8-10-2016

Investigations of Storage Root Development in Cassava (*Manihot esculenta* Crantz)

Yeetoh Chaweewan
University of Missouri-St. Louis, yczn9@umsl.edu

Investigations of Storage Root Development in Cassava (*Manihot esculenta* Crantz)

Yeetoh Chaweewan
M.Sc. in Agriculture (Plant Breeding), Kasetsart University, Thailand 1998
B.Sc. in Agriculture (Agronomy), Kasetsart University, Thailand 1994

A Dissertation Submitted to the Graduate School at

The University of Missouri-St. Louis
In partial fulfillment of the requirements for the degree

Doctor of Philosophy in Biology with an Emphasis in Cell Molecular Biology

December 2014

Dissertation Committee:
Wendy Olivas, Ph.D.
Advisor

Nigel Taylor, Ph.D.
Co-Advisor

Xuemin Wang, Ph.D.

Bethany Zolman, Ph.D.

Toni Kutchan, Ph.D

ABSTRACT

The tropical root crop cassava is cultivated for its large starchy storage roots. Understanding critical processes in root tuberization is essential if improvement programs are to secure future yields for farmers. Studies were undertaken to identify critical components of storage root development at the anatomical and gene expression levels. Two types of roots were identified from greenhouse-grown stem cuttings: basal roots, which develop from the stem cut end and are prolific in nature, and nodal roots, which originate from the region of the buried axillary bud and are limited to 3-5 per node. Only nodal roots develop to produce storage organs. Anatomical sectioning was performed to determine the origin of both root types. Basal roots were seen to develop from the cambium of the semi-woody stems, while nodal roots originated from deep within the secondary xylem or pith regions. This data contradicts accepted knowledge that storage roots develop from a subset of the fibrous roots. As a result, it is proposed here that storage and basal/fibrous roots are fundamentally different organs, originate through different rhizogenic processes, and are committed to their different developmental fates from initiation onwards. cDNA microarray analysis was performed on roots at different stages of storage root development. Gene Set Enrichment Analysis revealed up-regulation of the jasmonic acid biosynthesis pathway during the initiation stage of tuberization. K-means clustering identified three clusters of up-regulated genes at storage root initiation and later developmental stages, while Heatmap analysis revealed major latex allergen Hev b 4 proteins to be highly up-regulated at the initiation stage. Three candidate genes seen to be highly up-regulated at the later starch filling stages were identified as possible homologues of *Mec1*, cassava *ATDI21* and *ENOD40*-like genes.

RT-PCR analysis revealed their enhanced expression in storage roots compared to fibrous roots and leaves. *Mec1* has previously been associated with cassava storage roots, but no reports exist for the involvement of *ATDI21* or *ENOD40*. The homologues of the latter two genes require further characterization to determine their functional role in storage root development. Integration of anatomical studies with functional genomics tools has provided new knowledge of root tuberization in cassava and identified new avenues of research.

ACKNOWLEDGEMENT

My time within the doctorate program in St. Louis is the most invaluable experience in my life. Firstly, I would like to thank Dr. Claude Fauquet, who warmly accepts me to work on my thesis in his lab at Danforth Plant Science Center. It is an excellent chance for me to learn the specific skills in functional genomics research. I am very grateful for the kindness of Dr. Nigel Taylor, who always assists me for getting through all problems while I work and study here, especially in the anatomy study. I appreciate for the time he spends to correct my dissertation. Dr. Toni Kutchan is another committee member who deserves my thanks for her encouraging. I am also thankful to Dr. Howard Berg, Dr. Eliana Gaitan, Dr. Augustin Jorg and Matthew Kilgore for their guidance on data analysis. I also would like to thank Dr. David Horvath at USDA, Fargo, who always encourage and help me in the microarray experiment. Furthermore, I am grateful for the kindness of my committee at UMSL, Dr. Sam Wang, Dr. Wendy Olivas, and Dr. Bethany Zolman, who are always supporting me during my study time in UMSL. I will not be able to graduate without their reinforcement. I am thankful for my friends at UMSL, Danforth Center, IICI members who made my terrific life here. I appreciate Mrs. Surin Kayembe, Sylwia and Jack for their best friendship and their help, and for well taking care of me and my children. Thank you to everyone I know who I didn't mention here.

Lastly, I would like to dedicate this accomplishment to my parent and my family. I am grateful to my parent for their support and for being with me all the time. I am also grateful to my husband and my lovely children for their understanding and patience. I would not be successful in this program without their kindness and their love.

ABBREVIATIONS

ATDI21, *Arabidopsis thaliana* drought-induced 21

BLAST, Basic Local Alignment Search Tool

cDNA, Complementary DNA

ENOD40, Early nodulin 40 gene

EST, Expressed sequence tag

GA, Gibberellic acid

GO, The Gene Ontology

GSEA, Gene Set Enrichment Analysis

KI, Potassium iodide solution

MeATDI21, Cassava *ATDI21* gene

Mec1, Cassava *c1* gene

Lea3, Late embryogenesis abundant protein 3

Lea5, Late embryogenesis abundant protein 5

Pt2L4, Allergenic-related protein Pt2L4

RNAi, RNA interference

RNA-seq, RNA sequencing

RT-PCR, Reverse transcription polymerase chain reaction

Table of contents

Project Summary

Cassava storage roots play an important role in the world as a source of human food, animal feed and biofuels in more than 90 countries. The cassava tuberization process is a key mechanism that allows the plant to form storage roots to store very large quantities of starch, firstly for the benefit of the plant and for exploitation by humans. The yield component of cassava production is highly dependent on the size and number of storage roots per plants. This finite number of storage roots (3 to 14) is not only limited by genetic factors, but also by nutritional and environmental conditions. There are thus underlying, unknown molecular mechanisms in the young plant that drive the developmental switch from root to storage organ. Anatomical studies allow us to gain knowledge on storage root formation especially at the tissue level, including information on the origin of root initiation from the earliest stages of development. Transcriptome analysis is another way to approach this question and increase understanding of the molecular mechanisms involved, all of which are important for future genetic improvement of cassava.

Intellectual merit: The main goal of this thesis is to describe the process of storage root formation including the molecular mechanisms driving tuberization of cassava storage roots. Anatomical studies increased knowledge of root formation in cassava at the early stages to determine the process of switching nodal roots to storage roots. Gene expression profiling provides a powerful way to identify sets of genes that initiate and regulate initiation of storage root development and that play a role in subsequent steps leading to production of mature storage roots. The hypothesis of the present work is that this switch is correlated to a set of molecular changes triggered

by an unknown mechanism(s). A cDNA microchip containing 4,129 cassava unigenes and 19,808 unigenes of leafy spurge, a close relative species of cassava, was utilized to screen for regulatory genes and pathways involved in storage root tuberization. The integration of histological observations and microarray data over the early developmental stages of storage root formation pinpoint specific processes in which the developmental switch occurs. Specifically, comparisons of gene expression profiles was developed for four developmental stages of root formation, from prior to the storage root formation to maturation of storage roots, in order to identify a subset of biochemical pathways and regulated genes involved in root tuberization. Three putative candidate genes involved in storage root development are validated for their expression in five organs of the cassava plant.

Identification of key gene-regulated processes in storage root formation has significant implications for improvement programs targeting important quality traits in cassava, such as starch modification, fortified nutrition, post harvest shelf life and yield.

Chapter 1. Research background, goal and specific objectives

INTRODUCTION

Cassava (*Manihot esculenta* Crantz, Euphorbiaceae) is a staple crop in more than 90 countries in the world (CGIAR, 2014). It also is a major source of low-cost carbohydrates for populations in the tropics and is used in many fresh and processed forms for animal feed and as a source of biofuel (El-Sharkawy, 2003; Howeler *et al*., 2013, CGIAR, 2014). The cassava storage root is considered to be a modified adventitious root, enlarged to become a starch storage organ that is utilized by the plant for survival under adverse conditions (Barlow, 1994). As a crop, cassava is propagated from cuttings obtained from the woody stems of plants from the previous growing cycle. Plants establish from such cuttings by formation of adventitious fibrous roots from the basal region of the cut stem and development of the shoot system through axillary bud growth. Photosynthates produced in the developing canopy are translocated downwards and stored as starch within the large, specialized storage root organs. The process of storage root modification from the fibrous to storage state is called tuberization. Tuberization is a complex, poorly understood developmental process that requires the interaction of genetic, environmental and biochemical factors. Kizito (2006) showed that the growth pattern of cassava roots can be modified by changing the nutrient availability. For example, limiting mineral nutrients in a hydroponic system can induce storage root formation. This information raises the question of what is the internal tuberization trigger that induces the signal for storage root formation in

cassava and what approach can we take to elucidate these apparently complicated processes?

To enhance cassava crop quality and yield, it is important to understand the regulatory mechanisms of storage root initiation and tuberization. Cabral *et al.* (2000) demonstrated the existence of five different tissue layers in storage roots. The swelling of a root committed to becoming a storage organ occurs through secondary growth development to cause the proliferation of secondary xylem parenchyma in which starch is stored. Although cassava does not possess storage proteins as seen in potato (patatin) or sweet potato (sporamin), some minor cassava root protein genes have been identified to have a parenchyma-specific expression pattern in the storage roots (Cabral *et al.*, 2000; de Souza *et al.*, 2002; Beltran *et al.*, 2010). Furthermore, the opaque-2 (O2) transcription factors regulating endosperm-specific storage protein genes in maize were found to be differentially expressed in cassava storage roots (de Souza *et al.*, 2003). The *Mec1* gene coding for Pt2L4 glutamic acid-rich protein and a putative RING Zinc Finger and LEA protein genes were reported to be strongly induced in secondary xylem parenchyma of the cassava storage root (de Souza *et al.*, 2004). Moreover, de Souza *et al.* (2006) also reported that the RNA expression patterns of *Mec1*, encoding a glutamic acid-rich protein (Pt2L4), were correlated with maturation of the secondary xylem parenchyma in storage roots. Surprisingly, however, its promoter has been found in transgenic plants to express in a much wider range of tissues types (Beltran *et al.*, 2010). Therefore, the regulatory mechanism of storage root formation still needs to be established at the molecular level.

One way to study the regulatory mechanism of cassava tuberization is to analyze the gene expression profile of storage roots in comparison to non-tuberized roots. Microarray is an effective technology in evaluating the transcriptome of storage root formation. In 2007, Dr. Anderson (USDA, ND) isolated the cassava cDNA library from the model cassava cultivar 60444 (Lokko *et al*., 2007). A total of 4,129 cassava cDNAs as well as 19,808 cDNAs of leafy spurge (*Euphorbia esula*), a closely related species to cassava, were printed onto a microchip for microarray analysis. Utilizing such microarray in collaboration with the USDA laboratory may help elucidate the regulatory mechanism of cassava tuberization.

RESEARCH BACKGROUND AND SIGNIFICANCES

Cassava (*Manihot esculenta* Crantz) is a semi-perennial starchy root crop, which belongs to the family Euphorbiaceae (Kizito, 2006). It ranks second in terms of global production after maize (Howeler *et al.*, 2013; CGIAR, 2014). Although in developing countries the consumption of cassava is mainly in the form of processed flour such as farina and tapioca, it is also used in a wide range of industries including paper, adhesive, textiles, and as a source for the production of ethanol biofuel (Howeler *et al*., 2013). In Africa, cassava is the most important staple food after maize and serves as a food security crop for many sub-Saharan populations (Kizito, 2006; Howeler *et al*., 2013). A major reason for the popularity of cassava is the high starch content of its storage roots. Even though cassava is very valuable for more than 700 hundred million people, numerous constraints limit its potential productivity. The

most important of these are biotic stresses such as viruses, bacteria, mites and mealybugs (El-Sharkawy, 2003). Moreover, a major problem is the rapid postharvest physiological deterioration of the storage roots, which restricts distribution to markets and development of cassava as an industrial crop. (Huang *et al.*, 2001; Reilly *et al.*, 2007; Howeler *et al.*, 2013; Vanderschuren *et al.*, 2014).

The above constraints are being addressed by conventional breeding programs. However, as cassava sets limited seeds and is highly heterozygous, traditional cassava breeding systems take a long time to release improved new varieties (8-12 years) (O'Hair, 1995). Genetic modification via transgenic technologies is, therefore, a suitable approach to addressing these problems but requires effective candidate genes and tissue-specific promoters to generate desired phenotypes within the targeted genetic backgrounds. The storage root is the most important part of cassava for consumers, with the yield determined by the fresh and dry weigh per planting area, in addition to the number of storage roots produced per plant (El-Sharkawy, 2003). Thus a major target for cassava improvement via genetic modification is to study the processes involved in storage root development and thereby identify potential genes and control elements of interest to increase cassava productivity.

External and internal signals must be involved in order to switch a fibrous root, which is a source organ, into a storage root, which is a sink organ. The tuberization process in cassava is probably a complex mechanism, involving several phases with environmental and other biological factors affecting many genes. The storage root of

cassava is not a modified stem as in the case of potato, but it is a modified root that develops to become a large sink for the storage of starch (Shewry, 2003). Because the number of tuberous roots is limited and because they are established early in the growth of the plant, there must be a particular trigger(s) to initiate changes in the morphology and structure of a limited number of fibrous roots such that they develop into storage organs. A cascade of regulatory processes required to develop these organs and permit the storage of starch will follow this. Presently, the triggering system and tuberization process are poorly understood. Increasing knowledge of the genetic and molecular basis of the regulatory mechanism of tuberization will be of great advantage to manipulate the number and the filling of these roots and thus the yield of cassava.

Biology of cassava rooting system

The cassava root system consists of two different adventitious root types: fibrous roots to absorb water and nutrients, and storage roots to store starch (Medina *et al.*, 2007; Lebot, 2009). Vegetative propagation of cassava using mature woody stem cuttings, called stakes, leads to the development of adventitious roots from the nodes near the cut edges (nodal root). Thin fibrous roots also form at the base of the cutting (basal root). Both nodal roots and basal roots are considered to function as for absorption of nutrients and water from the soil at this early stage of development (El-Sharkawy, 2003). Unknown mechanisms cause a subset of the roots to develop secondary growth of xylem parenchyma to store starch and enlarge their size (de Souza *et al.*, 2003). This secondary xylem parenchyma tissue continuously develops until the storage

roots reach maturity. Mature storage roots consist of three layers: bark, peel, and flesh, which botanically are called periderm, cortex (cortical parenchyma and phloem), and xylem (xylem vessel, trachieds) and secondary xylem parenchyma. The starch storage tissue includes inner xylem layer, which has radially distributed xylem vessels (Lebot, 2009). The size and shape of storage roots are also dependent on environment condition and cultivars (Kizito, 2006; Lebot, 2009).

The physiological processes in relation to tuberization in cassava

Cassava has a high yield potential when growth conditions are optimum and plants are cultivated under optimized methods and conditions. A harvest index (the ratio of storage root mass to total plant biomass) higher than 0.5 can be achieved, and typically 6-12 storage roots are produced per plant at a planting density of 10,000 plants/ha (El-Sharkawy, 2003). The photosynthetic rate of the leaf canopy also affects the harvest index of cassava because the association between yield and leaf photosynthesis depends on the biological control of carbon assimilation in cassava (El-Sharkawy, 2003).

Cassava root-specific gene identification

As improvement of economic traits in cassava depends on the quality and the number of tuberous storage roots per plant, understanding the tuberization mechanism in cassava is required for a genetic modification approach. Storage proteins that are specifically expressed in storage organs of other economically important tuber crops have been identified (Shewry, 2003), but not in the case of cassava. Although

cassava has no known root specific storage protein, certain metabolically important proteins have been identified in this organ. However, the biological function of the most abundant root specific protein in cassava storage roots remains unclear. De Souza *et al*. (2002) isolated and first identified the Pt2L4 protein (glutamic acid-rich protein) because expression of this protein could be detected in the xylem storage tissues but not in the peel (periderm plus cortex) layer of the tuberized roots and not in the leaf. Characterization of a cDNA sequence of *Mec1,* which is the gene encoding for Pt2L4 protein, showed expression in vascular tissues of the storage root and might therefore be related to maturation of the secondary xylem parenchyma cells (de Souza *et al*., 2003; 2004; 2006). Zhang *et al*. (2003) identified two cassava storage root-specific cDNAs (c15 related to cytochrome P450 proteins and c54 related to Pt2L4 proteins) via differential screening and isolated its promoters for characterization in transgenic cassava plants. The results demonstrated the two promoters were related to vascular expression and secondary growth of storage root in cassava. Beltran *et al.* (2010) expressed a truncated version of the *Mec1* promoter in cassava and found strong expression in the root xylem parenchyma. However, in both studies expression of the transgenic marker gene also showed in other vascular tissues throughout the plant including leaves and stems (Beltran *et al.*, 2010). Co-expression of Pt2L4 proteins in both secondary xylem parenchyma of the cassava storage root and vascular tissues in stem indicated the role of *Mec1* gene in secondary growth of xylem development and is not specific to the root in cassava. Additionally, the promoter of *GBSSI* gene was characterized in cassava and was shown to be a good candidate for cassava storage root-specific gene (Putten *et al*., 2012). Attempts to identify storage root-specific genes in

cassava have proven difficult because the gene expression in the storage roots also showed in the stem, especially in xylem tissues.

Molecular mechanism model of cassava tuberization

The first attempt to develop a model for root tuberization in cassava was reported in 2007. Medina *et al.* (2007) studied the storage root system using *in vitro* cultures of cassava roots and demonstrated two types of fibrous roots: primary and secondary fibrous roots. Under their conditions, stem segments of cassava produce primary roots *in vitro* that can behave like storage organs. Anatomical analysis of storage organ-derived primary roots *in vitro* showed that proliferation and enlargement of parenchymatous cells occurred in the middle of the cortex and caused radial expansion to form storage root-like organs (Medina *et al.*, 2007). Moreover, Wechkrajang *et al.* (2006) characterized anatomical changes and protein profiles in adventitious and storage roots. Importantly, their results showed that initiation of starch deposition in the primary adventitious root occurs at 35 days after transplanting. These results may indicate the time and site of the initiation of tuberization processes. Carvalho (2010) studied a genome analysis in cassava biodiversity using the microarray technique. He proposed a model for cassava tuberization based on the domestication hypothesis, whereby a growth habit change in cassava is due to the process of domestication, which affects both flowering and storage root development patterns. This is based on the observation that most wild cassava grows in forested, shaded conditions and produce non-tuberous roots with

complete flowering. Conversely, domestic varieties, which are planted under open field conditions, produced tuberous roots with fewer flowers (Carvalho, 2010).

Gene expression analysis in storage roots of cassava.

A transcriptome analysis of gene expression during the process of storage root development is required to advance our understanding of its regulation. Several tools are currently available to facilitate functional genomic analysis in cassava such as a genetic map, cDNA libraries, and expressed sequence tags (EST) library (Anderson *et al.*, 2004; Lokko *et al.*, 2007; Raji *et al.*, 2009; Sojikul *et al.*, 2010; Mitprasat *et al*., 2011; Yang *et al.*, 2011). Moreover, a draft sequence of the cassava genome is currently available to access for cassava genome analysis (http://www.phytozome.net/cassava). An effective genetic transformation system has been developed for characterizing target genes including siRNA technology to facilitate gene knockdowns and for over-expression of desired genes in cassava (Taylor *et al.*, 2012; Zainuddin *et al*., 2012).

Transcriptome analysis using microarray

Microarray technology is a high throughput genome-wide analysis method (Hardiman, 2004; Nguyen and Williams, 2006; Wang *et al.*, 2008). Several microarray platforms have been developed to study genetic and cellular processes. The similarities and differences among these platforms depend on large data set manipulation and the complexity of their experimental target expression (Hardiman, 2004). Time-series microarrays provide information as multiple expression profiles at each time point for

continuous cellular processes. This technique has been routinely applied to identify expression patterns, detect differentially expressed genes, and construct gene networks (Bar-Joseph, 2004; Hardiman, 2004; Opgen-Rhein and Strimmer, 2007; Nguyen and Williams, 2006; Wang *et al.*, 2008). However, the limited sampling problems in the time-series microarrays cause increased potential for misleading analysis (Churchill, 2002; Jørstad *et al.*, 2007). Wang *et al*. (2008) developed short time-series microarrays to enhance the accuracy of data-series analysis, with limited sampling and address simplification-based approaches for integration of multi-source information. Lokko *et al.* (2007) characterized an 18,166 EST dataset for cassava for drought-responsive genes and demonstrated that these ESTs can be useful for developing microarrays and gene-derived molecular markers. Carvalho (2010) has also performed a cDNA- microarray platform in collaboration with USDA Fargo to demonstrate genomic analysis in cassava biodiversity and showed the transcriptomic diversity between wild and domesticated cassava.

The mechanism of tuber formation in potato was reported to be dependent on the regulation of two plant growth hormone families: Gibberellic acid (GA) and Jasmonic acid (JA), and the light quality conditions that regulate expression of GA-responsive plant growth and flowering (Vreugdenhil, 2004; Gottgens and Hedden, 2009). DELLA proteins also appear to be a critical factor in the regulation of GA expression in tuber formation (Carvalho, 2010; James Anderson, pers. com.). Furthermore, Yang *et al*. (2011) used a 60-mer oligonucleotide microarray representing 20,840 cassava genes to identify differentially expressed transcripts in

three types of cassava roots; fibrous roots, developing storage roots, and mature storage roots. This expression profiling of cassava storage root revealed an active process of glycolysis and gluconeogenesis involving sucrose and starch metabolism. However, all the above information is not sufficient to explain the tuberization process in cassava.

Although several research studies have generated informative data including new microarray platforms to increase understanding of many biological processes in plants, the tuberization mechanism in cassava still remains largely unknown. Knowledge of the genetic factors affecting storage root formation is required to explain how they interact with the external environment to switch fibrous roots to become tuberous roots. This research project aims to identify these critical components for storage root formation in cassava.

GOAL AND SPECIFIC OBJECTIVES

The goal of this research was to generate knowledge on cassava storage root formation in both anatomical and molecular contexts and to elucidate the molecular mechanism of cassava tuberization.

This research studied the anatomy and transcriptome profiles across different developmental stages of cassava storage root formation with the following specific objectives:

Objective 1. Determination of the anatomical change in specific root types causing storage root development by histology technique

The process of storage root development remains unclear at the anatomical and molecular levels. Anatomical studies of the various tissues involved in storage root formation provided understanding of where and when a specific type of root switches development into a storage root. This objective used anatomical investigations to study rhizogenesis from cassava stem.

The outcome of this objective was identification of the type of roots that have the potential to form storage tissues for further study of the underlying gene expression profiles in Objective 2. The process of the early xylem parenchyma development may reveal the specific time and zone for the switch to tuberization and lead to finding the triggering mechanism in the tuberization process. The anatomy and structure of cassava storage roots are also informative in elaborating the model of starch storage cell formation at the initiation point of storage root development.

Objective 2. Study of the transcriptome profiling of the tuberization process using microarrays

Our previous experiments attempted to determine the patterning of storage root formation using histology to define the type of root that is competent to become storage roots and also define the particular time and space where development switches from fibrous root to storage root. Information about the molecular triggering of this switching process is lacking. I hypothesized that

tuberization of cassava roots is triggered by the product of specific genes at a particular time and space. The goal of this objective was to identify the major biochemical pathways and/or subset of genes involved in the tuberization process. This research was undertaken in collaboration with scientists at USDA laboratory in Fargo, ND, who have expertise in microarray analysis and gene expression profiling in leafy spurge (*Euphorbiaceae esula*), a Euphorbiaceous species related to cassava. A microchip containing cDNAs of 4,129 cassava cDNA as well as 19,808 cDNAs of leafy spurge representing each specific unigene was used to screen for genes involved in cassava storage root formation. Although, it was proposed to start looking for major differences in gene expression between the developmental stages of storage root formation, an initial microarray analysis was performed comparing leaf and storage roots in order to evaluate the method. This work generated a matrix of genes/pathways involved in root tuberization and a shorter list of candidate genes for further study under Objective 3.

The goal of experiments in Objective 2 was to identify a set/subset of genes involved in the tuberization process of cassava storage roots. This tuberization process probably comprises several different phases, such as triggering, morphological change, filling, and is certainly tightly regulated. This outcome helped to narrow down the minimum number of genes/pathways in the initiation phase of that process and generate a short list of candidate genes for study in Objective 3.

Objective 3. Study of the gene expression of candidate genes involved in cassava root tuberization process

From analysis of the cDNA sequences that appear in expression profiles at each developmental stage, the putative pathway of these genes can reveal the expression of major target genes. However, the tuberization process in cassava may depend not only on up-regulated genes that are highly expressed in storage tissues, but also on down-regulated genes within the storage organ. The matrix data obtained from gene expression profiling using microarray analysis (Objective 2), as well as the expression of candidate genes in each organ, facilitated identification of putative storage root-specific genes. The sequence of each candidate gene was analyzed through the Cassava Genome (Phytozome.net) including the comparative analysis with the UniGene database (http://www.ncbi.nlm.nih.gov/unigene). The functional relationship between the cDNA sequence and the database was analyzed using BLAST (Basic Local Alignment Search Tools; http://blast.st-va.ncbi.nlm.nih.gov/Blast.cgi). The alternative splicing variants of each candidate gene were examined for their expression using RT-PCR to determine the role of each splice variant in different cassava organs, including various types of root organs. Early events that could correspond to the switch from nodal root to tuberous roots were targeted, as were common genes/pathways throughout the tuberization process.

The outcome of this objective was validation of gene expression of identified genes, including their splice variants, that putatively affect the tuberization process in cassava roots. The gene expression patterning of candidate genes in each organ type

helped to indicate their role in the regulation of storage root development. Furthermore, the function of unknown cassava-specific genes might reveal novel genes involved in cassava tuberization for further investigation and possible exploitation via biotechnological approaches.

REFERENCES

Anderson, JV., Delseny, M., Fregene, MA., Jorge, V., Mba, C., Lopez, C., Restrepo, S., Soto, M., Piegu, B., Verdier, V., Cooke, R., Tohme, J. and Horvath, DR. (2004). An EST resource for cassava and other species of Euphorbiaceae. *Plant Molecular Biology*, 56, 527-539.

Bar-Joseph, Z. (2004). Analyzing time series gene expression data. *Bioinformatics,* 20, 2493- 2503.

Barlow, PW. (1994). The Origin, Diversity and Biology of Shoot-Borne Roots. In *The Biology of Adventitious Root Formation.* (Davis, T. D. and Haissig, B. E., eds.). New York: Plenum Press, pp. 1-24.

Beltran, J., Prias, M., Al-Babili, S., Ladino, Y., Lopez, D., Beyer, P., Chavarriaga, P. and Tohme, J. (2010). Expression pattern conferred by a glutamic acid-rich protein gene promoter in field-grown transgenic cassava (*Manihot esculenta* Crantz). *Planta*, 231, 1413-1424.

Cabral, GB., Carvalho, LJCB. and Schaal, BA. (2000). The formation of storage roots in cassava. In *Cassava Biotechnology: Proceedings of the Fourth International Scientific Conference of the CBN, November 3-7, 1998.* (Carvalho, LJCB., Thro, AM., and Vilarinhos, AD., *eds*.). Brasilia: EMBRAPA, CENARGEN, CBN, pp. 345-356.

Carvalho, LJCB. (2010). Gene expression analysis in storage root of cassava using microarray data. http://www.slideshare.net/CIAT/carvalho-gines-mera2010.

CGIAR. (2014). CASSAVA. http://www.cgiar.org/our-research/crop-factsheets/cassava

Churchill, GA. (2002). Fundamentals of experimental design for cDNA microarrays. *Nature Genetics,* 32, 490-495.

De Souza, CRB., Carvalho, LJCB., de Almeida, ERP. and Gander, ES. (2002). Identification of cassava root protein genes. *Plant Foods for Human Nutrition*, 353-363.

De Souza, CRB., de Almeida, ERP., Carvalho, LJCB. and Gander, ES. (2003). Studies toward the identification of transcription factors in cassava storage root. *Brazilian Journal of Plant Physiology,* 15, 167-170.

De Souza, CRB., Carvalho, LJCB. and Cascardo, J. (2004). Comparative gene expression study to identify genes possibly related to storage root formation in cassava. *Protein Pept Lett.* 11: 577-582.

De Souza, CRB., Carvalho, LJCB., de Almeida, ERP. and Gander, ES. (2006). A cDNA sequence coding for a glutamic acid-rich protein is differentially expressed in cassava storage roots. *Protein Pept Lett.* 13, 653-657.

El-Sharkawy, MA. (2003). Cassava biology and physiology. *Plant Mol Biol* 53: 621-641.

Gottgens, EM. and Hedden, P. (2009). Gibberellin as a factor in floral regulatory networks. *J. Exp. Bot*. 60(7): 1979-1989.

Hardiman, G. (2004). Microarray platforms-comparisons and contrasts. *Pharmacogenomics,* 5, 487-502.

Howeler, R., Lutaladio, N. and Thomas, G. (2013). "Save and grow: cassava, a guide to sustainable production intensification." FAO, Rome (http://www.fao.org/3/a-i3278e.pdf) Cited 2014.

Huang, J., Bachem, C., Jacobsen, E. and Visser, RGF. (2001). Molecular analysis of differentially expressed genes during postharvest deterioration in cassava (*Manihot esculenta* Crantz) tuberous roots. *Euphytica*, 120, 85–93.

Jørstad, TS., Langaas, M. and Bones, AM. (2007). Understanding sample size: what determines the required number of microarrays for an experiment? *Trends in Plant Science*, 12, 46-50.

Kizito, EB. (2006). Genetic and root growth studies in Cassava (*Manihot esculenta* Crantz): Implications for Breeding. Acta Universitatis Agriculturae Sueciae 2006:82. Uppsala: Swedish University of Agricultural Sciences, pp. 127.

Lebot, V. (2009). Tropical root and tuber crops: cassava, sweet potato, yams and aroids. In: Crop Production Science in Horticulture, V. 17(eds) Wallingford, UK, CAB Books, CABI.

Lokko, Y., Anderson, JV., Rudd, S., Raji, A., Horvath, D., Mikel, MA., Kim, R., Liu, L., Hernandez, A., Dixon, AG. and Ingelbrecht, IL. (2007). Characterization of an 18,166 EST dataset for cassava (*Manihot esculenta* Crantz) enriched for drought-responsive genes. *Plant Cell Reports*, 26, 1605-1618.

Medina, RD., Faloci, MM., Gonzalez, AM. and Mroginski, LA. (2007). *In vitro* cultured primary roots derived from stem segments of cassava (*Manihot esculenta*) can behave like storage organs. *Annals of Botany*, 99, 409-423.

Mitprasat, M., Roytrakul, S., Jiemsup, S., Boonseng, O. and Yokthongwattana, K. (2011). Leaf proteomic analysis in cassava (*Manihot esculenta* Crantz) during plant development, from planting of stem cutting to storage root formation. *Planta*. 233:1209-1221.

Nguyen, NK., and Willams, ER. (2006). Experiment designs for 2-colour cDNA microarray experiments. *Applied Stochastic Models in business and Industry*. 22:631-638.

O'Hair, SK. (1995). Cassava. In *New Crop Fact Sheet.* Tropical Research and Education Center, University of Florida. http://www.hort.purdue.edu/newcrop/cropfactsheets/cassava.html.

Opgen-Rhein, R. and Strimmer, K. (2007). Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach. *Statistical Applications in Genetics and Molecular Biology*, 6, Article 9.

Putten, HJJK., Wolters, AMA., Pereira-Bertram, IM., Berg, HHG, Krol, ARVD., and Visser, RGF. (2012). Cloning and characterization of tuberous root-specific promoter from cassava (*Manihot esculenta* Crantz). *Planta*. 236 (6): 1955-1965.

Raji, AA., Anderson, JV., Kolade, OA., Ugwu, CD., Dixon, AG. and Ingelbrecht, I. L. (2009). Gene-based microsatellites for cassava (*Manihot esculenta* Crantz): prevalence, polymorphisms, and cross-taxa utility. *BMC Plant Biology,* 9, 118.

Reilly, K., Bernal, D., Cortes, D. F., Gomez-Vasquez, R., Tohme, J. and Beeching, JR. (2007). Towards identifying the full set of genes expressed during cassava post-harvest physiological deterioration. *Plant Molecular Biology,* 64, 187-203.

Shewry, PR. (2003). Tuber storage proteins. *Annals of Botany,* 91, 755-769.

Sojikul, P., Kongsawadworakul, P., Viboonjun, U., Thaiprasit, J., Intawong, B., Narangajavana, J. and Svasti, MR. (2010). AFLP-based transcript profiling for cassava genome-wide expression analysis in the onset of storage root formation. *Physiologia Plantarum,* 140, 189-198.

