2012

Genomic approaches to understand sweetpotato root development in relation to abiotic factors

Julio Solis Sarmiento
Louisiana State University and Agricultural and Mechanical College

GENOMIC APPROACHES TO UNDERSTAND SWEETPOTATO ROOT DEVELOPMENT IN RELATION TO ABIOTIC FACTORS

A Dissertation
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Plant, Environmental & Soil Sciences

by

Julio Solis-Sarmiento
B.S., Universidad Nacional Mayor de San Marcos, Peru, 1996
M.S., Universidad Peruana Cayetano Heredia, Peru, 2009
December 2012

Dedication

I especially dedicate this study to my beloved wife, Jeny Perez

To my mother, Benedicta Sarmiento; my father, Bernardino Solis; my grandmother, Alejandrina Chavez, and all my family, who supported me in every step.

Finally, I dedicate this to my boss and friend, Don Labonte.

Acknowledgments

First and foremost, I extend my deepest gratitude to Dr Don Labonte, a boss that I will always remember. He is really the "sweetpotato guy" since he has the sweetness in his life and in his talking; Don has a unique patience that every worker would like to have in their boss. Don, as an advisor, was nice, and through his mentorship and patience, he taught me the unique world of sweetpotatoes (Louisiana, and in particular Chase and Baton Rouge). He gave me the freedom to do this job and learn more of bioinformatics. Don always encouraged me not to dream to solve a puzzle; instead to be a piece in the path of getting "Beauregard and Evangeline". I just realized that sweetpotato breeders have the easiness to get the sweetness of the yams; you should know Dr. W. Gruneberg and Dr. R. Schafleitner to confirm this claim. I also extend special gratitude to the "Perl guy", Dr. Steven Brandt for instant help in perl programming, Dr. Charles Cavanaugh and Dr. Jaime Prilusky, who provided invaluable support to the scripting and parsing required for decoding the sweetpotato transcriptome. I cannot forget to give my unique appreciation to Dr. Roland Schafleitner for providing me ~ 50% of data that motivated my life and stay at LSU, specifically for sharing the raw sweetpotato sequences from leaves and stem for our analysis. I also want to thank to Dr Niranjan Baisakh and Dr. Arthur Villordon for helping me to develop scientific skills, and all people involved in this work. Finally, I thank Nurit Firon, our collaborative partner in the BARD Project Number US-4015-07, "Molecular and anatomical characterization of sweetpotato storage root formation" that funded this work. I am very thankful to Dr. Simon Chang and Dr. Bin Li that served as committee members and for their support provided during my work.

Table of Contents

List of Tables

List of Figures

Abbreviations

AP2/EREBP	APETALA2/ethylene-responsive element-binding protein
AREB	abscisic acid response element binding factor
BELL	BEL1-like gene
BELL1/BEL1	Arabidopsis homeodomain protein BELL1
bHLH	basic helix-loop-helix domain; basic helix-hoop-helix domain family protein
bp	base pair
CBP	calcium binding protein
CDPK	calcium-dependent protein kinase
dNTP	deoxynucleotidetriphosphate
DAP/DAT	days after (trans) planting
DET	differentially expressed transcripts
DGE	digital gene expression
DNA	deoxyribonucleic acid
DREB	dehydration responsive element-binding factor
EST	expressed sequence tag
FR	fibrous roots
GO	gene ontology
MAAMOVA	microarray analysis of molecular variance
KNOX	KNOTTED1-like homeobox gene
nt/nt	nucleotide (s)
PCR	polymerase chain reaction
qRT-PCR	quantitative reverse transcription PCR

RNA	ribonucleic acid
RT	reverse transcription
TF	transcription factor
SR/DSR	storage root/developing storage root
TSA	transcriptome shotgun assembly

Abstract

Storage root development is the most important physiological process in sweetpotato. Understanding the underlying genetics is the overall objective of this research. A secondary objective is examining the impact of abiotic stress on gene expression and storage root development. A comprehensive analysis of a high-throughput RNA sequencing data (sweetpotato root, stem, and leaf) and public Expressed Sequences Tags was done to generate a genome-wide transcriptome assembly. About 33 million sequences were assembled into 77,663 unigenes; 52,322 (69.55%) of these unigenes matched to a protein sequence from UniprotKB. Data from *de novo* root transcriptome enriched the existing sweetpotato gene index by 37,697 new sequences. Genes which control storage root formation under normal and drought and salt stress conditions were identified through a combination of (quantitative) reverse transcription polymerase chain reaction (qRT-RT-PCR) and microarray analysis.

Global gene expression analysis using a custom sweetpotato oligo array, including ~ 14,000 unigenes derived from the *de novo* transcriptome, identified 1,111 differentially expressed transcripts between fibrous and young storage roots; the majority of these transcripts are involved in basic cellular processes as well as development and differentiation. A set of regulatory genes, such as BEL1-like (BELL), basic helix-loop-helix (bHLH) and HD-Zip homeobox, and signal transduction genes, such as those encoding calcium binding proteins (CBP), calcium dependent protein kinases (CDPK) and genes involved in post-transcriptional modifications (protein phosphatase 2A and a phosphatase associated protein) were up-regulated in early developing storage roots in comparison to fibrous roots.

Thirteen out of 20 selected genes showed altered expression under drought stress and suppressed expression under salt stress. Interestingly, IbBEL1 and IbCRF1 (cytokinin response factor) were upregulated in two-week-old adventitious roots of the plants given drought stress at

planting but were downregulated in storage roots as revealed from sequence-based digital gene expression and microarray analyses. Field and greenhouse studies showed a significant reduction in storage root number and size under drought stress.

Altogether, this study furthers our knowledge in identification of new genes that are crucial in the physiological, metabolic, and molecular events during root organogenesis in sweetpotato under normal conditions and in response to external stimuli – drought and/or salt stress.

Chapter 1. Literature Review

1.1 Introduction

Sweetpotato ranks 7th as the most important food crop in the world when compared to wheat, rice, maize, potato, barley, and cassava, but ranks 12th when compared to all crops in the world (behind Sugar cane, Maize, Wheat, Rice, Potatoes, Cassava, Sugar beet, Soybeans, Oil palm fruit, Barley, and Tomatoes). It is the third most economically important root crop after potato and cassava (FAOSTAT data, 2007). Genetically it is a hexaploid (2n=6x=90) and is member of the Convolvulaceae (Morning Glory) family. The most agronomical important organ in sweetpotato production is a modified root called a storage root. The physiological processes and molecular events responsible for sweetpotato storage root formation are mostly unknown. Initially white fibrous roots develop, and some of these roots subsequently undergo sudden changes in their growth pattern and develop into storage roots. At the mature stage, the roots can be classified according to their morphology into thick storage roots, thick pencil roots, and thin fibrous root (Lowe and Wilson, 1974). The storage roots and the foliage are important economically for both human and animal feeding.

To date, the mechanism that initiates the formation of storage roots is yet to be elucidated. In contrast to sweetpotato, there is a more complete understanding of the physiological and molecular events that trigger the onset of storage organs for other starchy staple foods, such as tubers of potato and tuberous root of cassava (Sarkar, 2008; Sojikul et al., 2010). Significantly more molecular data exist for these crops (Kloosterman et al., 2005; Lokko et al., 2007; Lopez et al., 2004; Sojikul et al., 2010) and the polyploidy of sweetpotato is also a hinderance.

However, anatomical studies revealed that there are some specific changes that occur during the development of a storage root from a non-storage root. First, adventitious roots

develop the primary cambium between the protophloem and protoxylem. Second, a vascular cambium develops and helps to suppress the lignification of the stele. Later, anomalous cambia arise around the central cell and primary xylem elements (primary cambia) and secondary cambia are formed around secondary xylem elements derived from the vascular cambium. Cell division and expansion in these cambia regions lead to rapid thickening of the roots (Wilson and Lowe, 1973). The appearance of anomalous cambia represents the induction phase of storage root formation (Firon et al., 2009; Wilson and Lowe, 1973). By using this information, the timing of storage induction can be assessed during sweetpotato development. In this pursuit, a storage root is defined as young when the storage root is clearly formed and still expanding, and as developing storage root or initiating storage root when the thick roots are with visible pigmentation at early stages of storage root formation because most of them have the anatomical features of roots forming storage root. Most young storage roots are synonymous to initiated storage root that already have developed a complete primary vascular cambium and the anomalous cambium and they are identifiable at 4 weeks after planting (Villordon et al., 2009). Non-storage roots are the white fibrous roots with high lignification of the stele or like primary adventitious roots. Pencil roots are the adventitious non-storage roots with uniform thickening and high lignification; however, they are identifiable only at later stage of root development.

Although several candidate genes have been identified in sweetpotato during the last eight years, none are solely responsible for conversion of simple roots into storage organs (Kim et al., 2005; Kim et al., 2002; Ku et al., 2008; Noh et al., 2010; Tanaka et al., 2008; You et al., 2003).

1.2 Genes involved in sweetpotato root development

Different genetic and environmental factors are involved in root formation in sweetpotato; available data suggest that among the intrinsic factors are genes coding transcriptions factors (TFs) (Ahn et al., 2010; Kim et al., 2002; Ku et al., 2008; Tan7aka et al., 2008), proteins

involved in hormone signaling (McGregor, 2006), enzymes for carbohydrate metabolism, as well as “tuberizing” hormones such as cytokinins (CKs), jasmonic acid (JA) and “non- tuberizing” hormones as auxins (Noh et al., 2010), as well as external factors such as nitrogen supply and soil humidity play a role (Villordon A., personal communication). Most of the genes have been found differentially expressed not only in storage and fibrous roots, but also in all vegetative tissues, such as stem, leaves and flowers. Although, the exact mechanism by which these factors induce storage root formation in sweetpotato is largely unknown, the triggering events due to TFs are in the anomalous cambium region that promotes cell proliferation and thus thickening of the thin roots (Kim et al., 2002; Ku et al., 2008; Noh et al., 2010; Tanaka et al., 2008). The majority of genes identified currently are members of the homeobox family. Homeobox genes are a large family of genes coding transcription factors that have been identified in animals, plants, fungi, and yeast (Burglin, 1994), and contain a DNA binding domain known as a homeodomain (HD). The homeobox proteins can be classified into two large groups or superclasses: The HD non-TALE members, typical HD proteins, that contain the “typical” HD of 60 amino acids (aa), and the HD TALE members that have an “atypical” HD of 63 aa, with an extra three residues (Three Amino acid Loop Extension) in the homeodomain (Burglin, 1994; Burglin, 1997). Both TALE and non-TALE proteins have been found in all major eukaryotic lineages including plants. However, based on sequence similarities within the HD and in the flanking regions, homeodomain proteins were initially grouped into five classes (Kerstetter et al., 1994), then into seven classes (Bharathan et al.1997), and now 14 homeobox gene families have been recognized in plants (Mukherjee et al. 2009). Only two HD TALE gene families are known in plants, KNOX (knotted1-like homeobox) and BELL (BEL-Like) genes (Burglin, 1997; Hake et al., 2004). In *Arabidopsis*, there are four different class I KNOX genes named STM, KNAT1/BP (knotted-like proteins from *Arabidopsis thaliana*/BREVIPEDICELLUS), KNAT2,

and KNAT6, all of which are key factors for maintenance of meristematic tissue in the shoot and responsible for the formation of vegetative organs and also reproductive organs (Scofield et al., 2008).

One sub-class of KNOX proteins named the class I KNOX (KNOXI) is particularly interesting since genes for many of them have been isolated during the development of storage roots (Tanaka et al., 2008); Tanaka et al. (2008) found three KNOXI genes, named Ibkn1, Ibkn2 and Ibkn3, expressed differentially in the storage roots. The proteins encoded by these genes share high similarity with *Arabidopsis* STM (Ibkn1) and KNAT1/BP (Ibkn2, Ibkn3) proteins. Expression of Ibkn1 and Ibkn2 was upregulated in developing and mature storage roots compared with fibrous roots. However, they found that only Ibkn1 and Ibkn2, but not Ibkn3, share a similar pattern of expression among several cultivars.

The interaction of some KNOX and BELL members, to form a functional heterodimer, appears to be crucial for targeting KNOX proteins to specific genes during development; evidence supporting this has been shown in *Arabidopsis* (Bellaoui et al 2001), maize (Smith, 2002), and potato (Banerjee et al., 2006; Chen et al., 2003). KNOX proteins from *Arabidopsis thaliana* (KNATs) have been shown to require the heterodimerization with a member of the BELL family, BEL1 protein. Although the interaction of each KNOX is selective for binding a subset of BELL, not all KNOX proteins have a BELL partner, meaning that other KNOX proteins may form functional homodimers by themselves or between other KNOX proteins. Moreover, it appears that the recruitment of some BELL partners by KNOX proteins allows a spatial and temporal role to some widely distributed KNOX products. Hence, the case of potato is a good example. At least seven BEL1-like members are recognized in potato, and expressed differentially not only during tuber formation but also in other tissues (Chen et al., 2003). Interestingly, the *POTH*1 gene, a potato KNOX tuberization factor, requires a specific BELL,

StBEL5, for inducing the tuberization (Chen, 2004). Surprisingly, the authors found that the transcript of BEL5 accumulates in the stolon tips under short-day conditions, which is the site of tuber induction, after being delivered through the phloem and adding an unexpected new mechanism of gene expression. The remaining BEL1-like genes from potato probably modulate and target the action of POTH1 and other KNOX genes at different stages of development. Thus, the identification of potential BELL partners is likely in sweetpotato and requires further investigation.

Another interesting group of genes named MADS-box genes are widely distributed in all organisms including yeasts, insects, amphibians and mammals, and also in plants. The product of these genes are transcription factors, which have a conserved MADS domain [the letters refers to the four originally identified members: MCM1, AGAMOUS (AG), DEFA (DEFICIENS) and SRF (serum response factor)](Shore and Sharrocks, 1995). In flowering plants most of the MADS-box family genes are involved in regulating the development of floral organs (Riechmann and Meyerowitz, 1997). In potato and sweetpotato the expression of similar genes have been found in other vegetative tissues, such as stem and leaf (Carmona et al., 1998; Kim et al., 2002). Other MADS-box members have been found in sweetpotato storage root forming tissues (Kim et al., 2005; Kim et al., 2002; Noh et al., 2010). Recently, *IbMADS1* was found mostly expressed at early stages of the storage root formation (Ku et al., 2008); interestingly, this gene is induced by jasmonates and cytokinins, two known tuberization-related hormones. The pattern of expression of *IbMADS1* was located around meristematic cells within the stele and in lateral root primordial. Other MADS-box genes associated with initiation of tuber formation in sweetpotato are *IbMADS3* and *IbMADS4* (Kim et al., 2002). However, apart from root, they are expressed in other vegetative tissues, suggesting that they may have roles in the vegetative growth. Very recently, a MADS-box gene, *SDR1*, has been found in sweetpotato, the expression

of which was seen to increase in parallel to the increase of the auxin IAA, and considered a key triggering factor at early stages of the storage root development (Noh et al, 2010).

1.3 Candidate genes associated with lignin metabolism

In sweetpotato, the occurrence of lignification is related to reduced storage root and an increase in fibrous roots and pencil roots. Previous works showed greater lignification has been associated with pencil roots and fibrous root formation and a reduction in storage root formation (Belehu, 2003; Firon et al., 2009; Togari, 1950; Wilson and Lowe, 1973). Environmental factors such as a long photoperiod, nitrogen supply, water saturation of the soil, as well as genetic factors are linked in the lignification process (Belehu, 2003). A recent result correlates the exchange of oxygen with lignification and storage root formation (Eguchi and Yoshida, 2007), where hypoxia is shown to enhance the level of lignification. Therefore, lignification is a very important developmental process in sweetpotato and warrants further study.

Noteworthy is the BP (BREVIPEDICELLUS) gene associated with genes involved in lignin metabolism and secondary cell wall modifications. In sweetpotato, three KNOX I genes, Ibkn1, Ibkn2 and Ibkn3, were found preferentially upregulated in storage roots (Tanaka et al., 2008), a tissue which is poorly lignified; Ibkn2 and Ibkn3 are homologous to BP, while Ibkn1 is homologous to STM gene. For instance, three BP orthologous genes, Populus ARK2 (ARBOR-KNOX2), the peach ortholog KNOPE1, and the poplar KNAP2 genes, along with *Arabidopsis* BP, are all involved in secondary growth and lignin metabolism during xylogenesis processes in normal growth. Overexpression of any of these genes in transgenic plants down-regulate the expression of genes for lignin and cellulose synthesis (Du et al., 2009; Hertzberg et al., 2001; Mele et al., 2003; Testone et al., 2008).

The BP as well as KNOPE1 proteins have been demonstrated to interact with the BELL1 member from *Arabidopsis* (Mele et al., 2003; Testone et al., 2008), which is relevant since not

all KNOX genes may require a BELL partner to function. Both BP and KNOPE1 participate in the architecture of organs, the inflorescence and leaves, respectively by affecting the lignin metabolism and cell wall synthesis. BP in contrast to most MYB TFs, have been found to be a negative regulator of lignin formation (Mele et al., 2003). At least two different BP orthologous genes from poplar also appear to control the level of lignification in trees (Hertzberg et al. 2001), the afore-mentioned KNAP2 and ARK2 genes. ARK2 act as a negative regulator of lignin formation and cell wall formation since many genes encoding lignin biosynthetic enzymes, such as 4CL, C4H, C3CH, F5H, CAD and laccase, and genes required for the cellulose synthesis are all downregulated in transgenic tree plants overexpressing ARK2 (Du et al., 2009), in a similar way to the BP gene in *Arabidopsis* (Mele et al., 2003).

The relation between MADs-box genes and lignin formation is not clear, but there are two examples in *Arabidopsis* (Liljegren et al., 2000) and maize (Guillaumie et al., 2008). Two SHP genes (Shatterproof MAD-box) are present in *Arabidopsis* and required for lignification in siliques, whereas a maize SHP ortholog, ZmZAG5, has been found de-regulated in mutant plants of maize (bm3 plants); the outcome is the deficiency in COMT (caffeic acid *O*-methyl transferase), an enzyme required in an intermediate step of lignin synthesis (Guillaumie et al., 2008) . All these studies reveal the importance of KNOX genes for key morphogenetic processes during organ formation in plants.

How the three sweetpotato KNOX genes together with MADS-Box genes modulate the lignin synthesis and as whole the plant architecture is unknown. Post-transcriptional gene silencing mediated by microRNAs appears to be involved in lignin metabolism. For example, the microRNAs miR397, miR408 and miR857 are known to target the silencing of members of the laccase gene family in several plants (Abdel-Ghany and Pilon, 2008; Bonnet et al., 2004; Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004). Since one assigned role of laccase in plants is

in the last steps of lignin formation, it is possible that microRNAs may be involved in storage root formation.

1.4 Candidate genes in hormone signaling

The main components in auxin signaling pathways include a TF, ARF (Auxin response factor), a transcriptional repressor, Aux/IAA (auxin/indole-3-acetic acid), and an auxin receptor, F-box protein TIR1 (Transport inhibitor response 1). ARF and Aux/IAA exist as large gene families in plants. There are 22 ARF genes in *Arabidopsis* (Paponov et al., 2008) and 25 ARF genes in rice, which encode their respective proteins (i.e. ARF1, ARF2,…, ARF22; Wang et al., 2007); Aux/IAA genes are found encoded in the *Arabidopsis* and rice genomes and number 29 and 31, respectively. The expression of these factors occurs in specifics tissues and they have multiple functions. Both ARFs and Aux/IAA form a complex. Aux/IAA must be degraded to activate auxin response mediated by ARFs (Cohen and Gray, 2006). TIR1 protein is part of the ubiquitin ligase complex SCF, and upon binding to the hormone auxin, promotes the turnover of Aux/IAA. Three proteins are closely related to TIR1, the AFB proteins (Auxin Signaling F-Box) AFB1, AFB2, and AFB3, which acts in a similar way to TIR1. All contribute to auxin responsiveness in *Arabidopsis* (Dharmasiri et al., 2005). Thus, the repertoire of auxin receptors is encoded by a few genes in most plants, and interesting data may arise in case the studies are linked to miRNAs presence and their potential ARF targets during sweetpotato growth.

Aux/IAA proteins can form homodimers with other Aux/IAA members and heterodimers with ARF members. Likewise, ARF proteins also form homodimers with other ARF members for binding to their promoter binding element, the auxin response element (AuxRE, consensus TGTCTC), located in their targets, and thus activating the transcription of downstream genes. Aux/IAA proteins, when forming complexes with ARF, avoid the formation of ARF dimmers required for auxin action. ARFs can be activators, when they have a Glutamine-rich domain

(ARF5, ARF6, ARF7, ARF8, and ARF19), or repressors, by presence of a proline- and/or serine-rich region (ARF1, ARF2, ARF3, ARF4, and ARF9) (Tiwari et al., 2003). As stated before, all Aux/IAA proteins are transcriptional repressors, and have a short life which depends on the auxin concentration. The increase of auxin levels promotes the degradation of Aux/IAA proteins, which allow ARF proteins to form homodimers. This in turn leads to the activation of the transcription of auxin response genes.

Two recent works on sweetpotato, strongly suggested that the initiation of storage roots in sweetpotato involved interplay of *IbMADS1*, jasmonic acid (JA), and cytokinin (Ku et al., 2008; Wang et al., 2005). Certainly, combinations of cytokinin BA (6-benzylaminopurine) and JA induce the thickening of roots in vitro. Evidence from many studies reveals an interplay of KNOX genes with the metabolism of hormones in *Arabidopsis* (Jasinski et al., 2005), potato (Rosin et al. 2003) and sweetpotato (Tanaka et al, 2008), in which KNOX activity influences positively the synthesis of cytokinin. In *Arabidopsis*, the *IPT7* gene, coding the first enzyme for CK biosynthesis, is upregulated and targeted redundantly by two KNOX genes, *BP* and *STM*; similarly, proven targets of BP proteins in potato, *Arabidopsis*, tobacco, and tomato are genes for synthesis of giberellins (GAs). The action of repressing the accumulation of GA is by enhancing their catabolism, promoting expression of GA2 oxidases to catalyze the inactivation of active GAs to inactive forms, or by down-regulating key genes for GA synthesis (GA20 oxidases) (Jasinski et al., 2005; Thomas and Hedden, 2006) . The action of enhancing the expression of GA2 oxidase is downstream of the inductive action of KNOX on CK synthesis. GA action in part is dependent on MYB TFs. Two MYB TFs are downregulated in fibrous roots in sweetpotato (Desai, 2008). The expression of the three STM and KNAT1-like genes in sweetpotato roots also followed the pattern of distribution of endogenous trans-zeatin riboside (t-ZR) (Tanaka et al, 2008). Recent work (Nagata and Saitou, 2009) supports the role of sucrose

and cytokinin in the thickening of developing storage roots by promoting the activation of cell division D3 cyclin genes (CycD3). It was also shown that both sucrose and t-ZR acts synergistically to trigger storage root formation (Eguchi and Yoshida, 2008). In potato, the action of POTH1 and StBEL5 is mediated by regulating GA and cytokinin levels, two antagonistic hormones. The mechanism of action was shown by binding to the ga20 oxidase1 protein (ga20ox1), a key enzyme in the GA biosynthetic pathway (Chen et al. 2004), and in transgenic plants overexpression of either partner alone increased tuber yields by lowering gibberellin (GA) levels and increasing cytokinins. Expression of BP from *Arabidopsis* and peach is upregulated with external application of cytokinin (CK) and in the positive feedback enhancement of the synthesis of CK (Testone et al, 2008). A member of Calcium-dependent protein kinase (CDPK) and key enzymes for auxin-signaling have been found only upregulated at initiation of formation of storage roots and tuber in cassava (Sojikul et al., 2010) and potato (Gargantini et al., 2009; Raices et al., 2001). CDPK may have an important role as it relates GA-signaling during starchy organ formation in both crops.

Many genes required for the transduction of auxin have been observed during the onset of storage root formation. For example, Desai et al. (2008) showed four genes of the auxin response pathway genes that were differentially down- or up- regulated in fibrous and storage roots. Similar work using microarrays conducted by McGregor et al (2006) also stressed that the role of auxin in sweetpotato storage root formation needed to be studied in detail.

The overall objectives of the present study were to better understand the genetic mechanism underlying storage root development in sweetpotato and how abiotic stresses, such as drought and salinity effect gene expression in storage roots. This research was predicated on developing a genome-wide transcriptome of sweetpotato. Specific objectives of the present study were:

1.) Examine the impact of drought and salt stress on the development of storage roots; 2.) Characterize the expression of genes associated with storage root formation during: a.) the normal course of storage root development, b.) the imposition of drought stress and salt stress; 3.) To build consensus unigenes using sweetpotato transcriptome data and to identify novel candidate genes associated with storage root development by comparing gene expression of storage and non-storage roots; and 4.) To build a custom sweetpotato oligo-array (SPOArrayv1 = Sweetpotato Agilent Oligonucleotide Microarray) to study global gene expression profile in sweetpotato.

1.5 Literature cited

Abdel-Ghany S.E., Pilon M. (2008) MicroRNA-mediated systemic down-regulation of copper protein expression in response to low copper availability in arabidopsis. Journal of Biological Chemistry 283:15932-15945.

Ahn Y.O., Kim S.H., Kim C.Y., Lee J.S., Kwak S.S., Lee H.S. (2010) Exogenous sucrose utilization and starch biosynthesis among sweet potato cultivars. Carbohydr Res 345:55-60.

Banerjee A.K., Chatterjee M., Yu Y., Suh S.G., Miller W.A., Hannapel D.J. (2006) Dynamics of a mobile RNA of potato involved in a long-distance signaling pathway. Plant Cell 18:3443-57.

Belehu T. (2003) Agronomical and physiological factors affecting growth, development and yield of sweet potato in Ethiopia, Department of Plant Production and Soil Science, University of Pretoria, Pretoria.

Bonnet E., Wuyts J., Rouze P., Van de Peer Y. (2004) Detection of 91 potential in plant conserved plant microRNAs in Arabidopsis thaliana and Oryza sativa identifies important target genes. Proceedings of the National Academy of Sciences of the United States of America 101:11511-11516.

Burglin T.R. (1994) A Caenorhabditis elegans prospero homologue defines a novel domain. Trends Biochem Sci 19:70-1.

Burglin T.R. (1997) Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. Nucleic Acids Research 25:4173-4180.

Carmona M.J., Ortega N., Garcia-Maroto F. (1998) Isolation and molecular characterization of a new vegetative MADS-box gene from Solanum tuberosum L. Planta 207:181-188.

Chen H., Rosin F.M., Prat S., Hannapel D.J. (2003) Interacting transcription factors from the three-amino acid loop extension superclass regulate tuber formation. Plant Physiology 132:1391-1404.

Chen X.M. (2004) A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science 303:2022-2025.

Cohen J., Gray W. (2006) Auxin metabolism and signaling, in: P. Hedden and S. G. Thomas (Eds.), Plant Hormone signaling, Blackwell publishing. pp. 37-59.

Desai D.P. (2008) Understanding the genetic basis of storage root formation along with starch and betacarotene biosynthesis and their inter-relation in Sweetpotato (*Ipomoea batatas* LAM.). Department of Applied Plant Sciences and Plant Biotechnology, University of Natural Resources and Life Sciences, Vienna. pp. 144.

Dharmasiri N., Dharmasiri S., Weijers D., Lechner E., Yamada M., Hobbie L., Ehrismann J.S., Jurgens G., Estelle M. (2005) Plant development is regulated by a family of auxin receptor F box proteins. Developmental Cell 9:109-119.

Du J., Mansfield S.D., Groover A.T. (2009) The Populus homeobox gene ARBORKNOX2 regulates cell differentiation during secondary growth. Plant Journal 60:1000-1014.

Eguchi T., Yoshida S. (2007) Effects of Gas Exchange Inhibition and Hypoxia on Tuberous Root Morphogenesis in Sweetpotato (*Ipomoea batatas* (L.) Lam.). Environment Control in Biology 45:103-111.

Eguchi T., Yoshida S. (2008) Effects of application of sucrose and cytokinin to roots on the formation of tuberous roots in sweetpotato (*Ipomoea batatas* (L.) Lam.). . Plant Root 2:7-13.

Firon N., LaBonte D.R., Villordon A.Q., McGregor C., Kfir Y., Pressma E. (2009) Botany and Physiology: Storage Root Formation and Development, in: G. T. G. Loebenstein (Ed.), The Sweetpotato, Springer Science+Business Media B.V. pp. 13-26.

Gargantini P.R., Giammaria V., Grandellis C., Feingold S.E., Maldonado S., Ulloa R.M. (2009) Genomic and functional characterization of StCDPK1. Plant Molecular Biology 70:153-72.

Guillaumie S., Goffner D., Barbier O., Martinant J.P., Pichon M., Barriere Y. (2008) Expression of cell wall related genes in basal and ear internodes of silking brown-midrib-3, caffeic acid O-methyltransferase (COMT) downregulated, and normal maize plants. Bmc Plant Biology 8:71.

Hake S., Smith H.M.S., Holtan H., Magnani E., Mele G., Ramirez J. (2004) The role of KNOX genes in plant development. Annual Review of Cell and Developmental Biology 20:125-151.

Hertzberg M., Aspeborg H., Schrader J., Andersson A., Erlandsson R., Blomqvist K., Bhalerao R., Uhlen M., Teeri T.T., Lundeberg J., Sundberg B., Nilsson P., Sandberg G. (2001) A transcriptional roadmap to wood formation. Proceedings of the National Academy of Sciences of the United States of America 98:14732-14737.

Jasinski S., Piazza P., Craft J., Hay A., Woolley L., Rieu I., Phillips A., Hedden P., Tsiantis M. (2005) KNOX action in Arabidopsis is mediated by coordinate regulation of cytokinin and gibberellin activities. Current Biology 15:1560-1565.

Jones-Rhoades M.W., Bartel D.P. (2004) Computational identification of plant MicroRNAs and their targets, including a stress-induced miRNA. Molecular Cell 14:787-799.

Kerstetter R., Vollbrecht E., Lowe B., Veit B., Yamaguchi J., Hake S. (1994) Sequence-Analysis and Expression Patterns Divide the Maize Knotted1-Like Homeobox Genes into 2 Classes. Plant Cell 6:1877-1887.

Kim S.H., Hamada T., Otani M., Shimada T. (2005) Isolation and characterization of MADS box genes possibly related to root development in sweetpotato (*Ipomoea batatas* L. Lam.). Journal of Plant Biology 48:387-393.

Kim S.H., Mizuno K., Fujimura T. (2002) Isolation of MADS-box genes from sweet potato (*Ipomoea batatas* (L.) lam.) expressed specifically in vegetative tissues. Plant and Cell Physiology 43:314-322.

Kloosterman B., Vorst O., Hall R.D., Visser R.G., Bachem C.W. (2005) Tuber on a chip: differential gene expression during potato tuber development. Plant Biotechnol J 3:505-19.

Ku A.T., Huang Y.S., Wang Y.S., Ma D.F., Yeh K.W. (2008) IbMADS1 (*Ipomoea batatas* MADS-box 1 gene) is involved in tuberous root initiation in sweet potato (*Ipomoea batatas*). Annals of Botany 102:57-67.

Liljegren S.J., Ditta G.S., Eshed H.Y., Savidge B., Bowman J.L., Yanofsky M.F. (2000) SHATTERPROOF MADS-box genes control seed dispersal in Arabidopsis. Nature 404:766-770.

Lokko Y., Anderson J.V., Rudd S., Raji A., Horvath D., Mikel M.A., Kim R., Liu L., Hernandez A., Dixon A.G.O., Ingelbrecht I.L. (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., Verdier V. (2004) A unigene catalogue of 5700 expressed genes in cassava. Plant Molecular Biology 56:541-554.

Lowe S.B., Wilson L.A. (1974) Comparative Analysis of Tuber Development in 6 Sweet-Potato (Ipomoea-Batatas (L) Lam) Cultivars .1. Tuber Initiation, Tuber Growth and Partition of Assimilate. Annals of Botany 38:307-317.

McGregor C. (2006) Differential expression and detection of transcripts in sweetpotato (*Ipomoea batatas* (l.) lam.) using cdna microarrays, Louisiana State University, Baton Rouge, USA.

Mele G., Ori N., Sato Y., Hake S. (2003) The knotted1-like homeobox gene BREVIPEDICELLUS regulates cell differentiation by modulating metabolic pathways. Genes & Development 17:2088-2093.

Nagata T., Saitou K. (2009) Regulation of Expression of D3-type Cyclins and ADP-Glucose Pyrophosphorylase Genes by Sugar, Cytokinin and ABA in Sweet Potato (*Ipomoea batatas* Lam.). Plant Production Science 12:434-442.

Noh S.A., Lee H.S., Huh E.J., Huh G.H., Paek K.H., Shin J.S., Bae J.M. (2010) SRD1 is involved in the auxin-mediated initial thickening growth of storage root by enhancing proliferation of metaxylem and cambium cells in sweetpotato (*Ipomoea batatas*). Journal of Experimental Botany 61:1337-1349.

Paponov I.A., Paponov M., Teale W., Menges M., Chakrabortee S., Murray J.A.H., Palme K. (2008) Comprehensive transcriptome analysis of auxin responses in Arabidopsis. Molecular Plant 1:321-337.

Raices M., Chico J.M., Tellez-Inon M.T., Ulloa R.M. (2001) Molecular characterization of StCDPK1, a calcium-dependent protein kinase from Solanum tuberosum that is induced at the onset of tuber development. Plant Molecular Biology 46:591-601.

Riechmann J.L., Meyerowitz E.M. (1997) MADS domain proteins in plant development. Biol Chem 378:1079.

Sarkar D. (2008) The signal transduction pathways controlling in planta tuberization in potato: an emerging synthesis. Plant Cell Reports 27:1-8.

Scofield S., Dewitte W., Murray J.A. (2008) A model for Arabidopsis class-1 KNOX gene function. Plant Signal Behav 3:257-9.

Shore P., Sharrocks A.D. (1995) The Mads-Box Family of Transcription Factors. European Journal of Biochemistry 229:1-13.

Sojikul P., Kongsawadworakul P., Viboonjun U., Thaiprasit J., Intawong B., Narangajavana J., Svasti M.R. (2010) AFLP-based transcript profiling for cassava genome-wide expression analysis in the onset of storage root formation. Physiol Plant.

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

Tanaka M., Kato N., Nakayama H., Nakatani M., Takahata Y. (2008) Expression of class I knotted1-like homeobox genes in the storage roots of sweetpotato (*Ipomoea batatas*). Journal of Plant Physiology 165:1726-1735.

Testone G., Bruno L., Condello E., Chiappetta A., Bruno A., Mele G., Tartarini A., Spano L., Innocenti A.M., Mariotti D., Bitonti M.B., Giannino D. (2008) Peach [Prunus persica (L.) Batsch] KNOPE1, a class 1 KNOX orthologue to Arabidopsis BREVIPEDICELLUS/KNAT1, is misexpressed during hyperplasia of leaf curl disease. J Exp Bot 59:389-402.

Thomas G.S., Hedden P. (2006) Gibberellin metabolism and signal transduction, in: P. Hedden and S. G. Thomas (Eds.), Plant Hormone signaling, Blackwell publishing. pp. 147-176.

Tiwari S.B., Hagen G., Guilfoyle T. (2003) The roles of auxin response factor domains in auxin-responsive transcription. Plant Cell 15:533-543.

Togari Y. (1950) A study of tuberous root formation in sweet potato. 68:1–96. Bull. Natl. Agric. Exp. Stn. 68:1-96.

Villordon A.Q., La Bonte D.R., Firon N., Kfir Y., Pressman E., Schwartz A. (2009) Characterization of Adventitious Root Development in Sweetpotato. Hortscience 44:651-655.

Wang D.K., Pei K.M., Fu Y.P., Sun Z.X., Li S.J., Liu H.Q., Tang K., Han B., Tao Y.Z. (2007) Genome-wide analysis of the auxin response factors (ARF) gene family in rice (Oryza sativa). Gene 394:13-24.

Wang Q.M., Zhang L.M., Wang Z.L. (2005) Formation and thickening of tuberous roots in relation to the endogenous hormone concentrations in sweet potato. Scientia Agricultura Sinica 38:2414-2420.

Wilson L.A., Lowe S.B. (1973) Anatomy of Root System in West-Indian Sweet Potato (Ipomoea-Batatas (L) Lam) Cultivars. Annals of Botany 37:633-&.

You M.K., Hur C.G., Ahn Y.S., Suh M.C., Jeong B.C., Shin J.S., Bae J.M. (2003) Identification of genes possibly related to storage root induction in sweetpotato. Febs Letters 536:101-105.

Chapter 2. Drought and Salt Stress on Sweetpotato Storage Root Development

2.1 Introduction

Sweetpotato [*Ipomoea batatas (L.) Lam.*] is a perplexing crop to grow. The crop is established by vine cuttings survives in moderate to extreme environmental conditions. Yet, a solid stand does not translate into predictable yield. It is becoming evident that proper conditions during the initial stages of growth of the cuttings appear essential for obtaining high yield. The vine cuttings have no roots initially and produce adventitious roots at nodes 1-2 days after planting. These adventitious roots are now recognized as the progenitors of storage roots (Villordon et al., 2009). Unfortunately, these adventitious roots have alternative pathways which impact yield potential. They can bulk and form fleshy storage roots or form non storage forming roots of two types: thin feeder roots or roots with an intermediate level of thickening sometimes referred to as pencil roots. Anatomically the stele of the root undergoes cell division and expansion at the cambium region during the transition from an adventitious root to a storage root, where the anomalous and vascular cambium leads to rapid thickening of the roots (Wilson and Lowe, 1973). Lignification of the vascular cambium halts thickening of the roots, which results in the formation of pencil roots. Reports of the effect of environmental factors such as nutrients, temperature, and water availability on the formation of fleshy storage roots are limited (Belehu et al., 2004; Togari, 1950). Togari (1950) studied several environmental factors showing the importance of fertilizer, optimum temperature (23.4 °C), soil humidity and compactness during the first 20 days after trans-planting (DAT) towards storage root formation. It was the first report documenting the correlation of abiotic factors with the degree of lignification and cambium activity of the stele of the roots, both of which affect negatively or positively the formation of a storage root, respectively. The root architecture, measured by number and length of adventitious roots, and density of lateral root development of sweetpotato is genetic, which is highly affected

by the soil humidity (Pardales and Yamauchi, 2003). At 20-days of trans-planting those adventitious roots forming storage roots have been observed to have a higher density of lateral roots as compared to the count of lateral roots in pencil and lignified roots (Villordon et al., 2012). Although a number of molecular studies, during the past decade, have reported several candidate genes that are upregulated or expressed preferentially in storage roots, none have been shown to be solely responsible for conversion of the adventitious roots into storage organs (Kim et al., 2005; Kim et al., 2002; Ku et al., 2008; Noh et al., 2010; Tanaka et al., 2008; You et al., 2003) (Kim et al., 2005; Kim et al., 2002; Ku et al., 2008; Noh et al., 2010; Tanaka et al., 2008; You et al., 2003). Most of these genes have either regulatory roles as transcription factors, or are specific genes involved in carbohydrate and protein metabolism during the development and thickening of the storage roots.

