COMPARATIVE RNA-SEQ ANALYSIS OF PHENOTYPICALLY DIFFERENT SWEET POTATOES (IPOMOEA BATATAS [L.] LAM)

by

ELIZABETH FIEDLER

A THESIS

Submitted in partial fulfillment of the requirements for the degree of Master of Science in the Agricultural Sciences Graduate Program
of Delaware State University

DOVER, DELAWARE
May 2019

This Thesis is approved by the following members of the Final Oral Review Committee:

Dr. Venu (Kal) Kalavacharla, Committee Chairperson, Department of Agriculture and Natural Resources, Delaware State University
Dr. Richard Barczewski, Committee Member, Department of Agriculture and Natural Resources, Delaware State University
Dr. Marikis Alvarez, Committee Member, Department of Agriculture and Natural Resources, Delaware State University
Dr. Muthusamy Manoharan, External Committee Member, Department of Agriculture, Fisheries and Human Sciences, University of Arkansas at Pine Bluff

Comparative RNA-Seq Analysis of Phenotypically Different Sweet Potatoes (Ipomoea batatas [L.] Lam)

Elizabeth Fiedler

Faculty Advisor: Dr. Venu (Kal) Kalavacharla

ABSTRACT

Sweet potato is arguably one of Earth's top ten most important crops. It is relatively low maintenance, packed with essential vitamins and nutrients, and in addition to serving as an effective food crop, it has been suggested for use as a material for synthesizing plastics and as a replacement for corn as a source for bioethanol production. Sweet potatoes are difficult to bring to seed, so most sweet potato plants are grown from slips, which are cuttings from sweet potato vines. This makes it very easy for sweet potato viruses to spread from generation to generation. Currently, virus disease complexes, which are infections of two or more viruses with a synergistic interaction, pose the biggest threat to sweet potato yields. It is therefore crucial to have a better understanding of different sweet potato genotypes. Sweet potato genotypes tend to be very similar, since crossing and outbreeding strategies are all but impossible to carry out on sweet potatoes, as different genotypes tend to be very similar. Transcriptome profiling could prove to be an extremely useful method for understanding the causes of phenotypic differences in genotypes. This project details the construction of the transcriptome profiles of three phenotypically different genotypes. Illumina sequencing was carried out on the 2500HiSeq platform and the resulting reads were aligned to the sweet potato genome constructed by the Max Planck Institute. Bioinformatics software was used to carry out alignment and analysis. The analysis techniques were used to calculate and normalize gene expression, make fold change

comparisons, convert this data to log 2-fold change, determine which genes had the greatest expression differences, group these genes into functional groups, and perform other analyses.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGIURES

LIST OF ABBREVIATIONS

SPVD	Sweet Potato Virus Disease
SPCSV	Sweet Potato Chlorotic Stunt Virus
SPFMV	Sweet Potato Feathery Motte Virus
SPMMV	Sweet Potato Mild Mottle Virus
SPLCV	Sweet Potato Leaf Curl Virus
SPVG	Sweet Potato Virus G
SPLV	Sweet Potato Latent Virus
SPV2	Sweet Potato Virus 2
SPMSV	Sweet Potato Mild Speckling Virus
NARO/KARC	Kyushu Okinawa Agricultural Research Center, National Agriculture and Food Research Organization
SNP	Single Nucleotide Polymorphism
CNV	Copy Number Variation
SSR	Simple Sequence Repeat
NGS	Next Generation Sequencing
PCR	Polymerase Chain Reaction
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
SORC	Smyrna Outreach & Research Center
NCBI	National Center for Biotechnology Information
DBI	Delaware Biotechnology Institute
RIN	RNA Integrity Number
TPM	Transcripts Per Million

CHAPTER I: INTRODUCTION

The third most important root crop in the World, sweet potato, is preceded only by potato (*Solanum tuberosum*) and cassava (*Manihot esculenta*) (Valverde, Clark, & Valkonen, 2007). It is native to tropical and sub-tropical Middle and South America. Over 130 million tons of sweet potatoes are produced annually worldwide (Crop Trust, 2018).

Sweet potato is an essential crop to people from all walks of life. It is often referred to as a strategic crop because it is key to maintaining a stable food supply and also aids in the acquisition of proper nutrition for poverty-stricken individuals in developing countries. In addition to providing a stable food supply to subsistence farmers, sweet potato storage roots are packed with vitamin A and beta-carotene, which are essential nutrients for developing children and nursing mothers for combating malnutrition, which can incidentally cause blindness and other childhood diseases (Helen Keller International, 2018). Sweet potatoes are an especially convenient and economical solution for tackling the prevalence of malnutrition in developing countries of Africa and South America as many of these locations have ideal climates for year-round sweet potato cultivation (Helen Keller International, 2018). In addition to consuming storage roots, humans are also able to consume sweet potato leaves, which may also confront nutritional gaps in their regular diets (Helen Keller International, 2018).

The sweet potato plant has the potential to serve as so much more than just a food source for humans. The sweet potato foliage is a byproduct that may also be fed to farm animals and provide highly nutritious feed (UNCTAD, 2012). It has also been argued that sweet potato, because of its extensive root system, would be highly effective in preventing soil erosion due to both water and wind (UNCTAD, 2012). Since the actual sweet potato plant yields an enormous amount of biomass, this is a material that has been proposed to be used for producing plastics

(Ziska, et al., 2009) (Klass, 2004). Now more than ever, it is necessary to find ways to offset the production of biofuels from corn, as nearly a third of all corn produced in the U. S. is used for biofuel production and the inability to meet the demand for corn as a feedstock is now becoming a legitimate concern (Collins & Erickson, 2011). Sweet potato has been proposed as a potential source for manufacturing biofuel. Ziska et al. (2009) determined that higher amounts of carbohydrates could be produced in sweet potato than in corn per hectare, making sweet potato a much more desirable candidate for bioethanol production. A bonus feature for ethanol production is sweet potato has a relatively low lignin content for the amount of biomass that is yielded (Wang, et al., 2015).

Due to the sweet potato's potential as a stable food supply, a nutritious feedstock, and a source for biofuel production, it is important to develop superior breeding lines that will be most profitable to farmers and growers for whichever use they are growing. This will involve finding genes linked to desirable phenotypic qualities such as high yield, different types of abiotic stress resistance, broad disease resistance, high carbohydrate production, low lignin levels, desirable aesthetic features of the crop, and suitable nutrition. Viral disease complexes occur when two or more sweet potato viruses infect a single plant, and are currently one of the greatest threats to cultivated sweet potatoes. A plant infected with only one sweet potato virus usually is not significantly affected and often is not noticed until it is infected with another virus and a disease complex is formed (Ngailo, Shimelis, Sibiya, & Mtunda, 2013). For example, the main concern of sweet potato farmers is the Sweet Potato Virus Disease Complex (SPVD) which can cause up to 98% yield loss of storage roots (Ngailo, Shimelis, Sibiya, & Mtunda, 2013). SPVD occurs when a plant is infected with both Sweet Potato Chlorotic Stunt Virus (SPCSV), a crinivirus, and

Sweet Potato Feathery Mottle Virus (SPFMV), a potyvirus, and a symbiotic relationship is established between the two diseases (Ngailo, Shimelis, Sibiya, & Mtunda, 2013).

Here, we try to identify differences in gene expression of three distinct sweet potato genotypes. Since there is little genetic variation between sweet potato genotypes, genomic DNA sequencing would likely not reveal many differences related to phenotype. We stand to gain more information about the differences in gene expression at different levels that occur between these three genotypes. These genotypes have obvious phenotypic differences, as can be seen in **Figure 1**. The first genotype has leaves that have five pointed lobes and produces storage roots which are orange-fleshed. The second genotype has tri-lobal leaves and produces white-fleshed storage roots. Finally, the third genotype has both single-lobed, and trilobal leaves and produces purple-fleshed storage roots. It was also noticed that the leaves of the purple-fleshed genotype are considerably darker than the other two genotypes. Also, it should be noted, the purple-fleshed genotype produced considerably larger storage roots on average than the white-fleshed and the yellow-fleshed genotypes. There were also differences in flowering time, the white-fleshed plants produced a considerable amount of flowers fairly early and did not stop producing these flowers as the season went on, whereas the purple-fleshed plants did not produce any flowers, even though they had the largest biomass, until the very end of the season.

An overarching goal for this project is the need to understand the underlying cause of phenotypic variation. We needed to examine differences in gene expression between these three distinct genotypes that not only influence the observable characteristics, but also non-immediately noticed characteristics, such as biotic and abiotic stress resistance. It was also important to highlight gene expression differences involved in metabolic pathways, disease resistance, stress resistance, etc.

In order to gain a comprehensive look into differences in gene expression in these three distinct sweet potato genotypes, transcriptome profiles were created from the leaf tissue of each of these three genotypes. The profiles were then analyzed using the software program CLC Bio. We predict that since we did observe such extreme differences in pigments within the storage roots, more gene expression differences having to do with pigments and the manufacturing of pigments will be observed in this study. Since the purple sweet potatoes and the purple sweet potato plants were considerably larger than the others were, it is reasonable to assume that more genes would be up-regulated in the purple sweet potato that directly influence growth and storage root initiation.

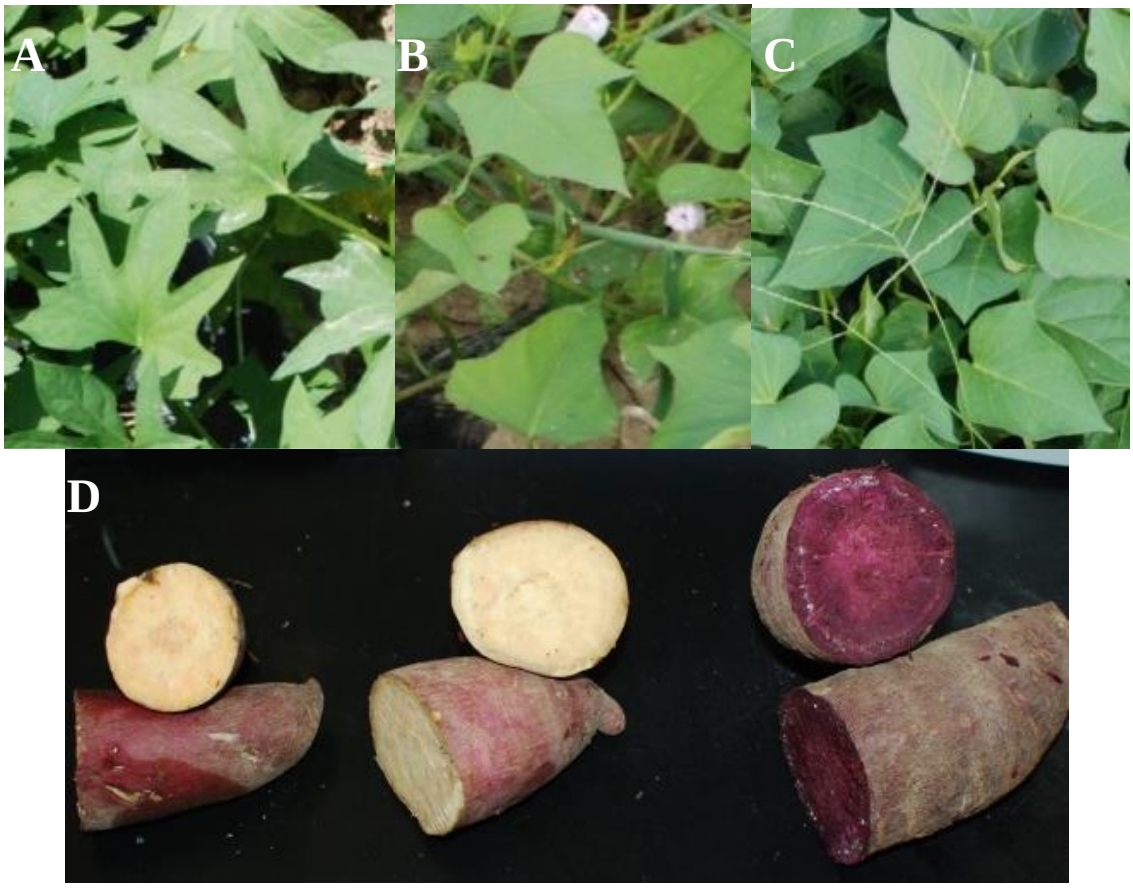


Figure 1: Three Distinct Sweet Potato Genotypes. Three distinct genotypes had their RNA sequences profiled in this experiment. **A**. Genotype #1 had skinny five-lobed leaf structure. **B**. Genotype #2 was tri-lobal, with a main central lobe. **C**. Genotype #3 had leaves that started out with a single lobe and as the plants matured, the leaves would develop into tri-lobal leaves. **D**. The three storage roots that were prouced by these plants genotype #1 (orange-fleshed storage root), genotype #2 (white-fleshed storage root), and genotype #3 (purple-fleshed storage root).

CHAPTER II: LITERATURE REVIEW

2.1. ECONOMIC RELEVANCE

Although more than 130 million tons of sweet potatoes are produced annually worldwide, 82.3% of all global sweet potato production occurs in Asia (U.S. National Library of Medicine, 2018) (UNCTAD, 2012) (Crop Trust, 2018). In 2010, China alone produced 81.5 million tons of sweet potatoes, with no other country coming anywhere close to this production level (the next highest producers of sweet potato that year were Uganda and Nigeria, each at 2.8 million tons in 2010, which is ~3.4% of China's annual production) (UNCTAD, 2012). China's production has fallen a bit, as only ~70.5 million tons were produced in 2013 and the same year Nigeria's production grew to ~3.6 million tons (Nag, 2017). The United States only produces about 1% of the World's sweet potatoes annually, with the leading production states being North Carolina, Mississippi, California, Louisiana, and Florida (UNCTAD, 2012) (North Carolina Department of Agriculture and Consumer Reviews, 2014).

As previously mentioned, sweet potato can be utilized as a strategic crop to defend against famine and poor nutrition (Helen Keller International, 2018). It can also be utilized as a strategic crop in Florida, in response to an economic problem. The Florida Institute of Food and Agricultural Sciences has developed an industrial variety preferable to table varieties for the production of ethanol (Mussoline & Wilkie, 2016) (IFAS, 2016). This newly developed variety is able to grow and flourish on land that was previously utilized for citrus production, but has been decimated due to greening disease (IFAS, 2016). Not only are the storage roots of this crop a sustainable source of biofuel, but the foliage may be used to supplement animal feed (Mussoline & Wilkie, 2016) (IFAS, 2016). As sweet potato is a relatively easy crop to establish,

this is one answer to minimizing a citrus grower's lost income during a time of crisis and giving them plenty of time to recuperate their operations.

2.2. GENETICS OF SWEET POTATO

Sweet potato is a hexaploid that contains 90 chromosomes (2n=6X=90). The genome size of sweet potatoes was reported as 4.38-5.3 pg/2C by Ozias-Akins et al (1994), which is approximately 2200-3000Mb (haploid) (Yan, et al., 2015) (Isobe, Shirasawa, & Hirakawa, 2017). Contrary to popular belief, sweet potatoes are not yams. While yams are usually larger and have a rough, scaly outside, sweet potatoes contain much less starch and contain a higher amount of beta-carotene (North Carolina Sweet Potato Commission, 2018). There is a wide range of sweet potato varieties as the storage roots can have flesh and skin that are red, purple, white, orange, yellow, tan, and colors in between. Beauregard and Jewell are the most common varieties that are typically seen in North American supermarkets (Sweet Potato Varieties, 2016). Some examples of more exotic sweet potatoes include Carolina Ruby (dark red-purple skin and dark orange flesh), Porto Rico (rose-pink skin and mottled orange flesh), Cherokee (bright copper skin and orange flesh), Hannah (light tan skin and light yellow flesh), and White Delight (purplish-pink skin and white flesh) (Sweet Potato Varieties, 2016).

2.3. EVOLUTION

Sweet potato is thought to have originated from the region located between the Yucatan Peninsula and the mouth of the Orinoco River in Northern Venezuela (Austin, 1988) (McDonald & Austin, 1990). The oldest known domesticated sweet potatoes were discovered at an excavation site in the Casma Valley of Peru and dated to about 2,500 B.C. (Ugent & Peterson, 1988). After extensive genome mapping, a hypothesis was formed that indicated an ancient event in which diploid gametes of three different plants fused together during formation of the

gametes (polyploidization) in *Ipomoea trifida* (diploid), *Ipomoea triloba* (diploid), and *Ipomoea tabascana* (tetraploid) (Reddy, Bates, Ryan-Bohac, & Nimmakayala, 2007).

Yang et al. (2017) proposes that a sub genome specialization took place approximately 1.3 million years ago in which an ancestor of the modern day *Ipomoea trifida* and *Ipomoea batatas* "broke off" into two pathways. One of these pathways would continue along and become the modern day plant *Ipomoea trifida*, while the other experienced a natural duplication of its entire genome (approx. 800,000 years ago) (Yang, et al., 2017). Meanwhile, a cross would occur approximately 300,000 years after this first duplication, in which *Ipomoea trifida* would cross with the now genomically-duplicated *Ipomoea trifida* to create a hexaploid genotype, which eventually, through natural and human influences over the last couple thousand years, became the modern-day sweet potato (Yang, et al., 2017).

2.4. MOLECULAR MARKERS, TRANSCRIPTOME SEQUENCING, & GENOME SEQUENCING

The complex hexaploid genome structure of the sweet potato makes it extremely difficult to carry out any genetic and genomic work, so quite often, *Ipomoea trifida* is analyzed as a reference genotype (or model species), as it is diploid and the most likely ancestor of sweet potato (Isobe, Shirasawa, & Hirakawa, 2017). Nakayama et al. (2010) constructed an AFLP linkage map of an F_1 crossing of two lines of *Ipomoea trifida* (0431-1 and Mx23-4) in order to create a reference linkage map for sweet potato. One of these two original plants (Mx23-4), was developed into a self-compatible line at the Kyushu Okinawa Agricultural Research Center, National Agriculture and Food Research Organization (NARO/KARC) (Isobe, Shirasawa, & Hirakawa, 2017).

The eleventh generation of plants (S11) of this newly developed self-compatible line and the other line of *Ipomoea trifida* involved in the crossing for Nakayama's linkage map (0431-1)

were each mapped with the Illumina HiSeq2000 platform (Hirakawa, et al., 2015). Although these resulting genomes were not of the greatest quality, they did reveal a number of single nucleotide polymorphisms (SNPs) and copy number variations (CNVs) that will help in the analysis of sweet potatoes (Hirakawa, et al., 2015) (Isobe, Shirasawa, & Hirakawa, 2017). Whereas SNPs are variations that occur at a single nucleotide and are a more traditionally utilized molecular marker, CNVs are becoming more popular and more useful in the context of crop improvement and breeding technologies as they refer to relatively large genome regions that have been deleted or duplicated (Eurofins, 2018). These markers can act as indicators for desirable genes and assist plant breeders when screening potential plant material either for crossing, or for plants that will grow more successfully in different environmental settings (Hirakawa, et al., 2015).

Transcriptome sequencing on the Illumina HiSeq2000 platform has also been carried out on yet another *Ipomoea trifida* genotype by Cao et al (2016). These sequencing results allowed for researchers to design primers to screen for simple sequence repeats (SSRs) and clone the believed-to-be drought tolerance gene ItWRKY1, which was then verified as a drought tolerance gene by its transformation into tobacco plants (this gene came from *I. trifida*) (Cao, et al., 2016). The fact that this gene was proven to increase the survival rate of the stressed tobacco plants, means that growers and breeders can potentially screen for this gene in order to pick preferable plants for drier environments (Cao, et al., 2016).

Root transcriptomes of both *Ipomoea trifida* and cultivated sweet potato have been created and comparatively analyzed by Ponniah et al. (2017) with the Illumina sequencing platform. Since *Ipomoea trifida* does not develop storage roots, even though it is considered the closest related ancestor to the sweet potato, it is probable that analysis of differing gene

expressions could indicate the crucial molecular mechanisms involved in sweet potato storage root formation. Among the many differences in gene expression between these two plants, the most significant are arguably the genes that were expressed in *Ib* and not in *It*, which included the transcription factor RKD, the RGL family proteins, the potassium efflux antiporter (KEA5), and the ERECTA protein kinase (Ponniah, Thimmapuram, Bhide, Kalavacharla, & Manoharan, 2017). All of these proteins have been previously linked to or have something to do with root proliferation, manufacturing of cell wall polysaccharides, starch formation, nitrogen regulation, phytohormonal regulation, etc. (Ponniah, Thimmapuram, Bhide, Kalavacharla, & Manoharan, 2017).

Transcriptome sequencing has also been utilized for determining genes that influence carotene biosynthesis in sweet potato storage roots (Li, et al., 2015). In this study, the transcriptome profiles of an orange-fleshed sweet potato cultivar and its mutant were compared in order to observe which genes were up-regulated (Li, et al., 2015). Illumina sequencing was carried out, followed by *De Novo* sequence assembly. As the mutant was much higher in carotenoid content, it was reasoned that genes, which were up-regulated in the mutant, were possibly directly influencing the manufacturing of carotenoids in the storage roots (Li, et al., 2015). From this research, 1,725 SSR markers were also discovered (Li, et al., 2015).

Next Generation Sequencing Technologies (NGS) allow for large-scale sequencing at dramatically lower costs allowing for much more complex organisms to sequenced, such as those with higher ploidy levels (Isobe, Shirasawa, & Hirakawa, 2017). In 2010, over 500,000 transcript sequence reads were created from stem and leaf cDNA collections in drought-stressed sweet potato plants with the 454 Pyrosequencing platform (Schafleitner, et al., 2010). From this

work, a number of unique genes were annotated through BLASTX and over 1600 microsatellite markers were identified by the researchers involved (Schafleitner, et al., 2010).

De novo sequencing on a large scale has also been carried out on the transcriptome of a purple sweet potato variety and utilized for the identification of genes involved in biosynthesis pathways, unigenes, and protein coding genes (Xie, et al., 2012). From this sequencing project, over 50,000 unigenes were identified along with 851 simple sequence repeats (SSRs) which, if they are within close proximity to a desired sequence, could potentially be used as markers for screening for desired genes (Xie, et al., 2012).

Genome sequencing was also recently performed on *in vitro* cultured sweet potato plants (Yang, et al., 2017). Total genomic DNA was extracted from Taizhong6, a sweet potato cultivar high in carotenoids (accession number 2013003), and sequenced on four different platforms; HiSeq2500, Nextseq500, Hiseq4000, and GS FLX+ (Yang, et al., 2017) (Illumina, 2018) (Illumina, 2018) (Illumina, 2018) (Roche, 2011). This work is crucial to scientists that investigate polyploid organisms. This paper details the procedures used to investigate the sweet potato genomic structure that do not involve working with another, simpler related species model from the *Ipomoea* genus (Yang, et al., 2017).