Taylor, N., Gaitan-Solis, E., Moll, T., Trauterman, B., Jones, T., Pranjal, A., Trembley, C., Abernathy, V., Corbin, D. and Fauquet, C. (2012). A high-throughput platform for the production and analysis of transgenic cassava (*Manihot esculenta*) plants. *Tropical Plant Biol*. 5:127-139.

Vanderschuren, H., Nyaboga, E., Poon, JS., Baerenfaller, K., Grossmann, J., Hirsch-Hoffmann, M., Kirchgessner, N., Nanni,P. and Gruissem W. (2014). Large-scale proteomics of the cassava storage root and identification of a target gene to reduce postharvest deterioration. *The plant cell*. 26:1913-1924.

Vreugdenhil, D. (2004). Comparing potato tuberization and sprouting opposite phenomena. *Amer J of Potato Res* (2004) 81:275:280.

Wang, X., Wu, M., Li, Z. and Chan, C. (2008). Short time-series microarray analysis: methods and challenges. *BMC Systems Biology*, 2, 58.

Wechkrajang, S., Vichukit, V., Toonkool, P. and Nakasathien, S. (2006). Anatomical characterization and protein profiles in adventitious and storage roots of two commercial Thai cassava varieties. *Kasetsart Journal (National Science),* 40, 297 - 305.

Yang, J., An, D. and Zhang, P. (2011). Expression profiling of cassava storage roots reveals an active process of glycolysis/gluconeogenesis. *Journal of Integrative Plant Biology,* 53, 193-211.

Zainuddin, IM., Schlegel, K., Gruissem, W. and Vanderschuren, HV. (2012). Robust transformation procedure for the production of transgenic farmer-preferred cassava landraces. *Plant Methods*. 8:24.

Zhang, P., Bohl-Zenger, S., Puonti-Kaerlas, J., Potrykus, I. and Gruissem, W. (2003). Two cassava promoters related to vascular expression and storage root formation. *Planta,* 218, 192-203.

Chapter 2. Anatomical assessment of root formation and tuberization in cassava (*Manihot esculenta* Crantz)

This chapter has been submitted to the international journal *Tropical Plant Biology* as Yeetoh Chaweewan and Nigel Taylor. YC and NT designed the experiments, YC performed the experimental work, and YC and NT wrote the manuscript.

ABSTRACT

Formation of storage root organs in cassava (*Manihot esculenta*) is poorly understood, but considered to occur when a subset of fibrous roots receive unknown signals to undergo secondary thickening. Large amounts of secondary xylem parenchyma are then produced in which starch is synthesized and stored. Anatomical studies were undertaken to examine rhizogenesis from greenhouse-grown cassava stem cuttings. Root formation was observed from the stem cut end (basal) and from close to the buried nodes (nodal) 5-10 days after planting. Transverse sectioning of the stem provided evidence that the basal roots were initiated from the cambium, while the nodal-derived roots developed from tissues deeper within the stem, at the boundary of the xylem and pith. Basal root anatomy remained constant with age, with minimal development of metaxylem. No tuberization was seen to occur from the basal roots. In contrast, nodal roots produced significant amounts of metaxylem and subsequently secondary xylem to form a large central stele. Further development established the storage organ in which secondary xylem parenchyma, tracheids and vessels were produced from the cambium. As a result, the nodal-derived roots were seen as precursors of the storage organs. It is proposed that nodal-derived and basal-derived fibrous roots are fundamentally different organs, that

they originate through different rhizogenic processes, and are committed to their respective developmental fates from the earliest stages of their initiation. These anatomical investigations offer new insight into root tuberization in cassava and should guide better focused studies into the underlying molecular and developmental control mechanisms.

Keywords: cassava, root anatomy, storage root development, tuberization, xylem differentiation

INTRODUCTION

Cassava (*Manihot esculenta* Crantz, Euphorbiaceae) is a staple crop and major source of low-cost carbohydrates in the tropics, where it is used in many fresh and processed forms for human food, animal feed, and increasingly as a source of biofuel (El-Sharkawy, 2003; Howeler *et al.*, 2013). When grown from seed, cassava produces a tap root which, along with some adventitious roots, develops to become storage organs. As a crop, however, cassava is almost exclusively propagated via woody stem section cuttings obtained from plants of the previous growing cycle. Plants establish by adventitious formation of roots from the basal region of the cut stem and by development of the shoot system through axillary bud growth. Once established, photosynthates produced in the developing canopy are diverted from the shoot, translocated downwards and stored as starch within large, specialized storage root organs. Depending on the cultivar and growing conditions, between three and fourteen storage roots are formed by each plant (Alves, 2002; Howeler *et al.*, 2013).

Despite its value for more than 700 hundred million people in the world, full potential of the crop is limited by numerous constraints, including biotic and abiotic stresses (El-Sharkawy, 2003; Sakurai *et al.*, 2007; Howeler *et al.*, 2013) and rapid postharvest physiological deterioration of the tuberous roots (Reilly *et al.*, 2007; Vanderschuren *et al.*, 2014). Improvement programs employing conventional breeding, and more recently biotechnology, have focused on addressing these constraints whilst also striving to improve dry matter content, starch quality and nutritional value of the storage roots (Sayre *et al.*, 2011; Ceballos *et al.*, 2012). In order to achieve these goals, a full understanding of the processes and biology of storage root development in cassava is required. Few reports are available describing the mechanisms underlying cassava storage root formation. As a result, knowledge of tuberization is limited to descriptions of secondary thickening within a subset of the fibrous roots, which subsequently become the storage organs (Alves, 2002; El-Sharkawy, 2003). It is not known which signals are responsible for triggering such conversion from fibrous to storage root or what controls which and how many roots switch development in this manner.

Lowe *et al.* (1982) describe two types of root production from newly planted stem cuttings. One takes place from the cut end (basal) and the other from close to the submerged nodes (nodal). They reported continual production of basal roots, while the nodal root numbers remained at around four per plant. Storage organs developed from both root types. Investigations of the early stages of storage root development indicate that the initial phase of tuberization takes place approximately six weeks after transplantation of stem cuttings into soil, at which time radial thickening becomes visible

in some root structures (Cock *et al.*, 1979; Lowe *et al.*, 1982). Anatomical studies have described radial swelling of the root to occur through the production of secondary xylem resulting in swelling and development of the storage organ (Lowe *et al.*, 1982; Wechkrajang *et al.*, 2006). Five distinct tissue layers are present within the developing and mature storage root (Cabral *et al*., 2000; de Souza *et al.*, 2002). Outermost is the peel consisting of the peridermal tissue, followed inwardly by the secondary phloem and cambium. The inner portion, or flesh, consists of secondary xylem parenchyma, xylem vessels and primary xylem. Starch is stored within parenchymatous cells of the secondary xylem, which undergoes massive proliferation to form the bulk of the mature storage root. Enrichment of Pt2L4, a protein associated with secondary growth of xylem parenchyma, is observed in this tissue (de Souza *et al.*, 2004; 2006).

We report here anatomical studies of root formation in cassava. Two types of adventitious root organs are described. We provide evidence that storage roots do not develop from the fibrous roots and that the two root types are produced as unique organs derived from different locations and tissue types within the stem.

MATERIALS AND METHODS

Plant material

Plants of cultivar 60444 were established from stem cuttings. Semi-woody stem sections 18-20 cm long consisting of six to eight nodes and a minimum diameter of 1 cm were obtained from greenhouse, pot-grown plants. Cuttings were planted 2-4 nodes deep in 12 cm pots containing Fafard 51 potting media. Stakes were uprooted every five days after

transferring potting media in order to visually examine development of the root system. Tissue culture-derived plants of cv. 60444 and TME204 were established in Fafard 51 as described by Taylor *et al*. (2012). All plants were fertilized twice weekly with 15-5-15 fertilizer (Jack's Professional LX, JR Peters Inc, PA) and micronutrient solution (MOST, JR Peters Inc, PA). Plants were grown on the open bench and under natural and supplemental lighting at 900 $\mu M/m^2$ for 16 hours per day, at 26-28^{o}C and 60% humidity.

Dissection of plant materials and histological staining

Plants were removed from their pots, and storage root and fibrous roots were washed thoroughly with tap water to remove soil. Roots were dissected by free-hand sectioning using a two-edged razor blade. In case of the harder tissues such as the semi-woody parts of the stem, a sliding microtome (Uchida Yoko, #162-3012 VWR) was used to obtain thin sections of 20-50 micron thickness. Woody samples were wrapped at one end with a soft supporting material such as a thin paper towel or artificial pith to avoid damaging samples while aligning in the microtome chamber. Sectioned samples were transferred to distilled water in a Petri dish to prevent desiccation prior to the staining procedure.

Toluidine blue O (C.I. 52040, Sigma) was used to study root anatomical structure. A 0.1% solution (w/v) was prepared in distilled water. Intact thin sections floating in water were chosen under the light box and transferred using fine forceps into separate clean Petri dishes. Two drops of Toluidine blue O solution were placed directly onto the thin section in the Petri dish using a pipette and left for one minute. Tissues were then thoroughly rinsed three times with distilled water or until the excess dye was completely

removed. Individual sections were removed from Petri dish and mounted on a microscopic slide using a cover slip with one drop of water for examination under the dissecting microscope. For observation of starch within storage tissues, sections were stained with potassium iodine (KI) reagent. A 2% (w/v) potassium iodine (Sigma Aldrich, #221945) solution was prepared in distilled water and stored in a tightly sealed dark glass bottle. The KI reagent was applied by dropping directly onto the thin sections and incubating for 3-5 minutes, depending on the thickness of specific samples.

Microscopy and photography

Stained sections of various parts of plants were examined under the microscope (Nikon SMZ1500). Images were recorded using a high-sensitivity CCD color camera system (QIMAGING RETIGA 1300) using the software program Qcapture (version 3.1.3.5) to capture the images in RBG mode.

RESULTS

Morphology of cassava rooting system

Stem cuttings were removed from their pots every five days and visually inspected for development of the adventitious root system. Roots first became visible five days after planting, developing from swollen tissues around the circumference of the basal cut end of the stem. These basal roots were white colored, fleshy, prolific and fragile in nature. They could easily be broken when handled and grew rapidly to fill the available space within the potting medium (Fig. 1A). A second type of root structure was seen to develop originating from, or very close to, the buried axillary buds. These structures became

visible from five to ten days after planting. The nodal root structures were lesser in number at three to five per node, and thicker in diameter than the fibrous roots produced from the basal cut end. Initially white in color, they developed to become brown with a corky surface, grew rapidly in length and were strongly connected to the stem. Lateral roots were seen to develop from both the nodal and basal roots and together developed to form the root mass (Fig. 1A). In contrast to the fragile nature of the fibrous (basal) roots, the whole plant could be picked up and suspended when held by one nodal root. Swollen storage roots were first observed 5-6 weeks after planting. In every cutting examined, the storage roots developed from radial expansion of the nodal-derived roots. In the majority of cases, the storage roots possessed a distinct narrow neck (peduncle) region closest to the stem that connected the latter from the swollen storage region of the tuberized root (Fig. 1B) (Lebot, 2009). In this study, no storage roots were seen to develop from the fibrous roots produced from the basal cut surface of the stem cutting.

Anatomy and origin of fibrous roots and storage roots structures

As described above, stem cuttings developed swollen tissues at their wounded cut edge after planting in soilless compost. The fibrous basal roots subsequently emerged from this tissue. Due to the semi-woody nature of stem, the outer layer consisting of the periderm could be easily peeled away and separated from the inner portion consisting of the xylem and pith. When performed at the swollen end undergoing fibrous root formation, all root structures could be removed from the stem along with the peel layer (Figs. 2A and B). The fibrous roots remained intact during this process and showed no structural connection through the cambium layer into the woody xylem tissue. A distinctly different

pattern was seen when the periderm was peeled away from around the submerged nodes. In this case, the nodal-derived roots were not removed but remained connected through the peel directly into the inner woody xylem tissues (Fig. 2C).

The anatomy of fibrous basal roots was examined by sectioning and histological staining with toluidine blue at increasing ages for up to two months after cuttings were transferred to potting compost. When sectioned transversely at five days after planting, the fibrous roots were seen to possess a distinct cortical region with a central stele. The stele carried an arch number of five or six with a highly visible, densely staining region associated with the pericycle (Figs. 3A and B). As the root aged, the cortex was maintained with production of metaxylem to produce a clear star shape consisting of xylem vessels (Fig. 3C). Sections taken from basal-derived fibrous roots up to 60 days after planting, as previously reported (Wechkrajang *et al.*, 2006), showed no further development of the metaxylem from that seen in Figure 3C. This pattern was maintained across more than 20 such roots examined, with the discreet pentarch system present at all positions from close to the root tip to the root's origin at the stem. Sectioning was also performed as close as possible to the cut edge of the stem in an attempt to identify the earliest stages and origin of the fibrous roots. Figures 3D and E show the presence of newly formed basal root primordial in transverse sections. These organs are seen developing from the vascular cambium of the stem. In both cases, they are yet to penetrate the peridermal region and have no visible connection to the underlying secondary xylem tissue. This is further illustrated in the longitudinal section of a young basal root in Figure 3F.

Similar sectioning was performed on nodal-derived root structures. Transverse sections of young nodal roots less than 2 cm long developing from the stem revealed an anatomy very similar to that of the basal roots. A distinct pericycle is visible with primary xylem seen as circular groups of protoxylem vessels present in a pentarch manner (Fig. 4A). As the nodal roots aged, development of the vascular tissue diverged from that of the basal roots, with production of metaxylem taking place to fill the center of the root structure (Fig. 4B). With increasing age of the organ, xylem tissues continued to be produced through secondary growth to form a distinct region of tracheids and vessels within the center of a greatly enlarged stele (Fig. 4C). Stem tissues in the region of the submerged node were sectioned transversely. Figure 4D shows root primordia within a cross section of the nodal region. The young root structure has yet to penetrate the periderm and can be seen growing through the lignified (blue stained), secondary xylem tissues. A similar root is shown in greater detail in Figure 4E. Likewise, it is not developing from the region of cambium, but appears derived from tissues deeper within the stem. Figure 4F provides further illustration, showing a transverse section of a peeled stem (cortex and phloem tissues have been removed) from which a more mature nodal root has emerged. The base of the root structure is seen to be completely embedded within the lignified xylem as deep as the edge of the central pith.

More mature developmental stages of the storage root structure were examined by both toluidine blue and potassium iodide (KI) staining. Transverse sectioning of the neck region (Fig. 1B) revealed the anatomy shown in Figure 5A, in which the vascular cambium has been established as a continuous cylinder running along the storage root's

longitudinal axis. Secondary phloem is seen external to the cambium. Significant production of secondary xylem has occurred internally to the cambium, consisting mostly of lignified tracheids and large vessels, with lesser amounts of cellulosic (purple stained) xylem parenchyma cells. Examination of sections taken along the storage root axis further from the stem reveals a change in composition of the secondary xylem. Concomitant with radial expansion of the organ, presence of tracheids and vessels remain prominent, but a significantly larger portion of the xylem tissue is composed of starch-containing parenchyma cells (Fig. 5B). Further progression distally from the stem culminates in the anatomy seen in Figure 5C, in which the vast majority of the tissue internal to the cambium consists of cellulosic secondary xylem parenchyma that stains deeply for presence of starch (Fig. 5F). Minimal xylem vessels are present scattered within the starch-containing cells and at the central xylem strand. In contrast to sections taken closer to the stem (Fig. 5A), tracheid elements are not conspicuous and may be absent within this storage tissue. Presence of starch in the tissues described above (Figs. 5D-F) is present in an inverse relationship between lignified xylem and cellulosic cells.

Plants were removed from their pots 12-16 weeks after planting and the stems split longitudinally in the region close to, and below, the soil surface. Storage roots at differing stages of radial development were examined. In all cases, a strong continuous connection was observed between the storage organs and the stem. Connection of the storage root can be seen to take place with the secondary xylem of the stem, internal to the vascular cambium such that the secondary xylem tissues of the storage root are continuous with the vascular xylem system of the semi-woody stem (Fig. 6).

DISCUSSION

A study was undertaken to determine patterns of root development from cassava stem cuttings. Observations confirmed earlier reports that roots develop both from the cut end of the stem and from the nodal regions closest to the stem base (Lowe *et al.*, 1982). Evidence is provided that the basal and nodal root structures originate in a similar timeframe but in different manners from different regions of the stem anatomy. While appearing at their earliest stages to be similar, their anatomies rapidly diverge, with only the nodal roots undergoing secondary thickening to develop as storage organs. In this study, basal roots developed from the swollen tissues initiated by the wound response at the stem cut end and were observed to originate from the cambium exterior to the xylem. In contrast, nodal roots originated from significantly deeper within the stem. The latter were seen to grow out through the secondary xylem before penetrating the cambium and periderm tissues. As a result, the nodal roots were deeply anchored into the central portion of the stem and could not easily be detached (Fig. 2B, Fig. 4F, Fig. 6). This contrasts with the more superficial basal roots that are easily stripped away from the stem along with the peel. The present study does not provide clarity concerning which meristematic cells the nodal roots are derived from. However, due to the depth of their origin and apparent location at the edge of the central pith and innermost xylem, they might be initiated from remnants of the pericycle associated with the protoxylem. These tissues are known to be associated with lateral root rhizogenesis, but are more commonly described in relation to fibrous roots (Esau, 1977). Further studies are required to investigate this in detail.

Although the present investigations were performed using relatively young, semi-woody stem cuttings and plants were rooted in 12 cm pots, important information is provided about the earliest stages of storage root formation in cassava. Established thinking states that the storage roots develop from a subset of the fibrous roots, but how such a conversion occurs has not been described. We provide evidence, and propose here, that the storage and fibrous roots are fundamentally different organs, that they originate through different rhizogenic processes, are almost certainly not interchangeable, and are committed to their respective developmental fates from their earliest stages of initiation. The nodal-derived roots are therefore proposed to be precursor structures for the storage organs. They are initiated early in the rooting process but do not undergo significant radial expansion until receiving signals from the developing stem. It is not clear from earlier work how the stem cuttings were prepared before planting and how this relates to observations that storage roots could arise from both basal and nodal-derived roots (Lowe *et al*., 1982). In the present study, care was taken to cut the stem at the internode. However, if the stem is cut close to a node, storage root structures would develop in close proximity with the basal fibrous roots and could cause confusion with regards to their respective origins.

Consideration that the two root types have fundamentally different origins and developmental fates has important implications for studies of storage root formation in the field and laboratory. Increasingly powerful tools are available to study plant development at the genomic, transcriptomic and proteomic levels. Such tools have been applied to investigate storage root formation and development in cassava (Li *et al.*, 2010;

Yang *et al.*, 2011; Vanderschuren *et al*., 2014). Like Yang *et al.* (2011), who utilized cDNA microarray technology to study storage roots at different developmental stages, previous experiments in our laboratory have failed to distinguish between the fibrous basal roots and early stage nodal roots. These two organ types can appear similar within a root mass at times before the latter undergoes secondary thickening and radial expansion. Failure to distinguish between them risks sampling a mixed population of organ types, with the resulting data likely to provide misleading results. Conversely, the ability to distinguish between fibrous roots and nodal-derived roots allows these organs to be segregated prior to analysis and facilitates improved studies on early tuberization processes. In this manner, modern molecular tools, bioinformatics and associated analytical tools could be utilized more effectively to elucidate the molecular mechanisms underlying root development in cassava.

In addition to shedding light on the origin of cassava storage roots, the anatomical studies of early storage root development described here raise questions as to how differentiation of the secondary xylem occurs within the storage organ. Initial development of the nodal root organ results in production of a central xylem consisting of tracheids and vessels (Fig. 4C). In the present study, the neck region, which connects the storage tissues to the shoot (Lebot, 2009), continues to lay down lignified xylem tissue, causing radial thickening with minimal presence of starch-containing parenchyma (Figs. 5A and D). As the root transitions distally away from the stem, the cambium still produces xylem, but a larger proportion of the new cells differentiate to form xylem parenchyma and not conducting vessels and tracheids. This trend continues with the vast

majority of the xylem differentiating as storage tissue and not lignified conducting tissues (Fig. 5C). The processes responsible for controlling transition of the nodal root into a storage organ are not understood. Hormonal signals are thought to be sent to the roots from the stem system, most likely via the phloem. However, what these are, how they stimulate the cambium, and how they influence cellular differentiation to produce lignified conducting cells or cellulosic storage cells within the nodal-derived roots is not known. Gaining better understanding of these questions is important as the underlying mechanisms most likely control the number, size, shape and timing of root tuberization in cassava. It could also provide an effective model of studies for cellular lignification, an area of significant importance (Bonawitz and Chapple, 2010).

Continual connection of storage root secondary xylem with the stem secondary xylem is described in these studies (Fig. 6) and may have implications for genetic manipulation of the storage root. The woody stem is also a starch storage organ in cassava, a trait most likely selected for because this organ is used as the propagule. This may explain difficulties, to date, in developing storage root specific promoters for cassava, and determine why transgene-promoter fusions engineered for expression in the storage root also express in the stem (Zhang *et al*., 2003; Beltran *et al*., 2010).

The anatomical investigations described here offer new insight into the root tuberization process in cassava. Evidence is provided that two types of roots develop from cassava stem cuttings and that storage organs develop only from the nodal-derived organs. It is hoped that this study generates new appreciation and interest in how cassava

storage roots are produced, and facilitates better designed studies on the underlying molecular and developmental control mechanisms.

ACKNOWLEDGEMENTS

This work was supported by a scholarship for Y. Chaweewan from the Royal Thai Government and the Donald Danforth Plant Science Center. The authors thank Drs. Howard Berg and Christopher Topp (DDPSC) for valuable discussions and DDPSC greenhouse staff for care of the plants.

REFERENCES

Alves, A. (2002). Cassava botany and physiology. In: Cassava: biology, production and utilization. Hilllocks R J, Thresh J M, Bellotti A C (eds), CAB International 67-90.

Beltran, J., Prias, M., Al-Babili, S., Ladino, Y., Lopez, D., Beyer, P., Chavarriaga, P. and Tohme, J. (2010). Expression pattern conferred by a glutamic acid-rich protein gene promoter in field-grown transgenic cassava (Manihot esculenta Crantz). *Planta* 231: 1413-1424.

Bonawitz, ND. and Chapple, C. (2010). The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annu Rev Genet* 44: 337-363.

Cabral, GB., Carvalho, LJCB. and Schaal, BA. (2000). The formation of storage roots in cassava. Cassava Biotechnology: Proceedings of the Fourth International Scientific Conference of the CBN, November 3-7, 1998. Carvalho LJCB, Thro AM, Vilarinhos A D. Brasilia, EMBRAPA, CENARGEN, CBN: 345-356.

Ceballos, H., Kulakow, P. and Hershey, C. (2012). Cassava breeding: current status, bottlenecks and the potential of biotechnology tools. *Trop Plant Biol* 5: 73–87.

Cock, JH., Franklin, D., Sandoval, G. and Juri, P. (1979). The ideal cassava plant for maximum yield. *Crop Science* 19: 271-279.

De Souza, CRB., Carvalho, LJCB., De Almeida, ER. and Gander, ES. (2002). Towards the identification of cassava root protein genes. *Plant Foods Hum Nutr* 57: 353-363.

De Souza, CRB., Carvalho, LJCB., and Cascardo J (2004). Comparative gene expression study to identify genes possibly related to storage root formation in cassava. *Protein Pept Lett* 11: 577-582.

De Souza, CRB., Carvalho, LJCB., de Almeida, ER., and Gander, ES. (2006). A cDNA sequence coding for a glutamic acid-rich protein is differentially expressed in cassava storage roots. *Protein Pept Lett* 13: 653-657.

El-Sharkawy, MA. (2003). Cassava biology and physiology. *Plant Mol Biol* 53: 621-641.

Esau, K. (1977). Anatomy of seed plants. New York, John Wiley & Sons.

Howeler, R., Lutaladio, N. and Thomas, G. (2013). "Save and grow: cassava, a guide to sustainable production intensification." FAO, Rome (http://www.fao.org/3/a-i3278e.pdf) Cited 2014.

Lebot, V. (2009). Tropical root and tuber crops: cassava, sweet potato, yams and aroids. In: Crop Production Science in Horticulture, V. 17(eds) Wallingford, UK, CAB Books, CABI.

Li, K., Zhu, W., Zeng, K., Zhang, Z., Ye, J., Ou, W., Rehman, S., Heuer, B. and Chen, S. (2010). Proteome characterization of cassava (*Manihot esculenta* Crantz) somatic embryos, plantlets and tuberous roots. *Proteome Sci* 8: 10.

Lowe, SB., Mahon, JD. and Hunt, LA. (1982). Early development of cassava (*Manihot esculenta*). *Can J Bot* 60(12): 3040-3048.

Reilly, K., Bernal, D., Cortes, DF., Gomez-Vasquez, R., Tohme, J. and Beeching, JR. (2007). Towards identifying the full set of genes expressed during cassava post-harvest physiological deterioration. *Plant Mol Biol* 64: 187-203.

Sakurai, T., Plata, G., Rodriguez-Zapata, F., Seki, M., Salcedo, A., Toyoda, A., Ishiwata, A., Tohme, J., Sakaki, Y., Shinozaki, K. and Ishitani, M. (2007). Sequencing analysis of 20,000 full-length cDNA clones from cassava reveals lineage specific expansions in gene families related to stress response. *BMC Plant Biol* 7: 66 doi 10.1186/1471-2229-1187-1166.

Sayre, R., Beeching, JR., Cahoon, EB., Egesi, C., Fauquet, C., Fellman, J., Fregene, M., Gruissem, W., Mallowa, S., Manary, M., Maziya-Dixon, B., Mbanaso, A., Schachtman, DP., Siritunga, D., Taylor, N., Vanderschuren, H. and Zhang, P. (2011). The BioCassava Plus program: biofortification of cassava for sub-Saharan Africa. *Annu Rev Plant Biol* 62: 251-272.

Taylor, N., Gaitan-Solis, E., Moll, T., Trauterman, B., Jones, T., Pranjal, A., Trembley, C., Abernathy, V., Corbin, D. and Fauquet, CM. (2012). A high-throughput platform for the production and analysis of transgenic cassava (*Manihot esculenta*) plants. Trop Plant Biol 5: 127-139.

Vanderschuren, H., Nyaboga, E., Poon, JS., Baerenfaller, K., Grossmann, J., Hirsch-Hoffmann, M., Kirchgessner, N., Nanni, P. and Gruissem, W. (2014). Large-scale proteomics of the cassava storage root and identification of a target gene to reduce postharvest deterioration. *Plant Cell* 26: 1913-1924.

Wechkrajang, S., Vichukit, V., Toonkool, P. and Nakasathien, S. (2006). Anatomical characterization and protein profiles in adventitious and storage roots of two commercial Thai cassava varieties. *Kasetsart J (Nat Sci)* 40: 297 – 305.

Yang, J., An, D. and Zhang, P. (2011). Expression profiling of cassava storage roots reveals an active process of glycolysis/gluconeogenesis. *J Integr Plant Biol* 53: 193-211.

Zhang, P., Bohl-Zenger, S., Puonti-Kaerlas, J., Potrykus, I. and Gruissem, W. (2003). Two cassava promoters related to vascular expression and storage root formation. *Planta* 218: 192-203.

Figure 1. Morphology of 60-day-old greenhouse grown plants derived from a stem cutting **A.** Shoot and root system **B.** Detail of root system, showing storage roots developing from the nodal region of the stem cutting. Massive proliferation of the fibrous root has occurred from base of the stem cutting. Three storage roots (1-3) are clearly visible. 1 and 2 have undergone radial development and are developing as storage organs, 3 has yet to initiate significant radial expansion. n indicates neck (peduncle) region that connects storage region of the tuberizing root to the stem

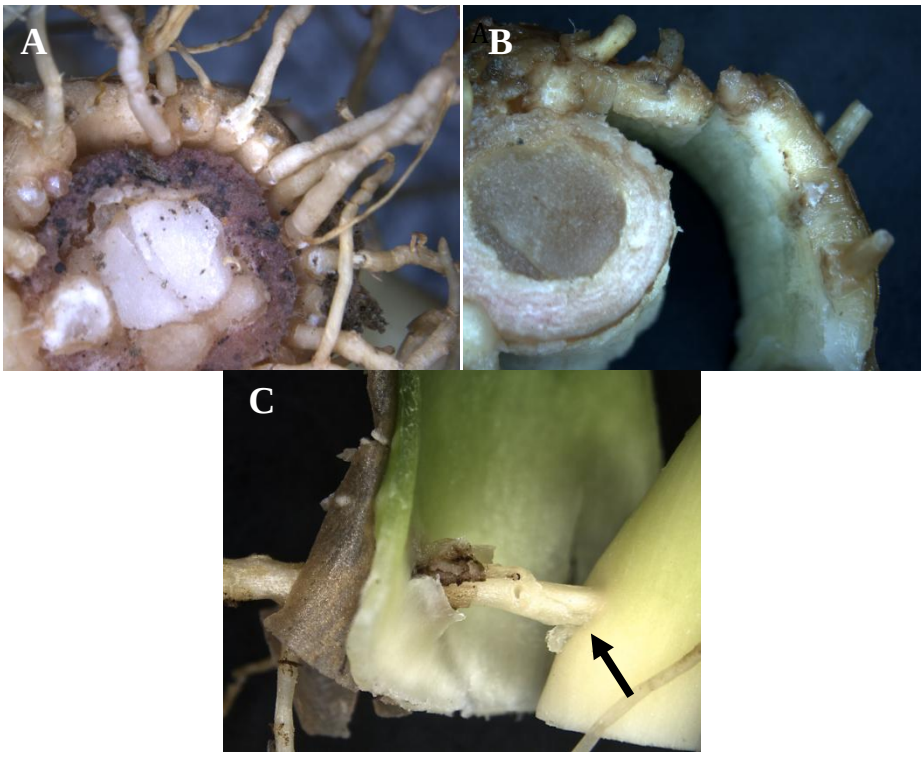

Figure 2. Attachment of basal and nodal derived roots to the stem **A.** Basal roots arising from swollen white tissues at the base of the semi-woody cut stem. **B.** Young roots removed with the periderm when the bark is stripped away from the lignified central tissues **C.** Nodal derived root. Peeling does not remove the root structure, which is seen to be emerging from within the stem (arrowed) and penetrating through the peeled bark.

