

**Genome-wide Associations Detect Allele Dose-
dependent Metabolism and Transport Genes as Basis
for Variation in Sweetpotato Storage Root Culinary
Profiles**

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Tara Elizabeth Rickman
December 2020

Copyright © 2020 by Tara Elizabeth Rickman
All rights reserved.

ACKNOWLEDGEMENTS

I would like to acknowledge the invaluable mentorship I received during my time in the department of Entomology and Plant Pathology [EPP] at the University of Tennessee Knoxville [UTK]. I have had the privilege to partake in various leadership roles during my time here and each experience proved to be another opportunity for my growth and development as both a scientist and human.

As a department, EPP is unique in their commitment to celebrating diversity and differences in a welcoming and open manner. I want to give specific recognition to my fellow graduate students who have made my life richer and gave me memories over the past couple of years that will not be soon forgotten. The academic, research, and personal advice and guidance was always welcomed and will be forever cherished.

Most importantly, I want to thank my wife, Madison, family, and friends. A strong support system is essential to any type of success and I would not be writing this if not for their unwavering love and support. Lastly, I would be remiss not to mention the important role of our little zoo of cats and dogs at home- without them I would be missing parts of myself.

ABSTRACT

Previously considered a poor man's crop, sweetpotato production is increasing globally due to its complex carbohydrates and other health benefits. Developing superior cultivars for various emerging market niches is imperative to sustaining and growing the production of sweetpotatoes. Culinary profiles such as sugar composition and storage root color are vital for selecting sweetpotatoes targeted for various value-added products. To determine the genetic architecture and candidate genes underlying these culinary profiles, a GWAS [genome wide association study] on twenty-three traits was performed in over 700 sweetpotato accessions representing global genetic diversity. A new NGS-based quantitative reduced representation sequencing approach (omeSeq/qRRS), an empirical-based quality filtering pipeline (ngsComposer), and a SNP [single nucleotide polymorphism] calling and filtering pipeline (GBSapp) were used to generate a high-density marker dataset comprised of about 80,000 dosage-based SNPs, indels, and CNVs [copy number variations]. The phenotypic data were obtained from 20 years of USDA [United States Department of Agriculture] field trials. A R software package, GWASpoly, was used to test marker-trait associations using a mixed linear model while controlling for population structure with a dosage-based additive relationship matrix. The association analysis accounted for additive and all hexaploid dosage-based dominance models. Candidate genes controlling sweetpotato storage root sugars were revealed to be involved in sugar metabolism and transport, while others underscored the role of sugars in plant biotic and abiotic stress response. While various gene dosage models account for the genetic architecture of the culinary profiles, the additive model was the predominant mode of gene action. This information will help enhance the implementation of next-generation marker and genomic assisted breeding in the genetically complex hexaploid sweetpotato.

TABLE OF CONTENTS

Chapter One Background.....	1
Genetic Analysis in Polyploid Sweetpotato.....	1
Crop Improvement in Polyploid Plants.....	2
Chapter Two Introduction.....	4
Importance of Sweetpotato	4
GWAS in Hexaploidy Plants	5
Chapter Three Materials and Methods.....	7
Phenotypic Data Generation and Exploration.....	7
Genetic Material Generation and SNP Calling.....	7
GWAS Performed using GWASpoly	8
Identification of Candidate Genes	8
Chapter Four Results and Discussion	10
Results.....	10
Discussion	11
List of References	14
Appendix.....	19
Vita.....	42

LIST OF TABLES

Table 1..	20
Table 2..	22
Table 3..	25

LIST OF FIGURES

Figure 1.	25
Figure 2.	26
Figure 3.	27
Figure 4.	28
Figure 5.	29
Figure 6.	30
Figure 7.	31
Figure 8.	32
Figure 9.	33
Figure 10.	34
Figure 11.	35
Figure 12.	36
Figure 13.	37
Figure 14.	38
Figure 15.	39

CHAPTER ONE

BACKGROUND

Genetic Analysis in Polyploid Sweetpotato

Genetic analysis in polyploid plants has been limited, despite advancements in sequencing technology, due to the development of novel bioinformatic tools tailored for diploid organisms. Another reason for the limitation is the comparatively poor understanding of genetic architecture in polyploids vs. diploids. Not surprisingly, this is because of the increased complexity of polyploid evolution (Dufresne et al. 2013). Most polyploids are either autopolyploids, the product of genome doubling within a single species, or allopolyploids, the product of hybridization between different species followed by genome doubling (Stebbins 1947). Autopolyploids represent each chromosome with two or more homologous copies; alternatively, allopolyploids represent each chromosome by two or more sets of divergent chromosomes. Sweetpotato (*Ipomoea batatas* (L.) Lam., Convolvulaceae) is a hexaploidy crop with unclear origins. The closest wild relative is *I. trifida* which is implicated in the origin of cultivated sweetpotato (Huang 2000, Srisuwan et al. 2006). One hypothesis is that a hybridization event between *I. trifida* and *I. triloba* resulted in the ancestor of sweetpotato; however, others speculate *I. trifida* formed an autopolyploid complex from which domestic sweetpotato ancestors developed (Austin 1988, Kobayashi 1984). Regardless of its origin, sweetpotato is a polyploid crop with limited tools for genetic analyses due to its ploidy level.

The major issue with existing software and most new software for genetic analyses is that they are tailored for diploid organisms and therefore lack the proper dosage models needed for polyploid analyses (Meirmans et al. 2018). In polyploids, dosage information is necessary to distinguish partial heterozygous genotypes (Meirmans et al. 2018). Anticipated dosage of alleles at individual loci differs depending on the mode of segregation making assumptions about inheritance patterns significant (Ferris & Whitt 1977, Allendorf 1978). Because of this, issues identifying markers for population genetic analysis arise due to uncertainty in predicting the correct number of gene copies and alleles. Existing population genetics models may be utilized but without segregation knowledge, expected dosage, and/ or allele number there are biases present (Dufresne et al. 2013, Meirmans et al. 2018).

A typical approach to SNP calling is to align DNA fragments of individuals from a population to a shared reference genome and identify regions of variation. For SNP calling, read depth is critical to ensure all alleles at a locus are sampled. Different molecular methods exist to ensure capturing read depth; specifically, RAD [restriction site associated DNA] sequencing has grown in popularity when sequencing to identify markers in polyploid crops. RAD sequencing uses restriction enzymes to shear DNA into fragments to be sequenced on next-generation sequencing [NGS] platforms. This method assigns each sample a unique barcode adapter for multiplex sequencing in a single flow-cell lane which increases the number of samples processed thus reducing costs (Andrews et al. 2016). Though popular for genomic research in crops, the amount of high-quality DNA needed for size selection steps makes RAD sequencing less ideal for frequent plant breeding routines (Wickland et al. 2017). A more feasible sequencing approach called GBS [genotyping-by-sequencing] requires less high-quality DNA, lacks a size selection step, and involves fewer steps overall compared to RAD sequencing (Wickland et al. 2017). The general approach for GBS is to digest DNA samples, ligate barcodes in single wells, pool samples, and enrich via PCR (Poland et al. 2012). An even more specialized GBS approach, GBSpoly, is optimized for highly heterozygous and polyploid genomes. This wet-lab protocol is coupled with a SNP calling and filtering bioinformatic pipeline called GBSapp. These two approaches have successfully been used to identify segregating SNPs in polyploid sweetpotato research (Wadl et al. 2018).

Crop Improvement in Polyploid Plants

As mentioned, to perform genome wide research a reference genome is vital. Due to the advancements in ancestral sweetpotato research, it has been reported that the genomes assemblies for *I. triloba* and *I. trifida* are acceptable reference genomes (Wu et al. 2018). The ability to generate dosage-dependent SNPs coupled with the availability of a reference genome leaves only phenotypic data left to have all starting data to perform a GWAS on sweetpotato. For the analyses performed in this study, the sweetpotato diversity panel, composed of 715 auto-allohexaploid ($2n = 6x = 90$) individuals from both the US sweetpotato germplasm collection maintained by the USDA, ARS, US Vegetable Laboratory (USVL) in Charleston, South Carolina, United States, and the USDA, ARS, Plant Genetic Resources Conservation Unit (PGRCU) in Griffin, Georgia, United States was used. Researchers at the University of Georgia

assessed the quality of storage traits in the USDA sweetpotato germplasm collection; of those traits, the culinary profile explored in this research is composed of 39 traits, shown in Table 1 located in the appendix (Kays et al. 1998).

The primary focus of crop improvement is to select specific individuals from a population with desirable traits to generate an improved variety. Selecting plants generally comes as a result of breeding program trials for one or more traits of interests where various phenotypic data is collected. This process is timely and expensive, as multiple generations of crops are needed to fully access crosses (Collard and Mackill 2008). Knowing with higher confidence which plants possess genes driving desired trait expression, saves time and money for breeding programs. A widely used breeding approach for crop improvement in polyploid plants is marker-assisted breeding. Marker-assisted breeding is described as using molecular markers, such as SNPs, to improve crop traits based on genotypic and phenotypic assays (Jiang 2013). Based on the principle of genetic recombination during meiosis, it is possible to detect SNPs in genomic regions and/or QTLs, quantitative trait loci, linked to traits of interest (Tanksley 1993). There are multiple breeding strategies this encompasses including marker-assisted selection (MAS), marker-assisted backcrossing (MAB), and genome wide selection (GWS) (Ribaut et al. 2010). These types of breeding allow for targeted selection while developing new crop varieties for the market. Sweetpotato marker-assisted breeding generally follows these steps: development of a breeding population with specific individuals included based on desirable traits, assessment of new genetic cultivars via breeding trials, and then release of optimal varieties.

CHAPTER TWO

INTRODUCTION

Importance of Sweetpotato

Sweetpotato, *Ipomoea batatas*, is an important crop globally with its international consumption and production increased significantly in the last fifteen years (Johnson et al, 2015). The increase in consumption is due, in part, to the realized nutritional benefits of sweetpotato in both developing and developed countries (FAO, 2014 and Johnson et al., 2015). Serving as a key source of carbohydrates, vitamins, dietary fiber, and protein, sweetpotato has long been produced and consumed in developing countries to combat invisible hunger (Hall & Phatak, 1993).

Historically, China and African American countries have been the top producers and consumers of sweetpotatoes. Through the nineties, the production of sweetpotato was decreasing in developed countries and increasing in developing countries (Allemann et al., 2004 and Hall & Phatak, 1993). The turn of the century brought a rejuvenated appreciation of the health benefits of sweetpotato and initiated a steady increase in consumption and production of sweetpotatoes in the United States resulting in an average 3.8% annual gain in production since 2001 (Hall & Phatak, 1993 and Johnson et al., 2015). In 2015, the Food and Agriculture Organization (FAO) reported the top six global producers of sweetpotatoes as China, 68.4 %, Nigeria, 3.3%, Uganda, 2.5%, Indonesia, 2.3%, Vietnam, 1.3%, and the United States, 1% (Johnson et al., 2015).

The nutritional benefits of sweetpotatoes are attributed to traits comprising the culinary profile such as sugars, including fructose and glucose, and storage root color, such as tristimulus colorimeter coordinate measurements. Sugars within the profile are important when considering consumers' desire for the availability of complex carbohydrates in their diet. Storage root color is similarly important in terms of capturing consumers' wants. Generally, in the United States, an orange flesh, sweet sweetpotato is more desirable compared to other parts of the world, Asia, Africa, and South America, where white/cream flesh sweetpotatoes are sought after (Jackson et al., 2018). The color variation is correlated to vitamin and mineral richness; for example, an orange flesh sweetpotato will be higher in vitamin A as compared to a white/cream flesh sweetpotato (Ameny and Wilson, 1997). Determining the association between the traits and

genes driving consumer demands, allows breeding programs to better maintain existing germplasm and develop new cultivars with desirable traits.

GWAS in Hexaploidy Plants

Genome-wide association studies (GWAS) are performed to determine associations between traits of interest and potential genes driving observed phenotypic expression. In the last 15 years, GWAS studies have become more widespread in human health research (Ikegawa 2012). Relative to medical research, agricultural use of GWAS is new, has boomed in the last five years (Begum et al. 2015). The delay is due to the fact crops of interests are non-model organisms with a small scientific community and therefore limited resources to develop high-density and genome-wide markers that are required for such analysis. Until now, this problem is more evident in polyploids that lack a robust genotyping platform due to the higher allele dosage (Rosyara et al. 2016; Tian et al. 2020). Genome-wide association studies performed using genome-wide SNP markers are increasing in efficiency and aid in genomic selection methods for crop breeding programs as they provide an exploration of the functional relevance of diverse and unique alleles absent in local breeding population that is often bottlenecked (Sumitomo et al. 2019; Brachi et al. 2011; Meuwissen et al. 2001). Compared to bi-parental populations and linkage-based quantitative trait loci (QTL) analysis that require more effort to design and that often have limited genome resolution, GWA- studies can effectively perform QTL and, in some cases, fine-map potential causal genes due to increased ancestral recombination events that break up the genome into small segments as a result of rapid decay in linkage disequilibrium (LD) (Korte and Farlow, 2013; Khan and Korban, 2012). While some inbred crops (e.g. rice) might lack fine-scale resolution and have extensive LD, outcrossing crops such as sweetpotato are more amenable to the GWAS approach. The rapid LD decay in sweetpotato has been reported to be below the 1kb scale, hence, providing gene-level resolution required for fine-mapping candidate genes (Bararyenya et al. 2020).