Sweetpotato is considered to be moderately drought tolerant (Saraswati et al., 2004), but the greatest sensitivity to water deficit is particularly during early establishment period including vine development and storage root initiation (Indira and Kabeerathumma, 1988). Yield loss due to drought could be up to 50-80% depending upon the timing, duration, and intensity of water stress. Although empirical evidence exist demonstrating yield reduction under drought conditions, it is unknown if drought conditions early on during establishment of adventitious roots impacts storage root formation regardless of plant survival. Salinity represents another abiotic stress, which likely impacts storage root formation since sweetpotato is highly susceptible to salt stress (Greig and Smith, 1962) and the negative effects of salt on growth were seen to be more prominent in roots than in leaves (Anwar et al., 2010). Salinity in many ways is similar to drought stress and affects yield, storage root formation and as a whole inhibit the root growing Responses to salinity stress of plants are determined by genetic factors and it is different from one to specie. Salinity affects yield, growth parameters and the timing of development in specific

way in plants (Shannon and Grieve, 1999). In general in plants at low salinities root growth is often less affected, or sometimes even stimulated by salinity, as compared to shoot growth. Salinity stress influence negatively the accumulation of beta carotene and root growth is much more sensitive to salinity than vine growth on sweetpotato (Greig and Smith, 1962). Besides, drought stress is known to increase weevil feeding in sweetpotato (Mao et al., 2004). Lower yield and increased susceptibility to pests under these stresses decreases the economic returns to the farmer. Both drought and salinity stresses represent two global constraints for sweetpotato production since most of the production of sweetpotato occurs in semi-arid regions. Considering the complexity of the physiological and genetic mechanisms associated with stress tolerance, a genomics-based understanding of the stress response of sweetpotato will have a significant impact on its productivity in stress environment (Boyer, 1982) . Candidate drought and salt stress responsive genes, identified through genomics research, will have great potential in widening the natural allelic variation and their possible utilization for crop improvement (Rus et al., 2006). However, sweetpotato has lagged behind with respect to studies leading to the identification of its drought and/or salt responsive genes. One study reported twelve genes in response to drought stress in white fibrous roots (Kim et al., 2009), and most of these genes are similar to genes that are known to be associated with dehydration response in many other plants. A dehydration responsive element-binding (DREB) protein gene encoding for an AP2/EREBP has also been characterized that responds to drought stress in roots and stems of sweetpotato plants (Kim et al., 2008). Evidence of genetic basis of sweetpotato tolerance to drought and salinity comes from different studies carried out by the use in vitro culture practice along with field and greenhouse experiments (Ekanayake and Dodds, 1993; Prabawardani et al., 2004; Ricardo, 2011; Saraswati et al., 2004) .

The present research examines the effect of water deprivation (drought stress) on the fate of adventitious roots under field and greenhouse conditions and the expression of a selected set of genes in developing root tissue. Results from pilot research examining the effect of salt stress on gene expression in adventitious roots as well as in leaves are also presented.

2.2 Materials and methods

2.2.1 Plant material

Plant cuttings were obtained from generation 0 (derived from in vitro cultures) virus tested, greenhouse grown 'Beauregard' for greenhouse studies and virus-tested generation 1 'seed roots' in plant beds for field studies.

2.2.2 Delayed watering studies under greenhouse conditions

Cuttings (n=27) were transplanted in dry sand in cylindrical tubes (50 x 9.82 cm) under greenhouse conditions. Plants from were grown inside the greenhouse under a day/night temperature regime of 27/21 °C and 28/25 °C at trial 1(Jan 30 to March 10, 2011) and trial 2 (May 9 to June 21, 2011), respectively.Water was withheld 5 and 10 days after transplanting (DAT) in two treatments and the control received water the same day of transplanting (0 DAT). Each treatment was replicated nine times. Watering (400 ml) occurred every three days after the initial 5 and 10 day treatment period. Peters 20–20–20 solution (Peters Professional Soluble Plant Food, Scotts-Sierra) was applied [0.374g /200 ml] at 12 and 22 days after transplanting. Four weeks after transplanting, the plants were evaluated for the number of storage roots (roots having maximum diameter > 1.5 mm), the maximum diameter or width of this root (MaxDiam) and the weight (WeightSR). Thin pigmented roots (maximum diameter < 1.5 mm) were included in the total count of storage roots (SRCount1) since we considered them as putative-forming/developing storage roots (DSR) (Villordon et al., 2009). A second count of storage roots (SRCount2) excluded these pigmented storage roots. Maximum diameter of storage roots were

measured with a caliper. Likewise, the plants were evaluated for number of non-storage roots (i.e. white fibrous roots), fresh shoot weight, fresh root weight but these measurements were not included in the analysis. The experiments were repeated twice.

A second study was carried out to evaluate gene expression of selected genes (Table 2.1) in total roots from plants that were under initial drought stress vs. well-watered (control) plants. For this purpose, the plant material was grown under greenhouse conditions and consisted of 9 control cuttings (watering started at 0 DAT, then every three days) and 9 drought stressed plants with a 5 DAT delay of watering. Fertilizer was applied at 12 DAT. Roots from these plants were sampled at 14 DAT in triplicate by pooling roots from three plants and then the roots were frozen in liquid nitrogen. Material was kept at -80 °C until further procession for RNA extraction. This experiment was repeated twice.

2.2.3 Delayed watering studies under field conditions

Field experiments were conducted in the summer of 2010 and 2011 in well-drained research fields in Chase, La, USA (32o 6'N, 91o 42'W). The soil taxonomic class was fine-silty, mixed, active, thermic Typic Glossaqualfs. In each year, natural rainfall deficits during May and June created conditions where soil moisture in the root zone was near the wilting point for the soil type used in the studies. Rickard and Fitzgerald (1969) have previously defined agricultural drought as existing when the soil moisture in the root zone is at wilting point or below. For the fields used in the studies, storage root formation does not occur below the hard pan (30 cm. depth to hard pan). We used these conditions to compare storage root yield from plots with delayed watering vs. plots that were irrigated to maintain soil moisture at 50% of field capacity during the critical plant establishment, storage root formation and early growth periods (Villordon et al., 2012). Field preparation activities, including fertilizer rates, herbicide, and insecticide applications, were very similar in each year. Supplemental overhead irrigation was

supplied with a traveling irrigation sprinkler if a rainfall event did not occur in irrigated plots. Stand counts were conducted between 15 and 35 days after transplanting (DAT). Plant stands ranged from 90 to 100% in all years.

In irrigated plots, in-season supplemental irrigation was based on soil moisture sensor data and irrigation was applied when soil moisture at the 15 cm depth approached 25% of field capacity. A 16% volumetric water content (VWC) represented ≈50% of field capacity in this study. This soil moisture range has previously been calibrated (Constantin et al., 1974) and validated (Villordon et al., 2010) as optimum for sweetpotatoes grown in north Louisiana. The amount of rainfall or supplemental irrigation applied in each growing season was generally equivalent to 25 mm per week for 15 to 16 weeks. Soil moisture was monitored with a HydroSense Soil Water Content Management System (CS-620, 20-cm probe; Campbell Scientific, Inc. Logan, UT). Soil moisture sensors (5 cm in length) were installed vertically at two depths (5 and 15 cm) in two plots in each year. All supplemental irrigation was delivered via a traveling irrigation sprinkler or furrow irrigation (after 35 DAT). For non-irrigated plots (delayed watering), soil moisture was consistently below or at levels defined as the wilting point during the critical plant establishment, storage root formation and early growth periods. In 2010, soil moisture at the 15 cm depth stayed below 10% VWC during the first 30 days of growth in non-irrigated plots; soil moisture slightly increased with subsequent late-season rainfall events. This range have been previously defined as the witling point for this soil type (Ley et al., 1994). In 2011, soil moisture at the 15 cm depth ranged from 9-11% VWC during the first 30 days of growth in non-irrigated plots; soil moisture slightly increased with subsequent late-season rainfall events. At harvest, storage roots were graded according to USDA standards (USDA, 2005): U.S.#1 (5.1 to 8.9 cm diameter and 7.6 to 22.9 cm in length), Canner (2.5 to 5.1 cm in

diameter and 5.1 to 17.8 cm in length), and Jumbo (larger than both groups). Total marketable yield was defined as the sum of U.S. #1, Canner, and Jumbo.

2.2.4 Salt stress studies

Twelve cuttings were transplanted in sand in cylindrical tubes (50 x 9.82 cm) under greenhouse conditions. An application of 200ml of 106mM NaCl solution to the sand from first day of transplanting to a set of cuttings (n=6) and then each day up to the 16th day of sampling. A set of control plants (n=6) was watered with 200ml of water every three days and as required to keep soil at saturation. Root and leaves at 16 DAT were collected from each set of plants (salt, control) in triplicate and frozen with liquid nitrogen and kept at -80 °C untilRNA extraction. In this study no fertilizer was applied during the growth of plants. Concentration of NaCl used in this study was in a range that was previously found to have significant effects on growth parameters (Anwar et al., 2010).

2.2.5 Gene expression profiling by quantitative reverse transcription PCR

The collected tissues were ground into powder in liquid nitrogen. RNA was extracted using the Qiagen RNeasy kit (Valencia, CA) following vendor protocol. An on-column DNase I treatment was done to eliminate any contaminating DNA. RNA was eluted in the provided water. RNA quality and RNA yield were determined taking the A260:A280 and A260:A230 ratios using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) with a 1.5 µl sample. cDNA synthesis required 2 µg of total RNA. An iSCRIPT strand cDNA synthesis kit (Bio-Rad, Hercules, CA) was used. Incubation was at 5 min at 25°C, 30 min at 42°C, 5 min at 85°C and hold at 4°C. The cDNA was then diluted 3 times with water, and 2 µl was used for the PCR using the iQSYBR green supermix (Bio-Rad, Hercules, CA). The conditions were: 1 cycle at 95 °C per 3 min; 45 cycles: 95 °C per 10 seconds, 65 °C per 30 seconds; and final step for melting curve analysis. The reference gene used for normalization

was the elongation factor 1-alpha (IbEF1a). Each qRT-PCR reaction was run in triplicate and each gene was tested twice using cDNA made from two independent RNA sample sets from total roots (drought vs. control). For salt stress a single set of triplicate samples (salt vs. control) was used to extract the RNA and cDNA preparation. The relative abundance of each transcript was calculated using the delta-delta-Ct method comparing stressed (drought, salt) vs. control samples. Table 2.1 describes all primer sets (20 genes) used for comparing the expression under drought vs. control samples; only a subset of these primer pairs (13 genes, including IbEF1a) was used for the study of salinity vs. control from leaf and root tissues. Table 2.4 describes the putative functional role of the genes targeted by these primers. Twenty genes were selected including 2 KNOX genes (Ibkn2, Ibkn3) (Tanaka et al., 2008) and the reference IbEF1a gene. The majority of these genes were selected from sweetpotato root transcriptome with either a role in transcriptional activity or in molecular signaling.

2.2.6 Experimental design and statistical analyses

The experiment under greenhouse conditions used a completely random design (CRD) by assigning the treatments to experimental units completely at random with six (salt vs. control under greenhouse conditions) and nine (drought vs. control under greenhouse conditions. The field experiment used a randomized complete block design (RCBD) with 4 replications (plots) and 10 plants per plot per treatment for each of the two treatments, one set of plants that were well-watered (irrigated) and another set of plants were under drought-stress (non-irrigated). Statistical analyses of greenhouse data were done using Proc GLM in SAS (SAS Institute Inc. Cary, NC). Field data were combined across 2010 (three planting dates) and 2011 (two planting dates) and subjected to statistical analysis using the PROC MIXED procedure in SAS (version 9.2; SAS Institute, Cary, NC).

2.3 Results

2.3.1 Delayed watering studies under greenhouse conditions

The results in separate trials showed a negative effect of the initial drought stress treatment on the growth of sweetpotato storage roots (Table 2.2). A reduction in the number, size and weight of storage roots was observed due to the imposed drought stress in the plants in comparison to the control plants. The number of storage roots decreased by 43% in trial 1 and by 30% in trial 2 under moderate 5 DAT drought stress (Table 2) using storage root counts (SRCount1) which include all thickened and pigmented (putative) storage roots. Using more conservative counts, SRCount2 [excluding pigmented putative forming storage roots], the reductions were 42.2% (5 to 2.89 roots) and 29.06% (3.44 to 2.44 roots) in trials 1 and 2, respectively, when comparing the 5 DAT drought stress versus control plants. The results at 10 DAT was consistent with the 5 DAT plants, showing a reduction in SRCount1 of ranging from 59% (5.89 roots in control to 2.44 roots at 10 DAT) to 67% (4 to 1.33 roots) in trial 2 and trial 1, respectively. SRCount2 showed a reduction of 63% (5 to 1.89 average roots) to 70.9% (3.44 roots to 1 root) in trial 2 and trial 1, respectively.

All plants reached similar foliar growth (no significant difference) at the time of harvest and no death of cuttings occurred as a result of drought stress. Moderate and severe drought stress effects are seen in the reduction of weight and the maximum diameter reached by the storage roots in plants that were under drought stress compared to the control. For example, the total weight of the storage roots were 50% (6.45/12.86) and 27% (3.49/12.86) in plants under 5 and 10 DAT of drought stress, respectively, less than control plants (Table 2.2). Reduction in root weight, WeightSR, was similar in the second experiment (Table 2.2).

2.3.2 Delayed watering studies under field conditions

Yield of sweetpotato plants in non-irrigated drought stressed plots showed a significant reduction of storage roots compared to the irrigated plots after 60 days of growing (Table 2.3). The results of field experiments are combined from 2010 and 2011 data with 2-3 planting dates within each year.

2.3.3 Gene expression under drought stress and salt stress

Nine genes were found upregulated in drought stressed roots having fold change rates of expression at least 1.5 and only two genes (IbCBP1 and IbCDPK3) were downregulated (fold changes ratio less than 1) (Table 2.4, Figure 2.3) . Of the upregulated genes, eight are transcription factors and the other is coding calcium binding protein (IbCBP2) and they are represented in Figure 2.3.

The greatest increase in abundance of transcripts under drought stress was observed for 3 orthologous genes that are known to mediate response to abiotic stress in other plants. IbHB2 (encoding a homeobox protein), IbCRF1 (encoding a protein similar to the *Arabidopsis* cytokinin response factor 1) and IbAREB1 (encoding an abscisic acid responsive elements-binding factor) are differentially expressed in drought stressed plants with a 5 DAT delay of water application in comparison to the control.

Table 2.1. Primer sets used for quantitative RT PCR analysis.

Gene Name[1]	Oligo Name	Sequence (5' to 3')	Product Length (bp)	cap3p80v3 (PBL) [cap3p90][2]	GenBank (SPGI) [PGDB][3]
IbGRF	GRF2_F	AGCTGACCTCGATCTGCAAC	210	(S_PBL_c1096), [Contig18728.2]	
IbGRF	GRF2_R	CAGTAGTGGCGATTTCCTGAGC			
IbDREB1	DR1_F	CGGCGATGATGAAGCCTGA	137	Contig35892.1	
IbDREB1	DR1_R	CATATCCACCAGCAAATTGTTC			
IbHB1	HB1_F	GGATGAAGAGGTTGTTGGTACG	210	Contig3374.1, (S_PBL_c3587)	
IbHB1	HB1_R	CAGTCCACCTCCGTCTGTTTC			
IbHB2	ATHB7_F	CAGGAAGCTGAAGATGAGTTATG	139	(S_PBL_c7043),[Contig529.2]	
IbHB2	ATHB7_F	CTCCTCATCAGTCTCCTTGTC			
IbCBP2	CBP2_F	CCTGAAACAATGGAGGAAGC	159	[Contig23250.2]	
IbCBP2	CBP2_R	CCAGTTTGCTTTTCCCAGTTC			
Ibkn2	Kn2F	TTTGAGTGAAGGCGATGCTATGAA	209		AB283028
Ibkn2	Kn2R	ATTGATCGAGTTCTGGGTCCTT			
Ibkn3a	KN_F1	CGCCTAGGTCCATAATCCAATCCA	212	Contig5774.1,(S_PBL_c31412), [Contig36256.2]	
Ibkn3a	KN_R1	CCACAAGTTGCAGATCAAAGATAA			
Ibkn3b	KN_F2	CGCCTAGGTCCATAATCCAACATA	215		AB283029
Ibkn3b	KN_R2	CCACAAGTTGCAGATATCAGATGAC			
IbEF1a	EF1a_F	TCCTGAAGAATGGTGATGCTGG	165	S_PBL_c12390	
IbEF1a	EF1a_R	CAGTTGGGTCCTTCTTGTCAAC			
IbTAP	TAP_F	TGCTGCGCGATCACAATACAT	219	Contig27416.1, Contig28125.2]	
IbTAP	TAP_R	CATGGTTGGCAATCTATAGTTTGG			
IbSnRK	SnRK_F	ATGTGGTGCTATTCGCATAATCC	217	S_PBL_c2856	
IbSnRK	SnRK_R	CACAGAAGTCCAGAAACAGAAATTG			
IbCRF1	CRF1_F	CCCTAAACGGAGACGAATTCAC	183	S_PBL_c3693	
IbCRF1	CRF1_R	GGGGGTGTTCTAACTGATTTTC			
IbCRF2	CRF4_F	CCCACAGATATCCCATTCATCG	211	S_PBL_c5300	
IbCRF2	CRF4_R	AATAATCGTCGACGTTCACGTC			

(Table 2.1 cont.)

Gene Name[1]	Oligo Name	Sequence (5' to 3')	Product Length (bp)	cap3p80v3 (PBL) [cap3p90][2]	GenBank (SPGI) [PGDB][3]
IbAREB	ABRE_F	GCAATCTAGTTCAAGGCTCTGC	188	[Contig41614.2], (S_PBL_c27146)	
IbAREB	ABRE_R	TTTCCTCCCCCTTACATTTCC			
IbCDPK3	CDPK3_F	CTTGATCGGAACCCGAAAAGGAGG	93	Contig28899.1[Contig34516.2]	
IbCDPK3	CDPK3_R	AAGAGGCTTGTCGGGGGCCATAC			
IbBEL2	BEL1_R2 (BEL1b)	TCCTCCATATACATCTCCTCCACCAT	248		DV038045.1, DC881164.1 [IBAT_PUT65, IBAT_PUT9796]
IbBEL2	BEL1_F2 (BEL1b)	AGGGCTTTGCAGCAGTTGGGAATGAT		Contig28899.1	
IbBEL1	BEL1_F1 (BEL1a)	TAGAGACCAAATTGAAGATTGCCAA	204		EE880927.1 [IBAT_PUT8950]
IbBEL1	BEL1_R1 (BEL1a)	GAAGGCAGATCAGAGCAGCATA	204		
IbBEL3	BEL13_R1	GCTTGTCTGTGTCTGTGGGATA	252		(Ib28923)
IbBEL3	BEL13_F1	CCATAACTGAGCAGCTGCAGTT			
IbCBP1	CBP1b_F	GACCGATCCGGTCGAGCTTAACTTG	259	RT_027290.1, RT_055908.1, (S_PBL_c13311)	JG699346
IbCBP1	CBP1_R	TGTTCTCATTCTCAACGGCTAAC			
IbMYB	MYB_F	CAGATGGTCCCTTATTGCAGG	233		EE877091 [IBAT_PUT985]
IbMYB	MYB_R2	CGTCGCTGATTTCTTTCCTCTG			

[1] Gene name is given based on the closest matching gene found by blastx analysis.

[2]Three transcripts assemblies named: cap3p80v3, PBL and cap3p90 were used to select the sequences. Assemblies details are presented in next chapter.

[3]Current sequence ID of the matching sequence in public databases: accession number from GenBank (http://www.ncbi.nlm.nih.gov/genbank/), customized ID (Ib#) where # is the number identifying the sequence in Sweetpotato Gene Index(SPGI) (Schafleitner et al, 2010); sequences from PlantGDB (http://www.plantgdb.org/) were renamed for example IBAT_PUT985 is a sequence deposited at PlantGDB as "PUT-169a-Ipomoea_batatas-985 PlantGDB-assembled Unique Transcript-fragment derived from *Ipomoea_batatas* mRNAs Jan_14_2009 (based on GenBank release 169)".

Table 2.2. Effect of drought stress on sweetpotato storage root growth. Data on sweetpotato storage root numbers and average storage root weight and maximum diameter after delaying water for 0 (control), 5 and 10 days after transplanting. Trial 1 from Jan 30, 2011 to March 10; Trial 2 from May 9 to June 21, 2011. SE=standard error. N=number of replicates

Treatment[1]	Variable	N	Mean (SE) (trial 1)[4]	Mean (SE) (trial 2)[4]
0	SRCount1[2]	9	5.89(0.42) a[4]	4.00(0.41) a[4]
	SRCount2[3]	9	5.00(0.33) a	3.44(0.41) a
	WeightSR	9	12.86(0.64) a	12.46(0.98) a
	MaxDiam	9	5.59(0.43) a	7.6(1.08) a
5	SRCount1[2]	9	3.33(0.24) b	2.78(0.22) b
	SRCount2[3]	9	2.89(0.31) b	2.44(0.24) b
	WeightSR	9	6.45(0.60) b	6.57(1.00) b
	MaxDiam	9	4.85(0.28) b	5.10(0.77) b
10	SRCount1[2]	9	2.44(0.44) c	1.33(0.24) c
	SRCount2[3]	9	1.89(0.42) c	1.00(0.24) c
	WeightSR	9	3.49(0.56) c	2.26(0.48) c
	MaxDiam	9	2.74(0.22) c	2.06(0.18) c

[1]Treatment are number of days without receiving first watering after transplanting.
[2]SRCount1 are the number of storage roots including the pigmented roots (i.e. putative initiating storage roots with an estimated maximum diameter less than 1.50mm).
[3]SRCount2 are the number of storage roots excluding the pigmented roots.
[4]Means following by same lower case letter within a column for a giving measurement were not significantly different at $P<0.05$.

Transcription factors of BELL (IbBEL1, IbBEL2 and IbBEL3) and the KNOX Ibkn3a which mediate thickening of the roots were also upregulated from 1.4 to 2.2 times in adventitious roots derived from plant that were under drought stress at planting. In comparison, only two genes, IbCDPK3 and IbCBP1, encoding a calcium dependent protein kinase and a calcium-binding protein, respectively, were found downregulated (1.3 and 4 fold, respectively) in roots in response to the drought stress. Three genes, named IbDREB1, IbMYB, IbGRF2 did not showed a significant differences in roots from drought stressed (from 5 DAT treatment) versus control plants; IbSnRK and IbTAP, genes showed inconsistent pattern of expression in the two independent sets of samples (data not shown).

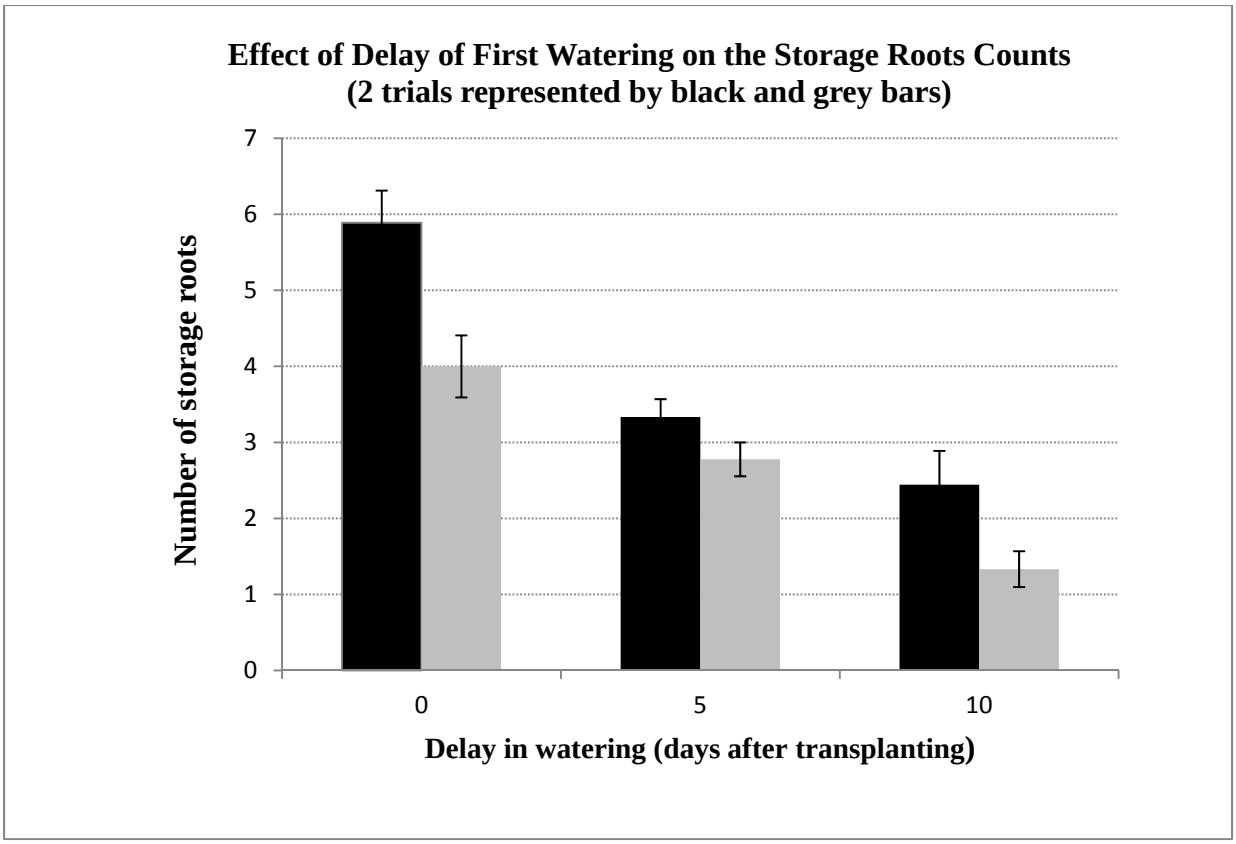


Figure 2.1. Effect of drought stress (delay of first watering 5 and 10 days after transplanting) vs. control plants (no delay) in the number of storage roots from two independent trials (black and gray columns are data from trial 1 and trial 2, respectively). Bars represent mean (n=9) and error bars represent standard error.

Of the downregulated genes, only IbCBP1 encoding a calcium binding protein showed the greatest reduction of expression up to 80% (fold changes 0.22) in roots under drought stress compared to control and a slight decrease of transcript abundance was observed for IbCDPK3 (fold changes 0.76), a gene putatively encoding a calcium dependent protein kinase (Table 2.4, Figure 2.3). Fold changes of IbBEL2, Ibkn2 and IbCRF2 on drought vs. control roots presented in Table 2.3 correspond to a single set of triplicate sample treatments.

A B

Figure 2.2 Storage root development at 65 DAT in irrigated (A) and non-irrigated (B) plots. Chase, La, 2010.

Table 2.3. Storage root yield response to irrigation treatments in field (Chase, Louisiana).

	Yield (t/ha) [1]			
Treatment	**U.S.#1[2]**	**Canner[2]**	**Jumbo[2]**	**TMY[2]**
Irrigated	26.45a	12.39a	6.35a	45.19a
Non-irrigated	12.98b	9.94a	2.82a	25.75a
P-value	<0.0001	0.2	0.2	<0.0001

[1] t/ha – tons per hectare

[2] U.S. #1 (51-89 mm in diameter, 76-229 mm long); Canner (25-51 mm in diameter, 51-178 mm long); and Jumbo (larger than U.S. #1 in diameter, length or both, and without objectionable defects). TMY (Total Marketable Yield) representing the summation of all grades.

Salinity stress mostly reduced the level of transcript abundance of tested genes in roots. For example, 8 genes out of 12 were downregulated with fold changes ranging from 0.37 to 0.82 (reduction of 18% to73% in comparison to the level in control plants) (Table 2.4), and 3 genes did not show a significant change in their expression in roots. The highest up regulation in in root

tissue in response to salinity was observed for Ibkn3a (fold change rate of 1.47) representing an increase in ~50% respect to control. Salinity stress in leaves differed. Many genes were upregulated. Notable were 3 genes: IbAREB (2.63-fold), IbHB2 (3.82-fold) and IbDREB1 (2.61-fold), that have the highest upregulation in leaf tissue in response to salt. All are orthologous of known plant stress response genes and two of them (IbAREB and IbHB2) showed increased expression under drought stress. These results document their role in abiotic stress response of sweetpotato. Six genes: IbBEL3, IbCBP1, IbCRF2, Ibkn3a, IbMYB and IbTAP, showed a reduction of their transcript levels in leaves under salinity stress (Table 2.4). All of these genes except IbCBP1 and IbTAP are transcription factors, not known previously to be associated with sweetpotato root development.

2.4 Discussion

Greenhouse studies: Drought stress in greenhouse grown plants profoundly reduced the number, size and weight of storage roots compared to control plants (Table 2.2) The 5 DAT treatments showed a 30-42% reduction in storage root number across the two studies. Storage root numbers were reduced further (up to 66%) by extending the drought period to 10 DAT. Results at 5 DAT and 10 DAT were consistent and demonstrated the effect moisture has on storage root formation. Preliminary experiments carried out in growth chambers in 2008 and 2009 under controlled conditions of humidity, daylight and temperature had a similar outcome (data not shown).

Field studies: The 2010 and 2011 growing seasons were characterized by prolonged periods without rainfall, especially during the critical period of storage root initiation, i.e., 1-20 DAT. This growing environment allowed for the comparison of irrigated vs. non-irrigated treatments on storage root initiation and subsequent storage root bulking without the

confounding effects from natural rainfall events. Under these conditions, storage root development was significantly delayed in non-irrigated plots when sampled at 60 days (Fig. 2.2).

In each year, there was marginal soil moisture (less than 25-50% of field capacity) during the transplanting phase in the non-irrigated plots. This allowed for some storage root initiation; however, the lack of additional soil moisture did not allow further development, resulting in low storage root count and delayed storage root development. At harvest (110-130 DAT), irrigated plots showed over a 100% increase in U.S.#1 yield relative to non-irrigated plots (Table 2.3).

The reduction of yield and quality of roots observed in the field and the lower weight, maximum diameter of the storage roots, and reduced number in greenhouse grown drought stressed plants present consistent trends and support our hypothesis that lack of soil moisture irreparably alters root development towards non storage forming roots. Pardales and Yamauchi (2003) demonstrated that soil moisture first 20 days after planting is important for root development. Our results are consistent with (Pardales and Yamauchi, 2003), although their work demonstrated a varietal dependent association of altered root growing and soil moisture.

Gene expression: Previous work in sweetpotato has demonstrated a number of key genes involved in storage root development by comparing storage and non storage forming roots (Kim et al., 2005; Kim et al., 2002; Ku et al., 2008; Noh et al., 2010; Tanaka et al., 2008; You et al., 2003). Although only one work focused primarily on the identification and study of genes from fibrous roots under drought stress (Kim et al., 2009), it was done in late growth. Our intent was to assess gene expression in total root pools from non-stressed and drought-stressed plants (5 DAT) at an early stage of growth. Two week old roots were used from two independent set of plants for cDNA preparation and the real-time PCR experiment. We selected nineteen genes (20

Table 2.4. Gene expression in sweetpotato roots from drought stressed plants vs. control plants and leaves and roots from salt vs. control plants estimated by quantitative RT-PCR (nd=no data available because the gene was not tested), ncd=no consistent data in independent set of replicates).

Gene name	Fold change in roots of Drought vs. Control plants[1]	Fold change in Salt stressed vs. Control in leaves (roots) [2]	Annotation
IbAREB	2.23	2.63(0.51)	AT1G45249 (AREB) abscisic acid responsive elements-binding factor 2, ABRE binding factor
IbBEL1	2.01	Nd	AT5G41410.1 POX (plant homeobox) family protein
IbBEL2[x]	1.45	1.01(0.59)	AT2G35940(BLH1,EDA29) BEL1-like homeodomain
IbBEL3	1.95	0.71(1.23)	AT5G02030 (BLH9,HB-6) POX (plant homeobox) family protein; BEL1-LIKE HOMEODOMAIN 9
IbCBP1	0.22	Nd	AT2G44310.1 Calcium-binding EF-hand family protein
IbCBP2	2.12	0.76(0.37)	AT2G33990.1 iqd9 (IQEdomain 9); calmodulin_binding family protein
IbCDPK3	0.76	Nd	AT4G09570(CPK4) Calcium dependent protein kinase 4
IbCRF1	5.93	0.95(1.16)	AT4G11140(CRF1) cytokinin response factor 1
IbCRF2	1.05	Nd	cytokinin response factor 2
IbHB1	1.44	0.66(0.54)	AT4G37790.1 Homeobox-leucine zipper protein family (HAT22)
IbHB2	8.56	3.82(0.40)	AT2G46680 (ATHB7, HB-7) homeobox 7
Ibkn2[x]	1.43	Nd	A8QXP6(IBKN2) Class-I knotted1-like homeobox protein IBKN2
Ibkn3a	1.91	0.27(1.47)	A8QXP7(IBKN3) Class-I knotted1-like homeobox protein IBKN3
Ibkn3b	0.88	Nd	A8QXP7(IBKN3) Class-I knotted1-like homeobox protein IBKN3
IbDREB1	1.18	2.61(0.46)	AT5G52020.1 encodes a member of the DREB subfamily A-4 of ERF/AP2 transcription factor family
IbMYB	nd	0.30(0.47)	MYB transcription factor family
IbSnRK	ncd	1.09(0.67)	AT3G01090 SNF1 kinase homolog 10(SKIN10); snf1-related protein kinase *Solanum tuberosum*
IbTAP	ncd	0.78(0.82)	2A phosphatase associated protein of 46 kD AT5G53000.1

(Table 2.4 cont.)

Gene name	Fold change in roots of Drought vs. Control plants[1]	Fold change in Salt stressed vs. Control in leaves (roots) [2]	Annotation
IbGRF	1.18	Nd	AT1G78300 (GRF2) general regulatory factor 2
Gene name	Fold change in roots of Drought vs. Control plants[1]	Fold change in Salt stressed vs. Control in leaves (roots) [2]	Annotation
IbEF1a	1.0	1.0(1.0)	EF1A_TOBAC Elongation factor 1-alpha Nicotiana tabacum

[1] Two independent sets of total root samples for each treatment (each treatment by triplicate).
[2] A single triplicate set of sample for each treatment.
[x] Fold changes of IbBEL2,Ibkn2 and IbCRF2 on drought vs. control roots correspond to a single set of triplicate sample treatments.

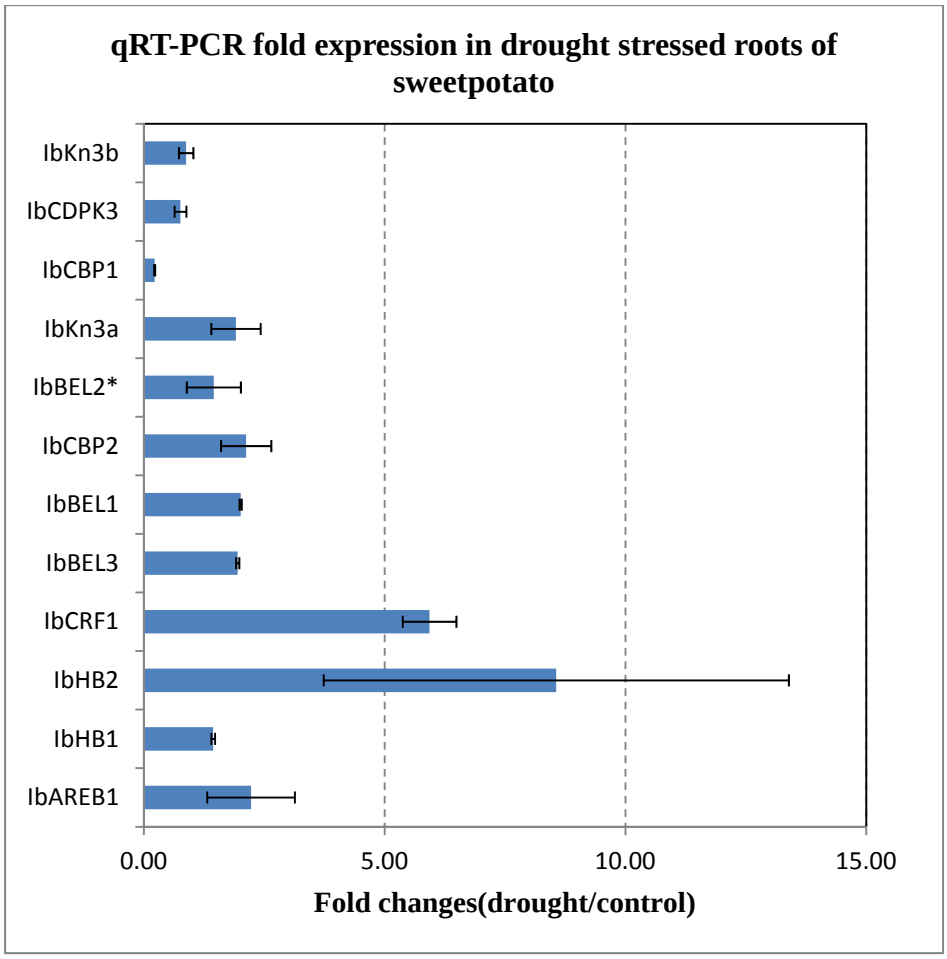


Figure 2.3.Gene expression profiling by qRT-PCR of sweetpotato roots from 14 day old plants under drought stress (5 days after transplanting) and a watered control.

total) for further scrutiny based on comparative analysis of the available sweetpotato transcriptome [i.e. the root library obtained by our group, sequences deposited at GenBank and leaf and stem libraries from drought stressed plants (Schafleitner et al., 2010)]. The reference gene coding an elongation factor protein (IbEF1a) is known to be among the most stable expression under many abiotic stresses and was included. Primers targeting alpha tubulin and actin as reference genes were also tested but IbEF1a was the most stable and demonstrated consistent expression in several stressed and unstressed tissues (data not shown). A digital northern of EF1a in a set of Illumina sequences from both storage and non-storage roots demonstrated that IbEF1a had the same number of reads in both tissues and further supports the IbEF1 as an adequate reference marker (N. Firon, unpublished data).

Homeobox leucine zippers, AP2/EREBP and abscisic acid responsive-like genes: Among the genes that have the highest up-regulation in sweetpotato roots from drought stressed plants were IbHB2, IbCRF1 and IbAREB (Table 2.4, Figure 2.3). IbHB2 and IbCRF1 are orthologous of the *Arabidopsis* genes ATHB7 (At2g46680) and CRF1 (At4g11140), respectively. Notably, we observed that expression of these 3 genes in drought stressed leaves followed the same trend as in the stressed roots, with fold changes ranging from ~ 3 to 12 times (data not show). Similarly, IbHB2 was the most upregulated gene in the leaves of sweetpotato under salt stress (~4-fold) compared to the control (Table 2.4). In addition, salt stressed sweetpotato leaves had the next most enhanced expression of IbAREB and IbDREB1 genes with fold changes of 2.63 and 2.61, respectively. IbAREB is an ortholog of AREB1 (At1g45249), a gene encoding the protein termed abscisic acid responsive elements-binding factor 2; IbDREB1 matches to an AP2/EREBP domain-containing transcription factor gene from *Populus trichocarpa* (DREB63).

