyploid like sweetpotato, it is important to utilize higher dose markers because each marker locus needs to be positioned in a homology group (Garcia, et al., 2013). Sweetpotato is also an outcrossing, highly heterozygous species; meaning that in a segregating population, different numbers of segregating alleles exist. In order to guarantee reasonable coverage of its genome for genetic characterization and analysis, a relatively large number of markers are required to capture the full spectrum of recombination events (Oloka, et al., 2018).
In this work, we describe the use of single- and multiple-dose SNP markers on the ‘Tanzania’ x ‘Beauregard’ (TB) genetic linkage map to identify QTL for the RKN, *Meloidogyne incognita*, race 3. The TB map is a SNP-based integrated genetic linkage map developed from a segregating mapping population consisting of 240 individuals derived from a cross between ‘Tanzania’, an African landrace cultivar, and ‘Beauregard’, a major US cultivar (Cervantes-Flores, 2006). Tanzania is highly resistant to 4 major RKN races, while Beauregard is highly susceptible to RKN infection (Cervantes-Flores, et al., 2002). Here, we describe the localization of five QTL associated with resistance, and their associated genetic effects. We further annotate the sequences surrounding the associated markers in order to search for putative candidate genes that may be involved in RKN resistance in sweetpotato.
**Materials and Methods**

**Germplasm**

The TB mapping population has been described in detail by Cervantes-Flores et al., (2008). It consists of 240 individuals of a cross between the RKN resistant African cultivar, ‘Tanzania’, and RKN susceptible USA cultivar, ‘Beauregard’. Since its development, the mapping population and parents have been maintained in the greenhouse in a vegetative state in virus-free conditions, with periodic propagation renewals with each clone planted, in a 20.3-cm diameter pot containing Fafard P4 soil mix (Fafard, Agawam, MA). For each individual clone, five three-node cuttings were taken and planted into 72-cell Landmark™ seedling trays (Stuewe & Sons, Corvallis, OR) containing Fafard P4 for propagation in the greenhouse. Along with the two parents, we included four checks of known RKN performance; ‘Covington’, ‘Hernandez’, ‘Jewel’ and ‘Porto Rico’. Plants were grown under greenhouse conditions at 25 to 28 °C and watered as needed. They were also fertilized to supplement their nutrient needs throughout the growing period.

**RKN screening**

A single ~15 cm long cutting of each genotype was transplanted into 4” Azalea pots (round) containing 50:50 pasteurized mix (by volume) of course sand and field soil (loamy sand: 88.9%, 8.3% silt, and 2.8% clay), respectively (Cervantes-Flores et al., 2008a, 2002). The cuttings were allowed to root before inoculation (14 days). The experiment was executed using a completely randomized design with five replications separated by time (i.e. one replication consisted of all 240 genotypes planted in a rep with reps repeated over time). All reps were planted and harvested in Fall 2016 on separate dates, with rep 1, rep 2, rep 3, rep 4 and rep 5 harvested 62, 55, 62, 69 and 62 days, respectively after planting.
Root knot nematodes eggs, *M. incognita* (race 3, the most predominant in North Carolina), that were previously cultured on ‘Rutgers’ tomato plants (*Solanum lycopersicum* L.) were extracted using NaOCl. This method has been described by Hussey and Barker (1973), and was used in this same population by Cervantes-Flores et al., (2006). A 15-ml inoculum solution containing ~ 20,000 RKN eggs was applied into soil mixture to infest each individual plant. Plants were grown under controlled greenhouse conditions of 25 to 28 °C and were watered and fertilized as needed.

At the end of each trial (rep), the plants were harvested and rated by counting the number of root-knot galls (egg masses) present on each root system for every genotype (Cervantes-Flores, et al., 2002). This was done beginning with the first rep, then after one week the second, then the third, and so on. The root tissue was stained with red food coloring (McCormick and Co., Baltimore, MD) for visual rating. Our laboratory routinely uses red food coloring in place of Phloxine B, a highly neurotoxic chemical, to stain the root-knot galls, and allows their lucid visualization on the roots. This method is non-toxic, and its results are comparable to those described by Hussey and Barker (1973). Analysis of variance (ANOVA) conducted on the number of egg masses per root system using SAS (version 9.3; SAS Institute, Cary, NC) showed that there were significant interactions between genotypes x reps in the population (result not shown). Analysis was performed on every rep and joint genotype means over the five replications were also analyzed.