2.5. IMPORTANT DISEASES/PESTS & CONTROL

Although sweet potatoes have a variety of different diseases and pests that could significantly impact yield in a growing season (i.e. clearwing moth, sweet potato stemborer, tortoiseshell beetles, armyworms, grasshoppers, aphids, whiteflies, nematodes, anthracnose, root rot, Fusarium wilt, bacterial wilt etc.), viral diseases tend to remain the main cause of concern for growers. This is primarily because viruses are not as immediately detected, they have relatively fast modes of infection, and there is currently no treatment available to the farmers and growers

once they know that their crop is infected (Ames, Smit, Braun, O'Sullivan, & Skoglund, 1997) (Ngailo, Shimelis, Sibiya, & Mtunda, Sweet potato breeding for resistance to sweet potato virus disease and improved yield: Progress and challenges, 2013). As of 2014, more than 30 different sweet potato viruses have been identified (Brunt, 1996) (Clark, et al., 2012) (Kwak, et al., 2014). Since sweet potatoes are typically grown from "slips" or vine cuttings, viruses are more likely to survive from generation to generation (Loebinstein, Thottappilly, Fuentes, & Cohen, 2003). Not only that, but as slips are transferred to new fields, in new locations, viruses will also be transferred (Loebinstein, Thottappilly, Fuentes, & Cohen, 2003).

Immediately diagnosing a single specific viral infection within a sweet potato plant is next to impossible as many viral infections occur simultaneously, some viruses mask other infections, and some viruses have very diverse strains (Valverde, Clark, & Valkonen, 2007). Serology, molecular hybridization, and PCR are the three most commonly used diagnosis techniques for diagnosing viral infection within sweet potato plants (Valverde, Clark, & Valkonen, 2007) (Clark & Adams, 1976) (Abad & Moyer, 1992) (Colinet, Kummert, & Lepoivre, 1998). An ELISA test is commonly used in serological testing for viruses, but accuracy is often cited as a crucial problem as sometimes viruses and virus particles are unevenly distributed within the plant or concentrations are sometimes too low to detect anything (Clark & Adams, 1976) (Esbenshade & Moyer, 1982). Inhibitors within the tissues have also been known to interfere with the test, giving inaccurate results (Abad & Moyer, 1992). Abad and Moyer (1992) developed a successful detection method with RNA probes, which they developed *in vitro*, specifically for detecting SPFMV. PCR is by far the most reliable testing method for viral detection because it is the most sensitive method available, it is not dependent upon the condition of proteins as ELISA typically is, and real-time PCR methods can be used to accurately quantify

viruses (Valverde, Clark, & Valkonen, 2007) (Abad & Moyer, 1992) (Kokkinos & Clark, 2006). There is also a special type of PCR called multiplex RT-PCR, which can be utilized to detect multiple viruses simultaneously during one PCR run, without compromising accuracy, saving time and resources (Kwak, et al., 2014).

The most effective way by far to prevent viral infections is to prevent an infestation of the natural transmitters for these diseases, which are insects (Ames, Smit, Braun, O'Sullivan, & Skoglund, 1997) (Kwak, et al., 2014). Aphids are typically the primary transmitters of sweet potato feathery mottle virus (SPFMV) and whiteflies are the primary transmitters of sweet potato chlorotic stunt virus (SPCSV) (Ames, Smit, Braun, O'Sullivan, & Skoglund, 1997) (Kwak, et al., 2014). Aphids and whiteflies are typically controlled using the pesticides diazinon and carbofuran (Ames, Smit, Braun, O'Sullivan, & Skoglund, 1997). However, by the time these insects are noticed and sprayed for, viral infection usually has already occurred (Ames, Smit, Braun, O'Sullivan, & Skoglund, 1997). Therefore, a preemptive spray is usually a grower's best hope. The transport of slips across long distances is also often discouraged. The thought process is that even if a cutting contains one type of virus, like SPFMV, then disease complexes like SPVD can be prevented by avoiding contact with plant material infected with SPCSV.

2.6. SWEET POTATO BREEDING

2.6.1. Challenges

Traditional breeding methods involving cross-breeding methods are typically not an option as sterility is a common phenomenon in sweet potatoes likely due to the high levels of ploidy, which is thought to cause defects in recombination during meiosis, resulting in a zygote that cannot continue developing (Martin & Jones, 1968) (Lebot, 2009). Some of these dysfunctions are mechanical in the seed fertilization process and can be immediately detected;

for instance, the plant can fail to produce pollen or even full flowers, stigma might not stimulate germination, or gametes can be deformed (Martin F. , 1982). Sweet potato is also a self-incompatible crop, it cannot fertilize itself and it has a high incidence of cross-incompatibility, which can be caused by both barriers that occur before fertilization as well as post-fertilization (Ngailo, Shimelis, Sibiya, & Mtunda, 2013) (Martin F. W., 1970).

2.6.2. Varietal Development Strategies

Currently, a few strategies available have been attempted for enhancing viral disease resistance across sweet potato populations. Since the mid-nineties, genetic engineering strategies have been put into action that specifically targeted sweet potato resistance to SPFMV (Clark, et al., 2012) (Wambugu, 2003). Coat protein obtained from the feathery mottle virus was transformed into the sweet potato to confer resistance to SPFMV (Clark, et al., 2012). This transformation of the coat protein was eventually repeated via electroporation by Okada et al. (2001), along with an additional protein from SPFMV called hygromycin phosphotransferase. However, in both of these trials, resistance eventually broke down via co-infection by SPCSV, further confirming a synergistic relationship between these viruses (Clark, et al., 2012).

Considering that past experiments have confirmed that transforming the rice gene for cysteine proteinase inhibitor to tobacco plants has conferred resistance to potyviruses, a similar experiment was conducted that involved transforming this very same gene into the sweet potato plant (Clark, et al., 2012) (Gutierrez-Campos, Torres-Acosta, Saucedo-Arias, & Gomez-Lim, 1999). The hope of this experiment was to give the sweet potato plant resistance to SPFMV, so even if a plant was infected with SPCSV, the SPVD complex would not be able to develop (Clark, et al., 2012) (Gutierrez-Campos, Torres-Acosta, Saucedo-Arias, & Gomez-Lim, 1999). Although there was an increased resistance initially to SPFMV, this resistance eventually broke

down upon infection with SPCSV, showing symptoms of SPVD (vein clearing, leaf strapping, leaf distortion, chlorosis, puckering, and stunting) (Clark, et al., 2012) (Gutierrez-Campos, Torres-Acosta, Saucedo-Arias, & Gomez-Lim, 1999) (Kokkinos, Clark, McGregor, & LaBonte, 2006).

Transgenic sweet potato plants have also been developed by both particle bombardment and by vector transmission of a bar gene for herbicide resistance (gusA) into the sweet potato plant (Prakash & Varadarajan, 1992) (Yi, et al., 2007). Particle bombardment was carried out in the early nineties, using tungsten micro-carriers that were coated with plasmid DNA and later confirmed through histochemical staining of the tissues propagated from cell culture (Prakash & Varadarajan, 1992). Southern blot testing along with spraying with a commercial herbicide confirmed successful Agrobacterium-mediated transmission of the gene by Yi et al. (2007). The strategy here is for growers to be able to spray their sweet potato fields and kill weeds that support the survival of harmful vectors (whiteflies and aphids) from season to season, without adversely affecting their sweet potato crop (Yi, et al., 2007) (Prakash & Varadarajan, 1992).

Transformation has been demonstrated on embryogenic suspension cell cultures of select sweet potato cultivars, using *Agrobacterium tumefaciens* as a mediator (Yang, et al., 2011). The hygromycin phosphotransferase (HPT) gene was utilized as a selective marker to allow for antibiotic resistance only to successfully transformed cells (Yang, et al., 2011). The transgenesis method that was developed for sweet potato can now be used for future endogenous gene transformations (Yang, et al., 2011).

Fan et al. (2012) described a procedure in which they transformed the BADH gene (betaine aldehyde dehydrogenase) from spinach (*Spinacia oleracea*) into the Sushu-2 sweet potato cultivar using an agrobacterium-mediated method. BADH is one of the genes involved in

the biosynthesis of glycine betaine, which has been shown to improve tolerance to various abiotic stresses once it has substantially accumulated in the plant (Fan, Zhang, Zhang, & Zhang, 2012). These stresses that high amounts of glycine betaine guards against include low temperature, high salinity, and high oxidative stress, which are the highest concerns of sweet potato growers today (Fan, Zhang, Zhang, & Zhang, 2012).

2.6.3. Genetic Editing

Using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) technology for targeted editing of genomes is one of the newest available and arguably the most revolutionary tool in genetic editing, allowing for the alteration of specific genes in the organism of interest through addition or removal (Reis, Hornblower, Robb, & Tzertzinis, 2014). CRISPR was first discovered in the genome of *Escherichia coli*, although their role in disease resistance was not confirmed until decades later (Reis, Hornblower, Robb, & Tzertzinis, 2014). In between these palindromic repeats are pieces of viruses that have previously infected the cell and these newly incorporated pieces of DNA are used to incur resistance to any future invasions of this virus (Reis, Hornblower, Robb, & Tzertzinis, 2014). These inserts were made with assistance of Cas9 proteins which cleave the organism's DNA and insert the new pieces into the genome (CRISPR-associated proteins) (Reis, Hornblower, Robb, & Tzertzinis, 2014).Once these new loci are transcribed, their sequence complementarity to invading viruses allow these transcripts to guide endonucleases in the destruction of these pathogens (Reis, Hornblower, Robb, & Tzertzinis, 2014). Scientists over the last decade have come up with several protocols on how to exploit the basic function of the Cas9 protein with success in a variety of different model organisms (Reis, Hornblower, Robb, & Tzertzinis, 2014). Since the sweet potato is next to

impossible to breed for desirable characteristics through traditional means, the CRISPR/Cas9 system could open up countless new possibilities for it and other multi-ploidy crops as well.

CHAPTER III: MATERIALS & METHODS

3.1. GROWING MATERIAL

Sweet potato plants were grown from slips from three genotypes that gave rise to orange-fleshed, white-fleshed, and purple-fleshed storage roots and were propagated by Dr. Conrad Bonsi and his team at Tuskegee University. The slips were grown from previously harvested storage roots in the Delaware State University greenhouse. The storage roots were covered completely with soil and watered daily for two months before the resulting vines were transplanted to an organic plot at the Smyrna Outreach and Research Center (SORC) in the location indicated in **Figures 2A and 2B** (Google, 2018).

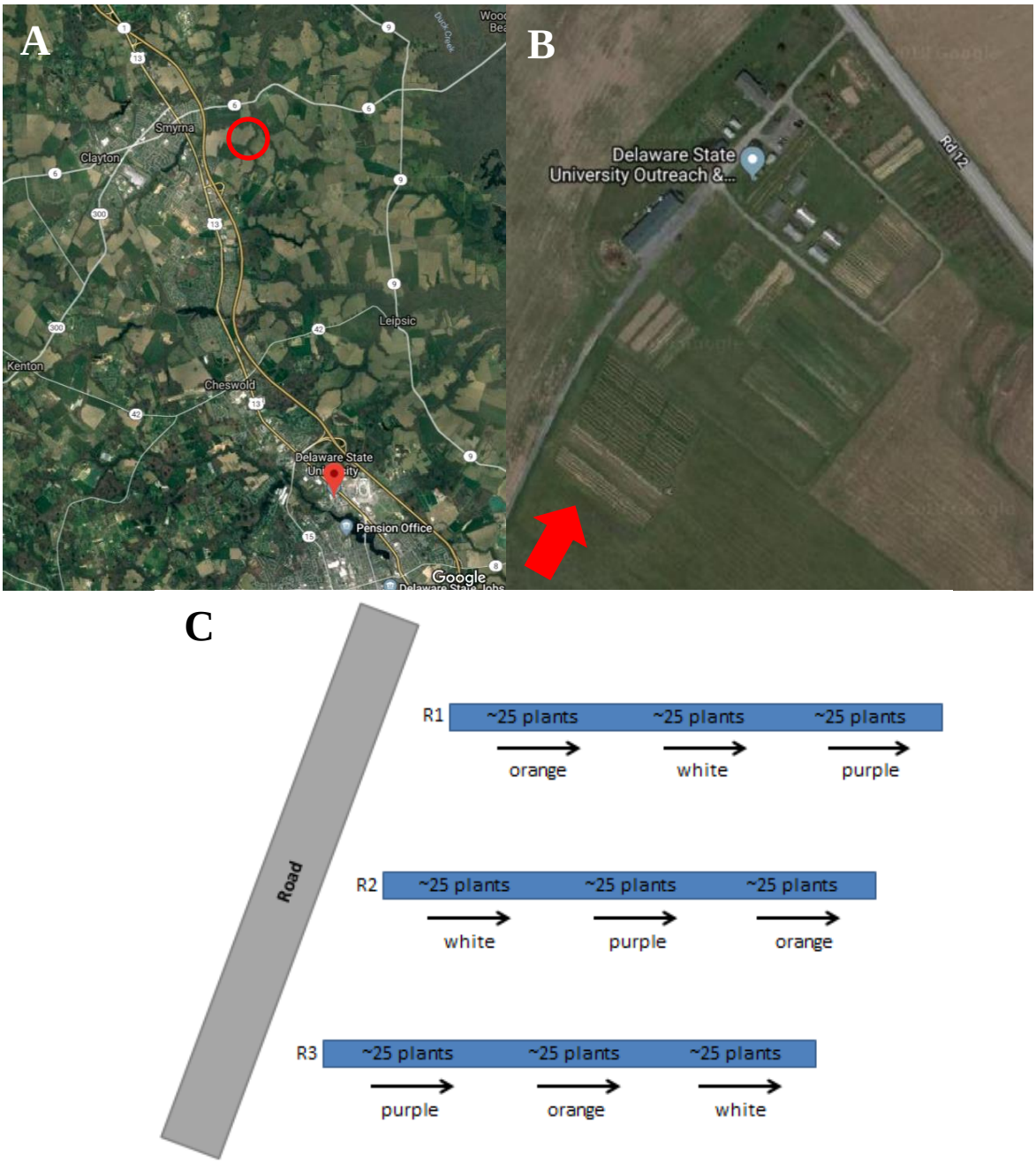


Figure 2: Smyrna Outreach and Research Center (SORC) (Google, 2018). **A.** Google Map view of SORC relative to Delaware State University. **B.** Close-up map of SORC; the red arrow indicates the location of our rows of sweet potato plants. **C.** A map of the orientation of the genotypes in each of the three rows.

The vines were cut into approximately 8-12 inch long slips with approximately 3-5 leaves on each and planted immediately into the soil on June 15th 2016. The plot was watered two to three times a week, depending on climate, for roughly four months before leaf samples were collected. The plot consisted of three rows (**Figure 2C**). Approximately 25-30 plants of each

genotype were planted in each row, therefore, there were ~225-270 plants in the entire plot. Leaf samples were separately collected from three different plants of each genotype in each row and immediately frozen in liquid nitrogen. In the end, leaf samples were taken from 27 different plants, three from each of the three genotypes in each of the three rows (3 plants X 3 genotypes X 3 rows).

3.2. PRIMER DESIGN

Eight sets of primers were used for screening for eight common sweet potato viruses. The viruses for which we screened included Sweet Potato Chlorotic Stunt Virus (SPCSV), Sweet Potato Feathery Mottle Virus (SPFMV), Sweet Potato Mild Mottle Virus (SPMMV), Sweet Potato Leaf Curl Virus (SPLCV), Sweet Potato Virus G (SPVG), Sweet Potato Latent Virus (SPLV), Sweet Potato Virus 2 (SPV2), and Sweet Potato Mild Speckling Virus (SPMSV). These primers were designed previously by Kwak et al. (2014), and used for determining how often and in which regions of Korea common sweet potato viruses tend to occur.

Four sets of intron-spanning primers were designed in order to make sure that there was no genomic DNA contamination in our cDNA after RNA was extracted, purified, and put through the process of reverse transcriptase. Intron-spanning primers are simply primers that will amplify a portion of an intron within a gene. This means that in pure cDNA, since introns are not transcribed, amplification will not occur. Genomic DNA could potentially affect our Real Time PCR validation and these primer sequences are a way to be sure we have eliminated this contamination. The sets of primers were designed from introns found within genes located on the National Center for Biotechnology Information (NCBI) website (U.S. National Library of Medicine, 2018). These introns were from genes including one duplicated gene (AB), a chloroplast rps16 gene for ribosomal protein S16 (AJ), a starch phosphorylase gene (L2), and a

cytosolic Cu/Zn-superoxide dismutase gene (L3) (U.S. National Library of Medicine, 2018). Eventually, AB2 primer set was selected to check for genomic DNA contamination, the other intron-spanning primers were not needed for further verification.

Housekeeping genes were found in a paper that purposefully sought out stable reference genes for Real Time PCR conducted on the sweet potato plant (Park, et al., 2012). The housekeeping genes were necessary for conducting Real Time PCR to verify sequencing results. Typically in this type of analysis, actin or tubulin primers are used for normalization. However, this paper explained that in sweet potato, these are not necessarily the best housekeeping genes to use and that it is preferable to use at least two housekeeping genes in each Real Time PCR analysis (Park, et al., 2012). We decided to use only one housekeeping gene for qPCR and this was ubiquitin extension protein (UBI). Although Park et al. (2012) suggested that cytochrome c oxidase subunit Vc (COX), glyceraldehyde-3-phosphate dehydrogenase (GAP), and ribosomal protein L (RPL) would also have been suitable for qPCR. The reason ARF2 was not considered a candidate can be seen in **Figure 3**, multiple bands were visualized in a single PCR run.

We designed the sequencing verification primers based on the reads that we received back from DBI. The primer design software from NCBI was used to design the primers for specific genes and only primer sets that would yield products between 100 and 200 base pairs were utilized (NCBI, 218).

Table 1 contains all of the primers that were acquired/designed for this project. The primers were checked to make sure that amplification will occur in genomic DNA. PCR was carried out on a Bio-Rad T100 Thermal Cycler (Step 1: 95°C five min, Step 2: 95°C 30 sec, 53°C 30 sec, 72°C 1 min (X34), and Step 3: 72°C 5 min).

All of the Primers			
Virus	**Primers**	**Product Size**	**Source**
Initial Virus Screening			
Sweet Potato Chlorotic Stunt Virus (SPCSV)	GGGAAGAMGAGAYATGGAGTTAA CCTTGTTACAAAGAGCGTTCCT	583	Kwak, et al., 2014
Sweet Potato Feathery Mottle Virus (SPFMV)	TACACACTGCTAAAACTAGG AGTTCATCATAACCCCATGA	356	Kwak, et al., 2014
Sweet Potato Mild Mottle Virus (SPMMV)	CCGCGCCAACAA AGGAACTA TTGATGGGGTAATAAAGCACT	298	Kwak, et al., 2014
Sweet Potato Leaf Curl Virus (SPLCV)	TCTGCCGTCGATCTGGAACTC GTGCCCGCCTTTGGTGGAC	507	Kwak, et al., 2014
Sweet Potato Virus G (SPVG)	CAATGCCAAATGGAAGAATAG GCATGATCCAATAGAGGTTTTA	286	Kwak, et al., 2014
Sweet Potato Latent Virus (SPLV)	GGAGTCAGTTCAATCAATGGTA AGTGGCTTTATTGGGTATGAT	184	Kwak, et al., 2014
Sweet Potato Virus 2 (SPV2)	ATGTGTTGAACCATCAGCTGAA GTAACTTGCCTTGGGCTACG	369	Kwak, et al., 2014
Sweet Potato Mild Speckling Virus (SPMSV)	GCCAAAACCAACAAGCATCA ATTCGCATTTCCTCATCATCT	275	Kwak, et al., 2014
Intron-Spanning			
Duplicated Gene GenBank: AB006185.1 (AB2)	CCACTTCAAGAGGGTAGACATAAA CCTATCAATTGTTGCCTTCACAC	579	Original Design
Chloroplast RPS 16 Gene for Ribosomal Protein S16 GenBank:AJ431071.1 (AJ2)	GAGGAAAGACCGCTCAAGAAA GGATAGATGTAGACGAGCAACAC	270	Original Design
Starch Phosphorylase Gene GenBank: L25626.2 (L22)	CTGTTTCGTTCGATCTCTCTCTT GCCACCAGCCACTCATATT	313	Original Design
Cytosolic Cu/Zn-Superoxide Dismutase Gene GenBank: L36229.1 (L32)	ACCATAGGATTATGGGTGGATTG GAGCCAGAGATTGTGAGTAAGG	340	Original Design
Housekeeping			
JX177359 (ARF)	CTTTGCCAAGAAGGAGATGC TCTTGTCCTGACCACCAACA	185	Park, et al., 2012
JX177358 (UBI)	TCGACAATGTGAAGGCAAAG CTTGATCTTCTTCGGCTTGG	209	Park, et al., 2012
S73602.1 (COX)	ACTGGAACAGCCAGAGGAGA ATGCAATCTTCCATGGGTTC	159	Park, et al., 2012
JX177362 (GAP)	ATACTGTGCACGGACAATGG TCAGCCCATGGAATCTCTTC	124	Park, et al., 2012
AY596742.1 (RPL)	TTTGACCGAAATGCCCTTAG TTCTGGTTCACCCCAACATT	160	Park, et al., 2012
Sequencing Verification			
G35591|TU58360(56.0)::chr8:2864259-2866482(+) (G35591)	GTCTCGGTTGATGACCTCATAG CACGAGAAGAATAGCAGGGTAG	191	Original Design
G13930|TU22754(432.5)::chr14:5165896-5167590(+) (G13930)	GATGGTTGTGCTGATGAGGT CTGATAGCGACCCAAGTGATG	185	Original Design
G4244|TU6996(0.7)::chr10:29406932-29408805(+) (G4244)	CACAGATTCTCTCCCGTTTCTC GAGCTTCTCTTCTCCCATCATC	110	Original Design
G26515|TU43560(2.4)::chr4:25005261-25007057(+) (G26515)	GTCGACAGTGTTGCTAGGTTAT TTCTGCTCAGGTTCTACAAGTG	150	Original Design
G33415|TU54771(26.1)::chr7:2000676-2009707(+) (G33415)	CTGAGCATAACACTCCTCCTTC AGCTTAGGGTGAGCATTGAC	152	Original Design
G28|TU51(3.8)::chr1:31800-34449(-) (G28)	CTTGCATTGTGAAGGCTGTG CGTTCCCTTTCACGGTTACT	113	Original Design

Table 1: Forward and Reverse Primers. Primers for viral screening, intron-spanning sections of DNA, housekeeping genes, and verification of sequence expression are listed here.