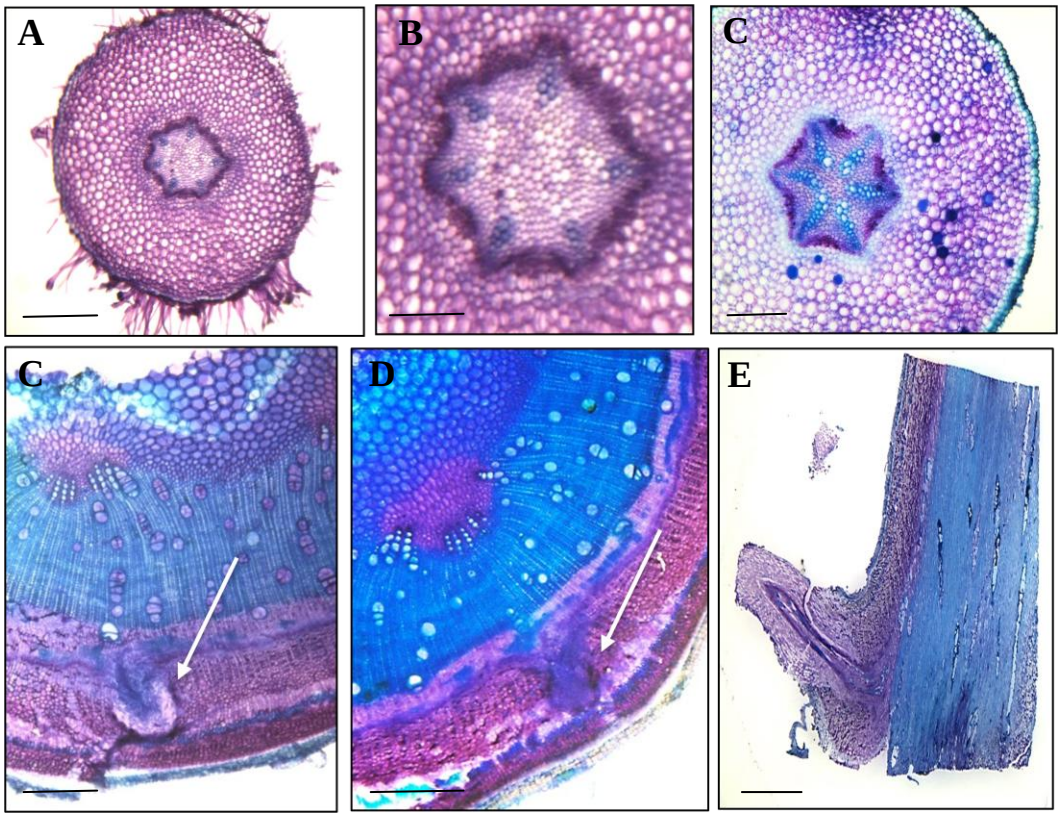

Figure 3. Thin sections of developing basal roots stained with toluidine blue. **A-C.** Transverse sections of basal roots. **A.** Section less than 1 cm from root tip showing distinct dark staining pericycle. (bar = 0.5 mm) **B.** Detail of vascular cylinder in young basal root, with six visible primary xylem poles. (bar = 0.25 mm) **C.** Root section 5-10 cm from tip showing development of metaxylem. (bar = 0.25 mm) **D&E**. Transverse sections of semi-woody stem showing basal root primordia (arrowed) developing from the cambium. (bar = 1 mm) **F.** Longitudinal section of young root developing from basal cut edge of stem. (bar = 2.5 mm)

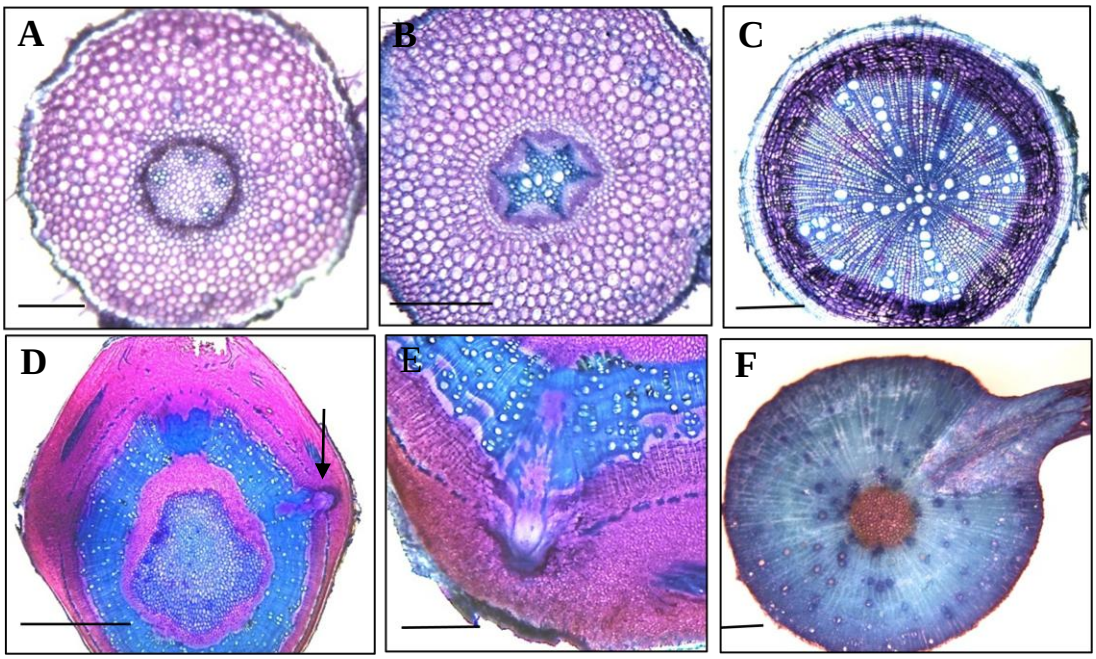

Figure 4. Transverse sections of developing nodal roots stained with toluidine blue. **A.** Section less than 1 cm from root tip showing distinct dark staining pericycle and five xylem poles. (bar = 0.25 mm) **B.** Section 2-3 cm from root tip showing development of metaxylem between the xylem poles. (bar = 0.5 mm) **C.** Root section 3-5 cm from tip. Significant secondary xylem has been produced within the central region of the root. (bar = 0.5 mm) **D&E**. Transverse sections at the nodal region of semi-woody stems. **D.** Development of a root primordia (arrowed) originating within the node. (bar = 5 mm) **E.** Detail of nodal root developing from within or deeper than the lignified xylem. Root has yet to penetrate the phelloderm tissues. (bar = 2.5 mm) **F.** Section showing nodal root that has emerged from the peeled stem, with connections deep into the stem through the secondary xylem to the central pith. (bar = 1 mm)

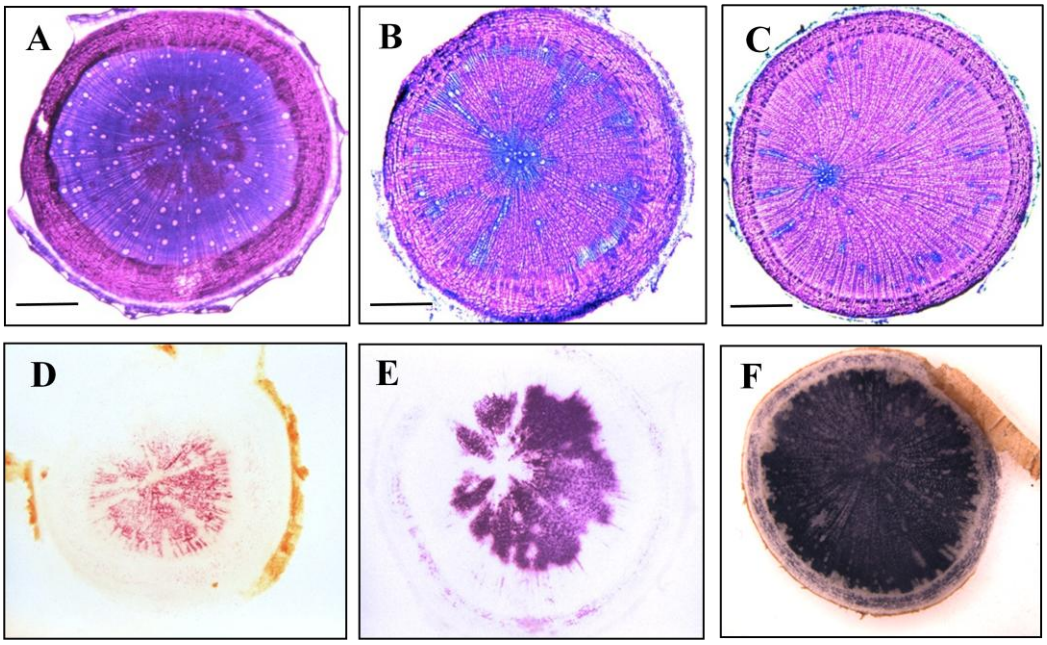

Figure 5. Transverse sections of different regions of the same storage roots stained with toluidine blue (A-C) and potassium iodide (D-F). **A.** Section through neck region of storage root showing predominant development of blue lignified tissues internal to the cambium. (bar = 1.5 mm) **B.** Section of storage root distal to the neck showing greater differentiation of xylem into pink staining cellulosic compared to blue staining lignified tissues. (bar = 2 mm) **C.** Section furthest from stem in which the vast majority of xylem cells are cellulosic and xylem present as scattered groups of vessels. (bar = 2.5 mm) **D-F.** Sections of same tissues shown in A-C but stained to show presence of starch. Dark staining starch is shown to reflect differentiation of cellulosic xylem parenchyma as the sections move distally away from the stem and neck regions.

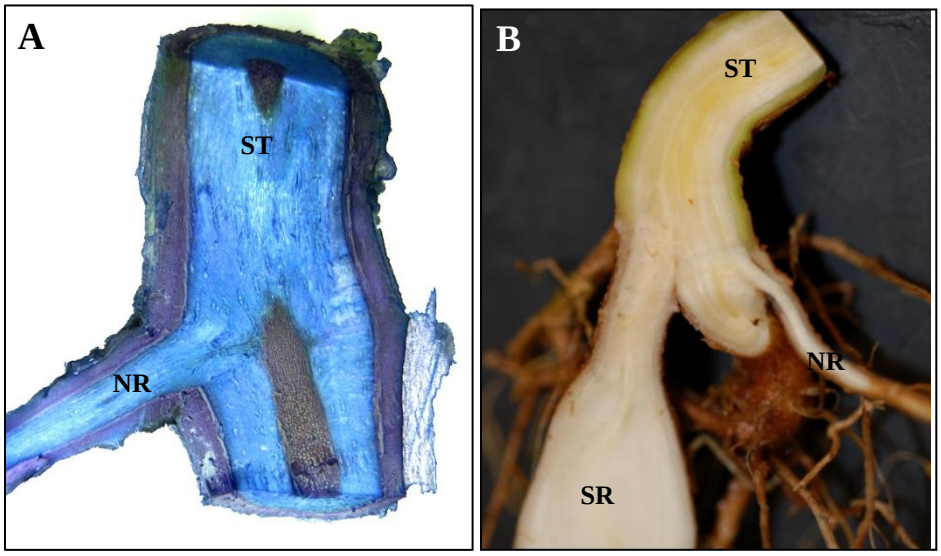

Figure 6. Stem (ST) and developing storage roots (SR) sliced longitudinally to the stem axis of tissue culture-derived plants. **A.** Semi-woody stem of cv. TME204 stained with toluidine blue showing continuous connection of lignified tissues from a young nodal root (NR) with the lignified stem tissues. **B.** A well-developed storage root (left) and nodal root (NR) yet to undergo substantial radial secondary thickening (right) of cv. 60444, shown with their vascular tissues continuously connected to the secondary vascular tissues of the stem.

Chapter 3. Microarray analysis of gene expression during storage root development in cassava (*Manihot esculenta* Crantz)

ABSTRACT

Cassava tuberization is a complex process that depends on genetics, environment and biological factors. Knowledge of the specific genes and pathways involved in storage root formation in cassava are needed in order to understand the underlying process and allow manipulation of storage root yields. cDNA microarray is a powerful tool to generate transcriptome profiles and was applied here to elucidate the molecular mechanisms of cassava root tuberization. A microarray consisting of 4,129 cassava and 19,808 leafy spurge cDNAs was used to generate transcriptome profiling across four developmental stages of storage root formation representing: prior to storage root formation, initiation of storage root formation, filling stage, and maturation stage. Gene Set Enrichment Analysis of up-regulated genes across the four developmental stages demonstrated up-regulation of the jasmonic acid pathway in the initiation stage. This might imply a role for jasmonic acid as a triggering factor in cassava storage root development and requires further investigation. K-means clustering and Heatmap analysis identified three candidate genes possibly involved in storage root development. *In silico* analysis through public databases and the Cassava Genome (Phytozome) indicated the three candidate genes to be homologues of cassava c1 (*Mec1*), cassava *ATDI21* and *ENOD40*-like genes. *Mec1* is related to secondary xylem parenchyma development in cassava storage roots, while cassava *ATDI21* encodes Lea3, a drought responsive gene, and *ENOD40*-like is a gene predicted to function as non-coding RNA.

INTRODUCTION

Tuberization is a critical mechanism in the life cycle of storage root crops to secure their survival under adverse environmental conditions. Several factors such as genetics, environment and unique physiology of each species affect their tuberization process. Due to the importance of tuber crops as sources of dietary energy, they fill the role of staple foods for half the world's population (Shewry, 2003; Howeler *et al*., 2013). The tuber crops are characterized by their different botanical origins. Tubers can be derived from diverse organs that are specific in each crop. Potato tubers are derived from underground stems, whereas taro is a corm. On the other hand, the storage organs of sweet potato and cassava are derived from root (Shewry, 2003). Cassava is distinct from the other tuber crops, however, in that the storage organ is not a propagule, but instead is a modified root structure with no additional known function. The tuber storage proteins have a major role in storage of nitrogen, sulfur and carbon for the purpose of aiding survival of the plant under harsh environments. These proteins have been identified in potato, sweet potato, taro, and yam as patatin, sporamin, tarin, and dioscorin, respectively. Uniquely, cassava has no known tuber storage protein (Shewry, 2003).

Cassava (*Manihot esculenta* Crantz) acts as a security crop for an estimated 700 million people in the tropical regions of Africa, Asia and South America (Howeler *et al*., 2013). Cassava is grown primarily for its storage roots that are used in fresh and processed forms for human food, livestock feed, and as a source of biofuel and industrial starch (IITA, 2009; Howeler *et al*., 2013). Although cassava is an excellent source of starch, full utilization of cassava storage roots is limited by inherent adverse traits such as

high cyanogenic content of many cultivars, rapid physiological deterioration after harvesting and low nutritional content (El-Sharkawy, 2003; Howeler *et al*., 2013; IITA, 2009).

Breeding programs have been addressing these limitations for decades but face numerous challenges. Cassava is highly heterozygous and suffers from sporadic flowering, poor seed set and low seed viability (Alves, 2002; El-Sharkawy, 2003). All these factors make modern transgenic technologies an attractive alternative for the improvement of otherwise high-performing and valued cassava varieties (Taylor *et al*., 2004, 2012; Jorgensen *et al*., 2005; Carvalho *et al*., 2011; Sayre *et al*., 2011; Zainuddin *et al*., 2012). Progress in this area includes the application of RNA interference technology (RNAi) to generate cyanide free cassava (Jorgensen *et al*., 2005). The BioCassava Plus program has also conducted biofortificaton of cassava to increase the iron and beta-carotene content in cassava storage roots (Sayre *et al*., 2011). The delay of postharvest physiological deterioration of cassava storage root was intensively studied by the manipulation of the reactive oxygen species production (Zidenga *et al.*, 2012; Xu *et al*., 2013). Furthermore, large-scale proteomic analysis in cassava storage root allowed identification of target genes to reduce postharvest deterioration (Owiti *et al*., 2011; Vanderschuren *et al*., 2014).

Despite recent progress many challenges remain, such as improvements to tuberous root harvest index, enhanced starch for industrial applications and the ethanol industry (Carvalho *et al.*, 2011; Howeler *et al*., 2013). All such improvement programs

would benefit from increased knowledge of the tuberization process in cassava. Cassava tuberization is a complicated processed. Cassava is usually propagated by stem cuttings. Two types of adventitious root are generated at the early stage of root development: basal roots and nodal roots (see Chapter 2; Lowe *et al*., 1982). According to the previous anatomical studies of this thesis (Chapter 2), storage roots develop from the nodal-derived roots. What genes and biochemical pathways are responsible for this process were investigated here.

Functional genomics is the study of gene products and how their interactions result in cellular and whole organism development, including phenotype and response to the environment (Butte, 2002). This approach enables the researcher to analyze genetic events on a genome-wide scale, suitable for using in gene discovery, marker assisted determination, trait classification and specific gene target identification (Chen *et al*., 2009). The approaches for functional genomics are performed in two ways: forward and reverse genetics. Forward genetics was initially performed using mutagenesis to generate randomly mutating population, resulting with genome-mapping to locate the target genes in the genome that correspond with the observed phenotype. Traditional forward genetics is a time-consuming method, due to the fact that large-scale screening of the saturated mutant phenotype is required. In contrast, the reverse genetic approach utilizes advanced sequencing technologies. Reverse genetics predicts the function of a gene based on the phenotype resulting from target gene alteration. Efficient techniques used in reverse genetics include gene silencing by RNAi technology, mutational analysis, genome-wide association studies, microarray analysis including RNAseq and proteomic analysis (Chen

et al., 2009; Till *et al*., 2013). Integrated functional genomics studies facilitate the identification of genetic significance, such as gene expression profiles and proteomic analysis, which can be used to predict the molecular mechanism of important traits. Complex biological systems can be elucidated using microarray technology, which facilitates identification of the expressed genes interacting with each other under given environmental conditions (Zhang *et al*., 2003). Transcriptional profiling from cDNA microarray analysis enables the prediction of putative pathways and candidate genes involved in target mechanisms (Duggan *et al*., 1999; Hardiman, 2004; Chen *et al*., 2009). However, cDNA microarray is a genome-wide analysis, thus the challenge for analysis is to interpret the large scale of the data produced. Experimental design and component analysis, including the minimum sample size, have been studied to improve efficiency of data analysis from microarrays (Butte, 2002; Churchill, 2002; Subramanian *et al*., 2005; Nguyen and Williams, 2006; Jørstad, *et al*., 2007).

A functional genomics approach was initially implemented for cassava research to develop expressed sequence tags (ESTs) from polymorphic transcript-derived fragments (TDFs) using the cDNA-AFLP (amplified fragment length polymorphism) technique (Suarez *et al*., 2000). Genetic diversity among domesticated varieties and wild population were examined to determine the evolutionary and geographical origin of cassava using the SNPs (single nucleotide polymorphisms) and SSRs (simple sequence repeat) (Olsen, 2004). Gene libraries constructed for cassava were developed and characterized using ESTs (Sakurai *et al.*, 2007, Lokko *et al.*, 2007). The draft genome sequence for cassava was announced in 2009 via JGI's Phytozome (www.phytozome.net/cassava), providing

new opportunities to improve cassava crop. Furthermore, a total of 9,600 cDNA and gene expression profiles at key growth stages of cassava were sequenced and established as another catalogue of expressed sequence tags (ESTs) (Li *et al*., 2010).

The cDNA coding for root protein Pt2L4 (the glutamic acid-rich protein), was reported to be differentially expressed in the starchy layer of cassava storage root (de Souza *et al*., 2002; 2006). Two cassava promoters related to vascular expression (cDNA of C15 and C54) have also been characterized (Zhang *et al*., 2003). Furthermore, a glutamic acid-rich protein promoter showed high expression in the storage root and stem when fused to a transgenic marker gene (Beltran *et al.,* 2010). However, these results still leave little understanding of what controls formation of storage roots in cassava.

In order to understand the pathway or subset of genes involved with each developmental stage of cassava storage root formation, genome analysis is required to provide information on the tuberization mechanism. Over the past decade, differential gene expression and transcriptome profiling provided sources of information about the gene regulated during tuberization in cassava. In addition, several cassava genome-wide expression analyses have been reported at different growth stages. Sojikul *et al*. (2010) compared gene expression profiles between fibrous roots and storage root in cassava using cDNA AFLP. The results indicated that sulfite reductase, calcium-dependent protein kinase, ent-kaurene synthase, and hexose transporter are involved in cassava storage root initiation because these genes showed specific expression in the storage root at an early stage. Leaf proteomic analysis was studied to elucidate the possible metabolic

switches in the leaf that may act as triggers for storage root formation (Mitprasat *et al*., 2011). Yang *et al*. (2011) studied genome-wide expression patterns during the tuberization stage using a 60 mer oligonucleotide microarray and showed dynamic changes in the active processes of glycolysis and gluconeogenesis during storage root formation. To date, although new technologies have been applied to study genome-wide expression and combined with improved analytical tools for interpretation of large-scale data, knowledge about cassava tuberization still remains unclear.

The present work attempts to elucidate the cassava tuberization mechanism using transcriptome analysis. A cDNA microarray containing 4,129 cassava-specific cDNA and the 19,808 leafy spurge unigenes (USDA-Bioscience research lab, Fargo, ND) was used to generate the transcriptome profiling across root developmental stages. Differential gene expression revealed the order and subset of genes and pathways involved at each specific developmental stage. The genes that were highly up-regulated in the initiation stage, starch filling stage and maturation stage were annotated of their putative function in cassava tuberization.

MATERIALS AND METHODS

Plant material

Cassava cultivar 60444 was micropropagated in tissue culture as described by Taylor *et al*. (2012). The media for cassava micropropagation consisted of Murashige and Skoog (1962) (MS) basal salt mixture and MS vitamin powder (Sigma-Aldrich, MO), and 20 g/l sucrose (MS2) solidified with 8 g/l of Difco Agar Noble. The pH of the media was

adjusted to 6.14 followed by addition of Difco Agar Noble and autoclaving at 121^{o}C and 15-20 psi for 20 minutes. The media was dispensed into 100 x 25 mm Petri dishes at 40 ml per dish and stored at room temperature until use. Apical cuttings carrying two to three nodes were excised from in vitro mother plants cultured on MS2 medium every four weeks by subculturing onto fresh MS2 culture media and incubating at 28^{o}C under a 16/8 hours photoperiod at 75 µMs/m^{2}.

To establish plants derived from tissue culture in the greenhouse, four-week-old *in vitro* plantlets were removed from the agar medium and transferred into the three-inch pots containing Fafard 51 potting media (Sun Gro Horticulture Canada Ltd.). The potting soil was soaked with 1.8 gram per 3.78 L. Gnatrol (Valent BioSciences Cooperation, IL), 14.7 gram per 3.78 L. of 15-5-15 fertilizer (Jack's Professional LX, JR Peters Inc, PA) and 5 ml per 3.78 L of micronutrient (MOST, JR Peters Inc, PA). Potted plants were placed on a mist bench for one week with bottom heat supplied before moving to the open bench and grown under natural and supplemental lighting at 900 µM/m^{2} for 16 hours per day, at 26-28^{o}C and 60% humidity.

Twenty *in vitro* plants were transferred out to greenhouse every 30 days for three months to be assured of the sufficient experimental material. Plants were removed from the pots without damaging the root system every two weeks through three months in order to collect root samples.

Sample collection for cDNA microarray analysis

Storage root vs. leaf microarray analysis

Cassava leaves were collected using gloved hands from three-month-old plants growing in the greenhouse, wrapped in aluminum foil and labeled. Samples were immediately frozen in liquid nitrogen until proceeding for RNA extraction. Storage roots from the same plants were removed and cleaned by rinsing under running tap water. The peel was removed by hand and the storage tissues cut into pieces approximately 125 mm^3 in size and placed in a clean 50 ml Falcon tube. Storage root tissues were then submerged in liquid nitrogen and freeze-dried using a lyophilizer (VirTis lyophilizer #FM 25ES-53, SP scientific) for 24 hours. Samples were stored at room temperature until RNA extraction. Three biological replications were represented by sampling from three individual plants and two technical replications were performed using the roller cycle dye swapping technique.

Microarray analysis of four root developmental stages

Storage root development from *in vitro* derived plants was examined at two weeks, one month, two months and three months after transfer to soil. This equated to the four developmental stages: 1. before tuberization (SR1); 2. storage root initiation (SR2); 3. starch filling (SR3); and 4. storage root maturation (SR4). Four biological replications and two technical replications were performed at each developmental stage. At SR1 and SR2 stages, five grams of fibrous root of two-week-old plants and one-month-old plants were collected for RNA extraction respectively. Collected root samples were gently rinsed in tap water to remove dirt, wrapped in aluminum foil and immediately kept in

liquid nitrogen and labeled. At SR3 and SR4, the storage roots were removed from two-month-old and three-month-old plants, taking care not to include any fibrous root structures. Storage roots were washed under running tap water, peeled and the inner storage tissues processed as described above. The processed storage root samples were immediately frozen in liquid Nitrogen and freeze-dried in a lyophilizer (VirTis lyophilizer #FM 25ES-53, SP Scientific) for 24 hours and stored at room temperature till required for RNA extraction.

RNA extraction

Leaf and fibrous root tissues

Samples from leaf and non-storage roots were homogenized in liquid N_2 in a DEPC-treated pestle to produce a powder and extracted using the modified CTAB protocol based on Lodhi *et al.* (1994). The CTAB buffer was modified to compose 2% CTAB (Sigma), 100 mM Tris-HCL (pH 8.0, Sigma), 20 mM EDTA (Sigma), 1.4 M Sodium Chloride (NaCl, Sigma), and 2% v/v beta-mercaptoethanol (Sigma) in RNase-free water. The 2% beta-mercaptoethanol was added into buffer just prior to use and then incubated at 65^{o}C before adding 0.5 g of ground tissue to the 15 ml centrifuge tube. Samples were incubated at 65^{o}C for 30 min with gentle mixing following by incubation at room temperature for 10 min. An equal amount of chloroform was added and mixed by inverting for 10 min at room temperature before centrifuged at 10000 rpm. After centrifugation, the aqueous phase containing nucleic acids was dispensed into a new centrifuge tube and the RNA purification performed using an equal amount of chloroform:isoamyl alcohol (24:1) and performed twice. The supernatant was carefully

transferred to the new 15 ml tube after centrifugation and precipitated by adding 0.7 volume of ice cold isopropanol. The RNA pellet was then washed with 500 µl 75% ethanol three times before dissolving in one ml RNase-free water. An additional overnight precipitation was then performed by adding 1/3 volume of 8 M LiCl (Ambion) and incubated at -20^{o}C. The final RNA pellet was washed with 75% ethanol three times. RNA quality and quantity was confirmed by denaturing agarose gel electrophoresis and spectrophotometer, respectively.

Storage root tissues

Total RNA from storage roots was extracted using a protocol modified for cassava storage roots based on Li and Trick (2005). The RNA extraction buffer consisted of 100 mM LiCl, 1% w/v SDS, 100 mM Tris-HCL pH 7.5, 100 mM EDTA, and 1% v/v beta-mercaptoethanol in diethylpyrocarbonate (DEPC) treated water, the latter being added into the buffer before extraction. The lyophilized storage root sample was briefly ground in liquid N_2 to produce a fine powder. Half a gram of ground lyophilized tissue was transferred into a 2 ml Eppendorf tube. First, 800 µl extraction buffer was added into the sample, followed immediately by the premix of 800 µl acidic phenol:chloroform (AM9720, Ambion). The extract was vortexed, incubated at room temperature for one hour and then centrifuged at maximum speed in a microcentrifuge for 30 min at 4^{o}C. The clear aqueous phase of the extract was carefully transferred into a new tube and centrifuged again for 5 min. The supernatant was carefully transferred into a new tube, 1/3 volume of 8M LiCl added and gently mixed followed by incubation at -20^{o}C for 16 hours. The extract was then centrifuged at 4^{o}C for 30 min to harvest the RNA pellet. The

RNA pellet was washed by ice cold 75% ethanol twice, then left to dry at room temperature before elution with 30 µl RNase-free water. The RNA was treated with DNase using DNA removal kit (DNA free, AM1906, Ambion), followed by cleanup using a spin column as described in RNA cleanup step manual (RNeasy mini kit,Qiagen). RNA quality and quantity was confirmed by denaturing agarose gel electrophoresis and spectrophotometer reading, respectively.

cDNA labeling and microarray analysis

The first strand cDNA synthesis of genomic RNA was performed using SuperScript indirect cDNA labeling system (L1014-02, Invitrogen). First, 20 µg of total RNA was reverse transcribed using SuperScript III reverse transcriptase with the anchored oligo(dT)20 as a primer and incorporating amino-modified dUTP into the synthesized cDNA. The cDNA synthesis mix was incubated at 46^{o}C for two hours. Then the template RNA was degraded by addition of 1M NaOH at 65^{o}C for 15 min and neutralized with 1M HCL and TRIS (1M pH 7), follow by clean up of the unincorporated nucleotides, primer and buffers using the PCR purification kit (Invitrogen, Carlsbad, CA). In the second step, the modified cDNA was coupled with Alexa Flour 555 and 647 (Invitrogen) according to manufacturer's protocol. The labeled cDNA was purified using the Purelink PCR purification kit (Invitrogen) to remove all unincorporated dye to be ready for hybridization to microarrays.

Labeled cDNA was hybridized to a custom made 23K element cDNA microarray containing 4,129 unigenes from a cassava EST database (Lokko *et al*., 2007) and 19,808

unigenes from leafy spurge EST database (Anderson *et al*., 2004). The signal intensities of 16 hybridized microarrays were read using a GenePix 4000X scanner (Axon Instruments/Molecular Devices Corp., Sunnyvale, CA) and GenePix Pro software. The intensities of each array were log2 transformed, then centered and normalized against each other.

Data Analysis

For the microarray analysis to study differential expression between leaf vs. storage root, gene expression profiles were statistically analyzed using GeneMath XT software (Applied Maths, NV) for normalization of expression values against each other, including principle analysis to predict the functional biological process based on GO term. For microarray expression analysis across the four different developmental stages of storage root formation, statistical analysis, including normalization of differentially expressed gene data sets, was performed using CLC Main Workbench software (CLCbio, a Qiagen company, MA). K-means clustering was used to partition the data set into distinct clusters with similar expression pattern across samples (Wagstaff *et al*., 2001).

To predict pathway networks involved in tuberization, Pathway Studio software (http://www.ariadnegenomics.com) was used for Gene Set Enrichment Analysis (GSEA) as described in Subramanian *et al.* (2005). The expression of genes significantly over-represented ($p < 0.05$) by up- and down-regulation in biological processes, molecular functions and cellular components were also identified using GSEA.

Candidate gene identification

Gene clusters resulting from differential expression analysis that were shown to be highly up-regulated in SR1, SR2 & SR3 and SR4 were selected to be the study groups. Non-redundant sequences including the cassava-specific genes were blasted into the TAIR database (http://www.arabidopsis.org/Blast/index.jsp) to identify the Arabidopsis orthologue and predict their function using BLASTN and BLASTP via NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Sequences of selected genes were retrieved from the EST database (NCBI) correlating to their unique accession number and then blasted into Cassava genome (www.Phytozome.net, JGI, CA) for analysis against the actual cassava sequence database.

RESULTS

Comparative study of gene expression profiles in storage root and leaf tissues

A microarray experiment was performed in order to validate efficacy of the cDNA microarrays and generate initial data for determining differential gene expression between leaf and storage root tissues. Young leaves and tuberous root tissues were sampled from three-month-old greenhouse grown plants.

An improved method for extraction of total RNA from cassava storage roots was developed to facilitate this study. When RNA was extracted from frozen tissues following the protocol of Li and Trick (2005), the resulting RNA was degraded and not suitable for cDNA synthesis. Addition of a lyophilizing step in sample preparation improved yields of intact RNA but purity remained problematic for labeling with

fluorescent dye. The purification steps were also modified by adding a chloroform cleaning step and increasing the time of LiCl precipitation from two hours to at least 12 hours. Use of these three modifications resulted in high yields of intact RNA suitable for production of cDNA and dye labeling (results not shown) and were used as adopted as the standard protocol for processing RNA from storage root and stem tissues.

Total RNA was extracted with the improved method, labeled and hybridized to a microchip consisting of 4,129 cassava and 19,808 leafy spurge cDNA. Data was log2 normalized and statistically analyzed using GeneMath software to display genes significantly (p-value <0.05) up- or down-regulated in leaves and storage roots. Using these criteria, 336 genes were seen to be up-regulated in leaf compared to storage root tissues, and 108 genes were up-regulated in storage root compared to the leaf. Putative biological function of up-regulated genes was predicted using Gene Set Enrichment Analysis based on gene ontology (GO). The majority of up-regulated genes fell within uncategorized biological function for both leaf and storage root profiles (Fig. 1A and B). Among 336 up-regulated genes in the leaf, approximately 67% of genes were undetermined, while 23% had biological process unknown. Data for the 108 up-regulated genes in storage root was similar to that of the leaf, with 67% undetermined and 22% of biological process unknown.

Figure 2 shows the number of up-regulated genes with known biological function identified from the leaf (35) and storage root (12). Twenty-six genes with known biological function were seen to be up-regulated in the leaf only and not the storage root.

Only three genes of known biological function were found up-regulated in roots tissues and not in the leaf. In the leaf, seven up-regulated genes were involved in photosynthesis, three in biosynthesis, five in signal transduction, three in defense response and one each in carbon utilization, cell communication, DNA methylation, growth, aging, secondary metabolism, sugar mediated signaling and unidimensional cell growth. Specific biological pathways up-regulated in storage roots and not in the leaf, including jasmonic acid mediated signaling pathway (1 gene), circadian rhythm (1 gene), and embryonic development (1 gene). The three biological functions – metabolism, transport and response to jasmonic acid stimulus – were found to have up-regulated genes from both leaf and storage roots. Metabolism was highly expressed and represented by six genes in leaf and two genes in the storage root. Likewise, seven genes belonging to the transport processes were represented in the leaf (2 genes) and storage root (5 genes). Genes involved in the response to jasmonic acid stimulus were represented in leaf and storage root as one gene and two genes, respectively. It should be noted that up-regulated genes involved in response to jasmonic acid stimulus were seen in both leaf and storage root tissues; genes involved in jasmonic acid mediated signaling pathway were presented only in storage root. This result may therefore indicate a likely role for jasmonic acid in storage root production and maturation and requires further study.

Differentially up-regulated genes unique to the storage root or leaf

Data was further analyzed in order to determine which genes were differentially up-regulated in the storage root and leaf. To achieve this, the root/leaf and leaf/root ratios was calculated from log2 normalized expression values and used to rank genes for

differential expression between the storage root and leaf (Table 1 and 2). BLAST searches were performed for the ten unigenes with highest differential expression in root and leaf tissues in order to identify their similarity to genes within the NCBI database. Because the microchip used in this study contained cDNAs from two organisms, leafy spurge and cassava, differentially up-regulated genes are shown as cassava-specific or spurge-specific.