We observed segregation for vine weight and storage root initiation amongst various other traits. At the end of the greenhouse trial and the RKN assay, individual clones were placed in labeled paper bags and dried in a hot air oven at 65-70°C for 72 hours. Measurements were taken for the weights (in grams) of both the dry vine and the roots (both storage and
adventitious roots combined) over four replications. A histogram of the average weight of the dried vines and roots were plotted (Figure 3).

**Genetic map and conditional probabilities**

SNP dosage calling was performed using SuperMASSA software (Serang, et al., 2012). The R-based software program, ‘MAPpoly’ ([https://github.com/mmolina/mappoly](https://github.com/mmolina/mappoly)) was used to produce 14,813 filtered SNPs and to construct an integrated 2,120-cM TB linkage map. The map is a high density, integrated SNP-based genetic linkage map containing all 15 linkage groups of hexaploid sweetpotato and marker density of 6.9 SNPs/cM. Genotype probabilities were estimated given the TB genetic map. This method has been comprehensively described by Mollinari et al., (in prep). In short, assuming no double reduction, there are 400 possible genotype configurations coming from independent inheritance of chromosomal segments from each parent. Having obtained the complete phase information for each parental homolog, the most likely genotype for each individual progeny was determined given the map. To infer the most likely combination of parental homologs in each individual offspring for each position along the chromosome, we used a Hidden Markov Model (Rabiner, 1989; Hackett, et al., 2013; Rak, et al., 2017). The genotype conditional probabilities were calculated every 1 cM for the whole genome (Mollinari et al., in prep). For each genotyped individual, we obtained conditional probabilities for all positions in the genome.
Phenotypic data and QTL analysis

Restricted maximum likelihoods (REML) variance components analysis of the phenotypic data was performed to generate marginal means of reps and joint adjusted means across all reps used for QTL analysis using the following mixed model:

\[ y_{ij} = \mu + r_j + g_i + e_{ij} \]

where \( y_{ij} \) is the observed phenotypic value of genotype \( i \) in rep \( j \), \( \mu \) is the population mean of the trait, \( g_i \) is the random effect of genotype \( i \) with \( g_i \sim N(0, \sigma_g^2) \) [\( g_i \) is normally distributed with mean 0 and variance \( \sigma_g^2 \)], \( r_j \) is the random effect of replication \( j \) with \( r_j \sim N(0, \sigma_r^2) \), and \( e_{ij} \) is the residual error with \( e_{ij} \sim N(0, \sigma_e^2) \). Marginal means of each rep were obtained and joint adjusted means across all five reps were also obtained. The phenotypic broad sense heritability of resistance to RKN in the greenhouse trial was estimated using the formula;

\[ H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2/R} \]

Where \( H^2 \) is the broad sense heritability, \( \sigma_g^2 \) is the variance of the genotype, \( \sigma_e^2 \) is the error variance, and \( R = 5 \) is the number of replications. For QTL mapping, we used a random effects model implemented in the R package QTLpoly [available at https://github.com/guilherme-pereira/qtlpoly]. For QTL detection, a multiple interval mapping model described by Kao and Zeng (1997) was adapted to a random-effect model framework referred to as random-effect multiple interval mapping [REMIM (Pereira, et al., in prep)]. The variance components associated with the QTL were estimated using the restricted maximum likelihood method. To identify QTL, the QTLpoly scripts performs the following functions: 1) a multiple QTL search using score-based statistics to assess significance and model optimization using forward and backward searches; 2) QTL search that estimates QTL
variance components using REML; 3) estimates of QTL allelic effects that include the prediction of individual breeding values of a given allele from the average effects of alleles; and 4) plots of QTL profiles for a given trait under evaluation. The whole genome is evaluated in a 1cM step size and the software fits QTL in positions that explain a significant part of the phenotypic variance (Pereira et al., in prep). QTL detection was followed by characterization. Up until this point, the joint conditional probabilities of markers were used for detecting and characterizing multiple QTL.

QTL peaks were identified from the most significant $\hat{p}$-value profiles and plotted along the linkage groups. Previously detected peaks were used to build these models using the R package ‘sommer’ (Covarrubias-Pazaran, 2016). A final analysis included the computation of QTL heritabilities ($H^2$) as a proportion of phenotypic variance explained by their associated genetic effects. This was simply calculated as the ratio of the genetic variances over the total phenotypic variance. Finally, the allele contribution to the population mean of the trait was got by averaging the genotypic values containing them. These REMIM functions and routines that involve plotting $\hat{p}$-value profiles, identifying and characterizing QTL peaks while building full and multiple QTL models, are implemented in the R package QTLpoly (Pereira, et al., in prep). Since QTL possibly span a large portion of the linkage group, QTLploy plots 95% confidence intervals of identified QTL. These support intervals show the limits on the linkage group within which there is high confidence of the presence of the QTL.