3.3. DNA EXTRACTION & VERIFICATION OF PRIMERS

DNA was extracted from leaf tissue based on the CTAB Buffer method commonly utilized in the Molecular Genetics and Epigenomics Laboratory (Doyle & Doyle, 1990). DNA was diluted to 100 ng/microliter. The diluted DNA was then used to test the total DNA to ensure that these primers would be effective. **Figure 3** is an example of primer testing using DNA from

the purple-fleshed genotype to test the housekeeping genes that we ordered to use for qPCR. The first lane contains the 100 base pair ladder. The second lane is our negative control, containing our PCR product performed with DNA and primers for SPFMV, which should and does give no band, even though we know this virus is present in this sweet potato, as SPFMV is a single-stranded RNA virus and its genetic information is never at any time in DNA form. The third through the seventh lanes show bands, indicating that our selected housekeeping genes, ARF, UBI, COX, GAP, and RPL respectively, do indeed work. The eighth lane is a second negative control, only run with genomic DNA and no primers, proving that this PCR is valid, as there is no band and no accidental amplification of any kind. The PCR program had an initial five minute step of 95°C. PCR was carried out on a Bio-Rad T100 Thermal Cycler (Step 1: 95°C five min, Step 2: 95°C 30 sec, 53°C 30 sec, 72°C 1 min (X34), and Step 3: 72°C 5 min). **Figure 3** is just a single example (in addition to others) of how we checked each housekeeping primer, intron spanning primer, and sequencing verification primer against each of the three genotypes. Since we do not have the virus in its pure form to verify the effectiveness of the virus screening primers, we had to accept that if they did not show a band, they were not present.

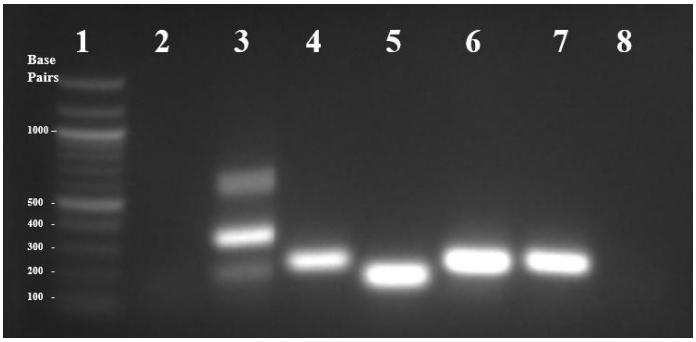


Figure 3: Primer verification example of UBI (housekeeping gene). Lane 1: 100 bp Ladder, Lane 2: DNA and SPFMV (negative control, SPFMV should not amplify in a DNA sample), Lane 3: DNA and ARF (185bp, however there are three different amplifications that occurred, making this a less than ideal primer to use for housekeeping purposes), Lane 4: DNA and UBI (209bp), Lane 5: DNA and COX (159bp), Lane 6: DNA and GAP (124bp), Lane 7: DNA and RPL (160bp), Lane 8: DNA only (no amplification). We decided that only one housekeeping gene would be needed for our validations, so we eventually opted to use the housekeeping gene, UBI. ARF ended up not being suitable because there were multiple amplifications, indicated by multiple bands in this lane and this could possibly throw off our qPCR results. PCR was carried out on a Bio-Rad T100 Thermal Cycler (Step 1: 95^0C five min, Step 2: 95^0C 30 sec, 53^0C 30 sec, 72^0C 1 min (X34), and Step 3: 72^0C 5 min).

3.4. RNA EXTRACTION & PURIFICATION

RNA was extracted using the Sigma Spectrum Plant Total RNA Kit, following the manufacturer's instructions (Sigma-Aldrich, 2010). The DNA Free Kit with Turbo DNAse from Invitrogen was then used to eliminate contaminating DNA as per the manufacturer's instructions (Thermo Fisher Scientific, 2018). **Figure 4** provides two examples of RNA genotypes visualized on an agarose gel before and after DNAse treatment is carried out.

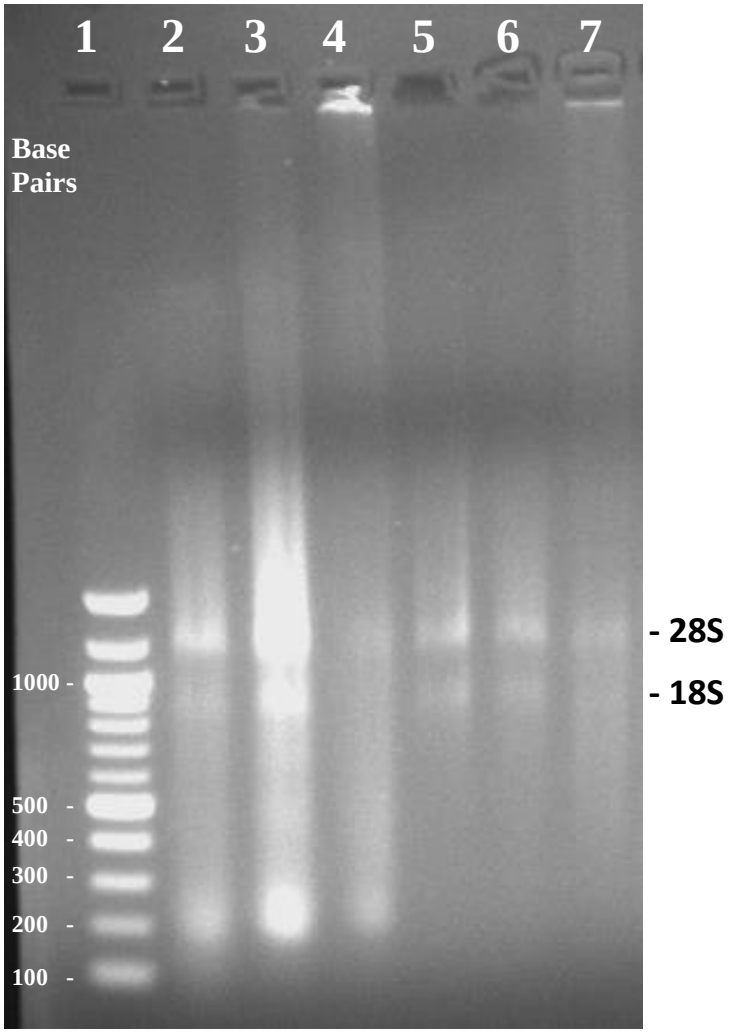


Figure 4: Example of RNA extraction before and after DNAse treatment. Lane 1: 100 bp Ladder, Lanes 2&3: genomic leaf RNA of white genotype, Lane 4: genomic leaf RNA of white genotype after DNAse treatment, Lanes 5&6: genomic leaf RNA of purple genotype, Lane 7: genomic leaf RNA of purple genotype after DNAse treatment. While all of the lanes contain the typical two visible subunits of ribosomal RNA (28S and 18S), they also have a slight smear, which contains the mRNA and residual DNA and proteins. However, the samples that have been put through DNAse treatment (lanes 4 and 7) leave a lighter smear in the gel, as residual DNA and proteins have been removed.

In order to verify that there was no DNA contamination in the purified batch of RNA, a small sample of RNA was converted to cDNA using the Protoscript II Reverse Transcriptase Kit (NEB, 2018) and one of the intron-spanning primers (AB2, 579 bp) was selected to test the cDNA with normal PCR procedures. If there is no DNA contamination in the RNA that was used to synthesize the cDNA, a band will not be observed because introns are not transcribed and translated to mRNA. If there is DNA contamination, amplification will occur and a band will be observed. The results of this verification method for three plant samples can be seen in **Figure 5**. Lanes 2, 6, and 10 contain PCR results of cDNA from each of the sweet potato genotypes (orange-fleshed, white-fleshed, and purple-fleshed, respectively). Amplification of housekeeping gene UBI (209 bp) occurs successfully, the observed bands serve as positive controls confirming that cDNA is in fact, viable. Lanes 3, 7, and 11 contain PCR results of genomic DNA from each of the sweet potato genotypes from the AB2 intron-spanning primer set, the resulting bands confirm that AB2 is in fact present in the genome of each of the sweet potato genotypes. Lanes 4, 8, and 12 contain PCR results of cDNA from each of the sweet potato genotypes along with primer set AB2, which have no band, as expected, since introns are not transcribed and translated. Had there been a resulting band, it would have indicated DNA contamination and results from sequencing and real time PCR verification would possibly have been flawed. PCR was carried out on a Bio-Rad T100 Thermal Cycler (Step 1: 95°C five min, Step 2: 95°C 30 sec, 53°C 30 sec, 72°C 1 min (X34), and Step 3: 72°C 5 min).

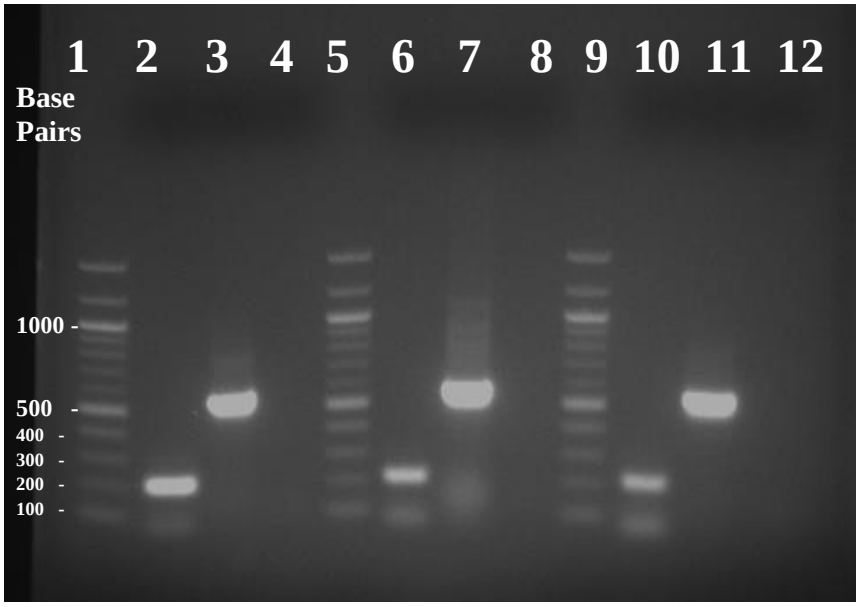


Figure 5: Verification of no DNA contamination. Lane 1: 100 bp ladder, Lane 2: orange genotype cDNA and UBI primer PCR results, Lane 3: orange genotype DNA and AB2 primer PCR results, Lane 4: orange genotype cDNA and AB2 primer PCR results, Lane 5: 100 bp ladder, Lane 6: white genotype cDNA and UBI primer PCR results, Lane 7: white genotype DNA and AB2 primer PCR results, Lane 8: white genotype cDNA and AB2 primer PCR results, Lane 9: 100 bp, Lane 10: purple genotype cDNA and UBI primer PCR results, Lane 11: purple genotype DNA and AB2 primer PCR results, Lane 12: purple genotype cDNA and AB2 primer PCR results. PCR was carried out on a Bio-Rad T100 Thermal Cycler (Step 1: 95^0C five min, Step 2: 95^0C 30 sec, 53^0C 30 sec, 72^0C 1 min (X34), and Step 3: 72^0C 5 min).

3.5. RNA LIBRARY PREPARATION

Out of the samples collected from the 27 plants, only nine plants were utilized for RNA sequencing and analysis. Once RNA sequencing was carried out, only one of each of the three plants collected from each genotype in each row was used in the final analysis. The plant used from each group was selected based on whichever one was determined to have the highest RNA integrity number (RIN), which was revealed by the bioanalyzer at DBI. This eliminated 18 of the samples for sequencing. Therefore, one plant from each genotype from each row was used for RNA Sequencing and analysis, effectively leaving three biological replicates for each genotype after RIN-based eliminations were made. RNA Library Preparation was carried out with the TruSeq RNA Library Prep Kit v2, distributed by Illumina (Illumina, 2017). Set A of the indexes was utilized and protocols were carried out according to the manufacturer's instructions (Illumina, 2017).

3.6. RNA SEQUENCING

Sequencing was carried out at the Delaware Biotechnology Institute (DBI) at the University of Delaware. Since each sample had a different index utilized for it, there was no need to separate the samples into different lanes. The samples were all run in the same lane on the Illumina Hi-Seq 2000 platform. **Figure 6** shows how the reads would appear once they were uploaded into the CLC software. Each of the reads were 101 nucleotides long and consisted of the first and last 101 nucleotides on each of the fragments. This sequencing format is known as "paired-end" sequencing. The pink bars below the nucleotides indicated in the sequence are quality scores and they indicate the likelihood of the nucleotide identity predicted by the sequencing platform. As can be seen here, this is only eleven of the millions of reads deciphered by the Illumina sequencing platform. In this experiment, the higher the pink bar, the more likely it is that the machine made an accurate prediction of the nucleotide. The fifth read down has relatively high quality scores for all of its nucleotides, which means that this read stood a better chance at being aligned to our reference genome. The fourth read, however, had almost none of its sequences read and probably was not able to be aligned to the reference genome.

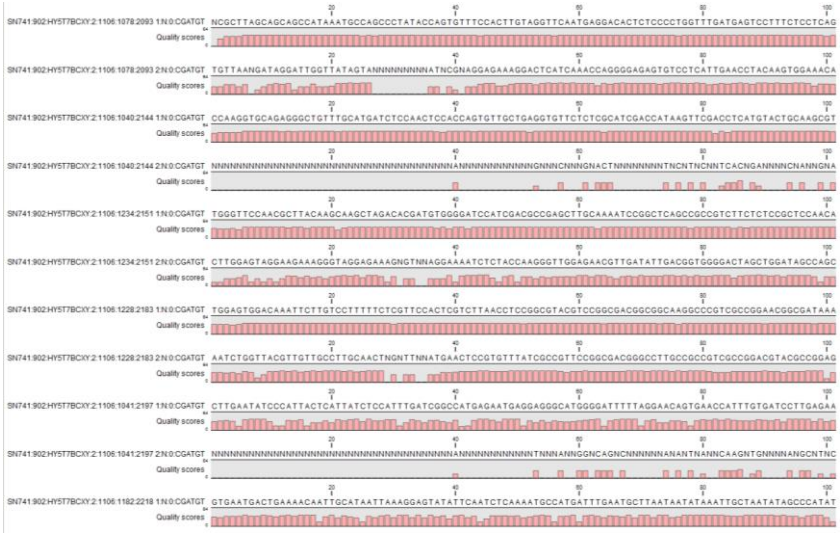
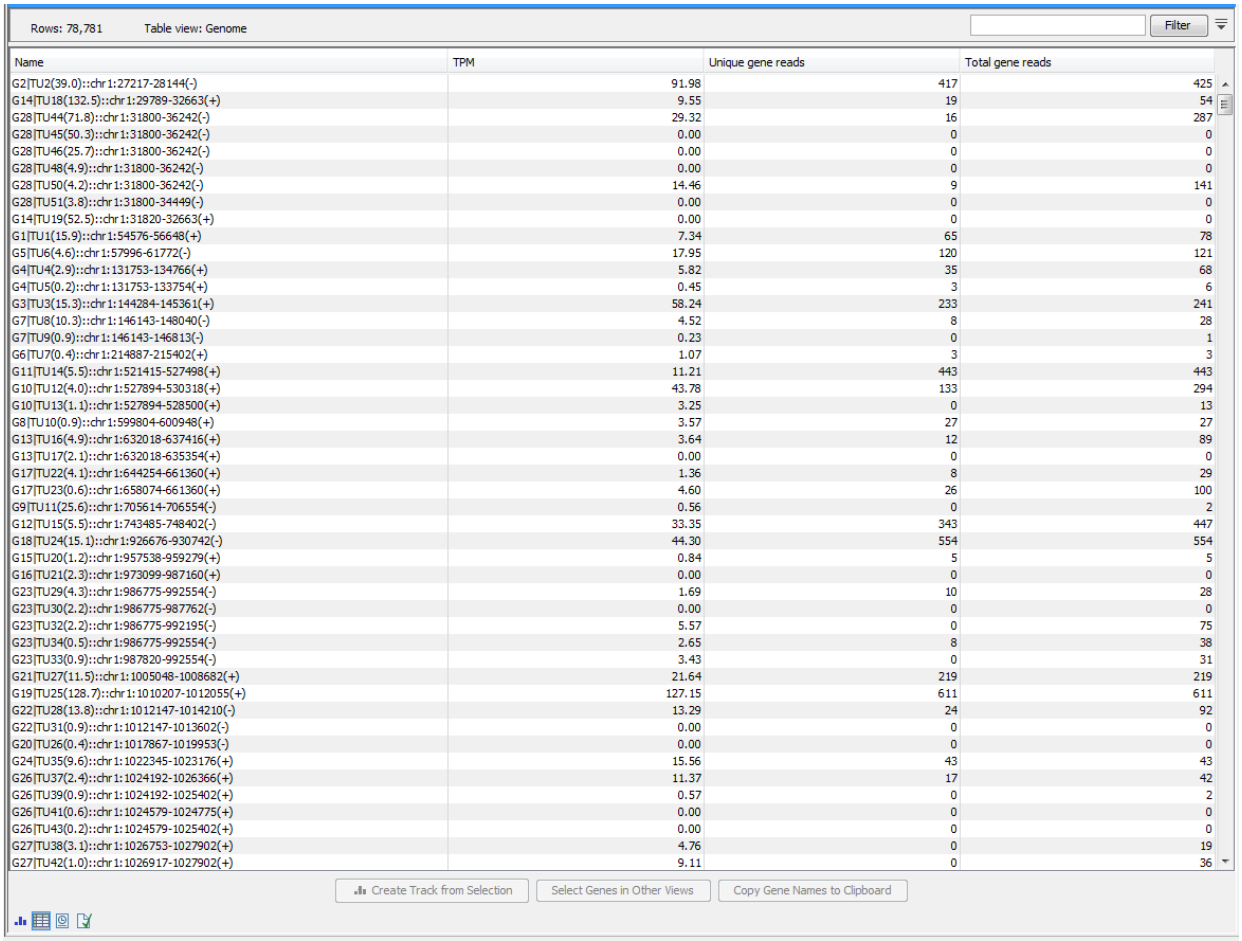


Figure 6: RNA Sequencing Reads. Here is an example of the RNA sequencing reads visualized in the CLC Bio software before alignment.

3.7. BIOINFORMATICS ANALYSIS

Reads from RNA sequencing were imported into the CLC Bio program (Qiagen, 2017). The sweet potato genome, along with the functional annotations of predicted genes available at the NGS Sequencing Core Facility at Max Planck Institute Molecular Genetics in Berlin, Germany was downloaded and imported into the CLC Bio program (NGS Sequencing Core Facility, 2017) (Yang, et al., 2017). The sweet potato genome assembled by the Max Planck Institute was used for mapping our reads to label and predict potential gene functions as well as estimate gene expression levels of the genes identified by the Max Planck Institute (NGS Sequencing Core Facility, 2017).

Figure 7 is an example of how our reads were aligned to the reference genome. In no particular order, 47 of the original 78,781 reference genes from the reference genome are displayed, with the gene calling numbers, the TPM values, the amount of unique reads aligned to each gene, and the total amount of reads aligned to each gene are included in this final report. TPM stands for transcripts per million and is now the standard method used for normalizing expression data. In other words, since each read is smaller than a gene, it makes sense, that larger genes will have more reads, but through TPM normalization, the size of the gene will not affect the expression value assigned to each gene. Just because some of these genes have a higher amount of gene reads, it does not necessarily mean they will have a higher TPM, this will depend entirely on the length of the gene relative to other genes contained within the genome. The default mapping parameters were utilized.



Rows: 78,781 Table view: Genome

Filter

Name	TPM	Unique gene reads	Total gene reads
G2|TU2(39.0)::chr 1:27217-28144(-)	91.98	417	425
G14|TU18(132.5)::chr 1:29789-32663(+)	9.55	19	54
G28|TU44(71.8)::chr 1:31800-36242(-)	29.32	16	287
G28|TU45(50.3)::chr 1:31800-36242(-)	0.00	0	0
G28|TU46(25.7)::chr 1:31800-36242(-)	0.00	0	0
G28|TU48(4.9)::chr 1:31800-36242(-)	0.00	0	0
G28|TU50(4.2)::chr 1:31800-36242(-)	14.46	9	141
G28|TU51(3.8)::chr 1:31800-34449(-)	0.00	0	0
G14|TU19(52.5)::chr 1:31820-32663(+)	0.00	0	0
G1|TU1(15.9)::chr 1:54576-56648(+)	7.34	65	78
G5|TU6(4.6)::chr 1:57996-61772(-)	17.95	120	121
G4|TU4(2.9)::chr 1:131753-134766(+)	5.82	35	68
G4|TU5(0.2)::chr 1:131753-133754(+)	0.45	3	6
G3|TU3(15.3)::chr 1:144284-145361(+)	58.24	233	241
G7|TU8(10.3)::chr 1:146143-148040(-)	4.52	8	28
G7|TU9(0.9)::chr 1:146143-146813(-)	0.23	0	1
G6|TU7(0.4)::chr 1:214887-215402(+)	1.07	3	3
G11|TU14(5.5)::chr 1:521415-527498(+)	11.21	443	443
G10|TU12(4.0)::chr 1:527894-530318(+)	43.78	133	294
G10|TU13(1.1)::chr 1:527894-528500(+)	3.25	0	13
G8|TU10(0.9)::chr 1:599804-600948(+)	3.57	27	27
G13|TU16(4.9)::chr 1:632018-637416(+)	3.64	12	89
G13|TU17(2.1)::chr 1:632018-635354(+)	0.00	0	0
G17|TU22(4.1)::chr 1:644254-661360(+)	1.36	8	29
G17|TU23(0.6)::chr 1:658074-661360(+)	4.60	26	100
G9|TU11(25.6)::chr 1:705614-706554(-)	0.56	0	2
G12|TU15(5.5)::chr 1:743485-748402(-)	33.35	343	447
G18|TU24(15.1)::chr 1:926676-930742(-)	44.30	554	554
G15|TU20(1.2)::chr 1:957538-959279(+)	0.84	5	5
G16|TU21(2.3)::chr 1:973099-987160(+)	0.00	0	0
G23|TU29(4.3)::chr 1:986775-992554(-)	1.69	10	28
G23|TU30(2.2)::chr 1:986775-987762(-)	0.00	0	0
G23|TU32(2.2)::chr 1:986775-992195(-)	5.57	0	75
G23|TU34(0.5)::chr 1:986775-992554(-)	2.65	8	38
G23|TU33(0.9)::chr 1:987820-992554(-)	3.43	0	31
G21|TU27(11.5)::chr 1:1005048-1008682(+)	21.64	219	219
G19|TU25(128.7)::chr 1:1010207-1012055(+)	127.15	611	611
G22|TU28(13.8)::chr 1:1012147-1014210(-)	13.29	24	92
G22|TU31(0.9)::chr 1:1012147-1013602(-)	0.00	0	0
G20|TU26(0.4)::chr 1:1017867-1019953(-)	0.00	0	0
G24|TU35(9.6)::chr 1:1022345-1023176(+)	15.56	43	43
G26|TU37(2.4)::chr 1:1024192-1026366(+)	11.37	17	42
G26|TU39(0.9)::chr 1:1024192-1025402(+)	0.57	0	2
G26|TU41(0.6)::chr 1:1024579-1024775(+)	0.00	0	0
G26|TU43(0.2)::chr 1:1024579-1025402(+)	0.00	0	0
G27|TU38(3.1)::chr 1:1026753-1027902(+)	4.76	0	19
G27|TU42(1.0)::chr 1:1026917-1027902(+)	9.11	0	36

Create Track from Selection | Select Genes in Other Views | Copy Gene Names to Clipboard

Figure 7: Alignment of Reads to Genome. This is an example of how our reads were aligned to the Max Planck Institute's genome.