Differentially up-regulated genes unique to the storage root

The ten genes with the highest differential root/leaf expression ratio were studied using BLAST searches against non-redundant database on NCBI. The up-regulated unigene DV445495 that ranked highest for expression ratio in storage root was allergenic-related protein Pt2L4. Pt2L4 is unique to *Manihot esculenta* and is a glutamic acid-rich protein previously reported to be related to secondary growth and storage root formation in cassava (de Souza *et al*., 2004; 2006). DV446014 was also seen to be highly differentially expressed in roots compared to leaves. The cDNA sequence of DV446014 was specific to cassava and not similar to any other organism. BLAST searching revealed no significant similarity to other nucleotide or protein sequences in any organism. Lea5 (Late embryogenic abundant) protein (DV446014), fiber protein Fb37 (*Gossypium barbadense*) (DV451141) and DnaJ protein (*Hevea brasiliensis*) (DV139532) were also shown to be highly differentially up-regulated in the storage root. Although, DV451141 had similarity to unknown hypothetical protein p85RF, its function in cassava is not known.

Differentially up-regulated genes unique to the leaf

The top ten differentially up-regulated genes found in the leaf from this array analysis are shown in Table 2. Predictably, five of the unigenes with highest leaf/root differential ratio showed high similarity with genes involved in photosynthesis. Highest was seen for DV441259, which is very similar to light harvesting chlorophyll a/b-binding protein in *Nicotiana sylvestris*. DV454456 was ranked ninth and is unique to *Manihot esculenta*, predicted to encode chloroplast latex aldolase-like protein.

Comparative study of gene expression profiles in different stages of storage root formation and development

In order to study differential gene expression during storage root initiation and development, cassava plants of cultivar 60444 were established in the greenhouse. Root tissues were harvested at four different time points representing known stages in development of the storage root system in this cultivar under these conditions. Root tissues were harvested two weeks after planting to represent the stage before storage root formation (SR1), at one month to represent the storage root initiation stage (SR2), at two months for radial expansion and starch filling stage (SR3) and at three months for storage root maturation stage (SR4) (Fig. 3).

Total RNA was extracted from four biological replicates for each developmental stage. Samples were processed, cDNA produced, labeled and hybridized to 16 microchips carrying 4,129 cassava and 19,808 leafy spurge cDNA respectively as described above. Data was log2 normalized and statistically analyzed using GeneMath software to display

genes significantly (p-value <0.05) up- or down-regulated at the different SR1-SR4 developmental stages.

Pathways significantly up- and down-regulated during storage root development

To identify significantly over-represented ($p<0.05$) pathways among the four developmental stages of storage root development, Gene Set Enrichment Analysis (GSEA) was employed to identify up- and down-regulated biological processes, molecular functions, and cellular components using gene ontology (GO) classification. Based on GO, the role of target gene sets were predicted through their biological function to provide putative pathways involved in the different stages of storage root development. The enrichment of each pathway was generated by the number of up-regulated genes in the pathway, and differential expression levels were determined by pairwise comparison between stages of storage root development, i.e. – SR1 vs SR2, SR2 vs SR3 and SR3 vs SR4. Table 3 shows pathways significantly over-represented ($p<0.05$) when comparing these developmental stages against each other.

The pathways that were highly up-regulated in SR2 were identified using comparative analysis between the pre-tuberization stage (SR1) versus the initiation stage (SR2). Seven significant pathways were found to be overrepresented in SR2 compared to SR1 (Table 3). These included jasmonic acid biosynthesis, different pathways of sucrose degradation, and systemin signaling. Up-regulation of jasmonic acid biosynthesis was seen to be unique to the SR2 developmental stage. Five of the same pathways involved in sucrose degradation, plus systemin signaling, were also found to be up-regulated in SR3.

An additional eight pathways were up-regulated in SR3 compared to SR2. These represented glycolysis I (plant cytosol), valine, isoleucine, suberin, flavonoid biosynthesis, and the superpathway of isoleucine and valine biosynthesis plus the superpathway of leucine, valine and isoleucine biosynthesis. Up-regulation of leucine, valine, and isoleucine biosynthesis superpathway was also evident in SR4, but this was the only up-regulated pathway common between SR3 and SR4. Ten additional up-regulated pathways were unique to SR4. These included auxin signaling, fatty acid oxidation, and the glyoxalate cycle. Surprisingly, for these achlorotic tissues the photosynthesis pathway was also represented here.

Transcriptome profiles among four developmental stages or root development

Microarray analysis was conducted across the four developmental stages (SR1-SR4) of storage root tuberization to determine transcriptomic changes at each developmental stage. The change of transcriptome profiles across the four developmental stages allowed determination of the subset of the genes involved in storage root formation. A total of 22,503 unigenes were analyzed using CLC Main Workbench program. 5,142 were significantly expressed at $p<0.005$ for normalized values and 1,239 at $p<0.005$ for maximum change greater than two fold. Analysis was performed using K-means cluster to group the 22,503 genes identified by CLC Workbench into different clusters that showed similar expression patterns across the four developmental stages. Analysis generating 12 different clusters did not provide sufficient resolution of the data to show differential expression. Increasing this to 16 clusters was also not successful. Therefore, analysis at 24 clusters was employed. Figure 4 shows a pie chart illustrating proportions

of genes up-regulated in the different SR stages. The top twenty unigenes showing highest fold changes were ranked for their normalized expression values across the four SR developmental stages (Fig. 5).

Because the main objective of this study was to uncover important biological processes involved in the developmental stages from fibrous roots to storage root initiation, filling and maturation, clusters showing significant changes in expression across, and between these stages were selected for further study. The three clusters – 15, 20 and 22 (Figs. 6A-C) – were chosen for further study because of their highly significant changes in normalized expression values (above 14,000) across the SR1 to SR4 stages. This was considered to increase the likelihood of their critical role in the stages of root tuberization. In addition, the number of genes within these clusters was reasonable to handle for subsequent analysis and study.

Cluster 20 (Fig. 6A) consisted of fifteen unigenes of which four pass significant filter at $p<0.005$ for expression. These four genes had significantly up-regulated expression at SR2 compared to the other developmental stages. Expression of unigenes in this group had low relative values in SR3 and SR4. Genes in this group could therefore be considered as candidates for regulators of early storage root formation. BLAST searches for these four leafy spurge unigenes obtained hits for putative homologues, including putative amino transferase from rice, major latex allergen Hev b 4 from rubber, cyclin A3 from tomato, and cysteine synthase from spinach (Table 4A).

Nine genes comprised cluster 15, of which eight had significant expression ($p<0.005$). Normalized expression levels were high compared to cluster 20 with up-regulation seen at SR3 and SR4 compared to earlier developmental stages (Fig. 6B). In contrast to cluster 20, in cluster 15, six out of the eight highly expressed genes were cassava unigenes. The unigene with highest expression levels in this cluster, and across all cira 24,000 unigenes investigated, was DV445495. DV445495 expressed most highly at SR4 and relates to the cassava-specific allergenic-related protein Pt2L4. This result corresponds with previous identification of this protein in the secondary xylem parenchyma of mature storage roots of cassava (de Souza *et al*. 2004; 2006). The second ranked unigene DV451479 in cluster 20 related to *LEA5*,; late embryogenic abundant protein in cassava. DV446014 had a greater than 10-fold change in expression during SR4 and is cassava-specific but of unknown function (Table 4B).

Cluster 22 consisted of seventeen unigenes (Fig. 6C). All seventeen had significant expression and were similar to those of cluster 15 being highly up-regulated expression at SR3 and SR4 compared to SR2. Six of the unigenes were specific to cassava. Eleven of the overrepresented unigenes were from leafy spurge and nine had no predicted functional annotation in cassava (Table 4C).

Gene expression involved in storage root development

In order to confirm the expression data described above, and to more fully analyze differential expression patterns across the four developmental stages, a Heatmap was created. CLC Main Workbench was used to produce a Heatmap of the 29 unigenes across

SR1 to SR4 using their normalized expression values. Hierarchical clustering was calculated from the average expression value in each developmental stage to illustrate their relationship to the root tuberization process (Fig. 7).

Using the Heatmap analysis, the 29 unigenes were separated into three distinct groups. In the first group expression was seen to be down-regulated in SR2 compared to SR1, SR3 and SR4. In the second group, the pattern was reversed with relative up-regulation of expression in SR2 compared to the other developmental stages. Lastly, a group was seen with low expression in SR1 and SR2 but high expression in SR3 and SR4. Within the second group, which showed high expression in SR2, DV127294 (major latex allergen Hev b 4) was distinct from all other unigenes showing elevated expression signal at this stage only. As SR2 is the developmental stage associated with transition from fibrous to storage root, this result indicates that the homologue gene for Hev b 4 in cassava is a possible candidate for involvement in the initiation of root tuberization.

It is notable that DV445495 (allergenic-related protein Pt2L4), which was the most highly differentially expressed gene as determined by fold change, was placed in the third group and showed increasing expression through SR3 and into SR4. DV451479 (late embryogenesis abundant protein LEA), although also placed in the third group, was less distinct by its up-regulation in the latter root developmental stages. DV446014 (cassava-specific 746, unknown function) showed the most distinct up-regulation in SR4 compared to earlier developmental stages of all 29 unigenes studies with the Heatmap method.

DISCUSSION

This study performed cDNA microarray analysis to determine the putative candidate genes involved in cassava root tuberization. The quality and quantity of RNA used in microarray is a first critical factor to achieve success. High starch content in cassava storage root tissue can cause difficulties in RNA extraction (Kumar *et al*., 2007, Xu *et al*., 2010**)**. For instance, the extended time required for grinding the fresh storage root tissue in liquid nitrogen at -80^{o}C leads to degradation of nucleic acids and decreases the yield of intact RNA. Furthermore, the conformational change of starch to form a gel in the high salts of the extraction buffer reduces the purity of intact RNA and obstructs downstream reactions. Lyophilization of starchy tissues, such as the cassava storage root, is an improved method that enhances the stability of nucleic acids and shortens the grinding time needed in liquid nitrogen before RNA extraction. LiCl and acidic phenol was used in the present protocol to eliminate the amount of DNA. Additional of purification steps and increasing the time for RNA precipitation under -20^{o}C resulted in a significant increasing in RNA quality and quantity. This improved protocol was not only used for cassava storage root RNA extraction, but can also be effectively used for RNA extraction from woody stem or lignified tissues (see Chapter 4).

Validation of the microarray methods and analysis was also important for the planned experiments. For example, the microarray employed in this study was composed a mixture of cassava and leafy spurge cDNA (Lokko *et al*., 2007). In order to assess efficacy of this array, differential gene expression profiling was performed to compare tissues derived from the storage root and leaves. Thus some up-regulated genes might be

shown to be specific to leafy spurge cDNA even when cassava cDNA was used as a probe for hybridization to the microarray chip. Data generated using this microarray produced predictable and informative results. As expected and logical, GSEA predicted that the majority of up-regulated pathways and genes in the leaf were involved in photosynthesis, including signal transduction and biosynthesis. In addition, the highest up-regulated cassavas-specific gene identified was *Mec1* that encodes allergenic-related protein Pt2L4. This result correlates with the findings of de Souza *et al*. (2004), who identified this gene to be involved with development of secondary xylem parenchyma in storage roots. Additionally, the gene encoding late embryogenesis abundant proteins (Lea) was shown to be highly expressed in storage root tissue in the present study, a result that also correlates with previous reports (de Souza *et al*., 2004; 2006). Information generated from this initial study comparing gene expression in storage root and leaves, therefore generated confidence in the cassava/leafy surge cDNA microarray and its utilization to study the stages to storage root development in greater detail.

An understanding of the processes controlling storage root formation in cassava is important for improvement programs aiming to improve the crop through enhanced dry matter content, starch quality, nutritional content and improved postharvest characteristics (Sayre *et al*., 2011; Zainuddin *et al*., 2012; Vanderschuren *et al*., 2014). However, the biological mechanisms involved in this process remain unclear. This present study utilizes the cDNA microarray analysis to investigate the transcriptional changes occurring across four developmental stages of storage root formation. These included “prior to tuberization” (SR1), initiation of tuberization (SR2), filling (SR3), and

maturation (SR4) (Fig. 3). GSEA was used to predict up-regulated pathways due to differential gene expression across the four developmental stages. Pathways associated with sucrose degradation including glycolysis were found to be up-regulated at the initiation of storage root formation stage (SR2) (Table 3). This result reveals the active process of glycolysis and gluconeogenesis in early cassava storage root development and corresponds with the report of Yang *et al.* (2011), who employed cDNA microarray analysis to specifically study starch biosynthesis in storage roots.

Interestingly, the jasmonic acid biosynthesis pathway was also found to be up-regulated in SR2 (Table 3). In the initial microarray experiment reported here, a gene involved in jasmonic acid mediated signaling pathway was shown to be highly up-regulated in storage roots compared to leaves (Fig. 1). Numerous reports have shown jasmonic acid to be involved in potato tuberization, where it interacts with gibberellic acid to promote tuber formation (Palacho and Castel, 1991; Takahashi *et al*., 1994; Castro *et al*., 2000**)**. In addition, it has been shown to have possible roles in tuberization in yams and sweet potato (Koda, 1997). The systemin signaling pathway, which is known to be involved in jasmonic acid biosynthesis (Sun *et al*. 2011), was up-regulated in SR2 and SR3 (Table 3). As SR2-SR3 represents the stages during which early storage roots can first be seen developing, it is possible that jasmonic acid, which is known to be an important signaling molecule, also plays a role in storage root formation and early development in cassava. Future experiments in which jasmonic acid would be exogenously applied to developing cassava root systems should be performed and could provide additional information about the role of this compound in root tuberization.

The superpathway of leucine, valine, and isoleucine biosynthesis, which is involved in glucosinolate production, was shown to be up-regulated in SR3 and SR4 (Table 3). This result might relate to the synthesis of cyanogenic compounds for which these amino acids are involved. Cassava storage roots are known to accumulate cyanogenic glycosides (Jorgensen *et al*., 2005, Binder. 2010**),** so up-regulation of this related pathways would be logical at these developmental stages in which the storage tissues have been laid down and are undergoing growth and filling. Prediction of up-regulated photosynthesis pathway in storage root stage SR4 is an unexpected result for these studies, but could indicate that this later stage of storage root development might require a subset of genes involved in photosynthesis, such as phytochrome interacting factor (*PIF*) to induce, or suppress, plant hormone biosynthesis (Alabadi and Blazquez, 2008).

Microarray data analysis poses challenges due to the large data sets that are generated. These must be analyzed using effective software tools in order to highlight the biologically important results. CLC Main Workbench was used to statistically analyze and generate predicted fold change for differential gene expression across the four root developmental stages. Normally, significance at p-value 0.05 is used to determine the cut off level for data interpretation. However, if the data set is still too large, significance at p-value 0.005 is used. This was the case for the present studies. K-means clustering was used to separate the up-regulated genes into distinct groups based on the expression value of each gene across the four developmental stages. Of the 24 clusters produced, 21 clusters were not considered further because the numbers of genes presented were too

large to handle and/or did not show sufficiently large fold changes of gene expression to be of interest. Three distinct clusters were selected for subsequent detailed analysis because the number of genes (29) was reasonable to study (Figs. 6A-C) and fold changes were large.

Twenty-nine genes represented from three clusters were analyzed by production of a Heatmap in CLC Main Workbench. By statistical analysis, highly expressed genes are usually shown as fold changes, but in Heatmaps the expression value is shown as the individual value compared across the other three developmental stages. The biological Heatmap demonstrated the actual expression of up-regulated genes across the four developmental stages (Fig. 7). Visualization in this manner allowed a second interpretation of the data. By this method *Mec1* (DV445495) was seen to be highly up-regulated in SR3, and especially SR4, compared to the earlier developmental stages. Likewise, DV4511479 (*ADT21*; Lea3) was also seen to be up-regulated in SR3, and especially SR4. In both cases this confirms data from microarray analysis of storage root vs leaf tissues (Table 1). The Heatmap and associated phylogenetic tree (based on expression value) reveals three groups of differential expression patterns for SR1-SR4. One group shows relative down-regulation in SR2, the second is up-regulated in SR2 and therefore might be associated with initiation of the tuberization process. The third appears as up-regulation in SR3 and SR4. Of interest is DV127294, which showed the most distinct change of expression across the four developmental stages by the Heatmap. This gene is predicted to be a homologue of major latex allergen protein (Hev b 4). As in cluster 20 (Fig. 6A), the Heatmap showed this gene to be highly up-regulated in SR2, and

relatively down-regulated in the developmental stages before and after this stage. This data contrasts with the statistical information, which showed fold changes for this gene to be relatively low at 1.4, compared to *Mec1* (DV445495) at (16.5) (Table 4A) and demonstrates the value of Heatmap analysis.

It should be noted that the microarray data described here was performed before full information from the anatomical studies in Chapter 2 was available. Thus separate collection of basal-derived and nodal-derived roots was not performed. Instead, all roots at SR1 and SR2 that visibly appeared fibrous in type were collected as one. As a result, tissues at these stages, and especially at SR2, are most likely composed of a mixture of basal- and nodal-derived root structures. This may therefore complicate identification of changes taking place at the RNA level in SR2, when the nodal-derived roots initiate secondary thickening and start to develop as storage organs. However, for SR3 and SR4 only distinct storage organs were collected and no fibrous root material was included, which allows clear comparison of these two types of root organ with previous developmental stages.

Due to this lack of understanding of the biology of early storage root initiation and development, the experimental design of this microarray study was likely not optimal for discovery of the genes and mechanism involved at the initiation stage (SR2) of storage root formation. However, jasmonic acid and major latex allergen Hev b 4 have been identified as having putative roles at, or close to, the SR2 storage root initiation stage. Data generated by this microarray study clearly indicate that during SR3 and SR4,

Mec 1 and *ATDI21* genes play important roles in these stages in which secondary thickening and starch accumulation are predominant. Functional analysis of these two genes is further investigated in Chapter 4.

REFERENCES

Alabadi, D. and Blazquez, MA. (2008). Integration of light and hormone signals. *Plant Signaling & Behavior* 3(7):448-449.

Alves, AAC. (2002). Chapter 5: Cassava botany and physiology. Cassava: Biology, Production and Utilization. Eds R.J. Hilllocks, J.M. Thresh and A.C. Bellotti. CAB International 2002.

Anderson, JV., Delseny, M., Fregene, MA., Jorge, V., Mba, C., Lopez, C., Silvia Restrepo, S., Soto, M., Piegu, B., Verdier, V., Cooke, R., Tohme, J. and Horvath, DP. (2004). An EST resource for cassava and other species of Euphorbiaceae. *Plant Molecular Biology* 56: 527–539.

Beltran, J., Prias, M., Al-Babili, S., Ladino, Y., Lopez, D., Beyer, P., Chavarriaga, P. and Tohme, J. (2010). Expression pattern conferred by a glutamic acid-rich protein gene promoter in field-grown transgenic cassava (*Manihot esculenta* Crantz). *Planta*, 231, 1413-1424.

Binder, S. (2010). Branched-Chain Amino Acid Metabolism in *Arabidopsis thaliana*. The Arabidopsis Book. e0137. 10.1199/tab.0137.

Butte, A. (2002). The use and analysis of microarray data. *Nature review*. 1:954-960.

Carvalho, LJCB., Vieira, EA., Fialho, JDR., and de Souza CRB. (2011). A genomic assisted breeding program for cassava to improve nutritional quality and industrial trails of storage root. *Crop breeding and applied biotechnology* 11:289-296.

Castro, G., Abdala, G., Aguero, C. and Tizio, R. (2000). Interaction between jasmonic and gibberellic acids on *in vitro* tuberization of potato plantlets. *Potato Research*. 43:83-88.

Chen, X., Jorgenson, E. and Cheung, ST. (2009). New tools for functional genomic analysis. *Drug Discov Today*. 14(15-16): 754-760.

Churchill, GA. (2002). Fundamentals of experimental design for cDNA microarrays. *Nature Genetics*, 32, 490-495.

De Souza, CRB., Carvalho, LJCB., de Almeida, ERP. and Gander, ES. (2002). Identification of cassava root protein genes. *Plant Foods for Human Nutrition*, 353-363.

De Souza, CRB Carvalho, LJCB. and Cascardo, JCDM. (2004). Comparative gene expression study to identify genes possibly related to storage root formation in cassava. *Protein and Peptide Letters*. 11(6):577-582.

De Souza, CRB., Carvalho, LJCB., de Almeida, ERP. and Gander, ES. (2006). A cDNA sequence coding for a glutamic acid-rich protein is differentially expressed in cassava storage roots. *Protein and Peptide Letters*, 13, 653-657.

Duggan, D., Bittner, M., Chen, Y., Meltzer, P. and Trent, JM. (1999). Expression profiling using cDNA microarrays. *Nature genetics supplement*. 21:10-14.

El-Sharkawy, MA. (2003). Cassava biology and physiology. *Plant Molecular Biology*, 53, 621- 641.

Hardiman, G. (2004). Microarray platforms – comparisons and contrasts. *Pharmacogenomics*, 5:487-502.

Howeler, R., Lutaladio, N., Thomas, G. (2013). "Save and grow: cassava, a guide to sustainable production intensification." FAO, Rome (http://www.fao.org/3/a-i3278e.pdf) Cited 2014.

IITA. (International Institute of Tropical Agriculture). Crops: Cassava. ttp://www.iita.org/cassava. Accessed 2009.

Jørgensen, K., Bak, S., Busk, PK., Sorensen, C., Olsen, CE., Puonti-Kaerlas, JP. And Moller, BL. (2005). Cassava plants with a depleted cyanogenic glucoside content in leaves and tubers. Distribution of cyanogenic glucosides, their site of synthesis and transport, and blockage of the biosynthesis by RNA interference technology. *Plant Physiology*. 139:363-374.

Jørstad, TS., Langaas, M. and Bones, AM. (2007). Understanding sample size: what determines the required number of microarrays for an experiment? *Trends in Plant Science*, 12, 46-50.

Koda, Y. (1997). Possible involvement of jasmonates in various morphogenic events. *Physiologia Plantarum*. 100: 639-646.

Kumar, GNM., Iyer, S. and Knowles, R. (2007). Extraction of RNA from fresh, frozen, and lyophilized tuber and root tissues. *J. Agric. Food. Chem*. 55:1674-1678.

Li, Z. and Trick, HN. (2005). Rapid method for high-quality RNA isolation from seed endosperm containing high levels of starch. *BioTechniques* 38:872-876.

Li, ZL., Pan, YH., Sun, CB., Dong, HT., Luo, XL., Wang, ZQ, Tang, JL. and Chen, B. (2010). An ordered EST catalogue and gene expression profiles of cassava (*Manihot esculenta)* at key growth stages. *Plant Mol Biol* 74:573-590.

Lokko, Y., Anderson, JV., Rudd, S., Raji, A., Horvath, D., Mikel, MA., Kim, R., Liu, L., Hernandez, A., Dixon, AG. and Ingelbrecht, IL. (2007). Characterization of an 18,166 EST dataset for cassava (*Manihot esculenta* Crantz) enriched for drought-responsive genes. *Plant Cell Reports,* 26, 1605-1618.

Lowe, SB., Mahon, JD. And Hunt, LA. (1982). Early development of cassava (*Manihot esculenta*). *Can J Bot* 60(12): 3040-3048.

Mitprasat, M., Roytrakul, S., Jiemsup, S., Boonseng, O. and Yokthongwattana, K. (2011). Leaf proteomic analysis in cassava (*Manihot esculenta* Crantz) during plant development, from planting of stem cutting to storage root formation. *Planta*. 233:1209-1221.

Murashige, T., Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol Plant* 15: 473–497.

Nguyen, NK., and Willams, ER. (2006). Experiment designs for 2-colour cDNA microarray experiments. *Applied Stochastic Models in business and Industry*. 22:631-638.

Olsen, KM. (2004). SNPs, SSRs and inferences on cassava's origin. *Plant Mol Biol*. 56(4):517-26.

Owiti, J., Grossmann, J., Gehrig, P., Dessimoz, C., Laloi, C., Hansen, MB., Gruissem,, W. and Vanderschuren, H. (2011). iTRAQ-based analysis of changes in the cassava root proteome reveals pathways associated with post-harvest physiological deterioration. *The plant Journal*, 67: 145-156.

Palacho, AM., and Castel, AMM. (1991). Jasmonic acid induces tuberization of potato stolons cultured in vitro. *Plant physiol*. 97:1253-1255.

Sakurai, T., Plata, G., Zapata, FR, Seki, M., Salcedo, A., Toyoda, A., Atsushi Ishiwata, A., Tohme, J., Sakaki, Y., Shinozaki, K. and Ishitani, M. (2007). Sequencing analysis of 20,000 full-length cDNA clones from cassava reveals lineage specific expansions in gene families related to stress response. *BMC Plant Biology* 2007, 7:66

Sayre, R., Beeching, JR,. Cahoon, EB., Egesi, C., Fauquet, C., Fellman, J., Fregene, M., Gruissem, W., Mallowa, S., Manary, M., Maziya-Dixon, B., Mbanaso, A., Schachtman, DP., Siritunga, D., Taylor, N., Vanderschuren, H. and Zhang, P. (2011). The BioCassava Plus Program: Biofortification of Cassava for Sub-Saharan Africa. *Annu. Rev. Plant Biol.* 62:251-272.

Shewry, PR. (2003). Tuber storage proteins. *Annals of Botany*, 91, 755-769.

Sojikul, P., Kongsawadworakul, P., Viboonjun, U., Thaiprasit, J., Intawong, B., Narangajavana, J. and Svasti, MR. (2010). AFLP-based transcript profiling for

cassava genome-wide expression analysis in the onset of storage root formation. *Physiologia Plantarum,* 140, 189-198.

Suarez, MC., Bernal, J., Gutterrez, J., Tohme, J., and Fregene, M. (2000). Developing expressed sequence tags (ESTs) from polymorphic transcript-derived fragments (TDFs) in cassava (*Manihot esculenta* Crantz). *Genome*. 43:62-67.

Subramanian, A., Tamayo, P., Mootha, VK., Mukherjee, S., Ebert, BL., Gillette, M A., Paulovich, A., Pomeroy, SL., Golub, TR., Lander, ES. and Mesirov, JP. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Science USA,* 102, 15545-15550.

Sun, JQ., Jiang, HL., and Li, CY. (2011). Systemin/jasmonate-mediated systemic defense signaling in tomato. *Molecular Plant* 4(4): 607-615.

Takahashi, K., Fujino, K., Kikuta, Y. and Koda, Y. (1994). Expansion of potato cells in response to jasmonic acid. *Plant Science* 100(1):3-8.

Taylor, N., Chavarriaga, P., Raemakers, K., Siritunga, D. and Zhang, P. (2004). Development and application of transgenic technologies in cassava. *Plant Molecular Biology,* 56, 671-688.

Taylor, N., Gaitan-Solis, E., Moll, T., Trauterman, B., Jones, T., Pranjal, A., Trembley, C., Abernathy, V., Corbin, D. and Fauquet, C. (2012). A high-throughput platform for the production and analysis of transgenic cassava (*Manihot esculenta*) plants. *Tropical Plant Biol*. 5:127-139.

Till, BJ., Amos, EK., Kinyua, MG., Bado, S., Jankowicz-Cieslak, J., Huynh, OA., Henry, I., Comai, L. and Lagoda, P. (2013). Mutation induction and reverse-genetics for functional genomics and breeding. Tagung der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreichs 5 – 6.

Vanderschuren, H., Nyaboga, E., Poon, JS., Baerenfaller, K., Grossmann, J., Hirsch-Hoffmann, M., Kirchgessner, N., Nanni,P. and Gruissem W. (2014). Large-scale proteomics of the cassava storage root and identification of a target gene to reduce postharvest deterioration. *The plant cell*. 26:1913-1924.

Wagstaff, L., Cardie, C., Rogers, S. and Schroedl, S. (2001). Constrained K-means clustering with background knowledge. Proceedings of the Eighteenth International Conference on Machine Learning. p. 557-584.

Xu, J., Aileni, M., Abbagani, S. and Zhang, P. (2010). A reliable and efficient method of total RNA isolation from various members of spurge family (Euphobiaceae). *Phytochemical analysis*. 21:395-398.

Xu, J., Duan, X., Yang, J., Beeching, JR. and Zhang, P. (2013). Enhanced reactive oxygen species scavenging by overproduction of superoxide dismutase and catalase delays postharvest physiological deterioration of cassava storage roots. *Plant Physiology* 161:1517-1528.

Yang, J., An, D. and Zhang, P. (2011). Expression profiling of cassava storage roots reveals an active process of glycolysis/gluconeogenesis. *Journal of Integrative Plant Biology*, 53, 193-211.

Zainuddin, IM., Schlegel, K., Gruissem, W. and Vanderschuren, HV. (2012). Robust transformation procedure for the production of transgenic farmer-preferred cassava landraces. *Plant Methods*. 8:24.

Zhang, P., Bohl-Zenger, S., Puonti-Kaerlas, J., Potrykus, I. and Gruissem, W. (2003). Two cassava promoters related to vascular expression and storage root formation. *Planta*, 218, 192-203.

Zidenga, T., Leyva-Guerrero, E., Moon, H., Siritunga, D. and Sayre, RT. (2012). Extending cassava root shelf life via reduction of reactive oxygen species production. *Plant Physiology*. 159:1396-1407.