Sweetpotatoes are a good example of a complex, polyploid plant in need of software that can meaningfully untangle their hexaploidy and highly heterozygous nature. Rosyara, 2016 successfully designed an R software package, GWASpoly, that incorporates different models of gene action and investigates different types of kinship models for autopolyploids. These features address the requirements for polyploid GWAS studies to be effective. The other need for

successful GWAS analyses is the detection of genetic markers such as single nucleotide polymorphisms (SNPs) which are used in this study. GBSapp is a pipeline that integrates various software to detect thousands of SNPs allowing for a more robust genetic analysis (Wadl et al. 2018).

Meaningful data has been generated by GWAS in Arabidopsis, maize, rice, wheat, tomato, and several other economically important crops (Andrés et al. 2020, Sekhon et al. 2019, Begum et al. 2015, Mathew et al. 2019, Zhao et al. 2019, Sadhukhan et al. 2017). These studies have identified genes involved in abiotic and biotic stress response, plant growth and development, cell signaling, and inter- and intracellular transport to name a few. Exploration of candidate genes has led to progress in multiple marker-assisted breeding programs (Olatoye et al. 2019, Sekhan et al. 2019, Frouin et al. 2019). The elucidation of the genetic architecture controlling desirable traits frequently results in discoveries between major plant systems involved in phenotypic expressions.

Inter- and intracellular transport are vital to plant processes directly linked to plant development, abiotic and biotic stress response, and nutrient transportation. Multiple genes encoding proteins involved in membrane channels with selectivity for cations transporting sugars; proteins involved in importing and exporting substrates involved in plant nutrition, growth, and stress response; and proteins that act as sugar transmembrane transporters have all been identified in Arabidopsis, maize, and/or rice (Andrés et al. 2020, Sekhon et al. 2019, and Begum et al. 2015). Existing annotation of genes involved in these plant processes is a valuable tool when embarking on understanding the gene function in other economically important plants such as sweetpotatoes.

The agronomic value of sweetpotato cannot be understated nor can the nutritional benefits. With recent progress in genome assemblies for reference genomes to use in sweetpotato-related research, determining the significant association between traits within the culinary profile and SNPs allows for more efficient next-generation marker and genomic assisted breeding programs. Here we describe how we performed a GWAS and identified candidate genes controlling sweetpotato storage root sugars putatively involved in sugar metabolism and transport and plant biotic and abiotic stress response.

CHAPTER THREE

MATERIALS AND METHODS

Phenotypic Data Generation and Exploration

The sweetpotato diversity panel is composed of 715 hexaploid ($2n = 6x = 90$) individuals from both the US sweetpotato germplasm collection maintained by the USDA, ARS, US Vegetable Laboratory (USVL) in Charleston, South Carolina, United States, and the USDA, ARS, Plant Genetic Resources Conservation Unit (PGRCU) in Griffin, Georgia, United States.

Researchers at the University of Georgia assessed the quality of storage traits in the USDA sweetpotato germplasm collection (Kays et al. 1998). Lines were grown at the University of Georgia Horticulture Farm in 1996 under standard commercial practices. At harvest, storage roots were cured 7 days at 29 °C and stored at 13 °C until further processed. For each trait measured, three storage roots per line were selected for sensory and chemical analysis.

The culinary profile is composed of 39 traits with 13 traits having at least one significant marker-trait association. To explore the relationship between those 13 traits within the culinary profile, the following analyses were performed. The raw phenotypic data for every trait was log-transformed in R to achieve a more normal distribution. A Spearman's correlation analysis for non-parametric data was performed to generate a correlogram with 3 ways to visualize the data. Fig.1 includes a histogram and scatter plot of each log-transformed trait. To explore significant correlations between traits within the culinary profile, a correlation plot was generated where blue indicates positive correlations, red indicates negative correlations, and white indicates little to no correlation.

Genetic Material Generation and SNP Calling

Total genomic DNA was isolated from freeze-dried leaf tissue using the DNeasy Plant Mini Kit (Qiagen). The integrity, purity, and concentration of the isolated genomic DNA were determined by 1% agarose gel electrophoresis and the fluorescence-based PicoGreen dsDNA assay using a Synergy HTX Multi-Mode Microplate Reader. The genomic library preparation was performed using a modified genotype-by-sequencing (qRRs) protocol with the OmeSeq restricted representation sequencing method optimized for highly heterozygous and polyploid genomes as

described by Wadl et al, 2018. ngsComposer performed stringent quality filtering of raw sequence reads before SNP discovery and filtering (Kuster et al., 2020).

The GBSapp pipeline was used for processing raw fastq reads followed by variant calling and filtering. GBSapp integrates GATK (Genome Analysis Toolkit) v3.7, which is applicable to organisms of any ploidy level, and optimizes it for highly heterozygous, polyploid species making it ideal for sweetpotato genomic analysis (Wadl et al. 2018). The physical reference genomes of sweetpotato's ancestral diploid progenitors, *I. trifida* and *I. triloba* were used in the variant calling step. Filtering parameters included read depth filtering for each data point and marker removal of markers. GBSapp generated ~80,000 high quality, dosage-dependent SNPs with maf of 0.02, and no more than 20% missing data.

GWAS Performed using GWASpoly

GWASpoly incorporates 8 models of gene action and operates using optimized kinship models for hexaploidy sweetpotato (Rosyara et al. 2016). The results are interpreted using Manhattan plots and quantile-quantile plots (Q-Q plots) (Pearson and Manolio, 2008). Manhattan plots show a significance level (- log-base-10 of p-value) for each SNP along the y-axis and the genomic position along the x-axis. The Bonferroni correction is used for significance to adjust for errors or false positives. Each dot represents a SNP in its location within the sweetpotato genome. The Bonferroni threshold runs horizontally on the plot with each SNP falling above it considered significant and thus necessary to further explore. The higher the SNP falls above the threshold, the stronger the trait-marker association. The Q-Q plots were used to explore the false positive rate of SNP associations due to confounding factors. The y-axis shows the significance level of the association from least to greatest. The x-axis shows the SNP markers plotted against the expected distribution if there were no association. The SNPs that deviate from the expected distribution, are considered true associations. These plots provide a visual confirmation of significant SNP associations.

Identification of Candidate Genes

Using the significant trait-associated SNPs identified by GWASpoly, the physical genome assembly of the diploids *I. trifida* and *I. triloba* (<http://sweetpotato.plantbiology.msu.edu/>) were used as the reference genomes for identifying candidate genes. The presumed functions of each

candidate gene were annotated using the Michigan State University *I. trifida* JBrowse feature. Additional gene annotations were performed using known gene function from the highly annotated genome of *Arabidopsis thaliana* as well as NCBI's BLAST (National Center for Biotechnology Information Basic Local Alignment Search Tool). An extensive literature review supports the proposed annotation of candidate genes.

CHAPTER FOUR

RESULTS AND DISCUSSION

Results

The traits within the culinary profile with significant associations include the concentration of baked fructose, raw fructose, baked glucose, raw glucose, baked maltose, raw maltose, total reducing sugar, cooked root colors *A and *L, uncooked root color *B, total baked hexoses, total raw hexoses, and total baked sugars. Tables 1 and 2 define how data for each trait was generated. The significant correlations between traits are shown in Fig. 1. Baked maltose has a slightly positive correlation with baked and raw fructose and glucose. This correlation may be due to the conversion of maltose to simple sugars during the baking process. Total hexoses raw and baked have a strongly negative correlation with baked and raw fructose and glucose. Cooked root color L has a positive correlation with baked and raw fructose, glucose, and total hexoses. All other traits have a medium to slight positive and negative correlations.

The primary charts used to interpret the significant marker-trait associations are Manhattan plots and quantile-quantile (Q-Q) plots. Dosage models (diplo- additive, additive, 1-dom-alt, 1-dom-ref, 2-dom-alt, 2-dom-ref, 3-dom-alt, 3-dom-ref) for this polyploid crop are represented by individual Manhattan plots for each trait. The results indicate each trait has optimal allele amounts necessary to produce a crop with a desirable culinary profile trait. The identified candidate genes we are reporting in this paper fall above either one or both horizontal lines representing FDR, false discovery rate, threshold (black line), and Bonferroni threshold (red line). Of the two thresholds, Bonferroni is more stringent at correcting for type 1 errors, false positives; however, FDR has greater power with the tradeoff of increased type 1 errors (Shaffer 1995). The Q-Q plots graphically show which SNPs, if any, are behaving abnormally and falling outside of the confidence interval for each dosage model. Thirteen of the culinary profile traits have at least one significant SNP-gene association with 2 or more SNPs within the genomic region of interest. The annotated genes and their putative functions are listed in Tables 1 and 2. The annotated candidate genes and their putative functions are shown in Tables 1 and 2 with the primary plant processes represented being: translation/transcription, sugar and protein transport, stress response, cell growth and development, and carbohydrate metabolism. Of the 13 traits with significant SNP associations, 9 have two or more candidate genes involved in different putative

plant functions driving their expression indicating the complexity of selecting for these traits (Fig. 2). Table 2 shows the candidate genes and their putative functions for the sugar traits within the culinary profile. Amongst the sugar traits, twelve candidate genes were identified with putative functions representing all primary plant processes. The root color traits, Table 3, have six candidate genes identified representing three of the five primary plant processes: translation/transcription, sugar and protein transport, and cell growth and development. Figure 3 shows the Manhattan plot (left) and Q-Q plot (right) for raw fructose. On the third chromosome, between the genomic region of Chr03_11254983-Chr03_1186727, there are five significant marker-trait associations with two or more SNPs in each candidate genomic region. Of the five genes, *VRN2* is involved in translation/transcription, *APKIA* is involved in stress response, *At3g17140* is involved in carbohydrate metabolism, and *SWEETII* and *UMAMIT16-psi* are involved in sugar and protein transport. Raw fructose, baked fructose, baked glucose, and raw glucose all have significant marker-traits associations with SNPs falling in genomic regions where candidate genes *UMAMIT16-psi* and *APKIA* are located. The *SWEETII* candidate gene has a significant marker-trait association in raw fructose, baked glucose, and raw glucose. The culinary profile can be broken into two sets of traits, sugars traits, and root traits, with the ten sugar traits being driven by four different major plant processes, sugar and protein transport, carbohydrate metabolism, stress response, and translation/ transcription, and a total of twelve candidate genes. The three root traits are driven by three major plant processes, sugar and protein transport, carbohydrate metabolism, and translation/ transcription, and a total of six candidate genes.

Discussion

The economic importance of sweetpotato, internationally and nationally, is increasing along with consumer demands. The identification of candidate genes driving the expression of traits within the culinary profile in sweetpotato provides meaningful guidance for marker-assisted breeding programs. The traits driving flavor are controlled by several biological processes indicating high trait complexity. Dosage models for this polyploid crop indicate each trait has optimal allele amounts necessary to produce a crop with a desirable trait. There are a diverse amount of candidate genes implicated in having a role in controlling culinary traits responsible for the eating quality of the sweetpotato. Identified genes have putative functions in carbohydrate

metabolism, cell growth & development, stress response, sugar and protein transport, and translation/ transcription.

Thirteen of the traits with significant marker-trait associations are driven by multiple genes with putative functions in three or more major plant processes indicating high trait complexity (Fig. 2). Carbohydrate metabolism and plant stress response are driving trait expression in seven of the twenty-two traits with significant associations. This suggests a meaningful connection between the expression of raw and baked glucose, fructose, total hexoses, and total baked sugars, and the plant processes carbohydrate metabolism and stress response (Fig. 2). Furthermore, this indicates a relationship between sweetpotato storage root sugar production and the plants' stress response as it pertains to carbohydrate metabolism and transport.

Figure 3 shows the annotated manhattan plot (left) and Q-Q plot (right) for raw fructose. The five candidate genes are *APKIA*, *At3g17140*, *SWEETII*, *VRN2*, and *UMAMIT16-psi*. These genes are involved in stress response, carbohydrate metabolism, and sugar and hormone transport.

Figures 4-12 show the remaining annotated manhattan and Q-Q plots for the sugar traits with the primary involvement of these genes being in one or more of the following plant processes: carbohydrate metabolism, stress response, sugar and protein transport, and translation/ transcription. Furthermore, of the ten sugar traits, six, baked, and raw fructose, glucose, and total hexoses have all significant associations localized on the third chromosome. The localization of candidate genes on the third chromosome and the importance of their putative functions on plant health in terms of carbohydrate metabolism and transport and stress response suggests a potential locus for breeders to select for to enhance both the eating quality of the sweetpotato as well as fortifying overall plant health.

The four traits with significant associations outside of the third chromosome are total reducing sugars, baked maltose, raw maltose, and total sugars, Figs. 2, 7, 8, and 11, respectively. Of these, total sugars, total reducing sugars, and baked maltose all have at least one significant association on chromosome thirteen demonstrating a localization on this genomic region. The putative function of the candidate genes in this genomic region is carbohydrate metabolism, sugar and protein transport, and translation/ transcription. Chromosome eleven has two traits with significant marker-trait associations, baked maltose, and total reducing sugars, with the same marker Chr11_2932505. The putative function of the gene this marker falls within is sugar

and protein transport, specifically the transfer of glycosyl groups which are involved in carbohydrate production.