**Candidate gene search**

A blast search of the regions within support intervals of the mapped QTL on linkage group 7 was performed using sequence information from the pseudomolecules of the *Ipomoea trifida* (NSP306) reference genome (Wu et al., 2018). Linkage group 7 was chosen because it
explained the highest variation for RKN resistance, at 57%. The *I. trifida* assembly is a 462 Mb reference sequence for hexaploid sweetpotato, with 32,301 annotated high confidence gene models (http://sweetpotato.plantbiology.msu.edu/). This reference sequence genome was repeat masked then annotated for protein-coding genes using a set of transcript and protein evidence. These sequence regions were also blast searched on the NCBI database.
Results

Phenotypic data

The RKN resistance reactions of the mapping population were highly significant \((p < 0.001)\) when analyzed with ANOVA. Histograms of individual replications of the data showed that most of the progenies were resistant (Figure 1). Genetic correlation of the distribution of resistance reactions showed that the different replications separated by time were highly correlated \((p^{***} < 0.001)\) to each other and that there were no significant differences among them (Figure 1). Just like in individual reps, the frequency distribution of joint means over all five trials demonstrated significant skewing towards the resistant parent (Figure 2) and was similar to that observed by Cervantes-Flores et al., (2008b), which suggested that resistance is quantitative but probably conferred by a few major genes. Transgressive segregation was observed [i.e., a few progenies (3%) exhibiting lower levels of RKN resistance than the susceptible parent (more than 125 EM per root system)]. A scatter plot and histogram of residuals by the SAS MIXED procedure (SAS Institute, Cary, NC, USA) (data not shown) indicated that the RKN scores were heteroscedastic as residual plots showed that susceptible genotypes showed much more variation across replications, whereas resistant genotypes were more consistently resistant. Broad-sense heritability in the greenhouse trial was estimated at 96%.

A histogram of the distribution of the average dry vine weight and the average dry root weight in the parents and progeny of the TB population (Figure 3) showed that most genotypes segregated towards the female parent, Tanzania. Average dry vine weight was normally distributed whereas average dry root weight was right skewed. The weight of vines for
Tanzania was high, but it didn’t have any storage roots formed by the end of the greenhouse trial.

**Linkage map and SNP marker dosage**

The completed TB linkage map (Figure 5), summarized in Table 1, consisted of 15 linkage groups containing all 90 homologous chromosomes mapped using 14,813 SNP markers, totaling 2,120 cM in length. This integrated genetic map includes both single dose and multiple dose markers evenly distributed on all 15 chromosomes. Linkage group 1 was the longest at 193.9 cM while linkage group 5 was the shortest at 99.5 cM.

**QTL analysis**

By using score-based statistics (Qu, et al., 2013) and a random effects model implemented in the R package *QTLpoly*, we detected 2 QTL for resistance to RKN, *M. incognita* race 3 on linkage groups 3 and 7 (Figure 5) explaining 8.7% and 57.6% of the variation (Table 2). QTL 2 on linkage group 7 explained the most variation and was the most significant (*p* < 1.00e-16). When the QTL model was fit for raw data of individual reps, 10 QTL were observed in rep 5, 8 of which were unique and absent in other reps (Figure 6). Raw data contain significant environmental variation and this was expected. QTL 2 on linkage group 7 was present in each of the reps and very consistent. Support intervals for mapped QTL (Figure 8) show the limits on the linkage groups where there is a 95% confidence for the QTL being in that location. The individual allele contribution obtained from averaging the genotypic values containing them showed that both Tanzania and Beauregard have QTL alleles conveying resistance and susceptibility in the segregating progeny (Figure 9). QTL with a positive effect demonstrated how much susceptibility to RKN can be increased by the presence
of those alleles in the population; whereas, those with a negative effect show how much resistance to RKN can be increased.