Our results showed that one sweetpotato gene (IbHB2) was upregulated under both drought and salt stress in sweetpotato adventitious roots and leaves, respectively (Table 2.4). IbHB2 and IbHB1 were downregulated due to salt stress in root tissues. The product of both IbHB1 and IbHB2 are highly similar to homeobox leucine zippers transcription factors (HD-Zip) ATHB6, ATHB12.and ATHB7 from *Arabidopsis*. Of particular interest is that the product of IbHB2 product is closely related ATHB7. ATHB7, was originally discovered to accumulate in response to water deficit in many organs of *Arabidopsis* (Soderman et al., 1996) and confirmed to be involved in water stress response (Olsson et al., 2004); Soderman et al., (1996) showed that ATHB7 expression was nominal in *Arabidopsis* under normal conditions. Further *Arabidopsis* studies have shown other HD-Zip genes are found to be involved in water stress such as ATHB6 (At2g22430) (Hjellström, 2002; Soderman et al., 1999), and ATHB12 (At3g61890) (Lee and Chun, 1998). HD-Zip proteins are implicated in both developmental changes and stress responses, and they are a common set of TFs in many plants required for tolerance to dehydration and abiotic stresses. For example, two transcripts similar to ATHB12 gene were induced under dehydration stress in cassava tissues (Lokko et al., 2007). HD-Zip genes are also very important for tolerance to dehydration in the root and leaves of resurrection plant (Deng et al., 2002). Likewise, ATHB7 and ATHB6 genes were found to be rapidly induced by abscisic acid (ABA) (Soderman et al., 1999; Soderman et al., 1996). ABA mediates the response of plants to many abiotic stresses such as cold, salt and drought stresses. Similarly, AtHB7 is found 14.6 times upregulated in salt stressed roots of *Arabidopsis* (Bio-Array Resource, University of Toronto, http://bar.utoronto.ca/welcome.htm). However, recent data suggested that HAT22 (At4g37790), an ortholog of IbHB1, is a master regulatory factor that integrates both carbon and light signals to control genes involved in amino acid metabolism, carbon metabolism and

gycolysis/gluconeogenesis (Gutierez et al., 2008). In *Arabidopsis,* ATHB7 is predicted to interact with other members of HD-Zip proteins (ATHB5), and other transcription factors predominantly from Agamous-like and KNOX families, totaling a network of up to 150 proteins [Figure 2.4, built using the *Arabidopsis* Interaction Viewer tool http://bar.utoronto.ca/welcome.htm) and Predicted *Arabidopsis* Interactome Resource (PAIR)]. The KNOX gene products that likely interact with ATHB7 are encoded by KNAT1 (Knotted-like from *Arabidopsis thaliana*), KNAT2 and KNAT3 genes, which are orthologs of the sweetpotato genes Ibkn2 and Ibkn3. Ibkn2 and Ibkn3 are sweetpotato genes associated with the onset of storage root formation (Tanaka et al., 2008). ATHB5 protein has been demonstrated to interact with ATHB7 protein in vivo. Interestingly, analysis of the *Arabidopsis* transcriptome by the gene co-expression analysis toolbox (GeneCAT, http://genecat.mpg.de/) (Mutwil et al., 2008) reveals that among the three top proteins co-expressed with HAT22 are dehydrin (ERD14), ATHB12 and a jasmonate-zim-domain protein 12 (At5g20900). Dehydrins are part of final targets of TFs during the response associated with drought stress. HD-zip genes are not solely implicated in response to abiotic but also in growth processes and developmental processes. For example, transgenic plants that constitutively expressing ATHB7 showed a reduction in elongation of the leaf and the inflorescence (Hjellström et al., 2003). A noteworthy observation was that sweetpotato plants under 5 or 10 days of drought stress showed a delay in growth, and at the time of harvest (14 days) the roots from stressed plants resemble the adventitious roots seen in sweetpotato plants during early stages of growth. ATHB6 appears to have a function related to cell division and/or differentiation in developing organs (Soderman et al., 1999). High expression of the cotton HD-Zip (GhHB1) gene has been seen in early stages of root growth and a lessening as roots develop further (Ni et al., 2008). In the same study, GbHB1 was found to be expressed mostly in roots

and hypocotyls, and induced by both ABA and salt stress, demonstrating the roles of HD-Zip in morphogenic processes and stress responses. In summary, this data suggest that homeobox leucine genes in sweetpotato might have similar roles as in other plants and specific roles during morphogenesis of storage roots.

The sweetpotato gene IbCRF1 that showed the second highest increase in expression (~ 6-fold) (Table 2.4) in adventitious roots from drought stressed plants is related to AP2(Apetala2)/EREBP (ethylene-responsive element binding protein) genes (Riechmann and Meyerowitz, 1998). The expression of CRF-like genes has not been shown previously in sweetpotato. On the contrary, IbCRF1 and IbCRF2 did not show significant changes in expression due to both drought and salt stress (Table 2.4, fold rates ~1) in both roots and leaf tissues. Many members of AP2/EREBP gene family work in response to environmental stimuli such as abiotic stresses (Chen et al., 2007; Kim et al., 2008; Kizis et al., 2001; Xiong et al., 2002) and also under biotic stress (Lin et al., 2007). IbCRF1 sequence is closely matched to a subfamily of cytokinin response genes (Rashotte and Goertzen, 2010). Cytokinins are needed for the induction of storage root organs in sweetpotato (Eguchi and Yoshida, 2008). IbCRF1 represents a new AP2/EREBP member family, which also includes the sweetpotato dehydration response gene swDREB1, was found upregulated in fibrous roots (Kim et al., 2008) in response to many abiotic stresses such as drought, salt and methyl viologen. Although further characterization in sweetpotato is needed, our work demonstrated that a CRF1-like gene has a stress-related response. IbDREB1 identified in this study is different from the swDREB1 gene previously found upregulated in several tissues of sweetpotato (Kim et al., 2008).

Homeobox Bell and KNOX I genes. Three KNOX (knotted1-like homeobox) genes, Ibkn1, Ibkn2 and Ibkn3 have been previously found associated with the formation of storage

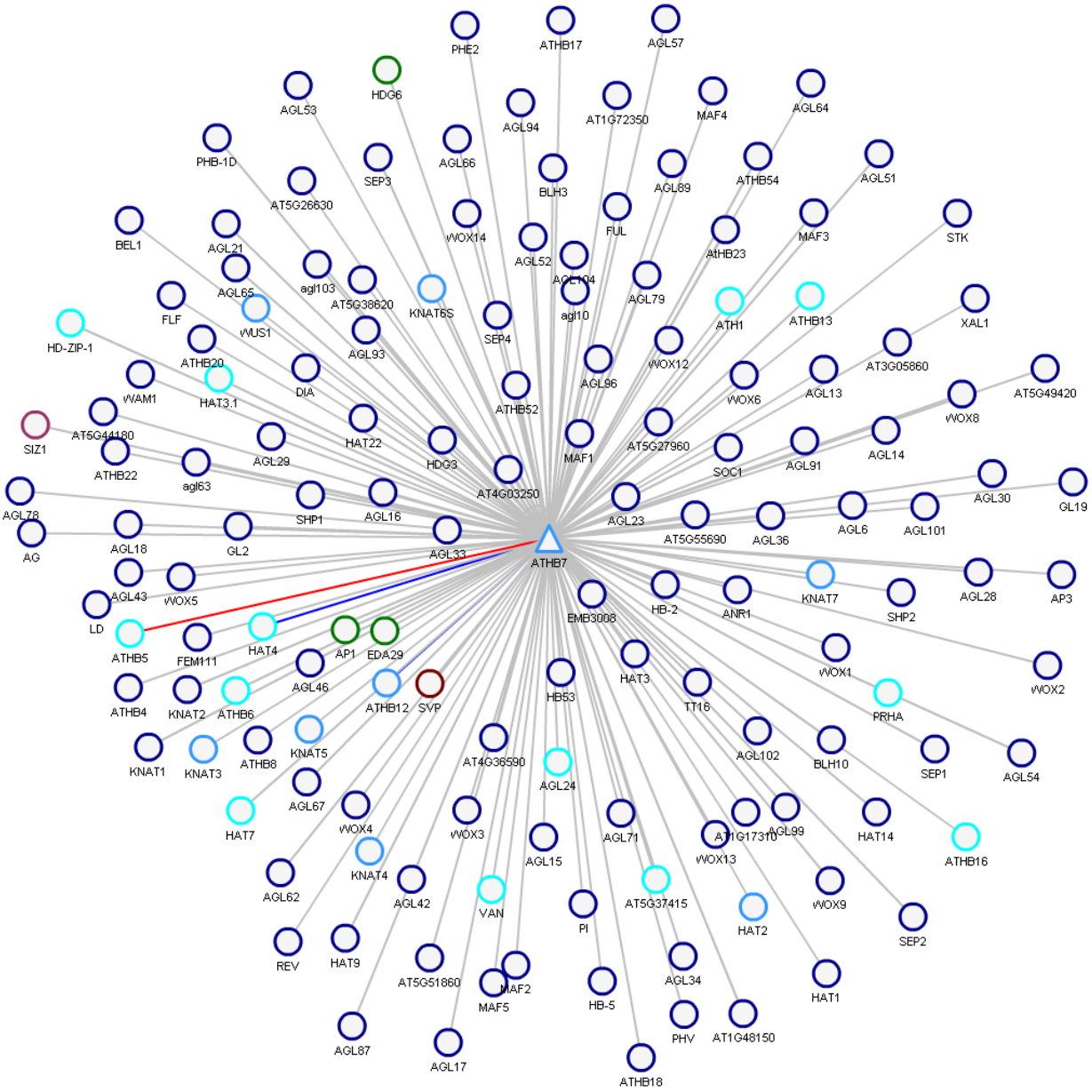


Figure 2.4. Interaction network of proteins predicted to interacts with ATHB7 from Predicted *Arabidopsis* Interactome Resource tool (http://www.cls.zju.edu.cn/pair/faces/home.xhtml) (http://www.cls.zju.edu.cn/pair/faces/searchinteraction.xhtml)

roots formation (Tanaka et al., 2008). BELL (BEL1-Like) and KNOX genes were upregulated ~ 1.5 to 2 times in 2-week old adventious roots in response to drought stress (Table 2.4). The designated names of the sweetpotato BELL genes are IbBEL1, IbBEL2 and IbBEL3. Salinity repressed the expression of IbBEL2. Sequence analysis identified two variants of Ibkn3 (termed

Ibkn3a and Ibkn3b) (please see Chapter 3). Ibkn3a showed enhanced expression in adventitious roots from both drought stress (5 DAT) and salt stress with ~ 1.5 to 2-fold changes over control (Table 2.4), whereas Ibkn3b was slightly downregulated in roots under drought stress (0.88-fold). The exact mechanism and specific role of KNOX and BELL genes is unknown and it is worth pursuing further functional characterization and expression of these genes in sweetpotato. BELL and KNOX gene action in potato tuberization involves their partnering (Chen et al., 2003); interestingly BELL products appear to be involved in a range of roles related to photosynthesis and wound response due to mechanical or insect damage in different tissues of potato (Chatterjee et al., 2007), and to confer disease resistance (Luo et al., 2005). Since KNOX genes act by affecting hormone levels at the site of action, it is significant that we also identified a second type of gene which, according to annotation, is associated to cytokinin signaling genes such as IbCRF1 and IbCRF2. Further work is required to validate the response of these genes in relation to cytokinins.

Calcium signaling genes: Calcium is a second messenger molecule and has many signaling roles stimulated by external and internal cues (Cheng et al., 2002; Kudla et al., 2010). Our results are consistent with these findings since we found that IbCBP2, an orthologous of *Arabidopsis iqd9* (At2g33990) encoding a calmodulin binding protein, was upregulated in the drought stressed roots (fold change of 2.12). Interestingly we also found decreased transcript levels of two genes, one encoding a calcium-binding protein (IbCBP1), which has the smallest fold change ratio in drought vs. control (0.22) and the other encoding a gene termed IbCDPK3, which has a fold change ratio of 0.76. Salt stress also repressed the expression of IbCBP2 in root and leaf tissues, with fold change ratio (salt stress vs. control) of 0.37 and 0.76, respectively. It is unknown if strict regulation is required for CBP-like genes to trigger storage root formation in

sweetpotato, but the expression of CBP2 in stressed roots at 14 days of growth and the impact of drought stress on the onset of storage roots might resemble patterns of CBP-like genes required for tuberization in potato (Poovaiah et al., 1996; Reddy et al., 2002).

There is a cross talk between calcium and ABA signaling in response to abiotic stresses in plants. For example, CDPKs are involved in ABA signaling in *Arabidopsis* (Zhu et al., 2007); similarly a calcium-binding protein required for ABA signaling is induced by both ABA and drought stress in seeds of *Fagus sylvatica* (Jimenez et al., 2008). The evidence in *Arabidopsis* suggests that the action of the CDPK proteins may be through phosphorylation of two ABA-responsive transcription factors, ABF1 and ABF4 (Zhu et al., 2007). However, a calcium signaling sensor, encoded by a calcineurin-B-like (CBL1) is activated in response to drought and salt stress in *Arabidopsis* (Albrecht et al., 2003; Cheong et al., 2003) independent of ABA signaling. The present study indicates calcium signaling is important in the drought stress response. The putative targets of CBP2 and CDPKs are unknown in sweetpotato. However, stress responsive genes, such as Rd29A/B, KIN1/2, and RD22, as downstream targets of *Arabidopsis* CBL1 signaling are documented (Cheong et al., 2003).

Other regulatory genes: Inconsistent results were obtained for SnRK and TAP-like genes between the two set of samples (replicates) for the drought stress study and hence the results are not described here in detail (Table 2.4). Differences in greenhouse temperature during growth stages and at sampling may have been a factor. Salt stress reduced the expression of both SnRK and TAP-like genes in roots with fold changes of 0.67 and 0.82, respectively. Similarly, the genes IbMYB, IbGRF2 (along with IbDREB1) did not show significant differences between roots from drought stressed (from 5 DAT treatment) versus control plants. These genes were

included in this study because digital expression analysis in the root transcriptome indicated their enhanced expression in storage root vs lignified roots (unpublished data).

In conclusion, the present work showed drought stress profoundly affects storage root number and development, and genes (IbHB2, IbCRF1 and IbAREB1), not previously documented in sweetpotato, were identified in response to drought stress that were consistent with research in other species. This work furthermore demonstrates, for the first time, that IbBEL1,-2 and -3, Ibkn3a were stress responsive genes, not identified in other crops, and were upregulated in roots under drought stress. IbHB2 was found to respond to both drought and salt stress. Moreover, salt stress negatively affected most of the tested genes in roots by suppressing their expression; only two homeodomain genes Ibkn3a and IbBEL3 showed upregulation, while five transcription factors and one calcium binding protein were strongly repressed. The overall results support the sensitivity of genes associated with the onset of bulking under drought stress and the consequence is a diminished number of storage roots in plants under stress. Upregulation of transcripts of calcium signaling (IbCBP2) and downregulation of post-transcriptional modification genes (IbCDPK3, IbCBP1) are most likely related to the reduced number of storage roots because alteration of these post-transcriptional regulatory activities affect the downstream targets. One consequence could be lack of active enzymes or transcriptions factors.

2.5 Literature cited

Anwar N., Kikuchi A., Watanabe K.N. (2010) Assessment of somaclonal variation for salinity tolerance in sweet potato regenerated plants. African Journal of Biotechnology 9:7256-7265.

Albrecht V., Weinl S., Blazevic D., D'Angelo C., Batistic O., Kolukisaoglu U., Bock R., Schulz B., Harter K., Kudla J. (2003) The calcium sensor CBL1 integrates plant responses to abiotic stresses. Plant Journal 36:457-470.

Belehu T., Hammes P.S., Robbertse P.J. (2004) The origin and structure of adventitious roots in sweet potato (*Ipomoea batatas*). Australian Journal of Botany 52:551-558. DOI: Doi 10.1071/Bt03152.

Boyer J.S. (1982) Plant productivity and environment. Science 218:443-448.

Chatterjee M., Banerjee A.K., Hannapel D.J. (2007) A BELL1-like gene of potato is light activated and wound inducible (1 C OA). Plant Physiology 145:1435-1443.

Chen H., Rosin F.M., Prat S., Hannapel D.J. (2003) Interacting transcription factors from the three-amino acid loop extension superclass regulate tuber formation. Plant Physiology 132:1391-1404.

Chen M., Wang Q.-Y., Cheng X.-G., Xu Z.-S., Li a.-C.L., Ye X.-G., Xia L.-Q., Ma Y.-Z. (2007) GmDREB2, a soybean DRE-binding transcription factor, conferred drought and high-salt tolerance in transgenic plants. Biochemical and Biophysical Research Communications 353:299-305.

Cheng S.H., Willmann M.R., Chen H.C., Sheen J. (2002) Calcium signaling through protein kinases. The *Arabidopsis* calcium-dependent protein kinase gene family. Plant Physiology 129:469-485.

Cheong Y.H., Kim K.N., Pandey G.K., Gupta R., Grant J.J., Luan S. (2003) CBL1, a calcium sensor that differentially regulates salt, drought, and cold responses in *Arabidopsis*. Plant Cell 15:1833-1845.

Constantin R.J., Hernandez T.P., Jones L.G. (1974) Effects of irrigation and nitrogen fertilization on quality of sweet potatoes. Journal of the American Society for Horticultural Science 99:308-310.

Deng X., Phillips J., Meijer A.H., Salamini F., Bartels D. (2002) Characterization of five novel dehydration-responsive homeodomain leucine zipper genes from the resurrection plant Craterostigma plantagineum. Plant Molecular Biology 49:601-610.

Eguchi T., Yoshida S. (2007) Effects of Gas Exchange Inhibition and Hypoxia on Tuberous Root Morphogenesis in Sweetpotato (*Ipomoea batatas* (L.) Lam.). Environment Control in Biology 45:103-111.

Eguchi T., Yoshida S. (2008) Effects of application of sucrose and cytokinin to roots on the formation of tuberous roots in sweetpotato (*Ipomoea batatas* (L.) Lam.). . Plant Root 2:7-13.

Ekanayake I.J., Dodds J.H. (1993) In vitro testing for the effects of salt stress on groth and survival of sweetpotato. Sci. Hort. 55:239-248.

Greig J.K., Smith F.W. (1962) Salinity effects on sweetpotato growth. Agronomy Journal 54:309-313.

Gutierez R.A., Stokes T.L., Thum K., Xu X., Obertello M., Katari M.S., Tanurdzic M., Dean A., Nero D.C., McClung C.R., Coruzzi G.M. (2008) Systems approach identifies an organic nitrogen-responsive gene network that is regulated by the master clock control gene CCA1. Proceedings of the National Academy of Sciences of the United States of America 105:4939-4944.

Hjellström M. (2002) Drought Stress Signal Transduction by the HD-Zip Transcription factors ATHB6 and ATHB7., Acta Universitatis Upsaliensis. Comprehensive Summaries of Uppsala Dissertat ions from the Faculty of Science and Technology 690.

Indira P., Kabeerathumma S. (1988) Physiological response of sweet potato under water stress. 1. Effect of water stress during the different phases of tuberization. Journal of Root Crops 14:37-40.

Jimenez J.A., Alonso-Ramirez A., Nicolas C. (2008) Two cDNA clones (FsDhn1 and FsClo1) upregulated by ABA are involved in drought responses in Fasus sylvatica L. seeds. Journal of Plant Physiology 165:1798-1807.

Kim S.H., Hamada T., Otani M., Shimada T. (2005) Isolation and characterization of MADS box genes possibly related to root development in sweetpotato (*Ipomoea batatas* L. Lam.). Journal of Plant Biology 48:387-393.

Kim S.H., Mizuno K., Fujimura T. (2002) Isolation of MADS-box genes from sweet potato (*Ipomoea batatas* (L.) lam.) expressed specifically in vegetative tissues. Plant and Cell Physiology 43:314-322.

Kim S.H., Song W.K., Kim Y.H., Kwon S.Y., Lee H.S., Lee I.C., Kwak S.S. (2009) Characterization of full-length enriched expressed sequence tags of dehydration-treated white fibrous roots of sweetpotato. Bmb Reports 42:271-276.

Kim Y.H., Yang K.S., Ryu S.H., Kim K.Y., Song W.K., Kwon S.Y., Lee H.S., Bang J.W., Kwak S.S. (2008) Molecular characterization of a cDNA encoding DRE-binding transcription factor from dehydration-treated fibrous roots of sweetpotato. Plant Physiology and Biochemistry 46:196-204.

Kizis D., Lumbreras V., Pages M. (2001) Role of AP2/EREBP transcription factors in gene regulation during abiotic stress. Febs Letters 498:187-189.

Ku A.T., Huang Y.-S., Wang Y.-S., Ma D., Yeh K.-W. (2008) IbMADS1 (*Ipomoea batatas* MADS-box 1 gene) is involved in tuberous root initiation in sweet potato (*Ipomoea batatas*). Annals of Botany 102:57-67.

Kudla J., Batistic O., Hashimoto K. (2010) Calcium Signals: The Lead Currency of Plant Information Processing. Plant Cell 22:541-563.

Lee Y.H., Chun J.Y. (1998) A new homeodomain-leucine zipper gene from *Arabidopsis thaliana* induced by water stress and abscisic acid treatment. Plant Molecular Biology 37:377-384.

Ley T.W., Stevens R.G., Topielec R.R., Neibling W.H. (1994) Soil water monitoring and measurement, Pacific Northwest Extension Publication 475, Washington State Universiy.

Lin R., Zhao W., Meng X., Peng Y.-L. (2007) Molecular cloning and characterization of a rice gene encoding AP2/EREBP-type transcription factor and its expression in response to infection with blast fungus and abiotic stresses. Physiological and Molecular Plant Pathology 70:60-68.

Lokko Y., Anderson J.V., Rudd S., Raji A., Horvath D., Mikel M.A., Kim R., Liu L., Hernandez A., Dixon A.G.O., Ingelbrecht I.L. (2007) Characterization of an 18,166 EST dataset for cassava (Manihot esculenta Crantz) enriched for drought-responsive genes. Plant Cell Reports 26:1605-1618.

Luo H., Song F., Goodman R.M., Zheng Z. (2005) Up-regulation of OsBIHD1, a rice gene encoding BELL homeo-domain transcriptional factor, in disease resistance responses. Plant Biology 7:459-468.

Mao L.X., Jett L.E., Story R.N., Hammond A.M., Peterson J.K., Labonte D.R. (2004) Influence of drought stress on sweetpotato resistance to sweetpotato weevil, Cylas formicarius (Coleoptera : Apoinidae), and storage root chemistry. Florida Entomologist 87:261-267.

Mutwil M., Obro J., Willats W.G.T., Persson S. (2008) GeneCAT - novel webtools that combine BLAST and co-expression analyses. Nucleic Acids Research 36:W320-W326.

Ni Y.X., Wang X.L., Li D.D., Wu Y.J., Xu W.L., Li X.B. (2008) Novel cotton homeobox gene and its expression profiling in root development and in response to stresses and phytohormones. Acta Biochimica Et Biophysica Sinica 40:78-84.

Noh S.A., Lee H.S., Huh E.J., Huh G.H., Paek K.H., Shin J.S., Bae J.M. (2010) SRD1 is involved in the auxin-mediated initial thickening growth of storage root by enhancing proliferation of metaxylem and cambium cells in sweetpotato (*Ipomoea batatas*). Journal of Experimental Botany 61:1337-1349.

Olsson A.S.B., Engstrom P., Soderman E. (2004) The homeobox genes ATHB12 and ATHB7 encode potential regulators of growth in response to water deficit in *Arabidopsis*. Plant Molecular Biology 55:663-677.

Pardales J.R., Yamauchi A. (2003) Regulation of root development in sweetpotato and cassava by soil moisture during their establishment period. Plant and Soil 255:201-208.

Poovaiah B.W., Takezawa D., An G., Han T.J. (1996) Regulated expression of a calmodulin isoform alters growth and development in potato. Journal of Plant Physiology 149:553-558.

Prabawardani S., Mark J., Ross C., Holtum J. (2004) Identification of drought tolerant sweet potato (*Ipomoea batatas* (L.) Lam) cultivars, in: R. Fischer (Ed.), New directions for a diverse planet.Proceedings of the 4th International Crop Science Congress, Brisbane, Australia.

Rashotte A.M., Goertzen L.R. (2010) The CRF domain defines Cytokinin Response Factor proteins in plants. BMC Plant Biology 10.

Reddy A.S.N., Day I.S., Narasimhulu S.B., Safadi F., Reddy V.S., Golovkin M., Harnly M.J. (2002) Isolation and characterization of a novel calmodulin-binding protein from potato. Journal of Biological Chemistry 277:4206-4214.

Ricardo J. (2011) Screening sweetpotato (*Ipomoea batatas* L.) for drought tolerance and high ß-carotene content in Mozambique, School of Agricultural Sciences and Agribusiness,Faculty of Science and Agriculture, University of KwaZulu-Natal Pietermaritzburg. pp. 110.

Riechmann J.L., Meyerowitz E.M. (1998) The AP2/EREBP family of plant transcription factors. Biological Chemistry 379:633-646.

Rus A., Baxter I., Muthukumar B., Gustin J., Lahner B., Yakubova E., Salt D.E. (2006) Natural Variants of AtHKT1 Enhance Na+ Accumulation in Two Wild Populations of *Arabidopsis*. PLoS Genet 2:e210.

Saraswati P., Johnston M., Coventry R., Holtum J. (2004) Identification of drought tolerant sweet potato (*Ipomoea batatas* (L.) Lam) cultivars., in: R. Fischer (Ed.), Proceedings of the 4th International Crop Science Congress, Brisbane, Australia, 2004.

Schafleitner R., Tincopa L.R., Palomino O., Rossel G., Robles R.F., Alagon R., Rivera C., Quispe C., Rojas L., Pacheco J.A., Solis J., Cerna D., Kim J.Y., Hou J., Simon R. (2010) A sweetpotato gene index established by de novo assembly of pyrosequencing and Sanger sequences and mining for gene-based microsatellite markers. Bmc Genomics 11: 604.

Shannon M.C., Grieve C.M. (1999) Tolerance of vegetable crops to salinity. Scientia Horticulturae 78:5-38.

Soderman E., Hjellstrom M., Fahleson J., Engstrom P. (1999) The HD-Zip gene ATHB6 in *Arabidopsis* is expressed in developing leaves, roots and carpels and upregulated by water deficit conditions. Plant Molecular Biology 40:1073-1083.

Soderman E., Mattsson J., Engstrom P. (1996) The *Arabidopsis* homeobox gene ATHB-7 is induced by water deficit and by abscisic acid. Plant Journal 10:375-381.

Tanaka M., Kato N., Nakayama H., Nakatani M., Takahata Y. (2008) Expression of class I knotted1-like homeobox genes in the storage roots of sweetpotato (*Ipomoea batatas*). Journal of Plant Physiology 165:1726-1735.

Togari Y. (1950) A study of tuberous root formation in sweet potato. 68:1–96. Bull. Natl. Agric. Exp. Stn. 68:1-96.

Villordon A., LaBonte D., Solis J., Firon N. (2012) Characterization of Lateral Root Development at the Onset of Storage Root Initiation in 'Beauregard' Sweetpotato Adventitious Roots. Hortscience 47:961-968.

Villordon A.Q., La Bonte D.R., Firon N., Kfir Y., Pressman E., Schwartz A. (2009) Characterization of Adventitious Root Development in Sweetpotato. Hortscience 44:651-655.

Wilson L.A., Lowe S.B. (1973) Anatomy of Root System in West-Indian Sweet Potato (*Ipomoea batatas* (L) Lam) Cultivars. Annals of Botany 37:633-&.

Xiong L.M., Schumaker K.S., Zhu J.K. (2002) Cell signaling during cold, drought, and salt stress. Plant Cell 14:S165-S183.

You M.K., Hur C.G., Ahn Y.S., Suh M.C., Jeong B.C., Shin J.S., Bae J.M. (2003) Identification of genes possibly related to storage root induction in sweetpotato. Febs Letters 536:101-105.

Zhu S.-Y., Yu X.-C., Wang X.-J., Zhao R., Li Y., Fan R.-C., Shang Y., Du S.-Y., Wang X.-F., Wu F.-Q., Xu Y.-H., Zhang X.-Y., Zhang D.-P. (2007) Two calcium-dependent protein kinases, CPK4 and CPK11, regulate abscisic acid signal transduction in *Arabidopsis*. Plant Cell 19:3019-3036.

Chapter 3. Comparative *De Novo* Assembly and Analysis of the Sweetpotato Transcriptome Identifies Candidate Genes for Regulatory Roles and Calcium Signaling in the Onset of Storage Root Formation

3.1 Introduction

Storage root formation is the most important growth process in sweetpotato. Unfortunately, its complex genetic structure has hindered our ability to understand this process. Sweetpotato genes involved in storage root are known. Mostly, these represent transcription factors (TFs), with KNOX (Ibkn1,-2,-3) (Tanaka et al., 2008) and MADS-box [IbMADS1 (Ku et al., 2008), IbAGL17, IbAGL20, and IbMADS79 (Kim et al., 2005), SDR1 (Noh et al., 2010) and IbMADS3, IbMADS4 (Kim et al., 2002a)] gene families. KNOX and MADS-box genes act as homeotic genes determining organ identity and meristem identity genes by regulating the cell fate during cell differentiation. Apart from TFs, genes encoding storage proteins, cell wall biogenesis, cell division, carbohydrate metabolism and molecular signaling have been associated with the progression of fibrous roots into storage roots (Desai, 2008; McGregor and LaBonte, 2006; Nagata and Saitou, 2009). Expression of sporamin and β-amylase genes occurs in developed storages roots and their value as genetic markers are only as indicators of the differentiation of white fibrous roots from a storage root. The main role of sporamin and β-amylase proteins are as storage proteins, with secondary defensive roles for sporamin (Yeh et al., 1997) and β-amylase has a role in sprouting (Chen et al., 2004b). Both sporamin and β-Amylase account for 60 to 80% and 5% of the total soluble protein in the fleshy roots, respectively. Another factor identified in the thickening of adventitious roots is the unsuspected interaction of starch synthesis via ADP-glucose pyrophosphorylase (AGPase), cell proliferation and nitrogen-mediated signaling (Kim et al., 2002b). Further, hormone and sugar signaling appears to be important for the onset of storage roots. For example, some genes coding for family NAC (NAM/ATAF/CUC) proteins, family of no apical meristem (NAM)-like proteins, and genes

associated with response to auxin were found differentially expressed in storage roots (McGregor, 2006; You et al., 2003); NAM/NAC genes are factors that act in the downstream pathway of auxin signaling with NAC1 product documented to be key for lateral root formation and auxin signaling (Xie et al., 2000); likewise ethylene responsive genes such as jasmonate and ethylene responsive factor (JERF3) (McGregor and LaBonte, 2006) and several AP2/EREBP transcription factors genes showed differential expression between fibrous and storage roots (Desai, 2008). Calcium (Ca^{2+}) signaling has been suggested to mediate cross-talk between sucrose and other stimuli in the expression of sporamin and β-Amylase genes (Ohto et al., 1995). Cytokinins and sucrose are among factors that triggers the expression of genes associated with the thickening of adventitious roots. Accumulation of the cytokinin trans-zeatin riboside (t-ZR) (Nakatani and Komeichi, 1991), and external application of sucrose and t-ZR have been correlated with the bulking of roots (Eguchi and Yoshida, 2008). There is evidence that the suggests that KNOX genes may mediate meristematic activity (cell division) found through promotion of the synthesis of cytokinins and the repression of biosynthesis of the growth regulator Gibberellin acid (GA) (Jasinski et al., 2005). Despite these studies, the molecular events underlying storage root development remains unknown. Certainly, it is worthy to study genes associated with molecular signaling, transcriptional activities and in response to stimuli.

The advent of next generation sequencing (NGS) technologies has been invaluable in describing and characterizing genes in many crops. Among them, 454 pyrosequencing and Illumina deep sequencing platforms are cost-effective for characterizing transcriptomes (Wall et al., 2009). 454 libraries of transcriptomes have increased genomic studies of non-models plants. Currently, around 23K Expressed sequence tags (ESTs) are available at GenBank for sweetpotato. NGS data has enriched the knowledge of the transcriptome of sweetpotato

(Schafleitner et al., 2010; Tao et al., 2012; Wang et al., 2010; Xie et al., 2012). However, no reference genome exits, so utilization of the NGS data toward the characterization of genes is less in comparison to other crops. Moreover, there is no known procedure for analysis of the transcriptome derived from NGS from non models plants and hexaploid species like sweetpotato.

Sweetpotato is propagated vegetatively by planting vine cuttings from which adventitious roots arise from the underground stem just after planting. Most of these roots have the potential to form storage roots (Villordon et al., 2009) but the fate of these adventitious is decided by both genetic and environmental factors. During the first stages of growth the roots develop into either the fleshy roots (i.e. storage roots, often wrongly referred to as tuberous roots), or the thin roots (i.e. non storage roots, hereinafter referred to as fibrous roots). Villordon et al. (2009) showed that the majority of 'Beauregard' (86%) and 'Georgia Jet' (89%) storage roots sampled at 60 to 65 DAT were traced directly to adventitious roots extant at 5 to 7 days after transplanting (DAT). Typical anatomical features (five or more protoxylem elements) associated with storage root development are observed in these initially established roots. A second feature recognized in these adventitious roots is the notably higher density of lateral roots (Villordon et al., 2012). Lignification is a third feature observed at advanced stages of growth (2-4 weeks) when the initiation has already occurred and the thickening of the storage roots is halted due to the lignification (Firon et al., 2009; Villordon et al., 2009). These characteristics of storage forming adventitious roots permit straightforward identification at a very early development for genomic expression research long before adventitious root exhibit thickening. One approach of a characterization of the sweetpotato root transcriptome was developed by application of NGS to initiating storage roots and non-storage roots (N. Firon, personal communication, 2010).

The objectives of the current study were twofold: first, build consensus unigenes using a combined set of sweetpotato transcriptome data; second, combine genomic data from the current study with existing gene indexes to identify novel genes putatively associated with developing of storage roots. An ancillary objective was to compare different available assemblers in *de novo* transcriptome assembly, to estimate the number of unigenes in sweetpotato and to demonstrate their utility to find putative genes associated with the storage root development.

3.2 Materials and methods

3.2.1 Biological material and RNA extraction

In this work only those young roots forming storage root (maximum diameter ranging from 2.5 to 3.5mm) that represent expanding storage root was used as storage root (SR); also included were thick, pigmented roots (maximum diameter >1.5-2.0 mm), which were called as developing storage root (DSR) or initiating storage root because most of them have the anatomical features of roots forming storage root. Most young storage roots are synonymous to initiated storage root that already have developed a complete primary vascular cambium and the anomalous cambium and hereon they are identifiable at 4 weeks of growing, in agreement with previously observed data in two varieties of sweetpotato (Villordon et al., 2009). Non-storage roots will be used as synonym of white fibrous roots, an agreement with the above study the anatomy of this roots are variable alike adventitious roots and also with high lignification of the stele. Adventitious roots that started but stopped the thickening with a uniform thickening are called pencil roots, however they are only identifiable at later stages of storage root development.

RNA was extracted using RNAzol (Molecular Research Center, Cincinnati, Ohio) from fibrous (FR), young storage roots (SR) and initiating developing storage roots (DSR) of 4-week-old plants followed by on-column DNase I digestion and RNA clean up using the RNeasy MinElute cleanup kit (Qiagen,Valencia, CA). RNA quality and RNA yield were determined

taking the A260:A280 and A260:A230 ratios using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) with a 1.5 µl sample and by agarose gel electrophoresis. These samples were used for quantitative reverse transcription-polymerase chain reaction (qRT-PCR) assay using SYBR green supermix (Biorad, CA) to quantify the expression of the selected genes. In addition, leaf samples were processed to extract RNA for semi-quantitative reverse transcription PCR (sqRT-PCR).

3.2.2 Transcriptome data, sequence analysis and assembly

In this chapter, we define a unique transcript (UT) as an EST singleton, single read or a contig assembled after EST/read assembly and clustering. The term unigene is also considered as a synonym to a UT and is used from hereon in this context. Although reads refer commonly to any sequence derived from the use of either the Illumina or the Roche-454 pyrosequencing technologies, we used the term "read" to also refer to each EST or singlets.

A combined dataset of raw sequences from multiple sources were used and is summarized in Table 3.1. One important dataset consisted of Illumina HiSeq 2000 (Solexa) reads from initiating storage roots (14,780,229) and lignified adventitious roots (17,703,982) of sweetpotato roots, totaling 32,484,160 reads. In addition, 454 pyrosequencing of an equimolar mixture of transcripts from both initiating and lignified adventitious roots yielded 524,403 sequences. Both Illumina and 454 NGS datasets of root was provided by N. Firon, and first time published in the present work. Prior to assembly, sequences were cleaned to vector and adaptor removal, trimmed of bad quality regions, short sequence filtering and removal of rRNA contaminants. Illumina data set (32,484,160 reads) was *de novo* assembled using AbySS and using 72 k-mer value, which we found was substantially better than lower values. A total of 514,202 contigs were generated (102.9 bp average size and total length of the contigs is

52,918,328 bp) from the AbySS assembly, and after cleaning using Seqclean program (http://seqclean.sourceforge.net/) only 201,447 contigs (126 bp average size total length of the contigs is 25,515,373 bp) were passed to the next step. Contigs were labeled by the string "abyssCtg_#", where # is a numerical index. These contigs were combined with public ESTs from sweetpotato (23,406 sequences, average length 552.1 bp) and three 454-derived libraries from stem (436,899 sequences, average length 316.2 bp), leaf (87,307 sequences, average length 243.9 bp) and from root (combined initiating storage roots and lignified roots, 524,403 sequences, average length 310.7 bp) tissues. Sweetpotato ESTs from Genbank were screened with cross_match against the Univec database and trimmed with Lucy and further screened with SeqClean and compared to rRNA sequences, reducing the final number to 22,362 sequences. The leaf, root and stem 454 transcriptome sequences were trimmed using NGS Backbone with screens for adaptor sequences. The files were further screened with SeqClean and compared to rRNA sequences, yielding 84,608 leaf sequences, 475,083 root sequences, and 327,510 stem sequences. Consensus sequences by clustering an assembly were created using Roche 454's Newbler (version 2.5) using default values of the parameters and CAP3 was using a minimum of 40-bp overlap length with 80% identity.

3.2.3 Functional annotation of the sequences

Sequence annotation was carried out by blast analysis (BLASTX and BLASTN). All unigenes were searched against protein databases UniProtKB (www.uniprot.org, release UniProt release 2011_04) April 5, 2011) and TAIR10 protein database (http://www.*Arabidopsis*.org/index.jsp) using a threshold of Expect (E) value<1e-6. In addition, Gene Ontology (GO) analysis was conducted using b2go tool to classify the unigenes according

to its GO terms. GO terms were also found via the UniprotKB accession associated to each unigene.

A set of 30 genes were selected for gene expression profiling including known genes associated with the formation of storage roots (Ibkn2, Ibkn3, IbAGP1a and β-Amylase) and the reference gene encoding an elongation factor1-alpha protein (IbEF1a). Sequences of gene-specific oligonucleotide primers were generated by the use of PRIMER3. A BLASTN search was done against non redundant nucleotide sequences at NBCI (http://www.ncbi.nlm.nih.gov/) to verify the novelty of the selected sequences and to further annotation. The detailed summary of the primers, expected product length, sequence accession and annealing temperature are in Table 3.2.

Table 3.1. Summary of sequences used to build d*e novo* consensus sequences

Tissue	Number of EST/Reads	Technology	Source
Leaves	84,608	454 pyrosequencing	Unpublished (R. Schafleitner, 2010)
Leaves	1,080	Sanger sequencing	dbEST (GenBank release 183)
Stems	327,510	454 pyrosequencing	Unpublished (R. Schafleitner,2010)
Others	1,626	Sanger sequencing	dbEST (GenBank release 183)
Roots	20,686	Sanger sequencing	dbEST (GenBank release 183)
Roots	475,083	454 pyrosequencing	Unpublished (N.Firon, 2010)
Roots	32,484,160	Illumina HiSeq 2000	Unpublished (N.Firon, 2010)
Total	33,397,459		

3.2.4 Quantitative RT-PCR analysis and reverse transcription PCR of selected unigenes

Primers were tested for amplification of 26 new candidate loci (excluding β-amylase, IbAGP1a and Ibkn2). The rationale for selection of the genes was because either they had highest number of reads from initiating storage roots or having key roles in developmental processes [i.e. transcription factor, molecular signaling or control of gene expression). BELL genes (IbBEL1,-2,-3,-4,-5,-6) and calcium binding protein encoding genes (IbCBP1,-2, -3, and

IbCDPK2, IbCDPK3) were selected because of its role in developmental processes in tissues and not necessarily because they derived from transcripts that were abundant in reads from the root transcriptome only. Known mechanism of tuberization in potato supported the selection of these genes. Two-step real-time PCR was performed for evaluating gene expression profiles. cDNA was synthesized from 2 µg of total RNA using the iScript cDNA synthesis kit (Bio-Rad, Hercules, CA), including 4 µl iScript reaction mix and 1 µl iScript reverse transcriptase and incubating 5 minutes at 25° C, 30 minutes at 42° C, and 5 minutes at 85° C. The iScript cDNA synthesis kit (Bio-Rad, Hercules, CA) contains an optimized blend of oligo (dT) and random primers. The cDNA was diluted 1:3 with nuclease-free water and qRT-PCR was performed using iQ^{TM} SYBR Green supermix (Bio-Rad, Hercules, CA), according to the manufacturer's instructions. All the primers used are listed in Table 3.2. For each reaction, 2 µl of diluted cDNA was mixed with 10 µl of 2× SYBR green reaction mix (iQ^{TM} SYBR Green supermix), and 3.8 pmol of the forward and the reverse primers were added to make a final volume of 20 µl. PCR was carried out in a MyiQ real-time PCR analysis system (Bio-Rad, Hercules, CA). The conditions for the PCR amplification were as follows: polymerase activation at 95°C for 3 min; followed by 45 cycles of 95°C for 10 s, 60°C for 20 s, and 72°C for 20 s. The fluorescence signal was measured once every 1°C. Negative PCR controls (no cDNA template) were included in all reactions to detect possible contamination and the specificity of the primer amplicons was checked by analysis of a melting curve. The Ct values were converted into fold change ratio using the elongation factor 1-alpha, IbEF1a, as the reference gene and the delta-delta-Ct method. The fold changes were done comparing the SR, DSR vs. FR. Amplicons of the predicted size were confirmed using agarose gel electrophoresis.