Figures 13-15 show the annotated manhattan and Q-Q plots for the root color traits. None of these subsets of traits have overlapping marker-trait associations. Uncooked root color *B, Fig. 15, has the following candidate genes: *MMI9.7*, *GYF*, and *pdxB*. These genes have been shown to have roles in independent plant processes and are each located on separate chromosomes. Most interesting out of the root color traits is the association between marker Chr03_2075979 and cooked root color *L, Fig. 14. This marker is in the same region as the SNPs on chromosome three with strong sugar trait associations. This marker falls within *MATE*, which is a gene involved in sugar and protein transport, specifically transporting organic acids, plant hormones, and secondary metabolites. Cooked root color *L is the measure of lightness of the sweetpotato periderm and Fig. 1 shows a positive correlation between this trait and baked and raw fructose and glucose. It is interesting to see a positive correlation between these traits as well as QTLs on the same chromosome; however, the marker associated with periderm lightness is over 9 million bp away from the markers associated with the sugar traits.

The successful identification of over 80,000 high-density markers and maintenance of 20-year-old USDA phenotypic data on ~700 sweetpotato accessions has provided the opportunity to analyze sweetpotato germplasm data in a meaningful way to contribute to next-generation marker-assisted breeding programs. The availability of bioinformatics software for polyploidy genomic analyses has allowed the utilization of old and new data to develop our understanding of organisms of all ploidy levels. Candidate genes controlling sweetpotato storage root sugars were revealed to be involved in sugar metabolism and transport, while others underscored the role of sugars in plant biotic and abiotic stress response. While various gene dosage models account for the genetic architecture of the culinary profiles, the additive model was the predominant mode of gene action. This information will help enhance the implementation of next-generation marker and genomic assisted breeding in the genetically complex hexaploid sweetpotato.

LIST OF REFERENCES

1. Allemann, J., Laurie, S.M., Thiart, S., Vorster, H.J., and Bornman, C.H. (2004). Sustainable production of root and tuber crops (potato, sweet potato, indigenous potato, cassava) in southern Africa. *South African Journal of Botany* 70, 1. [https://doi.org/10.1016/S0254-6299\(15\)30307-0](https://doi.org/10.1016/S0254-6299(15)30307-0)
2. Allendorf, F. W. (1978) Protein polymorphism and the rate of loss of duplicate gene expression. *Nature*. 272: 76-78.
3. Ameny M., Wilson, P. (1997) Relationship between Hunter color values and β -carotene contents in white fleshed African sweet potatoes (*Ipomoea batatas* Lam.). *J Sci Food Agric* 73:301–306
4. Andrés, F., Kinoshita, A., Kalluri, N. et al. (2020). The sugar transporter SWEET10 acts downstream of FLOWERING LOCUS T during floral transition of *Arabidopsis thaliana*. *BMC Plant Biol* 20, 53. <https://doi.org/10.1186/s12870-020-2266-0>
5. Andrews, K. R., Good, J. M., Miller, J. M., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews*. 17: 81-92.
6. Austin, D. F. (1988). The taxonomy, evolution and genetic diversity of sweet potatoes and related wild species. Lima, Peru. *International Potato Center* (CIP). 27–58.
7. Bararyenya, A., Olukolu, B. A., Tukamuhabwa, P. et al. (2020). Genome-wide association study identified candidate genes controlling continuous storage root formation and bulking in hexaploidy sweetpotato. *BMC Plant Bio*. 20 (3).
8. Begum, H., Spindel, J. E., Lalusin, A., Borromeo, T., Gregorio, G., Hernandez, J., Mccouch, S. R. (2015). Genome-Wide Association Mapping for Yield and Other Agronomic Traits in an Elite Breeding Population of Tropical Rice (*Oryza sativa*). *Plos One*, 10(3). doi: 10.1371/journal.pone.0119873
9. Brachi, B., Morris, G.P. & Borevitz, J.O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol* 12, 232. <https://doi.org/10.1186/gb-2011-12-10-232>
10. Chardon F, Bedu M, Calenge F, Klemens PA, Spinner L, Clement G, et al. (2013). Leaf fructose content is controlled by the vacuolar transporter SWEET17 in *Arabidopsis*. *Curr. Biol*. 23.
11. Chen L-Q, Hou BH, Lalonde S, Takanaga H, Hartung ML, Qu XQ, et al. (2010). Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature*. vol. 468 (pg. 527-532)
12. Chen L-Q, Qu XQ, Hou BH, Sosso D, Osorio S, Fernie AR, et al. (2012). Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science*. vol. 335 (pg. 207-211)
13. Collard, B. C. Y. and Mackill, D. J. (2007) Marker-assisted selection: an approach for precision plant breeding in the twenty first century. *Philos Trans R Soc Lond B Biol Sci*. 363: 557-572.

14. Denancé, N., Szurek, B., & Noël, L. D. (2014). Emerging Functions of Nodulin-Like Proteins in Non-Nodulating Plant Species. *Plant and Cell Physiology*, 55(3), 469–474. doi: 10.1093/pcp/pct198
15. Dufresne, F., Stift, M., Vergilino, R., & Mable, B. K. (2013). Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. *Mol Ecol*. 23:40-69. <https://doi.org/10.1111/mec.12581>
16. FAO. (2014). *Sweet potatoes – cultivating resilience, fighting malnutrition in Somalia*. Food and Agriculture Organization of the United Nations. <http://www.fao.org/emergencies/fao-in-action/stories/stories-detail/en/c/266043/>.
17. Ferris, M. & Whitt, G. S. (1977) Loss of duplicate gene expression after polyploidization. *Nature*. 265: 258-260.
18. Frouin, J., Labeyrie, A., Boissard, A., Sacchi, G. A., & Ahmadi, N. (2019). Genomic prediction offers the most effective marker assisted breeding approach for ability to prevent arsenic accumulation in rice grains. *Plos One*, 14(6). doi:10.1371/journal.pone.0217516
19. Hall, M.R. and Phatak, S.C. (1993). Sweet potato: *Ipomoea batatas* (L) Lam. *Genetic Improvement of Vegetable Crops*. <https://doi.org/10.1016/B978-0-08-040826-2.50051-5>
20. Harter, K., & Weber, A. P. M. (2013). Arabidopsis 2010 and beyond—big science with a small weed. *Frontiers in Plant Science*. doi: 10.3389/fpls.2013.00018
21. Huang, J. (2000) Genetic diversity and relationships of sweetpotato and its wild relatives in *Ipomoea* series *Batatas* (Convolvulaceae) as revealed by inter-simple sequence repeat (ISSR) and restriction analysis of chloroplast DNA. *Theor Appl Genet*. 100: 1050-1060.
22. Ikegawa S. (2012). A short history of the genome-wide association study: where we were and where we are going. *Genomics & informatics*, 10(4), 220–225. <https://doi.org/10.5808/GI.2012.10.4.220>
23. Jackson, M.D., Harrison, H.F., Jarret, R.L., Wadley, P.A. Color analysis of storage roots from the USDA, ARS sweetpotato (*Ipomoea batatas*) germplasm collection. *Genet Resour Crop Evol* 65, 1217–1236 (2018). <https://doi.org/10.1007/s10722-018-0609-6>
24. Jiang, G. (2013). Molecular Markers and Marker-Assisted Breeding in Plants. *Plant Breeding from Laboratories to Fields*. Croatia: InTech, doi: <http://dx.doi.org/10.5772/52583>
25. Johnson, T., Wilson, N., Worosz, M. R., Fields, D., Bond, J.K. (2015). Vegetables and Pulses Outlook: Special Article Commodity Highlight: Sweet Potatoes. *USDA: Economic Research Service*.
26. Kays, S. J., McLaurin, W. J., and Wang, Y. (1998). Sweetpotato. U.S. Nat. Plant Germplasm System. <https://npgsweb.ars-grin.gov/gringlobal/method.aspx?id=490810>
27. KEGG Orthology (KO) - Arabidopsis thaliana (*thale cress*). Retrieved from https://www.genome.jp/kegg-bin/get_htext?ath00001+AT4G23900
28. Khan, M.A., Korban, S.S. (2012). Association mapping in forest trees and fruit crops. *Exp. Bot*. 4045– 4060. doi:10.1093/jxb/ers105
29. Klemens P. A., Patzke K., Deitmer J., Spinner L., Le Hir R., Bellini C., et al. Overexpression of the vacuolar sugar carrier AtSWEET16 modifies germination, growth, and stress tolerance in Arabidopsis, *Plant Physiol.*, 2013, vol. 163

30. Kobayashi, M. (1984). The *Ipomoea trifida* complex closely related to sweet potato. In: Shideler SF, Rincon H, editors. Proceedings of the 6th Symposium of the International Society of Tropical Root Crop. Lima, Peru: CIP. 561–568.
31. Korte, A., Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9, 29. <https://doi.org/10.1186/1746-4811-9-29>
32. Layer, G., Heinz, D. W., Jahn, D., & Schubert, W.-D. (2004). Structure and function of radical SAM enzymes. *Current Opinion in Chemical Biology*, 8(5), 468–476. doi: 10.1016/j.cbpa.2004.08.001
33. Lilla, E., Yokoyama, K. (2016). Carbon extension in peptidynucleoside biosynthesis by radical SAM enzymes. *Nat Chem Biol* 12, 905–907. <https://doi.org/10.1038/nchembio.2187>
34. Mathew I, Shimelis H, Shayanowako AIT, Laing M, Chaplot V (2019) Genome-wide association study of drought tolerance and biomass allocation in wheat. *PLoS ONE* 14(12): e0225383. <https://doi.org/10.1371/journal.pone.0225383>
35. Meirmans, P. G., Liu, S., & Van Tienderen, P. H. (2018). The Analysis of Polyploid Genetic Data. *Journal of Heredity*, 109(3), 283–296. <https://doi.org/10.1093/jhered/esy006>
36. Merilo, E., Jalakas, P., Kollist, H., & Brosché, M. (2015). The Role of ABA Recycling and Transporter Proteins in Rapid Stomatal Responses to Reduced Air Humidity, Elevated CO₂, and Exogenous ABA. *Molecular Plant*, 8(4), 657–659. doi: 10.1016/j.molp.2015.01.014
37. Meuwissen, T. H., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), 1819–1829.
38. NCSU Dept. Hort. Sci. Sweetpotato and Potato Breeding and Genetics Program. Retrieved April 4, 2020, from <https://potatoes.ncsu.edu/Sweetpotato.html>
39. Olatoye, M.O., Clark, L.V., Wang, J. et al. (2019). Evaluation of genomic selection and marker-assisted selection in Miscanthus and energycane. *Mol Breeding*. <https://doi.org/10.1007/s11032-019-1081-5>
40. Otulak-Kozieł, K., Kozieł, E., & Valverde, R. A. (2019). The Respiratory Burst Oxidase Homolog D (RbohD) Cell and Tissue Distribution in Potato–Potato Virus Y (PVYNTN) Hypersensitive and Susceptible Reactions. *International Journal of Molecular Sciences*, 20(11), 2741. doi: 10.3390/ijms20112741
41. Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *Plos One*. doi: <https://doi.org/10.1371/journal.pone.0032253>
42. Pao, S. S., Paulsen, I. T., & Saier, M. H., Jr (1998). Major facilitator superfamily. *Microbiology and molecular biology reviews : MMBR*, 62(1), 1–34.
43. Pearson TA, Manolio TA. (2008). How to Interpret a Genome-wide Association Study. *JAMA*. 299(11):1335–1344. doi:10.1001/jama.299.11.1335
44. Ribaut, J. M., De Vicente, M. C., & Delannay, X. (2010). Molecular breeding in developing countries: challenges and perspectives. *Current Opinions in Plant Biology*. 13:213-218.
45. Rosyara, U.R., De Jong, W.S., Douches, D.S. and Endelman, J.B. (2016). Software for Genome-Wide Association Studies in Autopolyploids and Its Application to Potato. *The Plant Genome*, 9: 1-10. doi:10.3835/plantgenome2015.08.0073