**Candidate gene search**

Ideally, QTL span a wide range of the chromosome that could include a number of genes and several other molecular components that regulate them. Our search was performed on linkage group 7 within a few centimorgan (50 cM) of support intervals (confidence intervals) of the mapped QTL. The sequence regions on the *I. trifida* and *I. triloba* reference genomes (Wu, et al., 2019) that contain the markers associated with the identified QTL all lie within annotated genes. The sequence region containing marker S7_3375600 had 99.5% similarity to *major facilitator protein* (MFS) (e-value < 1e-127) characterized on the *I. trifida* reference. MFS is a superfamily of membrane transport proteins that facilitate movement of sugars and small molecules across cell membranes (Pao , et al., 1998). The sequence for marker S7_21032190 was associated with *protein kinase*, an enzyme that modifies other proteins by phosphorylating them to change their activity, cellular location, or association with other proteins. The nucleotide sequence for marker S7_3095392 was associated with *vacuolar protein sorting*, a family of proteins involved in a series of vital mechanisms including, but not limited to; autophagy, ion secretion by salt glands, abiotic/biotic stress responses, and so on (Surpin & Raikhel, 2004; Xiang , et al., 2013). When this sequence was searched on NCBI, the predicted protein with 100% identity and 100% positives and an e-value of 8e-31 was *phosphatidylinositol 3-kinase*, root isoform [*Ipomoea nil, nicotiana attenuate*, ...etc]. This enzyme is responsible for development and signaling; this includes its involvement in endocytosis, reactive oxygen species (ROS) production, and transcriptional activity (Park , et al., 2003; Joo , et al., 2005; Lee , et al., 2008). The nucleotide sequence for marker
S13_21933180 predicted tetraicopeptide repeat (TPR)-like superfamily protein (SRFR1), a protein involved in suppressing effector-triggered immunity (Kwon, et al., 2009).
Discussion

The TB mapping population has been studied extensively and is well suited for investigating the architecture of resistance to RKN for various reasons, as discussed by Cervantes-Flores et al., (2002, 2005, 2008b). In all screening studies of several sweetpotato genotypes, ‘Tanzania’ has consistently shown high levels of resistance to multiple species and races of nematodes (Cervantes-Flores, et al., 2002; Karuri , et al., 2017). ‘Beauregard’, on the other hand, is susceptible to the same RKN species (Cervantes-Flores, et al., 2002). The F1 TB mapping population is large enough at 240 individuals to increase statistical power of analysis and also allow for the detection of recombinant genotypes that allow better dissection of the trait and QTL analysis.

The frequency distribution observed in our study for RKN was consistent with that of other researchers (Cervantes-Flores, et al., 2008b; Nakayama , et al., 2012), and conforms to a 4:1 ratio of resistant: susceptible genotypes. This pattern can be explained by the action of one or two major genes/alleles that are most likely segregating independently (Cervantes-Flores, et al., 2008b), assuming hexasomic inheritance. However, since a continuum of resistance reactions is observed in most of the segregating populations studies to date, the resistance reaction to RKN is considered quantitative, but most likely conferred by a few major genes with differing levels of effect. Polygenic resistance to nematodes is very common in plants and has been reported in sweetpotato (Cervantes-Flores et al., 2008b) and potato [Solanum tuberosum] (Bryan, et al., 2002).

The frequency distribution observed for dry vine weight is characteristic of that observed for traits controlled quantitatively and by the action of several genes. This pattern was observed for storage root initiation as well, but with a slight skew to the right. The
distribution of trait means in the genotypes demonstrates the quantitative nature of their inheritance. Transgressive segregation was observed which means that these genotypes can be used as parents for population improvement for vine weight and early storage root initiation.

Molecular markers linked to QTL that have been effectively utilized in breeding programs for cultivar development include QTL for soybean resistance to cyst nematode \textit{[Heterodera glycines]} (Concibido, et al., 2004), the \textit{Fhb1} QTL for Fusarium head blight resistance in wheat \textit{[Triticum aestivum]} (Anderson, et al., 2008), and the \textit{Sub1} QTL for submergence tolerance in rice \textit{[Oryza sativa]} (Septiningsih, et al., 2009). In each of these cases, the favorable QTL allele had an effect that was large enough and consistent enough to be easily tracked and fixed by standard breeding procedures (Bernardo, 2016). This is difficult to fix in sweetpotato because of ploidy, high heterozygosity and a high degree of deleterious mutations which is difficult to flush out in asexually propagated species. Maintaining heterozygosity is therefore paramount to keep these mutations at bay, the downside of this being that a typical backcrossing breeding scheme to incorporate a single gene is difficult to achieve.

The broad sense heritability of resistance to RKN is high ($H^2 = 96\%$) and this is close to the 89\% observed by Cervantes-Flores et al. (2008b) considering that a greater portion of the progeny (4:1) would be considered resistant according to the number of root-knot galls present in the root system. Here, we identified two resistance QTL explaining a total of 66\% of the variation in the mapping population. The BLUPs for the allele effects showed that additive effects are the most important contributor to RKN resistance and that ‘Tanzania’ is the greatest contributor providing a combination of favorable alleles that increase resistance.