Semi-quantitative RT-PCR amplification was conducted on undiluted cDNA root samples (SD, DSR, FR) including an additional cDNA set of leaf tissues from same plants in 20 µl reactions. Each reaction consisted of 2µl undiluted cDNA extract, 0.75 U Taq DNA Polymerase and 1× KCl Reaction Buffer (Bioline, Taunton, USA), 1.5 mM $MgCl_2$, 250 µM dNTPs, and 0.5 mM amplification primers. PCR was carried out in a MyCycler ™ thermal cycler (BioRad, Hercules, CA) under the following conditions: initial denaturing at 95°C for 4 min, 35 cycles of 95°C for 30 sec-variable annealing temperature for 45 sec-72°C for 1 min, followed by a final extension at 72°C for 5 min. ADP-glucose pyrophosphorylase (IbAGP1a, accession AB271011) and β-Amylase (BMY1 gene, accession D01022.1) were included as positive controls. See Table 3.2 for annealing temperatures. PCR products were analyzed by electrophoresis on a 2.2 % agarose gel containing ethidium bromide (0.5 mg/ml).

3.3 Results

To build a consensus set of transcripts of sweetpotato and a comparative analysis against current sweetpotato gene indexes, sequencing data from different sources (tissues) were combined (Table 3.1). A deep coverage of transcripts from sweetpotato roots was used: 32 million sequences from Illumina HiSeq 2000 (Solexa) from initiating and lignified adventitious roots and 524,403 sequences (after cleaning up) from 454 pyrosequencing of a pooled equimolar concentration of cDNAs from initiating and lignified adventitious roots (unpublished methodology, N. Firon, 2011). ABySS was used to assemble the reads from Illumina. The resulting contigs (514,202) were reduced in number by excluding <80 nucleotides sequences and those matching rRNA sequences, yielding 201,447 contigs which were combined with cleaned and trimmed sequences consisting of public ESTs from sweetpotato (22,436 sequences) and three 454-derived libraries from stem (327,510 sequences), leaf (84,608 sequences) and from root (combined initiating storage roots and lignified roots, 475,083 sequences) tissues. The leaf,

root and stem 454 transcriptome sequences were trimmed using NGS Backbone with screening for adaptor sequences. The files were further screened with SeqClean to eliminate rRNA, analyzed with cross_match against the Univec database and trimmed with Lucy.

Sweetpotato ESTs from Genbank were screened with cross_match against the Univec database and trimmed with Lucy and further screened with SeqClean and compared to rRNA sequences, reducing the final number to 22,362 sequences. The leaf, root and stem 454 transcriptome sequences were trimmed using NGS Backbone with screens for adaptor sequences. The files were further screened with SeqClean and compared to rRNA sequences, yielding 84,608 leaf sequences, 475,083 root sequences, and 327,510 stem sequences.

Next, two assemblers were used get a final assembly, Newbler (2.5.3) and CAP3 approaches were used under default and customized parameters respectively. Contigs from ABySS, 454 pyrosequencing data and ESTs were used as input. The 454 pyrosequencing data from roots was also independently assembled into a unigene using MIRA software (Nurit Firon, personal communication) and referred to as *PBL assembly*. Unique transcripts from CAP3 and Newbler assembly are referred to as *cap3p80v3* and *newblerv1* assemblies, respectively. The summary of the description of the assemblies are in Table 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7.

The numbers of unique transcripts assembled for sweetpotato are presented in Table 3.3. CAP3 based assembly and clustering of the combined dataset (33,397,459 sequences) resulted in 56,270 contigs and 18,961 singlets, forming 75,231 unique transcripts (UTs). Newbler based assembly resulted in 35,069 contigs and 42,594 singlets, forming 77,663 UTs. The number of UTs of at least 100 bp from CAP3 assembly was 74,849. Although we made a third assembly, referred to as *cap3p90 assembly*, of the combined data set using CAP3 with 90% of identity in overlapping regions, the reduction of redundancy and length of consensus sequences appears to

be better using the 80% identity in overlapping. The average length of unique transcripts, UTs, of the sweetpotato assemblies is 594.5 bp for cap3p80v3 and 504.1 bp for newblerv1(Table 3.3, Figure 3.1); the N50 of all unigenes from *cap3p80v3* and *newblerv1* assembly are 633bp and 480bp, respectively. The vast majority of contigs, 60% (33,762 sequences) of cap3p80v3 assembly and 49.58% (17,386 sequences) contigs of newblerv1 assembly have at least 0.5Kbp (Table 3.3); 85% to 95% of contigs had a length ranging from 300 to 1000pb (Figure 3.1).

3.3.1 Novel unique sequence for root libraries

To assess the contribution of unique sequences from our 454 libraries from mixed root tissues (initiating storage roots and lignified roots), the unigene sequences from sweetpotato based on these sequences (PBL assembly) (unpublished data, Nurit Firon, personal communication, 2010) were analyzed by BLASTN versus the sweetpotato gene index (SPGI) (Schafleitner et al, 2010) and the set of sweetpotato UTs from the Plant Genome Database, Plant GDB (http://www.plantgdb.org/) (here referred to as SPPGDB). SPPGDB is built with ESTs sequences deposited at GenBank up to April 2009, we identified 88.4% (20,686 out of 23,392 sequences) of cleaned ESTs from GenBank were derived from roots (Table3.1) and the rest of ESTs were derived from leaf (1080), tissue culture cells (1411) and plantlets (215). In addition, 37,697 and 78,675 UTs of the *PBL assembly* (from 454 root libraries) did not have a matching sequence in the SPGI and in SPPGDB, respectively. The inclusion of 454 sequences from roots represents a gain of 56.75% (37,697/66,418) in the amount of UTs compared to SPGI and they represent novel UTs that increase the documentation of novel genes. Similarly, comparing the PBL assembly to the SPPGDB the amount of new unigenes are 78,675 sequences, and increased of ~6.4 times of the number of cleaned unigenes of SPPGDB (12,306) (Table 3.3). However, since the PBL assembly is build with 454 sequencing libraries sequences coming from specific

segments of transcripts such as 3'-UTR and terminal end of coding sequences, their consensus transcripts may represent a component of a whole transcript and lack of coverage of all available sequences. This was one justification of the present study which motivated us to a *de novo* assembly by combining all transcripts available for sweetpotato.

To assign putative function to UTs of each assembly, BLASTX search was done against the UniprotKB databases (Schneider et al., 2009) and *Arabidopsis* database (TAIR10). Summary of BLASTX results are in tables 3.5, 3.6 and 3.7. Putative functions were assigned to 69.55% (53,322 UTs) and 60.48% (46,967 UTs) of *cap3p80v3* and *newblerv1* assemblies, respectively, when searched against UniprotKB (Table 3.5) (E-value <=1E-6); while a lower fraction of sequences, 63.98% to 57.04% had a hit against *Arabidopsis*. Assignment of functions based on similarly to *Arabidopsis* was important to improve the functional annotation and the naming of the identified genes. Likewise, we carried out the annotation of the PBL assembly and found that it had 67.92% (68,944 UTs) and 61.11% (62,028) of TUs matching UniprotKB and a TAIR10, respectively. Most of the contigs in *cap3p80v3* (72.46%, 40,776 contigs) and in *newblerv1*(62.7%, 21,988 contigs) assemblies had a hit in UniprotKB (Table3.6) and a >=200 bp of length (Table 3.6). These contigs represented 99.9% and 95.98% of the total annotated contigs (Table 3.7) in *cap3p80v3* and in *newblerv1*, respectively. A total of 7,833 annotated contigs in *cap3p80v3* (and 6,371 in *newblerv1*) had a length of 1Kbp or greater (Table 3.6). The detailed distribution of sequences annotated in our assemblies and other existing assemblies of sweetpotato are in Table 3.6 and Table 3.7. In general, *cap3p80v3* assembly had the largest fraction of sequences (either unigenes or contigs) with a hit in UniprotKB and with the longest size (>= 1Kb).

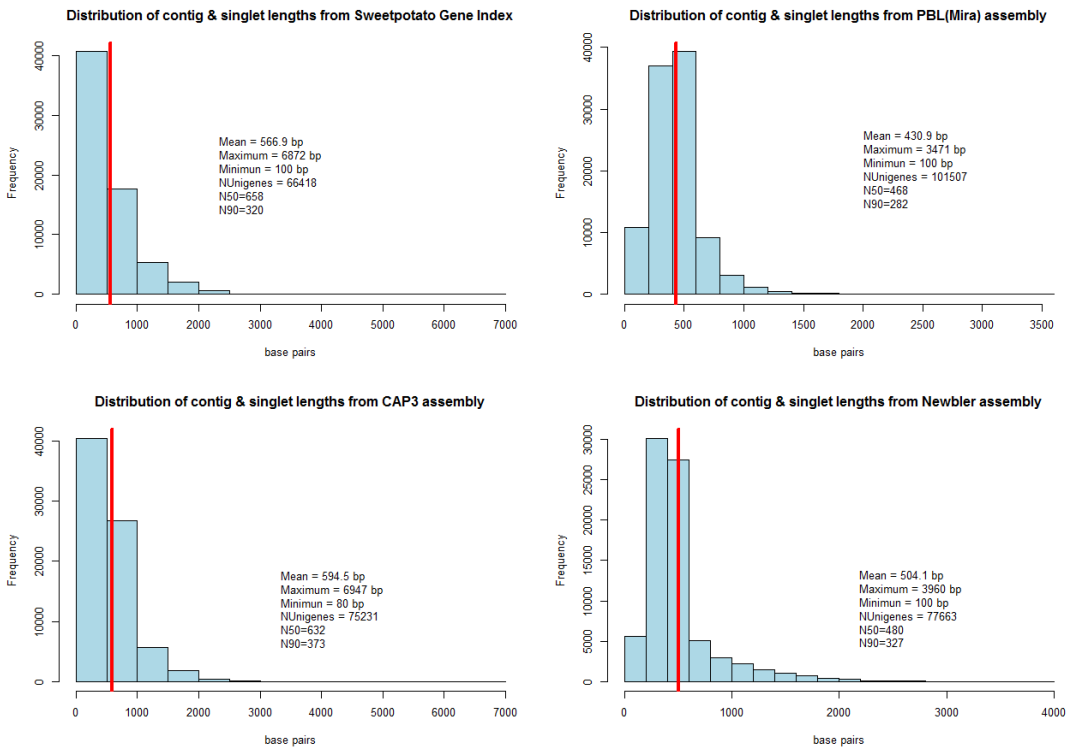


Figure 3.1. Summary of different gene indices of sweetpotato. Legend: Nun genes = number of contigs +singlets.

3.3.2 Expression analysis of genes possibly involved in the onset of storage roots

To identify novel genes involved in storage roots development, 30 unigenes were selected from the unique sequences classified in GO categories "nucleic acid binding transcription factor activity", "Molecular transducer activity", and "developmental process". Most of these unigenes were selected having as criteria that they were assembled by the overlapping of sequences coming from the root libraries. Primers were designed flanking a putative intron. Selected contigs from *cap3p80v3*, *newblerv1* and *PBL assemblies* were inspected to verify the details of sequences forming each consensus before the primer design and detect possible loss of information in selected contigs. Quantitative reverse transcription PCR assay was performed using the primers to compare the expression profiles in young storage root (SR)

Table 3.2. Primer sets used for reverse transcription and quantitative RT-PCR analysis.

Gene Name[1]	Oligo Name	Sequence (5' to 3')	Ta (°C)	Product Length (bp)	Accession in cap3p80 (PBL) [cap3p90] assemblies[2]	Accession at GenBank (SPGI)[PGDB][3]
IbBME	IbBME_F	agctagtccgacgtttgttcc	56	132	(S_PBL_c31897)	
IbBME	IbBME_R	catctcgtctctgcacctttct	56			
IbXX	IbXX_F	ggggcatagatttgaacatgg	56	226	(S_PBL_c3428)	
IbXX	IbXX_R	gaaataaccgctggcaacac	56			
IbBBR	IbBBR_F	ggatggcagtcgtcttgctg	60	200	Contig6685.1 (S_PBL_c182) [Contig8024.2]	
IbBBR	IbBBR_R	tttgttccatgcttggcccagta	60			
IbBEL4	IbBEL4_F	tcctacccaatcgaggaaatg	56	195	(S_PBL_c6271)	
IbBEL4	IbBEL4_R	tgcttcaccaatcacaactcc	56			
IbBEL5	IbBEL5_F	aagtcacgctgattggtaacac	56	192	(S_PBL_c24240)	
IbBEL5	IbBEL5_R	aatagcccccaaagtttgtaac	56			
IbHB1	IbHB1_F	ggatgaagaggttgttggtacg	56	210	Contig3374.1 (S_PBL_c3587)	
IbHB1	IbHB1_R	cagtccacctccgtctgtttc	56			
IbHB2	IbHB2_F	caggaagctgaagatgagttatg	56	139	(S_PBL_c7043) [Contig529.2]	
IbHB2	IbHB2_R	ctcctcatcagtctccttgtc	56			
IbPP2A	IbPP2A_F	cattcacgggcagttctacg	59	171	Contig5814.1 (S_PBL_c33740) [Contig7022.2]	
IbPP2A	IbPP2A_R	cacagctccatcatttagtaccc	59			
IbCBP2	IbCBP2_F	cctgaaacaatggaggaagc	60	159	[Contig23250.2]	
IbCBP2	IbCBP2_R	ccagtttgcttttcccagttc	60			
Ibkn3a	Ibkn3a_F	cgcctaggtccataatccaatcca	60	212	Contig5774.1 (S_PBL_c31412) [Contig36256.2]	
Ibkn3a	Ibkn3a_R	ccacaagttgcagatcaaagataa	60			
Ibkn3b	Ibkn3b_F	cgcctaggtccataatccaacata	60	215		AB283029
Ibkn3b	Ibkn3b_R	ccacaagttgcagatatcagatgac	60			

(Table 3.2 cont.)

Gene Name[1]	Oligo Name	Sequence (5' to 3')	Ta (°C)	Product Length (bp)	Accession in cap3p80 (PBL) [cap3p90] assemblies[2]	Accession at GenBank (SPGI)[PGDB][3]
IbEF1a	IbEF1a_F	tcctgaagaatggtgatgctgg	56	165	Contig35258.1 (S_PBL_c12390)	
IbEF1a	IbEF1a_R	cagttgggtccttcttgtcaac	56			
IbTAP	IbTAP_F	tgctgcgcgatcacaatacat	56	219	Contig27416.1 [Contig28125.2]	
IbTAP	IbTAP_R	catggttggcaatctatagtttgg	56			
Ibkn2	Ibkn2_F	tttgagtgaaggcgatgctatgaa	57	209		AB283028
Ibkn2	Ibkn2_R	attgatcgagttctgggtcctt	57			
IbSnRK	IbSnRK_F	atgtggtgctattcgcataatcc	56	217	(S_PBL_c2856)	
IbSnRK	IbSnRK_R	cacagaagtccagaaacagaaattg	56			
IbCRF1	IbCRF1_F	ccctaaacggagacgaattcac	56	183	(S_PBL_c3693)	
IbCRF1	IbCRF1_R	gggggtgttctaactgattttc	56			
IbCRF2	IbCRF2_F	cccacagatatcccattcatcg	58	211	(S_PBL_c5300)	
IbCRF2	IbCRF2_R	aataatcgtcgacgttcacgtc	58			
IbCDPK2	IbCDPK2_F	actactgagtggagttccgcca	56	170	Contig22153.1 [Contig26263.2]	
IbCDPK2	IbCDPK2_R	ctgtcaatctcttctctggatccct	56			
IbCDPK3	IbCDPK3_F	cttgatcggaacccgaaaaggagg	62	93	Contig28899.1 [Contig34516.2]	
IbCDPK3	IbCDPK3_R	aagaggcttgtcgggggccatac	62			
IbMAP65	IbMAP65_F	accagaagaagttccatgaccag	56	197	(S_PBL_c11096)	
IbMAP65	IbMAP65_R	cctcccatctttagatgttgacct	56			
IbBEL2	IbBEL2_F	agggctttgcagcagttgggaatgat	60	248		DV038045.1, DC881164.1 [IBAT_PUT65, IBAT_PUT9796][5]
IbBEL2	IbBEL2_R	tcctccatatacatctcctccaccat	60			
IbBEL6	IbBEL6_F	aaacccatggtggaagagatgtac	56	231		EE881365.1 [IBAT_PUT10648][5]
IbBEL6	IbBEL6_R	agcgttgatttcggatcttttgc	56			
IbBEL1	IbBEL1_F	tagagaccaaattgaagattgccaa	60	204		EE880927.1 [IBAT_PUT8950][5]

(Table 3.2 cont.)

Gene Name[1]	Oligo Name	Sequence (5' to 3')	Ta (°C)	Product Length (bp)	Accession in cap3p80 (PBL) [cap3p90] assemblies[2]	Accession at GenBank (SPGI)[PGDB][3]
IbBEL1	IbBEL1_R	gaaggcagatcagagcagcata	60			
IbBEL3	IbBEL3_F	gcttgtctgtgtctgtgggata	56	252		(Ib28923)
IbBEL3	IbBEL3_R	ccataactgagcagctgcagtt	56			
IbAGPase	IbAGP1ase_F	catcaaaagagcaatcattgaca	56	229		AY544766, AB271011 [IBAT_PUT667] [5]
IbAGPase	IbAGP1ase_R	tcattagtctctcaaacggctaca	56			
IbCBP3[6]	IbCBP3_F	ttaacttcgttcaagacgaaaaggtg	60	194	[Contig20410.2], (S_PBL_c1633), RT_038063.1 (corrected) [4], RT_483640.1, (S_PBL_c1633)	JG699262.1
IbCBP3[6]	IbCBP3_R	tgagtcataaactcggttaagctcggt	60			
IbCBP1[6]	IbCBP1_F	gaccgatccggtcgagcttaacttg	56	259	[Contig13109.2], RT_027290.1, RT_055908.1 (S_PBL_c13311)	JG699346.1
IbCBP1[6]	IbCBP1_R	tgttctcattctcaacggctaac	56			
IbNAM	IbNAM_F	cagtacggcttctcaaactcaac	56	201	(S_PBL_c19033)	
IbNAM	IbNAM_R	gtcataaccttgccctgatgc	56		(S_PBL_c19033)	
IbBHLH	IbBHLH_F	gcggacatagaagtgacaatgg	56	208	(S_PBL_c21769)	
IbBHLH	IbBHLH_R	catccacagatgccactttgca	56		(S_PBL_c21769)	
IbIAA	IbIAA_F	atgctcgttggtgatcttcc	56	204	(S_PBL_c15425, S_PBL_c19292)	
IbIAA	IbIAA_R	tagaacgccaacttcaaatgc	56		(S_PBL_c15425, S_PBL_c19292)	

[1] Gene name is given based on the closest matching gene found by BLASTX analysis.
[2]Three transcripts assemblies named: cap3p80v3, PBL and cap3p90 were used to select the sequences.
[3] Current sequence ID of the matching sequence in public databases: accession number from GenBank (http://www.ncbi.nlm.nih. gov/genbank/), customized ID (Ib#) where # is the number identifying the sequence in Sweetpotato Gene Index (SPGI) (Schafleitner et al, 2010).
[4]Putative error in sequence was detected and corrected after analysis with FrameDP (http://iant.toulouse.inra.fr/FrameDP/).
[5] Sequences from PlantGDB (http://www.plantgdb.org/) were renamed as IBAT_PUT#, for example IBAT_PUT667 is a sequence deposited at PlantGDB as "PUT-169a-Ipomoea_batatas-667 PlantGDB-assembled Unique Transcript-fragment derived from *Ipomoea_batatas* mRNAs Jan_14_2009 (based on GenBank release 169)".
[6] Single reads of IbCBP3 and IbCBP1 are in Contig20410.2 and Contig13109.2 from *cap3p90 assembly*, while most of all reads from these contigs are clustered in single consensus sequence, Contig10947.1, in the *cap3p80v3* assembly

Table 3.3. Summary of sweetpotato assemblies.

Assembly designation	Total Length	Unigenes	Contigs	Average Unigene Length (bp)	Average Contig Length (bp)	N50	N90	Max. Contig Length (bp)
Newblerv1	39,149,064	77,663	35,069	504.1	629.1	480	327	3960
CAP3P80v3 (including contigs<100pb)	44,724,637	75,231	56,270	594.5	659.5	632	373	6947
CAP3P80v3 (excluding contigs<100pb)	44,690,061	74,849	55,888	597.1	663.3	633	374	6947
PBL (MIRA)	43,741,302	101,507	55,296	430.9	519.4	468	282	3471
SPPGDB	6,985,150	12,464	2,117	560.4	nd	658	361	3236
SPPGDB (after cleaning)	6,923,640	12,306	1,960	562.6	nd	657	361	3236
SPGI	37,655,484	66,418	37,624	566.9	703.0	658	320	6872

nd =no data available

and developing storage roots (DSR) in contrast to fibrous roots (FR). The results are summarized in Table 3.8. The results indicated that the expression of 20 genes (including Ibkn2) was significantly upregulated in DSR with a fold change ratio of at least 1.5 times; similarly 17 genes were upregulated in SR, sharing 15 genes in common with those upregulated in DSR; in contrast, three unigenes were significantly downregulated in either or both DSR and SR stages. These upregulated genes in both SR and SDR stages were annotating as encoding different classes of proteins involved in calcium-mediating signaling, transcriptional activities and in molecular transduction. In brief, five out 17 upregulated genes are components of the cellular machinery of calcium signaling: two members of calcium dependent protein kinases (IbCDPK2, IbCDKP3), and two Ca^{2+} -binding proteins (IbCBP1, IbCBP2) showed fold changes ratio of ~1.5 to more than 30 times (Table 3.8). IbCBP1 showed the highest fold change ratio at SR stages and its specific expression in storage roots was confirmed by the semi-quantitative RT-PCR (Figure 3.2). Whereas the additional gene, IbCBP3, showed a specific expression in SR and SDR, supported by results of sqRT-PCR (Figure 3.2). Although we found that two members related to

the protein serine/threonine phosphatase 2A (IbPP2A) and its 2A phosphatase associated protein (IbTAP) were upregulated ~ two times in SR and DSR compared to FR, their expression is not specific to roots since their transcripts are abundant in the leaf tissues (Figure 3.2). A set of eleven transcription factors (TFs) were found that showed differential expression in three roots tissues. According to its annotation, TFs are members of BEL1-like (IbBEL1, IbBEL2, IbBEL3, IbBEL4 and IbBEL5), homeobox leucine zippers (IbHB1, IbHB2), GATA-type zinc finger (IbBME), basic helix-loop-helix (IbBHLH1), the cytokinin response factor (IbCRF1) and of the NAM-Like (IbNAM) protein families. In addition, transcripts of Ibkn2 gene were upregulated in both SR and DSR stages (2-4 times) compared to FR. In this study we were able to validate the expression of the identified variants of the sweetpotato KNOX gene, Ibkn3, designed Ibkn3a and Ibkn3b by sqRT-PCR (Figure 3.2); unfortunately we were not able to get the real time PCR results of Ibkn3b; Ibkn3a showed a fold change ratio of two times expression in SR and a slightly up-regulation in DSR (1.3 times) in comparison to FR. Consensus sequences S_PBL_c3141, Contig36256.2 and Contig5774.1 represent IbKn3a and AB283029 represents IbKn3b (Figure 3.2). Nucleotide variations such as SNP are also present mostly in the terminal untranslated region of both Ibkn3a and Ibkn3b alleles. AB283029.1 is the original IbKn3 gene. CAP3 with parameter p 80 generated a single contig (Contig5774.1, included 19 singlets) and CAP3 with parameter p 90 generated two contigs (Contig36256.2 and Contig6970.2 included 10 and 6 sequences, respectively). BM878851.1 (from plantlets) and ST_314184.1 (from stem) are sequences included in Contig36256.2 and Contig6970.2, respectively, but shown in the alignment (Figure 3.2). Similarly, S_PBL_c3141 is a contig assembled with root reads from 454 pyrosequencing (PBL assembly using MIRA).

Table 3.4. Number (percentage) of Unigenes of four sweetpotato assemblies and characteristics of their Contigs and Unigenes.

	Newblerv1 [1]	Cap3p80v3[1]	PBL [2](MIRA)	SPGI[3]
Number of Contigs (>=1Kbp)	6,469 (18.45%)	8,030 (14.27%)	2,222 (4.02%)	8,115 (21.57%)
Number of Contigs (>=0.5Kbp)	17,386 (49.58%)	33,762 (60%)	24,036 (43.47%)	21,991 (58.45%)
Number of Contigs (>=0.20Kbp)	29,796 (84.96%)	53,753 (95.53%)	52,314 (94.61%)	37,370 (96.97%)
Number of contigs	35,069 (100%)	56,270 (100%)	55,296 (100%)	31,685 (100%)
Number of Unigenes (>=1Kbp)	6,487 (8.35%)	8,072 (10.73%)	2,222 (2.19%)	8,182 (12.32%)
Number of Unigenes (>=0.5Kbp)	19,409 (24.99%)	34,957 (46.47%)	25,124 (24.75%)	25,912 (39.01%)
Number of Unigenes (>=0.20Kbp)	72,122 (92.87%)	72,525 (96.4%)	90,861 (89.51%)	61,584 (92.72%)
Number of Unigenes	77,663 (100%)	75,231 (100%)	101,507 (100%)	66,418 (100%)

[1] Assemblies designed NEWBLERv1 and CAP3P80v3 from the current study.
[2] PBL assembly of sweetpotato root transcriptome, provided by Nurit Firon (personal communication, December 2010) built using MIRA software (http://www.chevreux.org/projects_mira.html).
[3]Sweetpotato Gene Index (SPGI) (Schafleitner et al, 2010) built using MIRA software.

Table 3.5. Numbers of sequences with and without a matching sequence in UniprotKB and TAIR10 of three Sweetpotato assemblies (BLASTX E-value <=1e-6).

Assembly designation	#Sequences <100 bp	With_Hit	Without_Hit	Total	Database
Cap3p80v3	382 (0.51%)	52,322 (69.55%)	22,909 (30.45%)	75,231 (100%)	Uniprot
Cap3p80v3	382 (0.51%)	48,130 (63.98%)	27,101 (36.02%)	75,231 (100%)	TAIR10
Newblerv1	(0%)	46,967 (60.48%)	30,696 (39.52%)	77,663 (100%)	Uniprot
Newblerv1	(0%)	44,298 (57.04%)	33,365 (42.96%)	77,663 (100%)	TAIR10
PBL (>=100 bp)	(0%)	68,944 (67.92%)	32,563 (32.08%)	101,507 (100%)	Uniprot
PBL (>=100 bp)	(0%)	62,028 (61.11%)	39,479 (38.89%)	101,507 (100%)	TAIR10

Table 3.6. Number (Percentage) of sequences with and without a matching sequence in UniprotKB (E-value of BLASTX<1E-06) in relation to #Contigs and #Unigenes.

	Newblerv1	Cap3p80v3	PBL (MIRA)	SPGI
Number of Contigs (>=1Kbp) (with a BLASTX hit)	6,371 (18.17%)	7,833 (13.92%)	2,183 (3.95%)	7,864 (24.82%)
Number of Contigs (>=0.5Kbp) (with a BLASTX hit)	15,719 (44.82%)	29,292 (52.06%)	21,769 (39.37%)	19,180 (60.53%)
Number of Contigs (>=0.20Kbp) (with a BLASTX hit)	21,988 (62.7%)	40,776 (72.46%)	41,571 (75.18%)	23,927 (75.52%)
Number of contigs (any size)* (with a BLASTX hit)	22,910 (65.33%)	40,816 (72.54%)	42,184 (76.29%)	23,957 (75.61%)
Total Number of contigs (with & without a BLASTX hit)*	35,069 (100%)	56,270 (100%)	55,296 (100%)	31,685 (100%)
Number of Unigenes (>=1Kbp) (with a BLASTX hit)	6,382 (8.22%)	7,868 (10.46%)	2,184 (2.15%)	7,918 (11.92%)
Number of Unigenes (>=0.5Kbp) (with a BLASTX hit)	16,924 (21.79%)	30,081 (39.98%)	22,494 (22.16%)	22,011 (33.14%)
Number of Unigenes (>=0.20Kbp) (with a BLASTX hit)	46,026 (59.26%)	52,279 (69.49%)	65,672 (64.7%)	38,888 (58.55%)
Total Number of Unigenes (any size)* (with a BLASTX hit)	46,967 (60.48%)	52,322 (69.55%)	68,945 (67.92%)	38,918 (58.6%)
Number of Unigenes (with & without a BLASTX hit)*	77,663 (100%)	75,231 (100%)	101,507 (100%)	66,418 (100%)
Number of contigs (without a BLASTX hit,any size)*	12,159 (34.67%)	15,454 (27.46%)	13,112 (23.71%)	7,728 (24.39%)

* Newbler and CAP3 includes only singlets >=300nt (i.e. CAP3/Newbler includes NGS singlets>=300nt but all ESTs including (370 ESTs of 100 to 299nt in CAP3 and 542 in Newbler); PBL assembly includes sequences >=100nt.

.

Table 3.7. Number (Percentage) of sequences with a matching sequence in UniprotKB (E-value of BLASTX<1E-06) in relation to #Contigs or #Unigenes that have a matching sequence in UniprotKB

	Newblerv1	cap3p80v3	PBL (MIRA)	SPGI
Number of Contigs (>=1Kbp) (with a BLASTX hit)	6,371 (27.81%)	7,833 (19.19%)	2,183 (5.17%)	7,864 (32.83%)
Number of Contigs (>=0.5Kbp) (with a BLASTX hit)	15,719 (68.61%)	29,292 (71.77%)	21,769 (51.6%)	19,180 (80.06%)
Number of Contigs (>=0.20Kbp) (with a BLASTX hit)	21,988 (95.98%)	40,776 (99.9%)	41,571 (98.55%)	23,927 (99.87%)
Number of contigs (any size)* (with a BLASTX hit)	22,910 (100%)	40,816 (100%)	42,184 (100%)	23,957 (100%)
Total Number of contigs (with & without a BLASTX hit)*	35,069	56,270	55,296	31,685
Number of Unigenes (>=1Kbp) (with a BLASTX hit)	6,382 (13.59%)	7,868 (15.04%)	2,184 (3.17%)	7,918 (20.35%)
Number of Unigenes (>=0.5Kbp) (with a BLASTX hit)	16,924 (36.03%)	30,081 (57.49%)	22,494 (32.63%)	22,011 (56.56%)
Number of Unigenes (>=0.20Kbp) (with a BLASTX hit)	46,026 (98%)	52,279 (99.92%)	65,672 (95.25%)	38,888 (99.92%)
Total Number of Unigenes (any size)* (with a BLASTX hit)	46,967 (100%)	52,322 (100%)	68,945 (100%)	38,918 (100%)
Number of Unigenes (with & without a BLASTX hit)*	77,663	75,231	101,507	66,418
Number of contigs (without a BLASTX hit, any size)*	12,159 (25.89%)	15,454 (29.54%)	13,112 (19.02%)	7,728 (19.86%)

*Newbler and CAP3 includes only singlets >=300nt (i.e. CAP3/Newbler includes NGS singlets>=300nt but all ESTs including (370 ESTs of 100 to 299nt in CAP3 and 542 in Newbler); PBL assembly includes sequences >=100nt.

Table 3.8. Quantitative RT-PCR results (Fold change ratio) of genes in putative developing storage roots (DSR) and Storage Root (SR) vs. Fibrous root (FR), nd = no data.

GeneID	DSR/FR	SR/FR	Annotation
Ibkn2	2.392	4.160	A8QXP6(IBKN2) Class-I knotted1-like homeobox protein IBKN2
Ibkn3a	1.371	2.149	A8QXP7(IBKN3) Class-I knotted1-like homeobox protein IBKN3
Ibkn3b	nd	nd	A8QXP7(IBKN3) Class-I knotted1-like homeobox protein IBKN3
IbBEL2	2.467	1.054	AT2G35940(BLH1,EDA29) BEL1-like homeodomain
IbBEL3	3.819	11.447	AT5G02030(BLH9,HB-6) POX (plant homeobox) family protein; BEL1-like homeodomain 9
IbBEL1	0.950	0.468	AT5G41410.1 POX (plant homeobox) family protein
IbBEL4	4.271	2.313	AT2G35940 BEL1-like homeodomain 1
IbBEL5	2.741	1.175	AT4G36870.2 BEL1-like homeodomain 2
IbBEL6	1.713	0.792	AT2G23760(BLH4) BEL1-like homeodomain 4;BLH4 (SAWTOOTH 2);
IbBHLH1	7.641	37.237	AT1G72210 (bHLH096) basic helix-loop-helix (bHLH) DNA-binding superfamily protein
IbNAM	5.684	12.107	Q8LRL4_PETHY Nam-like protein 11 Petunia hybrida NH11
IbBBR	2.016	2.640	AT2G21240.1 basic pentacysteine 4(BBR,BPC4)
IbHB1	7.849	20.464	AT4G37790.1 Homeobox-leucine zipper protein family (HAT22)
IbHB2	4.242	7.797	AT2G46680(ATHB7,HB-7) homeobox 7
IbCRF1	3.390	0.918	AT4G11140(CRF1) cytokinin response factor 1
IbCRF2	0.547	0.450	AT4G27950.1(CRF4) cytokinin response factor 4
IbBME	2.709	4.537	AT3G54810.1(BME3,BME3-ZF) Plant-specific GATA-type zinc finger transcription factor family protein
IbPP2A	2.267	1.990	AT1G10430 (PP2A) protein phosphatase 2A-2; PP2A (serine/threonine protein phosphatase 2A)
IbTAP	2.011	2.133	AT5G53000(TAP46) 2A phosphatase associated protein of 46 kD
IbGRF2	1.148	1.027	AT1G78300 (GRF2) general regulatory factor 2
IbCDPK2	2.335	3.652	AT4G09570.1 calcium-dependent protein kinase 4
IbCDPK3	1.432	1.467	AT4G09570(CPK4) Calcium dependent protein kinase 4
IbMAP65	3.744	12.038	AT1G14690.2 microtubule-associated protein 65-7
IbSnRK	1.289	1.356	AT3G01090 SNF1 kinase homolog 10(SKIN10);snf1-related protein kinase
IbCBP1	7.714	67.380	AT2G44310 Calcium-binding EF-hand family protein
IbCBP2	3.084	5.155	AT2G33990.1 iqd9 (IQEdomain 9); calmodulin_binding family protein
IbCBP3	1.329	1.143	AT2G44310 calcium-binding EF hand family protein
IbIAA	2.963	2.061	AT4G29080 (PAP2, IAA27) phytochrome-associated protein 2;CAC84706.1| aux/IAA protein [Populus tremula x Populus tremuloides] ; indole-3-acetic acid inducible 16

A

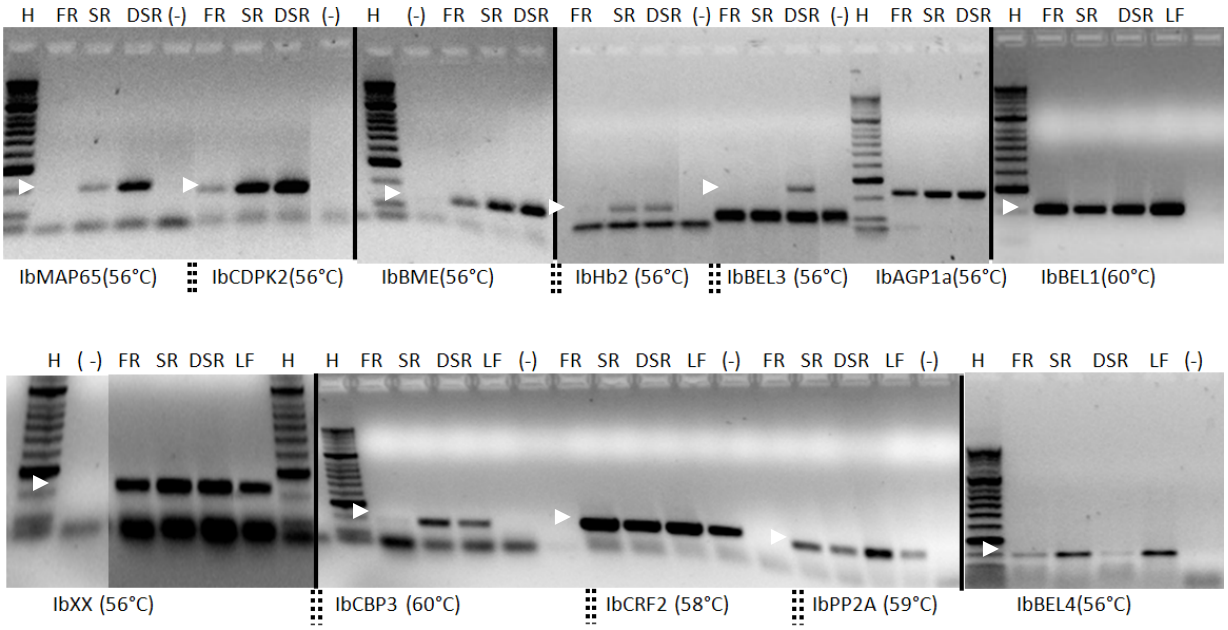
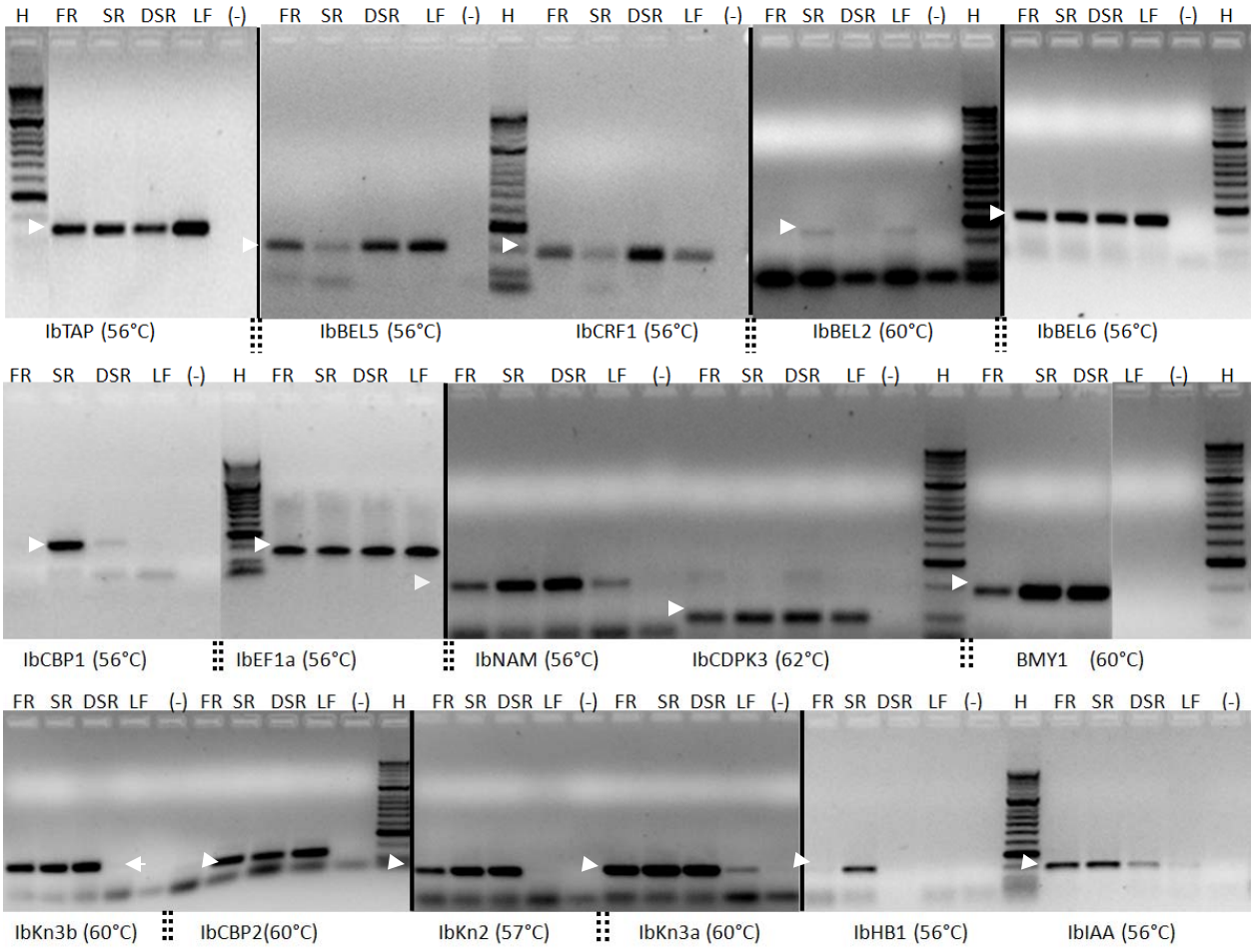


Figure 3.2 RT-PCR analysis showing the expression of genes in different tissues. SR=storage roots, DSR=initiating storage roots (thick, pigmented roots), FR=white, fibrous roots, LF=Leaf, (-) = non template control. H=Mol weight marker= Hyperladder II (50,100, 200, 300, 400, 500, 600, 700, 800, 1000, 1200, and 1400 base pairs) (Bioline, Taunton, USA). The products of sqRT-PCR were run in 2.2% of agarose gel electrophoresis. The specific annealing temperature of each gene is in brackets. The expected PCR product is indicated by a white arrowhead. **(A)** RT-PCR of sweetpotato genes IbMAP65 (microtubule-associated protein 65-7), IbCDPK2(calcium-dependent protein kinase), IbME (plant-specific GATA-type zinc finger transcription factor), IbHB2 (homeobox-leucine zipper protein), IbBEL3 (BEL1-like homeodomain), IbBEL1 (BEL1-like), IbXX (similar to accession XP_002510084.1 conserved hypothetical protein from *Ricinus communis*), IbCBP3 (calcium-binding protein), IbCRF2 (cytokinin response factor), IbPP2A(serine/threonine protein phosphatase 2A), and IbBEL4 (BEL1-like). **(B)** RT-PCR of sweetpotato genes IbTAP (2A phosphatase associated protein), IbBEL5 (BEL1-like), IbCRF1(cytokinin response factor), IbBEL2(BEL1-like), IbBEL6 (BEL1-like), IbCBP1 (calcium-binding protein), IbEF1a (elongation factor 1-alpha), IbNAM (Nam-like protein), IbCDPK3 (calcium dependent protein kinase), bMY1(β-amylase), Ibkn3b (Class-I knotted1-like homeobox protein IBKN3), IbCBP2 (calcium-binding protein), Ibkn2 (Class-I knotted1-like homeobox protein IBKN2), Ibkn3a (Class-I knotted1-like homeobox protein IBKN3), IbHB1 (homeobox-leucine zipper protein), and IbIAA (aux/IAA protein).