46. Sadhukhan A., Kobayashi Y., Nakano Y., Iuchi S., Kobayashi M., Sahoo L., and Koyama H. (2017). Genome wide Association Study Reveals that the Aquaporin NIP1;1 Contributes to Variation in Hydrogen Peroxide
47. Sensitivity in *Arabidopsis thaliana*. *Mol. Plant*. 10, 1082–1094.
48. Sekhon, R. S., Saski, C., Kumar, R., Flinn, B. S., Luo, F., Beissinger, T. M., Ackerman, A. J., Breitzman, M. W., Bridges, W. C., de Leon, N., & Kaeppler, S. M. (2019). Integrated Genome-Scale Analysis Identifies Novel Genes and Networks Underlying Senescence in Maize. *The Plant cell*, **31**(9), 1968–1989. <https://doi.org/10.1105/tpc.18.00930>
49. Shaffer, J. P. (1995) Multiple Hypothesis Testing. *Annual Review of Psychology*, **46**:561-584. <https://doi.org/10.1146/annurev.ps.46.020195.003021>
50. Stebbins, G. L. (1947). Types of polyploids: their classification and significance. *Advances in Genetics*. 1: 403-429.
51. Srisuwan, S., Sihackakr, D., Siljak-Yakovlev, S. (2006). The origin and evolution of sweet potato (*Ipomoea batatas* Lam.) and its wild relatives through the cytogenetic approaches. *Plant Sci*. 171: 424-433.
52. Sumitomo, K., Shirasawa, K., Isobe, S., Hirakawa, H., Hisamatsu, T., Nakano, Y., Yagi, M., and Ohmiya, A. (2019). Genome-wide association study overcomes the genome complexity in autohexaploid chrysanthemum and tags SNP markers onto the flower color genes. *Sci Rep* **9**, 13947. <https://doi.org/10.1038/s41598-019-50028-z>
53. Tanksley, S. D. (1993). Review Mapping polygenes. *Annu Rev Genet*. 1993; 27():205-33.
54. Tian, D., Wang, P., Tang, B., Teng, X., Li, C., Liu, X., Zou, D., Song, S., Zhang, Z. (2020). GWAS Atlas: a curated resource of genome-wide variant-trait associations in plants and animals. *Nucleic Acids Research*, **48**: 927-932. <https://doi.org/10.1093/nar/gkz828>
55. UniProt. (2020, April 22). Exocyst complex component EXO84B. Retrieved from <https://www.uniprot.org/uniprot/Q9LTB0>
56. U.S. Dept. of Agriculture (2019). Census of Agriculture, 2017. Retrieved from https://www.nass.usda.gov/Publications/AgCensus/2017/Full_Report/Volume_1,_Chapter_1_US/usv1.pdf
57. Wadl, P. A., Olukolu, B. A., Branham, S. E., Jarret, R. L., Yencho, G. C., & Jackson, D. M. (2018). Genetic Diversity and Population Structure of the USDA Sweetpotato (*Ipomoea batatas*) Germplasm Collections Using GBSpoly. *Frontiers in plant science*. 9, 1166. <https://doi.org/10.3389/fpls.2018.01166>
58. Wickland, D. P., Battu, G., Hudson, K. A., Diers, B. W., & Hudson, M. E. (2017). A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-eaSy. *BMC Bioinformatics*. 18, 586
59. Wu, S., Lau, K.H., Cao, Q. et al. (2018). Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. *Nat Commun*.
60. Zhao, J., Sauvage, C., Zhao, J. et al. (2019). Meta-analysis of genome-wide association studies provides insights into genetic control of tomato flavor. *Nat Commun* 10, 1534. <https://doi.org/10.1038/s41467-019-09462-w>

APPENDIX

Table 1. List of traits making up the full culinary profile with USDA- GRIN descriptors of each trait. Of these traits, 13 have significant SNP- trait associations.

Trait	Description
CITRIC_BAKED	Citric acid concentration in the baked root flesh
CITRIC_RAW	Citric acid concentration in the raw root flesh
FRUCTOSE_BAKED	Fructose concentration in the baked root flesh
FRUCTOSE_RAW	Fructose concentration in the raw root flesh
GALACTOSE_BAKED	Galactose concentration in the baked root flesh
GALACTOSE_RAW	Galactose concentration in the raw
GLUCOSE_BAKED	Glucose concentration in the baked root flesh
GLUCOSE_RAW	Glucose concentration in the raw root flesh
INOSITOL_BAKED	Inositol concentration in the baked root flesh
INOSITOL_RAW	Inositol concentration in the raw root flesh
MALIC_BAKED	Malic acid concentration in the baked root flesh
MALIC_RAW	Malic acid concentration in the raw root flesh
MALTOSE_BAKED	Maltose concentration in the baked root flesh
MALTOSE_RAW	Maltose concentration in the raw root flesh
QUINIC_BAKED	Quinic acid concentration in the baked root flesh
QUINIC_RAW	Quinic acid concentration in the raw root flesh
REDSUGTOT	Total reducing sugars in baked root flesh
SUCROSE_BAKED	Sucrose concentration in the baked root flesh
SUCROSE_RAW	Sucrose concentration in the raw root flesh
TOTHEXOSES_BAKED	Total hexoses baked represent the composite of glucose and fructose measured in G/100G dry matter
TOTHEXOSES_RAW	Total hexoses raw represent the composite of glucose and fructose measured in G/100G dry matter
TOTORGACID_BAKED	Total organic acids baked represent a composite of malic, citric and quinic acids measured in G/100G dry matter
TOTORGACID_RAW	Total organic acids raw represent a composite of malic, citric and quinic acids measured in G/100G dry matter
TOTSUGARS_BAKED	Total sugars of baked storage root
TOTSUGARS_RAW	Total sugars of raw storage root
ROOTCOLUNI	Root Color Uniformity
DRYMATLOSS	Dry Matter Loss
DRYMATTER	Percent dry matter
ROOTCOCOOK_A	Cooked Root Color a
ROOTCOCOOK_B	Cooked Root Color b
ROOTCOCOOK_L	Cooked Root Color l
ROOTCOUNCO_A	Uncooked Root Color a

Table 1 Continued

ROOTCOUNCO_B	Uncooked Root Color b
ROOTCOUNCO_L	Uncooked Root Color l
SUGAREQUIV	Baked Root Sugar Equivalents: General comparison of sweetness in baked root flesh. Sugars converted to their relative sweetness in relation to sucrose, using the equation: sucrose equivalents = 1.2 fructose + 1 sucrose + 0.64 glucose + 0.43 maltose
SWEETNESS	Sensory Root Sweetness
STARCHDWB	Starch dry weight basis
STARCHFVB	Starch Fresh Weight Basis
ANTHOCYAN	distribution of anthocyan pigment in flesh

Table 2. Table of annotated candidate genes identified within the sugar traits of the culinary profile. Twelve candidate genes have been identified with putative function in the following cell processes: abiotic and biotic stress response, plant growth and development, cell signaling, inter- and intracellular transport, and DNA/RNA processing/expression.

Ortholog	Putative Function	Trait(s)	Gene dosage model	$-\log_{10}P$ range	SNP ID	SNP proximity to gene (bp)
Protein Kinase Superfamily	Protein phosphorylation in responses to light, pathogen invasion, hormones, temperature stress, and nutrient deprivation	B- Fruc. R- Fruc. B- Gluc. R- Gluc. B- Tot. Hex. R- Tot. Hex.	Additive	6.4-7.54	Chr03_1125498 3	
Plant invertase/pectin methylesterase inhibitor superfamily	Hydrolyze sucrose into glucose and fructose.	B- Fruc. R- Fruc. B- Tot. Hex. R- Tot. Hex.	Additive	7.05-7.55	Chr03_1139367 9	
nodulin MtN21 /EamA-like transporter family protein	Amino acid, auxin, and sugar transport.	B- Fruc. R- Fruc. B- Gluc. R- Gluc. B- Tot. Hex. R- Tot. Hex.	Additive	6.37-7.07	Chr03_1186722 7	
VEFS-Box of polycomb protein	Controls HOX gene transcription	R- Fruc.	Additive	6.58-	Chr03_1159934 1	

Table 2 Continued

Major facilitator superfamily protein	Transport proteins that transport small solutes across cell membranes via chemiosmotic gradients.	B- Malt.	2-dom-alt	6.92-	Chr13_18612804	
Sugar Transporter (SWEET11)	Intracellular sugar transport.	B- Gluc. R- Gluc. B- Tot. Hex. R- Tot. Hex. B- Fruc. R- Fruc.	Additive	5.27-5.81	Chr03_11589347	
ataxia-telangiectasia mutated	Involved in DNA repair, RNA splicing, and cellular response to DNA damage stimulus.	R- Malt.	3-dom-alt	6.1	Chr02_14832741	
transferases transferring glycosyl groups	Catalysis of the transfer of glycosyl groups.	B- Malt. Tot. Red. Sug.	3-dom-alt	5.68-6.49	Chr11_2932505	
glycosyl hydrolase 9B13	Decomposition of cellulose.	B- Gluc. R- Gluc.	Additive	6.26-6.86	Chr03_11385900	58 bp downstream
cytosolic invertase	Hydrolyzes sucrose to glucose and fructose.	B- Malt. B- Tot. Sug.	Diplo-additive	6.17-8.4	Chr13_18532390	

Table 2 Continued

Phosphoinositide phosphatase family protein	Hydrolyzes phosphoric ester bonds	Tot. Red. Sug.	Additive	6.07	Chr13_18565303	
golgi nucleotide sugar transporter	Golgi-localized nucleotide sugar transporters (NSTs) are essential for the biosynthesis of wall polysaccharides and glycoproteins based on their characteristic transport of many nucleotide sugars to the Golgi lumen	Tot. Red. Sug.	2-dom-ref	5.09-	Chr14_8490664	

Table 3. Table of annotated candidate genes identified within the root color traits of the culinary profile. Six candidate genes have been identified with putative function in the following cell processes: growth and development, inter- and intracellular transport, and DNA/RNA processing/expression.

Ortholog	Putative Function	Trait(s)	Gene dosage model	$-\log_{10}P$ range	SNP ID	SNP proximity to gene (bp)
Erythronate-4-phosphate dehydrogenase family protein	Catalyzes the oxidation of erythronate-4-phosphate to 3-hydroxy-2-oxo-4-phosphonooxybutanoate. Step 2 in pyridoxine pathway responsible for vitamin B6 compounds.	Uncooked Root Color B	1-dom-alt	6.47	Chr08_2539913	
selenoprotein family protein	Unknown	Cooked Root Color A	2-dom-alt	8.74	Chr07_5845228	
homolog of yeast autophagy 18 (ATG18) H	Required for cytoplasm to vacuole transport, vesicle formation, and starvation induced autophagy.	Cooked Root Color A	2-dom-ref	5.64	Chr10_457314	
GYF domain-containing protein	GYF domain binds proline-rich sequences.	Uncooked Root Color B	3-dom-alt	6.17	Chr06_23391276	
MATE efflux family protein	Extrude organic compounds, transporting a broad range of substrates such as organic acids, plant hormones and secondary metabolites.	Cooked Root Color L	Additive	6.5	Chr03_2075979	

Table 3 Continued

Cell cycle regulated microtubule associated protein	Involved in activation of protein kinases and regulation of mitotic spindle organization.	Uncooked Root Color B	Additive	6.75	Chr03_11286710	149 bp downstream
--	---	-----------------------	----------	------	----------------	-------------------

List of Figures

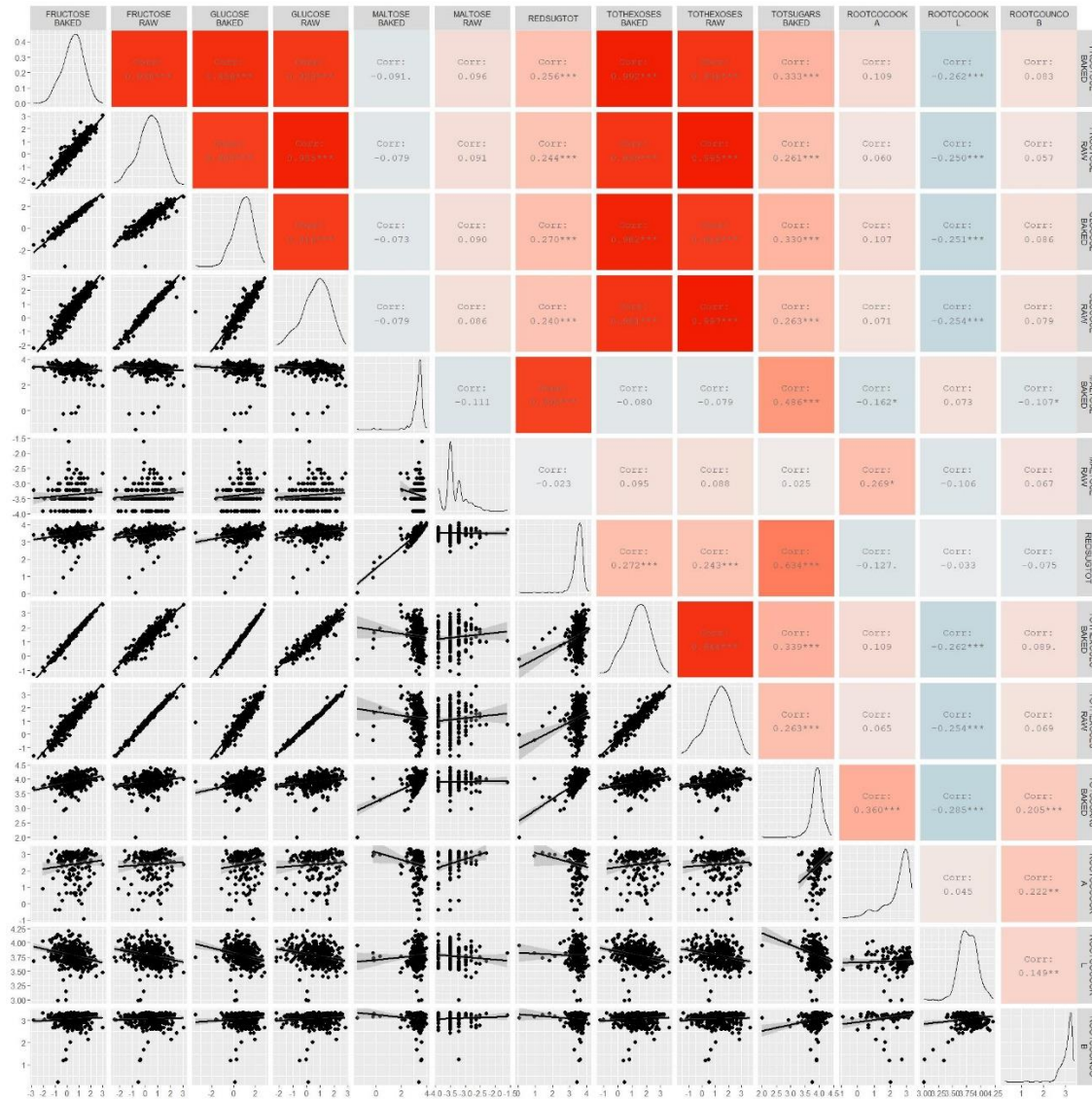


Fig. 1 Correlation plot depicting Pearson Correlation coefficients between culinary profile traits with significant candidate gene associations.