In their earlier work on RKN resistance in this TB population, Cervantes-Flores et al., (2008b) used AFLP markers and identified 9 minor effect QTL responsible for the resistance.
They reported that none of the mapped QTL explained more than 15% of variation, but interestingly observed 3 unmapped higher duplex markers that explained ~45% of the variation. With this information, they concluded that these higher dose markers are potentially associated with 1 or 2 major genes, and that this would agree with the observed segregation ratio of the phenotypes. Although they could not determine this hypothesis at that point in time, they recommended further analysis through the addition of more markers, sequencing, and other such approaches (Cervantes-Flores et al., 2008b). We were able to utilize high dose markers and identified 2 major QTL. QTL 2 on linkage group 7 explained 57% of the phenotypic variation, which is a great improvement compared to the work of Cervantes-Flores where none of their QTL explained more than 15% of the phenotypic variation. The total contribution of all 2 mapped QTL explained 66% of the variation and this agrees with the observation of Cervantes-Flores et al., that a few major genes are possibly involved in conditioning resistance.

Our integrated TB genetic map developed using single- and multi-dose SNP markers presents significant improvements in comparison to previous work described by Cervantes et al., (2008). Furthermore, the QTL described here are consistent and therefore considered reliable because we have taken into account the phenotypic joint adjusted means (predicted means containing BLUEs and BLUPs) of five replications to map QTL in a highly controlled greenhouse environment, and because we identified these QTL in the same chromosomal position using means of individual reps.

The mapped QTL locations on the linkage groups were considered as potential regions to search for putative candidate genes that might be involved in RKN resistance. QTLpoly gave us the confidence limits for the QTL within which a search for candidate genes could be
conducted. A strong marker-QTL association detected in full-sib progenies could be an important factor for crop improvement programs using clonal propagation because the possibility of crossover between the marker and QTL is low. Sweetpotato has high genetic complexity and its genome is not yet fully sequenced to date, therefore inferences about putative candidate genes could contribute to new insights and open new areas of research in mining and validation of QTL and genes of interest (Balsalobre, et al., 2017).

The QTL on linkage group 7 explained a variability of 57% for RKN gall counts. Upon inspection of the *I. trifida* reference genome we observed that the marker S7_3095392 on position 52.4 on linkage group 7 was centered on a sequence region that is similar with *vacuolar protein sorting*, a family of proteins involved in several vital cellular functions including; autophagy, ion secretion by salt glands and abiotic/biotic stress responses. In general terms, the protein sorting pathway begins in the ER, passes through the Golgi complex, and finally reaches either the vacuole, other cellular compartments or the cell surface (De Marcos Lousa, et al., 2012; Xiang, et al., 2013). Autophagy is a normal cellular process in an organism that destroys and recycles diseased cells as a consequence of effector triggered immunity (Goodman & Novacky, 1994). This makes sense because resistance mechanisms use similar defense responses and is expected. This sequence region was also similar to *phosphatidylinositol 3-kinase*, root isoform (*Ipomoea nil*), a protein that is particularly responsible for endocytosis and reactive oxygen species (ROS) production, vesicle trafficking, and is required for the activation of many signaling molecules (Tazearslan, et al., 2009). Resistance of field plants against RKN has previously been observed to be mediated through a hypersensitive response in cowpea (Das, et al., 2008), tomato (Meliillo, et al., 2006), and

Favorable alleles have an average effect that decrease the mean of the trait, thus increasing resistance were present in both parents. Tanzania, being the resistant parent, potentially has alleles that are able to mask its unfavorable alleles thereby resulting in resistance. When making selection in a breeding program, the use of an identified QTL requires the integration of breeding values estimated from its average effects of alleles. Selection is therefore made on favorable alleles only whose additive effects can pass on to the next generation for population improvement.

Based on our analyses of the genotypic value predictions in the TB population using QTLpoly, the top ten genotypes that have favorable alleles whose interactions have an average effect of decreasing the mean of RKN egg masses and thus increasing resistance are: TB 50, TB 163, TB 214, TB 213, TB 83, TB 235, TB 237, TB 5, TB 183, and TB 216. When these lines are used as parents for RKN population improvement, the average effects of their alleles would increase the genetic gain RKN resistance.

In conclusion, we have identified 2 QTL associated with RKN resistance in hexaploid sweetpotato that explain 66.3% of the variation in gall production. Our understanding of the genetic architecture of this important trait has significantly improved owning to the utilization of new molecular and bioinformatics tools. The search of possible candidate genes for mapped QTLs is a preliminary analysis highlighting the importance of new insights into the relationships between statistical genetics and biology. Further research is needed to expound on the observations made in this study, to identify and confirm the genes at these QTL regions through gene silencing, CRISPR technologies and/or transgenic analysis. The work presented
here is one aspect of a larger effort focused on the development and efficient use of new genomic, statistical and bioinformatics tools for sweetpotato improvement under the GT4SP project.
REFERENCES


Dean, J. L. & Struble, F. B., 1953. Resistance and susceptibility to root-knot nematodes in tomato and sweet potato. Phytopathology 53(290 (Abstr.)).