(Figure 3.2 cont.)

B



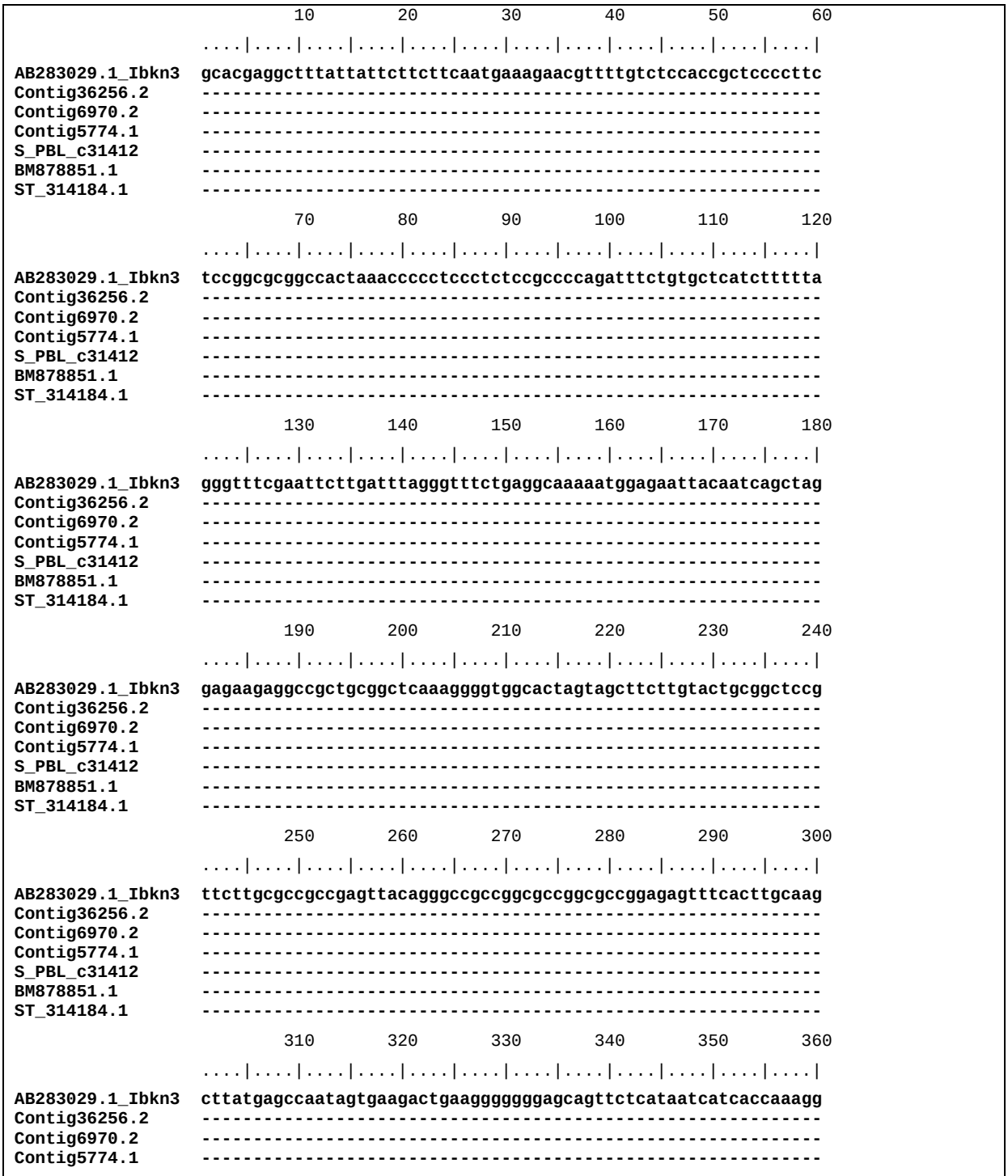


Figure 3.3 Alignment of Ibkn3-like sequences.
Three indels at positions 1179, 1296 and 1358 (numbering respect to AB283029) suggest existence of allelic variants of Ibkn3 designed as Ibkn3a and Ibkn3b. Consensus sequences S_PBL_c3141, Contig36256.2 and Contig5774.1 represent Ibkn3a and AB283029 represents Ibkn3b. Dots in alignment represent 100% identical residues and dashes are gaps.). In lowercase are regions of 5' and 3'-untranslated regions of transcripts as compared to AB283029.1 and in uppercase is the Coding region.

(Figure 3.3 cont.)

```
S_PBL_c31412      ------------------------------------------------------------
BM878851.1        ------------------------------------------------------------
ST_314184.1       ------------------------------------------------------------

                         370       380       390       400       410       420
                  ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3  tttgagagtgaaATGGAAGCTATAAAGGCTAAGATTATTGCCCATCCTCAGTATTCCAAT
Contig36256.2     ------------------------------------------------------------
Contig6970.2      ------------------------------------------------------------
Contig5774.1      ------------------------------------------------------------
S_PBL_c31412      ------------------------------------------------------------
BM878851.1        ------------------------------------------------------------
ST_314184.1       ------------------------------------------------------------

                         430       440       450       460       470       480
                  ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3  CTTTTGGAGGCTTACATGGACTGCCAAAAAGTTGGGGCTCCGCCGGAGGTGGTGGCGCGT
Contig36256.2     ------------------------------------------------------------
Contig6970.2      ----------------------------..........G.....................
Contig5774.1      ----------------------------..........G.....................
S_PBL_c31412      ------------------------------------------------------------
BM878851.1        ------------------------------------------------------------
ST_314184.1       ------------------------------------------------------------

                         490       500       510       520       530       540
                  ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3  CTTGCGGCGGTGCGCCAAGAGTTTGAGGCCCGGCAACGCGCCGCCGGTCTCGGTGGGAGA
Contig36256.2     ------------------------------------------------------------
Contig6970.2      ..............T..G.........T................................
Contig5774.1      ..............T..G.........T................................
S_PBL_c31412      ------------------------------------------------------------
BM878851.1        ------------------------------------------------------------
ST_314184.1       ------------------------------------------------------------

                         550       560       570       580       590       600
                  ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3  GATATTTCCTCCAAAGACCCGGAACTTGACCAGTTTATGGAAGCGTACTATGATATGTTA
Contig36256.2     ------------------------------------------------------------
Contig6970.2      ..............G.....A.....C................................
Contig5774.1      ..............G.....A.....C................................
S_PBL_c31412      ------------------------------------------------------------
BM878851.1        ------------------------------------------------------------
ST_314184.1       ------------------------------------------------------------

                         610       620       630       640       650       660
                  ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3  GTGAAGTACCGAGAAGAACTGACCAGGCCTTTACAAGAAGCAATGGAGTTCATGCGACGG
Contig36256.2     ------------------------------------------------------------
Contig6970.2      ............................................................
Contig5774.1      ............................................................
S_PBL_c31412      ------------------------------------------------------------
BM878851.1        ------------------------------------------------------------
ST_314184.1       ------------------------------------------------------------

                         670       680       690       700       710       720
                  ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3  ATCGAATCACAACTAAATATGCTTAGCAACGCCCCAGTCCGGGTCTTCACTTCGGATGAC
Contig36256.2     ------------------------------------------------------------
Contig6970.2      ........G...................................................
Contig5774.1      ........G...................................................
S_PBL_c31412      ------------------------------------------------------------
BM878851.1        -------.G...................................................
ST_314184.1       ------------------------------------------------------------
```

(Figure 3.3 cont.)

```
                        730       740       750       760       770       780
                    ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3    AAATGTGAGGGTGTCGGTTCCTCTGAAGACGACCAAGATAACAGCGGTGGTGAAACCGAG
Contig36256.2       ------------------------------------------------------------
Contig6970.2        ............................................................
Contig5774.1        ............................................................
S_PBL_c31412        ------------------------------------------------------------
BM878851.1          ..............T..............T..............................
ST_314184.1         ------------------------------------------------------------

                        790       800       810       820       830       840
                    ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3    CTTCCCGAGATTGATCCTCGGGCTGAAGACCGCGAGCTGAAGAACCACTTGCTGAGGAAG
Contig36256.2       ------------------------------------------------------------
Contig6970.2        ....................A.......................................
Contig5774.1        ....................A.......................................
S_PBL_c31412        ------------------------------------------------------------
BM878851.1          ............................................................
ST_314184.1         ------------------------------------------------------------

                        850       860       870       880       890       900
                    ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3    TACAGCGGCTACCTAAGCAGTCTGAAGCAAGAGCTTTCGAAGAAAAAGAAGAAAGGGAAA
Contig36256.2       ------------------------------------------------------------
Contig6970.2        ............................................................A...
Contig5774.1        ............................................................A...
S_PBL_c31412        ------------------------------------------------------------
BM878851.1          ........G...................................................A...
ST_314184.1         ----------------------------------NNNNNNNNNNNNNNNNNNNNNN
                        910       920       930       940       950       960
                    ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3    CTCCCGAAAGAAGCCAGGCAAAAGCTGCTCAACTGGTGGGAGTTGCACTACAAATGGCCT
Contig36256.2       --------TA.GT..............................................
Contig6970.2        .....A......................................................
Contig5774.1        .....A......................................................
S_PBL_c31412        ------------------------------------------------------------
BM878851.1          .....A......................................................
ST_314184.1         NNNNNNNNTA.GT..............................................

                        970       980       990       1000      1010      1020
                    ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3    TATCCCTCGGAAACCGAGAAGGTGGCTTTGGCTGAATCGACGGGGTTGGATCAGAAGCAG
Contig36256.2       ..............G.............................................
Contig6970.2        ............................................................
Contig5774.1        ............................................................
S_PBL_c31412        ------------------------------------------------------------
BM878851.1          ............................................................
ST_314184.1         ..............G.............................................

                        1030      1040      1050      1060      1070      1080
                    ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3    ATTAACAACTGGTTCATCAATCAAAGGAAACGGCACTGGAAGCCTTCCGAGGACATGCAG
Contig36256.2       ..............................A.............T.............
Contig6970.2        ..............................A.............T.............
Contig5774.1        ..............................A.............T.............
S_PBL_c31412        ------------------------------------------------..........
BM878851.1          ..............................A.............T.............
ST_314184.1         ..............................A............................
```

(Figure 3.3 cont.)

```
                        1090      1100      1110      1120      1130      1140
                    ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3    TTTATGGTGATGGATGGTCTGCATCCACAAAATGCAGCTCTTTAC-ATGGATGGTCACTA
Contig36256.2       ......................C.............A......-.............
Contig6970.2        ......................C.....................-.............
Contig5774.1        ......................C.....................-.............
S_PBL_c31412        ......................C.............A......-.............
BM878851.1          ......................C.....................-.............
ST_314184.1         ....C.................C.............A......A.............
                        1150      1160      1170      1180      1190      1200
                    ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3    CATGGGAGACGGTCCATATCGCCTAGGTCCAtaatccaa-----cataatataaaacgtc
Contig36256.2       ....................................tccaa.................
Contig6970.2        ...........................................-.................
Contig5774.1        ...........................................-.................
S_PBL_c31412        ....................................tccaa.................
BM878851.1          ...........................................-.................
ST_314184.1         ....................................tccaa.................
                        1210      1220      1230      1240      1250      1260
                    ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3    aacttccatatcgagtttttttaaggttgtaacttaggacatcatgtatctgcctgaga
Contig36256.2       g..............-.............................................
Contig6970.2        g......t..c.a.........-.....-.................................
Contig5774.1        g......t..c..........-.....-.................................
S_PBL_c31412        g..................-.....................................
BM878851.1          g......t..c..........-....g.........a----------------------
ST_314184.1         g............tag.............................................
                        1270      1280      1290      1300      1310      1320
                    ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3    ccgccttgaaatatacgcgtacactatgccttccggctagccggcttgccgggaatcttg
Contig36256.2       ............g.....................----.....................
Contig6970.2        ...............................t..----.....................
Contig5774.1        ............g.....................----.....................
S_PBL_c31412        ............g.....................----.....................
BM878851.1          ------------------------------------------------------------
ST_314184.1         ............g.......................c----------------------
                        1330      1340      1350      1360      1370      1380
                    ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3    aacgaggggagctagctaggttaa-cttctcgtttgtc----atctgatatctgcaactt
Contig36256.2       ......................-.........a..tttg....------.....
Contig6970.2        ..............gcta.g.t.a.........a..tttgt...------.....
Contig5774.1        ......................-.........a..tttg....------.....
S_PBL_c31412        ......................-.........a..tttg....------.....
BM878851.1          ------------------------------------------------------------
ST_314184.1         ------------------------------------------------------------
                        1390      1400      1410      1420      1430      1440
                    ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3    gtggtaaccttagcaagcaa-----gatttatcggcaactatgaaggaaagtagggcgtt
Contig36256.2       ..................agcaa.....c..c.a.....................
Contig6970.2        ..................agca--------------------------------
Contig5774.1        ..................agcaa.....c..c.a.....................
S_PBL_c31412        ..................agcaa.....c..c.a.....................
BM878851.1          ------------------------------------------------------------
ST_314184.1         ------------------------------------------------------------
                        1450      1460      1470      1480      1490      1500
                    ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3    tttatgtatttgtaacgagtttgatgtagtaattaagtgcaattgtatatggtgagactg
Contig36256.2       ............................................................
Contig6970.2        ------------------------------------------------------------
Contig5774.1        ............................................................
S_PBL_c31412        ............................................................
BM878851.1          ------------------------------------------------------------
ST_314184.1         ------------------------------------------------------------
```

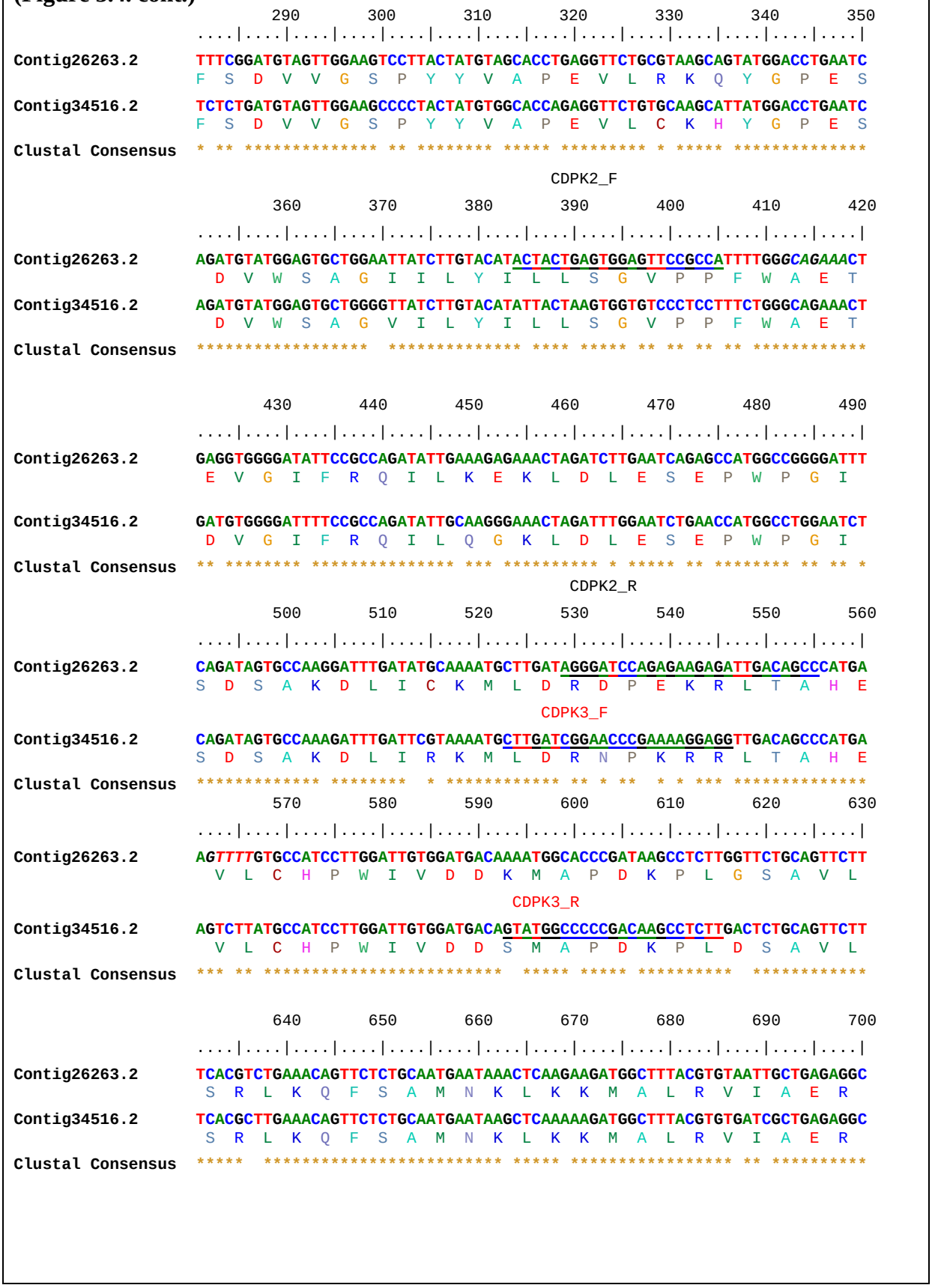
(Fig 3.3 cont.)

```
                          1510       1520       1530       1540       1550       1560
                   ....|....|....|....|....|....|....|....|....|....|....|....|
AB283029.1_Ibkn3   atgctagtaaataatttcagttagccctatataaaaaaaaaaaaaaaaaaaaaaaaaaaaa
Contig36256.2      ........................t.tatat..t.tt.tt.tg.tttt.gtccttt--
Contig6970.2       ------------------------------------------------------------
Contig5774.1       ........................tatatat..t.tt.tt.tg.tttt.gtccttt--
S_PBL_c31412       ........................tatatat..t.tt.tt.t.---------------
BM878851.1         ------------------------------------------------------
ST_314184.1        ------------------------------------------------------
                   ..
AB283029.1_Ibkn3   aa
Contig36256.2      --
Contig6970.2       --
Contig5774.1       --
S_PBL_c31412       --
BM878851.1
ST_314184.1
```

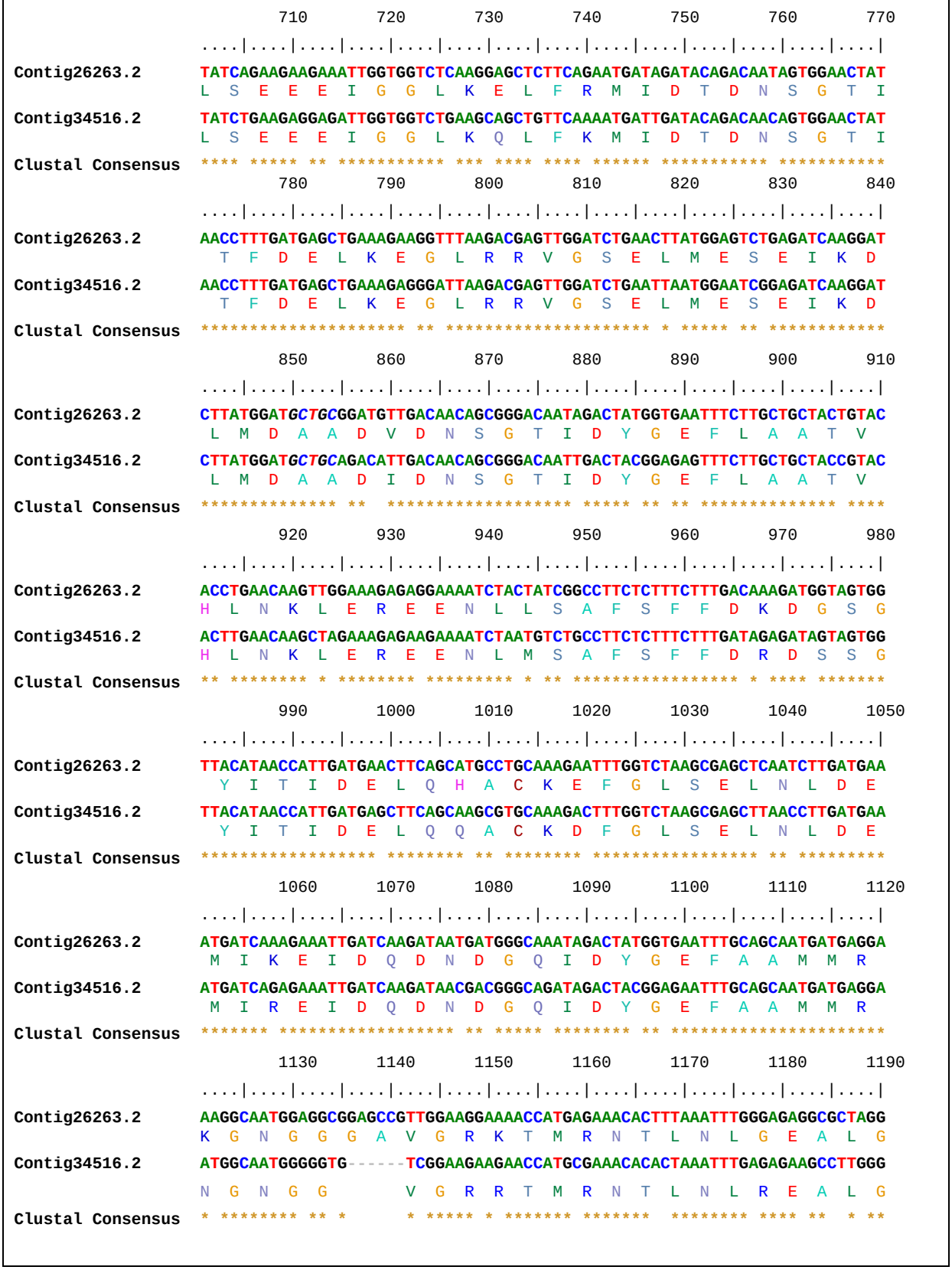
```
                            10         20         30         40         50         60         70
                     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Contig26263.2        AAGGGTACCTTTGAGGATGCTCTGTATGTACACATAGTCATGGAGCTCTGCGCCGGTGGGGAGCTTTTTG
                      K  G  T  F  E  D  A  L  Y  V  H  I  V  M  E  L  C  A  G  G  E  L  F
Contig34516.2        ---------- TAGGATGCGCTGTATGTGCACATAGTCATGGAGCTTTGCGCGGGCGGGGAGTTGTTTG
                                 *  D  A  L  Y  V  H  I  V  M  E  L  C  A  G  G  E  L  F
Clustal Consensus                ******* ******** ****************** ***** ** ****** * ****

                             80         90        100        110        120        130        140
                     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Contig26263.2        ATAGGATTGTGGAGAAGGGGCATTATAGTGAGAGGGAGGCTGCTAAGCTGCTTAAGACTATTGTTGGGGT
                     D  R  I  V  E  K  G  H  Y  S  E  R  E  A  A  K  L  L  K  T  I  V  G  V
Contig34516.2        ATAGGATTGTGGAGAAAGGGCAATACAGCGAGAGAGAGGCTGCTAAGCTAATCAAGACCATTGTTGGGGT
                     D  R  I  V  E  K  G  Q  Y  S  E  R  E  A  A  K  L  I  K  T  I  V  G  V
Clustal Consensus    **************** ***** ** ** ***** **************  * ***** ***********

                            150        160        170        180        190        200        210
                     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Contig26263.2        TGTGGAGGCTTGTCATTCCTTGGGGGTCATGCATAGAGATCTCAAGCCGGAGAACTTTTTGTGCCTTAGC
                       V  E  A  C  H  S  L  G  V  M  H  R  D  L  K  P  E  N  F  L  C  L  S
Contig34516.2        CGTCGAGGCTTGCCACTCTTTGGGGGTAATGCATAGAGATCTCAAGCCTGAGAACTTCTTGTTCCTTTGC
                       V  E  A  C  H  S  L  G  V  M  H  R  D  L  K  P  E  N  F  L  F  L  C
Clustal Consensus     ** ******** ** ** ******** ******************** ******** **** **** **

                            220        230        240        250        260        270        280
                     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Contig26263.2        ACTGATGAGGATGCTACTCTTAAGGCCATTGATTTTGGCCTTTCTGTTTTCTACAAACCAGGTGAAATAT
                      T  D  E  D  A  T  L  K  A  I  D  F  G  L  S  V  F  Y  K  P  G  E  I
Contig34516.2        TCCAACGAGGATGCTGCTCTCAAGGCCACTGATTTTGGTCTTTCTGTTTTCTATAAACCTGGGGAAACAT
                      S  N  E  D  A  A  L  K  A  T  D  F  G  L  S  V  F  Y  K  P  G  E  T
Clustal Consensus     *  * ********** **** ******* ********* ************** ***** ** **** **
```

Figure 3.4. Sequence comparison of two CDPK genes and their products**.** IbCDPK2 (Contig26263.2) and IbCDPK3 (Contig34516.2) assembled by CAP3. Contig34516.2 and Contig26263.2 were the result of the clustering and alignment of 139 and 49 reads derived from stem and root tissues. Underlined is the location of the primers designed for the reverse transcription of IbCDPK2 and IbCDPK3.

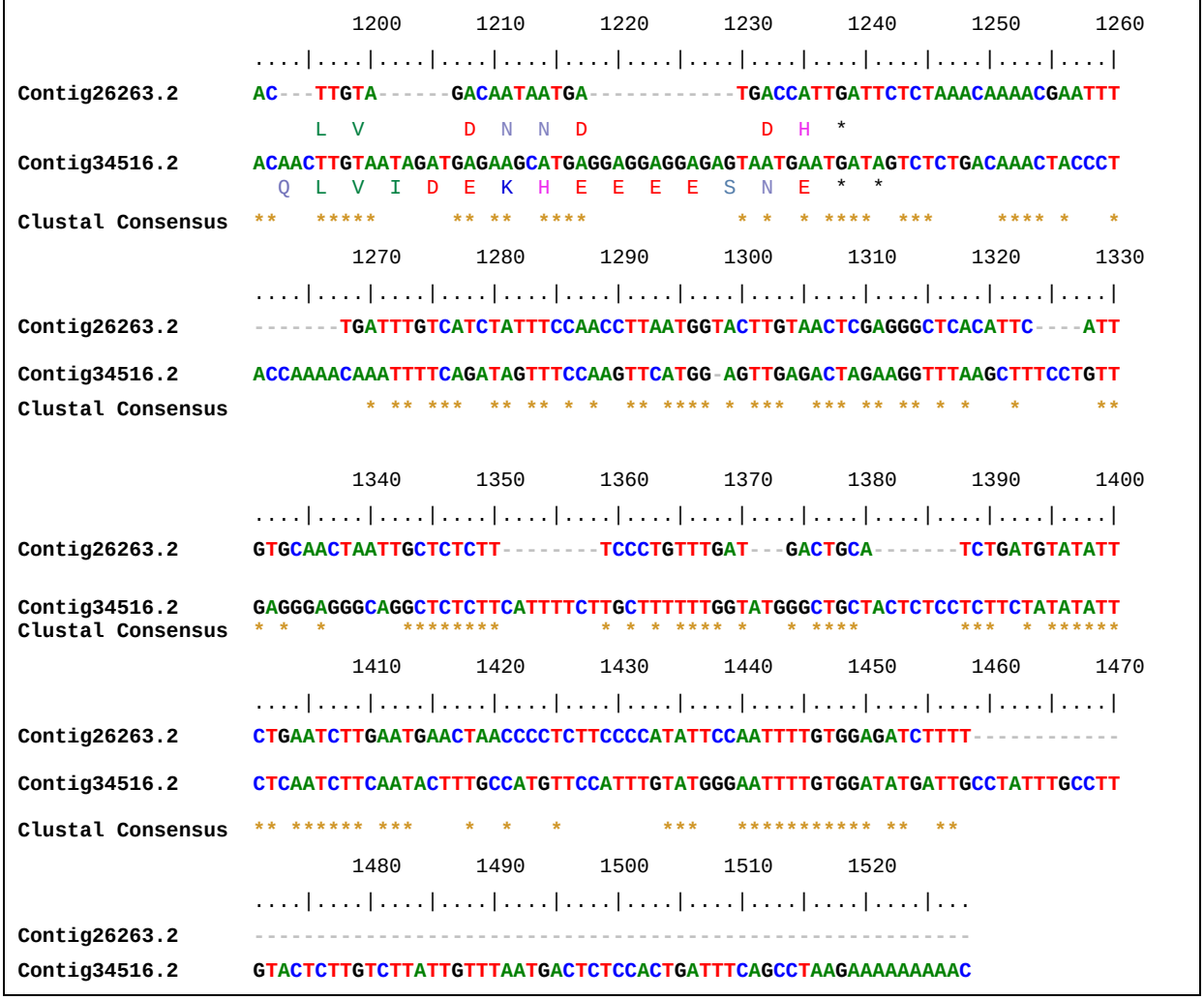
(Figure 3.4. cont.)



(Figure 3.4. cont.)



(Figure 3.4. cont.)



3.4 Discussion

De novo transcriptome analysis*:* Although the development of a fleshy root in sweetpotato and a tuberous root in potato is a complex trait, the understanding of the molecular mechanism have been facilitated by the availability of massive sequencing technologies and the use of *de novo* analysis of the generated data. By application of NGS technologies to both storage and non-storage roots it is possible to compare differences by digital gene expression analysis (DGE) and to assess the amount of differentially expressed transcripts (DET). In this study we first assembled the short reads from Illumina using ABySS, and then we used a NGS assembler, Newbler v2.5 and the typical CAP3 assembler following examples reported in the

literature (Kumar and Blaxter, 2010; Peng et al., 2007). It is known that the two most critical parameters that can greatly influence the output of CAP3 assembly are overlap length cutoff (default 40 bp) and overlap percent identity cutoff (parameter p) (Peng et al., 2007). We tested a combination of these parameters using a set of closely related sweetpotato paralogs of GenBank and from initial inspection of our data found that overlap percent identity cutoff was the most important criteria for assembly. Thus, we decide to build a CAP3 assembly using a overlap percent identity cutoff of 80% (we designed this assembly as *cap3p80v3;* contigs from this assembly are ending in suffix ".1" or ".3" [i.e., Contig1.2, Contig2.1, and so on]); Roche 454's Newbler (version 2.5) was also used under default parameters to build a second unigene (we designed this assembly as *newblerv1,*naming of contigs from the output of Newbler are all lower case and without a suffix [i.e., contig1, contig1, and so on]); Although, the latest version of Newbler or a combination of multiple assemblers are recommended when using *de novo* transcriptome assemblies, any other assembler such as MIRA, CAP3 is equally acceptable (Kumar and Blaxter, 2010; Prosdocimi et al., 2011). It cannot be concluded that the Newbler v2.5.3 assembly is not acceptable, but the fact that Newbler generates isotigs (i.e. putative alternative transcripts) by fragmenting EST in a non-model species like sweetpotato has made the Newbler assembly as an accessory software to this work.

The objective to build a unigene for sweetpotato was to discover novel genes associated with the onset of storage root formation. Key elements to this purpose were to combine transcripts derived from NGS of initiating storage roots and from non-storage, lignified, roots from our group (unpublished data), NGS from stem and leaf tissues (raw sequencing data provided by R. Schafleitner) and the ESTs from GenBank database (which were classified according to their source: storage root, fibrous roots, leaf or whole plantlets). Although, we focused our analysis on contigs assembled from a CAP3, we did two different assemblies from

this software, one being under an overlap percent identity cutoff of 90% (we designed this assembly as *cap3p90*, data not shown, contigs from this assembly are ending in suffix ".2" [i.e., Contig1.2, Contig2.1, so on]); we found *cap3p80v3* assembly was better in terms of reduction of redundancy and the minimal loss of information of reads (singlets) in the consensus sequence. The average length of unique transcripts of the sweetpotato assemblies is 594.5bp (*cap3p80v3*) and 504.1bp (*newblerv1*) (Table 3.3, Figure 3.1), similar to those previously reported for the sweetpotato gene index, SPGI (e.g. 566.9 bp in Schafleitner et al., 2010). A better comparison of different assemblies is done using the N50 metric (the smallest unigene size in which half the assembly is represented). The N50 of all unigenes in *cap3p80v3* assembly is 633bp (480bp in n*ewblerv1* assembly), while the estimated N50 value in SPGI and PGDB is 658bp and 657 bp, respectively. Although, the PBL assembly was built using the 454 pyrosequencing data obtained from roots, the average size of their contigs was 519.4 bp and possibly close or better than size of SPPGDB contigs, data not available (average of SPPGDB unigenes is 560.4 bp). The vast majority of contigs, 60% (33,762 sequences) of *cap3p80v3* assembly and 49.58% (17,386 sequences) contigs of n*ewblerv1* assembly have at least 0.5Kbp (Table 3.3); 84.96% to 95.53% of contigs had a length ranging from 300 to 1000pb (Figure 3.1). When focusing in sequences coming from the sweetpotato root transcriptome, we identified that most of ESTs (88.43%, 20,686/23,39) from GenBank forming the sweetpotato unigene deposited at PlantGDB, SPPGDB, are derived from libraries derived from root tissues (either or both fibrous and storage roots). Considering only the PBL assembly, ~78K transcripts are new and not matching to the known sequence data from sweetpotato roots (December 2010). A recent gene index of sweetpotato (Tao et al., 2012), that hereon referred to as SPTSA assembly (Sweetpotato TSA assembly), reported ~148K consensus transcripts, almost 2 times as in SPGI (Schafleitner et al., 2010) and 2 times as of our CAP3 based assembly reported here (~75K). Further analysis

revealed more than 1 consensus sequence (identity >90%) is present for each unigene from *cap3p80v3* assembly, based on comparison of unigenes from those listed in Table 3.1. The SPTSA assembly has shorter transcripts that overlap in a single, longer transcript from our dataset (data not shown). One example is the house keeping gene of elongation factor 1 alpha with Contig35258.1 (IbEF1a, 1925bp) vs. JP106582.1 (1692 bp). An improve of the sweetpotato transcriptome assembly is possible if raw sequencing data from multiple sources are combined and used to either further assembly or rebuild of a set of new unigenes. Our work and dataset, as well those from Schafleitner et al. (2010), Tao et al. (2012) and Xie et al., (2012) are extending sweetpotato genome data to comparable levels seen in other crops.

The sweetpotato annotated transcripts (unigenes) identified in the TSA assembly was 51,763. We estimated this number to be from ~48K to 52K by considering the number of unigenes matching an existing protein in UniprotKB (Table 3.5); this number is similar to the reported 39,031 protein-coding genes of potato (Potato Genome Sequencing Consortium, 2011), and much more than the ~21K annotated transcripts from coffee and the 26,346 annotated genes from grape (Jaillon et al., 2007); sweetpotato, potato and coffee belong to the Asterid I clade of dicot plant families, while grape belongs to the Rosid clade.

Discovery of Novel Genes: example of Ibkn3-like, CDPK-like and CBP-like genes: Although further intensive work is required to compare the results of the different assemblers, it is possible to use any potential biological information hidden in the consensus sequences by inspection of specific genes. For example, by comparing specific consensus sequences similar to the Ibkn3 gene (Tanaka et al., 2008) generated by the assemblers, CAP3, Newbler and MIRA (i.e. *cap3p80v3, newblerv1* and *PBL* assemblies), interesting putative sequence variations were thus identified. At least two consensus sequences were identified that appear to be alleles of Ibkn3 hereon referred to as Ibkn3a and Ibkn3b (Appendix A). Interestingly, these sequences are

almost 100% identical in their coding region and they differ only at their non-coding regions (3'-UTR) (Figure 3.2). The differences are mostly due to the presence of two insertion/deletions (indels). Gene-specific primer pairs were designed targeting these alleles using the observed sequence variations. Both Ibkn3a and Ibkn3b were identified by combining datasets from the different sources to make assemblies. We found a consensus sequence similar to Ibkn3a, (accession JP112770.1, Contig_8182) in the work of Tao et al. (2012) that is 99% identical in the aligned regions. In addition, the SPTSA assembly had several other Contigs that match almost 100% to both Ibkn3a and Ibkn3a in their coding region. Surprisingly, Tao et al. (2012), did not find a transcript overlapping to Ibkn3b (accession AB283029) in its whole extension. Similarly, two genes encoding calcium dependent protein kinases, CDPK genes, were found in the *cap3p90 assembly* (*i.e. C*AP3 using a highly stringent parameter of 40-bp overlap length with 90% identity). We designated these genes as IbCDPK2 (Contig26263.2) and IbCDPK3 (Contig34516.2); they are highly similar at a protein and nucleotide level (Figure 3.4). The consensus sequences (Contig22153.1 and Contig28899.1) corresponding to both IbCDPK2 and IbCDPK3, respectively, in the *cap3p80 assembly* were assembled not differently but only 3 additional sequences were included in Contig22153.1 and reads from both stem and root 454 pyrosequencing conformed to these contigs (data not shown); Further examples of hidden sequence variations and discovery of novel genes by manual comparison of sequences are the genes encoding calcium binding proteins (IbCBP1, IbCBP2 and IbCBP3) (data not shown) and the BELL-Like genes included in this study (Appendix B) .