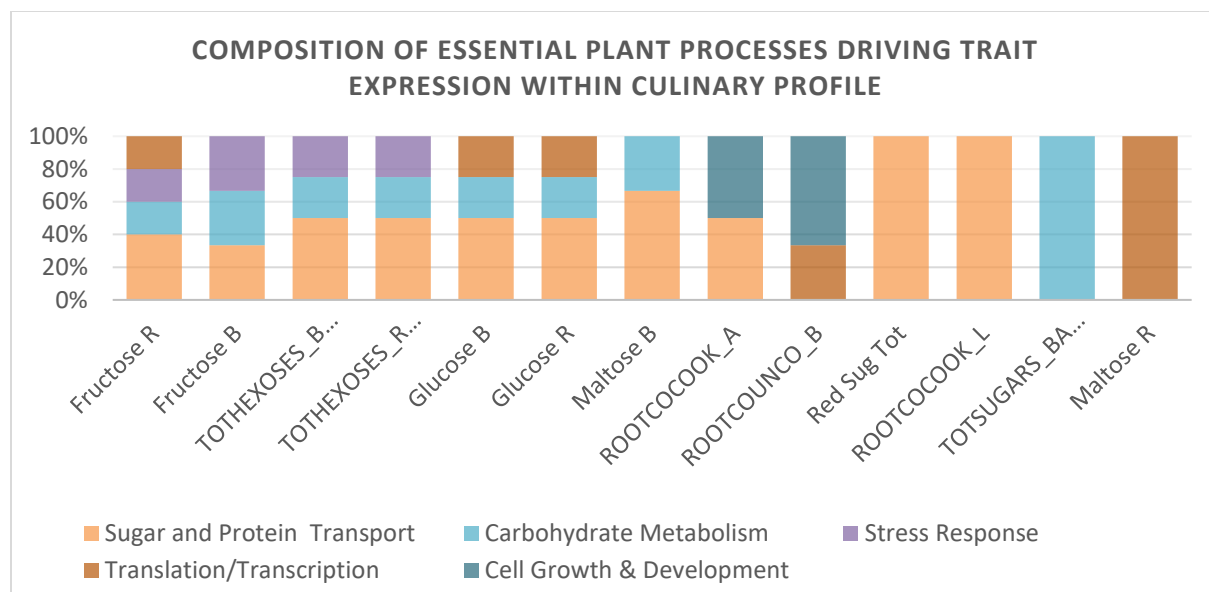


Fig. 2 Percent composition of essential plant processes: Carbohydrate metabolism, cell growth & development, stress response, and sugar and protein transport with candidate genes significantly associated with traits within the culinary profile of sweetpotato.

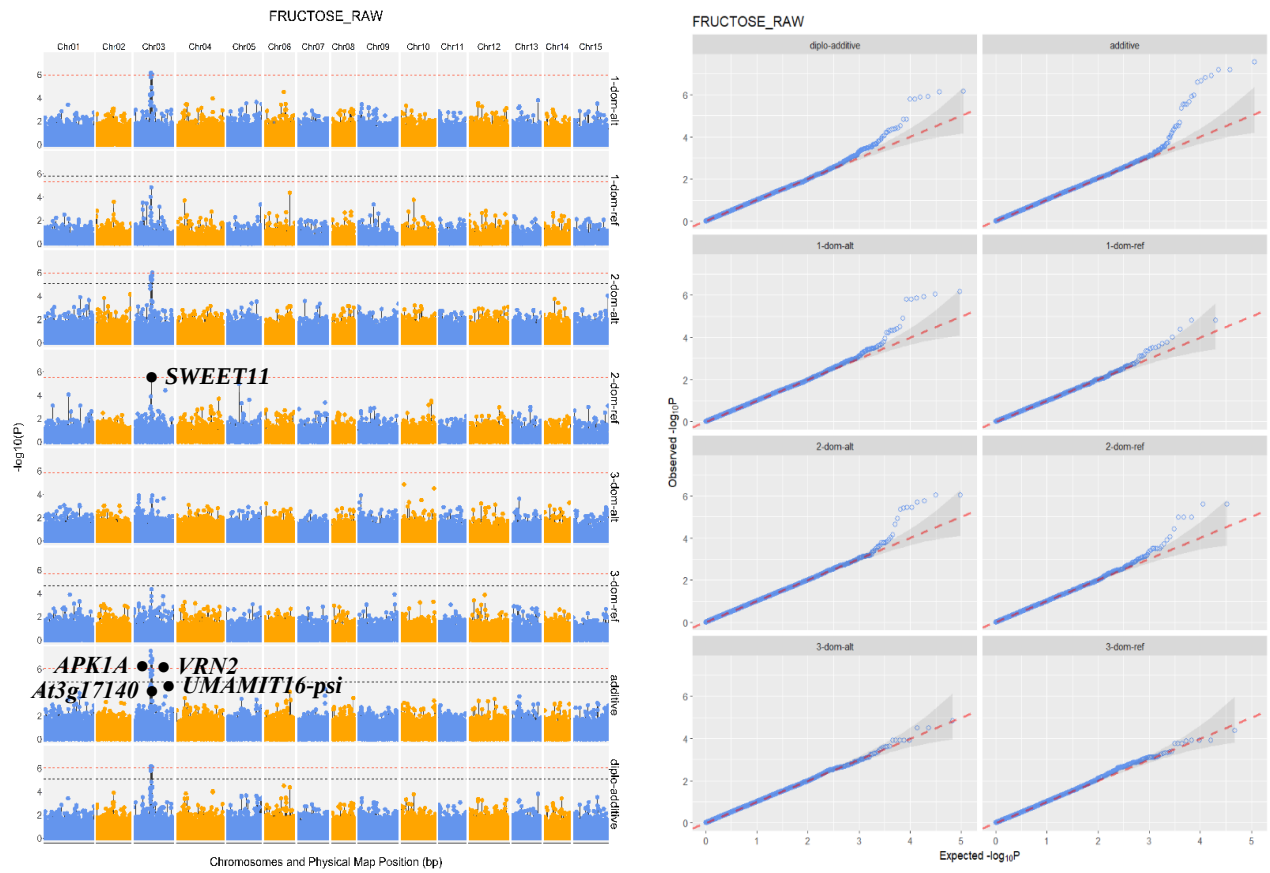


Fig. 3 Manhattan plot (left) showing all gene dosage models for SNP associations with fructose profile in raw sweetpotato storage roots. The five candidate genes with significant associations with raw fructose are: *APK1A*, *At3g17140*, *SWEET11*, *VRN2*, and *UMAMIT16-psi*. Q-Q plots (right) reveals the power of detection and extent of false positive rate for all gene dosage models.

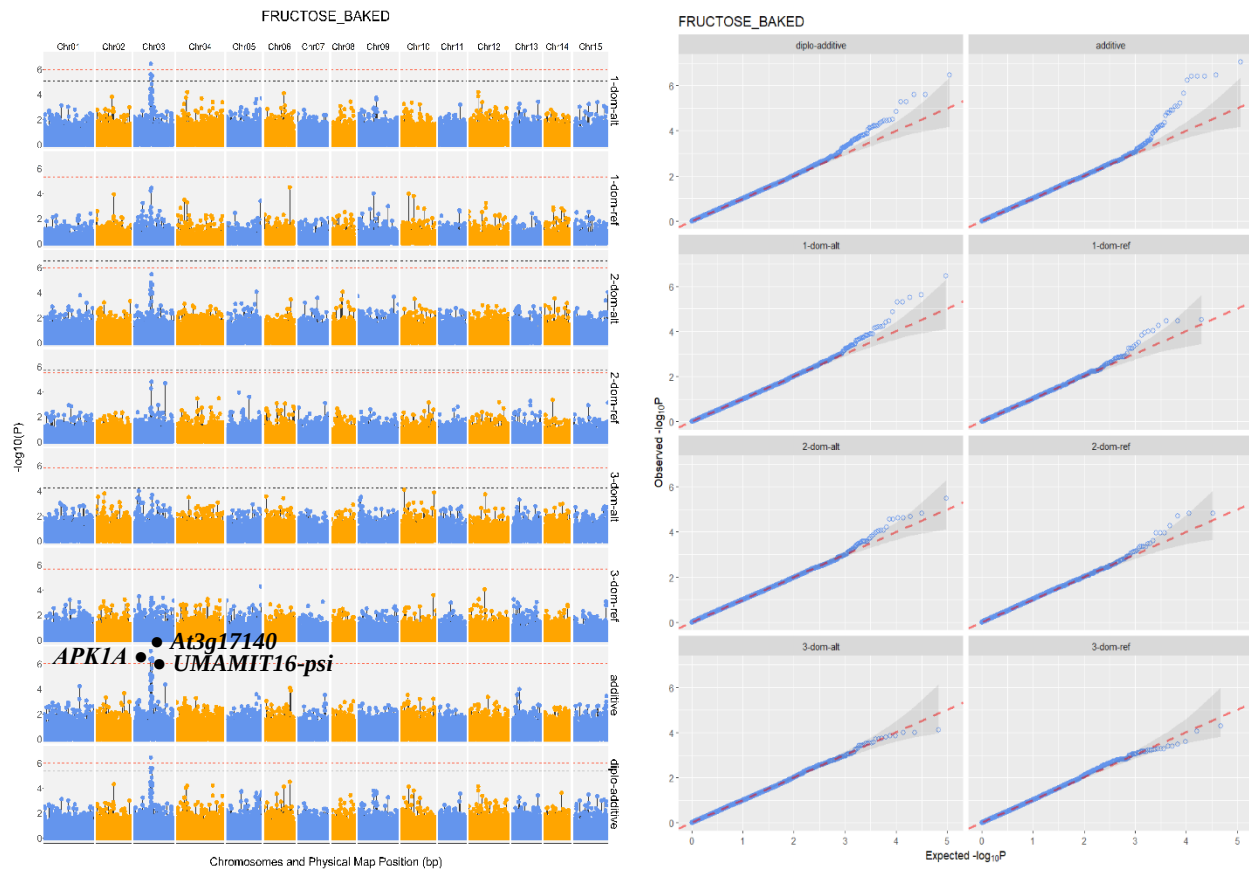


Fig. 4 Manhattan plot (left) showing all gene dosage models for SNP associations with fructose profile in baked sweetpotato storage roots. The three candidate genes with significant associations with baked fructose are: *APKIA*, *At3g17140*, and *UMAMIT16-psi*. Q-Q plots (right) reveals the power of detection and extent of false positive rate for all gene dosage models.

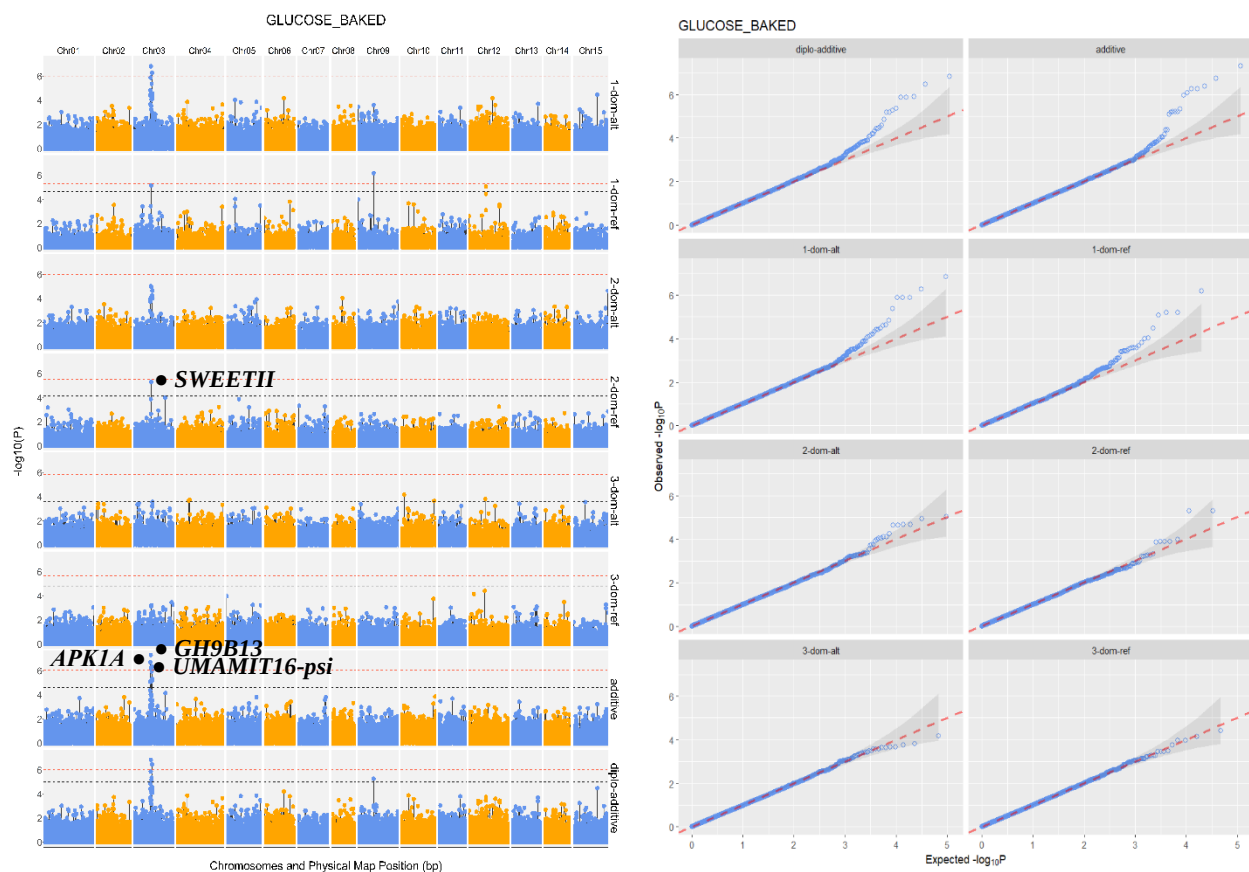


Fig. 5 Manhattan plot (left) showing all gene dosage models for SNP associations with glucose profile in baked sweetpotato storage roots. The four candidate genes with significant associations with baked glucose are: *APK1A*, *GH9B13*, *SWEETII*, and *UMAMIT16-psi*. Q-Q plots (right) reveals the power of detection and extent of false positive rate for all gene dosage models.