The Fold Change function within the CLC Bio program was then utilized to express the differences in gene expression for all the genes at once (NGS Sequencing Core Facility, 2017). Although each of the plants were sequenced separately, the sequencing reads were pooled together according to genotype for simplicity while conducting fold change comparisons. Two genotypes were compared each time, meaning that we made a Fold Change comparison between the orange-fleshed and the white-fleshed genotypes, the white-fleshed and the purple-fleshed genotypes, and the orange-fleshed and the purple-fleshed genotypes. From the fold change amounts, log 2-fold changes were generated using the log equation function in Microsoft Excel [=LOG(number, base)]. Also added to the gene calling numbers and the log 2-fold change

results, were the gene functions acquired from the Max Planck sweet potato website (NGS Sequencing Core Facility, 2017).

For the analysis, several genes were sorted into functional groups using keywords and the search function in Microsoft Office to search the actual gene function column. The resources used for this sorting included a paper on gene expression levels of sweet potatoes that were affected by Sweet Potato Virus Disease (SPVD), a paper on the transcriptome profiles of two switchgrass genotypes, another paper that profiles the transcriptomes of developing storage roots, and an encyclopedia resource. From this procedure, we were able to create tables and graphs to give us better visual aids for gene expression between the three different genotypes. We decided to sort the sequences into eleven functional groups, so eleven different tables were created. Examples of the keywords used for data mining can be located in **Table 2** (Ayyappan, et al., 2017) (Kokkinos, Clark, McGregor, & LaBonte, 2006) (Firon, et al., 2013) (The Gale Encyclopedia of Science, 2016).

Transcription Factors			
Acyl-CoA	Endonuclease	High mobility group	Pyrophosphatase
Acyltransferase	Growth Regulator	Indole	Regulator
Ankyrin	Heat Shock Protein	Mitochondrial transport	Squamosa
Cytokinin	Helix-Loop-Helix	Phytochrome	Tubulin
Stress-Related Transporters			
Aluminum sensitive	Chlorophyll A Apoprotein	Histidine kinase	Pleiotropic drug
Asparagine synthesis	Dehydration response	Multidrug resistance	Prenyltransferase
ATP-binding	Ethylene-responsive	NADPH	RNA polymerase sigma
ATPase	Glutamine dependent	Nucleoside triphosphate	Thylakoid lumenal
Various Stress-Responsive Genes			
Aldehyde dehydrogenase	Carotenoid cleavage	Epoxycarotenoid dioxygenase	Peroxiredoxin
Alpha-amylase	Cell division control	Expansin	Phenylalanine ammonia-lyase
Aminocyclopropane	Chaperone protein	Leucine-rich repeat	Scarecrow
Brassinosteroid	Cyclin	O-acetylserine	Tonoplast intrinsic
Other Stress Related Genes			
Beta-hydrolase	Catalytic chain	Ferredoxin	Recognition particle
Calcineurin	Chitinase	Kinesin	Splicing
Calcium-binding	Cycloisomerase	Motor	Tetratricopeptide
Calmodulin	Cytochrome	Phosphoenolpyruvate	Thioredoxin
Cell Rescue, Defense, and Virulence			
Allergen	Chaperone	Metallothionein	Rac-like
Catalase	GTP-binding	Peroxidase	Trigger factor
Disease Resistance Genes			
Bacterial	Extensin	Protease inhibitor	Syntaxin
Disease	Lipid transfer protein	RING	Virus
Pigment			
Anthocyanin	Chlorophyll	Lutein	Purple
Carotene	Chloroplast	Lycopene	Violaxanthin
Carotenoid	Green	Porphyrin	Zeaxanthin
Protein Synthesis and Protein Fate			
2B-like	Elongation factor	Polyubiquitin	Senescence
eIF-2B	Eukaryotic translation	Ribosome	Subtilase
Metabolism			
Adenine	Cinnamyl alcohol	Glucose-6-phosphate isomerase	Mitochondria
Adenosine kinase	Coproporphyrinogen	Glyceraldehyde	Phosphoglycolate
Aminotransferase	Eukaryotic translation	Glycine-cleavage	Rubisco
Bisphosphate carboxylase	Ferredoxin-thioredoxin reductase	Glyoxylate aminotransferase	Terpene synthase
Energy			
C oxidase	Chloroplast	Cytochrome	Rieske
Chlorophyll	COX	Iron-sulfur	Thylakoid
Initiation of Storage Roots			
4-coumarate	Caffeoyl-CoA	Fructose	Phosphoglucomutase
Beta-amylase	Cinnamoyl-CoA	Galactose	Serine-threonine
Beta-expansin	Ferredoxin	Glucosyltransferase	Starch

Table 2: Examples of Keywords for Functional Groups for Data Mining. Examples of the keywords for each of the eleven functional groups used for data mining. These keywords were acquired from journal articles and websites (Ayyappan, et al., 2017) (Kokkinos, Clark, McGregor, & LaBonte, 2006) (Firon, et al., 2013) (The Gale Encyclopedia of Science, 2016).

3.8. CDNA SYNTHESIS

cDNA synthesis was necessary to make sure there was no DNA contamination, as previously indicated in the "RNA Extraction & Purification" section. We still had to ensure that cDNA synthesis did, in fact, occur. For this, we would typically use housekeeping genes, as can be seen in **Figure 5** (lanes 2, 6, and 10). It was also needed for RNA-Seq verification and to screen for the eight selected viruses indicated in this paper, since PCR cannot be carried out on RNA.

3.9. RNA-SEQ VERIFICATION

In order to verify the accuracy of the RNA sequencing, qPCR needed to be utilized. qPCR could quantify gene expression, though not as fast and efficiently as the Illumina sequencing process. Here, qPCR was used as a way to double-check a couple of commonly expressed genes, and in a sense, "audit" all of the sequencing data.

3.10. VIRUS SCREENING

Eight different common sweet potato viruses were screened for in all 27 leaf samples collected from the sweet potato plot at SORC. The primers for each of these viruses are given in **Table 1**. RNA was extracted and purified the same way as above explained in "RNA Extraction and Purification." Purified RNA was first converted to the much more stable complementary DNA (cDNA) through the process of reverse transcriptase using the Protoscript II First Strand cDNA Synthesis Kit since most of these viruses are either single-stranded RNA viruses (SPLCV is a double-stranded DNA virus, but by now there are resultant mRNAs present) (New England Biolabs Inc., 2016). Protocols were followed as per the manufacturer's instructions (New England Biolabs Inc., 2016). PCR was carried out on a Bio-Rad T100 Thermal Cycler (Step 1: 95°C five min, Step 2: 95°C 30 sec, 53°C 30 sec, 72°C 1 min (X34), and Step 3: 72°C 5 min).

CHAPTER IV: RESULTS

4.1. PHENOTYPIC ANALYSIS

Aside from the storage roots being different colors, they are also different sizes and shapes. Also, different quantities of sweet potatoes are produced per plant. For calculating how many sweet potatoes were produced per plant, we used the minimum standard for the lowest marketable sweet potato grade set by the USDA (United States Department of Agriculture, 2018).

The lowest minimum standard for individual storage roots is a diameter of 1.5 inches or 3.81 cm; we called storage roots that met this diameter requirement marketable storage roots. This was how we differentiated sweet potatoes from fingerlings, which are essentially just thick feeder roots that may become sweet potatoes. There was no minimum size for this latter grade for storage root length or mass (United States Department of Agriculture, 2018). **Table 3** gives the total amount of plants we began with, the total amount of plants from which we were able to harvest, and how many sweet potatoes we were able to harvest as well as the total mass harvested from each plant. The orange-fleshed plants, on average, produced the highest amount of storage roots per plant at 3.4 (495grams/plant), the white-fleshed plants produced 3.2 storage roots per plant (590grams/plant), and the purple-fleshed plants actually produced the least amount of storage roots per plant at approximately 2.9 (981grams/plant). As far as diameter and mass goes, the purple-fleshed sweet potatoes were larger on average going by diameter and mass (5.5cm and 332g), the white-fleshed potatoes typically were somewhere in the middle (5.2cm and 182g), and the orange-fleshed potatoes were smaller overall (5cm and 146g). This table also has the average statistics for the sweet potatoes harvested in this experiment, including the individual storage root statistics like diameter, length, and mass.

Genotype	Total Plants Planted	Total Plants Harvested	Total Marketable Storage Roots	Total Mass of Marketable Storage Roots	Storage Roots/Plant	Storage Root Mass/Plant	Average Storage Root Measurements		
							Diameter (cm)	Length (cm)	Mass (g)
Orange	55	47	159	23284	3.4	495	5	18	146
White	72	64	207	37751	3.2	590	5.2	16	182
Purple	95	88	260	86284	2.9	981	5.5	20	332

Table 3: Average Harvest Sizes of Plants and Storage Roots. Total plants, storage roots, average storage roots harvested from each plant, and average storage root size and mass.

4.2. RNA SEQUNCING & BIOINFORMATICS ANALYSIS

The RNA sequencing conducted at the Delaware Biotechnology Institute (DBI) yielded 200,602,352 reads total (65,965,450 from the orange-fleshed genotype, 66,644,712 from the white-fleshed genotype, and 67,992,190 from the purple-fleshed genotype) (**Table 4)**. The sweet potato genome at the Max Planck Institute consisted of 78,781 predicted genes, as reported by Yang et al. (2017) from the Uniprot and NCBI protein databases. Of our total reads, just under 80% were able to be mapped to the Max Planck genome; 50,814,350 reads, 50,925,213 reads, and 52,439,608 reads, in the orange, white, and purple fleshed genotypes respectively, as can be seen in **Table 4.**

Genotype	Reads Mapped	% Mapped	Reads Unmapped	% Unmapped	Total
Orange-Fleshed	50,814,350	77.03	15,151,100	22.97	65,965,450
White-Fleshed	50,925,213	76.41	15,719,499	23.59	66,644,712
Purple-Fleshed	52,439,608	77.13	15,552,582	22.87	67,992,190

Table 4: Reads obtained from Delaware Biotech Institute. The mapped reads were actually mapped to the genome that had been functionally annotated with the predicted genes from the Max Planck Institute in Berlin Germany.

Reads were processed after they were fragmented and the sequencing platform mapped them to the reference genome. However, not all reads could be mapped. The reads which were unable to be matched to the Max Planck genome were either single reads or paired reads. The determining factor here is whether both the ends of a fragment is matched to the genome or only a single end of the fragment is unable to be matched to the genome. As can be seen in **Table 5**, there are approximately four times as many paired reads as single reads that are not mapped to the genome.

Raw Reads	Plant	Reads	Unmapped Single Reads	Unmapped Paired Reads
Orange Flesh	1	23,115,488	1,043,362	4,285,746
	2	22,914,038	1,016,059	4,197,204
	3	19,935,924	898,191	3,710,538
	Total	**65,965,450**	**2,957,612**	**12,193,488**
White Flesh	1	20,940,372	946,042	3,846,722
	2	21,386,800	942,592	4,208,784
	3	24,317,540	988,027	4,787,332
	Total	**66,644,712**	**2,876,661**	**12,842,838**
Purple Flesh	1	23,457,712	1,004,245	4,082,622
	2	23,105,320	1,019,996	4,294,742
	3	21,429,158	953,399	4,197,578
	Total	**67,992,190**	**2,977,640**	**12,574,942**
	Total Reads	**200,602,352**	**8,811,913**	**37,611,268**

Table 5: Mapped and Unmapped Reads. This table lists the total mapped and unmapped reads for each of the three plants from each genotype. There were 200,602,352 reads total between all nine of these plants (three plants from each genotype).

Table 6 shows the annotated genes that have the highest TPM or transcripts per million. Transcripts per million is now becoming the preferred standardized method for "normalizing" RNA-Seq reads. A single TPM means a read out of one million RNA molecules in a single gene or transcript. Basically, it is a method of accurately comparing two or more sequencing samples. **Table 6** contains the ten genes with the highest TPMs for each of the three genotypes.

Interestingly, the first gene in **Table 6** has the highest expression when it comes to all three of the genotypes. This gene is labeled with "putative ribulose bisphosphate carboxylase

small subunit protein precursor." There is also a second gene labeled with the same function within the top ten highest expression for all three genotypes as well. Another assembled fragment is labeled exactly the same and contains the aforementioned gene, being as the starting position of each assembled section start at the same position on the second chromosome (nucleotide position 12036053). Function is significant here because these assembled reads belonging to these two genes are involved with the assembly of Rubisco or ribulose bisphosphate carboxylase. Rubisco is an important enzyme involved in the Calvin Cycle, necessary for the conversion of carbon dioxide into energy-rich molecules (Salesse-Smith, et al., 2018), it adds CO2 to the carboneal structure ribulose-1,5-bisphosphate.

Another highly expressed gene across all three sweet potato genotypes has been labeled with the function "thiamine thiazole synthase 2, chloroplastic." This means that the reads assembled to this gene are key in the manufacturing of thiamine, specifically the conversion of glycine and nicotinamide adenine dinucleotide (NAD) to adenosine diphosphate, which is an intermediate that occurs during the biosynthesis of thiamine (UniProt, 2018). Thiamine, while involved in a number of biological processes, is essential for an organism's adaptation to abiotic stress conditions and DNA damage repair, so therefore, this aforementioned gene fragment is likely to be essential and common to the adaptation of all three of these genotypes (UniProt, 2018).

The three different feature IDs mentioned above have been highlighted in yellow, blue, and pink, respectively, in **Table 6** to better indicate the similarities in up-regulation between the different genotypes. In addition to these three being in the top five highest expressed genes when it comes to all three of the genotypes, there is one more that is in the top five overall, but its function is unknown at this time.

Many of the other genes listed here seem to have a lot to do with chloroplast/chlorophyll functions and general chloroplast maintenance and have been highlighted in green. Another interesting thing to note here is a feature ID labeled with "Metallothionein-like type 1 protein" was also one of the ten highest expressed genes in all three genotypes as well, which indicates that this gene has to do with the binding of heavy metals (UniProt, 2018).

Feature ID	TPM	Gene Function
Orange Sweet Potato		
G19726|TU32276(293.0)::chr2:12036053-12036886(+)	22741.13	putative ribulose bisphosphate carboxylase small subunit protein precursor [*Nicotiana tabacum*]
G25080|TU41151(268.8)::chr4:5570806-5571516(-)	11851.27	PREDICTED: thiamine thiazole synthase 2, chloroplastic-like [*Sesamum indicum*]
G19726|TU32275(389.6)::chr2:12036053-12125859(+)	8335.53	putative ribulose bisphosphate carboxylase small subunit protein precursor [*Nicotiana tabacum*]
G32716|TU53582(138.5)::chr6:27926871-27927134(-)	3842.65	Chlorophyll a/b binding protein (Fragment)
G40386|TU66249(545.7)::scaffold1010|size217550:113348-113871(-)	3624.64	Uncharacterized protein (Fragment)
G288|TU501(161.2)::chr1:5987717-5990401(-)	3572.56	PREDICTED: 29 kDa ribonucleoprotein A, chloroplastic [*Sesamum indicum*]
G37183|TU60947(2717.2)::chr8:28144285-28146077(-)	3211.04	Metallothionein-like type 1 protein n=1 Tax=*Ipomoea batatas* RepID=
G30898|TU50661(1051.0)::chr6:1081716-1195512(-)	2834.54	RNA-binding gricine-rich protein-1c [*Nicotiana sylvestris*]
G40847|TU66929(700.0)::scaffold1146|size198777:55021-56285(-)	2833.80	cytochrome P450 like TBP [*Nicotiana tabacum*]
G42592|TU69440(78.1)::scaffold171|size496903:180153-181554(-)	2572.43	PREDICTED: photosystem I subunit O-like [*Nicotiana tomentosiformis*]
White Sweet Potato		
G19726|TU32276(293.0)::chr2:12036053-12036886(+)	19381.98	putative ribulose bisphosphate carboxylase small subunit protein precursor [*Nicotiana tabacum*]
G19726|TU32275(389.6)::chr2:12036053-12125859(+)	6707.02	putative ribulose bisphosphate carboxylase small subunit protein precursor [*Nicotiana tabacum*]
G25080|TU41151(268.8)::chr4:5570806-5571516(-)	4056.01	PREDICTED: thiamine thiazole synthase 2, chloroplastic-like [*Sesamum indicum*]
G40386|TU66249(545.7)::scaffold1010|size217550:113348-113871(-)	3947.00	Uncharacterized protein (Fragment)
G40847|TU66929(700.0)::scaffold1146|size198777:55021-56285(-)	3403.19	cytochrome P450 like TBP [*Nicotiana tabacum*]
G27538|TU45153(50.8)::chr4:47423371-47423811(+)	2855.47	PREDICTED: calcium-dependent protein kinase 32-like [*Nicotiana tomentosiformis*]
G288|TU501(161.2)::chr1:5987717-5990401(-)	2830.86	PREDICTED: 29 kDa ribonucleoprotein A, chloroplastic [*Sesamum indicum*]
G32716|TU53582(138.5)::chr6:27926871-27927134(-)	2694.49	Chlorophyl a/b binding protein (Fragment)
G37183|TU60947(2717.2)::chr8:28144285-28146077(-)	2436.01	Metallothionein-like type 1 protein n=1 Tax=*Ipomoea batatas* RepID=
G32675|TU53521(62.0)::chr6:28079047-28079763(+)	2139.70	PREDICTED: chlorophyll a-b binding protein 21, chloroplastic [*Sesamum indicum*]
Purple Sweet Potato		
G19726|TU32276(293.0)::chr2:12036053-12036886(+)	10888.34	putative ribulose bisphosphate carboxylase small subunit protein precursor [*Nicotiana tabacum*]
G25080|TU41151(268.8)::chr4:5570806-5571516(-)	10274.50	PREDICTED: thiamine thiazole synthase 2, chloroplastic-like [*Sesamum indicum*]
G37183|TU60947(2717.2)::chr8:28144285-28146077(-)	5206.99	Metallothionein-like type 1 protein n=1 Tax=*Ipomoea batatas* RepID=
G40386|TU66249(545.7)::scaffold1010|size217550:113348-113871(-)	4895.55	Uncharacterized protein (Fragment)
G19726|TU32275(389.6)::chr2:12036053-12125859(+)	4742.28	putative ribulose bisphosphate carboxylase small subunit protein precursor [*Nicotiana tabacum*]
G40847|TU66929(700.0)::scaffold1146|size198777:55021-56285(-)	3917.79	cytochrome P450 like TBP [*Nicotiana tabacum*]
G30898|TU50661(1051.0)::chr6:1081716-1195512(-)	3832.47	RNA-binding gricine-rich protein-1c [*Nicotiana sylvestris*]
G288|TU501(161.2)::chr1:5987717-5990401(-)	2960.41	PREDICTED: 29 kDa ribonucleoprotein A, chloroplastic [*Sesamum indicum*]
G18851|TU30793(706.9)::chr15:32252122-32252678(-)	2814.01	PREDICTED: uncharacterized protein LOC102603354 [*Solanum tuberosum*]
G18851|TU41174(53.7)::chr4:5666408-5666867(+)	2711.26	

Table 6: Genes with the Highest Transcripts Per Million (TPM). TPM for each of the three genotypes in this study.

The fold change function in the CLC program was used to compare gene expression between the genotypes two at a time. For this function to be carried out, the mapped reads from the three orange storage root plants were combined, as well as the mapped reads from the three

white storage root plants, and also the mapped reads from the three purple storage root plants. The reads resulting from the orange-fleshed plants were then compared to the reads from the white-fleshed plants. This type of comparison was repeated for the reads from the white-fleshed plants and the reads from the purple-fleshed plants and finally for the reads from the orange-fleshed plants and the reads from the purple-fleshed plants. As a result, we were able to determine which genes were most up-regulated in each of these comparisons. We were also able to determine how many, not to mention, which genes were up-regulated in each comparison made within CLC.

Figure 8 shows overall results for log 2-fold changes of two or more (>4 fold change), as this is our cutoff for significantly up-regulated genes after using the fold change function in CLC Bio.. The first two columns consist of the genes that sweet potatoes, 2814 genes were up-regulated in orange-fleshed sweet potatoes while 3891 genes were up-regulated in white-fleshed sweet potatoes. The two middle columns include the genes that show greater than 2-fold up-regulation when comparing white- and purple-fleshed sweet potatoes, 4559 genes were up-regulated in white-fleshed sweet potatoes and 3772 genes were up-regulated in purple-fleshed sweet potatoes. Meanwhile, the last two columns represent the genes that show greater than 2-fold up-regulation when comparing orange- and purple-fleshed sweet potatoes, 3521 genes were up-regulated in orange-fleshed sweet potatoes and 4364 genes were up-regulated in purple-fleshed sweet potatoes. As the most differences in significantly up-regulated genes occur between white- and purple-fleshed sweet potatoes, it is fair to say that these two genotypes are probably the most different when it comes to gene expression. Meanwhile, the orange- and white-fleshed sweet potatoes have the least amount of significantly up-regulated genes between them in their genome profiles.