Sequences of IbCBP1 and IbCBP3 assemblies, corresponding to two sweetpotato calcium binding proteins, reported in Table 3.2, Table 3.8 and in the supplementary data at Appendix C are noteworthy. RT_055908.1, RT_027290.1, RT_483640.1 and JG699346.1 are the accession names of representative raw sequences included in Contig13109.2 (34 singlets) identified for

designing specific primers for IbCBP1. RT_038063.1, RT_483640.1 and JG699262.1 are also the accession names of representative raw sequences included in Contig20410.2 (18 singlets) identified for designing specific primers for IbCBP3. RT_038063.1 read from IbCBP3 was corrected, by deleting an additional nucleotide adenine (sequencing error), after visual inspection of alignment to the rest of sequences included in Contig20410.2 (accession JG699346.1 included in this contig). A comprehensive analysis was done in the selection of genes and their corresponding primer design, and both IbCBP1 and IbCBP3 are examples of the approach followed in this study. Surprisingly, some sequences such as JG699346.1 and JG699262.1 that were included in two separate contigs in *cap*3p90 assembly, were merged in a single Contig10947.1 (45 reads) of *cap3p80v3* assembly. Moreover, some single sequences forming Contig20410.2 (IbCBP3) were assembled in two separate contigs, Contig9867.1 (10 reads) and Contig10947.1 in the *cap3p80v3* assembly. Therefore, the best assemblies to representing IbCBP1 and IbCBP3 are derived from the conserved assembly (*cap3p90* assembly) using 40-bp overlap length with 90% identity and not 80% identity (*cap3p80v3* assembly).

Analysis of the transcriptome and differential expression of transcripts: Real time PCR analysis revealed that 21 unigenes (transcripts) are differentially expressed among SR, DSR, and FR (Table 3.8), and 20 represent novel Differentially Expressed Transcripts (DETs) (excluding Ibkn2, Ibkn3a, Ibkn3b) upregulated in DSR. Sixteen DETs are novel and upregulated in both young expanding storage roots (SR) and in the putative initiating storage roots (DSR). Semiquantitative RT-PCR confirmed the differential expression of these genes. Interestingly, the specific expression in roots and an apparent enhanced abundance of transcripts in either SR or DSR or both was observed for 11 transcripts corresponding to IbMAP65, IbCBP1, IbCBP2, IbCBP3, IbCRF1, IbHB1, IbHB2, IbBEL3, Ibkn3a, Ibkn3b, and IbIAA supported by the RT-PCR (Figure 3.2). In addition, a preferential upregulation in storage roots vs. fibrous roots was

apparent for five genes: IbNAM, IbCDPK2, IbCDPK3 IbPP2A, and IbTAP. The specific expression in roots was confirmed for a set of known genes corresponding to transcription factors, Ibkn2 and Ibkn3 (Tanaka et al., 2008), the starch-related enzymes, ADP-glucose pyrophosphorylase (AGPase) (Kim et al, 2002b, McGregor, 2006) and β-Amylase.

In total we found five calcium-binding proteins (CBP) with abundant transcripts in young storage roots and this represents the first report of this genes associated with the development of sweetpotato storage roots. These genes are IbCBP1, -2, -3 and IbCDPK1,-2. Previously, McGregor (2006) found a calcium-dependent protein kinase (DV037296) that is notably upregulated in storage roots, here referred to as IbCDPK1 (its corresponding contigs in PBL assembly are S_PBL_c3183 and S_PBL_c3183; Contig6182.2 in *cap3p90* assembly, Appendix D). Calcineurin-B-like (CBL) genes are among the top genes with the highest transcripts counts in the sweetpotato root transcriptome, PBL assembly (data not shown). Downstream targets of a calcineurin-B-like (CBL1) are stress responsive genes such as Rd29A/B, KIN1/2, and RD22 (Cheong et al., 2003). All three sweetpotato CBP-like proteins products (IbCBP1, IbCBP2, IbCBP3) were found to have a Ca^{2+} binding site resembling the typical "EF hand" motif of calcium (McCormack et al., 2005) composed of E and F helices, flanking a Ca^{2+}- binding loop (data not shown). The EF hand is a helix-loop-helix structural domain found in a large family of calcium-binding proteins http://pfam.sanger.ac.uk/family/PF00036). Further analysis of IbCBP1, IbCBP2, IbCBP3 revealed that they contain a pair of EF-hand motifs, similar to the majority of EF-hand calcium-binding proteins. A pair of EF-hand motifs is found in proteins implicated in endocytosis, vesicle transport, and signal transduction (http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cd00052). Typically only one EF-hand is canonical and binds to Ca^{2+}, while the other N-terminal is a pseudo EF-hand loop that does not bind to Ca^{2+}. IbCBP2 is an ortholog of At2g33990, a calmodulin binding protein found to be co-expressed in *Arabidopsis*

with two aquaporins (At1g01620, At4g23400) and one SEC14 cytosolic factor family proteins (At1g72160) (http://webs2.kazusa.or.jp/kagiana/cgi-bin/gcft.cgi? query=At2g33990 &organism=ath). Our analysis revealed that the sweetpotato root transcriptome (PBL assembly) is notably abundant in these types of genes, with 107 and 114 sequences for aquaporins and SEC14, respectively (data not shown). A tonoplast aquaporin (TIP) gene was found to be expressed in fibrous roots but not in pigmented (putative storage roots) and storage roots (Kim et al., 2008). A sequence similar to sweetpotato TIP, contig_8050 (JP112638.1) is also downregulated in storage root and studied by DGE analysis (Tao et al., (2012). Thus, aquaporins and a network of calcium signaling appears to occurs in the development of roots in sweetpotato and we could predict they have specific roles in both storage and non-storage roots. CDPKs and aquaporins encoding genes have been found involved in tuber formation in potato (Raices et al., 2003; Sarkar, 2008) and a CDPK is documented to be important during storage root initiation in cassava (Sojikul et al., 2010). A calmodulin gene termed PCBP (Potato calmodulin-binding protein) has been isolated from potato tubers (Reddy et al., 2002), and other work has demonstrated that expression of calmodulin genes are required for proper tuber formation (Poovaiah et al., 1996). The later work demonstrated that expression of CBP-like genes in potato produced elongated tubers. Importantly, many CDPK and calmodulin-binding proteins have been associated with the regulatory activities in the nucleus (Kim et al., 2009; Snedden and Fromm, 2001) that include DNA replication, DNA degradation during programmed cell death, cell cycle regulation, and transcription. CBP and CDPK in this study might participate in this kind of signaling from cytosol to nucleus signaling. Studies of PCBP suggest that its action may require nuclear localization (Reddy et al., 2002). It is unknown in sweetpotato the putative targets and the molecular events related to CBPs and CDPKs. The present work has discovered genes like CBP and CDPK and potential roles in storage root formation. Serine/Threonine-phosphatases

encoding genes are crucial in molecular signaling together with kinases and hormones. Both the products of IbPP2A and IbTAP genes, which were found upregulated in SR (Table 3.8), are examples of a signaling regulatory mechanism newly reported for sweetpotato, with IbPP2A being a serine-threonine phosphatase and IbTAP encoding a regulatory subunit of phosphatases. Apparently, PP2A-like genes are involved in the pathway of Gibberellic Acid (GA) signaling; supporting data comes from studies of the expression of a potato gene encoding a protein phosphatase type 2A catalytic subunits (PP2Ac) that is reported associated with source (leaf) to sink (tuber) signaling in tuber formation. GA is a recognized a negative regulator of the events that trigger the formation of potato tubers and sweetpotato storage root. Foliar expression of PP2Ac gene in potato was observed to be inhibited by GA, while tuber-specific genes were expressed in both leaf and potato tubers following activation of the PP2A protein (Pais et al., 2010). PP2A transcripts were reported to be enriched in the different stages of roots of sweetpotato (Tao et al., 2012). Equally, StCDPK1, a potato CDPK have been documented to play a role in a converging point of GA-signaling associated with modulating both promoting and inhibitory signals in the onset of potato tuberization (Gargantini et al., 2009). Moreover, sequential activation of CDPK protein products by post-transcriptional modifications might be critical regulatory steps of calcium signaling during potato tuberization (Raices et al., 2003). Two gibberellin-responsive protein transcripts similar to GASA5 (At3g02885.1) (accession DV034646) and GASA2 (At1g74670.1) (DV036052) were found slightly upregulated in storage roots of sweetpotato (McGregor and LaBonte, 2006). Whether IbPP2A and IbTAP products participate in the integration of signals that modulates the GA-signaling during the development of the storage organ is unknown in sweetpotato.

The BELL (BEL1-Like) genes detected in this study represent the first reported of characterization of these genes in sweetpotato (Table 3.8, Appendix B and Fig 3.2). Two KNOX-

like genes, Ibkn1 and Ibkn2 have previously been shown to be induced in developing and mature storage roots (Tanaka et al., 2008). Our data showed that transcripts of Ibkn2 gene were upregulated in both SR and DSR stages (2-4 times) compared to FR (Table 3.8), confirming its suspected role in the onset of storage root formation (Tanaka et al., 2008). Our analysis of the sweetpotato transcriptome revealed the existence of two putative alleles of the KNOXI gene Ibkn3, referred to as Ibkn3a and Ibkn3b. BELL and KNOX (knotted1-like homeobox) genes are part of the two HD TALE gene families known in plants (Burglin, 1997; Hake et al., 2004). Both KNOX and BELL are transcription factors that form homo- and hetero-dimmers in plants. In potato, the interaction of proteins from StBEL5 (BELL) and POTH1 (KNOX) genes have been found to be essential for repression of their target gene, *ga20 oxidase1* (*ga20ox1*) (Chen et al., 2004a). For example, the *Arabidopsis* At5G02030 gene product, orthologous of Ibkn3, is predicted to interact with other members of the KNOX gene family (BioGrid (http://thebiogrid.org/). Among these genes are the *Arabidopsis* ortholog of Ibkn1, STM (SHOOT MERISTEMLESS) gene, and BP (BREVIPEDICELLUS) gene (Bellaoui et al., 2001) ortholog of Ibkn2/Ibkn3. We hypothesize that up or downregulation of BELL genes might confound the formation of active dimmers and impact storage root formation, i.e., reduced number. Our analysis of the available transcriptome of sweetpotato reveals at least 5 different BELL genes some of which appears to be differentially expressed in storage roots and fibrous roots (data not shown). Surprisingly, expression of Ibkn3b in fibrous and storage roots was inconsistent as assessed by real time PCR; inadequate primer design is likely.

Digital gene expression analysis, DGE analysis in the PBL assembly corroborated the enhanced expression of all genes reported here (N. Firon, communication personal, 2011) and also decreased expression both IbBEL1-like transcripts (accession S_PBL_c43041, S_PBL_lrc26237 in PBL assembly) and IbCRF2 (accession S_PBL_c5300). In the same study,

abundance of IbCRF1 (accession S_PBL_c3693) is downregulated in the initiating storage roots vs. lignified roots. Both CRF1 and CRF2 genes encode cytokinin response factors that are known to participate in cell division in *Arabidopsis* (Day et al., 2008). Likewise, DGE analysis of the sweetpotato transcriptome by Tao et al. (2012) included many transcripts similar to the ones reported here. Their results of comparing the expression in the tagged tissues as "initial tuberous roots" (ITR) vs. "expanding tuberous roots" (ETR) confirms the upregulation of IbAGP1ase, IbBBR, IbBEL4, IbHB2, IbPP2A, IbCBP2, Ibkn3a, and IbTAP-like transcripts in ITR vs. ETR. IbBME gene, encoding a GATA-type zinc finger protein and IbHLH1 was upregulated in sweetpotato storage roots. Members of GATA, bHLH, AP2/EREBP and MYB TFs are known to be induced by cytokinins (Kiba et al., 2005). Although we were not able to test the IbBHLH1 gene in the sqRT-PCR, DGE analysis data (data not shown) results had 140 Illumina reads derived from storage root libraries mapped to IbHLH1 (accession S_PBL_c21769) with no mapped reads from lignified root libraries, supported our results of upregulation of this transcript in storage roots. IbBME (S_PBL_c31897) had with a fold change ratio of normalized counts in storage roots vs. lignified roots ~ 7.8 (103 reads from SR/13.2 reads from non-SR), while IbXX (S_PBL_c3428) had fold changes ratio ~100 (610/6) (N. Firon, personal communication, 2011). However, DGE results are informative but not conclusive in complex organism like sweetpotato. DEG is based counting the number of short reads from multiple tissues by mapping in a set of a "non-paradigm" reference assembled unigene of the transcriptome. DGE analysis can be confounded by the plasticity of the expressed genes and their similarities of paralogous and orthologous genes evident in the sweetpotato genome. Single experimental techniques like qRT-PCR, hybridization-based methods and cloning are still the golden options to clarify the role (s) of specific genes in sweetpotato.

Enhanced transcript populations of calcium signaling genes, novel transcription factors such as BELL genes in the developing storage roots are evident in the present study apart from previously described transcription factors. Our data showed that unigenes from our assemblies based on combined approach of ABySS and CAP3 apparently has multiple, shorter overlapping sequences with more than 90% identity in the SPTSA assembly (Tao et al., 2012). Comparative analysis of selected transcripts from this work with those of the SPTSA assembly (Tao et al., 2012) suggests that further improvement is possible by combining datasets. Although, the estimated number of expressed genes based on our work is around ~75K to 101K, an *Ipomoea* genome reference would aid in identifying protein-coding genes and to validate the de novo assemblies of sweetpotato. In general, the total protein coding unigenes that were annotated in the current assembly are similar to related plants.

3.5. Literature cited

Bellaoui M., Pidkowich M.S., Samach A., Kushalappa K., Kohalmi S.E., Modrusan Z., Crosby W.L., Haughn G.W. (2001) The *Arabidopsis* BELL1 and KNOX TALE homeodomain proteins interact through a domain conserved between plants and animals. Plant Cell 13:2455-2470.

Burglin T.R. (1997) Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. Nucleic Acids Research 25:4173-4180.

Chen H., Banerjee A.K., Hannapel D.J. (2004a) The tandem complex of BEL and KNOX partners is required for transcriptional repression of ga20ox1. Plant Journal 38:276-284.

Chen T.E., Huang D.J., Lin Y.H. (2004b) Isolation and characterization of a serine protease from the storage roots of sweet potato (*Ipomoea batatas* L. Lam). Plant Science 166:1019-1026.

Cheong Y.H., Kim K.N., Pandey G.K., Gupta R., Grant J.J., Luan S. (2003) CBL1, a calcium sensor that differentially regulates salt, drought, and cold responses in *Arabidopsis*. Plant Cell 15:1833-1845.

Day R.C., Herridge R.P., Ambrose B.A., Macknight R.C. (2008) Transcriptome Analysis of Proliferating *Arabidopsis* Endosperm Reveals Biological Implications for the Control of

Syncytial Division, Cytokinin Signaling, and Gene Expression Regulation. Plant Physiology 148:1964-1984.

Desai D.P. (2008) Understanding the genetic basis of storage root formation along with starch and betacarotene biosynthesis and their inter-relation in Sweetpotato (*Ipomoea batatas* LAM.). (Doctoral dissertation), University of Natural Resources and Life Sciences,Vienna. pp. 144.

Eguchi T., Yoshida S. (2008) Effects of application of sucrose and cytokinin to roots on the formation of tuberous roots in sweetpotato (*Ipomoea batatas* (L.) Lam.). . Plant Root 2:7-13.

Firon N., LaBonte D.R., Villordon A.Q., McGregor C., Kfir Y., Pressma E. (2009) Botany and Physiology: Storage Root Formation and Development, in: G. T. G. Loebenstein (Ed.), The Sweetpotato, Springer Science+Business Media B.V. pp. 13-26.

Gargantini P.R., Giammaria V., Grandellis C., Feingold S.E., Maldonado S., Ulloa R.M. (2009) Genomic and functional characterization of StCDPK1. Plant Molecular Biology 70:153-72.

Hake S., Smith H.M.S., Holtan H., Magnani E., Mele G., Ramirez J. (2004) The role of KNOX genes in plant development. Annual Review of Cell and Developmental Biology 20:125-151.

Jaillon O., Aury J.-M., Noel B., Policriti A., Clepet C., Casagrande A., Choisne N., Aubourg S., Vitulo N., Jubin C., Vezzi A., Legeai F., Hugueney P., Dasilva C., Horner D., Mica E., Jublot D., Poulain J., Bruyere C., Billault A., Segurens B., Gouyvenoux M., Ugarte E., Cattonaro F., Anthouard V., Vico V., Del Fabbro C., Alaux M., Di Gaspero G., Dumas V., Felice N., Paillard S., Juman I., Moroldo M., Scalabrin S., Canaguier A., Le Clainche I., Malacrida G., Durand E., Pesole G., Laucou V., Chatelet P., Merdinoglu D., Delledonne M., Pezzotti M., Lecharny A., Scarpelli C., Artiguenave F., Pe M.E., Valle G., Morgante M., Caboche M., Adam-Blondon A.-F., Weissenbach J., Quetier F., Wincker P., French-Italian P. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463-U5.

Jasinski S., Piazza P., Craft J., Hay A., Woolley L., Rieu I., Phillips A., Hedden P., Tsiantis M. (2005) KNOX action in *Arabidopsis* is mediated by coordinate regulation of cytokinin and gibberellin activities. Current Biology 15:1560-1565.

Kiba T., Naitou T., Koizumi N., Yamashino T., Sakakibara H., Mizuno T. (2005) Combinatorial microarray analysis revealing *Arabidopsis* genes implicated in cytokinin responses through the His -> Asp phosphorelay circuitry. Plant and Cell Physiology 46:339-355.

Kim M.C., Chung W.S., Yun D.-J., Cho M.J. (2009) Calcium and Calmodulin-Mediated Regulation of Gene Expression in Plants. Molecular Plant 2:13-21.

Kim S.H., Hamada T., Otani M., Shimada T. (2005) Isolation and characterization of MADS box genes possibly related to root development in sweetpotato (*Ipomoea batatas* L. Lam.). Journal of Plant Biology 48:387-393.

Kim S.H., Mizuno K., Fujimura T. (2002a) Isolation of MADS-box genes from sweet potato (*Ipomoea batatas* (L.) lam.) expressed specifically in vegetative tissues. Plant and Cell Physiology 43:314-322.

Kim S.H., Mizuno K., Sawada S., Fujimura T. (2002b) Regulation of tuber formation and ADP-glucose pyrophosphorylase (AGPase) in sweet potato (*Ipomoea batatas* (L.) Lam.) by nitrate. Plant Growth Regulation 37:207-213.

Kim Y.H., Yang K.S., Ryu S.H., Kim K.Y., Song W.K., Kwon S.Y., Lee H.S., Bang J.W., Kwak S.S. (2008) Molecular characterization of a cDNA encoding DRE-binding transcription factor from dehydration-treated fibrous roots of sweetpotato. Plant Physiology and Biochemistry 46:196-204.

Ku A.T., Huang Y.-S., Wang Y.-S., Ma D., Yeh K.-W. (2008) IbMADS1 (*Ipomoea batatas* MADS-box 1 gene) is involved in tuberous root initiation in sweet potato (*Ipomoea batatas*). Annals of Botany 102:57-67.

Kumar S., Blaxter M.L. (2010) Comparing de novo assemblers for 454 transcriptome data. BMC Genomics 11.

McCormack E., Tsai Y.-C., Braam J. (2005) Handling calcium signaling: *Arabidopsis* CaMs and CMLs. Trends in Plant Science 10:383-389.

McGregor C. (2006) Differential expression and detection of transcripts in sweetpotato (*Ipomoea batatas* (l.) lam.) using cdna microarrays, Louisiana State University, Baton Rouge, USA.

McGregor C.E., LaBonte D.R. (2006) Differential expression of genes between storage roots of sweetpotato cultivars jewel and white jewel. J Am Soc Hortic Sci. 131:798-805.

Nagata T., Saitou K. (2009) Regulation of Expression of D3-type Cyclins and ADP-Glucose Pyrophosphorylase Genes by Sugar, Cytokinin and ABA in Sweet Potato (*Ipomoea batatas* Lam.). Plant Production Science 12:434-442.

Nakatani M., Komeichi M. (1991) Changes in the Endogenous Level of Zeatin Riboside, Abscisic-Acid and Indole Acetic-Acid during Formation and Thickening of Tuberous Roots in Sweet-Potato. Japanese Journal of Crop Science 60:91-100.

Noh S.A., Lee H.S., Huh E.J., Huh G.H., Paek K.H., Shin J.S., Bae J.M. (2010) SRD1 is involved in the auxin-mediated initial thickening growth of storage root by enhancing proliferation of metaxylem and cambium cells in sweetpotato (*Ipomoea batatas*). Journal of Experimental Botany 61:1337-1349.

Ohto M.A., Hayashi K., Isobe M., Nakamura K. (1995) Involvement of Ca^{2+} signaling in the sugar-inducible expression of genes-coding for sporamin and beta-amylase of sweetpotato. Plant Journal 7:297-307.

Pais S.M., Garcia M.N.M., Tellez-Inon M.T., Capiati D.A. (2010) Protein phosphatases type 2A mediate tuberization signaling in Solanum tuberosum L. leaves. Planta 232:37-49.

Peng F.Y., Reid K.E., Liao N., Schlosser J., Lijavetzky D., Holt R., Zapater J.M.M., Jones S., Marra M., Bohlmann J., Lund S.T. (2007) Generation of ESTs in Vitis vinifera wine grape (Cabernet Sauvignon) and table grape (Muscat Hamburg) and discovery of new candidate genes with potential roles in berry development. Gene 402:40-50.

Poovaiah B.W., Takezawa D., An G., Han T.J. (1996) Regulated expression of a calmodulin isoform alters growth and development in potato. Journal of Plant Physiology 149:553-558.

Potato Genome Sequencing Consortium X.X., Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, Wang J, Orjeda G, Guzman F, Torres M, Lozano R, Ponce O, Martinez D, De la Cruz G, Chakrabarti SK, Patil VU, Skryabin KG, Kuznetsov BB, Ravin NV, Kolganova TV, Beletsky AV, Mardanov AV, Di Genova A, Bolser DM, Martin DM, Li G, Yang Y, Kuang H, Hu Q, Xiong X, Bishop GJ, Sagredo B, Mejía N, Zagorski W, Gromadka R, Gawor J, Szczesny P, Huang S, Zhang Z, Liang C, He J, Li Y, He Y, Xu J, Zhang Y, Xie B, Du Y, Qu D, Bonierbale M, Ghislain M, Herrera Mdel R, Giuliano G, Pietrella M, Perrotta G, Facella P, O'Brien K, Feingold SE, Barreiro LE, Massa GA, Diambra L, Whitty BR, Vaillancourt B, Lin H, Massa AN, Geoffroy M, Lundback S, DellaPenna D, Buell CR, Sharma SK, Marshall DF, Waugh R, Bryan GJ, Destefanis M, Nagy I, Milbourne D, Thomson SJ, Fiers M, Jacobs JM, Nielsen KL, Sønderkær M, Iovene M, Torres GA, Jiang J, Veilleux RE, Bachem CW, de Boer J, Borm T, Kloosterman B, van Eck H, Datema E, Hekkert BL, Goverse A, van Ham RC, Visser RG. (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189-195.

Prosdocimi F., Bittencourt D., da Silva F.R., Kirst M., Motta P.C., Rech E.L. (2011) Spinning Gland Transcriptomics from Two Main Clades of Spiders (Order: Araneae) - Insights on Their Molecular, Anatomical and Behavioral Evolution. PLoS ONE 6:e21634.

Raices M., Gargantini P.R., Chinchilla D., Crespi M., Tellez-Inon M.T., Ulloa R.M. (2003) Regulation of CDPK isoforms during tuber development. Plant Molecular Biology 52:1011-24.

Reddy A.S.N., Day I.S., Narasimhulu S.B., Safadi F., Reddy V.S., Golovkin M., Harnly M.J. (2002) Isolation and characterization of a novel calmodulin-binding protein from potato. Journal of Biological Chemistry 277:4206-4214.

Sarkar D. (2008) The signal transduction pathways controlling in planta tuberization in potato: an emerging synthesis. Plant Cell Reports 27:1-8.

Schafleitner R., Tincopa L.R., Palomino O., Rossel G., Robles R.F., Alagon R., Rivera C., Quispe C., Rojas L., Pacheco J.A., Solis J., Cerna D., Kim J.Y., Hou J., Simon R. (2010) A sweetpotato gene index established by de novo assembly of pyrosequencing and Sanger sequences and mining for gene-based microsatellite markers. BMC Genomics 11:604.

Schneider M., Lane L., Boutet E., Lieberherr D., Tognolli M., Bougueleret L., Baiyoch A. (2009) The UniProtKB/Swiss-Prot knowledgebase and its Plant Proteome Annotation Program. Journal of Proteomics 72:567-573.

Snedden W.A., Fromm H. (2001) Calmodulin as a versatile calcium signal transducer in plants. New Phytologist 151:35-66.

Sojikul P., Kongsawadworakul P., Viboonjun U., Thaiprasit J., Intawong B., Narangajavana J., Svasti M.R.J. (2010) AFLP-based transcript profiling for cassava genome-wide expression analysis in the onset of storage root formation. Physiologia Plantarum 140:189-298.

Tanaka M., Kato N., Nakayama H., Nakatani M., Takahata Y. (2008) Expression of class I knotted1-like homeobox genes in the storage roots of sweetpotato (*Ipomoea batatas*). Journal of Plant Physiology 165:1726-1735.

Tao X., Gu Y.H., Wang H.Y., Zheng W., Li X., Zhao C. (2012) Digital Gene Expression Analysis Based on Integrated De Novo Transcriptome Assembly of Sweet Potato [*Ipomoea batatas* (L.) Lam.]. PLoS ONE 7:e36234.

Villordon A., LaBonte D., Solis J., Firon N. (2012) Characterization of Lateral Root Development at the Onset of Storage Root Initiation in 'Beauregard' Sweetpotato Adventitious Roots. Hortscience 47:961-968.

Villordon A.Q., La Bonte D.R., Firon N., Kfir Y., Pressman E., Schwartz A. (2009) Characterization of Adventitious Root Development in Sweetpotato. Hortscience 44:651-655.

Wall P.K., Leebens-Mack J., Chanderbali A.S., Barakat A., Wolcott E., Liang H.Y., Landherr L., Tomsho L.P., Hu Y., Carlson J.E., Ma H., Schuster S.C., Soltis D.E., Soltis P.S., Altman N., dePamphilis C.W. (2009) Comparison of next generation sequencing technologies for transcriptome characterization. BMC Genomics 10:-.

Wang Z., Fang B., Chen J., Zhang X., Luo Z., Huang L., Chen X., Li Y. (2010) De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). BMC Genomics 11:726.

Xie F., Burklew C.E., Yang Y., Liu M., Xiao P., Zhang B., Qiu D. (2012) De novo sequencing and a comprehensive analysis of purple sweet potato (*Ipomoea batatas* L.) transcriptome. Planta 236:101-113.

Xie Q., Frugis G., Colgan D., Chua N.H. (2000) *Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. Genes Dev. 14:3024-3036.

Yeh K.W., Chen J.C., Lin M.I., Chen Y.M., Lin C.Y. (1997) Functional activity of sporamin from sweet potato (*Ipomoea batatas* Lam): A tuber storage protein with trypsin inhibitory activity. Plant Molecular Biology 33:565-570.

You M.K., Hur C.G., Ahn Y.S., Suh M.C., Jeong B.C., Shin J.S., Bae J.M. (2003) Identification of genes possibly related to storage root induction in sweetpotato. Febs Letters 536:101-105.

Chapter 4. Construction and application of a sweetpotato microarray for gene expression profiling during sweetpotato root development

4.1 Introduction

The level of expression of a gene is commonly estimated using two analytic approaches referred to as 'analog' and 'digital' (Audic and Claverie, 1997; Eujayl and Morris, 2009). The analog methods are based on oligonucleotide probe hybridizations such as Northern blotting, mRNA differential display, and DNA microarrays (Eujayl and Morris, 2009). The digital methods are based on high throughput sequencing and bioinformatic analysis of transcripts from different libraries. Two analog methods commonly used in expression profiling are quantitative reverse transcription polymerase chain reaction (qRT-PCR) and microarrays. In contrast to microarray, qRT-PCR is a gene-specific assay and requires specific pairs of primers and/or a probe. Both microarray and qRT-PCR have been used to identify genes in the development of storage roots in sweetpotato (Desai, 2008; McGregor, 2006). A third technology termed RNA-seq (Wang et al., 2009) emerged as a consequence of availability of deep sequencing technologies and allows for digital quantification of the expression of genes. This approach is called digital gene expression (DGE) analysis. By using this technology, two recent studies demonstrated a comprehensive descriptions of the sweetpotato transcriptome in both multiple tissues (Tao et al., 2012) and those related to the secondary metabolism of purple sweetpotato roots (Xie et al., 2012) . However, results were inconsistent for genes previously described and for closely related transcripts.

In the present study, a new custom Sweetpotato Oligo Microarray based on the 4x44K format was developed on the Agilent platform. The oligo array was built using probes for 14K transcripts from sweetpotato, and was used to analyze gene expression profiles in storage and non storage roots of sweetpotato. The objectives were to build a custom microarray and to

analyze gene expression levels between early developed storage roots and white fibrous, non storage roots from 4 weeks old plants. The work was carried out in order to identify potential genes associated with the formation of storage roots by generating a transcriptome for sweetpotato, including information from sequencing libraries of initiating storage forming roots and lignified roots.

4.2 Materials and methods

4.2.1 Biological sample and extraction of RNA

Six independent plants from 'Beauregard' were grown inside the greenhouse under a day/night temperature regime of 29.6/26.8 °C and 14 h day light. Storage root and fibrous roots were sampled at 4 weeks after planting. RNA was prepared using the Direct-zol RNA miniprep kit (Zymo Research, Irvine, CA) and Trizol reagent (Invitrogen, Carlsbad, CA) according to the manufactures' instructions including a DNase treatment using 200 mg of ground tissue of storage (SR) and fibrous roots (FR). The concentration was determined by spectrophotometry, and RNA integrity was assessed by agarose gel-electrophoresis. Maximum width (diameter) SR and FR roots were <1.5mm for fibrous roots and >1.5 to ~3 mm, respectively, representing early stages of the induction of storage root formation and most were expanding storage roots.

4.2.2 Custom microarray development and probe design

A custom Sweetpotato Agilent 4x44K array was designed using ~14K (13,843 unique probesId/features) replicated three times in each array. This oligo array is referred to as SPOArrayv1 (Sweetpotato Agilent Oligonucleotide Microarray). Sequences (>13K) derived from a multiple gene index were used for the probe design. The majority of probes (13,788) were from unigenes from CAP3, Newbler and PBL assemblies, while the remaining were from unigenes of the Sweetpotato Gene Index (Schafleitner et al., 2010) and Expressed sequence Tags (ESTs) from GenBank. A total of 12,899 sequences were derived from CAP3 (Huang and

Madan, 1999) and Newbler-based assemblies, but all 12,899 probes matched identifiable unigenes in CAP3 assembly. A specific set of 53 selected unigenes (Appendix E) were treated differently. Two to three different probes were designed for each transcript (Appendix E, Table 5.1). Included in this set are sequences that were used as positive, negative or reference controls. Positive control probes were designed by selecting genes known to be up or downregulated in the storage root development such as sequences of Ibkn2, Ibkn4, Sporamin A and B, β-Amylase, small subunit of ADP-glucose pyrophosphorylase alpha subunit (IbAGPa1), cysteine proteinase inhibitor, UDP-glucose glucosyltransferase (Sucrose synthase, SuSys), starch phosphorylase, tonoplast instrinsic protein (TIP), fructokinase-like and Cyclin D3(CYCD3;2). The reference control included probes for alpha tubulin, Actin, Glyceraldehyde 3-phosphate dehydrogenase (GADPH), ubiquitin-conjugating enzyme E2 (UBCE) and elongation factor-1 alpha (IbEF1a). Two negative control sequences were also included in the microarray corresponding to 2 sequences: *Escherichia coli* hygromycin B phosphotransferase (Accession NM_003140.1, seven probeid in the array for a single probe sequence) and *Homo sapiens* sex determining region Y (Accession NM_003140.1, two probeid in the array for a single probe sequence). These sequences are additional to Agilent's internal controls.

Probes were designed using the coding strand (sense strand) and evaluated on an Agilent Array platform using Picky software (Chou, 2010). Oligos (60-mer) were preferentially derived from the 3′-end of each transcript. Multiple probes were designed for each transcript and at least one probe for each transcript was chosen to be included in the custom array. Included in the ~14K unique features were 837 sequences from non-coding unigenes from the PBL assembly. Each probe was randomly replicated 3 times in the array.

4.2.3 Sample preparation for microarray hybridizations, measurement of fluorescence signals and data analysis

Two hundred ng of total RNA derived from storage roots and fibrous were labeled using Agilent's Low Input Quick Amp Labeling Kit (Agilent Technologies, Santa Clara, CA, USA) and other reagents following the manufacturer's instructions. The kit included the dyes Cyanine 3-CTP (Cy3) and Cyanine 5-CTP (Cy5), which were used to generate cyanine 3- or cyanine 5-labeled cRNA. Each of the six replicate samples was dye-swapped by labeling with Cy3 and Cy5. The experimental hybridization was done in a balanced design which included a dye swap totaling 24 arrays (i.e. 12 arrays contained cRNA of FR mixed with cRNA samples of SR derived from same plant, and 12 arrays contained cRNA of FR mixed with cRNA samples of SR derived from a biological replicate plant). Following hybridization, slides were washed following the manufacturer's protocols, slide images were captured using the Packard Bioscience ScanArray Express (PerkinElmer Inc., Waltham, MA, USA), and spot intensities were assayed for each channel (Cy3 and Cy5) using ImaGene software (Biodiscovery Inc., El Segundo,CA, USA). Spots that showed artifacts due to intrinsic errors during the printing of probes or the experiment (e.g. air bubbles, spread of dies) were removed from the analysis during the acquisition of data. The fluorescence signal of each probe reported by ImaGene software was measured as arbitrary units. Data analysis was implemented in R2.14.2 software (R Development Core Team 2012) as described earlier (Baisakh et al., 2012). In brief, the $\log_2$ transformed data were normalized within slides using a global loss with no background correction (Ritchie et al. 2007). The data were then normalized between slides; all normalization procedures were done in limma (Smyth and Speed 2003). After normalization, the $\log_2$ mean value for each probe on each array for each dye was computed; the resulting values were used for further analysis. For differential expression analysis, a mixed model analysis of variance (ANOVA) using shrinkage

estimators (Cui et al. 2005) with tissue and dye as fixed effects and biological replicate, plant sample within biological replicate, slide, and array within slide as random effects was fit using ANOVA implemented in MAANOVA (MicroArray ANalysis Of Variance) package (Cui and Churchill 2003). Significance was determined from permutation-based P-values (from 1000 samples) for the tissue effect. P-values were adjusted for the large number of probes tested using the False Discovery Rate (FDR) test; Storey and Tibshirani 2003). Mean and standard error values are reported in $\log_2$ scale after normalization. Fold-changes of each transcript was calculated in a linear scale, where '0.5' indicates 50% larger, '1' means 100% larger, and '2' means 200% (2-fold) larger, and '-0.5' means 50% smaller. The estimation of fold change ratio for each probeid was calculated as follow, respect to fibrous roots, using the above mentioned mean values and the examples showed in Table 4.1:

-Fold-change ratio (FCR) (SR vs. FR) (in linear scale) = $2^{\text{mean (Fibrous root) –mean (storage root),}}$ if mean value (from SR) < mean value (FR), or

-Fold-change ratio (FCR) (SR vs. FR) (in linear scale) = (-1) x $2^{\text{mean (Storage root) –mean (Fibrous root),}}$ if mean value (SR) ≥ mean value (FR)

If the sign of the calculated fold change ratio is negative and the P-value is significant (FDR<0.01 or <0.05), it means that the gene is downregulated in storage roots and if the calculated fold change ratio is positive, the corresponding gene is upregulated in storage roots.

For example: fold change for Gene A is: (-1) $x2^{14.53-13.19}$= -2.53; similarly, fold change for Gene D is: (-1) $x2^{13.18-11.06}$= 4.34. Looking at the above example, a gene (i.e. transcript) is declared to be upregulated in storage roots if the fold change is positive and the associated adjusted P-value is less that the significance threshold (FDR<0.01 or FDR<0.05).

Table 4.1. Example of raw data for calculating fold change.

Gene (probeid)	Mean FR	Mean SR	Fold change	Adjusted P value	Annotation	Probeid (from real data)
A	13.19	14.53	-2.53	0.0012	Nitrate Transporter	contig27517
B	10.00	10.03	-1.03	0.3098	WRKY transcription factor	CUST_35912_PI427086615
C	11.98	13.23	2.38	0.0034	Fasciclin-like arabinogalactan protein 2	CUST_44129_PI427086615
D	13.18	11.06	4.34	0.0002	Sporamin	DQ195758.1_control1
E	14.26	12.92	2.53	0.0102	KNOX, Ibkn2	Ibkn2_AB283028.1probe1
F	14.63	14.60	1.02	0.4701	Elongation factor-1 alpha	IbEF1a_Contig35258.1probe1_control1
G	10.25	10.19	1.05	0.3701	Cop9 complex subunit, putative	IbCOP8_S_PBL_c26931probe1
H	10.13	8.95	2.28	0.0596	Polygalacturonase (RT_085368.1)	CUST_82798_PI427086615

4.2.4 Bioinformatics analysis

Sequence annotation was carried out by BLAST analysis (Altschul et al., 1998). All unigenes were searched against protein databases UniProtKB (www.uniprot.org, release UniProt release 2011_04) April 5, 2011) and TAIR10 protein database (http://www.arabidopsis.org/index.jsp) using a threshold of Expect (E) value <1e-6. In addition, Gene Ontology (GO) analysis was conducted using B2GO tool (Conesa et al., 2005) and carried out to classify the genes according to GO terms associated with Molecular Function, Biological Process and Cellular Component categories. GO terms were also found via the UniprotKB accession associated to each unigene.

4.2.5 Quantitative reverse transcription PCR

Twelve genes were selected based on the custom microarray results to confirm their expression changes by quantitative reverse transcription PCR (qRT-PCR). Primer pairs were designed using PRIMER3 and listed in Table 4.2. Amplification was carried out using the same RNA samples used for microarray hybridizations with the iScript cDNA synthesis kit (Bio-Rad,

Hercules, CA). In Brief, 1.3 µg of total RNA including 4 µl iScript reaction mix and 1 µl iScript reverse transcriptase were used for cDNA synthesis by incubating 5 minutes at 25° C, 30 minutes at 42° C, and 5 minutes at 85° C. The cDNA was diluted 1:3 with nuclease-free water and qRT-PCR was performed using iQ^{TM} SYBR Green supermix (Bio-Rad, Hercules, CA), according to the manufacturer's instructions. The conditions for the PCR amplification were as follows: polymerase activation at 95°C for 3 min; followed by 45 cycles of 95°C for 10 s, 60°C for 20 s, and 72°C for 20 s. The fluorescence signal was measured once every 1°C. Negative PCR controls (no cDNA template) were included in all reactions to detect possible contamination and the specificity of the primer pairs was checked by analysis of a melting curve. Each reaction was performed in triplicate, and the mean threshold cycle (Ct) was used to estimate the fold change ratio according to the delta-delta Ct method (Baisakh et al., 2012). The expression of each transcript was quantified relative to elongation factor-1a (IbEF1a) transcript levels. The fold changes were determined comparing the storage root vs. fibrous root.