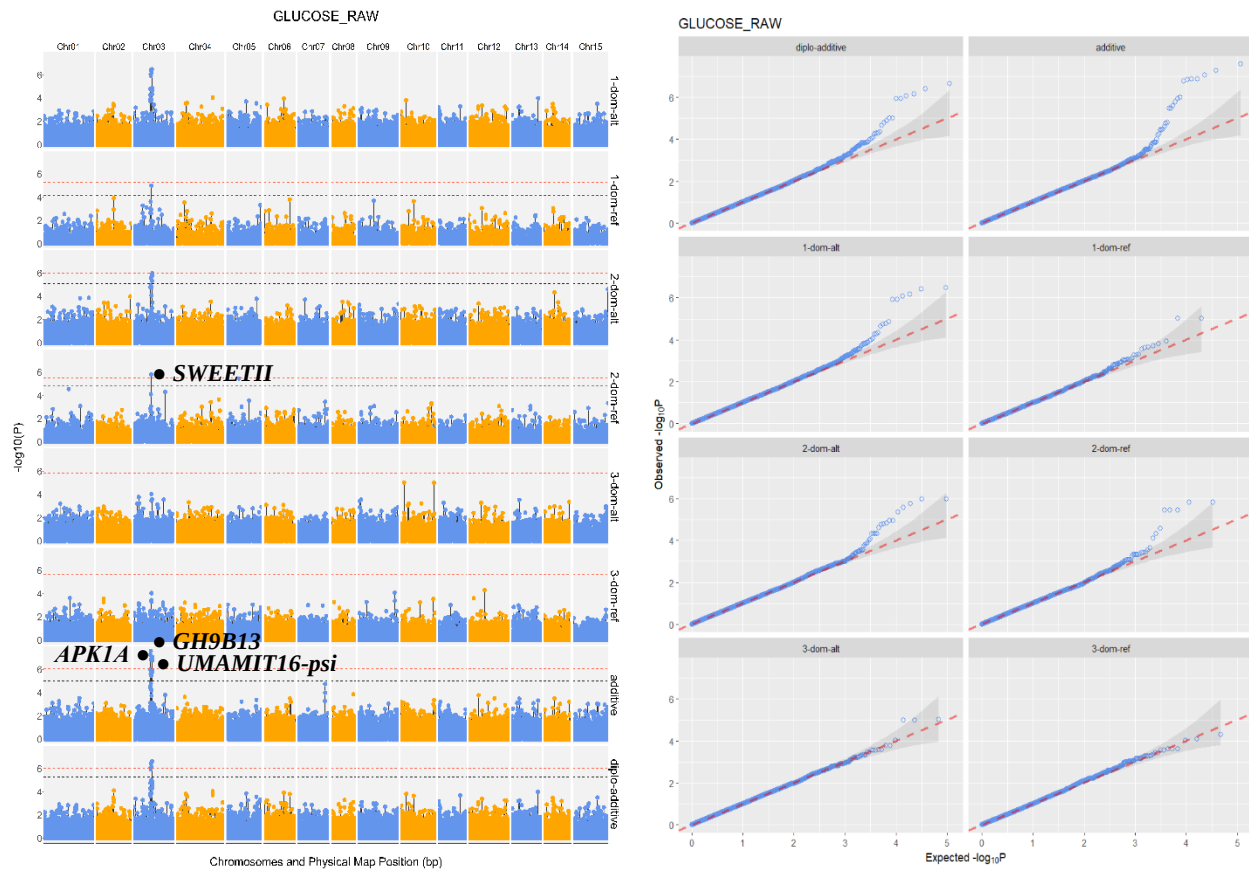


Fig. 6 Manhattan plot (left) showing all gene dosage models for SNP associations with glucose profile in raw sweetpotato storage roots. The four candidate genes with significant associations with raw glucose are: *APK1A*, *GH9B13*, *SWEETII*, and *UMAMIT16-psi*. Q-Q plots (right) reveals the power of detection and extent of false positive rate for all gene dosage models.

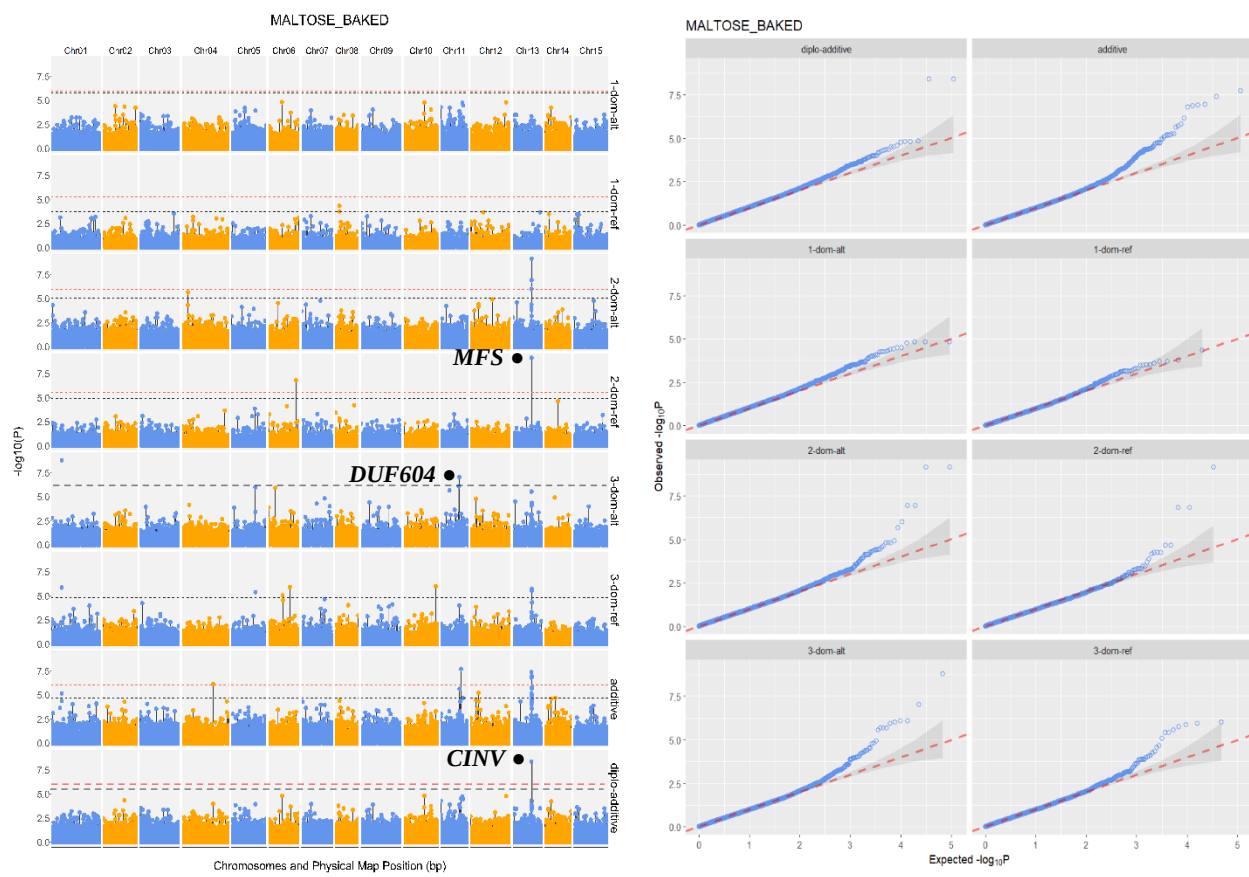


Fig. 7 Manhattan plot (left) showing all gene dosage models for SNP associations with maltose profile in baked sweetpotato storage roots. The three candidate genes with significant associations with baked maltose are: *DUF604*, *CINV*, and *MFS*. Q-Q plots (right) reveals the power of detection and extent of false positive rate for all gene dosage models.

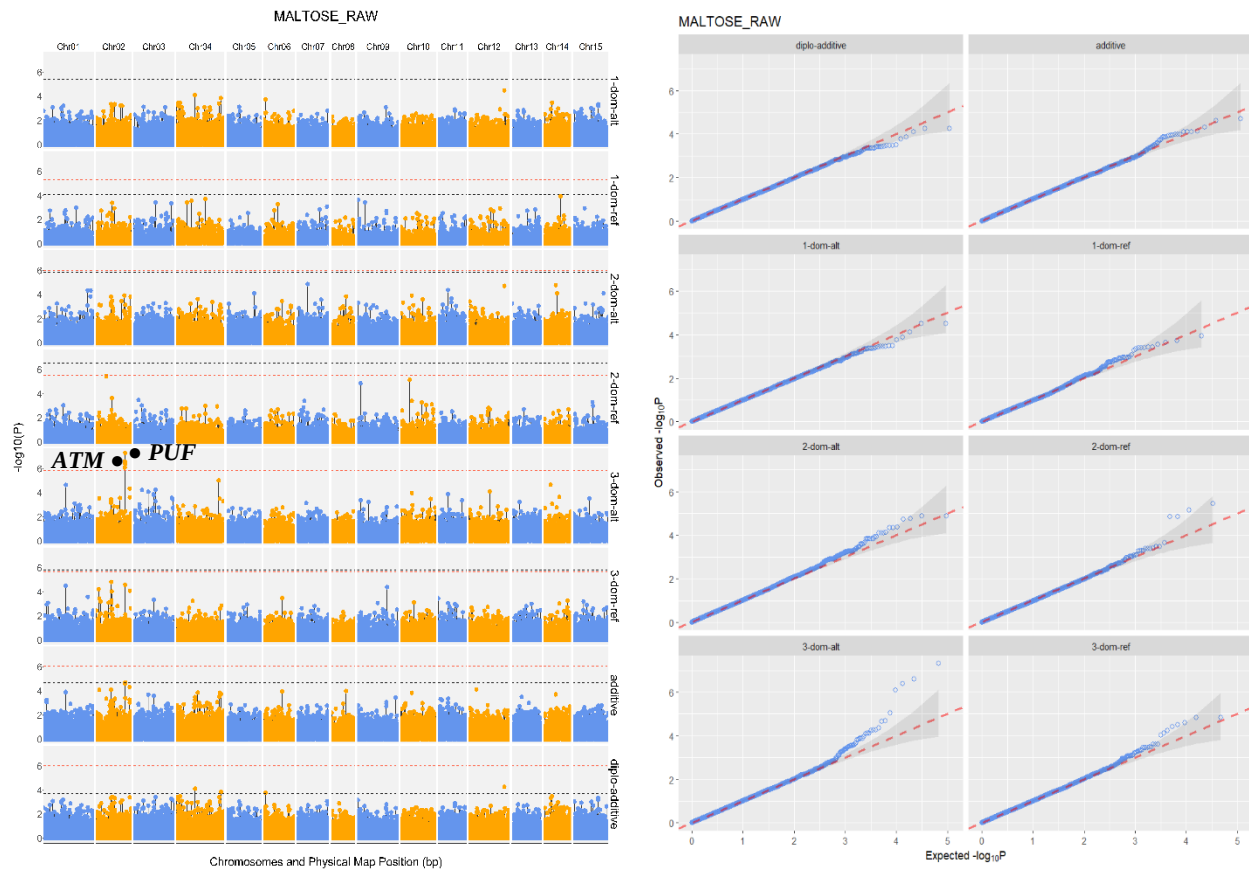


Fig. 8 Manhattan plot (left) showing all gene dosage models for SNP associations with maltose profile in raw sweetpotato storage roots. The two candidate genes with significant associations with raw maltose are: *ATM* and *PUF*. Q-Q plots (right) reveals the power of detection and extent of false positive rate for all gene dosage models.