Table 1. GeneMath determined top ten significantly up-regulated genes in mature storage roots

GB_accession (unigene)	Organism	Blast_hit	Arabidopsis Orthologue	Root/Leaf ratio
DV445495	Cassava	Allergenic-related protein Pt2L4 [Manihot esculenta]	Unknown	2.82379985
DV446014	Cassava	Cassava specifc 746	Unknown	1.6413579
DV451141	Cassava	Unknown hypothetical protein p85RF	At1g51200	1.54818425
DV451479	Cassava	Lea5 protein [Citrus sinensis]	At4g15910	1.46267635
DV447512	Cassava	Fiber protein Fb37 [Gossypium barbadense]	At3g52800	1.3702485
DV447813	Cassava	Cassava-specific 878	Unknown	1.3664046
DV139865	Leafy spurge	Spurge specific 3836	Unknown	1.35062855
DV447978	Cassava	Hypothetical protein [Cicer arietinum]	At1g72150	1.24675385
DV449550	Cassava	Cassava-specific 2240	Unknown	1.14256645
DV139532	Leafy spurge	DnaJ protein [Hevea brasiliensis]	At3g44110	1.12819925

Table 2. GenMath determined top-ten significantly up-regulated genes in leaves

GB_accession (unigene)	Organism	Blast_hit	Arabidopsis Orthologue	Leaf/root ratio
DV441259	Cassava	Light harvesting chlorophyll a/b-binding protein	At1g29930	2.6269454
DV134008	Leafy spurge	Carbonic anhydrase, Chloroplast precursor (Carbonate dehydratase)	At3g01500	2.62582885
DV450878	Cassava	Chloroplast oxygen-evolving enhancer protein [Manihot esculenta]	At4g05180	2.2122086
DV452214	Cassava	Chlorophyll a/b-binding protein [Pisum sativum]	At5g54270	2.19743425
DV455684	Cassava	Questionable orf [Candida albican] SC5314]	Unknown	2.15877195
DV444628	Cassava	Oxygen evolving enhancer protein 1 precursor	At3g50820	2.11052345
DV458083	Cassava	Chloroplyll a/b-binding protein type III precursor - tomato	At1g61520	2.0925479
DV449255	Cassava	Unnamed protein product [Lycopersicon esculentum]	At1g76100	1.90737655
DV454456	Cassava	Chloroplast latex aldolase-like protein [Manihot esculenta]	At4g38970	1.8830498
DV136998	Leafy spurge	Unknown protein [Arabidopsis thaliana]	At5g16520	1.86234195

Table 3. Gene ontology determination for biological processes significantly up-regulated ($p<0.05$) in each stage of storage root development

Differentially up-regulated pathway	Initiation (SR2)	Starch filling (SR3)	Maturation (SR4)
Jasmonic acid biosynthesis	√		
Sucrose degradation	√		
Sucrose degradation to ethanol and lactate (anaerobic)	√	√	
Superpathway of sucrose degradation to pyruvate	√	√	
Superpathway of starch degradation to pyruvate	√	√	
Systemin Signaling	√	√	
Glycolysis II (plant plastids)	√	√	
Glycolysis I (plant cytosol)		√	
Valine biosynthesis		√	
Suberin biosynthesis		√	
Flavonoid biosynthesis		√	
Isoleucine biosynthesis		√	
Triacylglycerol degradation		√	
Superpathway of isoleucine and valine biosynthesis		√	
Superpathway of leucine, valine, and isoleucine biosynthesis		√	√
Oxidative ethanol degradation			√
Fatty acid omega-oxidation			√
Glyoxalate cycle			√
Superpathway of pantothenate and coenzyme A biosynthesis			√
Photosynthesis, light reaction			√
Leucine biosynthesis			√
Phospholipases			√
Photosynthesis			√
Auxin Signaling			√
Salvage pathway of purine nucleoside			√

√ = significantly up-regulated in this stage compared to previous stage

Table 4A. Significant genes in cluster 20 determined by K-means cluster to be over-represented as up-regulated in the initiation stage (SR2) of storage root development

Feature ID	Organism	Fold change	Blast_hit	Arabidopsis orthologue	functional annotation*
DV128628	Leafy spurge	1.4	putative aspartate aminotransferase [Oryza sativa (japonica cultivar-group)]	At2g22250	Aminotransferase class I and II
DV127294	Leafy spurge	1.29	major latex allergen Hev b 4 [Hevea brasiliensis]	At1g54030	GDSL-like Lipase/Acylhydrolase
DV125594	Leafy spurge	1.17	cyclin A3 [Lycopersicon esculentum]	At1g47230	G2/MITOTIC-SPECIFIC CYCLIN
DV129011	Leafy spurge	-1.92	cysteine synthase [Spinacia oleracea]	At3g59760	Cystathionine beta-synthase and related enzymes

*Predictive functional annotation (Cassava Genome, Phytozome)

Table 4B. Significant genes in cluster 15 determined by K-means cluster to be over-represented as up-regulated in the maturation stage (SR4) of storage root development

Feature ID	Organism	Fold change	Blast_hit	Arabidopsis orthologue	Functional annotation*
DV445495	Cassava	16.15	Cassava-specific 1668	Unknown	Manihot esculenta allergenic-related protein Pt2L4 (c1)
DV451479	Cassava	11.47	Lea5 protein [Citrus sinensis]	At4g15910	Late embryogenesis abundant protein
DV446014	Cassava	10.67	Cassava-specific 746	Unknown	No functional annotation
DV444641	Cassava	9.19	68418.m06286 expressed protein	At5g50730	Populus EST from severe drought-stressed opposite wood
DV443475	Cassava	8.35	ubiquitin [Pisum sativum]	At4g02890	ubiquitin like protein
DV121679	Leafy spurge	7.31	ubiquitin extension protein [Cucumis sativus]	At2g47110	ribosomal protein s27a + Ubiquitin family domain
DV442863	Cassava	6.45	fw2.2 [Lycopersicon esculentum]	At1g58320	PLAC8 family
DV141376	Leafy spurge	6.42	S-adenosyl-methionine-sterol-C- methyltransferase [Ricinus communis]	At5g13710	SAM-dependent methyltransferases

*Predictive functional annotation (Cassava Genome, Phytozome)

Table 4C. Significant genes in cluster 22 determined by K-means cluster to be over-represented as up-regulated in the filling stage (SR3) of storage root development

Feature ID	Organism	Fold change	Blast_hit	Arabidopsis orthologue	Predictive functional annotation (Cassava Genome)
DV117936	Leafy spurge	10.09	Spurge Specific 166	Unknown	Not found any region in Cassava
DV131468	Leafy spurge	7.56	putative GAR1 protein [Arabidopsis thaliana]	At3g03920	
DV120061	Leafy spurge	4.64	Spurge Specific 335	Unknown	Not found any region in Cassava
DV445389	Cassava	4.24	cassava-specific 699	Unknown	Protein tyrosine kinase/leucine rich repeat
DV441408	Cassava	4.12	cassava-specific 317	Unknown	Not found any region in Cassava
DV155281	Leafy spurge	3.64	Spurge Specific 2845	Unknown	Not found any region in Cassava
DV444942	Cassava	3.52	cassava-specific 668	Unknown	Sodium sulfate symporter and related arsenite permeases
DV443227	Cassava	3.51	cassava-specific 501	Unknown	Not found any region in Cassava
DV129279	Leafy spurge	3.28	Spurge Specific 1299	Unknown	Not found any region in Cassava
DV119603	Leafy spurge	3.2	Spurge Specific 280	Unknown	Not found any region in Cassava
DV118797	Leafy spurge	3.16	40S ribosomal protein S25 (RPS25E) [Arabidopsis thaliana]	At4g39200	S25 ribosomal protein
DV126381	Leafy spurge	3.03	Spurge Specific 1031	Unknown	Not found any region in Cassava
DV441855	Cassava	2.89	cassava-specific 364	Unknown	NmrA-like family (NITROGEN METABOLIC REGULATION PROTEIN NMR-RELATED)
DV120250	Leafy spurge	2.81	Spurge Specific 356	Unknown	RNA binding
DV119585	Leafy spurge	2.77	Spurge Specific 275	Unknown	Not found any region in Cassava
DV442549	Cassava	2.75	cassava-specific 1893	Unknown	Glyceraldehyde 3-phosphate dehydrogenase
DV131392	Leafy spurge	2.62	unknown [Arabidopsis thaliana]	At5g26940	Exonuclease

*Predictive functional annotation (Cassava Genome, Phytozome)

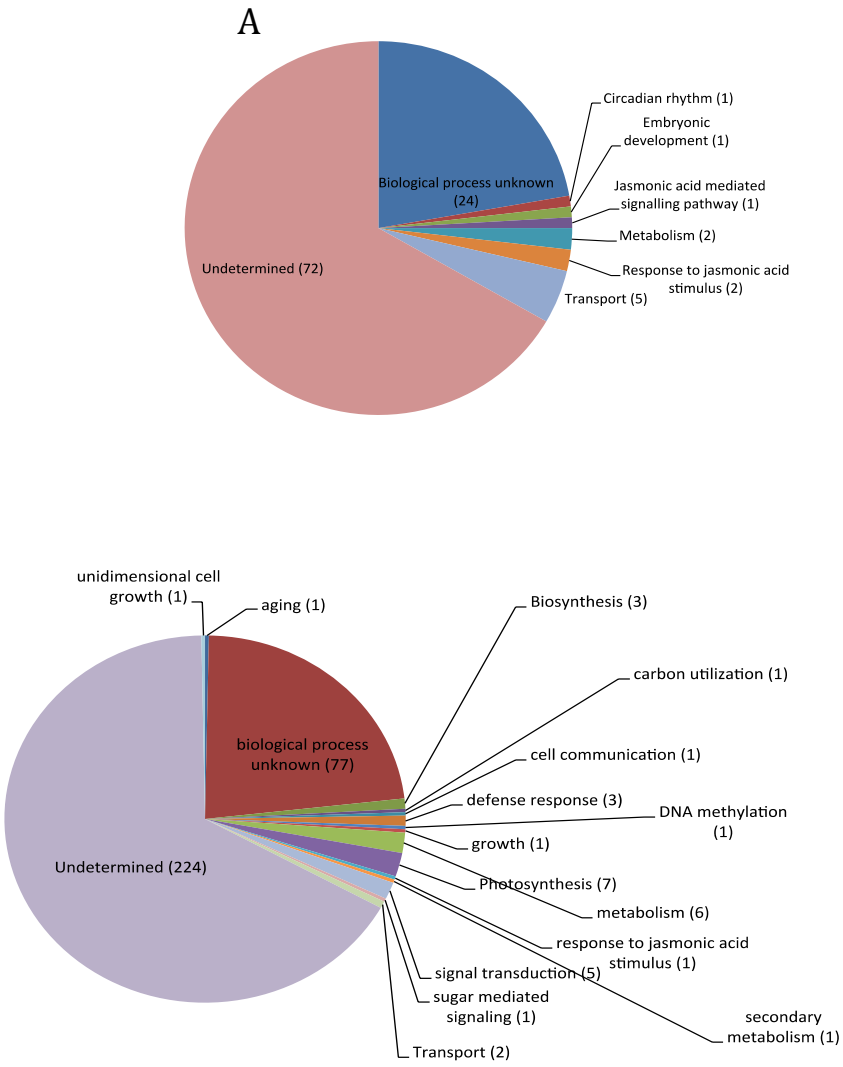

Figure 1. Predicted biological function of significantly up-regulated genes ($p<0.05$) in leaves and roots as determined by Gene Ontology. Microarray analysis was performed to compare gene expression in leaves and storage roots in three-month-old plants. Gene Set Enrichment Analysis was done using Pathway Studio software to categorize pathways represented by up-regulated genes. **A**. Up-regulated pathways in storage roots **B.** Up-regulated pathways in leaf. Number of genes in each pathway is shown in brackets.

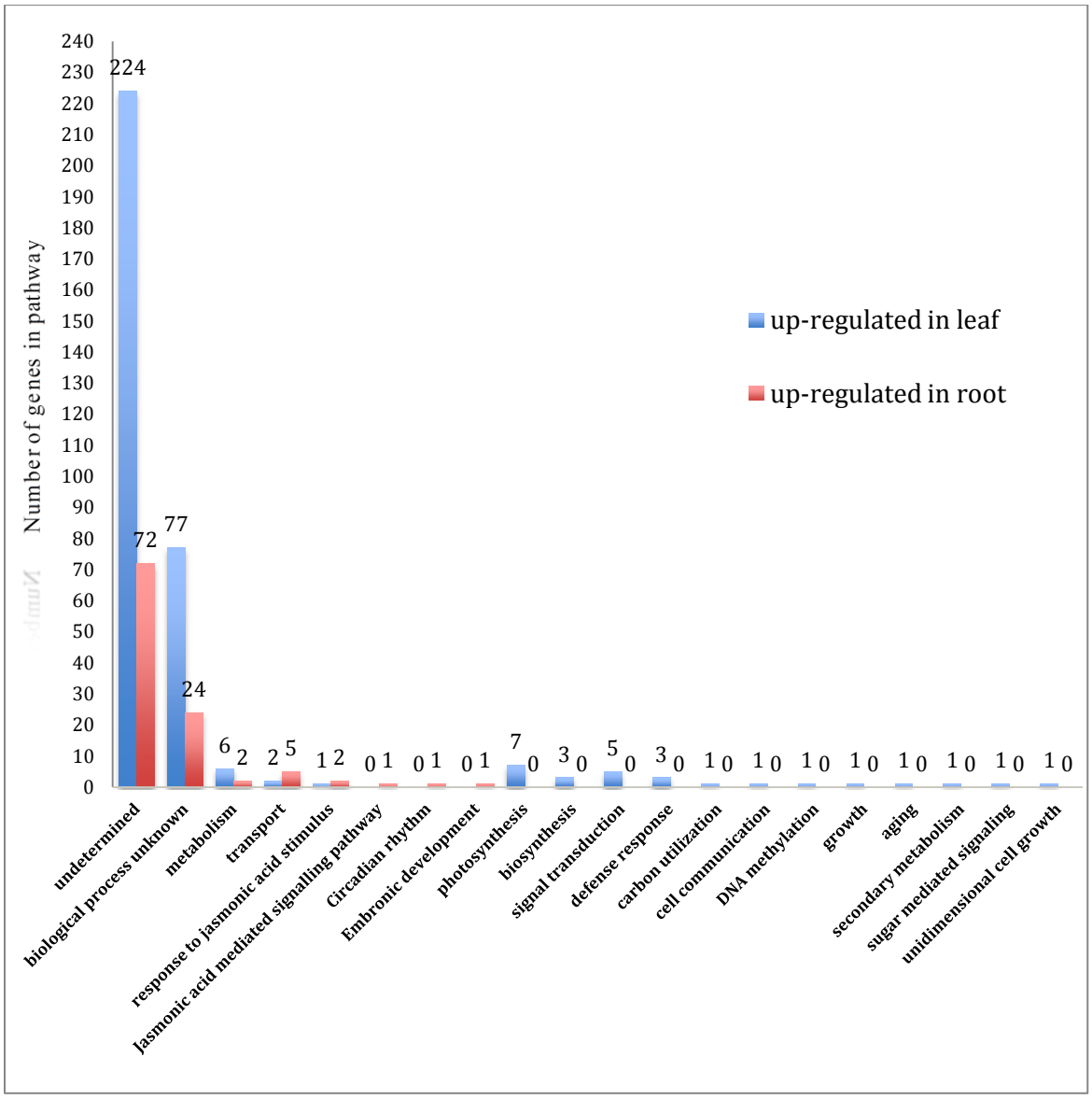

Figure 2. Enrichment analysis based on Gene Ontology to compare pathways containing up-regulated genes over-represented in storage root and leaf tissues.

Figure 3. Developmental stages in storage root development with comparative size of plants. SR1 – fibrous roots prior to tuberization, plants at two weeks after planting; SR2 – initiation stage, plants at one month after planting; SR3 – starch filling stage, plants at two months after planting; and SR4 – maturation stage, plants at three months after planting.

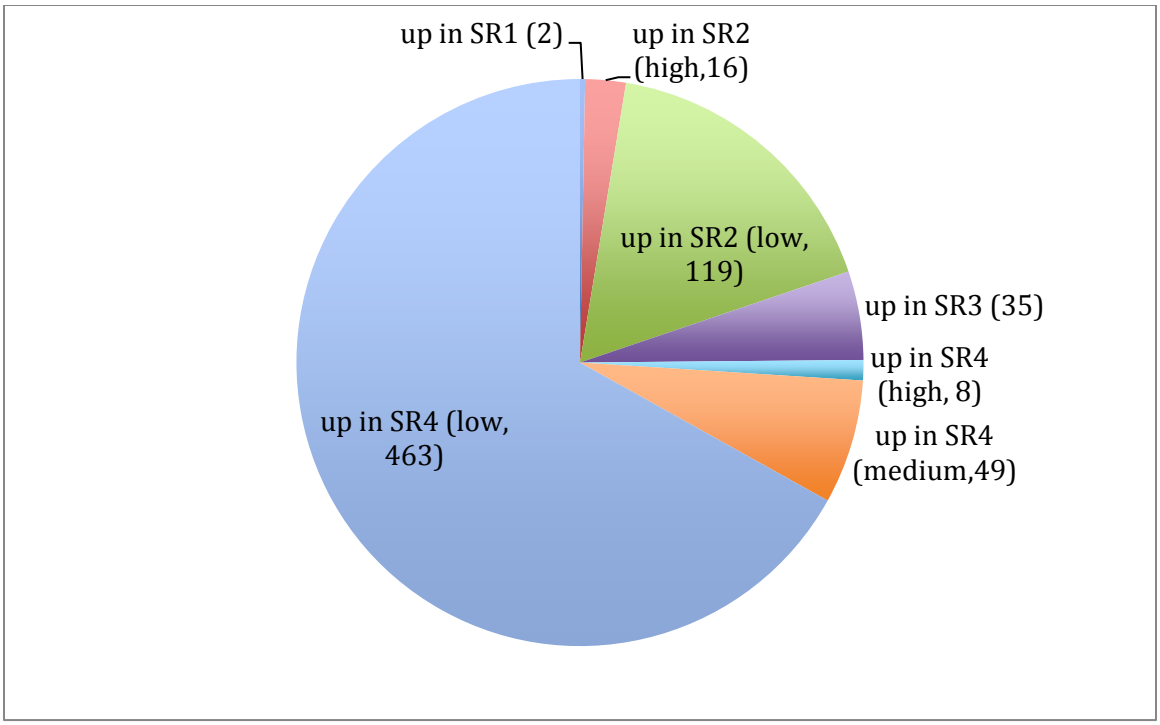

Figure 4. Significantly up-regulated genes ($p<0.005$) in each developmental stage of storage root formation determined by k-mean clustering. Number of genes and expression level (high, medium or low) is represented in each different developmental stage (SR1-SR4) as shown in brackets. Low expression equates to less than a normalized expression value below 5,000, medium to 5,000 – 10,000 and high to expression values above 10,000.

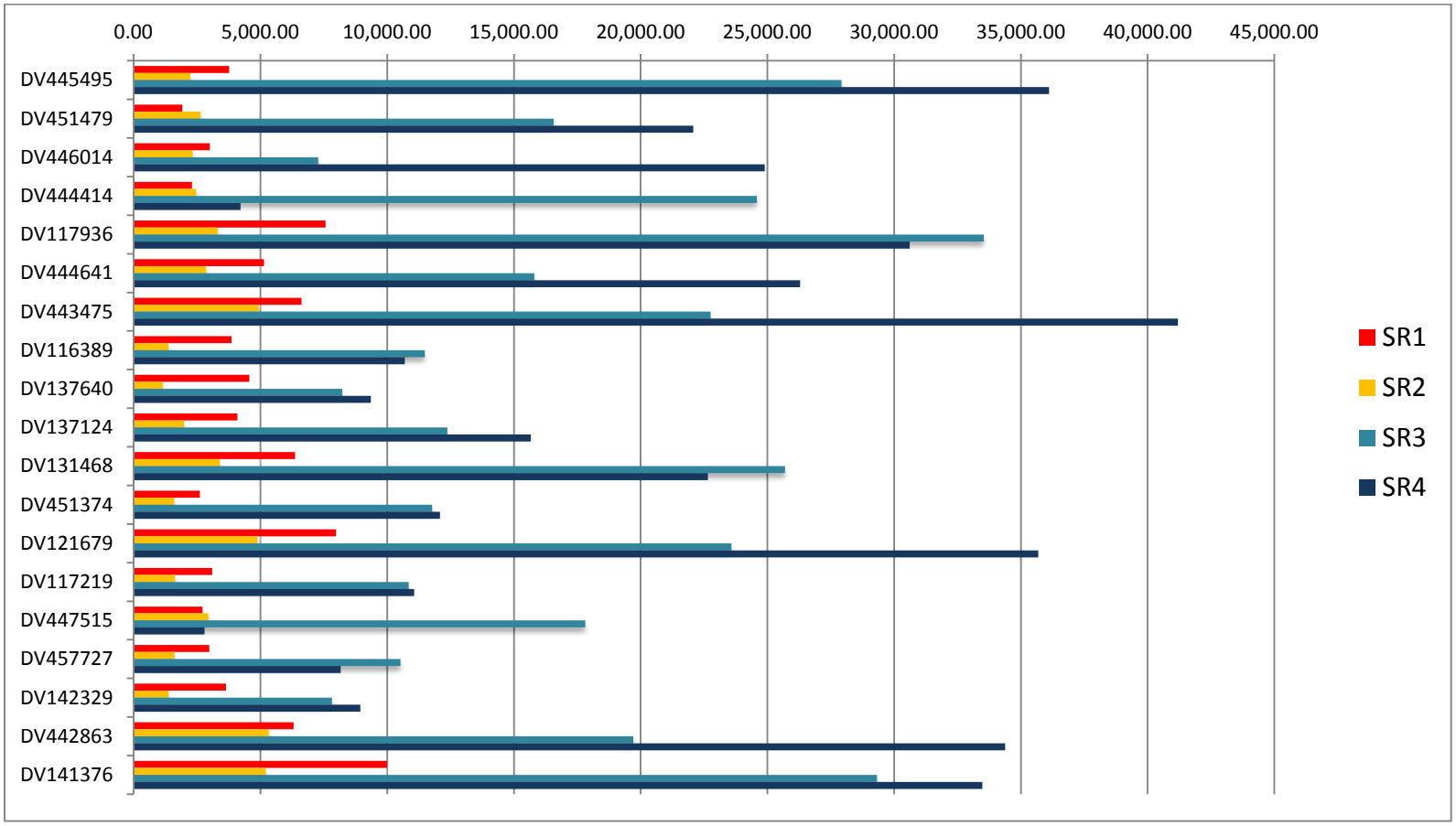

Figure 5. The average original expression value of twenty highest expression genes across the four root developmental stages (SR1-SR4) ranked by fold change calculated by GeneMath software.

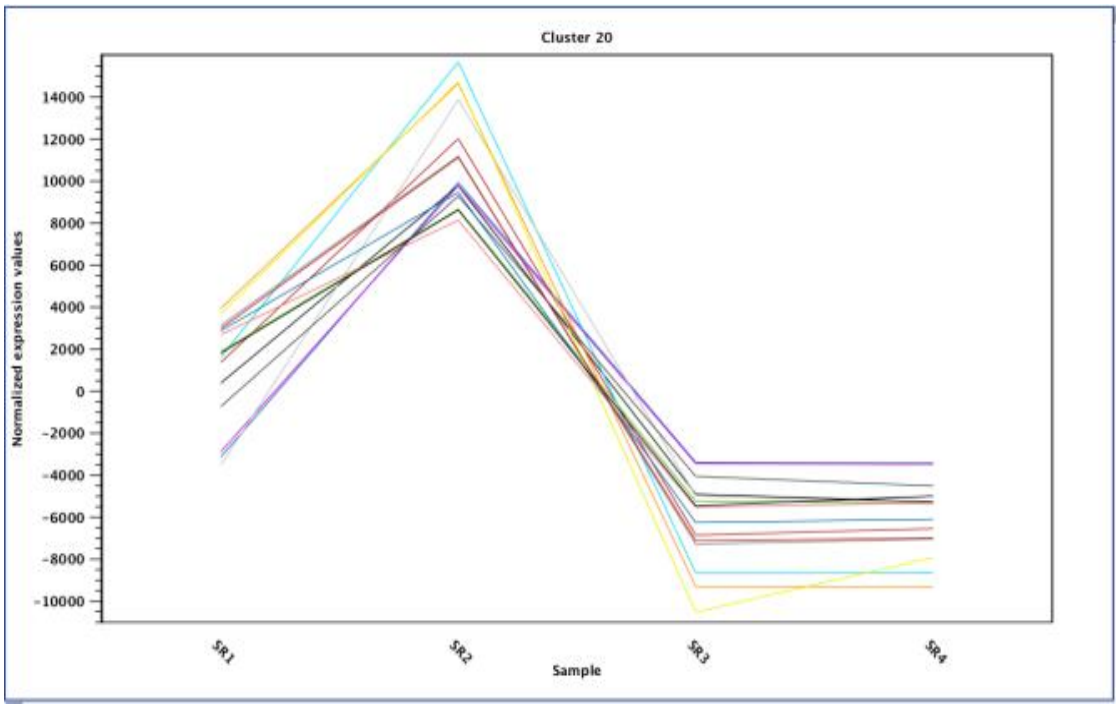

Figure 6A. Pattern of gene expression in cluster 20 across four developmental stages of root development (SR1-SR4). Four genes in this cluster showed significantly higher expression in the initiation stage (SR2) of storage root development. The expression pattern was shown as normalized expression value.

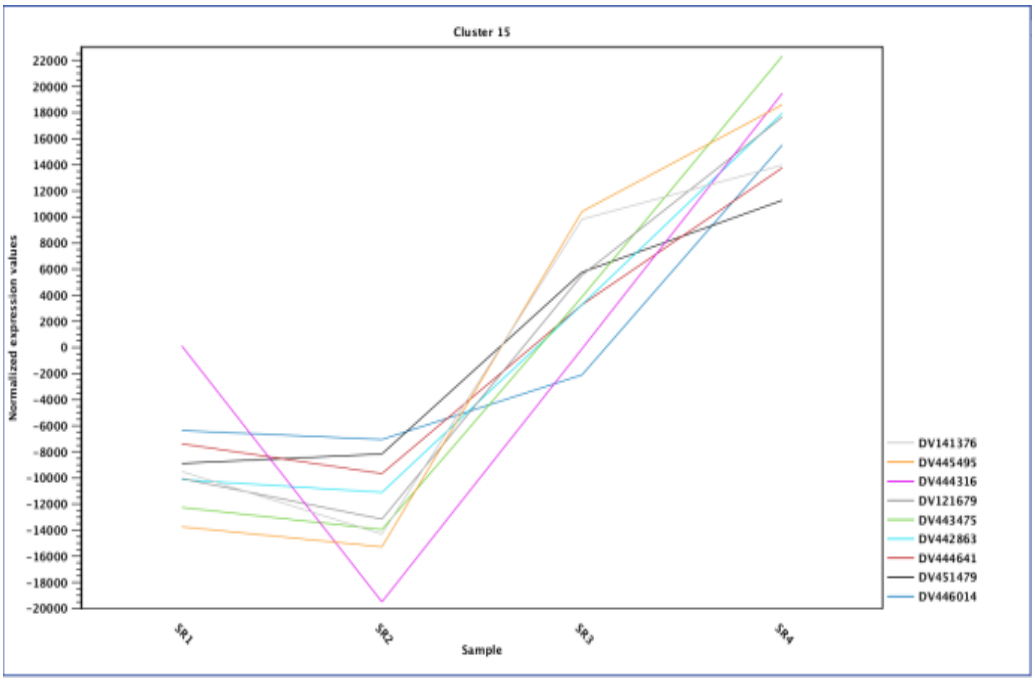

Figure 6B. Pattern of gene expression in cluster 15 across four developmental stages of root development (SR1-SR4). Eight genes in this cluster showed significantly higher expression in the maturation stage (SR4) of storage root development. The expression pattern was shown as normalized expression value.

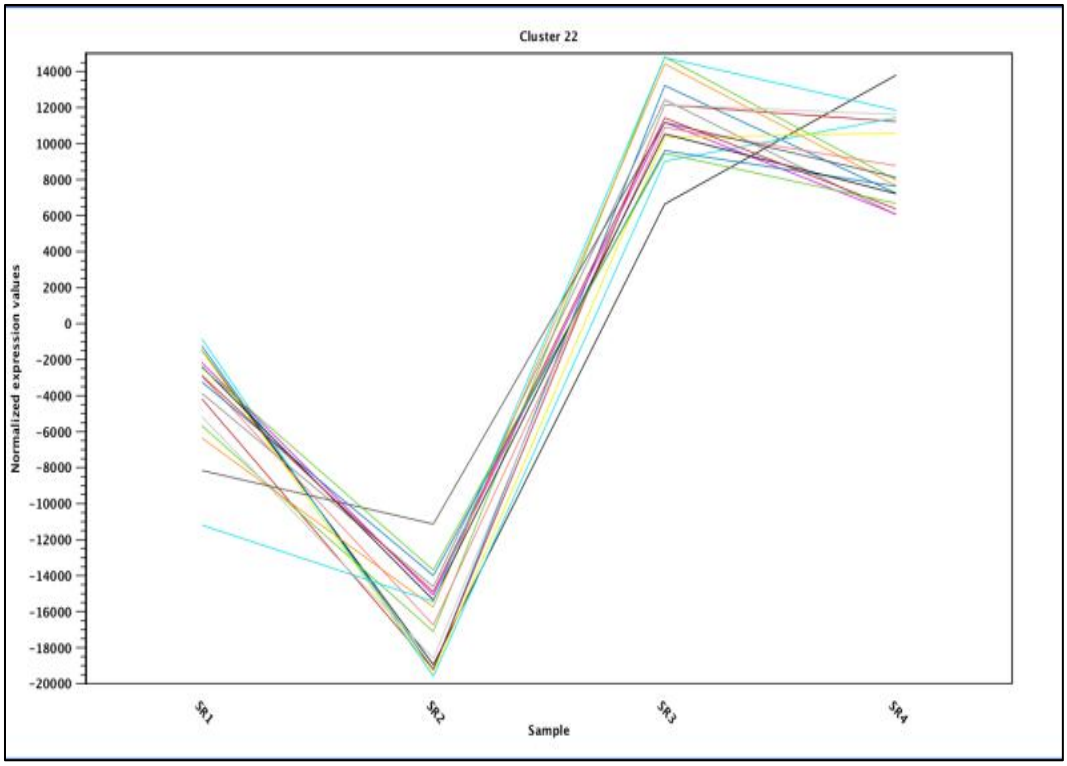

Figure 6C. Pattern of gene expression in cluster 22 across four developmental stages of root development (SR1-SR4). Seventeen genes in this cluster showed significantly higher expression in the filling stage (SR3) of storage root development. The expression pattern was shown as normalized expression value.

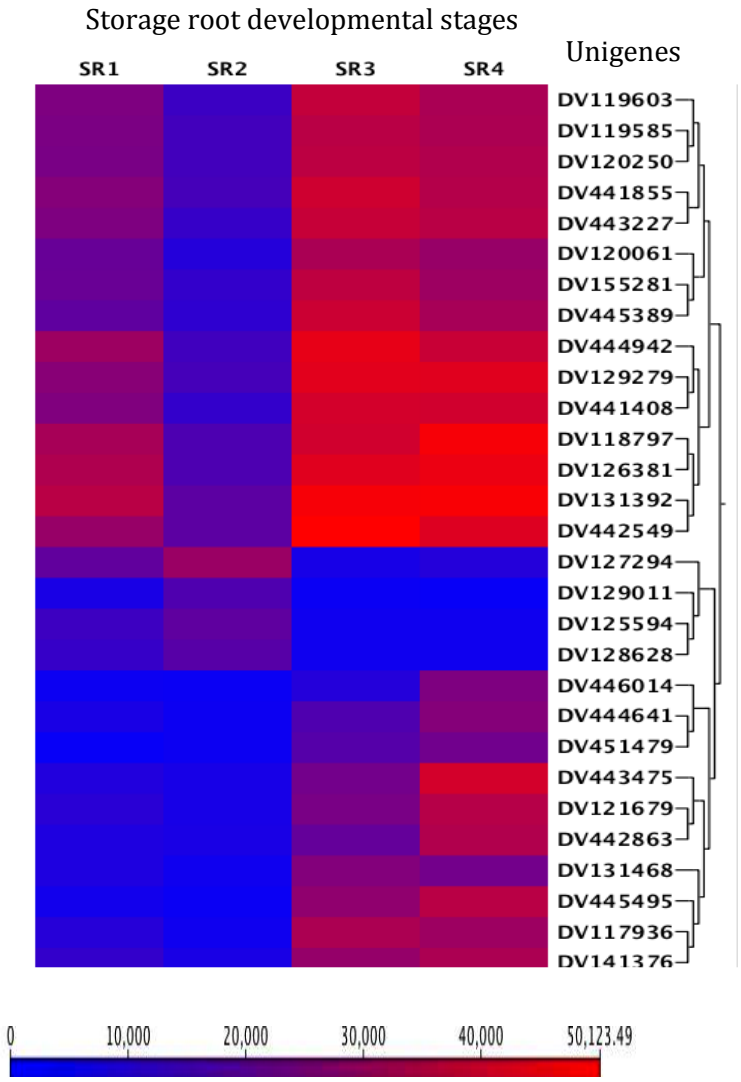

Figure 7. Heatmap analysis of expression profiles of unigenes identified in clusters 15, 20 and 22 across each of the four root development stages SR1-SR4. Phylogenetic tree is shown for unigenes displaying similar patterns of expressions.

Chapter 4. Gene expression analysis of candidate genes enriched during development of cassava storage roots

ABSTRACT

Molecular mechanisms controlling root tuberization in cassava can be revealed by studying differential gene expression between the tuberous and non-tuberous roots and across timing of the tuberization process. Microarray analysis across four different stages of storage root development identified three candidate genes involved in the maturation stage of storage root formation. These three candidate genes were identified as *Mec1* (DV445495), *MeATDI21 (*DV451479) and *ENOD40*-like genes (DV446014). In the present study, the functional role of these candidate genes was investigated in cassava using *in silico* analysis against the genome sequence available on public databases and the Cassava Genome sequence. In addition, the expression of each gene was examined by RT-PCR, Northern blotting and tissue printing in leaf, fibrous root, nodal root, stem, and storage root tissues, in order to validate actual expression in cassava plants. Northern blots and tissue printing indicated that *Mec1* is highly expressed in the storage root and stem tissues, while RT-PCR revealed a high signal of *Mec1* expression in the storage root, stem, and nodal root, with the low expression in leaf and fibrous root. *MeATDI21* was highly expressed in storage roots and nodal root with the low expression seen in fibrous roots and the stem. Finally, RT-PCR indicated that the cassava *ENOD40*-like gene was highly expressed in nodal roots with the low expression in storage root, stem,

and fibrous root. The further analysis is required for these candidate genes in order to annotate their putative role in cassava root tuberization.

INTRODUCTION

Root tuberization in cassava (*Manihot esculenta* Crantz) is a complex process that is dependent on genetics and environments. Diverse conditions including biotic stress and abiotic stress clearly affect the process of storage root development (El-Sharkawy, 2003, Lebot, 2009). The modification of nodal-derived roots into storage root occurs around one month after propagation via stem cuttings with visual thickening observed at approximately six weeks after planting under the optimum conditions in the greenhouse (see Chapter 2). However, various planting conditions such as low light intensity and high nitrogen fertilization can cause delays in storage root development. Conversely, early development of storage roots can be found under stress conditions such as low level of N fertilization and drought as well as the long day length period (Taylor *et al*., 2012). Therefore stress conditions may play a role in the initiation of storage root development, although the specific genes and pathways involved need to be clarified.

Gene expression profile analysis is an efficient approach to elucidate biological mechanisms in plants (Jung *et al*., 2014) and could have great value if applied to the study of cassava root tuberization. Several efficient functional genomics tools have been used to investigate gene expression in plants (Butte, 2002). Reverse genetics approaches have been used to study gene functions based on alteration of phenotype. Efficient reverse genetic tools were developed to investigate the role of target genes for validation of differential gene expression between controls and a mutated phenotype; for example,

gene silencing by RNAi technology, recombinant mutagenesis, microarray analysis, RNA-seq and proteomic analysis (Chen *et al*., 2009; Till *et al*., 2013). Gene expression profiling and proteomic analysis can be useful tools for identification of genes involved in significant traits, whereas microarray technique or RNA-seq can be powerful for elucidating the complex biological systems under controlled environmental conditions (Butte, 2002; Zhang, 2003). Factors affecting efficiency of microarray experiments such as experimental design, component analysis, and the minimum sample size have been improved in order to increase the accuracy of data interpretation (Butte, 2002; Churchill, 2002; Subramanian *et al*., 2005; Nguyen and Williams, 2006; Jørstad *et al.*, 2007). As a result, transcriptional profiling from cDNA microarrays have been shown to be an efficient method for prediction of putative pathways and candidate genes involved in the studied mechanisms (Duggan *et al*., 1999; Hardiman, 2004; Chen *et al*., 2009).