Garcia, A. A. F. et al., 2013. SNP Genotyping allows an in-depth characterization of the genome of sugarcane and other complex autopolyploids. SCIENTIFIC REPORTS 3(3399).


Table 1. Summary of individual chromosomes for the integrated TB linkage map created using MAPpoly.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>No. of markers</th>
<th>Length</th>
<th>SNPs/cM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHR 1</td>
<td>1265</td>
<td>193.9</td>
<td>6.5</td>
</tr>
<tr>
<td>CHR 2</td>
<td>926</td>
<td>161.4</td>
<td>5.7</td>
</tr>
<tr>
<td>CHR 3</td>
<td>1137</td>
<td>151.7</td>
<td>7.5</td>
</tr>
<tr>
<td>CHR 4</td>
<td>1414</td>
<td>181.6</td>
<td>7.8</td>
</tr>
<tr>
<td>CHR 5</td>
<td>668</td>
<td>99.4</td>
<td>6.7</td>
</tr>
<tr>
<td>CHR 6</td>
<td>967</td>
<td>126.5</td>
<td>7.6</td>
</tr>
<tr>
<td>CHR 7</td>
<td>833</td>
<td>128.2</td>
<td>6.5</td>
</tr>
<tr>
<td>CHR 8</td>
<td>596</td>
<td>105.7</td>
<td>5.6</td>
</tr>
<tr>
<td>CHR 9</td>
<td>1318</td>
<td>167.2</td>
<td>7.9</td>
</tr>
<tr>
<td>CHR 10</td>
<td>1130</td>
<td>141.4</td>
<td>8.0</td>
</tr>
<tr>
<td>CHR 11</td>
<td>694</td>
<td>116.8</td>
<td>5.9</td>
</tr>
<tr>
<td>CHR 12</td>
<td>1316</td>
<td>151.0</td>
<td>8.7</td>
</tr>
<tr>
<td>CHR 13</td>
<td>666</td>
<td>124.0</td>
<td>5.4</td>
</tr>
<tr>
<td>CHR 14</td>
<td>835</td>
<td>131.4</td>
<td>6.4</td>
</tr>
<tr>
<td>CHR 15</td>
<td>1048</td>
<td>140.1</td>
<td>7.5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>14,813</strong></td>
<td><strong>2,120.9</strong></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Number of markers per dosage in both parents. T (Tanzania) dosage is the marker dosage in Tanzania. B (Beauregard) dosage is the marker dosage in Beauregard. The dosage groups 0, 1, 2, 3 and 4 correspond to nulliplex, simplex, duplex, triplex and quadruplex marker groups respectively.

<table>
<thead>
<tr>
<th>T. dosage</th>
<th>B. dosage</th>
<th>No. SNPs</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>5973</td>
<td>31.56</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>783</td>
<td>5.02</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>87</td>
<td>0.60</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>16</td>
<td>0.07</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>5</td>
<td>0.02</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>5690</td>
<td>37.01</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1417</td>
<td>9.63</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>374</td>
<td>2.01</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>66</td>
<td>0.60</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>22</td>
<td>0.26</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>2</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>798</td>
<td>5.48</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>254</td>
<td>2.64</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>108</td>
<td>1.13</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>61</td>
<td>0.56</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>11</td>
<td>0.35</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>99</td>
<td>0.58</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>73</td>
<td>0.60</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>38</td>
<td>0.79</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>18</td>
<td>0.19</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>12</td>
<td>0.19</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>22</td>
<td>0.14</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>24</td>
<td>0.21</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0.02</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>4</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Table 3. Summary statistics of QTL for RKN resistance on the TB map. $\sigma^2_{QTL}$ is the variance associated with the QTL. $h^2_{QTL}$ is the heritability of the QTL. QTL limit is the support interval for the QTL on the linkage group.