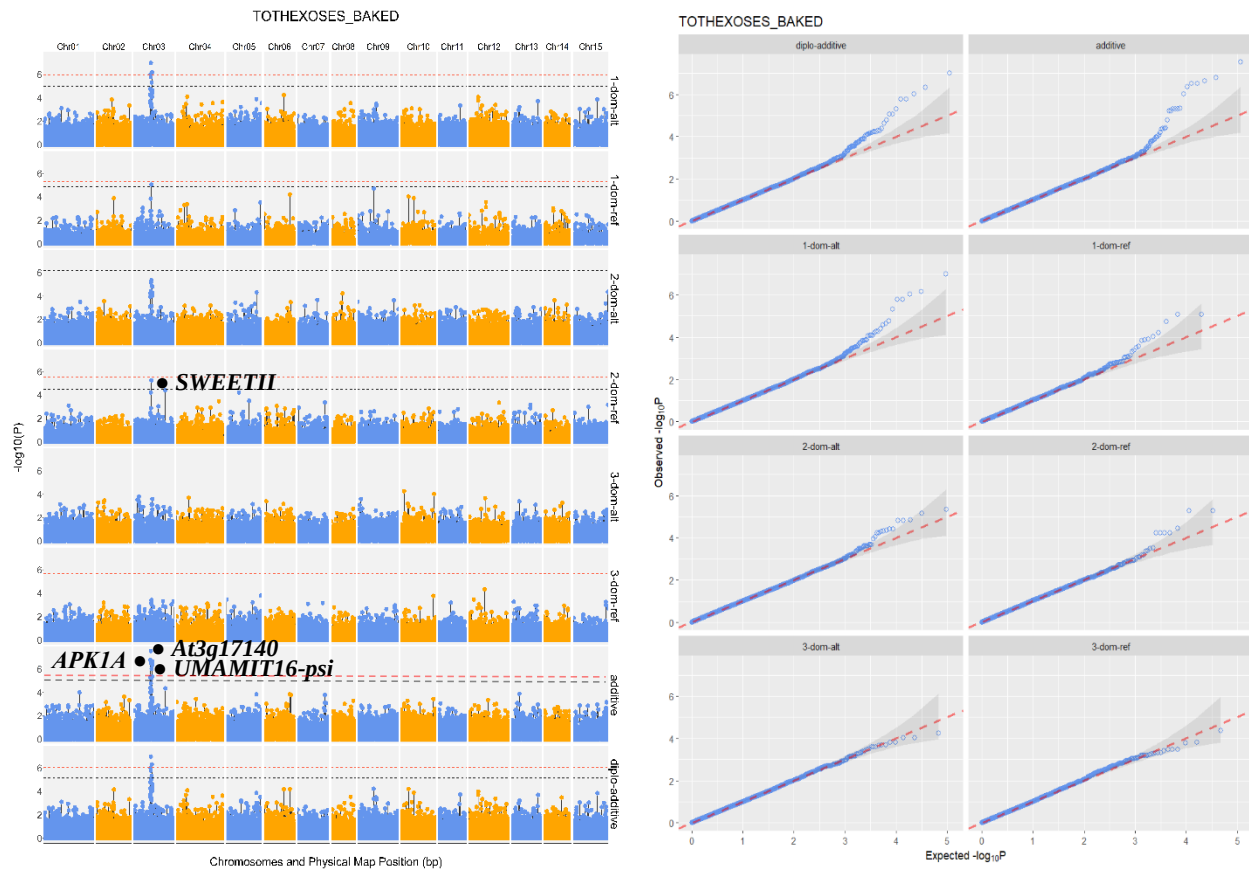


Fig. 9 Manhattan plot (left) showing all gene dosage models for SNP associations with the total hexoses profile in baked sweetpotato storage roots. The four candidate genes with significant associations with total hexoses baked are: *APK1A*, *At3g17140*, *SWEETII*, and *UMAMIT16-psi*. Q-Q plots (right) reveals the power of detection and extent of false positive rate for all gene dosage models.

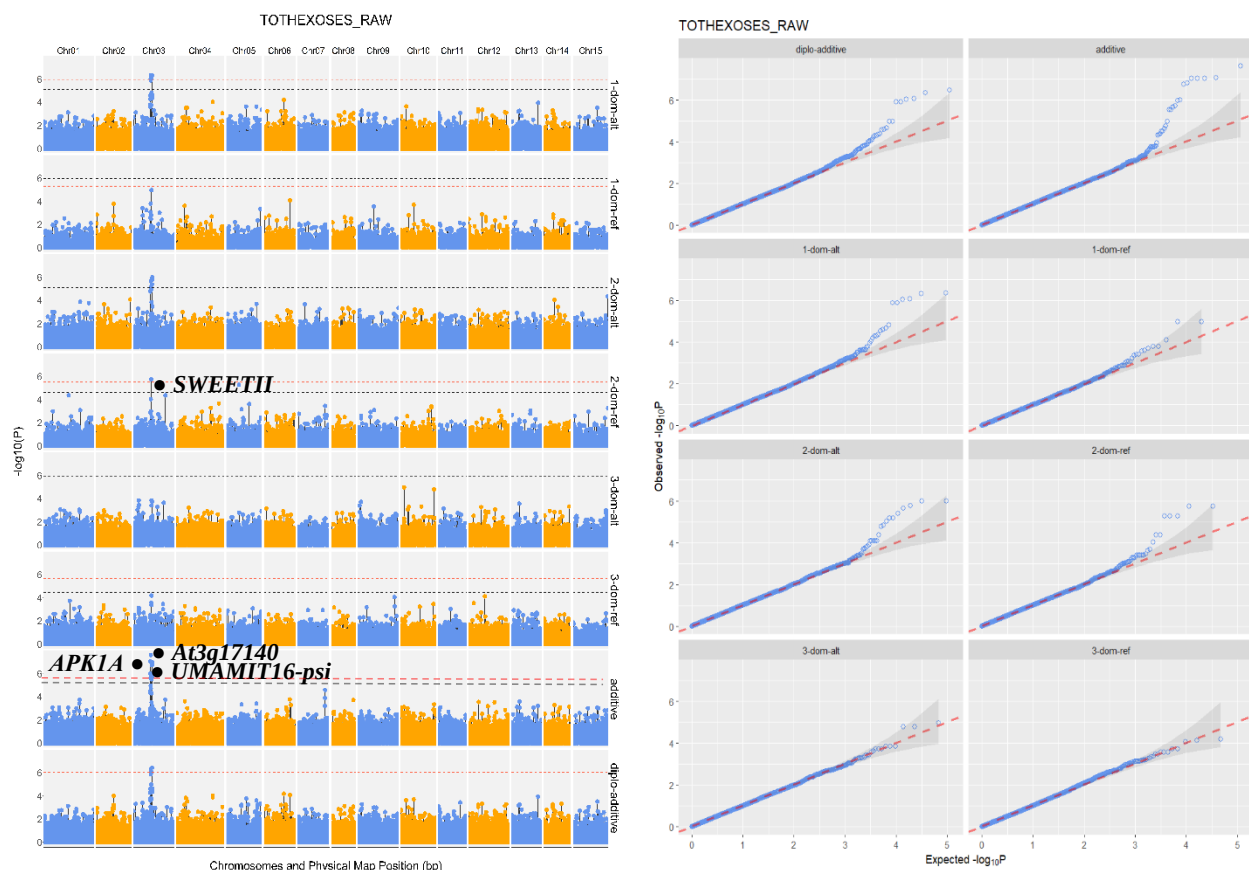


Fig. 10 Manhattan plot (left) showing all gene dosage models for SNP associations with the total hexoses profile in raw sweetpotato storage roots. The four candidate genes with significant associations with total hexoses raw are: *APK1A*, *At3g17140*, *SWEET11*, and *UMAMIT16-psi*. Q-Q plots (right) reveals the power of detection and extent of false positive rate for all gene dosage models.

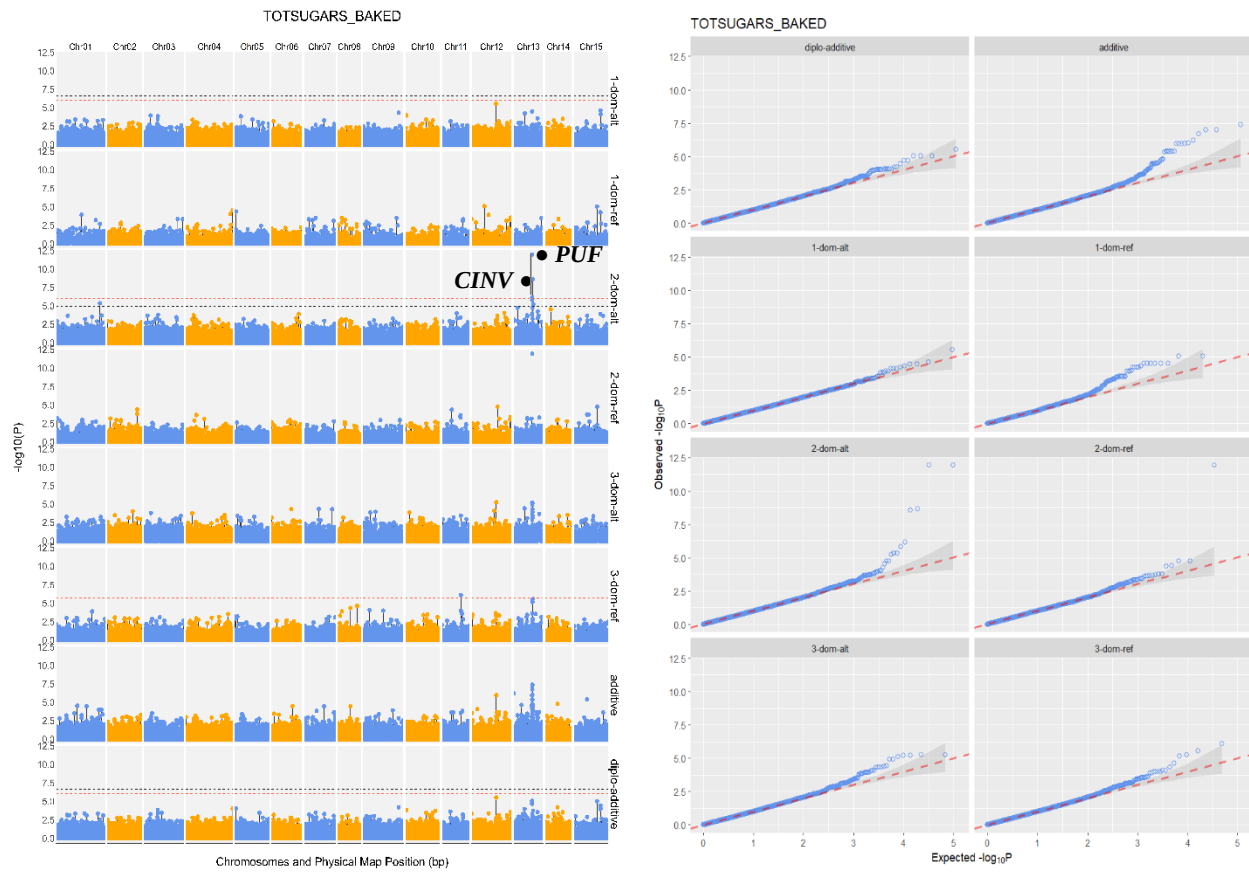


Fig. 11 Manhattan plot (left) showing all gene dosage models for SNP associations with the total sugars profile in baked sweetpotato storage roots. The two candidate genes with significant associations total sugars baked are: *CINV* and *PUF*. Q-Q plots (right) reveals the power of detection and extent of false positive rate for all gene dosage models.

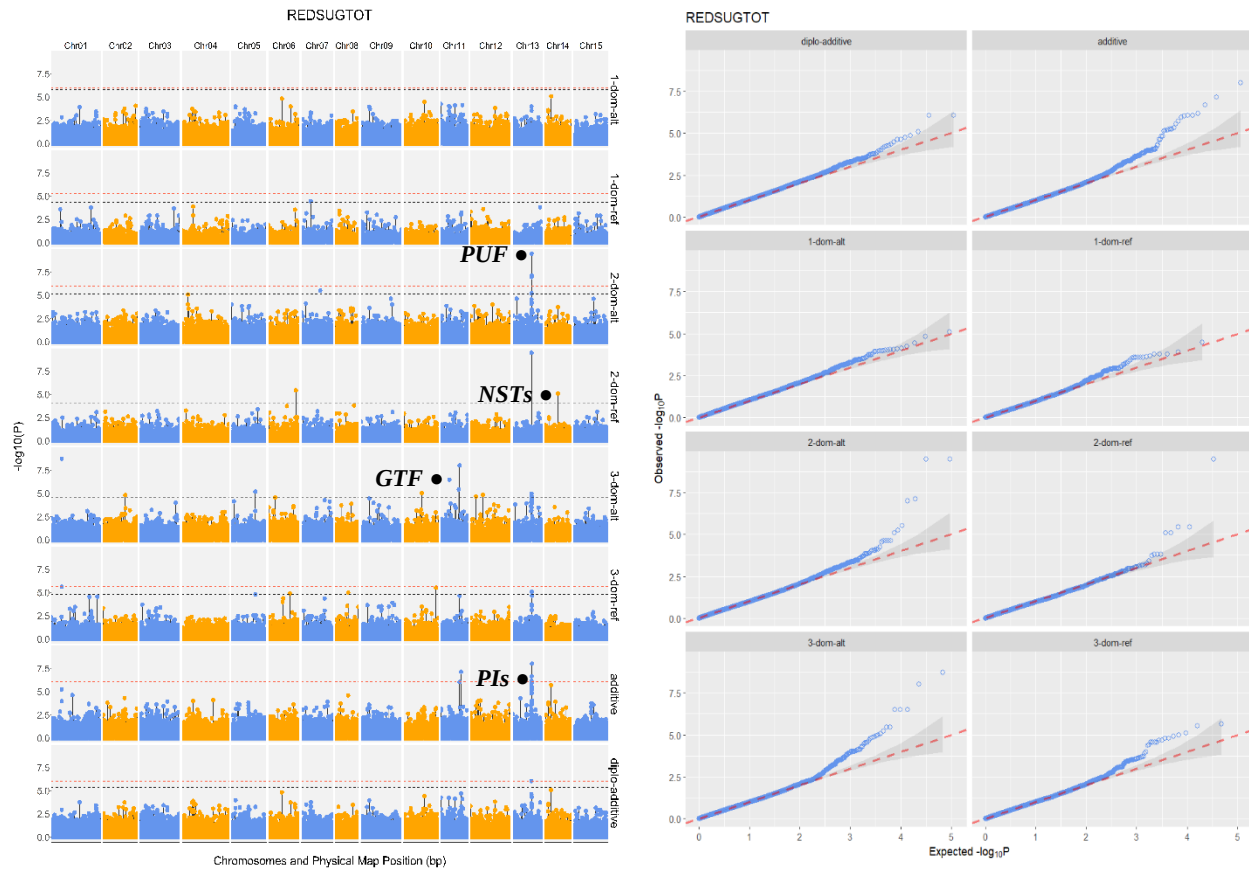


Fig. 12 Manhattan plot (left) showing all gene dosage models for SNP associations with total reducing sugars profile in sweetpotato storage roots. The four candidate genes with significant associations with total reducing sugars are: *GTF*, *PIs*, *PUF*, and *NSTs*. Q-Q plots (right) reveals the power of detection and extent of false positive rate for all gene dosage models.