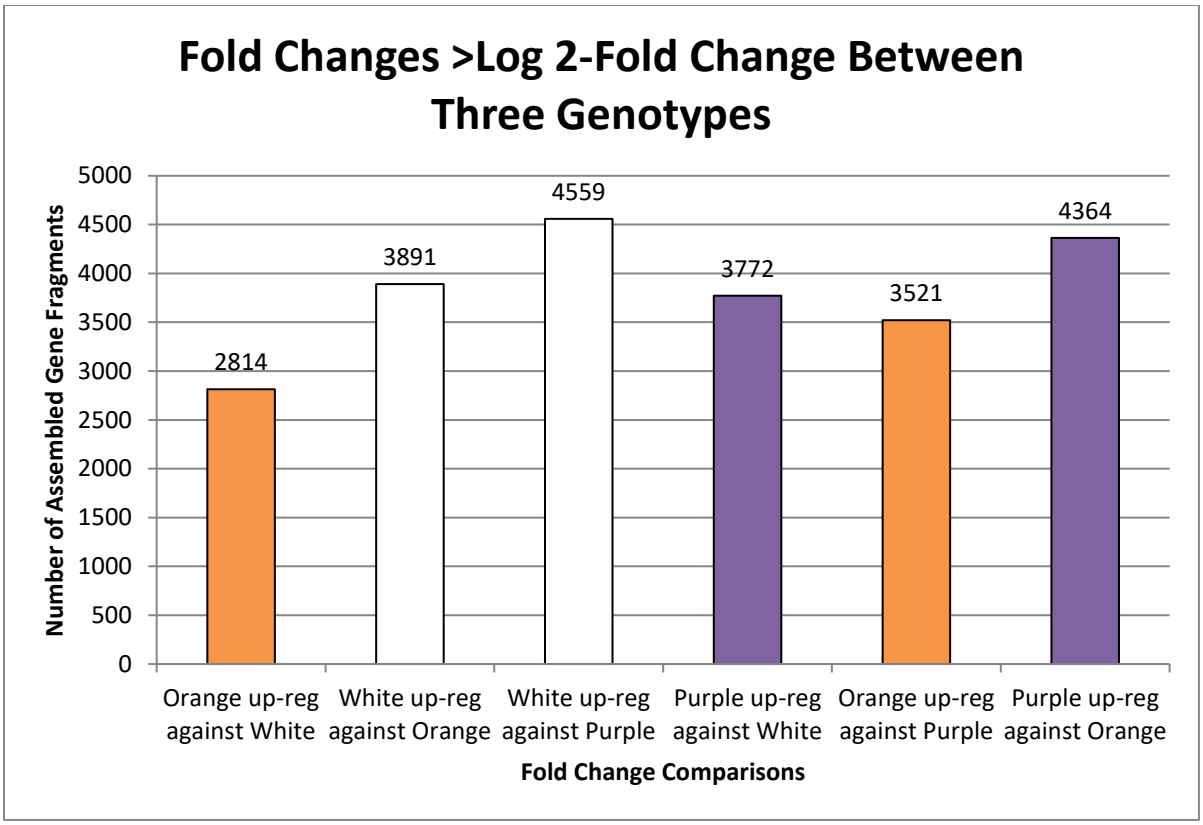


Figure 8: Comparison of Up-Regulated Genes Between Genotypes. Up-regulation of at least log 2-fold change in comparisons of orange-fleshed genotypes to white-fleshed genotypes, white-fleshed genotypes to purple-fleshed genotypes, and orange-fleshed genotypes to purple-fleshed genotypes.

Alongside the 78,781 predicted genes, the gene functions are indicated in adjacent columns after being transferred into Microsoft Excel, and through a system of data mining, we were able to link these predicted genes, to different functional groups such as disease resistance genes, pigment genes, stress-responsive genes, etc. This data mining process involved identifying key words relating to functional groups like pigments, disease resistance, abiotic stress tolerance, transcription factors, etc., which were used to search the fold change data and determine the differences in gene expression for each of the fold change comparisons made. Using the search function in Microsoft Excel, we were able to efficiently and quickly search through all of the predicted genes so that this data could be sorted into these different functional groups indicated in the horizontal axis in **Figure 9**. This figure displays the amount of genes in each functional group that have a log 2-fold change equal to or greater than 2 (>4 fold change) as

a result of each of the fold changes carried out. There is some overlap, as some genes can be included in more than one functional group. For example, a gene relating to the synthesis of chlorophyll could be grouped into both the Energy and Pigment functional groups.

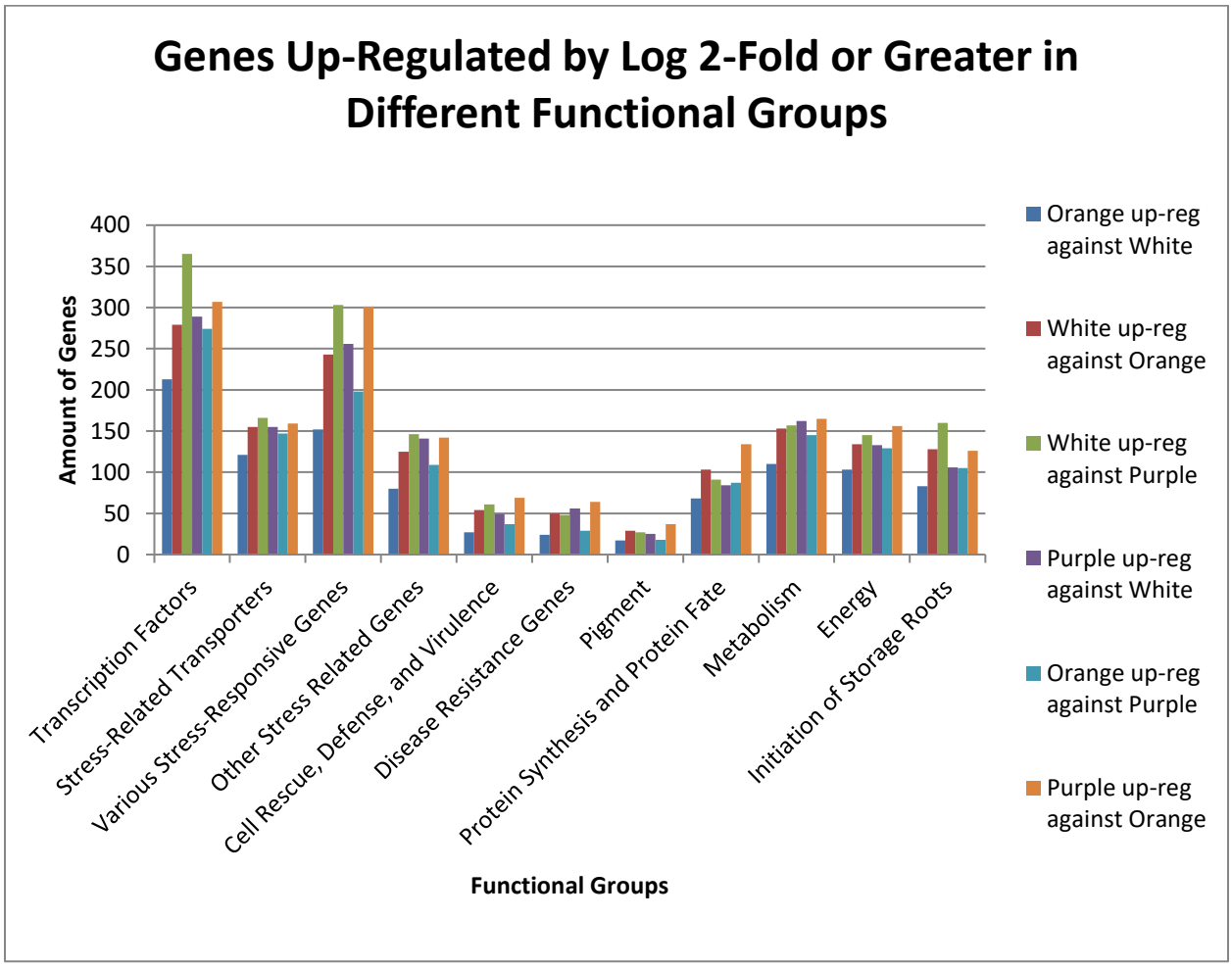


Figure 9: Up-Regulated Genes of Functional Groups. > log 2-fold up-regulated genes for different functional groups.

Table 7 contains the top five up-regulated genes in each of the two genotypes in each fold-change comparison made (5 Genes X 2 Genotypes X 3 Fold Change Comparisons = 30 Genes). The function of nine of these 30 genes are currently unable to be predicted. This means that Yang et al. (2017) was unable to match these contigs to homologous genes in the NCBI and UniProt databases, in their original sequencing work, so while our reads could be aligned with their gene, a gene function could not be matched (NCBI, 2018) (UniProt, 2018). As was previously mentioned, the fold-change values have been converted to log2-fold change.

As can be seen, many of these genes are labeled as "Predicted," because these genes were matched up to similar genes which had already been defined with functions through past research on other organisms. This is also why many of these gene descriptions contain scientific names in brackets, to indicate the model organism that has been used in past experiments to prove the indicated function of this gene. Until a gene's function is certified in sweet potato, it must remain merely as "predicted," since the only evidence we have for function is gene homology.

In **Table 7**, we can see that in the first fold change comparison (first and second sections of the chart) two genes having to do with translation mechanisms are upregulated and another involved in the Calvin Cycle were up-regulated in the orange-fleshed genotype when it was compared with the white-fleshed genotype. The two other genes in this first section of the table remain unidentified. The first gene having to do with transcription was labeled with "elongation factor 1-beta," meaning that it has to do with the mechanism of binding aminoacyl-tRNAs (tRNAs with attached amino acids) to the A-site on the ribosome, while the other gene has to do with the manufacturing of a subunit of a ribosomal protein (EMBL-EBI, 2018). The gene involved in the Calvin Cycle is labeled with "fructose-bisphosphate aldolase cytoplasmic isozyme-like," meaning that this gene likely has to do with glycolysis and gluconeogenesis,

which is likely to be heavily involved in various stress-responses (Lv, et al., 2017). The up-regulated genes in the white sweet potato included two currently unidentified genes. One of these genes is predicted to have to do with chloroplasts. However, the function of this gene, "uncharacterized aarF domain-containing protein kinase At1g79600", is not yet clear, it is merely predicted to have some regulating power over UbiI, meaning that it could potentially control if and when aerobic coenzyme Q biosynthesis occurs (UniProt, 2017). Another gene is labeled as a probable aquaporin, a gene likely contributing to the construction of a cellular water channel, which is often up-regulated in response to drought conditions (Zargar, et al., 2017). The final one able to be predicted is labeled as a "Kunitz-type trypsin inhibitor," literally meaning that it halts the enzymatic function of trypsin, or stops it from breaking down proteins (Bendre, Suresh, Shanmugam, & Ramasamy, 2018). Kunitz-type inhibitors are an important factor in biotic stresses or insects and other pests (Bendre, Suresh, Shanmugam, & Ramasamy, 2018).

The second fold change comparison (third and fourth sections of **Table 7**) shows us the most up-regulated genes in white-fleshed and purple-fleshed genotypes when these two are compared. The highest up-regulated gene in the white genotype was labeled "RD22-like protein," which the UniProt database identifies as a protein that is commonly activated when dehydration in the plant occurs, suggesting that whatever the currently unknown function is, it is in all likelihood, a drought-tolerant gene (UniProt, 2017). This is confirmed to some extent being that the next highest up-regulated gene is also labelled as a "dehydration-responsive protein" and the fifth highest up-regulated gene is labeled as a mitochondrial subunit of "pyruvate dehydrogenase E1," which has been shown to be up-regulated in drought-tolerant barley genotypes (Guo, et al., 2009). The third highest up-regulated gene in the white-fleshed genotype is glycerate dehydrogenase HPR, peroxisomal, which is crucial for the photorespiratory

cycle during the conversion of hydroxypyruvate to glycerate (UniProt, 2018b). The fourth up-regulated gene for the white genotype in this comparison is a Ribonuclease (RNase)-like protein. Ribonucleases are important for natural processes like maturation of other RNA molecules and degrading mRNA molecules that are no longer needed to be translated at that time. Meanwhile, ribonucleases are a major defense against viruses and other pathogens (BiochemDen, 2018). The first and fourth genes are the only ones identified out of the five highest up-regulated genes in the purple-fleshed genotype when compared against the white-fleshed genotype. The highest regulated gene is labeled as "hypersensitive-induced response protein 1," meaning that it is a positive regulator of cell-death, which is necessary to be carried out for both old and infected cells (UniProt, 2018c). The fourth up-regulated gene for the purple-fleshed genotype labeled "ubiquitin-60S ribosomal protein L40," which means that it is a gene which instructs the formation of a ribosomal subunit (Next Prot, 2018).

The third fold change comparison (fifth and sixth sections of **Table 7**) contains the most up-regulated genes in orange-fleshed and purple-fleshed genotypes from when these two genotypes were compared. The highest expressed gene in the orange-fleshed genotype is proline-rich protein 4, which has been shown to be up-regulated in early lateral root formation in Arabidopsis (Fowler, Bernhardt, & Tierney, 1999). Glycerate dehydrogenase HPR, peroxisomal, is the second highest orange-fleshed up-regulated gene which was also one of the highest up-regulated genes mentioned in the previous fold change comparison in the white sweet potato profile. This gene is part of the photorespiratory cycle which sees the conversion of hydroxypyruvate to glycerate (UniProt, 2018b). "Alpha-glucan phosphorylase, H isozyme isoform XI" is the third highest up-regulated gene for orange sweet potatoes, which is an important enzyme for carbohydrate metabolism (UniProt, 2017). The fourth highest up-

regulated gene for orange-fleshed sweet potato for this comparison is also up-regulated in the previous comparison for the white sweet potatoes (dehydration-responsive protein RD22). The last gene in **Table 7** that is up-regulated is heavily involved in protein transport (Ras-regulated protein RABD1-like), especially between the Golgi apparatus and the endoplasmic reticulum (UniProt, 2018). The highest up-regulated and the third highest up-regulated genes for the purple genotype when it is compared against orange are unidentified. The second is labeled as "sulfiredoxin, chloroplastic/mitochondrial," which aids in the case of oxidative stress (UniProt, 2018d). The fourth highest up-regulated gene was also one of the five highest up-regulated genes in the white genotype during its fold change comparison with the orange genotype and controls the biosynthesis of coenzyme Q. The final up-regulated gene for the purple genotype here is Fen (Flap Endonuclease)-interacting protein 3, which cleaves specific nucleotides and along with the help of other enzymes, possibly helps to guard against duplication mutations in DNA (Liu, Qiu, Finger, Zheng, & Shen, 2006).

Feature ID	Log 2-Fold Change	Description
Orange up-regulated against White		
G35591|TU58360(56.0)::chr8:2864259-2866482(+)	9.2890	PREDICTED: elongation factor 1-beta-like [*Nicotiana sylvestris*]
G16512|TU26978(21.1)::chr14:46273757-46275226(+)	9.1951	PREDICTED: 40S ribosomal protein S8-like [*Solanum tuberosum*]
G48714|TU78206(1.8)::scaffold8940|size6732:18-1856(+)	9.0178	
G34273|TU56195(54.7)::chr7:17482979-17484933(-)	8.9067	
G38771|TU63543(141.7)::chr9:9777771-9778233(-)	8.4582	PREDICTED: fructose-bisphosphate aldolase cytoplasmic isozyme-like [*Sesamum indicum*]
White up-regulated against Orange		
G13930|TU22754(432.5)::chr14:5165896-5167590(+)	9.8567	
G25485|TU41849(7.4)::chr4:10711280-10713652(+)	9.7319	PREDICTED: uncharacterized aarF domain-containing protein kinase At1g79600, chloroplastic [*Nicotiana sylvestris*]
G37406|TU61276(9.3)::chr8:31562184-31563698(-)	9.3503	PREDICTED: probable aquaporin TIP1-1 [*Nicotiana tomentosiformis*]
G13337|TU21827(63.5)::chr13:38859439-38860441(+)	9.0056	Kunitz-type trypsin inhibitor n=1 Tax=*Ipomoea batatas* RepID=
G44915|TU72600(0.3)::scaffold2665|size76638:0-1932(-)	9.0056	
White up-regulated against Purple		
G4244|TU6996(0.7)::chr10:29406932-29408805(+)	11.7291	RD22-like protein n=1 Tax=*Camellia sinensis* RepID=
G4281|TU7069(2.3)::chr10:29985984-29988278(+)	9.6922	PREDICTED: dehydration-responsive protein RD22 [*Nicotiana sylvestris*]
G16401|TU26776(4.3)::chr14:45152914-45153595(+)	9.4510	PREDICTED: glycerate dehydrogenase HPR, peroxisomal [*Sesamum indicum*]
G37014|TU60698(6.3)::chr8:24136457-24196828(+)	9.3398	RNase-like protein n=1 Tax=*Calystegia sepium* RepID=
G34800|TU57074(0.9)::chr7:27195781-27197339(+)	9.2318	PREDICTED: pyruvate dehydrogenase E1 component subunit alpha, mitochondrial [*Sesamum indicum*]
Purple up-regulated against White		
G2384|TU3963(13.6)::chr1:34564218-34595052(-)	9.5248	PREDICTED: hypersensitive-induced response protein 1 [*Sesamum indicum*]
G34273|TU56195(54.7)::chr7:17482979-17484933(-)	9.1728	
G26515|TU43560(2.4)::chr4:25005261-25007057(+)	9.1075	
G31837|TU52226(40.0)::chr6:14388810-14390649(-)	8.7610	PREDICTED: ubiquitin-60S ribosomal protein L40 [*Nicotiana tomentosiformis*]
G18189|TU29733(32.4)::chr15:15173099-15291057(-)	8.6596	
Orange up-regulated against Purple		
G33415|TU54771(26.1)::chr7:2000676-2009707(+)	10.8120	PREDICTED: proline-rich protein 4 [*Nicotiana sylvestris*]
G16401|TU26776(4.3)::chr14:45152914-45153595(+)	9.9737	PREDICTED: glycerate dehydrogenase HPR, peroxisomal [*Sesamum indicum*]
G33597|TU55078(18.8)::chr7:4136879-4137320(-)	9.5824	PREDICTED: alpha-glucan phosphorylase, H isozyme isoform X1 [*Nicotiana sylvestris*]
G4281|TU7069(2.3)::chr10:29985984-29988278(+)	9.5315	PREDICTED: dehydration-responsive protein RD22 [*Nicotiana sylvestris*]
G30988|TU50792(4.1)::chr6:4594513-4595428(-)	9.2561	PREDICTED: ras-related protein RABD1-like [*Nicotiana sylvestris*]
Purple up-regulated against Orange		
G38162|TU62538(36.5)::chr9:1418820-1420397(-)	9.0710	Uncharacterized protein n=5 Tax=core eudicotyledons RepID=
G18344|TU29975(4.4)::chr15:18956215-18958694(+)	9.0010	PREDICTED: sulfiredoxin, chloroplastic/mitochondrial-like [*Solanum tuberosum*]
G11744|TU19337(1.6)::chr13:7667029-7670489(+)	8.9533	
G25485|TU41849(7.4)::chr4:10711280-10713652(+)	8.8376	PREDICTED: uncharacterized aarF domain-containing protein kinase At1g79600, chloroplastic [*Nicotiana sylvestris*]
G14501|TU23692(49.8)::chr14:15190383-15191307(-)	8.5923	Fen-interacting protein 3 [*Solanum lycopersicum*]

Table 7: Highest Log 2-Fold Changes Detected with CLC Bio Software. The Fold-Change function in the CLC Bio software was used three times to compare the orange-fleshed sweet potato sequencing results to those of the white-fleshed sweet potato, the white-fleshed sweet potato sequencing results to those of the purple-fleshed sweet potato, and the orange-fleshed sweet potato sequencing results to those of the purple-fleshed sweet potato. The fold-changes were all converted to log 2-fold changes. This table displays the five highest log 2-fold changes of each of the two genotypes in each of the three comparisons.

From this log 2-fold change data, we were able to group some of these genes into a stress responsive gene category and the results of each of the three fold changes amongst these grouped genes are displayed in **Table 8**. The first two sections are the results of the first fold change

(orange-fleshed genotype vs white-fleshed genotype). The third and fourth sections are the second fold change (white-fleshed genotype vs purple-fleshed genotype). The third fold change results (orange-fleshed genotype vs purple-fleshed genotype) are contained in sections five and six. Each of these sections contains the top five up-regulated genes in each comparison, expressed as the log 2-fold change. This table is set up similarly to **Table 7**, but all these genes either are or are directly linked to various stress responsive genes. The way we were able to find these genes was through data mining, as described in the Bioinformatics Analysis section of the materials and methods section, using keywords involving stress response genes (eg. zinc finger, aquaporin, LRR, heat shock protein, RING, etc.) (Ayyappan, et al., 2017).

When the orange and white genotypes are compared with the CLC fold change function, it is revealed that there are 152 significantly up-regulated (>2 log 2-fold change) genes in the orange genotype and 243 significantly up-regulated genes in the white genotype directly related to stress responses. The next fold change function carried out was between the white and purple sweet potato genotypes and there were 303 significantly up-regulated genes in the white sweet potato along with 256 significantly up-regulated genes in the purple sweet potato. The final fold change comparison was between the orange and purple genotypes with 198 and 301 genes were significantly up-regulated respectively. The highlighted gene indicates that this was one of the highest upregulated genes overall included in **Table 7**.