4.3 Results

In the present study, we developed and used a new constructed custom Agilent microarray (hereinafter on referred to as SPOArrayv1), consisting of 13,843 unique probeid/features from sweetpotato. The majority of the probes were from unigenes from CAP3, Newbler and PBL assemblies, while the remaining was from unigenes of Sweetpotato Gene Index and ESTs from Genbank. 12,899 sequences were derived from CAP3 and Newbler, but all 12,899 probes have a matching contig in CAP3 assembly. A specific set of 53 unigenes were selected and 2-3 different probes were designed for each transcript (Appendix E) because these represent transcripts associated with differential expression in storage and fibrous roots (based on available literature or digital expression analysis of the PBL assembly and qRT-PCR, presented

in previous chapter). These probes are useful for analyzing gene expression profiles in storage and non-storage roots of sweetpotato.

A dye swap design was done with 12 hybridizations using 6 biological replicate samples of SR and FR from 4 week old plants. The samples used for this study were different than the samples used for the selected unigenes with qRT-PCR and sqRT-PCR in previous chapter. After the analysis, no arrays contained saturated spots. Of the 13,843 sweetpotato probes/features on the array, 372 (2.7%) had a signal indistinguishable from the background (defined as less than the mean of 153 negative control spots plus 2.6 times their standard error, or the upper bound of the 99% confidence interval) for 90% or more of all spots. All data were included in the normalization procedures and statistical analysis. Difference in expression for each transcript (each probe identified by a unique probeid) was based on difference in expression between tissues (storage root vs. fibrous root) and the variation within each tissue for each transcript to get the false discovery rate (FDR). Using a FDR cut off of <0.05, 4,948 (35.6%) of all features were significantly different between tissues (SR and FR), while the 0.01 FDR suggested 1,569 features (11.3%) were different for the 13,912 probes tested. The statistically significant differences in expression between fibrous and storage roots ranged from small (1.04 fold) to large. Relative to fibrous tissue, storage tissue had up to 2.98-fold higher expression or up to 12.61-fold lower expression. A detailed estimation of the number of differentially expressed transcripts (DETs) was done using an FDR test ($p<0.05$) combined with a cutoff fold change of 1.5 (for upregulated transcripts) or a cutoff fold change of -1.5 (for downregulated transcripts). The complete list of upregulated and downregulated transcripts in SR is given in the appendix appendix F and G. We found that 797 features corresponding to 783 unique transcripts were significantly upregulated in SR at FDR of 0.05 (Table 4.6 and Appendix F), while 342 features were downregulated in storage roots at FDR of 0.05 (Table 4.7 and Appendix G), and these

represented ~ 328 unique transcripts. The number of unannotated features with significant expression in storage root was 68 of out 783, corresponding to same number of unique transcripts (genes). Whereas, six out the 328 significantly downregulated features had no annotation. All of these DETs are potentially derived from genes specific for sweetpotato and hence are worth further consideration.

GO enrichment analysis was done using only 505 unique transcripts out of 797 features of upregulated features (fold change ratio >1.48), corresponding to different probeid sequences from the gene indexes used to built SPOArrayv1 array (i.e. each from a different annotated unigene).. Based on the GO classification of genes by "molecular function" a significant amount of upregulated genes were assigned in the following GO terms: kinase (74 genes, GO:0016301), transcription factor activity (51 features, GO:0003700), catalytic regulator activity (28 features, GO:0030234) and carbohydrate binding (22 features, GO:0030246). All the above mentioned GO terms, including those in Table 4.3, Table 4.4, and Table 4.5 were found to be significantly overrepresented after the GO enrichment analysis. The GO classification by "cellular component", showed that within the upregulated genes in storage root was overrepresented for cell wall (166 features, GO:0005618), intracellular membrane-bounded organelle (409, GO:0043231) and cytosol (103 features GO:0005829) and this included encoding products participating in basic cellular activities as well as the starch and sugar metabolism. Similarly, overrepresented GO terms of "biological process" of the upregulated genes corresponded to developmental process, anatomical structure morphogenesis, response to stress, carbohydrate metabolic process, and signal transduction (data not shown). Enrichment analysis was not done for downregulated genes due to their small numbers and lack of relevance to this work.

A list of selected transcripts that are detected significantly upregulated in storage root with fold change ratio >1.5 and downregulated in storage root with a fold change ratio <-1.5 are

in table 4.6 and 4.7. Similarly, known genes associated with storage roots are listed in table 4.5 We found at least five different transcripts of sucrose synthase with fold change ratio of 1.5 to ~3.7 time and β-Amylase which have a notable expression in SR vs. FR with 2.5 to 5 times. Enhanced expression was also revealed for novel genes (Table 4.5, Table 4.6, and Table 4.7) in sweetpotato, including a new transcription factor identified as BEL1-like transcript, which was termed as IbBEL7. Downregulation of IbBEL1 and IbCRF1 and IbCRF2 (two members of the AP2/EREBP protein family of transcription factors) transcripts was also observed in storage roots (Table 4.7). It may be noted that IbBEL1 and IbCRF1 transcription factors showed an enhanced expression in 2-week old roots from drought stressed plants (see chapter 2). In addition, transcripts of AP2/EREBP, bHLH and MYB transcription factors were differentially expressed in the SR and FR. Reduction in expression was observed for MYB (6 transcripts), WRKY (12 transcripts) and bHLH (1 transcript) in SR (Table 4.7 and Appendix G). Apart from many transcription factors, components of response mechanisms to biotic and abiotic stimuli and cellular transport were significantly upregulated in storage root, including the particular abundance of transcripts of nine aquaporins, one kinase (Contig1100.1, termed IbKIN1), four proline-rich proteins and 21 peroxidases (Table 4.6). Detail review of the consensus sequences of aquaporin transcripts showed that two out of the nine different aquaporin genes that were upregulated in storage roots (Contig046040.3 and Contig5724.1) were represented in the root libraries and they are partial sequences of the coding sequences and 3'-end of their putative genes. A single transcript of a gibberellin regulated protein (RT_091599.1) (Table 4.6, Table 4.8) was highly upregulated in SRcompared to FR and constitutes a novel report for this type of protein for sweetpotato. Further examples of genes associated with hormone signaling, also novel finding of this study, were the six auxin response transcription factors that were slightly upregulated in storage root. Most notably, among transcripts apparently significantly

downregulated in storage root (i.e. upregulated in fibrous roots; Table 4.7) were: a.) nitrate transporters (~ 10 unique transcripts); b.) ABC transporters [P-multidrug resistance/P-glycoprotein (MDR/PGP) (MDR/PGP)] (~ 20 unique transcripts) that mediate multiple basic developmental processes. An additional single transcript encoding a protein phosphatase 2c, was also slightly downregulated (had a fold change ratio of -1.5). In addition, transcripts associated with structural changes of the cell wall that were upregulated includes member of expansin (which is among the gene family with highest expression in expanding storage roots) and xyloglucan endotransglucosylase/hydrolase (XTH/EXGT) gene family (13 transcripts).

Twelve genes, including expansin, that were differentially expressed in microarray experiments, were selected for validation by qRT-PCR (Table 4.2 and Table 4.8). Out of these twelve genes, 11 were upregulated transcripts (Table 4.6) and one, corresponding to an invertase/pectin methylesterase (IbPMI) gene, was downregulated (Table 4.7). The consensus sequences of the majority of these 12 selected sequences were represented in the root transcriptome (data not shown). The results of qRT-PCR validated the observed data of the microarray (Table 4.8), although the fold change of expression in SR over FR was slightly higher than that observed in microarray results.

Interestingly, further review of the aligned singlets (reads) that clustered into the Contig16599.1, annotated as endo-xyloglucan transferase (sweetpotato gene termed IbXTH1) (Table 4.8), showed that the reads separated into two types with an insertion/deletion (indel) (Figure 4.1). The two types of sequences may represent allelic variants that were missed in the consensus sequence. The selected probe of Contig16599.1 (CUST_48523_PI427086615) overlapped nearly all reads. We targeted the primer pairs for only one allele. The qRT-PCR results showed upregulation of the transcript and support that a specific allele of this contig is upregulated in 4-week-old expanding storage roots.

Figure 4.1. **Detail view of Contig16599.1 showing the location of an indel in the 3'-UTR that suggests the presence of two alleles. The location of one designed probe (CUST_48523_PI427086615 in SPOArrayv1 oligoarray) is underlined in those sequences matching 100%. The annotation of this contig is endo-xyloglucan transferase, XHT/EXGT (accession O80431_TOBAC at Uniprot)**.

Out of the unannotated transcripts that were significantly upregulated in storage root, all probes derived from the next generation sequencing (NGS) data are derived from initiating storage roots vs. lignified roots from this study (unpublished data), one example corresponding to the transcript S_PBL_lrc26197 had 3.5 times as much expression in FR (Table 4.6). However, intensive BLAST analysis found that S_PBL_lrc26197 has an ORF of 174 nucleotides in length (Figure 4.2) that matches a short peptide (XP_003611068.1 – a hypothetical protein MTR_5g010060 from *Medicago truncatula*) of 90 amino acids (aa), having 52.8% identity (88.7% similarity) with 53 aa overlap. The putative protein product of S_PBL_lrc26197 was predicted to comprise of 58 aa. The difficulty of annotation of this transcript is explained due to its small size, and this may constitute a novel gene specific for sweetpotato. Similarly, we were able to annotate a second sequence (S_PBL_lrc49266), which had 3.95 times significant upregulation in SR, and correspond to the non-coding region of an expansin gene (DQ515800.1).

It was found that IbBEL7 had a similar transcript from sweetpotato (accession JP106340.1 at Transcriptome Shotgun Assembly databases at GenBank). Organellar sequences appear to be included in the set of initially unannotated sequences from the PBL assembly, because two of these sequences have a matching entry in sequences derived from chloroplast and mitochondria.

In order to validate the design of probes and to evaluate the usefulness of the developed custom microarray, a set of 42 known genes known to be upregulated or downregulated in storage roots were included. The results of this work are summarized in Table 4.5. Two to three different probes for each of these genes were replicated as different features in the array. The results of the microarray analysis presented a matching profile to the expected fold changes for 32 out of 42 genes (i.e. their expression showed concordance with the results of previous, independent studies) with a support of P value <0.05. For example, transcripts of sporamin A, sporamin B, β-amylase and Ibkn2 had the greatest fold change ratio (SR over FR) of 3.60, 4.3, 4.21, and 2.53 based on their corresponding probes (Table 4.5). Surprisingly, alternative probes for some genes such as Ibkn2 (Table 4.5) and alpha tubulin (data not shown), showed an apparent notable variation of their fold change ratio values.

Through this study, suitability of four commonly used housekeeping genes [alpha tubulin, actin, elongation factor-1 alpha (IbEF1a), and glyceraldehyde phosphate dehydrogenase (GAPDH)] was also investigated to be included as potential internal reference for normalizing qRT-PCR data in the sweetpotato root development. From the microarray results (Table 4.5), only IbEF1a showed stable expression in both SR and FR. The custom microarray contained a total of 588 spots (308, 56, and 56 spots of each of the three probes) for IbEF1a. Hence, the three features of IbEF1a were not significantly up/downregulated in the microarray analysis, suggesting that IbEF1a represents a reliable housekeeping gene for gene expression analysis of

A.

```
DNA:  aaggcaacctggtttcttctctaagtttggcaacacatcagcttccaaaatcaaggtctaagggcaatcctttgagctata
DNA:  atcaatcATGAATCAGAAAGTGAGTCAGCAGGCCTCAGCAGAGGCAATTAAGGGTTTCAGGAGCAGAAGGGCATTACCAAA
+2fr:  ·····M··N··Q··K··V··S··Q··Q··A··S··A··E··A··I··K··G··F··R··S··R··R··A··L··P··K·
DNA:  GAGGGGGCAGATCAAGTCAAGAATAGTGGCAAATACCTTCAAGTCCATTGTTTCTGTGCTCTCCAGAGCACCTTCTCATCA
+2fr:  ·R··G··Q··I··K··S··R··I··V··A··N··T··F··K··S··I··V··S··V··L··S··R··A··P··S··H··H·
DNA:  CTTGTTCCATCATGAATTCtaagtaatggtgattagagaacttctgayagtagtgtattaatggaagcatgtgtgatcatt
+2fr:  ·L··F··H··H··E··F··*··V··M··V··I··R··E··L··L··X··V··V··Y··*··W··K··H··V··*··S··F·
DNA:  taaattttggtttcccattttgttttcatccctgttctctcagctaacttatagtgttataatcatgttgtcatagcagca
DNA:  taaaaattaaatatatatatatatcattcaatagtaaacaaaaaaaaaaaaaaaanaaaa
```

B.

>S_PBL_lrc26197 protein product (matches two ESTs: EE884430.1 and EE884476.1)
MNQKVSQQASAEAIKGFRSRRALPKRGQIKSRIVANTFKSIVSVLSRAPSHHLFHHEF

>EE884430.1|EE884476.1 protein product
MNHKVSQQASAEAIKGFRSRRALPKRGQIKSRIVANTFKSIVSVLSRAPSHHLFHHEF

>Q5VS61_ORYSJ Putative uncharacterized protein P0644B06.43 OS=Oryza sativa subsp. Japonica

MGEKACKVASAHLIKVYKGEKQMRVRPLPRRGQVKSRIARIVMSAITSALVRALSQLPV
L

>XP_003611068.1

MNTRMGRVASAEVVNNRNRFRGKKSLPKRGQIKSKIAASAFHSIVSVISRASTSSGLVG
NVALVEKEFMFDSKSWSFEEQSNFMCKKLYV

Figure 4.2. Sequence analysis of transcript S_PBL_lrc26197: (A) Putative ORF of 59 amino acids in frame +2; (B) Predicted conserved protein similar to product of S_PBL_lrc26197 in accession sequences from Sweetpotato (EE884430.1, EE884476.1), grape (XP_003611068.), and rice (1Q5VS61_ORYSJ).

the sweetpotato storage root development. Therefore, qRT-PCR analysis was performed by evaluating the levels of expression of target genes for validating the microarray results by standardized expression of IbEF1a in the identical cDNA samples. The synthesized probes of IbEF1a transcripts in the custom microarray are:

IbEF1a_Contig35258.1probe1 ctctgagtacccaccacttggtcgttttgctgtgagggacatgagacaaacggttgcagt

IbEF1a_Contig35258.1probe2 ggttgtgatgggagcttttctggatgccatttattatacccatttgctattttcgatgtg

IbEF1a_Contig35258.1probe3 ggcaagacttttgtgggtgaacagttacaatttacctgttcaatgtcaagttttatttag

However, qRT-PCR experiments using primer targeting all the reference genes also confirmed the results of the microarray (data not shown).

Table 4.2. Primer sequence of selected genes for qRT-PCR validation and its probe used in the microarray.

Gene Symbol	Transcript (Unigene)	Primer	Primer Sequence	PCR product (bp)	Probe Sequence
IbCBF	Contig4451.1	Forward	cgcagcaagaacaacagaatgc	227	acgtggaagactgaaaacctaggaatttacgtactacaatcacgaatcacgacctagatt
		Reverse	tgtcgtgggaatcaagaatttgg		
IbPIP2	Contig5724.1	Forward	attggatttgcggtgttcatgg	242	tgggcactgtgagtgtgagccatgtgagtgtagatagatagggagtgcgtttttttataaa
		Reverse	cagacttgagaaccgctcctg		
IbEXP2	Contig9244.1	Forward	gtccttcggccagacctactc	240	aaggtaagatgaatgaaaccttcaagcttgttcagactttattgtcattcttataacccc
		Reverse	gctttcgatcaaaaccaatcc		
IbARF	Contig13201.1	Forward	catggcacgagttctgtgaag	193	tcttatggtactgtactatgagaatgccactaaaaagcttgtcctatgaatcccctctttt
		Reverse	ccttaaatcctgatcactgttgga		
IbPRP	Contig13241.1	Forward	ctcctcaatgcttgctccaagg	198	tgtatgcttgatctgagtggaaacagtcggaaagatactattctactttgctaatttcta
		Reverse	gaattccatccgacattcaaca		
IbXTH1	Contig16599.1	Forward	aagtccaggtgtgcaacacagt	204	actgttctattctatatatccatcgtcctgttggataatatatgggggagagatattgtt
		Reverse	gaacagtccctagcttcccttaaa		
IbBGL	Contig18425.1	Forward	tccatgaccatacatcctcgtt	239	acaagaagtgttccaacctagtcaattgagacattacattacctactatgtacttgtttg
		Reverse	tccaggttttaactgcccttgt		
IbBHLH2	Contig28401.1	Forward	atatgctgcgcaacaaaacctg	215	cccactgtttaagcggacatacatgtctcttattcacaagaatttttgcaagtcttttaa
		Reverse	tacagacaccttcctccatagcc		
IbRSI	RT_091599.1	Forward	acttacggcaacaagcagacct	239	cgagcgactttccatcatttcgatttgcttttccagtaaagttaaggagaaagctaaatg
		Reverse	agaattcaagcatacattgatcg		
IbIB	Contig7506.1	Forward	caaagtcccctacgagtccaaa	167	acgaggagatcgattaagaaagtttgtcttacctcatagtatgtatgcatgcatgttatt
		Reverse	gaagcctgacatggtcaacaac		
IbKIN1	Contig1100.1	Forward	aggtcctttactaggcgtttcc	200	ctgatgaatttggtggcaagaattctctgtaatctcgtcagtagtgttgtaataaacaaa
		Reverse	aattcttgccaccaaattcatc		
IbPMI	Contig15056.1	Forward	gagacgtgggtgagtactgcta	154	accaagtctagctaagaattccttcatcctacgttacacttgtcctcatggtgtatgttc
		Reverse	agttgttaaccaaagcgagagc		

Table 4.3. GO classification by Molecular function (F) of Upregulated genes from the GO enrichment analysis.

GO term	Name	Type	Number of features in test group	Number of unigenes in reference group
GO:0030246	carbohydrate binding	F	22	350
GO:0003677	DNA binding	F	56	2151
GO:0030234	enzyme regulator activity	F	28	482
GO:0016787	hydrolase activity	F	158	6342
GO:0016301	kinase activity	F	74	3546
GO:0008289	lipid binding	F	9	243
GO:0005515	protein binding	F	251	6030
GO:0005102	receptor binding	F	19	6
GO:0003700	transcription factor activity	F	51	1129
GO:0016740	transferase activity	F	154	7356

Table 4.4. GO classification by Cellular component (C) of upregulated genes from the GO enrichment analysis.

GO term	Name	Type	Number of features in test group	Number of unigenes in reference group
GO:0005618	cell wall	C	166	736
GO:0005829	Cytosol	C	103	1210
GO:0005783	endoplasmic reticulum	C	53	917
GO:0005768	Endosome	C	24	141
GO:0005615	extracellular space	C	3	23
GO:0005794	Golgi apparatus	C	38	742
GO:0043231	intracellular membrane-bounded organelle	C	409	14147
GO:0043232	intracellular non-membrane-bounded organelle	C	48	1640
GO:0005739	mitochondrion	C	96	3907
GO:0005730	Nucleolus	C	22	438
GO:0005654	Nucleoplasm	C	18	276
GO:0005634	Nucleus	C	144	4166
GO:0005777	Peroxisome	C	14	356
GO:0005886	plasma membrane	C	229	3103
GO:0009536	Plastid	C	173	5725
GO:0005773	Vacuole	C	112	929

Table 4.5. Comparison of expression of known genes differentially expressed in storage roots vs. fibrous roots. Observed fold change from microarray results.

Accession (Gene)	Annotation	Expected fold change	Observed fold change	Significant & Agreement	Adjusted permuted P
AB283028.1 (Ibkn2)	Class-I knotted1-like h omeobox protein IBKN2	up (Tanaka et al., 2008)	2.53	yes	0.0102
AB283028.1 (Ibkn2)	Class-I knotted1-like homeobox protein IBKN2	up (Tanaka et al., 2008)	1.48	yes	0.0389
AB283028.1 (Ibkn2)	Class-I knotted1-like homeobox protein IBKN2	up (Tanaka et al., 2008)	1.13	yes	0.0268
AB369253.1 (Ibkn4)	Class-I knotted1-like homeobox protein IBKN4	up (Tanaka et al., 2008)	1.34	yes	0.0360
DQ195758.1	sporamin B	up (Desai et al., 2008; McGregor et al, 2006)	4.34	yes	0.0002
DQ195765.1	sporamin A	up (Desai et al., 2008; McGregor et al, 2006)	3.60	yes	0.0195
AB271011.1	ADP-glucose pyrophosphorylase alpha subunit,small suunit (IbAGPa1)	up (Desai et al., 2008; McGregor et al, 2006)	1.44	no	0.1535
CB330696.1	fructokinase-like	up (McGregor et al,2006)	1.00	no	0.3778
Contig1062	similar to protein induced upon tuberization [Solanum demissum]	up (unpublished data, N. Firon, 2011)	1.06	no	0.3503
Contig1288	sucrose synthase	up (McGregor et al., 2006)	1.32	yes	0.0051
Contig2826	Probable glutathione S-transferase	down (Mc Gregor et al., 2006)	-1.49	yes	0.0043
D12882.1	β-Amylase gene	Up	4.21	yes	0.0083
EE875156.1	cysteine proteinase inhibitor(Cystatin)	Up	1.36	yes	0.0140
EE878453.1	starch phosphorylase	Up	2.53	yes	0.0097
EE879297.1	UBQ3 (POLYUBIQUITIN 3); protein binding	down (Desai et al., 2008	-1.31	yes	0.0154

(Table 4.5 cont.)

Accession (Gene)	Annotation	Expected fold change	Observed fold change	Significant & Agreement	Adjusted permuted P
EE879745.1	RPL18 (ribosomal protein L18)	up (Desai et al., 2008)	1.13	no	0.1565
EE880451.1	60S ribosomal protein L26 (RPL26A)	up (Desai et al., 2008)	1.09	no	0.3465
EU784152.1	Escherichia coli hygromycin B phosphotransferase	negative control	1.01	yes	0.4429
Contig35258.1 (IbEF1a)	Elongation factor 1-a (IbEF1a)	stable (present work, in silico)	1.02	yes	0.4701
NM_003140.1	Homo sapiens sex determining region Y	negative	-1.00	yes	0.3652
S_PBL_c337	Osmotin	down (unpublished data, N. Firon, 2011)	1.09	no	0.2505
S_PBL_c3493	ADP/ATP carrier 2-AAC2	down (unpublished data, N. Firon, 2011)	1.08	no	0.2835
Contig34910.1	GADPH	stable (unpublished data, N. Firon, 2011); up (McGregor et al., 2006)	1.24	no	0.0408
S_PBL_c5869	Actin	stable (unpublished data, N. Firon, 2011)	1.31	no	0.0257
Contig14376.1	Alpha tubulin	up (McGregor et al., 2006)	1.81	yes	0.0153
EE880927.1(IbBEL1)	AT5G41410.1 POX (plant homeobox) family protein	down (J.Solis, present work)	-1.79	yes	0.0082
Contig22795.1(IbBEL7)	AT5G02030(BLH9,HB-6) POX (plant homeobox) family protein; BEL1-LIKE HOMEODOMAIN 9	up (J.Solis, present work)	1.25	yes	0.0195
Contig22795.1(IbBEL7)	AT5G02030(BLH9,HB-6) POX (plant homeobox) family protein;BEL1-LIKE HOMEODOMAIN 9	up (J.Solis, present work)	1.26	yes	0.0178

(Table 4.5 cont.)

Accession (Gene)	Annotation	Expected fold change	Observed fold change	Significant & Agreement	Adjusted permuted P
Contig22795.1(IbBEL7)	AT5G02030(BLH9,HB-6) POX (plant homeobox) family protein;BEL1-LIKE HOMEODOMAIN 9	up (J.Solis, present work)	1.75	yes	0.0166
EE881365.1(IbBEL6)	AT2G23760(BLH4) BEL1-like homeodomain 4;BLH4 (SAWTOOTH 2); DNA binding	up (J.Solis, present work)	-1.36	yes	0.0437
Contig39444.1 (IbHLH1)	AT1G72210 (bHLH096) basic helix-loop-helix (bHLH) DNA-binding superfamily protein	up (J.Solis, present work)	1.68	yes	0.0450
Contig39444.1 (IbHLH1)	AT1G72210 (bHLH096) basic helix-loop-helix (bHLH) DNA-binding superfamily protein	up (J.Solis, present work)	1.50	yes	0.0500
Contig39444.1 (IbHLH1)	AT1G72210 (bHLH096) basic helix-loop-helix (bHLH) DNA-binding superfamily protein	up (J.Solis, present work)	2.06	yes	0.0316
S_PBL_c31897 (IbBME)	AT3G54810.1(BME3,BME3-ZF) Plant-specific GATA-type zinc finger transcription factor family protein	up (J.Solis, present work)	1.42	yes	0.0317
Contig23250.2 (IbCBP2)	AT2G33990.1 iqd9 (IQEdomain 9); calmodulin_binding family protein	up (J.Solis, present work)	1.34	yes	0.0365
Contig23250.2 (IbCBP2)	AT2G33990.1 iqd9 (IQEdomain 9); calmodulin_binding family protein	up (J.Solis, present work)	1.42	yes	0.0185
Contig22153.1 (IbCDPK3)	AT4G09570.1 calcium-dependent protein kinase 4	up (J.Solis, present work)	1.16	yes	0.0167
Contig34516.2 (IbCDPK3)	AT4G09570(CPK4) Calcium dependent protein kinase 4	up (J.Solis, present work)	-1.39	no	0.0046

(Table 4.5 cont.)

Accession (Gene)	Annotation	Expected fold change	Observed fold change	Significant & Agreement	Adjusted permuted P
S_PBL_c1655	IbCKR_S_PBL_c1655probe2	up (unpublished data, N. Firon, 2011)	1.24	yes	0.0319
S_PBL_c3693 (IbCRF1)	AT4G11140(CRF1) cytokinin response factor 1	down (unpublished data, N. Firon, 2011)	-1.40	yes	0.0059
S_PBL_c5300 (IbCRF2)	AT4G27950.1(CRF4) cytokinin response factor 4	down (unpublished data, N. Firon, 2011)	-1.43	yes	0.0054
AB478416.1	*Ipomoea batatas* CycD3;1 mRNA for cyclin D3, complete cds	up (Nagata and Saitou, 2009)	1.64	yes	0.0322
Contig3374.1 (IbHB1)	AT4G37790.1 Homeobox-leucine zipper protein family (HAT22)	up (J.Solis, present work)	1.50	yes	0.0482
Contig401.1 (IbHB2)	AT2G46680(ATHB7,HB-7) homeobox 7	up (J.Solis, present work)	1.49	yes	0.0020
Contig401.1 (IbHB2)	AT2G46680(ATHB7,HB-7) homeobox 7	up (J.Solis, present work)	1.41	yes	0.0073
S_PBL_c15425 (IbIAA)	Auxin-responsive Aux/IAA family protein	up (J.Solis, present work)	1.43	yes	0.0091
S_PBL_c11096 (IbMAP65)	AT1G14690.2 microtubule-associated protein 65-7	up (unpublished data, N. Firon, 2011; J.Solis, present work)	1.94	yes	0.0110
S_PBL_c19033 (IbNAM)	Q8LRL4_PETHY Nam-like protein 11 Petunia hybrida NH11	up (unpublished data, N. Firon, 2011; J.Solis, present work)	1.98	yes	0.0438
EF192419.1 (IbPIP1)	tonoplast instrinsic protein (TIP)	down (Kim et al., 2009)	-1.27	yes	0.0280
S_PBL_c3428 (IbXX)	conserved hypothetical protein	up (unpublished data, N. Firon, 2011)	1.69	yes	0.0232

Table 4.6. **List of selected transcripts significantly upregulated in storage roots. P=biological process, F=molecular function, C=cell component, and NA=no data available.**

Unigene	Fold Change	Annotation	Probeid	GO group	Accession in TSA	TSA_ annotation	DGE analysis in SPTSA
S_PBL_c3576	12.61	NoMatchingHitfound	S_PBL_c3576	NA	gb|JP130210.1 Contig_25622	No Matching Hit found	down
S_PBL_c2430	7.61	B9NDK1 B9NDK1_POPTR Predicted protein OS=Populus trichocarpa GN=POPTRDRAFT_789085	S_PBL_c2430	NA	gb|JP121220.1 *Ipomoea batatas* Contig_16632	No Matching Hit found	
Contig33633.1	7.25	Q93X99_IPOBA Proteinase inhibitor	CUST_76135_PI427086615	F			
Contig14347.1	6.55	B9R8R1_RICCO Glucan endo-1,3-beta-glucosidase, putative	CUST_45040_PI427086615	P			
Contig056161.3	6.34	A9ZSX7_9ASTE Alcohol dehydrogenase (Fragment)	CUST_22924_PI427086615	P			
Contig13858.1	6.06	B9RJV7_RICCO Aspartic proteinase nepenthesin-1, putative	CUST_44268_PI427086615	F			
Contig28401.1	5.92	B9SSC2_RICCO DNA binding protein, putative	CUST_67623_PI427086615	P			
Contig34683.1	5.79	OLP1_SOLLC Osmotin-like protein	CUST_77751_PI427086615	P			
Contig16599.1	5.76	O80431_TOBAC Endo-xyloglucan transferase (EXGT)	CUST_48523_PI427086615	P			
S_PBL_c3396	5.65	No Matching Hit found	S_PBL_c3396	NA			
Contig35533.1	5.23	B3IWI0_ORYSJ QLTG-3-1 protein; Low-temperature germinability 3-1	CUST_79137_PI427086615	P			
Contig29781.1	1.56	B9SRA9_RICCO Auxin:hydrogen symporter, putative	CUST_69911_PI427086615	C	Sequence not compared to SPTSA		

(Table 4.6 cont.)

Unigene	Fold Change	Annotation	Probeid	GO group	Accession in TSA	TSA_ annotation	DGE analysis in SPTSA
Contig35525.1	5.18	B9S3J8_RICCO Repetitive proline-rich cell wall protein 2, putative	CUST_79121_PI427086615	P			
Contig045390.3	5.1	U497E_VITVI UPF0497 membrane protein 14	CUST_6323_PI427086615	C			
Contig9244.1	5.03	B7P071_IPOBA Expansin	CUST_37159_PI427086615	P			
Contig23824.1	4.99	AMYB_IPOBA Beta-amylase	CUST_60183_PI427086615	C			
Contig046841.3	4.72	PRP3_SOYBN Repetitive proline-rich cell wall protein 3	CUST_8445_PI427086615	NA			
Contig8425.1	4.71	PER4_VITVI Peroxidase 4	CUST_35950_PI427086615	P			
Contig18155.1	4.39	B9T3K3_RICCO Protein P21, putative	CUST_51006_PI427086615	P			
Contig13241.1	4.37	Q2VT58_CAPAN Proline-rich protein	CUST_43270_PI427086615	P			
Contig35601.1	4.3	Q8W112_ARATH Beta-D-glucan exohydrolase-like protein	CUST_79231_PI427086615	NA	Sequence not compared to SPTSA		
Contig7506.1	4.28	O24329_RICCO Putative uncharacterized protein	CUST_34536_PI427086615	NA			
S_PBL_c20147	4.17	NoMatchingHitfound	S_PBL_c20147	NA			
S_PBL_c19740	4.16	NoMatchingHitfound	S_PBL_c19740	NA	Sequence not compared to SPTSA		
Contig5724.1	4.05	Q8W1A6_PETHY Aquaporin-like protein	CUST_31637_PI427086615	P			
S_PBL_lrc49266	3.95	DQ515800.1 *Ipomoea batatas* expansin mRNA, complete cds	S_PBL_lrc49266	NA	gb|JP135257.1 | *Ipomoea batatas* Contig_30669	NoMatching Hitfound	down

(Table 4.6 cont.)

Unigene	Fold Change	Annotation	Probeid	GO group	Accession in TSA	TSA_annotation	DGE analysis in SPTSA
S_PBL_c40350	3.76	No Matching Hit found	S_PBL_c40350	NA	gb|JP142955.1 *Ipomoea batatas* Contig_38367	NoMatching Hit found	down
Contig9836.1	3.71	A8QXP6_IPOBA Class-I knotted1-like homeobox protein IBKN2	CUST_38048_PI427086615	F	Sequence not compared to SPTSA		
RT_091599.1	3.7	A2Q374_MEDTR Gibberellin regulated protein	CUST_82993_PI427086615	F	gb|JP120229.1 *Ipomoea batatas* Contig_15641		
Contig15133.1	3.67	B3F8H6_NICLS Sucrose sythase	CUST_46239_PI427086615	P			
contig16960	3.66	Q42936_NICPL Pectinesterase (Fragment)	contig16960	C			
Contig2866.1	3.51	B3SHI0_IPOBA Anionic peroxidase swpa7	CUST_27186_PI427086615	P			
S_PBL_lrc26197	3.5	XP_003611068.1 hypothetical protein MTR_5g010060 [Medicago truncatula]	S_PBL_lrc26197	NA	gb|JP125047 *Ipomoea batatas* Contig_20459	NoMatching Hitfound	up
Contig14099.1	3.49	Q9FHM9_ARATH AT5g51550/K17N15_10	CUST_44677_PI427086615	NA	Sequence not compared to SPTSA		
Contig18258.1	3.42	Q6RJY7_CAPAN Elicitor-inducible protein EIG-J7	CUST_51170_PI427086615	F			
S_PBL_c46290	3.04	NoMatchingHitfound	S_PBL_c46290	NA	no similar sequence in TSA		
Contig18425.1	3.07	Q0H8W0_TOBAC Endo-beta-1,4-glucanase	CUST_51397_PI427086615	P	Sequence not compared to SPTSA		

(Table 4.6 cont.)

Unigene	Fold Change	Annotation	Probeid	GO group	Accession in TSA	TSA_ annotation	DGE analysis in SPTSA
Contig23250.1	3.22	Q2HPE5_CAPAN CAPIP2, involved in disease resistance and drought and salt stress tolerance.	CUST_59160_PI427086615	NA			
Contig8256.1	3.06	Q9SML5_CAPAN Knolle	CUST_35693_PI427086615	P	Sequence not compared to SPTSA		
Contig11521.1	3.24	XTH1_SOLLC Probable xyloglucan endotransglucosylase/hydrolase 1	CUST_40601_PI427086615	P			
Contig1100.1	3.03	B6TKU2_MAIZE Protein kinase	CUST_24543_PI427086615	NA	gb|JP131535.1|TSA: *Ipomoea batatas* Contig_26947	gi|118488581|gb|ABK96103.1| unknown [Populus trichocarpa]	up (in HTR)
Contig18487.1	2.97	VATG2_TOBAC V-type proton ATPase subunit G 2	CUST_51527_PI427086615	C	Sequence not compared to SPTSA		
Contig9874.1	2.97	BH096_ARATH Transcription factor bHLH96	CUST_38111_PI427086615	F			
Contig32758.1	2.89	Q07446_SOLLC Peroxidase	CUST_74751_PI427086615	P			
RT_327455.1	2.86	C7C5S8_NICBE PME inhibitor	CUST_88753_PI427086615	C			
Contig18822.1	2.77	Q6AVU2_ORYSJ Endoplasmic oxidoreductin 1, putative, expressed	CUST_51974_PI427086615	P	Sequence not compared to SPTSA		
RT_344595.1	2.52	B9MTB7_POPTR Aquaporin, MIP family, PIP subfamily	CUST_89245_PI427086615	P	Sequence not compared to SPTSA		
Contig4451.1	2.5	Q6SA75_TOBAC ANT-like protein	CUST_29641_PI427086615	F			

(Table 4.6 cont.)

Unigene	Fold Change	Annotation	Probeid	GO group	Accession in TSA	TSA_ annotation	DGE analysis in SPTSA
Contig25858.1	2.3	D2DGW4_SOYBN Aux/IAA protein	CUST_63543_PI427086615	P			
Contig5206.1	2.16	B9R873_RICCO Endo-1,4-beta-glucanase, putative	CUST_30801_PI427086615	P			
S_PBL_c22205	2.09	EEF31071.1|aspartate aminotransferase, putative [Ricinus communis]	S_PBL_c22205	C			
Contig705.1	2.05	Q8RVF8_TOBAC Thioredoxin peroxidase	CUST_23950_PI427086615	P	Sequence not compared to SPTSA		
Contig13201.1	2.01	Q2LAI9_SOLLC Auxin response factor 4	CUST_43212_PI427086615	P	Sequence not compared to SPTSA		
S_PBL_c22019	1.95	No Matching Hit found	S_PBL_c22019	NA	gb|JP122384.1| TSA: *Ipomoea batatas* Contig_17796	gi|222834016| gb|EEE72493.1| predicted protein [Populus trichocarpa]	Up
contig27450	1.85	Q564D7_SOLLC Pectinesterase	contig27450	C	Sequence not compared to SPTSA		
Contig23828.1	1.81	B9SYJ3_RICCO Acetylglucosaminyltransferase, putative	CUST_60190_PI427086615	NA	Sequence not compared to SPTSA		
S_PBL_c40771	1.79	emb|AJ429702.1| *Ipomoea batatas* chloroplast atpE gene (partial)	S_PBL_c40771	C	gb|JP112616.1| TSA: *Ipomoea batatas* Contig_8028		NA
Contig19145.1	1.78	B9RKN2_RICCO Sugar transporter, putative	CUST_52513_PI427086615	C	Sequence not compared to SPTSA		

(Table 4.6 cont.)

Unigene	Fold Change	Annotation	Probeid	GO group	Accession in TSA	TSA_ annotation	DGE analysis in SPTSA
Contig18798.1	1.74	Q18PQ3_IPOBA Starch branching enzyme I B9SWB6_RICCO	CUST_51947_PI427086615	C	Sequence not compared to SPTSA		
RT_351811.1	1.74	Structural constituent of cell wall, putative	CUST_89437_PI427086615	NA	Sequence not compared to SPTSA		
S_PBL_c11962	1.73	NoMatchingHitfound	S_PBL_c11962	NA	gb|JP130843.1| TSA: *Ipomoea batatas* Contig_26255	No Matching Hit found	Up
EE876438.1	1.7	B9RYY6_RICCO Auxin-induced protein 5NG4, putative	CUST_93273_PI427086615	C	Sequence not compared to SPTSA		
S_PBL_c3428	1.69	Putative uncharacterized protein Ricinus communis GN=RCOM_1587270	IbXX_S_PBL_c3428probe3	F	gb|JP120261.1| TSA: *Ipomoea batatas* Contig_15673	No Matching Hit found	up
RT_357758.1	1.68	Q40473_TOBAC PS60 protein	CUST_89609_PI427086615	C	Sequence not compared to SPTSA		
Contig15375.1	1.64	LAX4_MEDTR Auxin transporter-like protein 4	CUST_46600_PI427086615	P	Sequence not compared to SPTSA		
S_PBL_c26605	1.62	NoMatchingHitfound	S_PBL_c26605	NA	gb|JP121220.1| TSA: *Ipomoea batatas* Contig_16632	No Matching Hit found	up
Contig16851.1	1.6	A3F771_IPONI Auxin response factor 8	CUST_48939_PI427086615	P	Sequence not compared to SPTSA		
Contig29781.1	1.56	B9SRA9_RICCO Auxin:hydrogen symporter, putative	CUST_69911_PI427086615	C	Sequence not compared to SPTSA		
Contig23980.1	1.54	B9SZH5_RICCO Auxin-induced protein	CUST_60408_PI427086615	P	Sequence not compared to SPTSA		
Contig33353.1	1.53	Q66GR8_ARATH At1g03470	CUST_75648_PI427086615	NA	Sequence not compared to SPTSA		

Table 4.7. List of transcripts significantly downregulated in storage roots.