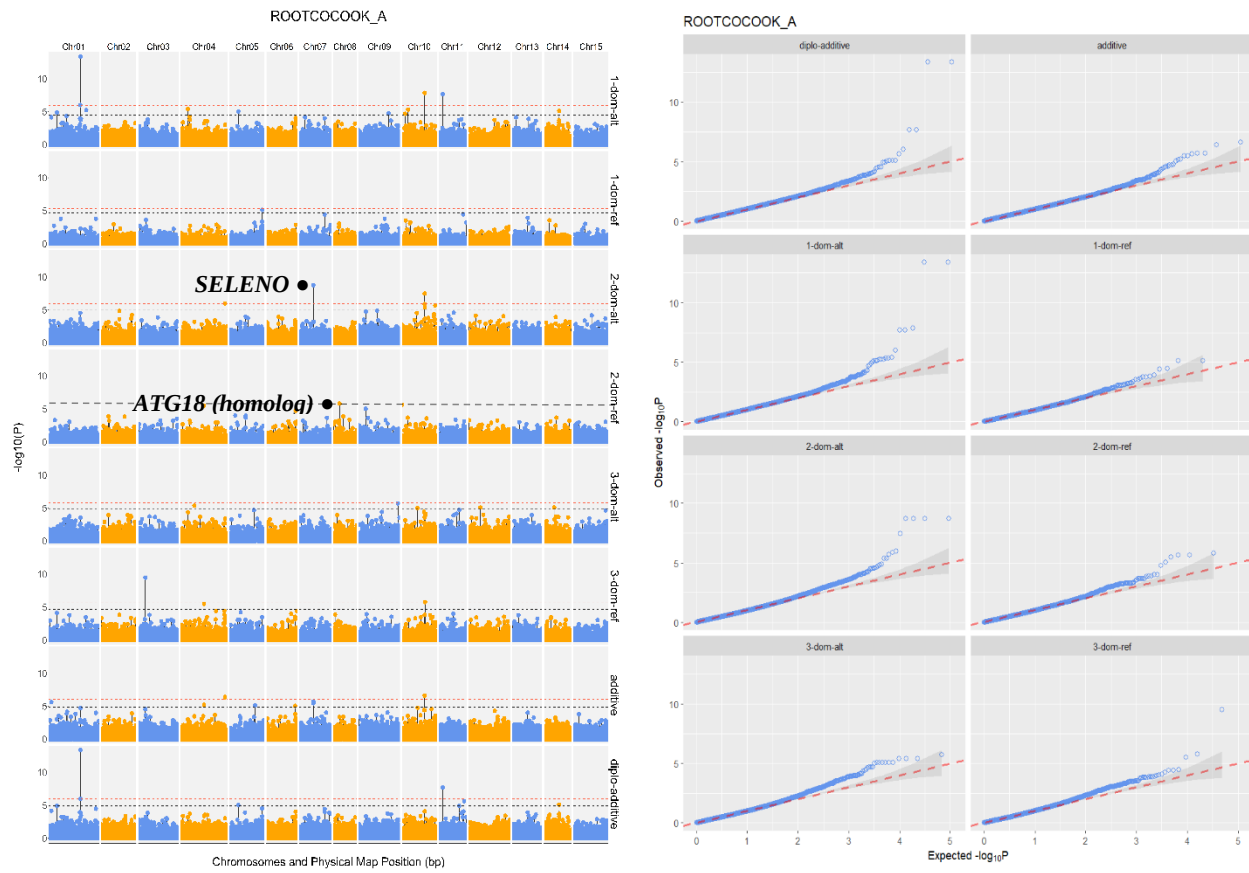


Fig. 13 Manhattan plot (left) showing all gene dosage models for SNP associations with the cooked root color *A in sweetpotato storage roots. The two candidate genes with significant associations with cooked root color *A are: *SELENO* and *ATG18 (homolog)*. Q-Q plots (right) reveals the power of detection and extent of false positive rate for all gene dosage models.

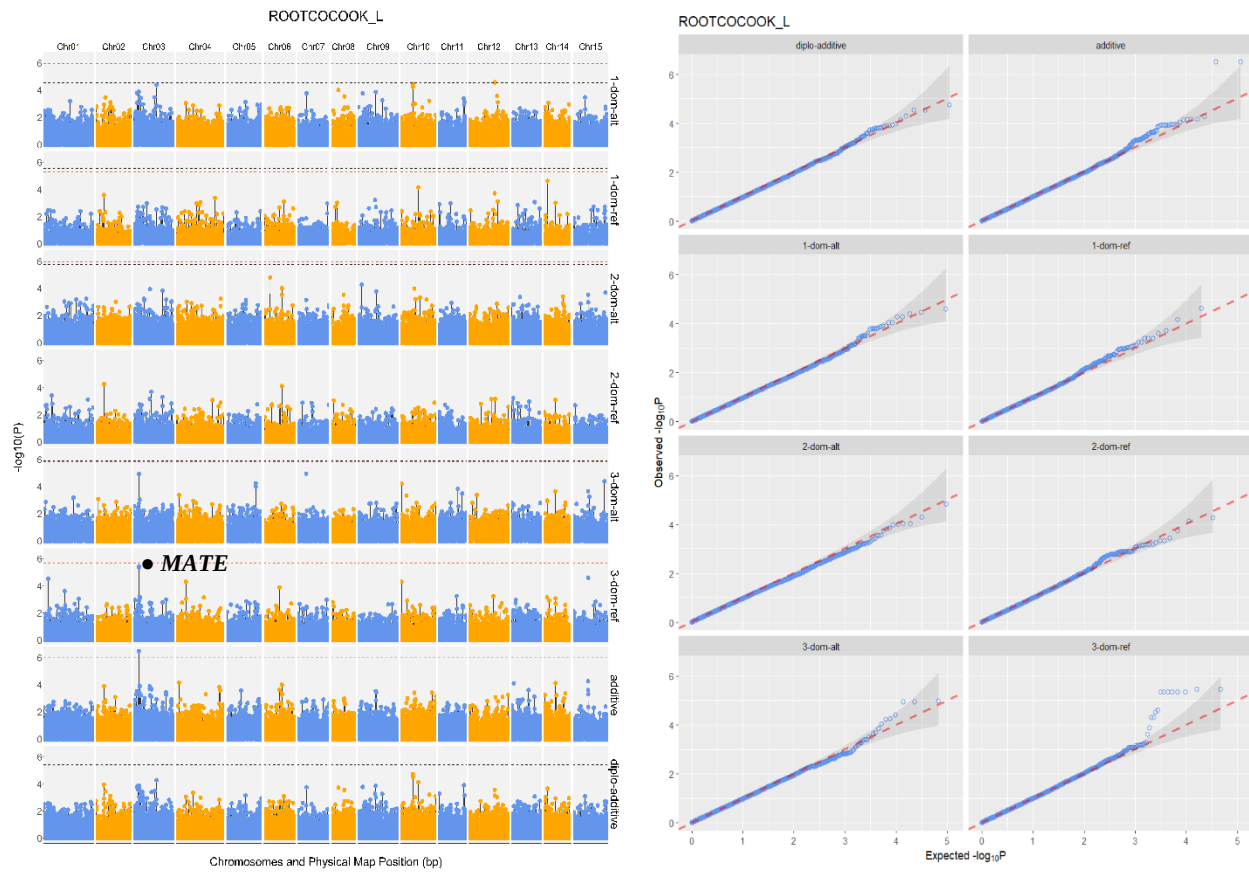


Fig. 14 Manhattan plot (left) showing all gene dosage models for SNP associations with the cooked root color *L in sweetpotato storage roots. The one candidate gene with a significant association with cooked root color *L is: *MATE*. Q-Q plots (right) reveals the power of detection and extent of false positive rate for all gene dosage models.

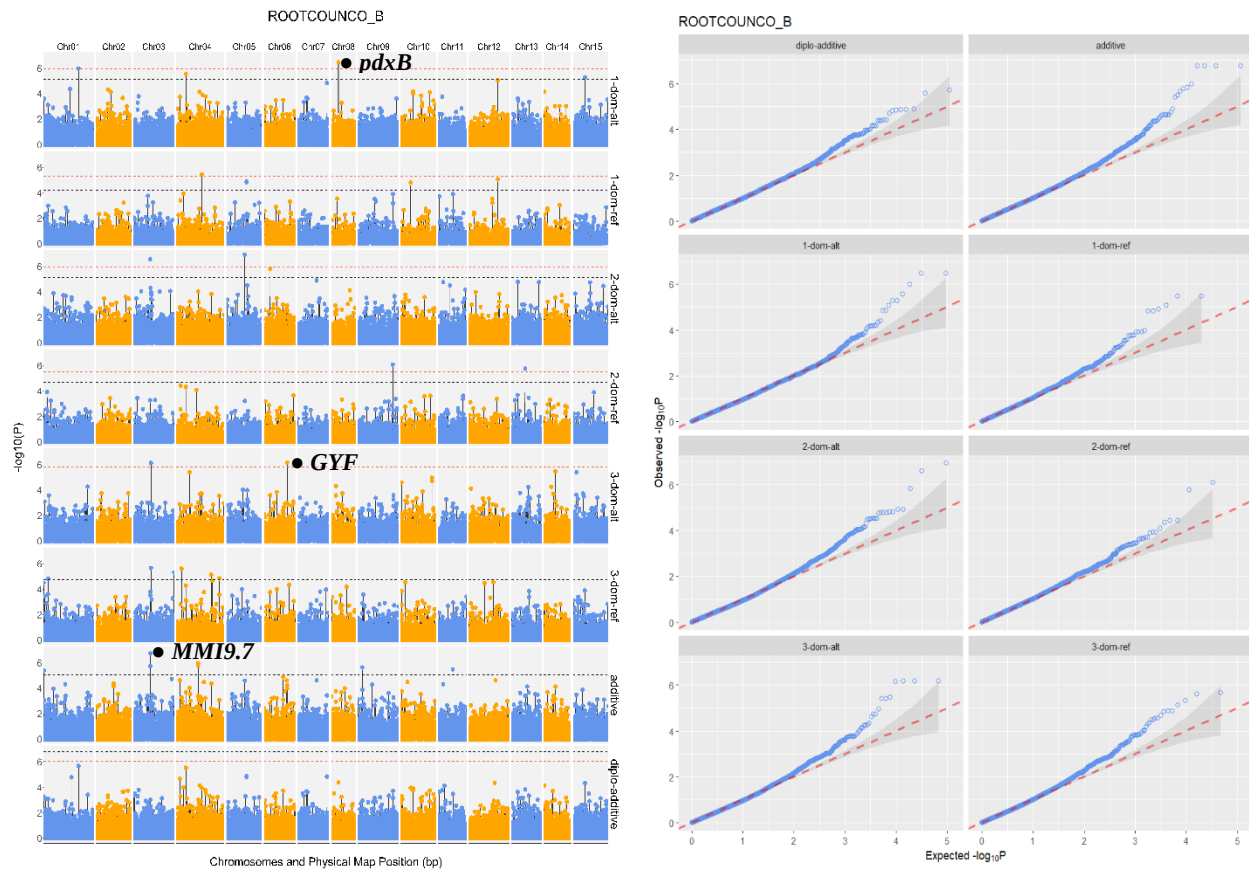


Fig. 15 Manhattan plot (left) showing all gene dosage models for SNP associations with the uncooked root color *B in sweetpotato storage roots. The three candidate genes with significant associations with uncooked root color *B are: *MMI9.7*, *GYF*, and *pdxB*. Q-Q plots (right) reveals the power of detection and extent of false positive rate for all gene dosage models.

VITA

Tara E. Rickman was born in Mississippi in January of 1996. She attended Millsaps College in Jackson, MS and graduated with a Bachelor's of Science in Biology in December of 2017. During her time at Millsaps, she participated in research projects at Millsaps College, Loyola University Chicago, and North Carolina State University. She attended the University of Tennessee Knoxville from 2018-2020 where she graduated with a Master's of Science in Entomology and Plant Pathology in December of 2020.