<table>
<thead>
<tr>
<th>QTL</th>
<th>Chr.</th>
<th>Position</th>
<th>QTL limit (cM)</th>
<th>Marker</th>
<th>p-value</th>
<th>$\sigma^2_{QTL}$</th>
<th>$h^2_{QTL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>120.38</td>
<td>143.68</td>
<td>S3_12962376</td>
<td>4.89e-03</td>
<td>99.6</td>
<td>0.087</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>6.74</td>
<td>49.01</td>
<td>S7_4668738</td>
<td>&lt;1.00e-16</td>
<td>661.4</td>
<td>0.576</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.663</td>
</tr>
</tbody>
</table>
Figure 1. Genetic correlation of the distribution of resistance reactions in the progeny and parents of TB mapping population among different replications separated by time. Raw1 to Raw5 are the raw data for the number of RKN egg masses counted per root system from Rep1 to Rep5. Joint is the joint adjusted mean. Pearson correlation ($p^{** *} < 0.001$) calculated among the adjusted means of 5 replications.
Figure 2. Histogram and boxplot of the distribution of resistance reactions in the TB population in terms of the number of galls of RKN egg masses per root system. Most of the progeny 70% reacted like Tanzania, the resistance parent while 2% were more susceptible than Beauregard. The red bar on the boxplot shows where half of the population (50%) is located.
Figure 3. Histogram of the distribution of a) average dry vine weight and b) average dry root weight in the parents and progeny of the TB population. Most of the genotypes segregated towards the female parent, Tanzania. Average dry vine weight was normally distributed whereas average dry root weight was right skewed.
Figure 4. Boxplot showing the distribution of marker dosages in the TB mapping population. Tanzania 0-1 Beauregard means nulliplex by simplex for Tanzania x Beauregard, 0-2 is nulliplex by duplex, 2-1 is duplex by simplex, and so on.
Figure 5. Integrated SNP-based genetic linkage map of Tanzania x Beauregard mapping population.
Figure 6. QTL profile for the TB population showing the raw data for no. of egg masses (RKN galls) per root system in rep 1 to 5 plotted for QTL analysis. Raw1 is the data for Rep 1, Raw 2 is for Rep 2 and so on. Mean is the adjusted mean for all the five reps. LOP is the logarithm of probability values.
Figure 7. QTL profile of the entire genome for no. of egg masses (RKN galls) per root system in the TB population obtained with the joint model with weighted averages of the five replications (Mean). QTL were identified on linkage groups 2 and 7. LOP is the logarithm of probability values.
Figure 8. Confidence interval (95%) for the mapped QTL for resistance to RKN in the TB mapping population.
Figure 9. Breakdown of genotypic values showing additive effects from each allele at the QTL using their decomposed BLUPs in the population. Alleles a, b, c, d, e, f are from Beauregard whereas alleles g, h, i, j, k, l are from Tanzania. Numbers on the y-axis show the genetic value (deviation from population mean) of the allele, or its average effect. Zero (0) is the population mean (sum of allele frequency * values) of the trait from a normal distribution.
CHAPTER 6

What are the Next Steps for Sweetpotato Genomics Research in SSA?
**What are the Next Steps for Sweetpotato Genomics Research in SSA?**

Sweetpotato is a staple food for a large proportion of the population in many parts of sub-Saharan Africa (SSA). It is of greatest importance in the food systems of Uganda, Rwanda, Burundi and eastern Congo where it forms a major part of the diet. In this region, sweetpotato is also grown as a major income generator for small holder peasant farmers. With an annual production of 2.7 million metric tons – representing almost 3% of the total world production, Uganda is the world’s third largest producer of sweetpotato. In Uganda, sweetpotato is the third most important source of carbohydrates after banana and cassava and is grown in all parts of the country, mainly by subsistence farmers on small plots that rarely exceed 0.5 ha.

Over the past three decades in SSA, the area under sweetpotato production has greatly increased by 64.4% to 5.9 million hectares, but productivity has remained static at around 4.4 ton/ha compared to a world average of 12.7 ton/ha. By far the most important production constraint of sweetpotato in SSA is plant damage caused by sweetpotato virus disease (SPVD) and sweetpotato weevils (SPW), viz. *Cylas* spp. These biotic stresses damage every harvestable part of the plant in a crop, with devastating consequences for poor farmers in SSA leading to lower income and reduced food security.

Due to sweetpotato’s genetic complexity, the need for improved breeding tools to facilitate its improvement is great. This research was conducted to identify QTL conditioning resistance to SPVD, SPW, and RKN and also storage root β-carotene, starch and dry matter content. This helps the sweetpotato breeding community better understand the genetic basis of variation of these complex traits in order to inform breeding decisions for sweetpotato improvement in SSA.
Using the previously developed full sib NKB and TB populations, together with the new genomic tools developed under the Bill and Melinda Gates funded GT4SP project, we have obtained the core resources needed to bring sweetpotato into the genomics era. The new breeding tools include: two fully sequenced and annotated diploid lines (*I. trifida* and *I. triloba*) used as reference genomes for cultivated sweetpotato; a sequencing-based genotyping platform for highly heterozygous hexaploid sweetpotato, GBSpoly, with supporting bioinformatics tools; three high density genetic maps of hexaploid sweetpotato – BT, TB and NKB map; dosage-dependent SNP calling, phasing and linkage mapping algorithms for autopolyploids; robust polyploid QTL analysis and linkage mapping R packages, QTLploy and MAPpoly respectively; and a breeding program database, SweetpotatoBase. All these genomics resources are extremely necessary, along with the highly trained human resource capacity in SSA to put these priceless resources to good use for faster genetic gain of key traits.