Various Stress Responsive Genes		
Feature ID	**Log 2-Fold Change**	**Description**
Orange up-regulated against White		
G46243|TU74519(25.9)::scaffold347|size381502:192798-200749(+)	7.6467	PREDICTED: BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1-like [*Nicotiana sylvestris*]
G44002|TU71369(1.6)::scaffold2245|size99336:38-374(-)	7.4164	PREDICTED: zinc finger BED domain-containing protein RICESLEEPER 1-like [*Sesamum indicum*]
G36336|TU59595(5.2)::chr8:11947165-11949130(+)	6.0760	PREDICTED: chaperone protein dnaJ 6-like [*Nicotiana tomentosiformis*]
G3510|TU5760(0.4)::chr10:16584236-16585100(-)	6.0196	PREDICTED: zinc finger BED domain-containing protein RICESLEEPER 2-like [*Nicotiana sylves*tris]
G11451|TU18866(18.3)::chr13:1535886-1539809(-)	5.9157	PREDICTED: heat shock protein 90-1-like [*Sesamum indicum*]
White up-regulated against Orange		
G37406|TU61276(9.3)::chr8:31562184-31563698(-)	9.3503	PREDICTED: probable aquaporin TIP1-1 [*Nicotiana tomentosiformis*]
G15943|TU26048(10.6)::chr14:38285435-38289604(-)	7.3236	PREDICTED: probable LRR receptor-like serine/threonine-protein kinase At1g56140 isoform X1 [*Nicotiana tomentosiformis*]
G6622|TU10919(2.9)::chr11:20051739-20053986(+)	6.7786	PREDICTED: zinc finger CCCH domain-containing protein 48 [*Nicotiana sylvestris*]
G17844|TU29168(10.9)::chr15:9450732-9786115(+)	6.5822	PREDICTED: transcription factor MYB44-like [*Nicotiana sylvestris*]
G10690|TU17613(2.9)::chr12:46132473-46134545(-)	6.4729	PREDICTED: probably inactive leucine-rich repeat receptor-like protein kinase At5g48380 [*Nicotiana tomentosiformis*]
White up-regulated against Purple		
G40380|TU66240(4.3)::scaffold10249|size5466:15-883(+)	6.9654	PREDICTED: F-box/FBD/LRR-repeat protein At1g13570-like [*Nicotiana sylvestris*]
G9206|TU15159(1.7)::chr12:25550289-25557031(+)	6.8470	PREDICTED: DNA-directed RNA polymerase I subunit rpa1 [*Nicotiana tomentosiformis*]
G10690|TU17613(2.9)::chr12:46132473-46134545(-)	6.3871	PREDICTED: probably inactive leucine-rich repeat receptor-like protein kinase At5g48380 [*Nicotiana tomentosiformis*]
G2910|TU4798(2.8)::chr10:6266989-6269085(-)	6.2087	PREDICTED: zinc finger BED domain-containing protein RICESLEEPER 2-like [*Nicotiana sylvestris*]
G22208|TU36303(5.7)::chr2:47175922-47215437(-)	6.1001	PREDICTED: expansin-A4-like [*Nicotiana tomentosiformis*]
Purple up-regulated against White		
G11451|TU18866(18.3)::chr13:1535886-1539809(-)	7.5984	PREDICTED: heat shock protein 90-1-like [*Sesamum indicum*]
G30033|TU49213(1.2)::chr5:37567124-37567429(-)	7.5354	PREDICTED: probable LRR receptor-like serine/threonine-protein kinase At1g07650-like [*Solanum tuberosum*]
G44002|TU71369(1.6)::scaffold2245|size99336:38-374(-)	7.4849	PREDICTED: zinc finger BED domain-containing protein RICESLEEPER 1-like [*Sesamum indicum*]
G31260|TU51267(17.5)::chr6:4565010-4567527(-)	7.4610	PREDICTED: stromal 70 kDa heat shock-related protein, chloroplastic [*Solanum lycopersicum*]
G3510|TU5760(0.4)::chr10:16584236-16585100(-)	6.6098	PREDICTED: zinc finger BED domain-containing protein RICESLEEPER 2-like [*Nicotiana sylvestris*]
Orange up-regulated against Purple		
G40380|TU66240(4.3)::scaffold10249|size5466:15-883(+)	6.4219	PREDICTED: F-box/FBD/LRR-repeat protein At1g13570-like [*Nicotiana sylvestris*]
G17353|TU28386(8.5)::chr15:4425014-4426535(-)	6.3695	PREDICTED: beta-amylase 8-like, partial [*Solanum tuberosum*]
G9206|TU15159(1.7)::chr12:25550289-25557031(+)	6.0740	PREDICTED: DNA-directed RNA polymerase I subunit rpa1 [*Nicotiana tomentosiformis*]
G39547|TU64877(0.4)::chr9:21743253-21745835(+)	5.9838	PREDICTED: probable WRKY transcription factor 26 [*Sesamum indicum*]
G47143|TU75876(1.8)::scaffold500|size414498:126694-127461(+)	5.9604	PREDICTED: F-box/LRR-repeat protein 15 [*Nicotiana tomentosiformis*]
Purple up-regulated against Orange		
G35328|TU57918(7.0)::chr7:32489204-32491769(-)	7.3308	PREDICTED: cyclin-B1-2-like [*Nicotiana sylvestris*]
G27457|TU45017(1.7)::chr4:46757350-46761030(-)	6.8892	PREDICTED: RING finger and CHY zinc finger domain-containing protein 1 [*Sesamum indicum*]
G8605|TU14153(4.8)::chr12:8097274-8099691(-)	6.4861	PREDICTED: heat shock cognate protein 80 [*Sesamum indicum*]
G12004|TU19733(2.1)::chr13:11320325-11322612(+)	6.0956	PREDICTED: probable LRR receptor-like serine/threonine-protein kinase At3g47570-like [*Solanum tuberosum*]
G46555|TU74970(0.6)::scaffold3846|size37512:24408-25573(-)	5.9736	NBS-LRR type disease resistance protein (Fragment)

Table 8: Various Stress Responsive Genes. After the Fold-Change function in CLC Bio was utilized to rate expression differences overall between two genotypes at a time, we were able to sort some of these genes into the

Various Stress Responsive Genes functional group (Ayyappan, et al., 2017). The fold-changes were all converted to log 2-fold changes. This table displays the various stress responsive genes with the five highest log 2-fold changes of each of the two genotypes in each of the three comparisons.

Table 9 contains the top five pigment related genes that are upregulated in each fold change comparison, formatted the same way as **Table 8**. Each of these genes were linked to pigments through data mining using pigment-specific keywords (eg. anthocyanin, chloroplast, chlorophyll, purple, etc.) (The Gale Encyclopedia of Science, 2016).

Only 17 genes were found to be significantly up-regulated (>2 log 2-fold change) in the orange genotype, while 29 were significantly up-regulated in the white genotype in the first fold change comparison. The second fold change carried out in CLC revealed that 27 and 25 genes were significantly up-regulated in white and purple sweet potato genotypes respectively. Finally, for the final fold change comparison orange vs. purple indicated that 18 genes were significantly up-regulated in the orange genotype and 37 genes were significantly up-regulated in the purple genotype.

Pigment Genes		
Feature ID	**Log 2-Fold Change**	**Description**
Orange up-regulated against White		
G45725|TU73746(4.8)::scaffold31167|size1262:453-1245(-)	5.6752	1-O-acylglucose:anthocyanin-O-acyltransferase-like protein n=1 Tax=*Gentiana triflora* RepID
G12946|TU21218(1.5)::chr13:32412639-32413387(+)	4.4947	Putative anthocyanin regulator n=1 Tax=*Ipomoea hederacea* RepID
G6361|TU10483(2.3)::chr11:16078267-16082867(+)	4.0472	PREDICTED: zeaxanthin epoxidase, chloroplastic [*Nicotiana sylvestris*]
G3121|TU5139(1.2)::chr10:9045357-9046703(+)	3.8773	carotenoid isomerase [*Nicotiana tabacum*]
G45273|TU73111(2.0)::scaffold28579|size1705:362-794(+)	3.7577	Zeta-carotene desaturase n=1 Tax=Ipomoea sp. Kenyan RepID
White up-regulated against Orange		
G35948|TU58934(25.9)::chr8:7198733-7200506(+)	6.0795	PREDICTED: chlorophyll a-b binding protein CP26, chloroplastic [*Sesamum indicum*]
G34094|TU55916(2.7)::chr7:12610271-12623738(+)	5.7422	PREDICTED: protochlorophyllide-dependent translocon component 52, chloroplastic [*Sesamum indicum*]
G41507|TU67869(2.8)::scaffold13671|size3859:834-2997(-)	4.9944	PREDICTED: carotenoid 9,10(9',10')
G5142|TU8464(23.7)::chr11:2464218-2465771(+)	4.6787	chlorophyll a-b binding protein 13, chloroplastic-like protein *[Nicotiana tabacum*]
G11162|TU18375(0.9)::chr13:892364-892998(-)	4.3924	PREDICTED: probable inactive purple acid phosphatase 29 isoform X1 [*Nicotiana sylvestris*]
White up-regulated against Purple		
G34094|TU55916(2.7)::chr7:12610271-12623738(+)	6.5603	PREDICTED: protochlorophyllide-dependent translocon component 52, chloroplastic [*Sesamum indicum*]
G46673|TU75149(0.5)::scaffold4032|size33464:17944-18653(+)	4.8723	1-O-acylglucose:anthocyanin-O-acyltransferase-like protein n=1 Tax=*Gentiana triflora* RepID
G3517|TU5772(1.6)::chr10:16732125-16738235(+)	4.0268	putative red chlorophyll catabolite reductase [*Nicotiana tabacum*]
G28813|TU47257(8.9)::chr5:17993661-17995425(-)	3.9178	PREDICTED: chlorophyll a-b binding protein CP29.1, chloroplastic [*Sesamum indicum*]
G32701|TU53562(84.4)::chr6:27746177-27929237(-)	3.5772	PREDICTED: chlorophyll a-b binding protein 21, chloroplastic-like [*Sesamum indicum*]
Purple up-regulated against White		
G4181|TU6907(1.1)::chr10:28060999-28062034(-)	4.9844	PREDICTED: carotenoid 9,10(9',10')
G45725|TU73746(4.8)::scaffold31167|size1262:453-1245(-)	4.8801	1-O-acylglucose:anthocyanin-O-acyltransferase-like protein n=1 Tax=Gentiana triflora RepID
G10953|TU18049(1.0)::chr12:51329785-51330862(+)	4.1696	PREDICTED: probable inactive purple acid phosphatase 27 [*Nicotiana tomentosiformis*]
G11164|TU18380(8.6)::chr13:748184-750104(-)	3.4326	PREDICTED: probable inactive purple acid phosphatase 29 [*Solanum lycopersicum*]
G12946|TU21218(1.5)::chr13:32412639-32413387(+)	3.3655	Putative anthocyanin regulator n=1 Tax=*Ipomoea hederacea* RepID
Orange up-regulated against Purple		
G3121|TU5139(1.2)::chr10:9045357-9046703(+)	4.1671	carotenoid isomerase [*Nicotiana tabacum*]
G3517|TU5773(0.9)::chr10:16732125-16733387(+)	3.4891	putative red chlorophyll catabolite reductase [*Nicotiana tabacum*]
G6361|TU10483(2.3)::chr11:16078267-16082867(+)	3.4891	PREDICTED: zeaxanthin epoxidase, chloroplastic [*Nicotiana sylvestris*]
G40868|TU66958(3.9)::scaffold11589|size4596:1-4577(-)	3.3150	PREDICTED: probable inactive purple acid phosphatase 27 isoform X1 [*Nicotiana tomentosiformis*]
G32701|TU53562(84.4)::chr6:27746177-27929237(-)	3.1076	PREDICTED: chlorophyll a-b binding protein 21, chloroplastic-like [*Sesamum indicum*]
Purple up-regulated against Orange		
G35948|TU58934(25.9)::chr8:7198733-7200506(+)	5.8283	PREDICTED: chlorophyll a-b binding protein CP26, chloroplastic [*Sesamum indicum*]
G17607|TU28806(0.7)::chr15:6720621-6730746(+)	5.2056	PREDICTED: zeaxanthin epoxidase, chloroplastic-like [*Nicotiana tomentosiformis*]
G10953|TU18049(1.0)::chr12:51329785-51330862(+)	4.7517	PREDICTED: probable inactive purple acid phosphatase 27 [*Nicotiana tomentosiformis*]
G15072|TU24603(5.0)::chr14:22184063-22213929(+)	4.6214	PREDICTED: probable inactive purple acid phosphatase 1-like [*Solanum tuberosum*]
G16438|TU26850(0.4)::chr14:45453384-45454964(+)	4.4939	PREDICTED: probable inactive purple acid phosphatase 1-like [*Solanum tuberosum*]

Table 9: Pigment Genes. After the Fold-Change function in CLC Bio was utilized to rate expression differences overall between two genotypes at a time, we were able to sort some of these genes into the Pigment Genes functional group (The Gale Encyclopedia of Science, 2016). The fold-changes were all converted to log 2-fold changes. This table displays the pigment genes with the five highest log 2-fold changes of each of the two genotypes in each of the three comparisons.

Genes that were related to the initiation of storage root genes were also found through data mining procedures, with keywords for genes that have been demonstrated to be heavily involved in the sweet potato vegetable root development (Firon, et al., 2013). The top five up-regulated genes related to the initiation of storage root genes for each genotype, between each fold change comparison are displayed in **Table 10**.

Through data mining, we found that the first fold change comparison between orange and white sweet potato plants indicated that 83 genes were significantly up-regulated (>2 log 2-fold change) in the orange genotype and that 128 genes were significantly up-regulated in the white genotype. The second fold-change comparison between the white and purple sweet potato plants showed 160 significantly up-regulated genes in the white genotype and 106 significantly up-regulated genes in the purple genotype. The final fold change here indicates that there were 105 and 126 significantly up-regulated genes in the orange and purple genotypes, respectively. The two highlighted genes indicate that these were also in the top five up-regulated genes for each fold change comparison displayed in **Table 7.**

Initiation of Storage Root Genes		
Feature ID	**Log 2-Fold Change**	**Description**
Orange up-regulated against White		
G38771|TU63543(141.7)::chr9:9777771-9778233(-)	8.4622	PREDICTED: fructose-bisphosphate aldolase cytoplasmic isozyme-like [*Sesamum indicum*]
G32976|TU54011(20.2)::chr6:40881698-40882304(+)	6.0734	Cytochrome b6-f complex iron-sulfur subunit n=1 Tax=*Populus trichocarpa* RepID
G5773|TU9512(0.7)::chr11:10040079-10040464(-)	5.6370	PREDICTED: probable caffeoyl-CoA O-methyltransferase At4g26220 [*Nicotiana sylvestris*]
G19358|TU31634(4.9)::chr2:6657450-6660014(+)	5.6175	PREDICTED: ADP-ribosylation factor-like protein 8B isoform X1 [*Nicotiana tomentosiformis*]
G6438|TU10612(3.8)::chr11:17552654-17554340(+)	5.0903	PREDICTED: E3 ubiquitin-protein ligase At3g02290-like [*Nicotiana sylvestris*]
White up-regulated against Orange		
G34367|TU56356(1925.1)::chr7:19506263-19510374(+)	7.2260	Sporamin n=1 Tax=*Ipomoea batatas* RepID
G29992|TU49139(0.8)::chr5:37212343-37219991(-)	6.2067	Coumarin glucosyltransferase 1 n=1 Tax=*Ipomoea nil* RepID
G8403|TU13834(37.6)::chr12:10672981-10694522(+)	5.5073	PREDICTED: peroxidase 21 [*Nicotiana sylvestris*]
G46475|TU74850(2.7)::scaffold3731|size39962:27944-30342(-)	5.5050	beta-expansin precursor [*Solanum lycopersicum*]
G33355|TU54684(0.4)::chr7:824683-826419(-)	5.3895	PREDICTED: peroxidase 39-like [*Solanum tuberosum*]
White up-regulated against Purple		
G21551|TU35226(1.9)::chr2:44344865-44355334(+)	9.5005	PREDICTED: iron-sulfur assembly protein IscA-like 1, mitochondrial [*Nicotiana tomentosiformis*]
G42816|TU69748(24.4)::scaffold18033|size3201:1056-2454(-)	7.1766	ADP-ribosylation factor-like 8a [*Nicotiana tabacum*]
G5761|TU9492(4.6)::chr11:7837270-7839222(-)	6.5660	PREDICTED: UDP-galactose/UDP-glucose transporter 3 [*Nicotiana tomentosiformis*]
G23113|TU37813(10.3)::chr3:19341469-19345276(+)	6.4992	ascorbate peroxidase [*Nicotiana tabacum*]
G12023|TU19757(9.1)::chr13:11499533-11501885(-)	5.6993	PREDICTED: ADP-ribosylation factor GTPase-activating protein AGD3-like isoform X2 [*Solanum tuberosum*]
Purple up-regulated against White		
G38771|TU63543(141.7)::chr9:9777771-9778233(-)	8.5006	PREDICTED: fructose-bisphosphate aldolase cytoplasmic isozyme-like [*Sesamum indicum*]
G18065|TU29523(1.3)::chr15:14263543-14265386(+)	6.5581	PREDICTED: E3 ubiquitin-protein ligase CCNB1IP1 homolog isoform X3 [*Solanum lycopersicum*]
G7411|TU12193(0.9)::chr11:41113331-41114162(+)	6.4723	PREDICTED: beta-amylase isoform X1 [*Nicotiana tomentosiformis*]
G41570|TU67959(9.7)::scaffold13942|size3816:2480-3484(+)	5.9418	PREDICTED: peroxidase 15-like [*Nicotiana sylvestris*]
G10295|TU16935(21.7)::chr12:42313438-42320609(-)	5.8249	PREDICTED: putative E3 ubiquitin-protein ligase XBAT31 [*Nicotiana sylvestris*]
Orange up-regulated against Purple		
G21551|TU35226(1.9)::chr2:44344865-44355334(+)	9.8720	PREDICTED: iron-sulfur assembly protein IscA-like 1, mitochondrial [*Nicotiana tomentosiformis*]
G12023|TU19757(9.1)::chr13:11499533-11501885(-)	6.4972	PREDICTED: ADP-ribosylation factor GTPase-activating protein AGD3-like isoform X2 [*Solanum tuberosum*]
G17353|TU28386(8.5)::chr15:4425014-4426535(-)	6.3695	PREDICTED: beta-amylase 8-like, partial [*Solanum tuberosum*]
G5761|TU9492(4.6)::chr11:7837270-7839222(-)	6.0183	PREDICTED: UDP-galactose/UDP-glucose transporter 3 [*Nicotiana tomentosiformis*]
G42816|TU69748(24.4)::scaffold18033|size3201:1056-2454(-)	5.9485	ADP-ribosylation factor-like 8a [*Nicotiana tabacum*]
Purple up-regulated against Orange		
G15695|TU25650(2.8)::chr14:33294251-33297330(-)	7.5672	PREDICTED: probable E3 ubiquitin-protein ligase HERC2 [*Nicotiana sylvestris*]
G34367|TU56356(1925.1)::chr7:19506263-19510374(+)	5.9704	Sporamin n=1 Tax=*Ipomoea batatas* RepID
G48695|TU78171(1.1)::scaffold88|size591879:551547-555799(+)	5.7777	Sporamin B n=1 Tax=*Ipomoea batatas* RepID
G5691|TU9395(7.5)::chr11:6176185-6177451(+)	5.7777	PREDICTED: thiosulfate sulfurtransferase 18-like isoform X1 [*Nicotiana sylvestris*]
G34367|TU56358(199.5)::chr7:19506263-19510374(+)	5.7022	Sporamin n=1 Tax=*Ipomoea batatas* RepID

Table 10: Initiation of Storage Root Genes. After the Fold-Change function in CLC Bio was utilized to rate expression differences overall between two genotypes at a time, we were able to sort some of these genes into the Initiation of Storage Root Genes functional group (Firon, et al., 2013). The fold-changes were all converted to log 2-fold changes. This table displays the initiation of storage root genes with the five highest log 2-fold changes of each of the two genotypes in each of the three comparisons.

4.3. SEQUENCING VERIFICATION WITH REAL TIME PCR

In order to ensure overall accurate expression levels, it was necessary to perform a sort of "audit," using real-time PCR, as is seen in many sequencing papers. This means quantifying a couple of select genes to check and make sure that the expression levels match the fold changes indicated by the CLC program. The way that we selected these genes, was to select the ones with the highest difference in expression, or highest fold-change. We did this for each fold change comparison that we set up. For example, let us say that the highest fold change while comparing the orange-fleshed sweet potato to the white-fleshed sweet potato was 500, with orange having the higher expression. We would then select the highest fold change for the white sweet potato as well. By selecting from the two extreme ends of ends of the fold change table, we would be able to see the most expansive differences in gene expression between genotypes through qPCR. Therefore, primers were designed for these genes, as can be seen in the last section of **Table 1**, using NCBI's primer design software (NCBI, 2018).

Real-time PCR was carried out following the default parameters of the real time PCR machine (Step 1: 50^0C two min, Step 2: 95^0C 10 min, Step 3: 95^0C 15sec and 60^0C 1 min (X40 cycles), Step 4: 95^0C 15 sec, 60^oC 1 min, 95^oC 30 sec, and 60^oC 15 sec) and unlike in the TPM gene expression values of the sequencing data, the lower the CT value for real-time PCR actually means that a higher expression has occurred for the corresponding gene to the utilized primer set. This is because a lower CT value essentially means that a lower amount of PCR cycles were needed to cross the expression threshold, meaning that the reaction began with a higher amount of target sequence. As can be seen in **Table 11**, only two of our sequencing data expression value patterns correspond to our real time expression results. PCR for oliogs of G33415 indicate that expression for this gene is highest in the orange sweet potato genotype and lowest in the

purple sweet potato genotype, which is exactly what the sequencing data indicates. The other PCR oligo set that shows similar expression patterns between the sequencing data and the real-time PCR data is G4244, with the highest expression value in the white sweet potato plant and the lowest expression value in the purple sweet potato plant. Each of the other real-time PCR results do not correlate to the sequencing data.

Primers	Expression Values of Sequencing Data (TPM)			Expression Results of Real-Time PCR (CT)		
	Orange	White	Purple	Orange	White	Purple
G28	0.04	28.1	38.76	28.96	30.72	29.75
G33415	84.43	22.96	0.05	24.73	27.16	31.2
G35591	102.34	0.16	58.22	22.37	23.58	24.65
G4244	18.92	255.8	0.08	31.3	28.97	37.08
G13930	0.05	49.33	14.23	27.89	29.58	0
G26515	10.39	0.06	32.52	25.75	36.4	26.99

Gene Expression Key		
High	Med	Low

Table 11: Average Overall Expression Values of the Sequencing Data. The average overall expression values of the sequencing data in TPM and the average overall expression results of the real-time PCR data in CT values for each of the six selected primer sets. To make it easier to decipher this data, the expression amounts of genes on the left side of the table have been indicated with the color key as well as the concentration of these genes on the right side of this table.

Figure 10 includes three examples of the six oligo primer sets used for gene expression verification through real time PCR. The first electrophoresis gel contains the amplification results of the G33415 primer. The first three lanes contain our negative, the next three contain PCR product from the orange-fleshed genotype, followed by the next three which contain PCR product from the white-fleshed genotype, and the last three with the PCR product from the purple fleshed genotype. From this gel, we can see that the highest amplification occurs in the orange genotype and the lowest amplification occurs in the purple genotype and we can clearly see that this visualization on the gel corresponds to the expression values indicated in **Table 11** on both the sequencing data side and the real-time PCR data side. The second gel also indicates the same expression pattern as both the sequencing data and the real-time sides of **Table 11** (the

highest expression occurs in the white genotype and the lowest expression occurs in the purple genotype). The third gel is an example of a real-time PCR amplification pattern does not correlate directly to the expression value pattern of either the sequencing data or the real-time data in **Table 11**.