Genome-wide analysis, especially transcriptome analysis in cassava research, was first implemented using the cDNA-AFLP to develop cassava-specific ESTs (Suarez *et al*., 2000). The evolutionary and geographical origin of cassava between domesticated varieties and wild populations were determined using SNPs (single nucleotide polymorphisms) and SSR (simple sequence repeat) in order to study their genetic diversity (Olsen, 2004). Moreover, databases of ESTs derived from various cassava cDNA libraries were developed and are available through the public database (Lopez *et al*., 2004; Lokko *et al*., 2007; Sakuria *et al*., 2007; Li *et al.*, 2010). ESTs which are shown to be predicted unigenes located at the same transcribed locus on UniGene database facilitate faster interpretation of putative function of candidate genes.

The draft genome sequence for cassava was announced in 2009 via JGI's Phytozome (www.phytozome.net/cassava) for providing new opportunities for improving the cassava crop. This cassava draft genome sequence does not only facilitate the gene analysis and annotation for cassava-specific genes, but also allows the advancement of data on genome-wide expression analysis for traits such as storage root formation. Furthermore, a total of 9,600 cDNAs and gene expression profiles at the key growth stages of cassava were sequenced and established as another catalogue of expressed sequence tags (ESTs) (Li *et al*., 2010). The library of a unigene set of 5,700 EST sequences was generated using differential expression in a cassava variety that differed in starch content and bacterial blight resistance (Lopez *et al*., 2004). These large cassava EST resources publically available on published databases are useful for mining the genomic knowledge of cassava.

To date, different databases for *in silico* analysis is available, for example the UniGene database (http://www.ncbi.nlm.nih.gov/unigene), to determine groups of cDNAs that share the same transcribed locus. The large EST libraries in UniGene database provide the possibility to identify transcripts from the same locus of a given candidate gene, and then allow prediction of the biological function based on annotated cDNAs in the same transcribed locus. However, most cassava-specific cDNAs still lack effective functional annotation. An alternative tool for analysis is the genome-wide analyses of alternative splicing variants. Alternative splicing occurs under certain condition such as diverse growth and development to create multiple transcripts, or

isoforms, from a single gene (Barbazuk *et al*., 2008). Isoform derived alternative splicing has been reported to play a role in biological function (Stamm *et al*., 2005; Reddy, 2007). Alternative splicing causes an increase of mRNA diversity and also affects the localization, enzymatic properties, and ligand interaction of proteins (Kelemen *et al*., 2013). Moreover, alternative splicing is also shown to play a role in plant tissue identity, stress adaptation, and defense responses (Gassmann, 2008; Mastrangelo *et al*., 2012; Thatcher *et al*., 2014).

According to the microarray results described in Chapter 3, three unigenes were selected as candidate genes with putative relation to storage root development in the latter maturation stage (SR4). Those unigenes are DV445495, DV451479, and DV446014 that were shown to be enriched in the maturation stage (Ch3, Fig. 7/Table 4B). The functional protein domain of DV445495 was identified through Cassava Genome to be allergenic-related protein Pt2L4, which has been proposed previously to be related to storage root and vascular tissue of cassava (de Souza *et al*., 2004; 2006). The gene corresponding to DV451479 was annotated as the homologue of *ATDI21* (Arabidopsis thaliana Drought-induced 21) in cassava. However, the functional annotation of DV446014 remains to be identified.

The cDNA of DV445495 was characterized as the *Mec1* gene encoding allergenic-related protein Pt2L4 and identified as the gene related to cassava storage root formation (de Souza *et al*., 2002; 2003; 2004; 2006). Two cassava promoters related to storage root formation and vascular expression were developed from cDNA of c15 and

c54 and characterized in cassava (Zhang *et al*., 2003). Furthermore, a glutamic acid-rich protein Pt2L4 promoter was shown to drive high transgenic expression of a marker gene in storage roots and stem (Beltran *et al.*, 2010). DV451479 is the homologue of *ATDI21* gene in cassava. The *ATDI21* gene is located in Chromosome 4 in Arabidopsis and its transcription level of *ATDI21* was reported to be elevated by changes in abscisic acid levels (Gosti *et al*., 1995). Because the coding sequence consisted of the late embryogenesis abundant protein 3 (Lea3), the putative function of DV45149 was predicted as the response to biotic and abiotic stress (Costa *et al*., 2011). DV446014 is one candidate unigene that requires further study to determine its functional role in cassava tuberization.

In this study, I attempted to analyze gene expression of the three candidate unigenes (DV445495, DV451479, and DV446014) related to storage root development at maturation stage of the tuberization process. The expression of alternative splicing variants of each candidate gene was also studied in different tissue types of cassava to predict the putative role in tissue-specific regulation.

MATERIALS AND METHODS

Plant materials

Three-month-old plants derived from tissue culture were transferred to greenhouse as described in Chapter 2. Young leaves consisting of the second to fourth leaf below the

apical point were collected. Samples were taken with gloved hands, wrapped with aluminum foil, labeled and immediately frozen in the liquid nitrogen.

Plants were removed from their pots, the whole rooting system gently removed and soil rinsed away from the root tissues under running tap water. The whole root tissues were briefly dried with paper towel. The fibrous roots (FR), nodal roots (NR), and storage roots (SR) were carefully separated and cut from the plant using a sharp blade. Tissues were placed into 15 ml centrifuge tubes and kept on dry ice until RNA extraction. Storage roots were removed from the stem using a sharp blade and rinsed again with tap water. Cleaned storage roots were quickly peeled and cut into small pieces approximately 125 mm^3 in size and transferred into a 50 ml plastic Falcon tube and closed using aluminum foil. Storage root samples were freeze-dried in a lyophilizer (VirTis lyophilizer #FM 25ES-53, SP Scientific) for 24 hours and then stored at room temperature till required for RNA extraction. Stem tissues was cut into small pieces approximately 125 mm^3, placed in 50 ml Falcon tunes, and freeze-dried in the same manner as the storage roots. The freeze-dried samples were kept at room temperature till proceeding with RNA extraction.

RNA extraction for each organ type

Total RNA from two grams of leaf sample, one gram of fibrous roots, and one gram of nodal roots were separately extracted using a modified CTAB protocol adapted from Lodhi *et al.*, (1994). The CTAB extraction buffer was modified to be composed of 2% CTAB (Sigma), 100 mM Tris-HCL (pH 8.0, Sigma), 20 mM EDTA (Sigma), 1.4 M NaCl

(Sigma), and 2% v/v beta-mercaptoethanol (Sigma) in RNase-free water, with the beta-mercaptoethanol added to the buffer immediately before extraction. The buffer was incubated at 65^{o}C before adding onto 0.5 g of ground sample tissue in the 15 ml centrifuge tube. Samples were incubated on 65^{o}C for 30 min with gently mixing. An equal amount of chloroform was then added and the sample mixed for 10 min at room temperature before centrifuge at 10000 rpm. After centrifugation, the aqueous phase containing nucleic acid was dispensed to a new 15 ml centrifuge tube and purification steps performed twice using equal volumes of Chloroform:isoamyl alcohol (24:1). Total RNA was precipitated from the supernatant after centrifugation with 0.7 volumes of ice cold isopropanol in a 2 ml Eppendorf centrifuge tube. The RNA pellet was washed with 500 µl 75% ethanol three times and then dissolved in 1 ml RNase-free water before an overnight precipitated with 1/3 volume of 8 M LiCl (Ambion) at -20^{o}C. The final RNA pellet was washed with ice cold 75% ethanol three times.

Total RNA of storage roots and stem was extracted using a modified protocol for cassava storage root RNA as described in in Chapter 3. The RNA extraction buffer was composted of 100 mM LiCl, 1% (w/v) SDS, 100 mM Tris-HCL pH 7.5, 100 mM EDTA in diethylpyrocarbonate (DEPC) treated water, and 1% v/v of beta-mercaptoethanol, which was added into the buffer before extraction. The lyophilized storage root sample was briefly ground in liquid N_2 to a fine powder. Half a gram of ground, lyophilized tissue was transferred into 2 ml Eppendorf microtube. First, 800 µl extraction buffer was added into the sample, then the premix of 800 µl acidic phenol:chloroform (AM9720, Ambion) immediately added to the same tube. The extract was well vortexed and

incubated at room temperature for one hour. The extract was centrifuged at max speed for 30 min at 4^{o}C. The clear aqueous phase of the extract was carefully transferred into the new centrifuge tube and then centrifuged again for 5 min. The supernatant was transferred into a new tube, 1/3 volume of 8M LiCl was added to the supernatant with gently mixing and incubated at -20^{o}C for 16 hours. The extract was centrifuged at 4^{o}C for 30 min to harvest the RNA pellet, followed by two washing steps using ice cold 75% ethanol. The pellet was then dried before elution with 30 µl RNAse-free water. The RNA was treated with DNase using DNA removal kit (DNA free, AM1906, Ambion) following by the cleanup steps as describe in the manual (RNeasy mini kit,Qiagen). RNA quality and quantity was confirmed by denaturing agarose gel electrophoresis and a spectrophotometer, respectively.

RT-PCR

The expression of selected unigenes DV445495 (allergenic-related protein Pt2L4), DV445495 (cassava *ATDI21*), and DV446014 (No functional annotation) was performed by reverse transcriptase PCR (RT-PCR) from total RNA extracted from different organs of three-month-old plants. Primers were custom-designed in order to amplify cDNA fragments corresponding to candidate gene sequences (Table 1) using the PrimerQuest tools (http://www.idtdna.com). The first strand cDNA were synthesized from total RNA using the SuperscriptIII first-strand cDNA synthesis system (#18080-051, Invitrogen, Life Technology). cDNA from the genomic RNA was synthesized using Oligo(dT)20 as a primer and the synthesis mix incubated at 50^{0}C for 50 min. The reaction was then incubated at 85^{o}C for 5 min to terminate the reaction. The RNA duplex was degraded by

RNase H at 37^{o}C for 20 min, and 2 µl cDNA used as the template in PCR reactions with the primers and the optimized annealing temperature specific for each candidate gene (Table 1). The expression of selected gene was amplified by PCR using One tag DNA polymerase (NEB, MA). The amplified fragments of candidate unigene were run on a 1% agarose gel electrophoresis to reveal expression within the different organ types.

***In silico* analysis of alternative splicing forms of candidate genes**

The three unigenes DV445495 (Pt2L4 protein), DV445495 (cassava *ATDI21*), and DV446014 were analyzed by blasting the non-redundant sequences, including the cassava-specific unigenes, against the TAIR database (http://www.arabidopsis.org/Blast/index.jsp) in order to identify the Arabidopsis orthologue and predict their function using BLASTN and BLASTP via NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The sequence of candidate genes was retrieved from the ESTs database (NCBI) correlating to their unique accession number, then these sequences were blasted into Cassava genome (www.Phytozome.net, JGI, CA) for analysis of the actual cassava sequence and their alternative splicing forms (RNA Isoform). The alignment of nucleic and protein sequences were analyzed through CLC Main Workbench software (CLCbio, a Qiagen company, MA).

Because information for the DV446014 transcript was not available on the Cassava Genome database (Phytozome), nor through BLASTN search on NCBI, the UniGene database (http://www.ncbi.nlm.nih.gov/unigene) was used to identify unigenes from the same transcribed locus. The Rfam database, which is a collection of non-coding

RNA families, was used to identify RNA families with conserved RNA secondary structure domain (Griffiths-Jones *et al*., 2003, Burge *et al*., 2012). The RNA sequence of DV446014 transcript was submitted to the Rfam database (http://rfam.xfam.org). Information of the match RNA family was then further analyzed by comparing to the reference sequence to identify the regulatory domains within DV446014.

Northern blot hybridization of *Mec1* RNA

To examine the expression of the *Mec1* gene in different cassava organs, Northern blot analysis was performed to compare gene expression between the three-month-old plants derived from tissue culture. Fibrous roots (FR), leaves (L), stems (ST), and storage roots (SR) were sampled as described above. Large storage roots, approximately 6 cm in diameter and woody stems were also harvested from field grown eight-month-old plants of cv. 60444 and imported from the Tropical Agricultural Research Station Mayaguez, University of Puerto Rico. Field grown storage roots were peeled, chopped into 125 mm^3 size pieces, lyphilized and processed in the same manner as greenhouse derived materials. Stems were peeled to remove the bark, cut into 125 mm^3 size pieces and processed likewise. Total RNA of leaf, fibrous root, stem, and storage root was extracted as described above.

Ten µg of genomic RNA from each tissue was run on a 1% denaturing agarose gel under 80 eV for 2 hrs. RNA was transferred to a Hybond-N+ nylon membrane (Amersham Pharmacia Biotech, USA) with transfer buffer (20x SSC). A DIG-labeled probe was prepared by PCR amplification of the coding sequence of Pt2L4 protein using

primers DV445495-F (5'-CCTTGGCTGACTATGGCTACTGCTGAGGTAGT-3') and DV445495-R (5'-TCAGATTTCTTCTCATCACCTTCTTCCTCC-3') by PCR DIG-probe synthesis kit (Roche Life Science). Optimum PCR conditions used were: initial denaturation at 94^{o}C for 4 min, followed by 35 cycles of 30 s at 94^{o}C, 60 s at 58^{o}C, and 60 s at 72^{o}C, and incubation at 72^{o}C for 5 min. DIG-labeled probe was checked for quality on agarose gel electrophoresis and was kept at 20^{o}C till hybridization.

The membrane was pre-hybridized at 60^{o}C for 15 min, follow by hybridization with buffer containing the DIG-labeled probe at 60^{o}C for 16 hrs. The hybridized membrane was washed twice at 60^{o}C for 20 min. The DIG blocking step was incubated at room temperature for 30 min using a shaker, after which the anti-DIG antibody (1:15,000) was added to the DIG blocking buffer and incubated at 25^{o}C for 1 hr. The membrane was washed three times with 1x maleic buffer at room temperature for 20 min using a shaker and then incubated in equilibration buffer at room temperature for 5 min. Chemiluminescence detection was done by applying chemiluminescent substate (CDP-star, Roach Life Science) for alkaline phosphatase (10 µl per cm of membrane) and incubated at room temperature for 5 min. The membrane was covered with cling film and signal detection performed by exposure to X-ray film for 20 min.

In situ hybridization of *Mec1*

To examine localization of *Mec1* RNA in developing storage roots and stem, in situ hybridization was performed using a tissue printing technique following Pluskota *et al*. (2011). Fresh storage roots and stems were obtained from three-month-old greenhouse

grown plants. Storage roots were sliced into 50 μm thick sections using a sliding microtome (Uchida Yoko, VWR). Sections were cut from the neck region of the storage root and fleshy tuberized tissues and from the semi-woody stem. RNA printing was performed immediately by placing the thin sections on a Hybond-N+ nylon membrane (Amersham Pharmacia Biotech, USA), and then covering the membrane with parafilm. Each tissue section was pressed hard down onto the membrane using the thumb for 20 seconds ensuring that all areas of the tissue contacted the membrane. After the pressure was released, tissue sections were removed using forceps taking care not to touch the membrane. The membrane was then cross-linked using a UV cross-linker and processed for hybridization with the DIG-labeled DNA probe as described above for the Northern blot analysis.

RESULTS

Selection of candidate genes involved in cassava storage root maturation for further analysis

According to the results of microarray analysis described in Chapter 3, differential gene expression across the stages of storage root formation demonstrated transcriptional changes in each developmental stage from SR1 (before tuberization), SR2 (initiation), SR3 (filling), and SR4 (maturation). Three distinct clusters of differential gene expressions generated by k-means clustering were selected to identify candidate genes involved in storage root formation. Cluster 15 consisted of unigenes that were highly up-regulated in SR4, showing fold changes of 6.4-16.2 (Ch3, Table 4B/Fig. 6B). Cluster 20 and cluster 22 consisted of genes showing high expression in SR2 but the fold change of

their expression was only a maximum of 1.4 (Ch3, Table 4A&C and Fig. 6A&C). Corresponding sequences for unigenes in cluster 22 could not be found in any region of the published cassava genome. In contrast, in cluster 15, of the eight significantly expressed unigenes, all could be found to have similar annotated sequences in the cassava genome. The three unigenes (DV445495, DV451479, and DV446014) that showed highest fold change across the four developmental stages and highest expression in storage root tissues were therefore selected for further study (Table 2).

Sequence similarity analysis of candidate genes

BLAST searches (performed 10/22/14) were performed for cDNA sequences of unigenes DV445495, DV451479, and DV446014 against the non-redundant database (NCBI) and Cassava Genome Database (Phytozome.net). The BLAST search showed the cassava-specific allergenic-related protein Pt2L4 (*Mec1*) to be highly similar to the DV445495 (E value 0.0) both in NCBI database and Cassava Genome. Cassava *ATDI21*, the homologue of *Lea3* was shown to be highly similar to DV451579 (E value 0.0). In contrast, there was no significant similarly found for DV446014 against the non-redundant database on NCBI. However, one region was found to be highly similar to DV446014 on Cassava Genome (E value =0.0), but no transcript or functional annotation was available.

DV445495 (Allergenic-related protein Pt2L4)

Functional annotation of DV445495 in cassava

BLAST searches for DV445495 against the NCBI database predicted five highly similar sequences as shown in Table 3. The results showed that DV445495 falls within the

distinct group of *c1* genes that is specific in cassava and called *Mec1*. *Mec1* encodes the allergenic-related protein Pt2L4 previously reported to be involved in development of secondary xylem parenchyma in cassava storage roots (de Souza *et al*., 2004).

The DV445495 cDNA sequence was used as the query sequence for a BLAST search against the Cassava Genome through Phytozome. Two significant regions were predicted to have the homologue sequence of DV445495. The sequence in the first region was located in scaffold06844 and shown to be highly similar to the reference sequence (E-value=0, score 1058.1), whereas the sequence in the second region located in scaffold 03131 was showed to have lower similarity (E-value=1.8e-70, score 270.0) (Fig. 1A). Two alternative splicing variants were predicted in the first region on scaffold06844 (Fig. 1B), while the genes located on scaffold03131 showed no alternative splicing forms. Although the cDNA sequence of DV445495 was shown to be highly similar to cassava c1 (*Mec1*) encoding allergenic-related protein c1 (Pt2L4) on NCBI database, no functional annotation for this locus was presented on cassava genome.

Expression of Mec1 *in storage root and stem*

The full length cDNA of DV445495 was used as a probe to investigate expression of *Mec1* gene by Northern blot analysis. Total RNA from leaves, storage root, fibrous roots, and stem was extracted from three-month-old tissue culture derived plants and hybridized with DIG-labeled probe. Expression of *Mec1* was shown to be high in tissues of storage root and stem but was absent or below detectable levels in the leaf and fibrous root (Fig. 2). Expression of *Mec1* was also examined in storage roots, stem, and leaf tissues of

eight- month-old field grown roots. Likewise, *Mec1* was seen to be highly up-regulated in storage root and stem tissues but not detectable in leaves collected from these fully grown plants (Fig. 2). This data indicate that expression of the *Mec1* gene encoding for the Pt2L4 protein was not specific to storage root tissues but was present also in the stem. Interestingly, its pattern of expression was similar in relatively juvenile greenhouse grown plants and almost mature plants cultivated under tropical conditions in the field. The latter information is important as it provides confidence that although studies in the greenhouse are performed on small plants under artificial conditions, the data generated does relate well to the field and can therefore be treated with some confidence.

Expression of Mec1 *in secondary xylem parenchyma in storage roots and stem vascular tissue*

Mec1 expression was confirmed at the tissue level by *in situ* hybridization using the tissue printing technique. The same full length cDNA of DV445495 was used as a label for this Northern blot analysis. The neck of storage root (see Ch2, Figs. 1 & 5) through to the widest area of storage root, as well as stem of three-month-old cassava plants were freshly cut into sections approximately 50 μm thick, pressed onto a nylon membrane and blotted as described in Material and Methods. Expression of *Mec1* was seen within both storage root and stem tissues especially associated with the starch storage tissue in roots (Fig.3A) and the vascular tissue in stem (Fig. 3B). Within the storage root, expression was seen least in the neck region of storage root, where the majority of the tissues are known to be lignified xylem (Ch 2, Fig. 4) with little starch storage, while the expression was highest within the more cellulosic, starch-containing storage region located further

from the stem (Fig. 3). The result therefore corresponds with those of de Souza *et al.* (2004) who originally proposed the function of Pt2L4 protein to be associated with secondary xylem parenchyma of storage root. Expression in the more mature stem would also be explained by this information as the stem of cassava is lignified and does store starch within these xylem tissues.

Alternative splicing variants prediction of Mec1

Two homologous transcripts showing high similarity with *Mec1* were predicted in Cassava Genome Database (Fig. 1A). The first hit region was located in scaffold06844 (SC06844) which showed two alternative splicing variants (*Mec1*-1, and *Mec1*-2), while the second region was located in scaffold03131 (SC03131) and had no alternative splicing (*Mec1*-like). SC06844 region was further studied to test whether the putative splice variants affected gene expression in different organs or tissue types of cassava. The coding sequences of two splice variants (*Mec1*-1 and *Mec1*-2) of the first homologue were retrieved from the Cassava Genome Database and then aligned against the reference sequence of *Mec1*. From the alignment, the coding sequence of *Mec1*-1 maintained its internal intron, whereas this intron was spliced out in the second splice variant, *Mec1*-2. The results showed that the transcript of *Mec1* was more likely similar to the splice variant 1 (*Mec1*-1) than to the splice variant 2 (*Mec1*-2), as shown in Figure 4.

Gene expression analysis of Mec1 *and its predicted alternative splicing variants in different tissue types*

Primers were designed to amplify the entire coding sequence of DV445495 (*Mec1*) and specifically amplify each predicted variant (Table 1). RT-PCR was performed on total RNA extracted from leaves, stems, fibrous roots, nodal roots (see Chapter 2), and storage root tissues. Amplification of the complete coding sequence showed high expression of *Mec1* in storage roots, nodal roots, and stem compared with lower expression in leaf and fibrous roots (Fig. 5). The same expression pattern was seen for *Mec1* homologue 2 (*Mec1*-like) and for splicing variant 1 (*Mec1*-1). In contrast, *Mec1* splice variant 2 (*Mec1*-2) showed a different pattern with a weak signal of expression in most tissues. Signals were present for expression in storage root, but were barely detectable for fibrous root and leaves. In addition, non-specific amplification was shown for this primer pair.

DV451479 (cassava *ATDI21;* Late embryogenesis abundant protein 3)

Functional annotation of DV451479 in cassava

The cDNA sequence of DV451479 was retrieved from EST database through NCBI. A BLAST search was performed to predict the homologue gene against the non-redundant database on NCBI. The cassava *ATDI21* (*MeATDI21*) mRNA sequence (accession no. JQ807808) was found to be highly similar to cassava-specific unigene DV451479 with the max score 1094 and E value 0.0. The Arabidopsis *ATDI21* gene encoding *ATDI21* (*Arabidopsis thaliana* drought-induced 21 protein) contains one Late embryogenesis abundant protein 3 (Lea3) domain in the transcript. The function of Lea protein is known to be involved with stress response in various plants (Olvera-Carrillo *et al.*, 2011) as well as in cassava (Costa *et al.*, 2011). BLAST searches performed before 2012 returned the sequence of unigene DV451479 to be as similar to late embryogenesis abundant protein 5

(Lea5). After that time, this cassava-specific DV451479 was identified as cassava *ATDI21*. The similarity between DV451479 and *MeATDI21* is shown in Figure 6. The transcript sequence of DV451479 consisted of 294 nucleotide coding sequence which is 100% identical to the transcript of *MeATDI21*. Thus, DV445495 is exactly the cassava-specific *ATDI21* gene.

To examine the predictive function for *ATDI21* in cassava, the coding sequence of DV451479 was BLAST searched against the cassava genome via Phytozome.net. Two significant regions were predicted to be its homologue. The first matched region was located in scaffold01945 (SC01945:558330-559288) and second region in scaffold01551 (SC01551:92609-92764) with the scores 677.5 (E value 0) and 21.6 (E value 8.7e-11), respectively (Fig. 7A). Due to the obviously higher similarity of *ATDI21* gene located in SC01945 compared to a gene located in SC01551, the transcript of cassava-specific *ATDI21* gene in SC01945 was chosen for further study.

Alternative splicing variant prediction in cassava ATDI21

The cassava *ATDI21* gene contains two different introns and its transcription produces two alternative spliced variants, as shown in Figure 7B. The alternative splicing occurred at the 5' UTR in mRNA of the splice variant 1, whereas the variant 2 did not show splicing in the 5'UTR. Both alternative splicing variants share the same 294 coding sequence and contained the late embryogenesis abundant protein 3 (Lea3) domain (Pfam:0342), which is known to be involved in drought stress response based on GO term (Phytozome). Transcripts and coding sequences of both alternative splicing variants

were aligned to review structural mismatch against DV451479 (Fig. 8). Cassava-specific unigene DV451479 was shown to have higher similarity to alternative variant 2, which did not contained splicing in the 5'UTR of the transcript.

Gene expression analysis of DV451479 (MeATDI21) in different tissue types

The expression of the *MeATDI21* gene containing Lea3 domain was examined using RT-PCR with a primer set that amplified the full length coding sequence of the cassava-specific unigene DV451479. The expression level was examined for leaf, fibrous root, nodal root, storage root, and stem of three-month-old greenhouse grown plants. The expression of this gene was found to be non-detectible in leaf tissue, low in fibrous roots, moderate in the stem, and highly expressed in storage root tissues and nodal roots (Fig. 9). Unlike *Mec1*, although the transcript of this gene possibly produces two alternative splicing variants, the splicing occurred in the 5'UTR region. Therefore the coding sequence still maintains the same translational information.

DV446014 (*ENOD40*-like gene)

Sequence similarity analysis of DV446014

As the results of microarray analysis described in Chapter 3, DV446014 was one of the cassava-specific unigenes in cluster 15 showing high expression in the maturation stage of cassava tuberization (SR4). A BLASTN search was performed for DV446014. Unlike DV445495 and DV451479 that returned known putative functions in cassava, the cDNA sequence of DV446014 did not show similarity to any gene or protein in the NCBI database. Furthermore, no open reading frame was found for translated protein in the

DV446014 transcript. However, BLASTN search against the draft Cassava Genome through Phytozome identified five regions in cassava to be similar to the EST sequence of DV446014. The result showed 100% similarity to a gene locus on scaffold 03823 at E value 0.0 with Max score at 870.5 (Fig. 10). In this region, although the EST sequence of DV446014 showed 100% similarity, the transcript information was not available and therefore no putative function was available for any of the five matching regions.

Expression pattern of DV446014 transcript

Because information about the putative function of DV446014 unigene was not available in Cassava Genome or other genome databases, expression of its transcript was investigated by RT-PCR to assess its expression in different tissue and organ types. Total RNA from five different organs – leaf (L), fibrous root (FR), nodal root (ND), stem (ST), and storage root (SR) – of three-month-old plants was obtained and extracted as described in Material and Methods. RT-PCR was optimized at the annealing temperature at 58^0C for 30 cycles to amplify the full length (508 bp) of the cDNA sequence available in the EST database. Expression of DV446014 transcript were detected in all five organs, but clearly showed highest expression in the nodal root with lowest expression seen in the leaf (Fig. 11).

Putative functional identification of DV446014 transcript

Due to the absence of an open reading frame (ORF) in the transcript of DV446014, it was not possible to use a BLASTP search to identify the functional protein motif of this transcript. DV446014 may, therefore, behave as a non-coding RNA gene. The EST

sequence of DV446014 was converted to RNA sequence and a search performed for the RNA family match using the Rfam database (http://rfam.xfam.org/). Rfam database is a collection of RNA families that categorizes the functional class of RNA families as non-coding RNA, primarily RNAs with a conserved RNA secondary structure, including both RNA genes and mRNA cis-regulatory elements (Burge *et al.*, 2012). Based on this search the RNA structure of DV446014 showed only one hit to *ENOD40* (accession no. RF01845) with score of 82.6 on the plus strand and E value 1.7e-16. The *ENOD40* (early nodulin 40) is a dual RNA containing both a short open reading frame (sORF) and two functional RNA domains. It is known that ENOD40 is involved in root nodule organogenesis in legumes (Crespil *et al*., 1994; Campalans *et al*., 2004; Bardou *et al*., 2011).

Sequence analysis of functional domains between MeENOD40 *and DV446014*

The peptide sequence logo of *ENOD40* coding sequence and RNA secondary structure for common leguminous plant is shown in Figure 12 (http://rfam.xfam.org/family/enod40). *ENOD40* gene encoding the short peptide and two conserved domain is known to be activate during nodule organogenesis in legumes. Their putative role has also been studied further in non-leguminous plants such as Arabidopsis (Guzzo *et al.*, 2005). Gultyaev and Roussis (2007) analyzed the conserved short ORF and possible RNA secondary structure of the *ENOD40* transcript through the nucleotide sequence database to identify *ENOD40* homologues in various plant families. Their results suggest that the encoded RNA structure was necessary to determine the common

function of *ENOD40* and that the short peptide might be responsible for the diverse function in plant development (Gultyaev and Roussis, 2007; Bardou *et al*., 2011).

Gultyaev and Roussis (2007) identified the unigene CK643649 (cDNA library submitted by Lopez *et al.*, 2004) as a *ENOD40* homologue in cassava. These workers also identified three domains and their locations of *ENOD40* in the transcript of CK643649. In the present study, this information was used as a reference for cassava *ENOD40* (*MeENOD40*). The peptide sequence of the conserved sORF, including secondary structure domains, were identified in the RNA sequence of CK643649 (Fig. 13). The cDNA sequence of unigene CK643649 was used as the reference to align with cDNA of DV446014 in order to find sORFs and secondary structure domains in DV446014 (Fig. 14). The comparison of RNA sequence with the translated ORF between DV446014 and CK643649 was also performed to investigate the putative functional domains on the transcript (Fig. 15). Two major differences were seen between CK643649 and DV446014. No sORF was found in DV446014 and a tandem repeat of domain 3 in RNA structure was seen to be present in DV446014 (Figs. 14 and 15). The functional domain structure of the DV446014 transcript was demonstrated as shown in Figure 16.

Identification of DV446014 as ENOD40-like gene

Differences in functional domain structure were seen between transcripts of DV446014 and CK643649 (*MeENOD40*) (Fig. 15). In order to study this further, unigenes represented by DV446014 and *MeENOD40* were compared to examine their library information. The cDNA sequences of unigene DV446014 and CK643649 were retrieved

from EST database and then submitted to UniGene database through NCBI (http://www.ncbi.nlm.nih.gov/unigene) to identify unigenes in the same transcribed locus. The result showed the group of unigenes from various EST and/or cDNA libraries that shared the same transcribed locus and contained the same functional domains (UniGene, NCBI). The cassava transcribed locus corresponding to DV446014 was represented by 20 ESTs from six cDNA libraries. Likewise, cassava transcribed locus corresponding to CK643649 was represented by 20 ESTs from five cDNA libraries. UniGene database is composed of six cDNA libraries produced from different tissue sources and genetic backgrounds of cassava (Table 4). Three ESTs from the same transcribed locis of DV446014 and three from CK63649 were chosen from different libraries produced from different sources of tissue.

Three ESTs (FF379705, CK643649, and FG805321) represented the transcribed locus of *MeENOD40* based on the identification of cassava *ENOD40* described by Gultyaev and Roussis (2007). Two unigenes that shared the same transcribed locus with DV446014 were CK646520 and DR084027. All six unigenes from the two transcribed loci were aligned for domain analysis (Fig. 17). The results clearly showed two groups of cassava *ENOD40*. The EST corresponding to the same transcribed locus with CK643649 (*MeENOD40*) showed one sORF with 12 amino acids. In contrast, the transcribed locus containing DV446014 was absent of sORFs. The differences between these two transcribed loci show not only the lack of sORF, but also the non-splicing intron in domain 2 which was clearly shown in DV446014 (Fig. 17). These results indicated that DV446014 might function as cassava *ENOD40*-like gene.