Using QTLploy, we were able to predict the genotypic value of full-sib genotypes in both the NKB and TB populations. These methods can be extended to predict individual breeding values for non-phenotyped individuals in the full-sib family, provided they are genotyped and are included in the linkage mapping analysis. Haplotype segments inherited from the parents can be distinguished and used to predict breeding values of the individual containing them, based on our multiple QTL model. This work therefore sets the stage for genomic selection in sweetpotato in SSA!

Compared to the current system of clonal propagation prior to evaluation, GS would allow accurate evaluation at the seedling stage and evaluation of many seedlings, limited only by the program’s capacity for DNA extractions and funds available for genotyping. Lines selected as seedlings, based on their genotypes, could be included in the crossing block in the
second year, providing new segregating progenies for another round of selection. In addition to this, selected seedlings would be looked at more closely in more sites and in more replications thus increasing the odds of making the right call during selection. This ability to effectively evaluate a seedling during year one, rather than waiting until year three or four, would drive the genetic gain of selected traits upward faster and reduce on breeding cycle time, which is cost-effective in the long run. In other words, recurrent mass selection, i.e., multiple cycles of crossing and selection prior to variety release, would become feasible, resulting in a greatly increased rate of genetic gain of all key traits in the program.

Using genomic selection would allow for phenotypic evaluation of individuals, predicted to have good breeding values, in multiple environments and replications therefore increasing efficiency of phenotyping and selection. Since the GS model estimates the effects of alleles, and not clones, not every clone would need to be tested in every environment to determine its performance because its alleles would be represented and its effects estimated accurately. The other consequence of a substantial acceleration of the breeding cycle would be lower cost of producing superior new varieties, thus making them more likely to be adopted by farmers because they can be evaluated in more of their fields. In order to establish a sustainable seed system for sweetpotato in SSA, it is imperative to more frequently release clean improved varieties to replace old ones.

The findings from our research could help breeders in SSA to accelerate sweetpotato breeding by using these developed new genomic tools whose utility, potential and benefits have been clearly demonstrated in this research. The QTL that we have identified for SPVD resistance, RKN resistance, SPW resistance, storage root dry matter, starch and \(\beta\)-carotene
content will need to be evaluated and validated in different populations so that they can be used in routine breeding for sweetpotato in SSA.

Prior to the GT4SP project, the availability of adequately dense markers has been the limiting step for QTL analysis in sweetpotato. Accordingly, the remaining limitations for QTL analysis in sweetpotato in SSA are now primarily at the level of phenotyping and sample size. Small sample sizes may fail to detect QTL of small effect and result in an overestimation of effect size of the few QTL that are identified. Additionally, only those differences that are captured between the initial parents can be mapped. The differences contained within segregating alleles should be large enough at every locus contributing to the variation otherwise some loci will remain undetected. In order to circumvent these challenges, we recommend use of large sample size for QTL mapping. For phenotypic data, make sure that the parents are obviously diverse and use metabolomic / biochemistry data as phenotypes or develop a greenhouse / laboratory assay for the trait to be mapped.

In conclusion, there is need for greater collaboration among the sweetpotato breeding community in SSA and globally in order to strengthen partnerships and leverage resources for a common good. The GT4SP project brought together several partners and formed a stellar team who worked together to develop a molecular breeding toolbox that can be applied in breeding programs across SSA for sweetpotato improvement. The exposure and immediate integration of these tools will encourage long-term, reproducible and sustained use. The high predictive power of these tools will accelerate breeding, save operational costs and lead to adoption of improved sweetpotato varieties.
Appendix A

Storage root color of ‘New Kawogo’, ‘Beauregard’ selected progeny (© Yada, 2014)
Appendix B

Root-Knot Nematode damage on roots of sweetpotato

Galls formed in fibrous roots due to the establishment of female RKN in the root to produce egg masses (bright red knots stained with red food coloring);

Cracking on storage roots due to root-knot nematode (© G. Lawrence, APS).