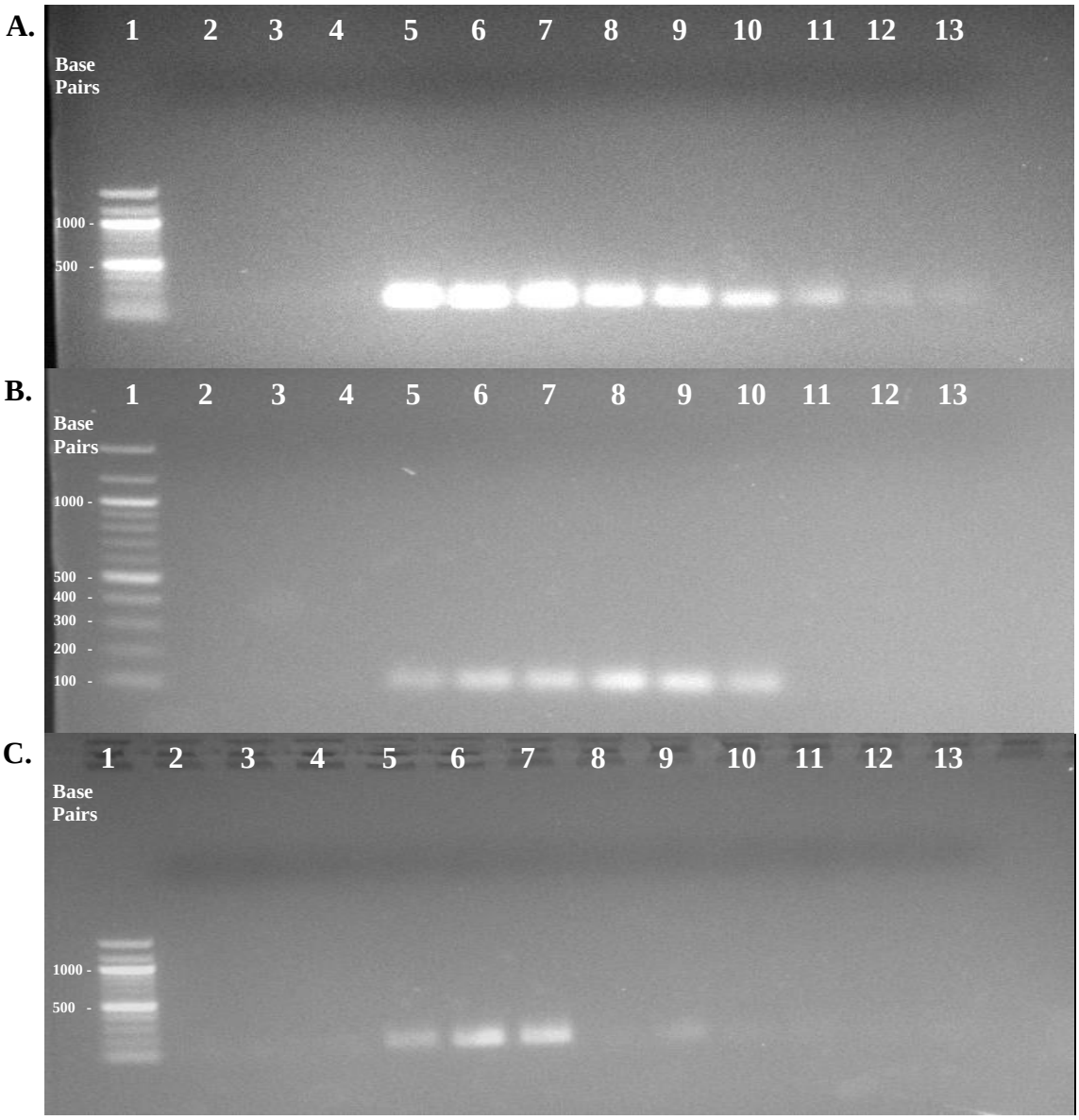
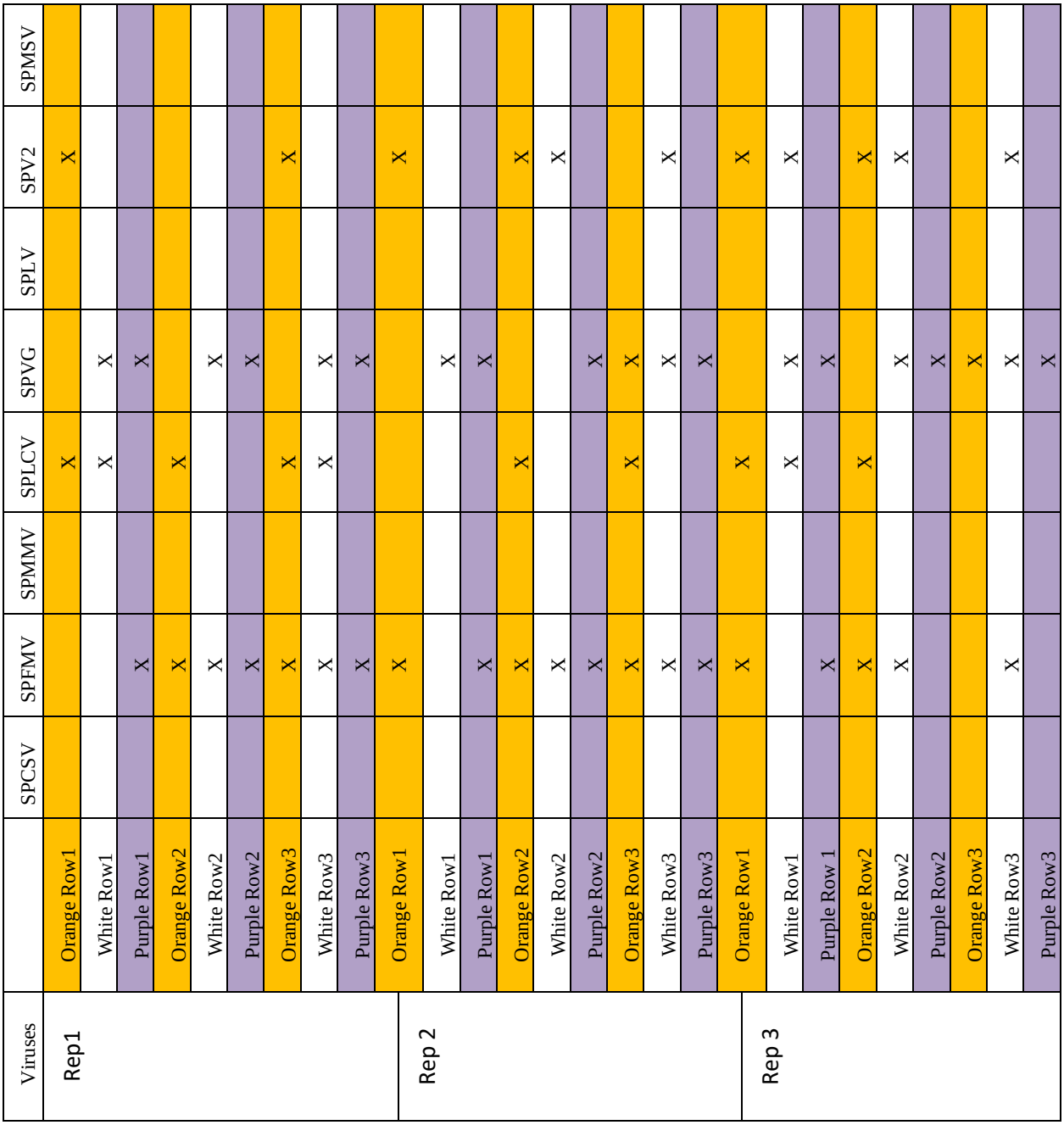


Figure 10: Sequencing Verification of Real Time PCR. Verified through agarose gel electrophoresis. **A.** Real-Time PCR verification of gene G33415 expression. Lane 1: 100bp, Lanes 2-4: G33415 and no cDNA, Lanes 5-7: G33415 and orange-fleshed sweet potato cDNA, Lanes 8-10: G33415 and white-fleshed sweet potato cDNA, Lanes 11-13: G33415 and purple-fleshed sweet potato cDNA. **B.** Real-Time PCR verification of gene G4244 expression. Lane 1: 100bp, Lanes 2-4: G4244 and no cDNA, Lanes 5-7: G4244 and orange-fleshed sweet potato cDNA, Lanes 8-10: G4244 and white-fleshed sweet potato cDNA, Lanes 11-13: G4244 and purple-fleshed sweet potato cDNA. **C.** Real-Time PCR verification of gene G26515 expression. Lane 1: 100bp, Lanes 2-4: G26515 and no cDNA, Lanes 5-7: G26515 and orange-fleshed sweet potato cDNA, Lanes 8-10: G26515 and white-fleshed sweet potato cDNA, Lanes 11-13: G26515 and purple-fleshed sweet potato cDNA.

4.4. VIRUS SCREENING

Since no previous work has been conducted to get an idea of what sweet potato viruses commonly prevail in Delaware, it was important that we conduct a screening of plants that have been cultivated in the field for a season, for some common viruses. When the eight different sweet potato viruses were screened for, instead of only utilizing only nine samples as we did for transcriptome sequencing, we tested all 27 leaf samples collected from the sweet potato plot (three genotypes x 3 rows x 3 replicates). Using regular PCR techniques, we discovered that only four of the viruses were present in these 27 samples and the results of the screening process are shown in **Table 12**. Overall results for the genotypes are given in **Table 13**. Seven of the orange plants are infected with Feathery Mottle Virus (SPFMV), seven are infected with Leaf Curl Virus (SPLCV), only two are infected with Virus G (SPVG), and six are infected with Virus 2 (SPV2). Six of the white plants are infected with SPFMV, three are infected with SPLCV, all but one are infected with SPVG, and five are infected with SPV2. Meanwhile, there are only two patterns of infection in the purple-fleshed plants, seven of the plants are infected with both SPFMV and SPVG while only two plants are infected with just SPVG.



Viruses		SPCSV	SPFMV	SPMMV	SPLCV	SPVG	SPLV	SPV2	SPMSV
Rep1	Orange Row1				X			X	
	White Row1				X	X			
	Purple Row1		X			X			
	Orange Row2		X		X				
	White Row2		X			X			
	Purple Row2		X			X			
	Orange Row3		X		X			X	
	White Row3		X		X	X			
	Purple Row3		X			X			
	Orange Row1		X					X	
Rep 2	White Row1					X			
	Purple Row1		X			X			
	Orange Row2		X		X			X	
	White Row2		X					X	
	Purple Row2		X			X			
	Orange Row3		X		X	X			
	White Row3		X			X		X	
	Purple Row3		X			X			
	Orange Row1		X		X			X	
Rep 3	White Row1				X	X		X	
	Purple Row 1		X			X			
	Orange Row2		X		X			X	
	White Row2		X			X		X	
	Purple Row2					X			
	Orange Row3					X			
	White Row3		X			X		X	
	Purple Row3					X			

Table 12: Results of Virus Screening for Each Sweet Potato Leaf Sample. X's indicate a positive result from our PCR tests, meaning that the virus is present. SPCSV=Sweet Potato Chlorotic Stunt Virus, SPFMV=Sweet Potato Chlorotic Stunt Virus, SPMMV=Sweet Potato Mild Mottle Virus, SPLCV=Sweet Potato Leaf Curl Virus, SPVG=Sweet Potato Virus G, SPLV=Sweet Potato Latent Virus, SPV2=Sweet Potato Virus 2, SPMSV=Sweet Potato Mild Speckling Virus.

Sweet Potato Viruses	Orange	White	Purple
SPCSV			
SPFMV	7X	6X	7X
SPMMV			
SPLCV	7X	3X	
SPVG	2X	8X	9X
SPLV			
SPV2	6X	5X	
SPMSV			

Table 13: Simplified Chart of Overall Virus Infections Detected in Three Genotypes. Simplified chart of overall virus infections detected in the three genotypes. The orange- and white-fleshed sweet potatoes have similar virus infection profiles (positive for SPFMV, SPLCV, SPVG, and SPV2). The purple-fleshed sweet potato is only infected with SPFMV and SPVG.

CHAPTER 5: DISCUSSION

The amount of total reads from each sweet potato genotype are approximately the same, as well as the percentage of mapped and un-mapped reads. These read totals and percentages for each genotype in **Table 4** are each made up of three individual plants. These aligned reads have been aligned to genes from the assembled genome of the Max Planck Institute. Alignment happens when there is homology between the reads which are fractionated pieces of the transcribed genes from the sample, all 101 nucleotides in length. Transcripts per million (TPM) is the normalized measurement of the amount of reads that are aligned to each individual gene from the Max Planck genome. The reason there must be a normalized measurement is because genes come in many different sizes, and it stands to reason that the longer genes will inevitably have more reads than the shorter reads. Since we do not want gene expression to be influenced by more reads due to size, and we wanted relatively lower numbers to work with, we converted the amount of reads to TPM.

Probably the most interesting thing in this study is the fact the three genotypes have seven out of the ten genes with the highest TPMs in common. These include the two genes labeled "putative ribulose bisphosphate carboxylase small subunit precursor," which is essential to the assembly of Rubisco, the molecule that allows the fixation of atmospheric carbon dioxide to occur during photosynthesis. It is hypothesized that Rubisco is the most plentiful protein on the planet, so it makes sense that Rubisco is one of the top ten up-regulated in all three genotypes (Cooper, 2000). Two of these genes are, at least, directly related to abiotic stress resistance. These genes are labeled "thiamine thiazole synthase 2, chloroplastic-like" and "metallothionein-like type 1 protein A chloroplastic." The thiamine thiazole synthase 2 gene is key to the biosynthesis of thiamine, obviously, which has proven to be crucial for tolerating abiotic stresses

and DNA damage (UniProt, 2018). The metallothionein gene is related to heavy metal stress tolerance (UniProt, 2018). There is also another gene that has one of the top 10 highest TPMs, but it is uncharacterized as of yet. Another one of these genes is related to chloroplast functions and maintenance. The final gene expressed as one of the top ten in all three genotypes is labeled "cytochrome P450 like TBP," meaning that it is likely a gene heavily involved in the synthesis of hormones and the generation of defensive compounds, which may be crucial against biotic and abiotic stresses (Miao, et al., 2017).

Since many of the aforementioned genes are influential in different stress tolerances, it would be fair to say that the plants are probably, to some extent, stressed, but to what extent, we cannot say, since all of our plants were grown out in the open, subject to external weather conditions. A scale has also yet to be created to estimate the level of stress based on gene expression.

Figure 8 shows us that although our reads were mapped to the 78,781 genes, only a couple thousand of these genes were significantly up-regulated in each genotype in each fold change comparison that was made. By significantly up-regulated, we mean that the log 2-fold change calculated after the fold change comparison carried out in CLC Bio is equal to or greater than two (had we not calculated log 2-fold change, the fold change value for significantly up-regulated genes would have been any value equal to or greater than four). The first fold change resulted in there being 2814 genes significantly up-regulated in the orange genotype and 3891 genes were significantly up-regulated in the white genotype (total of 6705 significantly up-regulated genes in the first fold change function). The second fold change resulted in 4559 genes significantly up-regulated in the white genotype and 3772 genes significantly up-regulated in the purple genotype (total of 8331 genes significantly up-regulated in the second fold change). The

final fold change results included 3521 genes significantly up-regulated in the orange genotype and 4364 genes significantly up-regulated in the purple genotype (total of 7885 genes significantly up-regulated in the third fold change comparison). From this information, we may deduce that there are more extreme differences in gene expression profiles between the white and purple sweet potato genotypes, followed by the differences between the orange and the purple sweet potato genotypes, and then finally we can observe the least extreme differences in gene expression profiles between the orange and white sweet potato genotypes.

In **Figure 9**, the blue and red bars represent the up-regulated genes in the orange and white genotype fold change comparison, the green and purple bars represent the up-regulated genes in the white and purple genotype fold change comparison, and the light blue and orange bars represent the up-regulated genes in the orange and purple genotype fold change comparison. As can be seen, from first glance at this graph, most of these differences in significantly up-regulated genes occur between the white and purple sweet potato genotypes and the least amount of differences tend to occur between the orange and white sweet potato genotypes. This is what was also observed in **Figure 8**, so it makes sense that overall, this pattern would continue into the individual functional groups.

Figure 9 gives more detail on the up-regulated genes across all three fold change comparisons. Here, we have sorted some of our up-regulated genes into functional groups using keywords like those in **Table 2**. As would probably be assumed, since there are such stark phenotypic differences, the most obvious being the identifying feature, the storage root color, the bulk differences in gene expression do not lie within the pigment functional group. In fact, this group has the lowest amount of significantly up-regulated genes in all of the functional groups at which we looked. Meanwhile, the functional group with the most significantly up-regulated

genes was transcription factors, which does make sense, as transcription must constantly occur within an organism, so the machinery to make this possible must continuously be manufactured.

From **Table 7**, we can see that one of the highest up-regulated genes in the orange genotype from the first fold change comparison is related to tolerance of a wide variety of stresses. Therefore, this plant could have been adapting to any number of, or even multiple stresses. However, as we look at the up-regulated genes in the first fold change comparison of **Table 8**, we observe that among the five up-regulated genes is one specifically related to heat tolerance (heat shock protein 90-1-like), so it is likely, though not confirmed that excessive heat was what stressed this plant. Meanwhile, back at **Table 7**, the white genotype indicated that one of the five highest up-regulated genes was one related to drought-tolerance and another up-regulated that is key for pest/insect tolerance. This information indicates that the white sweet potato genotype likely has a lower tolerance for drought stress, at least when compared to the orange genotype. Since the white genotype in this same comparison also displayed up-regulation of the gene related to tolerating pests and insects, the plant also likely has a lower tolerance for damage due to insect interference. Another reason for this up-regulation to occur could be that possibly the white sweet potato has some type of inherent allure for insects and other pests, leading to a higher rate of attack and a need for up-regulation of this gene.

The purple genotype in the third fold change comparison, is shown to have a gene up-regulated that likely prevents damage due to oxidative stress, not only that, but it is one of the top five up-regulated genes within the fold change comparisons. Oxidative stress occurs when during photosynthesis, excess light is absorbed and reactive oxygen species (ROS) are created, which may cause harm to the plant (Sejima, Takagi, Fukayama, Makino, & Miyake, 2017). It is interesting when it is considered that the purple sweet potato plants were unusual in that they did

not begin to flower until much later in the year, around the time when summer had officially ended. Is it possible that longer days were causing excess stress on the plants and as a result, they were unable to produce flowers? Perhaps, and this type of information would indicate that perhaps this plant is a better candidate for later plantings, once the longest days of the summer are over. Also, in **Table 8**, in the second and third change comparison, heat shock proteins were up-regulated in the purple genotype in both instances, making it evident that the purple sweet potato genotype may actually be more practical for overall warmer climates.

Table 9 yielded results that I expected for the most part. Anthocyanin proteins were mostly found to be up-regulated in the purple genotype and carotenoids were mostly upregulated in the orange-fleshed genotypes. Of course, these differences would probably be more apparent if we had created RNA profiles of the sweet potato storage roots themselves. However, there were a couple of times in the white genotype when purple acid phosphatase, red chlorophyll, carotenoid-related proteins, and anthocyanins were up-regulated, which was completely unexpected.

The initiation of storage roots functional group was interesting in that it was expected that purple sweet potato plants would have more up-regulated genes. This proved not to be the case for the fold change comparison between the white and purple genotypes. In fact, in both of the fold changes involving the white genotype saw more genes up-regulated in the white genotype. This was interesting because it is evident that by mass there was a higher amount of marketable storage roots harvested, and a higher amount of storage roots by number harvested on average from each of the orange plants, so it would stand to reason that one of these genotypes would have a higher amount of up-regulated genes. This is not the case, the white sweet potato genotype has the highest amount of up-regulated genes related to the initiation of storage roots.

The sequencing verification that we carried out involved performing a sort of "audit," with Real Time PCR and utilizing primers related to genes that are differentially regulated in the three genotypes. For example, if the orange sweet potato genotype has a high TPM for one gene, it is reasonable to assume that the real time PCR will reveal that there will be a higher quantity of that same gene present in the purple sweet potato genotype as well. Unfortunately, only two out of the six primer sets correlate directly to the sequencing data (G33415 and G4244). Of course, PCR is a very sensitive tool, so there could have been other factors involved here.

Viral distribution was found to be distinct to different sweet potato genotypes. While the white and orange sweet potatoes were found to have some incidences of infection with Sweet Potato Feathery Mottle Virus (SPFMV), Sweet Potato Leaf Curl Virus (SPLCV), Sweet Potato Virus G (SPVG), and Sweet Potato Virus 2 (SPV2). There was a higher incidence of SPLCV in the orange genotype and a higher incidence of SPVG in the white genotype. Meanwhile, the purple sweet potato plant was only found to have incidences of infection of SPFMV and SPVG. The fact that the purple sweet potatoes seem to have some type of immunity to SPLCV and SPV2 is a significant phenotypic difference that should be further explored.

CHAPTER VI: CONCLUSIONS

From these transcriptome profiles generated here, we are able to get a better sense of differences in gene expression than we would have been able to do with full genome sequencing. These transcriptomes will also serve as reference transcriptomes for future sweet potato transcriptome work within our collaborative work. We hope to carry out transcriptome sequencing to elucidate differences in gene expression between two or more generations of the same plant. Also, since there is an extreme interest in how a plant is affected by viruses at the transcriptome level, this is the most likely direction that our research will head. More specifically, we will examine transcriptome effects of the synergistic interaction of the feathery mottle virus and the chlorotic stunt virus. Comparing healthy plants, plants infected with both of these viruses, and also plants that are singularly infected, will help us to understand what resistance mechanisms begin to break down once there is a co-infection within the plant. This research is essential for preventing future devastating global losses that could possibly occur due to accidental transport of infected plants with one of these viruses, to a growing region where the other virus is present. Once the transcriptome business is taken care of, it will be necessary to closer explore what mechanisms control these changes in gene expression once the plant is dually infected. Chromatin Immunoprecipitation (ChIP) will play a key role in determining how chromatin conformation plays a role in gene expression by creating epigenome profiles of healthy plants, plants inoculated with SPFMV, plants inoculated with SPCSV, and plants inoculated with both of these sweet potato viruses. Transcriptome and epigenome results combined together, can reveal to us specific genes that are up-regulated or down-regulated due to infection with one of these viruses and how severe breakdowns in immunity from other viruses occurs. Histone methylation and/or histone acetylation are how we can visualize chromatin

accessibility throughout the entire genome and how this changes due to these different infection states, negatively influencing yields. Through these future studies, we will soon be able to preemptively apply strategies and methods that could prevent these mechanism breakdowns that inevitably breakdowns causing major crop losses.