DV446014 may act as a tissue-specific RNA regulator

The region of short coding sequence was analyzed to reveal the feature of conserved domains in *ENOD40* gene (CK642649) and *ENOD40*-like gene (DV446014) in cassava. The presence of a short ORF was obvious in the group of cassava *ENOD40* (FF379705, CK643649, and FG805321), whereas the absence of start and stop codons was shown in the group of cassava *ENOD40*-like transcripts (Fig. 18). The sequence logo of peptides showed two distinct groups determined by the source of tissue used for the construction of cDNA libraries (Table 4). The unigenes related to CK643649 were presented in stem, leaf, and meristem tissue. FG80321 was generated from root tissue of seven-month-old plants (Li *et al*., 2010). The other isoform which did not contain start and stop codons clearly showed that the unigenes in this group derived from root tissues (Fig. 18). Due to the sequence logo showing the conserved protein motif of DV446014 is clearly identified as a cassava *ENOD40*-like gene.

DISCUSSION

The present work aimed to study the putative biological function of three candidate genes determined by microarray analysis to be involved in the development of cassava storage roots (Chapter 3). Focus was mostly on stage SR4, during which the storage root has been formed and is filling with starch, and its comparison to the earlier stages of storage root initiation and early development. Although microarray results allowed ranking of up-regulated genes by fold changes of differential gene expression across the four developmental stages (SR1-SR4), those genes identified as up-regulated needed to be validated for their actual expression in the relevant cassava tissues. To date, several

genome databases have been published for utilization of genome-wide analysis of plant species (Li *et al*., 2010; Jung *et al*., 2014; Thatcher *et al*., 2014). *In silico* functional prediction is relatively accurate for determining the putative function of candidate genes. However, prediction through genome databases alone might result in errors for data interpretation because it deals with very large genome data sets. Hence the integration of multiple BLAST searches across the cassava genome and non-redundant database NCBI, including the UniGene database, are a more efficient way to identify the putative function of candidate genes.

The first candidate gene to be studied was DV445495, a unigene predicted to be a homologue to cassava-specific *Mec1* gene which, based on BLASTN search (NCBI), encodes allergenic-related protein Pt2L4 with 100% identity (Table 2). We can have confidence that the unigene DV445495 is, therefore, the cassava-specific *Mec1* gene. This Pt2L4 protein was previously reported as the enriched protein related to secondary xylem parenchyma in cassava storage roots (de Souza *et al*., 2004; 2006). Two homologue genes of DV445495 (*Mec1*) were predicted on the Cassava Genome. Alternative splicing was shown to form splice variant in homologue 1 (*Mec1* 1). To study further, the full length cDNA sequence corresponding to DV445495 was cloned from cassava genomic DNA of cv. 60444, sequenced, and then aligned to its predicted homologues. The DV445495 transcript was found to be 100% identical to splice variant 1 of *Mec 1* 1 (DV445495 1), which has no intron splicing (Fig. 4). The expression of both homologues including the alternative splicing variants were analyzed by RT-PCR and tissue printing. As expected, high expression of *Mec1* gene appears in storage root, stem,

and nodal root tissues (Figs. 3 and 5). Only low expression was seen in the leaves and fibrous roots. This is the first time that expression of *Mec1* has been reported for the nodal-derived root and lignified stem tissues. This correlates with the report of Beltran *et al*. (2010), who describe transgenic expression in vascular tissue of the stem and storage root when a marker gene was under control of the *Mec1* promoter.

Expression of the splice variant 2 of *Mec1 1* predicted from Cassava Genome was weak and unpredictable when assessed by RT-PCR (Fig. 5). It is possible, therefore, that this splice variant might not be an alternative splicing variant of *Mec1*, but may be due to annotation error on the genome database. Interestingly, presence of *Mec1* expression in the stem and nodal root as confirmed by Northern blot analysis and tissue printing indicated that this *Mec1* gene is not specific to the storage root. Its relatively high expression in both stem and storage root would therefore associate it with secondary growth of the xylem. Both of these organs, and especially the storage root at SR4, are undergoing significant development of the xylem through secondary growth from active cambium (Ch2, Figs. 4 and 5). However, this process is not substantial in fibrous roots or leaves.

The second candidate gene, DV451479 was first predicted as the homologue of the gene encoding late embryogenesis abundant protein 5 (Lea5) in rubber tree (NCBI). To date, BLASTN search predicted DV451479 was 100% identical to cassava *ATDI21* (*MeATDI21*) based on the similarity of cDNA sequence (Fig. 8). *ATDI21* contains the domain encoding late embryogenesis abundant 3 (Lea3) and is known to be involved in

stress response, especially water deprivation (Costa *et al*., 2011). Furthermore the splice variant at 5' UTR was predicted in the first homologous gene of DV451479, meaning that the coding sequence was not being affected by the splicing. The effect of splicing on gene expression needs to be further studied in order to validate the effect of alternative splicing on the biological processes. The expression of this *MeATDI21* was strongly shown in the storage root and nodal root, with low expression in stem and fibrous root and below detectable level in the leaf (Fig. 9). Again, relatively high co-expression of genes involved in stress response was shown in the storage root and in the nodal roots but seen as low expression in stem and fibrous root. This might imply a role for drought stress-like processes in cassava storage root maturation.

DV446014 is the third candidate gene identified to be involved in storage root formation. Unlike the *Mec1* and *MeATDI21* genes, no gene was predicted to be homologous to DV446014 on the NCBI database. However, BLASTN search showed one of five possible regions located on the cassava genome having 100% identical to the cDNA sequence of DV446014 (Phytozome) (Fig. 10). Furthermore, DV446014 cDNA was shown to have a transcribed locus on UniGene database (NCBI). Presence of a homologous sequence to DV446014 found in the cassava genome, including presence of a transcribed locus, indicates that DV446014 is a novel cassava-specific gene. This gene has not been previously identified and its biological function has yet to be annotated for storage root development or other processes. It is noticeable that the transcript sequence of DV446014 contains no ORF or putative regulatory domain. Thus this gene may not be regulatory protein, but might be coding for regulatory RNA. The DV446014 RNA was

shown to be homologous to the transcripts of *ENOD40* RNA family that is known to be involved in early nodulation in leguminous plants (Crespil *et al*., 1994). This result could have significance in advancing knowledge about cassava storage root formation by relating the root tuberization to nodulation mechanisms described in other species.

The cassava *ENOD40* gene (CK643649 unigene) and its structure was predicted and identified by Gultyaev and Roussis *et al*. (2007). Apparently, DV446014 is the unigene located in the different transcribed locus with CK643649 and differs from CK643649 in two structural domains (Fig. 15). DV446014 does not contain sORF and has a tandem repeat at domain 3 of *ENOD40*. The similarity in RNA sequence, but difference in structural domains, is predicted in both cassava *ENOD40* and DV446014, indicating that DV446014 can be confidently predicted as an *ENOD40*-like gene in cassava. The lack of sORF in *ENOD40*-like genes might relate to tissue specific function because the conserved peptide sequence showed differences between the two groups of unigenes derived from tuberous and non-tuberous tissue. These results demonstrate the possible role of *ENOD40*-like genes in cassava storage root formation. Further characterization is needed to identify and annotate the actual function for this *ENOD40*-like gene. Ambiguous results related to the expression of DV446014 were shown by RT-PCR in the present study. According to microarray results, DV446014 was highly expressed in SR4, the stage of maturation of storage root development. High expression of DV446014 is therefore expected to be present in storage root tissue and was confirmed. Expression of this *ENOD40*-like was shown to be low in storage root and stem, including the fibrous roots. Interestingly, however it appears to be highly expressed

in nodal-derived roots prior to radial thickening. Results described in Chapter 2 postulate that the young nodal root is the precursor structure for storage roots. Common expression of putative storage root-associated genes between these two root structures would therefore be logical. Further research using techniques such as qRT-PCR should be undertaken to evaluate more accurately expression of DV446014 RNA in these organ types and ages.

REFERENCES

Barbazuk, WB., Fu, Y. and McGinnis, K. (2008). Genome-wide analyses of alternative splicing in plants: Opportunities and challenges. *Genome Res*. 18:1381-1392.

Bardou, F., Merchan, F., Ariel, F. and Crespi, M. (2011). Dual RNAs in plants. *Biochimie*. 93(11):1950-1954.

Beltran, J., Prias, M., Al-Babili, S., Ladino, Y., Lopez, D., Beyer, P., Chavarriaga, P. and Tohme, J. (2010). Expression pattern conferred by a glutamic acid-rich protein gene promoter in field-grown transgenic cassava (*Manihot esculenta* Crantz). *Planta,* 231, 1413-1424.

Burge, SW., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, EP., Eddy, SR., Gardner, PP. and Bateman, A. (2012). Rfam 11.0: 10 years of RNA families. *Nucleic Acids Research* 1-7.

Butte, A. (2002). The use and analysis of microarray data. *Nature review*. 1:954-960.

Campalans, A., Kondorosi, A. and Crespi, M. (2004). *Enod40,* a short open reading frame-containing mRNA indeces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. *The plant cell,* 16: 1047-1059.

Chen, X., Jorgenson, E. and Cheung, ST. (2009). New tools for functional genomic analysis. *Drug Discov Today*. 14(15-16): 754-760.

Churchill, GA. (2002). Fundamentals of experimental design for cDNA microarrays. *Nature Genetics,* 32, 490-495.

Costa, CNM., Brigida, ABS., Borges, BDN., Neto, MAM., Carvalho, LJCB. and de Souza, CRB. (2011). Levels of *MeLEA3*, a cDNA sequence coding for a atypical late embrogenesis abundant protein in cassava, increase under in vitro salt stress treatment. *Plant Mol Biol Rep*. 29:997-1005.

Crespil, MD., Jurkevitch, E., Poiret, M., d'Aubenton-Carafa, Y.,Petrovics, G., Kondorosi, E. and Kondorosi, A. (1994). *Enod4O*, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. *The EMBO Journal*. 13(21):5099-5112.

De Souza, CRB., Carvalho, LJCB., de Almeida, ERP. and Gander, ES. (2002). Identification of cassava root protein genes. *Plant Foods for Human Nutrition*, 353-363.

De Souza, CRB., de Almeida, ERP., Carvalho, LJCB. and Gander, ES. (2003). Studies toward the identification of transcription factors in cassava storage root. *Brazilian Journal of Plant Physiology,* 15, 167-170.

De Souza, CRB., Carvalho, LJCB. and Cascardo, J. (2004). Comparative gene expression study to identify genes possibly related to storage root formation in cassava. *Protein Pept Lett* 11: 577-582.

De Souza, CRB., Carvalho, LJ., de Almeida, ER. and Gander, ES. (2006). A cDNA sequence coding for a glutamic acid-rich protein is differentially expressed in cassava storage roots. *Protein Pept Lett* 13: 653-657.

Duggan, D., Bittner, M., Chen, Y., Meltzer, P. and Trent, JM. (1999). Expression profiling using cDNA microarrays. *Nature genetics supplement*. 21:10-14.

El-Sharkawy, MA. (2003). Cassava biology and physiology. *Plant Mol Biol* 53: 621-641.

Gassmann, W. (2008). Alternative splicing in plant defense. *Curr Top Microbiol Immunol*. 326:219-33.

Gosti, F., Bertauche, N., Vartanian, N. and Giraudat, J. (1995) Absicisic acid-dependent and -independent regulation of gene expression by progressive drought in Arabidopsis thaliana. *Mol Gen Genet* 246(1):10-8.

Griffiths-Jones, S., Bateman, A., Marshall, M., Ajay, M. and Eddy, SR. (2003). Rfam: an RNA family database. *Nucleic Acids Research*, 31(1):439-441.

Gultyaev, AP. and Roussis, A. (2007). Identification of conserved secondary structures and expansion segments in enod40 RNAs reveals new enod40 homologues in plants. *Nucleic Acids Research*. 35(9): 3144-3152.

Guzzo, F., Portaluppi, P., Romeo Grisi. R., Barone. S., Stefania Zampieri, S., Henk Franssen, H. and Levi, M. (2005). Reduction of cell size induced by enod40 in Arabidopsis thaliana. *Journal of Experimental Botany*. 56(412):507–513.

Hardiman, G. (2004). Microarray platforms-comparisons and contrasts. *Pharmacogenomics,* 5, 487-502.

Jørstad, TS., Langaas, M. and Bones, AM. (2007). Understanding sample size: what determines the required number of microarrays for an experiment? *Trends in Plant Science.* 12, 46-50.

Jung, KH., Kim, SR., Giong, HK., Nguyen, MX., Go, HJ. And An, G. (2014). Genome-wide identification and functional analysis of genes expressed ubiquitously in rice. *Mol Plant.* 7 (10).

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M. and Stamm, S. (2013). Function of alternative splicing. *Gene.* 514 (1):1-30.

Lebot V. (2009). Tropical root and tuber crops: cassava, sweet potato, yams and aroids. In: Crop Production Science in Horticulture, V. 17(eds) Wallingford, UK, CAB Books, CABI.

Li, ZL., Pan, YH., Sun, CB., Dong, HT., Luo, XL., Wang, ZQ, Tang, JL. And Chen, B. (2010). An ordered EST catalogue and gene expression profiles of cassava (Manihot esculenta) at key growth stages. *Plant Mol Biol* 74:573-590.

Lodhi, MA., Ye, GN., Weeden, NF. and Reisch, BI. (1994). A simple and efficient method for DNA extraction from grapevine cultivars, Vitis species and Ampelopsis. *Plant Molecular Biology Reporter*. 12(1):6-13.

Lokko, Y., Anderson, JV., Rudd, S., Raji, A., Horvath, D., Mikel, MA., Kim, R., Liu, L., Hernandez, A., Dixon, AG. and Ingelbrecht, IL. (2007). Characterization of an 18,166 EST dataset for cassava (*Manihot esculenta* Crantz) enriched for drought-responsive genes. *Plant Cell Reports,* 26, 1605-1618.

Lopez, C., Jorge, V., Piegu, B., Mba, C., Cortes, D., Restrepo, S., Soto, M., Laudie, M., Berger, C., Cooke, R., Delseny, M., Tohme, J. and Verdier, V. (2004). A unigene catalogue of 5700 expressed genes in cassava. *Plant Mol. Biol.* 56(4).

Mastrangelo, AM., Marone, D., Laido, G., de Leonards, AM. and de Vita, P. (2012). Alternative splicing: Enhancing ability to cope with stress via transciptome plasticity. *Plant Science*. 185-186 :40-49.

Nguyen, NK., and Willams, ER. (2006). Experiment designs for 2-colour cDNA microarray experiments. *Applied Stochastic Models in business and Industry*. 22:631-638.

Olsen, KM. (2004). SNPs, SSRs and inferences on cassava's origin. *Plant Mol Biol*. 56(4):517-26.

Olvera-Carrillo, Y., Reyes, JL. and Covarrubias, AA. (2011). Late embryogenesis abundant proteins: Versatile players in the plant adaptation to water limiting environments. *Plant Signaling & Behavior*. 64:586-589.

Pluskota, WE., Bradford, KJ., and Nonogaki, H. (2011). Tissue-printing methods for localization of RNA and proteins that control seed dormancy and germination. *Methods Mol Biol*. 773:329-39.

Reddy, ASN. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu Rev Plant Biol*. 58:267-94.

Sakurai, T., Plata, G., Zapata, FR, Seki, M., Salcedo, A., Toyoda, A., Atsushi Ishiwata, A., Tohme, J., Sakaki, Y., Shinozaki, K. and Ishitani, M. (2007). Sequencing analysis of 20,000 full-length cDNA clones from cassava reveals lineage specific expansions in gene families related to stress response. *BMC Plant Biology*, 7:66.

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, TA. and Soreq, H. (2005). Functional of alternative splicing. *Gene.* 344:1-20.

Suarez, MC., Bernal, J., Gutterrez, J., Tohme, J., and Fregene, M. (2000). Developing expressed sequence tags (ESTs) from polymorphic transcript-derived tragments (TDFs) in cassava (*Manihot esculenta* Crantz). *Genome.* 43:62-67.

Subramanian, A., Tamayo, P., Mootha, VK., Mukherjee, S., Ebert, BL., Gillette, MA., Paulovich, A., Pomeroy, SL., Golub, TR., Lander, ES. and Mesirov, JP. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Science U S A,* 102: 15545-15550.

Taylor, N., Gaitan-Solis, E., Moll, T., Trauterman, B., Jones, T., Pranjal, A., Trembley, C., Abernathy, V., Corbin, D., and Fauquet, C. (2012). A high-throughput platform for the production and analysis of transgenic cassava (*Manihot esculenta*). Plants. *Tropical Plant Biol*. 5:127-139.

Thatcher, SR., Zhou, W., Leonard, A., Wang, BB., Beatty, M., Zastrow-Hayes, G., Zhao, x., Baumgarten, A, and Li, B. (2014). Genome-wide analysis of alternative splicing in Zea mays : landscape and genetic regulation. *Plant Cell.* 26(9):3472-87.

Till, BJ., Amos, EK., Kinyua, MG., Bado, S., Jankowicz-Cieslak, J., Huynh, OA., Henry, I., Comai, L., and Lagoda, P. (2013). Mutation induction and reverse-genetics for functional genomics and breeding. Tagung der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreichs 5 – 6.

Zhang, P., Bohl-Zenger, S., Puonti-Kaerlas, J., Potrykus, I. and Gruissem, W. (2003). Two cassava promoters related to vascular expression and storage root formation. *Planta*, 218, 192-203.

Table 1. Primer sequences used for functional analysis of candidate genes and their alternative splicing variants

Candidate cDNAs	Vairant	Primer name	Primer Sequence	Annealing temperature	Functional annotation
DV445495	1-1	Mec1-1F	GAGATTGTAACAGAAGAGGCAGCA	58	Pt2L4
		Mec1-1R	TCTTCTCAGCTTCAACTTCTGC		(*Mec1*)
	1-2	Mec1-2F	AGCTGAAGAAGTGAAGGAGG	58	
		Mec1-2R	TCTTCTCAGCTTCAACTTCTGC		
	2-1	Mec1-like F	AGTCAAGGTTCCAGAGGCA	58	
		Mec1-like R	TTCTTCTGCTTCAGGCTTCTTCTTCTC		
	full length	Mec1-entire F	CTTGGCTGACTATGGCTACTG	55	
		Mec1-entire R	ACCTTGCAGAGCTATCTCATTAC		
DV451495	1-1	MeATDI21 1-1F	GGTGCTGCAGAAGAAAATAG	55	Lea3
		MeATDI21 1-1R	AGAGTGAACACCACCACAGA		(MeATDI21)
	1-2	MeATDI21 1-2F	TGCTCTTCTTGAGAATCCAT	55	
		MeATDI21 1-2R	AGAGTGAACACCACCACAGA		
DV446014	ND*	DV446014-F	AGAATCCATCCTTGGGTCTTC	55	Not detected
		DV446014-R	CGTTTGGCTGAGATTCAGTTG	55	

*ND- Isoform not detected

Table 2. Function annotation of three candidate cassava-specific unigenes predicted from genome databases

Accession No. [a]	Functional annotation[b]	Fold change[c]	Blast hit[d]	Arabidopsis Othologue[e]	No. of regions[f]	variants[b]
DV445495	Manihot esculenta allergenic-related protein Pt2L4 (c1)	16.15	Allergenic-related protein [*Manihot esculenta*]	Unknown	2	3
DV451479	Late embryogenesis abundant protein	11.47	ATDI21 [*Manihot esculenta*]	At4g15910	2	2
DV446014	No functional annotation	10.67	Cassava-specific 746	Unknown	1	ND*

[a]: ESTs database (Lokko et al., 2007) through NCBI corresponding to the array
[b]: Cassava genome database (www.Phytozome.net)
[c]: Fold change calculated as normalization value of overall experiment
[d]: Sequence similarity (http://blast.ncbi.nlm.nih.gov/Blast.cgi)
[e]: Arabidopsis Othologue (http://www.arabidopsis.org/Blast/index.jsp)
[f]: No of regions found in Cassava genome
*Not determined

Table 3. Five cassava-specific cDNAs showing high similarity to DV445495 sequence using BLASTN search against non-redundant database on NCBI

Accession	Description	Gene	Organism	E value	Reference (submission)
EU249994	Allergenic-related protein	ND*	Cassava	0.0	Guo and Zhang (2007)
AY101376	Allergenic-related protein Pt2L4 (c1)	*Mec1*	Cassava	0.0	de Souza *et al.* (2002)
JF710639	Manioc Glu	ND*	Cassava	0.0	Santos *et al*. (2011)
AY217354	Glutamic acid-rich protein (c54)	*c54*	Cassava	1e-25	Zhang *et al*. (2003)
FJ688171	Glutamic acid-rich protein Pt2L4 (c1)	*Mec1*	Cassava	2e-23	de Souza *et al*. (2009)

*ND – Not determined

Table 4. Six unigenes from different cDNA libraries selected for domain analysis of two transcribed loci corresponding to CK643649 (*MeENOD40*) and DV446014

Accession no.	Size (bp)	Tissue type	Library name	Reference
FF379705*	257	Leaf, stem meristem	CASL	Hearne *et al*. 2008
CK643649*	481	Stem	MBra685 cassava lambda zap	Lopez *et al*. 2004
FG805321*	452	Root	Cassava root 210-day-old plants cDNA library	Li *et al*. 2010
CK646520	503	Root	Cassava lambda zap	Lopez *et al*. 2004
DR084027	411	Storage root	Cassava tuber	Emmersen *et al*. 2005
DV446014	508	Mix tissue	CV01-normalized library	Lokko *et al*. 2007

*unigenes represented in the same transcribed locus with CK643649

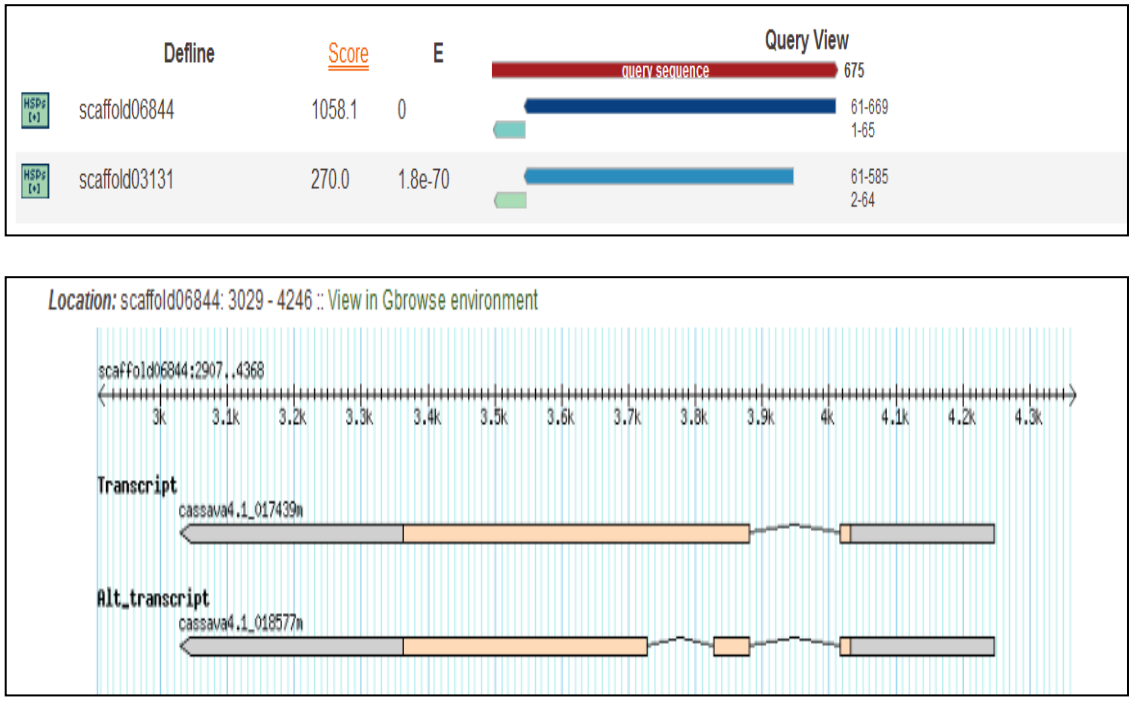

Figure 1. Predicted homologues of cassava-specific unigene DV445495 located on the cassava genome. **A.** Two homologues of unigene DV445495 predicted in scaffold06844 at location 2907-4368 and scaffold03131 at 86914–88042. **B.** Two alternative splicing variants of DV445495-1 on scaffold06844.

(Source : Phytozome.net)

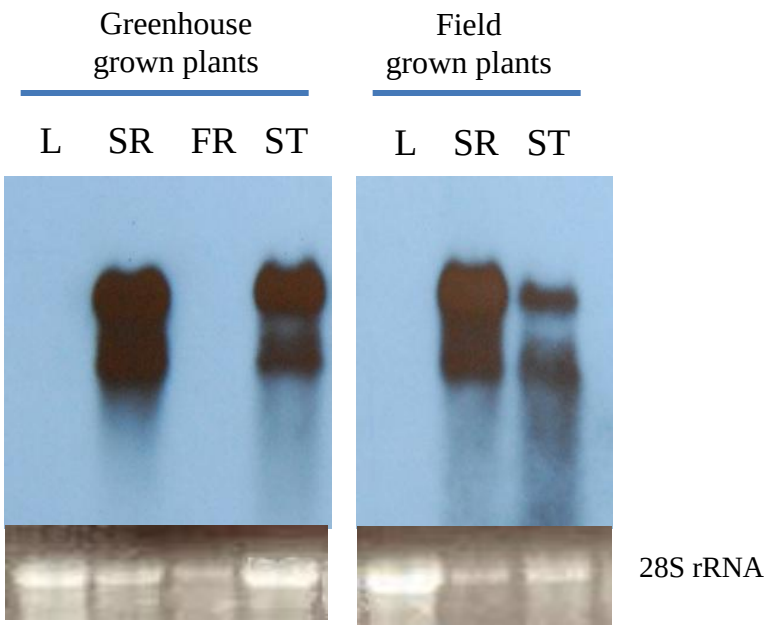

Figure 2. Northern blot analysis showing expression of *Mec1* in tissues of greenhouse and field grown cassava plants. *Mec1* specific DNA probe was generated from sequence derived from DV445495 cDNA. RNA was isolated from L-leaf, SR-storage root, FR-fibrous root, and ST-stem tissues of three-month-old greenhouse grown and six-month-old field grown plants. 28S rRNA band was used for RNA quality and quantity control.

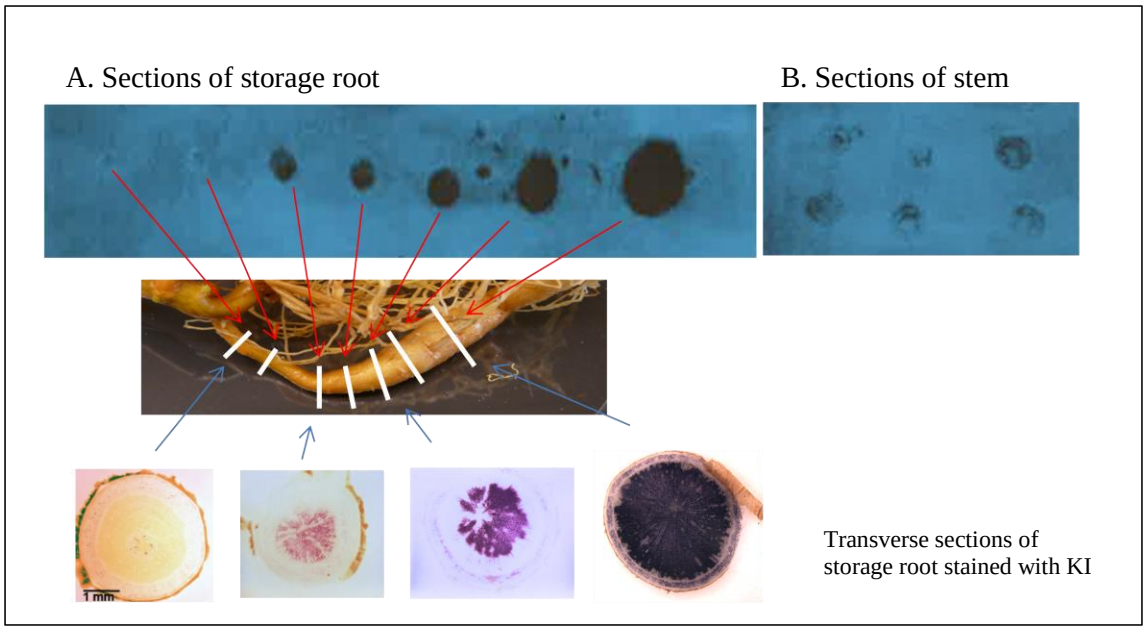

Figure 3. *In situ* hybridization to reveal expression of *Mec1* gene encoding allergenic-related protein Pt2L4 in storage root and stem sections using tissue printing technique of three-month-old greenhouse grown plants. **A.** Tissue cross sections of the storage root (left-to-right) from the neck region into the starch storage regions. Blotting signal increases as the tissues transition from primarily lignified to starch storage. **B.** Transverse sections of semi-woody stem. *Mec1* specific DNA probe was generated from sequence derived from DV445495 cDNA.

Figure 4. Alignment of cDNA sequences showing similarity between DV445495 unigene and its two alternative splicing variants. The highlighted red shaded area in the coding sequence represents the splicing intron in splice variant 2. The highlighted areas at the beginning and the end of reference sequence of DV445495 represented the 5'UTR and 3'UTR region, responsively.

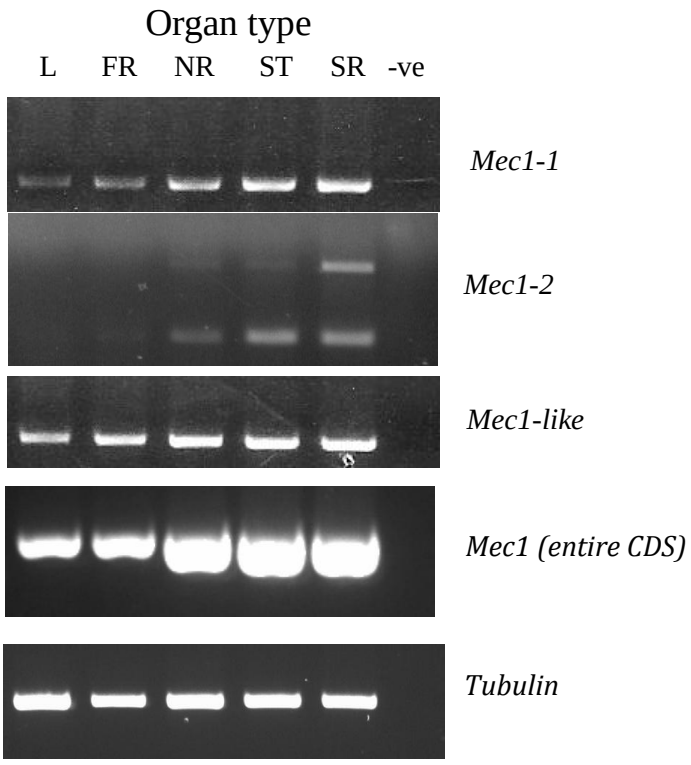

Figure 5. RT-PCR determination of *Mec1* transcript expression comparing alternative splicing variants in tissue from five different organ types: leaf (L), fibrous root (FR), nodal root (NR), stem (ST) and storage root (SR) in three-month-old plants. Tubulin was used as the quality control of cDNA synthesis, -ve = water negative control.

Figure 6. Alignment of *MeATDI*21 cDNA sequence from NCBI and DV451479 showing 100% similarity between coding sequences.

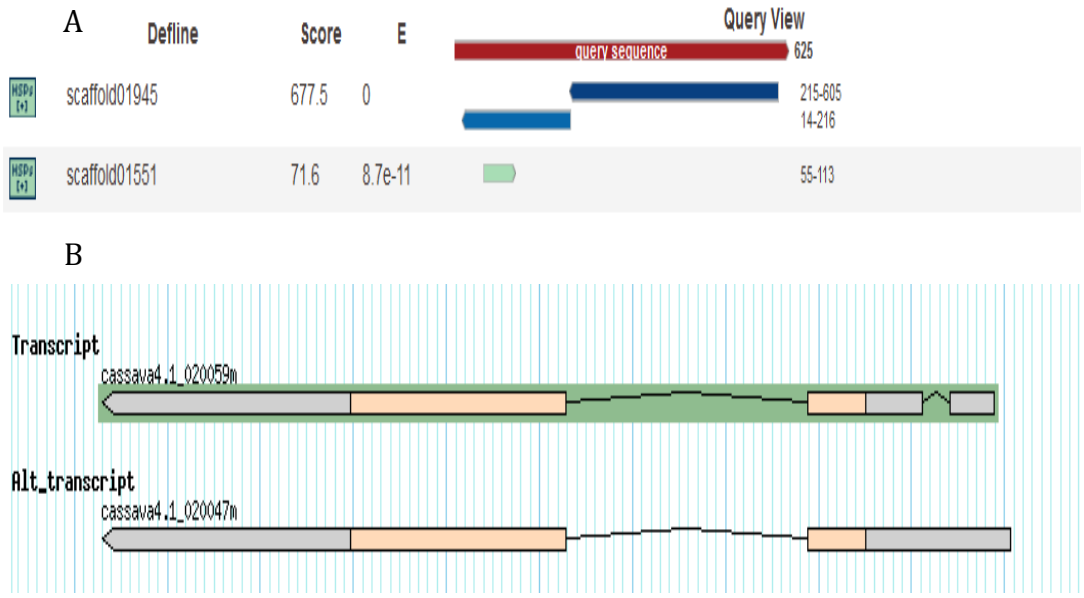

Figure 7. Predictive homologue of cassava-specific unigene DV451479 located on Cassava Genome. **A.** Two homologues matching the cDNA sequence of DV451479 were predicted in Cassava Genome database. **B.** Two alternative splicing variants were found in the region 1 (scaffold01945).