REFERENCES

Abad, J., & Moyer, J. (1992). Detection and distribution of sweetpotato feathery mottle virus in sweetpotato by in vitro-transcribed RNA probes (riboprobes), membrane immunobinding assay, and direct blotting. *Phytopathology*, 300-305.

Ames, T., Smit, N., Braun, A., O'Sullivan, J., & Skoglund, L. (1997). *Sweetpotato: Major Pests, Diseases, and Nutritional Disorders.* (B. Hardy, Ed.) Lima, Peru: International Potato Center (CIP). Retrieved from http://www.sweetpotatoknowledge.org/wp-content/uploads/2016/02/SP-ames-et-al.pdf

Austin, D. (1988). The taxonomy, evolution and genetic diversity of sweepotato and related wild species. *Exploration, Maintenance, and Utilization of Sweet Potato Genetic Resources* (pp. 27-59). Lima, Peru: International Potato Center (CIP).

Ayyappan, V., Saha, M. C., Thimmapuram, J., Sripathi, V. R., Bhide, K. P., Fiedler, E., . . . Kalavacharla, V. (2017). Comparative transcriptome profiling of upland (VS16) and lowland (AP13) ecotypes of switchgrass. *Plant Cell Reports*, 129-150.

Bendre, A., Suresh, C., Shanmugam, D., & Ramasamy, S. (2018). Structural insights into the unique inhibitory mechanism of Kunitz type trypsin inhibitor from Cicer arietinum L. *Journal of Biomolecular Structure and Dynamics*, 1-38.

BiochemDen. (2018). *What is RNase? What is the importance in living things?* Retrieved from BiochemDen: https://www.biochemden.com/what-is-rnase-what-is-importance-in/

Brunt, A. A. (1996). *Viruses of plants: descriptions and lists from the VIDE database.* Wllingford, Oxon, UK: CAB International. Retrieved from https://trove.nla.gov.au/work/22068129?selectedversion=NBD12216566

Cao, Q., Li, A., Chen, J., Sun, Y., Tang, J., Zhang, A., . . . Gao, S. (2016). Transcriptome

Sequencing of the Sweet Potato Progenitor (Ipomoea trifida (H.B.K.) G. Don.) and Discovery of Drought Tolerance Genes. *Tropical Plant Biology*, 63-72.

Clark, C. A., Davis, J. A., Abad, J. A., Cuellar, W. J., Fuentes, S., Kreuze, J. F., . . . Valkonen, J. P. (2012, February). Sweetpotato Viruses: 15 Years of Progress on Understanding and Managing Complex Diseases. *96*(2), 168-185. doi:10.1094/PDIS-07-11-0550

Clark, M., & Adams, A. (1976). Characteristics of the Microplate Method of Enzyme-Linked Immunosorbent Assay for the Detection of Plant Viruses. *Journal of General Virology*, 475-483.

Colinet, D., Kummert, J., & Lepoivre, P. (1998). The nucleotide sequence and genome organization of the whitefly transmitted sweetpotato mild mottle virus: a close relationship with members of the family Potyviridae. *Virus Research*, 187-196.

Collins, G., & Erickson, A. (2011). *The Saudi Arabia of Corn: Reduce U. S. ethanol production and boost corn exports to China to trim the trade deficit and ensure China's food security.* China SignPost.

Cooper, G. M. (2000). Chloroplasts and Other Plastids. In *The Cell: A Molecular Approach.* Sunderland: Sinauer Associates.

Crop Trust. (2018). *Sweetpotato*. Retrieved from Crop Trust: https://www.croptrust.org/crop/sweet-potato/

Doyle, J., & Doyle, J. (1990). Isolation of plant DNA from fresh tissue. *Focus*, 13-15.

EMBL-EBI. (2018). *Eukaryotic Translation Elongation Factor 1 Complex*. Retrieved from Quick GO: https://www.ebi.ac.uk/QuickGO/term/GO:0005853

Esbenshade, P., & Moyer, J. (1982). Indexing system for sweet potato feathery mottle virus in sweet potato using enzyme-linked immunosorbant assay. *Plant Disease*, 911-913.

Eurofins. (2018). *Copy Number Variation-Important Markers Beyond SNPs*. Retrieved from Eurofins: https://www.eurofinsgenomics.eu/en/genotyping-gene-expression/genotyping-services/copy-number-variation/

Fan, W., Zhang, M., Zhang, H., & Zhang, P. (2012). Improved Tolerance to Various Abiotic Stresses in Transgenic Sweet Potato (Ipomoea batatas) Expressing Spinach Betaine Aldehyde Dehydrogenase. *PLOS One*.

Firon, N., LaBonte, D., Villordon, A., Kfir, Y., Solis, J., Lapis, E., . . . Nadir, L. A. (2013). Transcriptional profiling of sweetpotato (Ipomoea batatas) roots indicates down-regulation of lignin biosynthesis and up-regulation of starch biosynthesis at an early stage of storage root formation. *BioMed Central Genomics*.

Fowler, T. J., Bernhardt, C., & Tierney, M. L. (1999). Characterization and expression of four proline-rich cell wall protein genes in Arabidopsis encoding two distinct subsets of multiple domain proteins. *Plant Physiology*, 1081-1092.

Google. (2018). *Google Maps*. Retrieved from Google Maps Satellite: https://www.google.com/maps

Guo, P., Baum, M., Grando, S., Ceccarelli, S., Bai, G., Li, R., . . . Valkoun, J. (2009). Differentially expressed genes between drought-tolerant and drought-sensitive barley genotypes in response to drought stress during the reproductive stage. *Journal of Experimental Botany*, 3531-3544.

Gutierrez-Campos, R., Torres-Acosta, J., Saucedo-Arias, L., & Gomez-Lim, M. (1999). The use of cysteine proteinase inhibitors to engineer resistance against potyviruses in transgenic tobacco plants. *Nature Biotechnology*, 1223-1226. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10585723

Helen Keller International. (2018, May 16). *Our Sweet Potato Solution*. Retrieved from http://www.hki.org/sweetpotato#.Wvx3be8vyUl

Hirakawa, H., Okada, Y., Tabuchi, H., Shirasawa, K., Watanabe, A., Tsuruoka, H., . . . Isobe, S. N. (2015). Survey of genome sequences in a wild sweet potato, Ipomoea trifida (H.B.K.) Don. *DNA Research*, 171-179.

IFAS. (2016, August 5). UF/IFAS study: Sweet potato crop shows premise as feed and fuel. *Ethanol Producer Magazine*. Florida. Retrieved from http://ethanolproducer.com/articles/13597/uf-ifas-study-sweet-potato-crop-shows-promise-as-feed-and-fuel

Illumina. (2017). TruSeq RNA Library Prep Kit v2. Retrieved from https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/truseq-rna-v2.html

Illumina. (2018). HiSeq 2500 System Guide. San Diego, California, U.S.A. Retrieved from https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/hiseq2500/hiseq-2500-system-guide-15035786-02.pdf

Illumina. (2018). HiSeq 4000 System Guide. San Diego, California, U.S.A. Retrieved from https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/hiseq4000/hiseq-4000-system-guide-15066496-05.pdf

Illumina. (2018). NextSeq 500 System Guide. San Diego, California, U.S.A. Retrieved from https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/nextseq/nextseq-500-system-

guide-15046563-04.pdf

Isobe, S., Shirasawa, K., & Hirakawa, H. (2017). Challenges to genome sequence dissection in sweetpotato. *Breeding Science*.

Klass, D. L. (2004). Biomass for Renewable Energy and Fuels. *Encyclopedia of Energy*, 193-212.

Kokkinos, C., & Clark, C. (2006). Real-Time PCR Assays for Detection and Quantification of Sweetpotato Viruses. *Plant Disease*, 783-788.

Kokkinos, C., Clark, C., McGregor, C., & LaBonte, D. (2006). The Effect of Sweet Potato Virus Disease and its Viral Components on Gene Expression Levels in Sweetpotato. *Journal of the American Society for Horticultural Science, 131*(5), 657-666. Retrieved from http://journal.ashspublications.org/content/131/5/657.full.pdf+html

Kwak, H.-R., Kim, M.-K., Shin, J.-C., Lee, Y.-J., Seo, J.-K., Lee, H.-U., . . . Choi, H.-S. (2014). The Current Incidence of Viral Disease in Korean Sweet Potatoes and Development of Multiplex RT-PCR Assays for Simultaneous Detection of Eight Sweet Potato Viruses. *Plant Pathology Journal, 30*(4), 416-424. doi:10.5423/PPJ.OA.04.2014.0029

Lebot, V. (2009). *Tropical Roots and Tuber Crops: Cassava, Sweet Potato, and Aroids.* London, UK: MPG Books Group.

Li, R., Zhai, H., Kang, C., Liu, D., He, S., & Liu, Q. (2015). De Novo Transcriptome Sequencing of the Orang-Fleshed Sweet Potato and Analysis of Differentially Expressed Genes Related to Carotenoid Biosynthesis. *International Journal of Genomics*.

Liu, R., Qiu, J., Finger, L. D., Zheng, L., & Shen, B. (2006). The DNA-protein interaction modes of FEN-1 with gap substrates and their implication in preventing duplication mutations. *Nucleic Acids Research*, 1772-1784.

Loebinstein, G., Thottappilly, G., Fuentes, S., & Cohen, J. (2003). Virus and Phytoplasma Diseases. In G. Loebenstein, & G. Thottappilly, *Virus and Virus-like Diseases of Major Crops in Developing Countries* (pp. 105-134). Kluwer Academic Publishers. doi:10.1007/978-1-4020-9475-0 8

Lv, G.-Y., Guo, X.-G., Xie, L.-P., Xie, C.-G., Zhang, X.-H., Yang, Y., . . . Xu, H. (2017). Molecular Characterization, Gene Evolution, and Expression Analysis of the Fructose-1, 6-bisphosphate Aldolase (FBA) Gene Family in Wheat (Triticum aestivum L.). *Frontiers in Plant Science*.

Martin, F. (1982). Analysis of the incompatibility and sterility of sweet potato. In: Villareal, R. L. and Griggs, T. D. (eds) Sweet Potato. *Proceedings of the First International Symposium*, 275-284. Taiwan: AVRDC Publication No. 82-172.

Martin, F. W. (1970). Self And Interspecific Incompatibility In The Convolulaceae. *Botanical Gazette*, 139-144.

Martin, F., & Jones, A. (1968). Breeding sweet potatoes. *Plant Breeding Review*, 313-345.

McDonald, J., & Austin, D. (1990). Changes and additions in Ipomoea section Batatas (Convolvulaceae). *Brittonia*, 116-120.

Miao, Q., Deng, P., Saha, S., Jenkins, J. N., Hsu, C.-Y., Abdurakhmonov, I. Y., . . . Ma, D.-P. (2017). Genome-wide identification and characterization of microRNAs differentially expressed in fibers in a cotton phytochrome A1 RNAi line. *PLOS One*.

Mussoline, W. A., & Wilkie, A. C. (2016). Feed and fuel: the dual-purpose advantage of an industrial sweetpotato. *Journal of the Science of Food and Agriculture*. doi:10.1002/jsfa.7902

Nag, O. S. (2017, April 25). *World Leaders In Sweet Potato Production*. Retrieved from

Economics: https://www.worldatlas.com/articles/world-leaders-in-sweet-potato-production.html

Nakayama, H., Tanaka, M., & Takahata, Y. (2010). An AFLP-based genetic linkage map of Ipomoea trifida (H.B.K.) G. Don., a diploid relative of sweetpotato, I. batatas (L.) Lam. *Tropical Agriculture and Development*, 9-16.

NCBI. (2018). *NCBI Home*. Retrieved from National Center for Biotechnology Information: https://www.ncbi.nlm.nih.gov/

NCBI. (218). *Primer-BLAST*. Retrieved from NIH U.S. National Library of Medicine: https://www.ncbi.nlm.nih.gov/tools/primer-blast/

NEB. (2018). *First Strand cDNA Synthesis (Standard Protocol) (NEB #M0368)*. Retrieved from New England Biolabs: https://www.neb.com/protocols/2016/04/26/first-strand-cdna-synthesis-standard-protocol-neb-m0368

New England Biolabs Inc. (2016). Protoscript II First Strand cDNA Synthesis Kit. Retrieved from https://www.neb.com/products/m0368-protoscript-ii-reverse-transcriptase#Product%20Information

Next Prot. (2018, September 3). *UBA52- Ubiquitin-60S ribosomal protein L40*. Retrieved from Next Prot: https://www.nextprot.org/entry/NX_P62987/

Ngailo, S., Shimelis, H., Sibiya, J., & Mtunda, K. (2013). Sweet potato breeding for resistance to sweet potato virus disease and improved yield: Progress and challenges. *African Journal of Agricultural Research*, 3202-3215. doi:10.5897/AJAR12.1991

Ngailo, S., Shimelis, H., Sibiya, J., & Mtunda, K. (2013). Sweet potato breeding for resistance to sweet potato virus disease and improved yield: progresses and challenges. *African Journal of Agricultural Research*, 3202-3215.

NGS Sequencing Core Facility. (2017). *Ipomoea Batatas Genome Browser*. (Max Planck Institute Molecular Genetics) Retrieved from NGS Sequencing Core Facility: http://public-genomes-ngs.molgen.mpg.de/SweetPotato/

North Carolina Department of Agriculture and Consumer Reviews. (2014). *Marketing North Carolina Sweet Potatoes.* Federal-State Market News Service.

North Carolina Sweet Potato Commission. (2018). *What is the Difference Between A Sweet Potato and a Yam?* Retrieved from North Carolina Sweet Potatoes Naturally Healthy: https://www.ncsweetpotatoes.com/sweet-potatoes-101/difference-between-yam-and-sweet-potato/

Okada, Y., Saito, A., Nishiguchi, M., Kimura, T., Mori, M., Hanada, K., . . . Murata, T. (2001). Virus resistance in transgenic sweetpotato [Ipomoea batatas L. (Lam)] expressing the coat protein gene of sweet potato feathery mottle virus. *Theory of Applied Genetics*, 743-751.

Ozias-Akins, P., & Jarret, R. L. (1994). Nuclear DNA Content and Ploidy Levels in the Genus Ipomoea. *Journal of the American Society for Horticultural Science*, 110-115.

Park, S.-C., Kim, Y.-H., Ji, C. Y., Park, S., Jeong, J. C., Lee, H.-S., & Kwak, S.-S. (2012). Stable Internal Reference Genes for the Normalization of Real-Time PCR in Different Sweetpotato Cultivars Subjected to Abiotic Stress Conditions. *PLoS One*. doi:10.1371/journal.pone.0051502

Ponniah, S. K., Thimmapuram, J., Bhide, K., Kalavacharla, V. (., & Manoharan, M. (2017). Comparative analysis of the root transcriptomes of cultivated sweetpotato (Ipomoea batatas [L.] Lam) and its wild ancestor (Ipomoea trifida [Kunth] G. Don). *BioMedCentral Plant Biology*.

Prakash, C., & Varadarajan, U. (1992). Genetic transformation of sweet potato by particle bombardment. *Plant Cell Report, 11*(2), 53-57. doi:10.1007/BF00235252

Qiagen. (2017). CLC Bio. Aarhus, Denmark.

Reddy, U., Bates, G., Ryan-Bohac, J., & Nimmakayala, P. (2007). Genome mapping and molecular breeding in plants. In C. Kole, *Sweetpotato* (pp. 237-247). Berlin: Springer-Verlag.

Reis, A., Hornblower, B., Robb, B., & Tzertzinis, G. (2014). *CRISPR/Cas9 & Targeted Genome Editing: New Era in Molecular Biology*. Retrieved from New England Biolabs Inc.: https://www.neb.com/tools-and-resources/feature-articles/crispr-cas9-and-targeted-genome-editing-a-new-era-in-molecular-biology

Roche. (2011, May). Sequencing Method Manual for the GS FLX+ Instrument. Retrieved from http://netdocs.roche.com/PPM/GS_FLXplus_Sequencing_XLR70_Kit_Method_Manual_May_2011.pdf

Salesse-Smith, C. E., Sharwood, R. E., Busch, F. A., Kromdijk, J., Bardal, V., & Stern, D. B. (2018). Overexpression of Rubisco subunits with RAF1 increases Rubisco content in maize. *Nature Plants*, 802-810.

Schafleitner, R., Tincopa, L. R., Palomino, O., Rossel, G., Robles, R. F., Alagon, R., . . . Simon, R. (2010). A sweetpotato gene index established by de novo assembly of pyrosequencing and Sanger sequences and mining for gene-based microsatellite markers. *BMC Genomics*.

Sejima, T., Takagi, D., Fukayama, H., Makino, A., & Miyake, C. (2017). How plants oxidative stress. *Plant Cell Physiology*, 1184-1193.

Sigma-Aldrich. (2010). Spectrunm Plant Total RNA Kit. St. Louis, MO, United States. Retrieved from https://www.sigmaaldrich.com/content/dam/sigma-

aldrich/docs/Sigma/Bulletin/strn250bul.pdf

Sweet Potato Varieties. (2016). Retrieved from All About Sweet Potatoes: http://www.all-about-sweet-potatoes.com/sweet-potato-varieties.html

The Gale Encyclopedia of Science. (2016). *Plant Pigment*. Retrieved from Encyclopedia.com: https://www.encyclopedia.com/science/encyclopedias-almanacs-transcripts-and-maps/plant-pigment-0

Thermo Fisher Scientific. (2018, June 29). *Turbo DNA-free Kit*. Retrieved from User Guide: https://www.thermofisher.com/order/catalog/product/AM1907

U.S. National Library of Medicine. (2018, May 16). *National Center for Biotechnology Information*. Retrieved from GenBank: https://www.ncbi.nlm.nih.gov/

Ugent, D., & Peterson, L. W. (1988). Archaeological Remains of Potato and Sweet Potato in Peru. *CIP Circular*, 1-10.

UNCTAD. (2012, May 2). *United Nations Conference on Trade and Development*. Retrieved October 6, 2015, from Infocomm Commodity Profile Sweet Potato: http://www.unctad.info/en/Infocomm/AACP-Products/COMMODITY-PROFILE---Sweet-potato/

UniProt. (2017, December 20). *Alpha-glucan phosphorylase, H isozyme*. Retrieved from UniProt: https://www.uniprot.org/uniprot/P32811

UniProt. (2017, June 7). *Dehydration-induced protein RD22-like protein 1*. Retrieved from UniProt: https://www.uniprot.org/uniprot/Q6DUZ1

UniProt. (2017, November 22). *Probable Protein Kinase UbiB*. Retrieved from UniProt: https://www.uniprot.org/uniprot/A0A0H3NJQ5

UniProt. (2018). *Metallothionein-like protein type 2*. Retrieved from UniProtKB-Q43513

(MT2Y_SOLLC): https://www.uniprot.org/uniprot/Q43513

UniProt. (2018, July 18). *Ras-related protein RABD1*. Retrieved from UniProt: https://www.uniprot.org/uniprot/Q9ZRE2

UniProt. (2018, May 23.). *Thiamine thiazole synthase 2, chloroplastic*. Retrieved from UniProtKB-Q41739 (THI42_MAIZE): https://www.uniprot.org/uniprot/Q41739

UniProt. (2018). *UniProt Home*. Retrieved from https://www.uniprot.org/

UniProt. (2018b, September 12). *Glycerate dehydrogenase HPR, peroxisomal*. Retrieved from UniProt: https://www.uniprot.org/uniprot/Q9C9W5

UniProt. (2018c, April 25). *Hypersensitive-induced responsive protein 1*. Retrieved from UniProt: https://www.uniprot.org/uniprot/Q9FM19

UniProt. (2018d, May 23). *Sulfiredoxin, chloroplastic/mitochondrial*. Retrieved from UniProt: https://www.uniprot.org/uniprot/Q8GY89

United States Department of Agriculture. (2018). *Sweetpotatoes Grades and Standards*. Retrieved from Agricultural Marketing Service: https://www.ams.usda.gov/grades-standards/sweetpotatoes-grades-and-standards

Valverde, R., Clark, C., & Valkonen, J. (2007). Viruses and Virus Disease Complexes of Sweetpotato. *Plant Viruses*, 116-126.

Wambugu, F. M. (2003). Development and Transfer of Genetically Modified Viru-resistant Sweet Potato for Subsistence Farmers in Kenya. *International Life Sciences Institute*, S110-S113.

Wang, T., Dong, X., Jin, Z., Su, W., Ye, X., Dong, C., & Lu, Q. (2015). Pyrolytic characteristics characteristics of sweet potato vine. *Bioresource Technology*, 799-801.

Xie, F., Burklew, C., Yang , Y., Liu, M., Xiao, P., Zhang, B., & Qui, D. (2012). De novo

sequencing and a comprehensive analysis of purple sweet potato (Ipomoea batatas L.) transcriptome. *Planta*, 101-113.

Yan, L., Lai, X., Li, X., Wei, C., Tan, X., & Zhang, Y. (2015). Analysis of the Complete Genome and Gene Expression of Chloroplast of Sweet Potato [Ipomoea batata]. *PLOS One*.

Yang, J., Bi, H.-P., Fan, W.-J., Zhang, M., Wang, H.-X., & Zhang, P. (2011). Efficient embryogenic suspension culturing and rapid transformation of a range of elite genotypes of sweet potato (Ipomoea batatas [L.] Lam.). *Plant Science*, 701-711.

Yang, J., Moeinzadeh, M.-H., Kuhl, H., Helmuth, J., Xiao, P., Haas, S., . . . Vingron, M. (2017, August). Haplotype-resolved sweet potato genome traces back its hexaploidization history. *Nature Plants*, *3*, 696-703. doi:10.1038/s41477-017-0002-z

Yi, G., Shin, Y.-M., Choe, G., Shin, B., Kim, Y. S., & Kim, K.-M. (2007). Production of herbicide-resistant sweet potato plants transformed with the bar gene. *Biotechnology Letters*, *29*, 669-675. doi:10.1007/s10529-006-9278-1

Zargar, S. M., Nagar, P., Deshmukh, R., Nazir, M., Wani, A. A., Masoodi, K. Z., . . . Rakwal, R. (2017). Aquaporins as potential drought tolerance including proteins: Towards instigating stress tolerance. *Journal of Proteomics*, 233-238.

Ziska, L. H., Runion, G. B., Tomecek, M., Prior, S. A., Torbet, H. A., & Sicher, R. (2009). An evaluation of cassava, sweet potato and field corn as potential carbohydrate sources for bioethanol production in Alabama and Maryland. *Biomass and Bioenergy*, 1503-1508. doi:10.1016/j.biombioe.2009.07.014