(Source: Phytozome.net)

Figure 8. Alignment of cDNA sequences showing similarity between unigene DV451479 and its alternative splicing variants. Transcripts of both variants showed alternative splicing located in the 5'UTR. The coding sequence of DV451479 remained identical to both splice variants. The highlighted red shaded area in the 5'UTR represents the alternative splicing (DV451479 2) that occurred in transcript variant 1. DV451479 was similar to transcript variance 2 which had no splicing region.

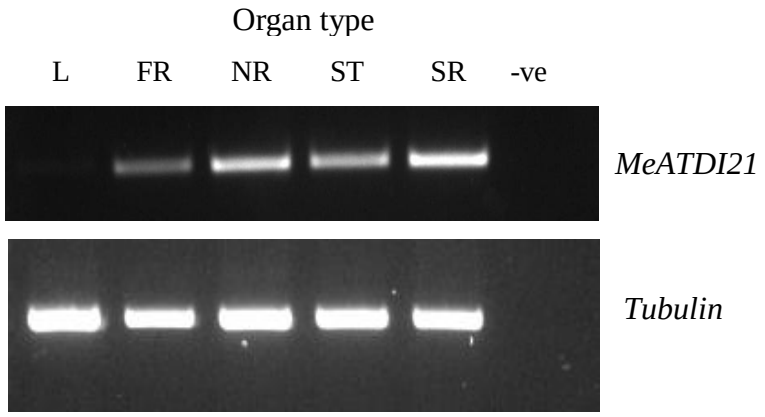

Figure 9. RT-PCR determination of DV451479 (*MeATDI21*) transcript expression comparing tissue from five different organ types: leaf (L), fibrous root (FR), nodal root (NR), stem (ST) and storage root (SR) in three-month-old plants. Tubulin was used as the quality control of cDNA synthesis, -ve = water negative control.

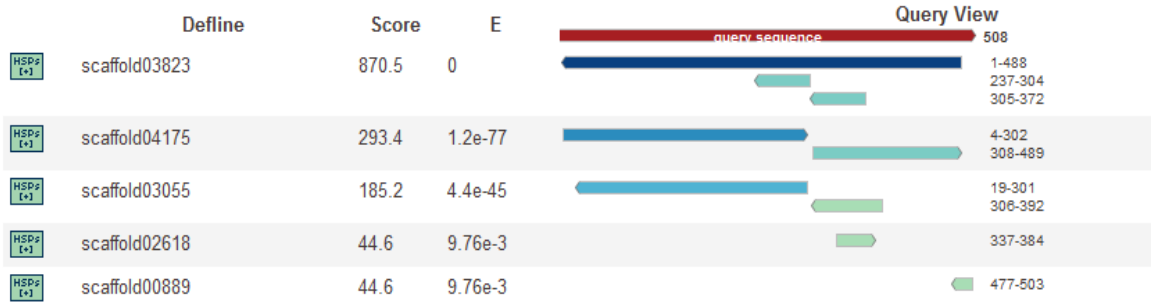

Figure 10. Five predicted homologues of cassava-specific unigene DV446014 on the Cassava Genome. The homologue located on scaffold03823 was shown to have greatest similarity to DV446014 (E value 0).

(Source : Phytozome.net)

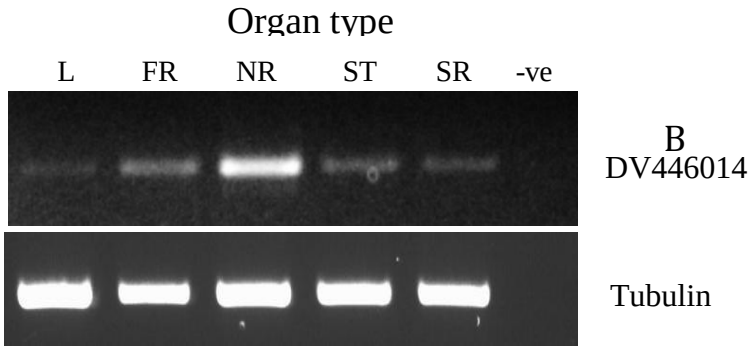

Figure 11. RT-PCR determination of DV446014 transcript expression comparing tissue from five different organ types: leaf (L), fibrous root (FR), nodal root (NR), stem (ST) and storage root (SR) in three-month-old plants. Tubulin was used as the quality control of cDNA synthesis, -ve = water negative control.

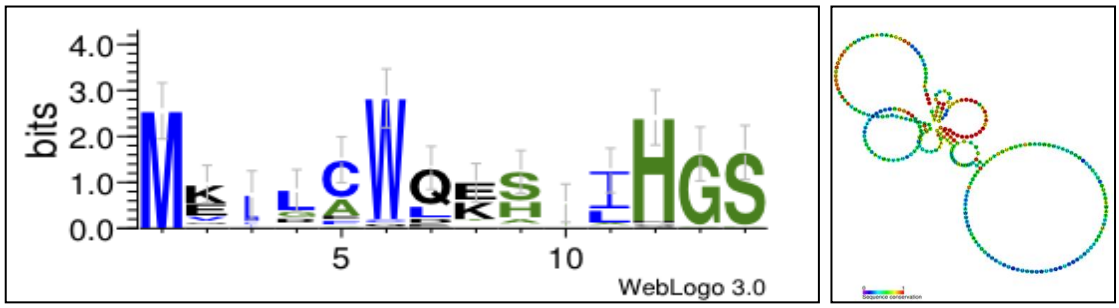

Figure 12. Functional domain and secondary structure of *ENOD40* RNA. **A.** Peptide sequence logo showing the conserved functional domain of short open reading frame (sORF) of *ENOD40* in leguminous plant. **B.** Secondary structure of *ENOD40* RNA, which function as the regulatory structure to interact with binding proteins.

(Source: http://rfam.xfam.org/family/enod40)

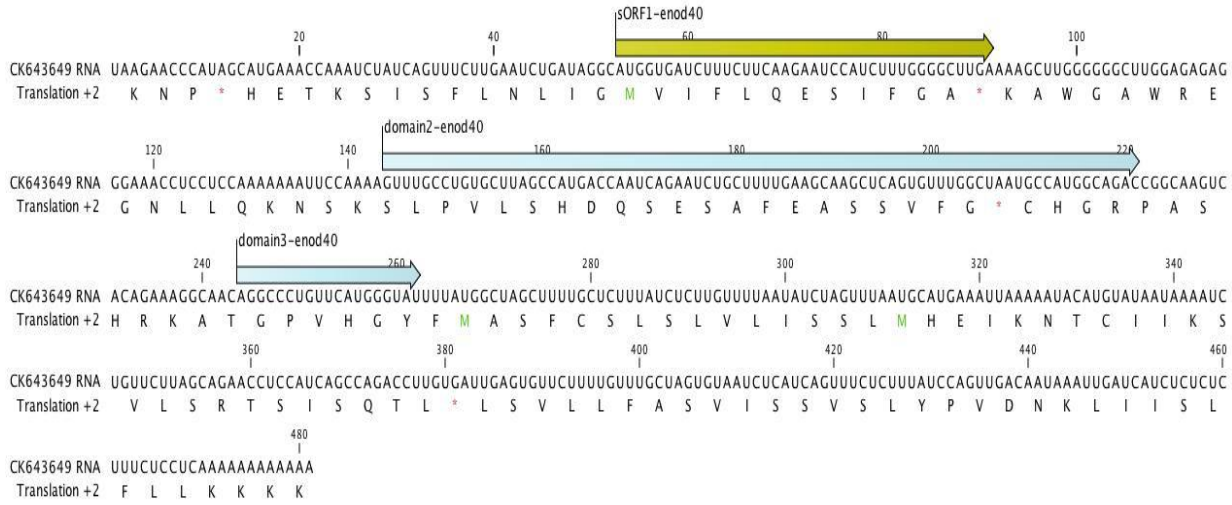

Figure 13. RNA sequence analysis of cassava *ENOD40* (accession no. CK643649) showing three main domains: short open reading frame (sORF), domain 2 and domain 3. The sORF consisted of twelve amino acids. No coding sequence is found in RNA domain 2 and 3.

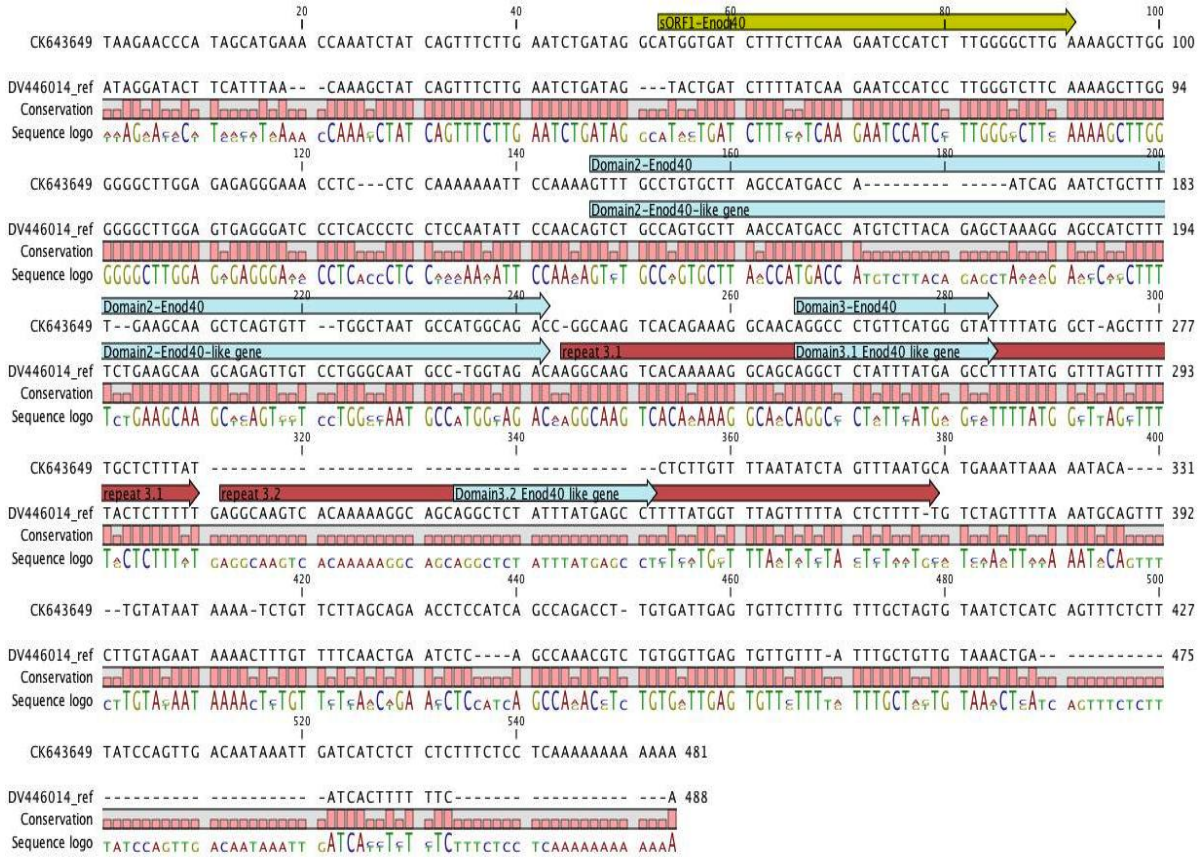

Figure 14. Alignment of CK643649 and DV446014 cDNA to determine different structures between both *ENOD40* homologues.

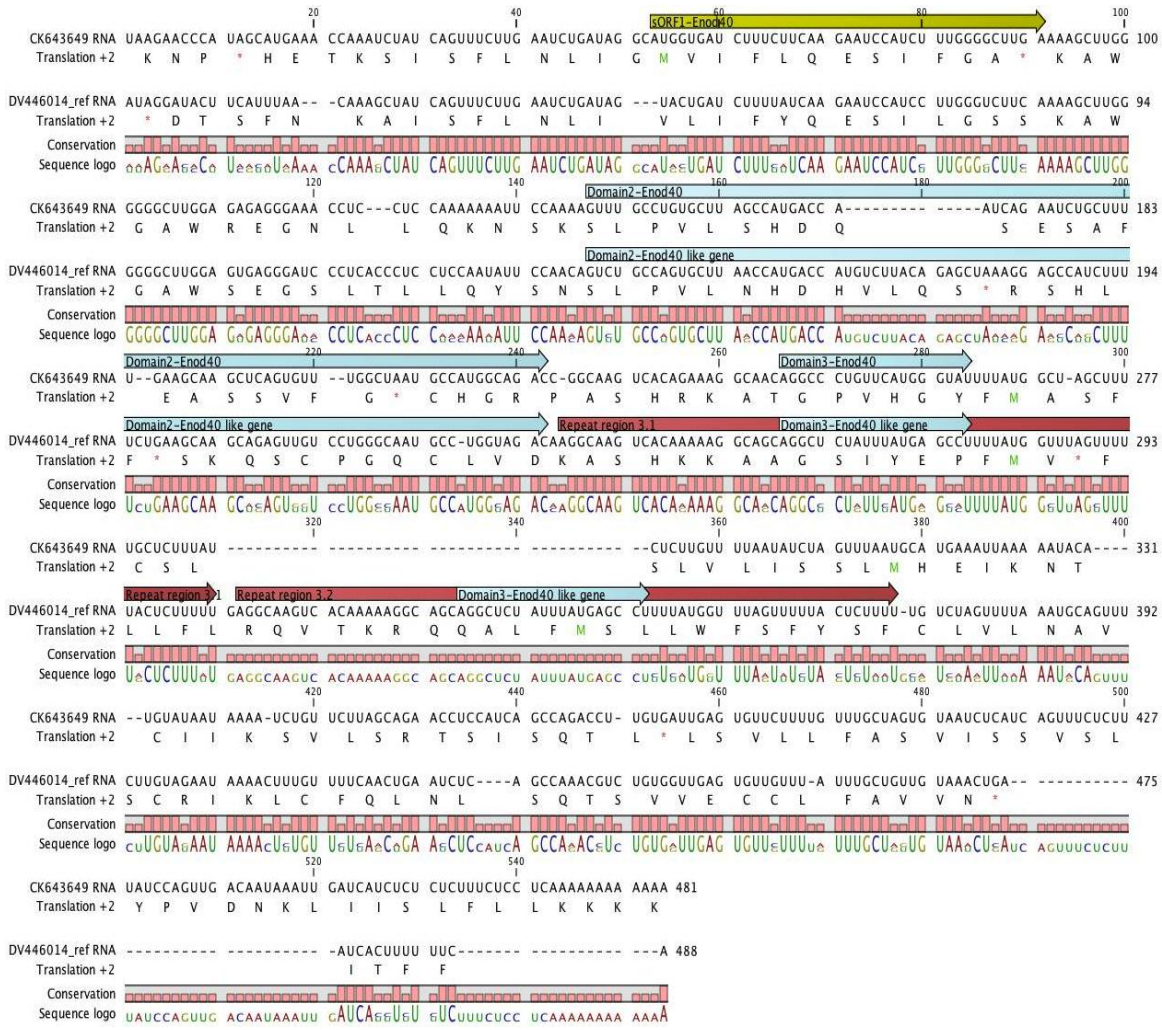

Figure 15. Alignment of RNA sequences between cassava-specific unigene DV446014 and reference *MeENOD40* (CK643649) showing differences of structural RNA domains against *MeENOD40* (CK643649). DV446014 did not show the same sORF as found in reference *MeENOD40*.

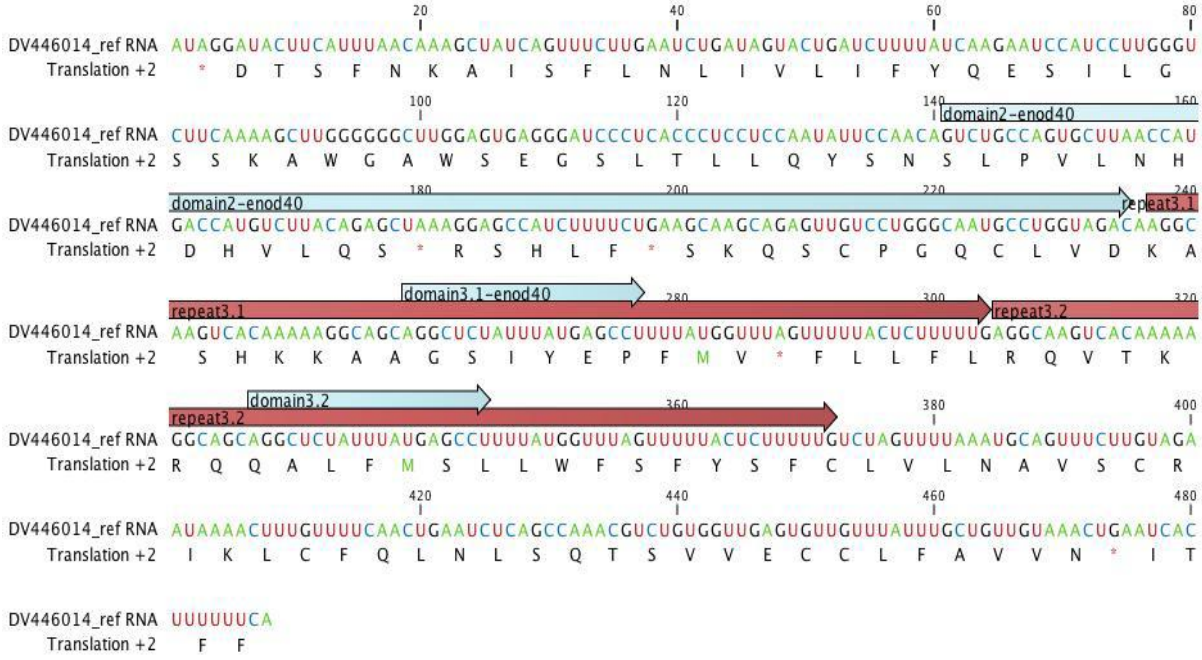

Figure 16. RNA sequence analysis of cassava-specific unigene DV446014 locating the regulatory RNA structure domains (domain 2 and domain 3) of *ENOD40*. DV446014 does not obtain sORF but contains the tandem repeat of domain 3 at the end of RNA sequence.

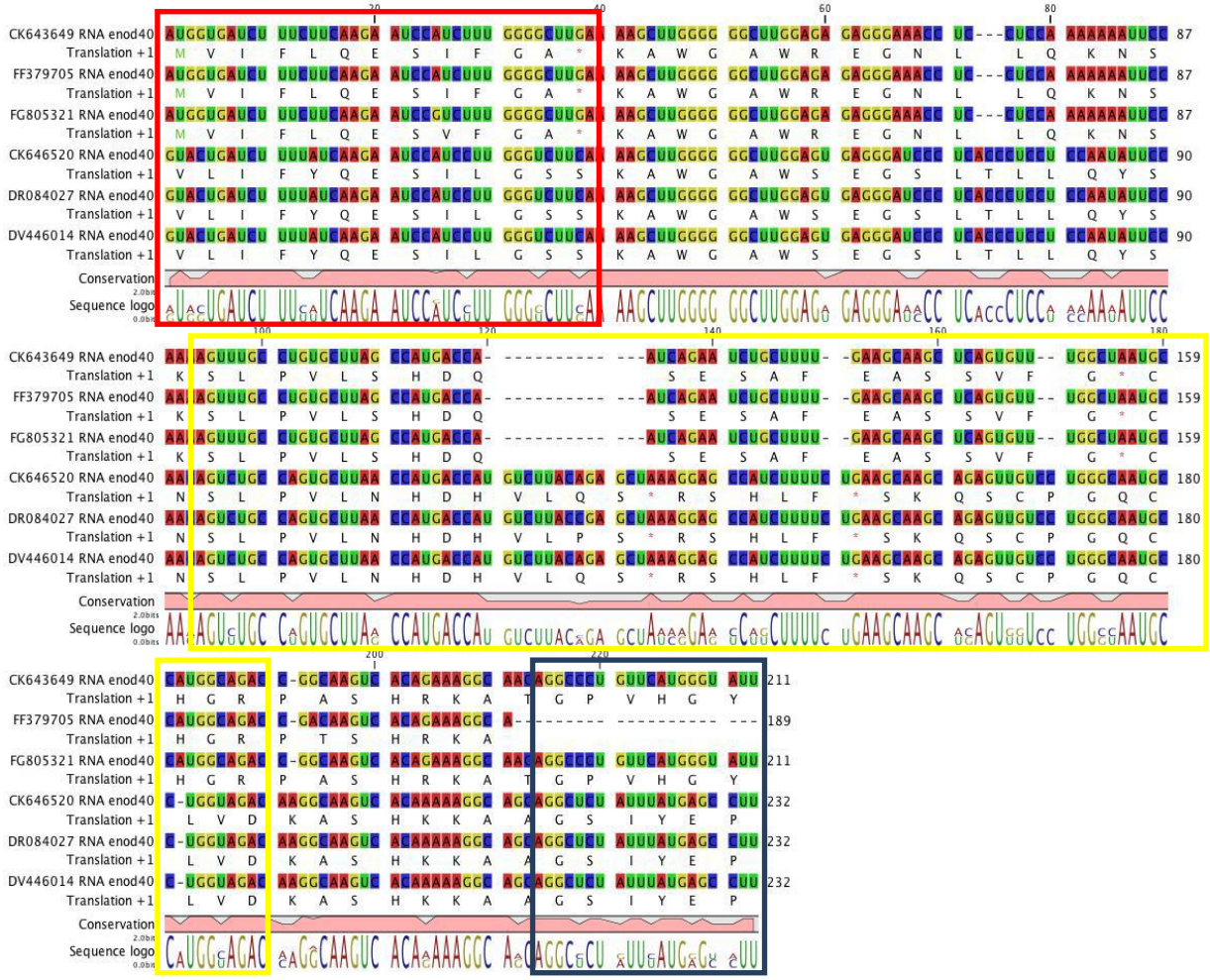

Figure 17. Comparison analysis between six transcripts in the same transcribed locus of DV446014 and CK643649 (*MeENOD40*) showing two different groups of *ENOD40* RNA gene.

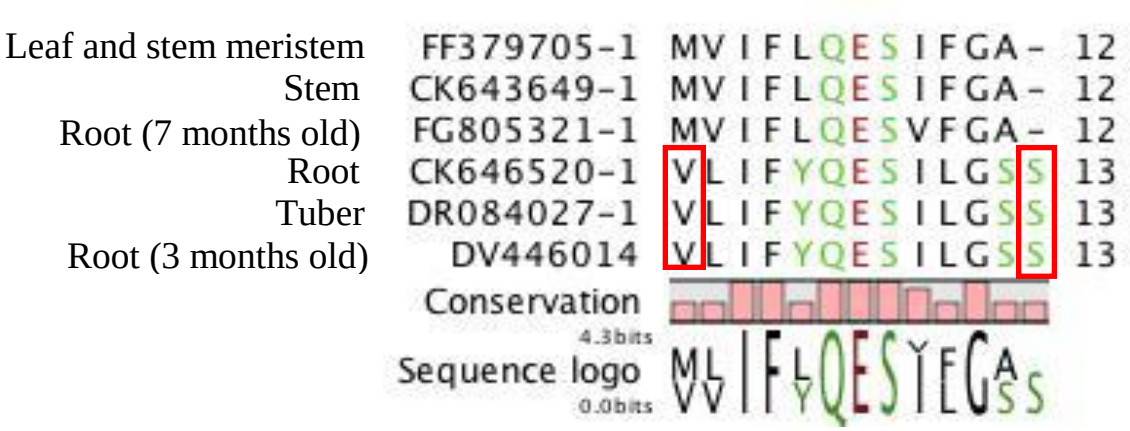

Figure 18. Two groups of peptide sequence logo showing the conserved functional domain of sORF of six cassava-specific unigenes from different cDNA libraries including DV446014. The source of tissue derived cDNA libraries is shown on the left. The positions lacking start and stop codons are shown in red block.

Chapter 5. Conclusions and perspectives

Cassava storage roots are used for food security and are increasingly important as an industrial commodity in developing economies. However, the biology of root tuberization in cassava is poorly understood. The main hypothesis of this research is that the structural modification of a root to become a storage root in cassava is under control of unknown molecular and developmental mechanisms. To elucidate the mechanisms involved with storage root formation in cassava, an anatomical study of cassava root formation was undertaken along with transcriptome analysis. cDNA microarray analysis was performed to investigate the regulatory genes related to the different stages involved in storage root development. The integration of the anatomical study and molecular analysis was shown to be an effective tool to investigate and elucidate unknown mechanisms of cassava tuberization, which is a unique process among the major crop species.

Firstly, the anatomy of root and stem was studied to determine how and where the storage roots develop in comparison with the fibrous root system. The results confirm earlier reports that two types of root develop from stem cuttings of cassava (Lowe *et al*., 1982). Basal roots were seen to be produced from the wounded cut end of the stem while nodal roots are produced from the regions close to the buried axillary buds. While appearing to be similar at the earliest stages of development, these two root types were shown to rapidly diverge in their anatomy, with only the nodal roots undergoing secondary thickening. Importantly, evidence was provided to show that the two root types were derived from different tissue layers within the stem. The nodal-derived root primordia were produced from deep within the stem at the boundary of the xylem and

pith, while basal-derived fibrous roots were initiated from the cambium. Only the nodal root type was seen to develop into storage organs.

This information contradicts previous assumptions that cassava storage roots develop from a subset of the fibrous roots. From data described in these studies we propose that the storage and fibrous roots are fundamentally different organs, originate through different rhizogenic processes, and are committed to their different developmental fates from their earliest stages of initiation. This hypothesis, supported by the data presented in Chapter 2, offers new appreciation of the root tuberization process in cassava and has important implications for the research community. Powerful tools are available to study plant development at the genomic, transcriptomic and proteomic levels and have been applied to investigate storage root formation and development in cassava. Failure to distinguish between the two root types when sampling for such studies results in analysis of a mixed population of organ types, with the resulting data possibly providing misleading results. This has almost certainly been the case for published reports to date (Wechkrajang *et al*., 2006; Mitparasat *at el*., 2011; Yang *et al*., 2011;). The ability to distinguish between fibrous roots and nodal-derived, and to specifically study the storage organ precursor (nodal) roots as an individual organ will greatly facilitate better studies to be performed on the early tuberization processes.

An unexpected result of this study was illustration of xylem differentiation taking place within the storage roots. Depending on age and location along the length of the storage organ, differing proportions of xylem cells newly formed by the cambium

differentiated to become lignified, conducting tracheids and vessels, or to become cellulosic starch storing cells (Ch2, Fig. 5). Understanding how this process occurs, and subsequently the ability to manipulate the underlying control mechanisms, has important implications and could lead to technologies for controlling timing, number, shape and maturation of tuberization in cassava. The study of lignification is hugely important with respect to wood and biomass production and for the development of biofuels and associated industries. The tissue system described here, where location of xylem differentiation is clearly identified, provides a potentially important tool to study the molecular mechanisms controlling this process. Sampling tissues close to the cambium that are undergoing lignified or cellulosic fates would be simple and facilitate transcriptomic, proteomic and other studies. Data generated would be important not only for the improvement of cassava but also for understanding lignification in other plant species.

Further studies can build on the new information described in Chapter 2 and are needed to better understand the specific origin and control of nodal root development. While a cambium-derived origin of the basal roots (Ch2, Fig. 3) can be appreciated, the exact nature of the progenitor cells of the nodal roots at the pith/xylem boundary (Ch2, Fig. 4) is not obvious at this time. Also, it is not known why these roots are produced only from the nodal and not the internodal regions. Once more, knowledge of this type could be of importance for enhancing cassava production.

Finally, the continual connection of storage root secondary xylem with the stem secondary xylem described in these studies (Ch2, Fig. 6) may have implications for genetic manipulation of the storage root. The woody stem is also a starch storage organ in cassava, a trait most likely selected for by farmers because this organ is used as the propagule. These two storage organs are shown though the anatomical studies in Chapter 2 to be connected and to appear almost as one continuous organ. The difficulty in developing storage root specific promoters for cassava biotechnology may be explained by this observation and determine why transgene-promoter fusions engineered to be expressed in the storage root also express in the stem (Zhang *et al*., 2003; Beltran *et al*., 2010).

To predict the putative pathways or subset of genes involved in cassava storage root formation, cDNA microarray analysis was used as a tool to investigate transcriptome profiling related to the developmental stages of cassava tuberization. A data set of microarray analysis across four developmental stages of storage root formation was generated and CLC Main Workbench used for analysis and interpretation of the data. The jasmonic acid biosynthesis pathway was shown to be a significant pathway associated with the early stage of cassava storage root development and may therefore be a key pathway to initiate storage root formation. This result indicated that jasmonic acid might possibly play a role as the trigger in the initiation of storage root formation within the nodal-derived roots, because its signal was highly upregulated in SR2 (initiation stage) but remained relatively low signal at the other developmental stages. Reports of jasmomic acid regulation in potato and other tuberizing species (Palacho and Castel,

1991; Koda *et al*., 1997) make this signaling molecule an important target for future study of storage root formation in cassava. Such investigations could be as simple as studying effects of jasmonic acid application to the shoot and roots are various ages and effects in tuberization. At the molecular level, identification of candidate genes within jasmonic acid pathway is needed for further analysis. Silencing of such genes either by RNAi technology or virus-induced gene silencing (VIGS) would be a good way to test their role in storage root formation.

K-means cluster analysis identified three clusters containing genes involved in the SR2, SR3 and SR4 root developmental stages. The Heatmap analysis was also performed to visually display changes in the expression for 29 genes identified from three clusters produced by k-means analysis. Three best candidate genes involved in cassava storage root formation are DV445495 (*Mec1*), DV451479 (*MeATDI21*), and DV446014 (*ENOD40*-like protein). Up-regulation of *Mec1* encoding allergenic-related protein, Pt2L4, was observed mostly in the SR3 and SR4 stages. This is expected and would correlate with production of secondary xylem in the developing storage root. This also agrees with de Souza *et al*. (2004) who identified this gene from mature storage roots. In the present study, *Mec1* expression was also shown to be up-regulated in the stem and in the nodal roots. This result is not surprising due to the knowledge gained in Chapter 2 describing the similarity and continuity of the xylem tissues of the storage root and the stem and evidence that the nodal root is the precursor of the tuberized root. What is not known at this time is whether *Mec1* is part of the triggering process received by the nodal

root to become a storage organ or if it purely operates downstream to associate with production of secondary xylem parenchyma.

The two genes *MeATDI21* and *ENOD40*-like were shown for the first time in these studies to be candidate genes for involvement in cassava storage root tuberization. *MeADTI21* was shown to be up-regulated at SR3 and SR4 while *ENDO40*-like was up-regulated in SR4 (Ch4, Figs. 9 and 11). Further work using qRT-PCR should now be performed to better characterize these genes and their expression at different stages of tuberization and different tissues within the storage organ and nodal root.

Finally a gene that is a homologue of major latex allergen protein Hev b 4 in rubber was found to be highly up-regulated in SR2, storage root initiation stage. Previously, Souza *et al.* (2008) reported that the *Mec1* gene is a homologue of Hev b 5, while patatin, which is the storage protein in potato, is a homologue of Hev b 7. It appears possible that this protein family is involved in the tuberization process. Further work is required to confirm this hypothesis.

The significance of the studies reported here is a gain of novel knowledge on root formation, including identification of putative genes involved in storage root tuberization in cassava. Information from the anatomy study brings a new appreciation of rhizogenesis in cassava and identifies the origin and specific root type that undergoes tuberization. New candidate genes with possible roles in regulating storage root development in cassava have also been identified, opening multiple new avenues for

further study. It has become apparent that an understanding of the basic biology of a system is required before implementing complex, modern analytical tools to its investigation. Correlating results between the expression profile analysis and anatomy works confirms the intellectual merit gained from these studies and opens the new approaches for the research of the tuberization process in cassava.