Probeid	Fold Change	Annotation	Unigene	GO term	GO category	PBL assembly
CUST_11464_PI427086615	-2.04	D7M111_ARALY Invertase/pectin methylesterase inhibitor family protein	Contig048963.3	GO:0005618	cell wall	S_PBL_c27094
CUST_16127_PI427086615	-1.91	C9WF05_GOSHI Class III peroxidase	Contig052093.3	GO:0006950	response to stress	GMNYDFC01AOXA9
CUST_25807_PI427086615	-2.21	O23190_ARATH MAP3K-like protein kinase	Contig1960.1	GO:0007165	signal transduction	S_PBL_c6661
CUST_31615_PI427086615	-2.08	B9S775_RICCO Peroxidase 10, putative	Contig5710.1	GO:0006950	response to stress	NoHit
CUST_37581_PI427086615	-2.37	A1E0X8_SOLTU NAC domain protein NAC2	Contig9539.1	GO:0030528	transcription regulator activity	GMNYDFC02ENP0N
CUST_40954_PI427086615	-1.55	B8LFG9_IPOBA Tonoplast intrinsic protein	Contig11751.1	GO:0005215	transporter activity	S_PBL_c11202
CUST_44967_PI427086615	-2.25	B3SHI0_IPOBA Anionic peroxidase swpa7	Contig14301.1	GO:0006950	response to stress	NoHit
CUST_54963_PI427086615	-2.11	D1M7W9_SOLLC Heat stress transcription factor A3	Contig20679.1	GO:0030528	transcription regulator activity	S_PBL_c51261
CUST_13187_PI427086615	-1.69	Q40090_IPOBA SPF1 protein	Contig050116.3	GO:0030528	transcription regulator activity	S_PBL_c12020
contig10788	-1.94	B9RC22_RICCO Multidrug resistance pump, putative	Contig13553.1	GO:0030528	transcription regulator activity	S_PBL_c1253

(Table 4.7 cont.)

Probeid	Fold Change	Annotation	Unigene	GO term	GO category	PBL assembly
contig24525	-1.58	B9SAP4_RICCO Multidrug resistance-associated protein 2, 6 (Mrp2, 6), abc-transoprter, putative	Contig20352.1	GO:0005215	transporter activity	NoHit
contig36064	-1.80	PDR1_TOBAC Pleiotropic drug resistance protein 1	RT_381219.1	GO:0005215	transporter activity	S_PBL_c20743
CUST_13887_PI427086615	-1.69	B9I191_POPTR Multidrug resistance protein ABC transporter family	Contig050604.3	GO:0005215	transporter activity	GMNYDFC02ENJE5
CUST_14631_PI427086615	-1.82	B9SX20_RICCO NAC domain-containing protein, putative	Contig051116.3	GO:0005215	transporter activity	S_PBL_c26038
CUST_36993_PI427086615	-1.85	Q76CU1_TOBAC PDR-type ABC transporter 2 (Fragment)	Contig9131.1	GO:0005215	transporter activity	S_PBL_c1744
CUST_37299_PI427086615	-2.98	B9SCG8_RICCO Nitrate transporter, putative	Contig9348.1	GO:0005215	transporter activity	NoHit
CUST_41967_PI427086615	-1.80	PDR1_TOBAC Pleiotropic drug resistance protein 1	Contig12402.1	GO:0005215	transporter activity	S_PBL_c20743
CUST_43782_PI427086615	-1.97	B9RC22_RICCO Multidrug resistance pump, putative	Contig13553.1	GO:0005215	transporter activity	S_PBL_c1253
CUST_4579_PI427086615	-1.69	Q8GU64_ORYSJ MRP-like ABC transporter	Contig044185.3	GO:0005215	transporter activity	S_PBL_c24235

(Table 4.7 cont.)

Probeid	Fold Change	Annotation	Unigene	GO term	GO category	PBL assembly
CUST_46118_PI427086615	-2.26	D7M111_ARALY Invertase/pectin methylesterase inhibitor family protein	Contig15056.1	GO:0005618	transporter activity	S_PBL_c27094
CUST_48781_PI427086615	-2.88	Q84MZ8_TOBAC High affinity nitrate transporter protein	Contig16753.1	GO:0005215	transporter activity	GMNYDFC 01BGDJQ
CUST_49258_PI427086615	-1.53	B9N4E6_POPTR ABC transporter family protein (Fragment)	Contig17052.1	GO:0005215	transporter activity	S_PBL_c36025
CUST_49343_PI427086615	-1.53	B9RQF2_RICCO ATP-binding cassette transporter, putative	Contig17112.1	GO:0005215	transporter activity	S_PBL_lrc30858
CUST_54420_PI427086615	-1.72	B9I9S5_POPTR Multidrug resistance protein ABC transporter family	Contig20352.1	GO:0005215	transporter activity	NoHit
CUST_23161_PI427086615	-1.52	Q0WSR2_ARATH Putative peroxidase	Contig144.1	GO:0006950	response to stress	NoHit
CUST_42875_PI427086615	-1.93	B9T8I2_RICCO Peroxidase N, putative	Contig12997.1	GO:0006950	response to stress	S_PBL_c29652
CUST_11378_PI427086615	-1.59	B6V6Z9_POPEU GAST-like protein	Contig048896.3	GO:0006950	response to stress	S_PBL_c15608
CUST_25604_PI427086615	-1.60	B9R887_RICCO Gibberellin-regulated protein 1, putative	Contig1829.1	GO:0009725 GO:0005554	response to hormone stimulus molecular function unknown	S_PBL_c15463
CUST_39798_PI427086615	-1.68	B9HPF8_POPTR GRAS family transcription factor	Contig11000.1	GO:0030528	transcription regulator activity	S_PBL_c3199

(Table 4.7 cont.)

Probeid	Fold Change	Annotation	Unigene	GO term	GO category	PBL assembly
CUST_42099_PI427086615	-1.62	B9HZY3_POPTR NAC domain protein, IPR003441	Contig12488.1	GO:0030528	transcription regulator activity	S_PBL_c16979
CUST_4519_PI427086615	-1.92	D5LHU3_IPONI Gibberellin 2-oxidase 2	Contig044148.3	GO:0009685	gibberellin metabolic process	GMNYDFC 02DQFTG
CUST_14117_PI427086615	-1.58	C7ENF8_IPOBA WRKY1	Contig050769.3	GO:0030528	transcription regulator activity	S_PBL_c10601
CUST_26266_PI427086615	-1.53	Q9FXS1_TOBAC WRKY transcription factor NtEIG-D48	Contig2257.1	GO:0030528	transcription regulator activity	S_PBL_c8597
CUST_26353_PI427086615	-1.74	D3GDP7_SOLLC JA-induced WRKY protein	Contig2317.1	GO:0030528	transcription regulator activity	GMNYDFC 01B9OG4
CUST_26489_PI427086615	-1.94	C7E5X8_CAPAN Transcription factor WRKY	Contig2412.1	GO:0030528	transcription regulator activity	NoHit
CUST_36533_PI427086615	-1.86	B6U788_MAIZE WRKY39v2-superfamily of TFs having WRKY and zinc finger domains	Contig8824.1	GO:0030528	transcription regulator activity	GM0Z85L06 GZ71X
CUST_3846_PI427086615	-2.17	C7E5X8_CAPAN Transcription factor WRKY	Contig043659.3	GO:0030528	transcription regulator activity	S_PBL_c12895
CUST_47820_PI427086615	-1.70	B9S164_RICCO WRKY transcription factor, putative	Contig16173.1	GO:0030528	transcription regulator activity	GMNYDFC02 ES2AX
CUST_30987_PI427086615	-1.90	Q5GA67_SOLLC BHLH transcriptional regulator	Contig5325.1	GO:0030528	transcription regulator activity	S_PBL_c29834

(Table 4.7 cont.)

Probeid	Fold Change	Annotation	Unigene	GO term	GO category	PBL assembly
CUST_34219_PI427086615	-1.51	D7P233_TOBAC MYC1a transcription factor	Contig7297.1	GO:0030528	transcription regulator activity	NoHit
CUST_38571_PI427086615	-1.86	C3W4Q3_VITVI R2R3 transcription factor MYB108-like protein 1	Contig10181.1	GO:0030528	transcription regulator activity	S_PBL_c7416
CUST_4141_PI427086615	-1.81	D7P234_TOBAC MYC1b transcription factor	Contig043877.3	GO:0030528	transcription regulator activity	S_PBL_c3888
CUST_4142_PI427086615	-1.83	D7P234_TOBAC MYC1b transcription factor	Contig043877.3	GO:0030528	transcription regulator activity	GMNYDFC02D75T1
CUST_48150_PI427086615	-1.59	D7P233_TOBAC MYC1a transcription factor	Contig16379.1	GO:0030528	transcription regulator activity	S_PBL_c7035
CUST_54086_PI427086615	-1.54	B9SYQ1_RICCO R2r3-myb transcription factor, putative	Contig20112.1	GO:0030528	transcription regulator activity	S_PBL_lrc35891
CUST_24041_PI427086615	-1.93	ERF2_TOBAC Ethylene-responsive transcription factor 2	Contig769.1	GO:0030528	transcription regulator activity	S_PBL_c5522
CUST_26785_PI427086615	-2.03	Q53JG2_ORYSJ AP2 domain containing protein, expressed	Contig2597.1	GO:0030528	transcription regulator activity	GMNYDFC01BT8YT
CUST_26786_PI427086615	-2.11	Q53JG2_ORYSJ AP2 domain containing protein, expressed	Contig2597.1	GO:0030528	transcription regulator activity	GMNYDFC01BT8YT
CUST_28007_PI427086615	-1.71	A3F770_IPONI APETALA2-like protein	Contig3392.1	GO:0030528	transcription regulator activity	S_PBL_c10820

(Table 4.7 cont.)

Probeid	Fold Change	Annotation	Unigene	GO term	GO category	PBL assembly
CUST_32801_PI427086615	-1.73	A9PL97_POPTR AP2/ERF domain-containing transcription factor	Contig6424.1	GO:0030528	transcription regulator activity	GMNYDFC02DN6NH
CUST_33975_PI427086615	-1.89	B9RG00_RICCO AP2 domain transcription factor RAP2.3, putative	Contig7149.1	GO:0030528	transcription regulator activity	S_PBL_c124
CUST_35057_PI427086615	-2.11	A9PL97_POPTR AP2/ERF domain-containing transcription factor	Contig7824.1	GO:0030528	transcription regulator activity	GM0Z85L06HCXP7
CUST_24632_PI427086615	-1.51	B9RJK7_RICCO Protein phosphatase 2c, putative	Contig1169.1	GO:0007165	signal transduction	

Table 4.8. Validation of results from microarray (SPOArrayv1 array) by qRT-PCR.

Unigene	Annotation	Gene Symbol	Fold-change (qRT-PCR)	Fold-change (Array)
Contig4451.1	Q3L8J0_VITVI CBF-like transcription factor Vitis vinifera/Q6SA75_TOBAC ANT-like protein	IbCBF	6.19	2.5
Contig5724.1	Q8W1A6_PETHY Aquaporin-like protein	IbPIP2	12.44	4.05
Contig9244.1	B7P071_IPOBA Expansin	IbEXP2	24.59	5.03
Contig13201.1	Q2LAI9_SOLLC Auxin response factor 4	IbARF	3.64	2.01
Contig13241.1	Q2VT58_CAPAN Proline-rich protein	IbPRP	46.42	4.37
Contig16599.1	O80431_TOBAC Endo-xyloglucan transferase (XTH/EXGT)	IbXTH1	23.48	5.76
Contig18425.1	Q0H8W0_TOBAC Endo-beta-1,4-glucanase	IbBGL	5.35	3.07
Contig28401.1	transcription factor bHLH93-like/B9SSC2_RICCO DNA binding protein	IbBHLH2	43.71	5.92
RT_091599.1	A2Q374_MEDTR Gibberellin regulated protein	IbRSI	24.88	3.7
Contig7506.1	Q9LYE7_ARATH Putative uncharacterized protein At5g11420 Arabidopsis thaliana (DUF642 family, associated with cell wall modifications)/O24329_RICCO Putative uncharacterized protein/ B6TKU2_MAIZE Protein	IbIB	24.03	4.28
Contig1100.1	kinase D7M111_ARALY	IbKIN1	8.42	3.03
Contig15056.1	Invertase/pectin methylesterase inhibitor family protein	IbPMI	0.14	-2.26

4.4 Discussion

In order to discover putative genes associated with the formation of storage roots, a custom array was developed for sweetpotato by selecting a relevant set of candidate genes having a role in sweetpotato root development. The selected transcripts (genes) encompass a

wide range of molecular signaling pathways in roots. Gene Ontology and KEEG (Kyoto Encyclopedia of Genes and Genomes) annotation was used to select unigenes for probe design and included this in the customized sweetpotato Agilent microarray. Gene selection for probe design was based on the ontology Molecular Function, Biological Process, and Cellular Component with their description terms being "Transcription Factor, Carbohydrate Metabolism, Kinase, Molecular Signaling, Membrane and Cell wall". Most of the selected sequences for probe design were from CAP3-based assembly (12,397 unique transcripts or unigenes of "cap3p80v3 assembly") unless a longer homologous sequence was present in a Newbler-based assembly, in which case the longer unigene (502 unique transcripts or unigenes from "newblerv1 assembly") was chosen. In addition, a set of probes were designed from 837 unannotated unigenes from the PBL assembly; each of these unigenes were found to be differentially expressed and abundant in the initiating storage roots compared to lignified adventitious roots by *in-silico* digital transcriptome profiling (N. Firon, communication personal, 2011).

Microarray data was analyzed using an ANOVA implemented in a MAANOVA package and differential expression determined using an FDR test. FDR is more suitable for microarray data analysis and allows multiple hypothesis testing; testing the differential expression of each gene in an array involves one hypothesis and FDR. The ANOVA model from MAANOVA was applied to transformed intensity data (i.e., logarithmic transformation of raw intensity data). It allows one to account for sources of variation in the data that are attributable to factors other than differential expression of genes, thus it effectively normalizes the data. There is no loss of information, as is the case when raw intensity data are converted to ratios. Furthermore, it allows one to combine the information from many different arrays into a single analysis. Moreover, MAANOVA allows missing data if non-background subtracted data is used as input in the analysis. The 0.05 FDR indicated that 4,948 (35.6%) of the features were significantly different

between tissue types, while the 0.01 FDR suggested 1,569 features (11.3%) were different for the 13,912 probes tested. Considering FDR<0.01; only 1% of the features were significant for these cutoffs. Thus, for the 1,569 features that were significant with a FDR of 0.01, about 16 of these features are likely false positives and the other 1,553 are significantly different. Fold change ratio of each transcript was defined in a linear scale as explained in the methods section; this is not a traditional logarithmic scale used in describing the results of a microarray to facilitate the identification of up or downregulated genes as explained in the material and methods section.

Although the probe design with eARRAY and Picky yielded a high number of potential functional probes for each selected unigene placed in the custom microarray, our results from the hybridization experiments showed that many of them did not have a detectable signal (data not shown). Both technical and biological reasons such as location of probe, errors in assembly, alternative splicing and nucleotide variations in the genes of sweetpotato help to explain these negative results. The position of the probe relative to the target may cause variation in signal intensity among different oligonucleotide probes. About 53 transcripts (Table 4.5) included in this customized microarray had multiple probes and their estimated fold change ratio appears to be influenced by the position of the probe in the transcript. Some examples are shown in table 4.5. The following probeids: IbTAP_Contig2072.2probe1, IbTAP_Contig2072.2probe2, IbTAP_Contig2072.2probe3, were wrongly named and do not correspond to the IbTAP gene previously reported to be upregulated in storage roots. These three probes were designed based on Contig2072.2 from a CAP3 assembly, which includes the GenBank accession EE880072.1. A probe for IbTAP gene (Contig27416.1, which contains accessions DV037856.1 and DV038130.1) was not included in the custom array. Background subtraction for the analysis of numeric data coming from the SPOArrayv1 array was not done and represents an advantage of the MAANOVA package.

Assessing the quality of the array could be done using the internal controls provided by Agilent but also using custom probes derived from known genes. The consistency of the signals was evaluated by placing multiple probes for known genes (4.5). Reference controls included probes for alpha tubulin, actin, elongation factor-1 alpha (IbEF1a), GADPH and UBCE because digital expression analysis showed their stable expression in both storage and non-storage roots (data not shown). The microarray data showed that only IbEF1a transcripts had consistent stable expression in both storage and fibrous root tissues.

Novel transcription factors and previously known genes for sweetpotato storage root formation: This study revealed that the most of differentially expressed transcripts (DETs) at FDR<0.05, were upregulated (797 features corresponding to 773 unique transcripts) in the storage roots with a fold change ratio of at least 1.49, while a smaller set of genes showed downregulation (342 features corresponding to 328 unique transcripts) with a fold change ratio ranging from of -1.50 to -2.98. Table 4.5 includes the list of 53 genes that were identified in this study as associated with storage root formation. These set of genes were also used as indicators of the quality of the microarray when used in the study of expression of genes in storage roots vs. non storage roots. For example, transcripts corresponding to IbCRF1 and IbBEL1 were found downregulated in the microarray and this result is supported by DGE analysis derived from the analysis of sweetpotato root transcriptome (N. Firon, personal communication, 2011, unpublished data). Transcripts for starch phosphorylase, fructokinase, AGPase1 and SuSys, all involving the metabolism of starch in plant cells, were upregulated (Table 4.5). However, increased expression of these genes does not necessarily correlate with increased activity of their products. For example, the activity of a plastidial starch phosphorylase, Pho1, is reported to be regulated at post-transcriptional level by the 20S proteasome (Lin et al., 2012). Similarly, in the study of Tao et al. (2012), the top five upregulated transcripts in storage root tissues were

sporamin A, sporamin B, expansin 2, sucrose-phosphate synthase, and an aspartyl protease. Among the most interesting genes which were significantly upregulated in storage root are transcription factors, genes for molecular signaling and response to stimulus, and genes encoding proteins and enzymes for carbohydrate metabolism and cell wall-related activities (Table 4.5 and Table 4.6. The roles for the product of these transcripts are basic cellular processes such as primary and secondary metabolism. Genes involved in starch metabolism are many and have the highest differential expression between storage and fibrous roots, and in regulating developmental and differentiation processes. One component, SPF1, specific for sweetpotato is downregulated consistently for a number of similar probes (derived from highly similar transcripts) in storage roots. There are three transcripts of SPF1encoding a DNA-binding protein that recognizes the promoter of sporamin and β-Amylase genes which were found downregulated in storage roots (Table 4.7). SPF1 is a repressor that binds to SP8 motif of the downstream genes and regulates storage root development; the present custom array was able to recognize expression of SPF1 and correlate it to related genes, demonstrating the robustness of the oligo array-based technology for sweetpotato.

Comparison of the present global gene expression profile study with those of Desai (2008), McGregor (2006), and Tao et al. (2012) reflect differences in approach. For example, Desai (2008) and McGregor (2006) studied expression profiles using the same PICME microarray based on EST sequences, but they included samples from multiple (Desai et at., 2008) and single time points (McGregor, 2006); the present results could not be comprehensively compared to these earlier works due to high redundancy of ESTs in the PICME microarray and more importantly difference in the age of the storage roots used in the latter studies. Microarray results were compared to a newly released Transcriptome Shotgun Assembly (TSA) database at GenBank for sweetpotato (SPTSA assembly; Tao et al., 2012). A comparison against this dataset

was done for a select few transcripts identified by our work. For example, Contig_1752 (JP106340.1) from the SPTSA database has a 97.5% identity to IbBEL7 in an overlapping region of 751 nucleotides; two studies based on DGE profiling showed contrasting results for IbBEL7-like transcripts, one study showed upregulation in storage roots (N.Firon, communication personal, 2011) whereas the other study showed its downregulation (Tao et al., 2012). It is likely becasue different stages of roots were used for gene expression profiling without a biological replicate; and fibrous (FR) and storage roots (SR) were collected at multiple time points in the study of Tao et al. (2012). They classified the SR in three stages, termed initial (ITR), expanding (ETR) and harvesting tuberous roots (HTR) collected at 1.5, 3 and 5 months after planting, while FR were from 1-month old plants. However, the present results from the SPOArrayv1 array correlated partially with the DGE results of Tao et al. (2012), when using the values comparing ITR vs. FR or alternatively comparing ETR vs. ITR. Indeed, the majority of genes upregulated in storage roots listed in Table 4.6, were also found upregulated in the study of Tao et al (2012); in addition, two unannotated transcripts of SPOArrayv1 array (S_PBL_c16992 and S_PBL_c22019) showed similar expression of their homologous transcripts at SPTSA in ETR or ITR vs. FR. Despite this correlation between the microarray and the DGE results, the study of Tao et al. (2012) lacks a key feature because they did not compare the expression of transcripts from SR and FR at the same time point. The present study is unique because the expression of genes was compared between young storage root vs. fibrous roots sampled at same time point (a time at which the initiation of storage root is peaking; Villordon et al., 2009). Similarly, root transcriptome sequencing and the corresponding DGE analysis used in this work were done in samples of initiating storage root and lignified storage root (data not shown). The results of this approach are strongly supported by the significantly high positive correlation of the microarray and this DGE analysis (Table 4.5). Although, currently RNA-seq technology (Wang et al., 2009)

has an advantage over microarrays, the concordance from our work with the digital gene expression profile of sweetpotato, validates the current developed oligo array.

Among the set of transcription factors that are novel for storage root development, homeobox genes were found with a significant enhanced expression. Although multiple probes were designed and used in the SPOArrayv1 for these types of genes and others listed in Table 4.5, their relative expression based on their replicate probes showed a relatively consistent fold change ratio. For example gene Contig39444.1 (encoding a basic helix-loop-helix, bHLH family protein) had 3 different probes on the array, each is tagged as a different feature, and fold change ratios were significantly upregulated in storage roots (1.68, 1.50 and 2.06); similarly, the three probes from KNOX gene, Ibkn2 (accession AB28302) had fold change ratio of 1.48, 1.13 and 1.34 which were significant at FDR <0.05. Ibkn2 is a well-known example of genes previously known upregulated in storage root. Another Contig401.1 (IbHB2) showed significant upregulation in storage roots with a fold change ratio of 1.41at FDR<0.01 and Contig22795.1 (BEL1-LIKE) had three probes with fold change ratios of 1.25, 1.26, and 1.75, all of which are significant at FDR<0.05. The above mentioned transcript, Contig28401.1, was named as the IbHLH2 gene. This is a second bHLH-type transcription factor found in this work, while the IbHLH1 gene was derived from transcript Contig39444.1, which was found upregulated in storage roots (data not shown). Both represent newly reported genes for sweetpotato; we found that a homologous sequence of IbHLH2 in SPTSA assembly encodes a gene for anthocyanin synthesis and is also upregulated in storage roots (Tao et al., 2012).

Genes for cell wall rearrangements: In rapidly expanding tissues like storage roots the rigidity of the cell wall must be relaxed in order to allow continuous growth and expansion. Known genes that target the components of cell wall (i.e. cellulose, hemicellulose, and pectins) during developmental processes and in response to abiotic stimuli are expansins (EXP),

xyloglucan endotransglycosylase /hydrolases (XTHs) and pectin methylesterases (PMEs). Although, we found seven transcripts of expansin notably upregulated in storage roots (Table 4.6), not all expansins appear to be involved in the thickening processes of storage roots. Indeed, of three reported expansin genes (IbEXP1, IbEXP2, and IbEXP3) isolated initially from young storage roots (You et al., 2003). IbEXP1was found to have a functional role in blocking the initial thickening growth of the storage root by suppressing metaxylem and cambium cell proliferation in fibrous roots (Noh et al., 2012). We validated this observed unique expression of an expansin gene, Contig9244.1 (IbEXP2-like) and found it to be 4 to ~24 times upregulated in storage roots compared to fibrous roots (Table 4.8). Expansin are encoded by multifamily genes consisting of four gene subfamilies: α-expansin, β-expansin, expansin-like A, and expansin-like B (Sampedro and Cosgrove, 2005); the notably abundance of expansin transcripts in this work suggests that they are encoded by a multifamily of genes. XTHs are also occurring in families of glycosyl hydrolases in plants and 13 and 20 unique transcripts of XTH and pectinesterases were found in this study to be significantly upregulated in storage root (Table 4.6 and Appendix F). The interaction of PME with proteins that inhibit its activity (pectin methyesterase inhibitors, PMIs), contribute to the modulation of the degree of the methylesterified state of pectin in the cell wall during different developmental processes (Pelloux et al., 2007). We had validated the observed downregulation of a PMI gene (Contig15056.1, IbPMI) and upregulation of a transcript of a endo-xyloglucan transferase (IbXTH1, Contig16599.1) by qRT-PCR (Table 4.8) and they represent new for sweetpotato.

Among the genes that were found highly upregulated in storage roots, are four transcripts encoding proline-rich proteins (PRP) with fold changes ratio values >4; two of these PRP transcripts, Contig35525.1 and Contig13241.1 merit special attention. The product of these transcripts is related to a cell wall linker protein (CWLP) which is conserved across *Arabidopsis,*

Populus and *Ricinus*. Tolerance to drought due to overexpression of CWLP has been reported in potato (Honig et al. 2007, http://abstracts.aspb.org/pb2007/public/P23/P23006.html). A detailed analysis of these proline-rich proteins-coding sequences from sweetpotato revealed that the assembled consensus transcript mostly includes singlets derived from NGS (data not shown). For example, Contig35525.1 consists of 502 sequences all of which are derived from root libraries. Four hundred and ninety eight out of these 502 are from the root transcriptome and the remaining four sequences were deposited in GenBank (DV035554.1, CO501130.1, DV038158.1, EE879878.1). Similarly, for Contig13241.1 235 out of 253 sequences are from root transcriptome and fourteen from GenBank (CB330729.1, BU690280.1, EE880043.1, EE877558.1, BU690861.1, BU690688.1, BU691825.1, BU690149.1, BU691866.1, BU691764.1, CB330152.1, EE881604.1, DV036357.1 and CB330348.1). Only six out of 253 sequences in Contig13241.1 came from leaf transcriptome.

Genes in cellular and hormone signaling: Another set of genes, associated with the phytohormone signaling pathway, and recognized to play a role during differentiation processes were found differentially expressed in storage and fibrous roots. Specifically, both qRT-PCR and microarray results validated the enhanced expression of transcript RT_091599.1, referred to as IbRSI gene (Table 4.8). IbRSI gene matches to two sequences, one from potato (RSI1_SOLLC protein RSI-1), encoding a gene expressed very early in lateral root development and induced by auxin, and the other from *Ricinus communis* (B9R887_RICCO gibberellin-regulated protein 1). Among other top transcripts with the highest downregulation in sweetpotato storage roots, transcripts for 13 nitrate transporters were observed. Downregulation of transcripts of nitrate transporters in plants is given by the action of microRNAs and mechanisms associated with auxin signaling. A single transcript encoding a nitrate transporter was also found downregulated in storage and pencil roots by Desai et al. (2008). We do not know the mechanism of action of

these genes, but they might be related to the development of lateral roots. Relevant for sweetpotato is the observation that lateral root development is different in initiating storage roots vs. non-storage roots (Villordon et al., 2012) in sweetpotato. Phytohormones are now recognized as part of the regulatory and signaling mechanisms which modulate many developmental processes including the induction of lateral roots; phytohormones include: auxin (indole-3-acetic acid,IAA), ethylene, cytokinin (trans-zeatin riboside,t-ZR), jasmonic acid (JA), abscisic acid (ABA) and gibberellins (GA). Both auxin and low levels of ethylene are documented to be involved in the stimulation of lateral root growth (Ivanchenko et al., 2008; Ruzicka et al., 2007). Recent work clarifies how these hormones are inter-acting during lateral root formation. The process of lateral root formation in plants begin when quiescent pericycle cells which are activated to undergo a precise series of divisions that leads to formation of a new lateral root (Fukaki and Tasaka, 2009; Lewis et al., 2011). The exact mechanism of action of auxin and ethylene reveals that a maximal local accumulation of the auxin IAA at the location of these quiescent pericycle cells is required to promote the development of lateral roots, while presence of ethylene precursor 1-aminocyclopropane-1-carboxylic acid (ACC) suppress this local accumulation of auxin (Lewis et al., 2011). The transport of auxin from leaf to root tissues is accomplished by transporters in the cell, with auxin influx mediated by proteins encoded by auxin resistant 1 (AUX1) and Like AUX(LAX1,2 and 3) genes, and auxin efflux mediated by PIN-formed (PIN) and P-multidrug resistance/P-glycoprotein (MDR/PGP) (MDR/PGP) genes. The negative effects of ACC was documented to correlate with enhanced expression of PIN3 and PIN7 genes in root tissues (Lewis et at., 2011); a hypothesis suggests these PIN transporters take auxins out of cells in basipetal transport (to the root tip) thus causing a lack of a localized accumulation of auxin needed to drive lateral root formation. We found seventeen different transcripts of a resistance/P-glycoprotein (MDR/PGP) family of ABC transporters MDR/PGP as

well as and six transcripts encoding ethylene response factors (AP2/ERF) domain-containing transcription factors upregulated in fibrous roots; in addition, seven unique transcripts encoding Auxin response factors were found significantly overexpressed in storage roots (Table 4.6). Together these results emphasize the importance of auxin and ethylene signaling mechanisms in development of storage roots. Cytokinins and Gibberellic acid (GA) are just emerging to be known as modulators of the action of auxin in lateral root development in many plant species (Fukaki and Tasaka, 2009; Gou et al., 2010; Pernisova et al., 2009; Ruzicka et al., 2009; Zimmermann et al., 2010), and although unraveling the exact mechanism of these hormones in sweetpotato is complex, t-ZR is a known bulking inducing factor at the onset of storage root development (Eguchi and Yoshida, 2008). Relevant for sweetpotato is the observation that lateral root development is notably different in initiating storage roots vs. non-storage roots (Villordon et al., 2012). Apart from auxin and ethylene interactions in the development of roots, high levels of gibberellic acid, GA, is known to counter the development of tuberous roots in potato, cassava and sweetpotato. Three GA-inducible genes were found upregulated in fibrous roots: a GRAS family transcription factor (probeid CUST_39798_PI427086615), and gibberellin 2-oxidase [GA2-oxidase (GA2ox)] and three GASA gibberellin regulated proteins (CUST_11378_PI427086615, CUST_25604_PI427086615). In addition, a single transcript, Gibberellin regulated protein (RT_091599.1), was notably upregulated in storage roots in the SPOArrayv1. GA2ox and GA20ox enzymes are dioxygenases that degrade and inactive GA and encode multiple gene families. At least 4 genes coding for different GA2ox isozymes (GA2ox1,-2,-3,-4) exists in Arabidopsis, and the upregulation of these genes by application of GA support the existence of a positive feedforward regulation of these gene. This means that an increase in the concentration of GA in the tissues trigger the synthesis of GA degradation enzymes. However, other external and internal factors are involved in the homeostatic changes of GA. For

example, under a short-day photoperiod potato tuberization is induced and is apparently due to increased expression of catabolic dioxygenases (Carrera et al., 1999) and GRAS transcription factors are implicated in lateral root development at early stages of root development (Zimmermann et al., 2010).

Transcription factors related to lignification and secondary metabolism: A unique set of genes differentially expressed between storage root and fibrous roots are transcription factors and enzymes associated with the arrangement of the cell wall, synthesis of secondary metabolites. For example, genes downregulated in storage roots (alternatively upregulated in fibrous roots) were a set of transcription factors (TF) from WRKY and MYB/MYC families. Twelve WRKY and six MYB/MYC transcription factors were upregulated in fibrous roots rather than in storage roots. Additionally a single bHLH transcriptional regulator is downregulated in fibrous roots. Complexes of MYB-bHLH proteins are known to target the genes involved in the biosynthesis of flavonoids and have been studied in greater detail in many plants like petunia, snapdragon and maize (Vom Endt et al., 2002). In general, MYB proteins alone or combined with a specific bHLH protein regulate the secondary metabolism in plants (Yang et al., 2012). Enhanced expression of these genes in fibrous roots, many of them harboring an anatomy of adventitious roots (data not shown), is expected since plant secondary metabolites are synthesized at early developmental stages. However, MYB and bHLH factors also act in activating and repressing secondary metabolic pathways. MYB transcription factors like WRKY and KNOX have been implicated in the transcriptional activation of genes encoding enzymes for lignin biosynthetic pathways in many species (Guillaumie et al., 2008; Guillaumie et al., 2010). For example, the VvWRKY2 gene is involved in altered expression of genes involved in the lignin biosynthesis pathway and cell wall formation in both grapevine and transgenic tobacco plants (Guillaumie et al., 2010). Lignification halts the thickening of the primary adventitious roots, preventing storage

root formation and converting them to pencil roots (Villordon et al., 2009). Noteworthy is the BP (BREVIPEDICELLUS) gene associated with genes involved in lignin metabolism and secondary cell wall modifications. In sweetpotato, three KNOX I genes, Ibkn1, Ibkn2 and Ibkn3, were found preferentially upregulated in developing tubers (Tanaka et al., 2008), a tissue which is poorly lignified; Ibkn2 and Ibkn3 are homologous to BP, while Ibkn1 is homologous to the STM gene. For instance, three BP orthologous genes, Populus ARK2 (ARBOR-KNOX2), the peach ortholog KNOPE1, and the poplar KNAP2 genes, along with *Arabidopsis* BP, are all involved in secondary growth and lignin metabolism during xylogenesis processes in normal growth. Overexpression of any of these genes in transgenic plants will downregulate the expression of genes for lignin and cellulose synthesis (Du et al., 2009; Hertzberg et al., 2001; Mele et al., 2003; Testone et al., 2008). There is a notable abundance of transcripts encoding enzymes involved in xyloglucan metabolism; 13 and 12 XHT transcripts were significantly downregulated and upregulated in storage roots, respectively (Table 4.6 and Table 4.7) and upregulation suggests that XTHs exhibit a central role in sweetpotato root development; similarly, XTHs are documented to be abundant in monocots, e.g., 29 genes in rice, where they are key to cell wall restructuring and also expressed in different organs in a temporally and spatially controlled fashion (Yokoyama et al., 2004); A*rabidopsis,* is also rich in XTH genes (33 AtXTH genes) and four of them are expressed specifically in roots (Vissenberg et al., 2005). This study reported evidence that supports four root-XTH genes were generated by gene duplication and that they subsequently diversified their expression profile within the root in such a way that they acquired different physiological roles in the cell wall dynamics. Although we validated the expression of a single sweetpotato XTH gene, IbXTH1 (Contig16599.1) (Table 4.8), the sequences clustered in this contig are highly similar, and interestingly, the presence of an indel in the reads (Figure 4.1)

is close to the location of the designed probe included in SPOArrav1. This suggests a similar evolutionary path as in monocots and *Arabidopsis*.

Genes for transport and molecular signaling: We also found an abundance of transcripts required for transport and molecular signaling, such as aquaporins and calcium binding proteins. Further review of the aquaporin sequences revealed that two out of the nine different aquaporin transcripts were upregulated in storage roots (Contig046040.3 and Contig5724.1) and derived from root library sequences and partial sequences of the coding regions or the non-coding terminal end of their putative genes. On the other hand, four aquaporins (Contig16304.1, Contig30548.1, Contig515.1 and Contig989.1) have been assembled with sequences coming from different sources (stem, leaves and roots). Surprisingly, most of these aquaporin sequences are highly conserved at the nucleotide level across many plant species. For example, the aquaporin sequence (RT_344595.1) from sweetpotato has a matching sequence in *Ipomoea nil, Malus prunifolia* and *Juglans regia*. An aquaporin gene, previously reported, IbPIP1, was found downregulated in storage roots in a similar way to increased expression in fibrous roots in response to drought stress (Kim et al., 2009); we observed in the SPTSA assembly, the homologous sequence corresponding to IbPIP1 contig_8050 (accession JP112638.1) is also downregulated in storage roots (Tao et al., (2012). A second aquaporin gene, IbPIP2, is newly reported in this study and it has an enhanced expression in storage roots and was validated by qRT-PCR (Table 4.8). Although the observed expression of IbPIP1 and IbPIP2 genes in our work and others (Kim et al., 2009) suggests a wide range of roles for aquaporins for sweetpotato, it is unknown whether this enhanced expression of aquaporins in sweetpotato participates directly in the initiation of storage roots, but most probably it is related to cellular processes occurring in the whole plant; another, interesting set of new genes reported for first time in sweetpotato is calcium dependent protein kinases (CDPK), and calcium binding proteins (CBP).

Here we validated the enhanced expression of IbCBP2 in storage roots, but two probes for IbCDPK3 showed discordant results (Table 4.5) with one of them being slightly upregulated in storage roots. Unfortunately, we did not include a probe for IbCBP1 in the SPOArrayv1; but IbCBP1 is another gene notably upregulated in early developing stages of storage root. CDPKs and aquaporins encoding genes have been found involved in tuber formation in potato (Raices et al., 2003; Sarkar, 2008). An additional component for molecular signaling and response to stimulus are protein kinases. We found a single protein kinase, (IbKIN1, Contig1100.1, 407 bp), with ~4 to 24 times expression in storage roots supported by the microarray and qRT-PCR results (Table 4.8). IbKIN1 is 72% identical to a transcript encoding a predicted wall-associated receptor kinase-like 10-like (LOC100258149, XM_002270779.1) from grape and 89% identical to a transcript from *Ipomoea nil* (CJ770533.1). We also found that IbKIN1n1 is matching (97%) two shorter sequences from the SPTSA assembly, Contig_26947 of 398 bp (JP131535.1) and Contig_50224 of 220bp (JP154812.1); although the supporting data suggests an enhanced expression in mature storage roots (Tao et al., 2012), both transcripts occurred in leaf tissues. Four novel genes identified in the current work and related to molecular signaling and associated with storage root formation are those encoding protein phosphatase 2A-like (PP2A) and their regulatory subunits. We previously (chapter 2 and chapter3) reported for first time a gene (IbTAP) encoding a PP2A regulatory subunit with enhanced regulation in storage roots, and the microarray revealed a new transcript[Contig053096.3 (Protein phosphatase-2c, putative)] slightly upregulated in storage root (Table 4.6) and correspond to probeid of CUST_17325_PI427086615. Although, the ubiquitous nature of PP2A-like transcripts may nullify its specific role in roots, their role in molecular signaling and post-transcriptional modifications make them a novel type of proteins that merits further study in sweetpotato. Thus,

aquaporins, calcium binding proteins, and member of the protein kinases and protein phosphatases appear to occur in the development of the storage roots in sweetpotato.

Novel genes derived from the sweetpotato root transcriptome: By DGE analysis of the sweetpotato root transcriptome we have found 1,795 out of 55,296 contigs representing sweetpotato unigenes (PBL assembly) are upregulated in initiating storage roots vs. lignified roots (N. Firon, unpublished data). The evidence is based on at least 10 Illumina reads (normalized counts) derived of initiating storage roots mapped in the contigs, compared to zero or up to 2 reads (ratio 5 to 1) derived from non-storage, lignified roots (data not shown). A set of 889 contig sequences derived from PBL assembly were included in the SPOArrayv1 and the majority of them (837 sequences) were initially known to have no annotation at the completion of the design of the array and the experimental work reported here. We carried out further BLASTN analysis in order to better describe probes without annotation from unigenes from the PBL assembly. Indeed, five out of 837 unannotated sequences had a matching sequence, leaving 832 sequences (and its corresponding features in the SPOArrayv1) as unannotated. The following are sequence names(in bold) and their annotation:

S_PBL_c2430	Predicted protein OS=*Populus trichocarpa*
S_PBL_lrc49266	*Ipomoea batatas* expansin mRNA
S_PBL_lrc26197	XP_003611068.1 hypothetical protein MTR_5g010060 [*Medicago truncatula*]
S_PBL_c10460	P83241 IP23_CAPAN Proteinase inhibitor PSI-1.2 OS=*Capsicum annuum*
S_PBL_c22205	P46643.1 AAT1_ARATH Aspartate aminotransferase, mitochondrial
S_PBL_c25632	gar2_schpo ame: full=protein gar2
S_PBL_c40771	AJ429702*.1 Ipomoea batatas* chloroplast atpE gene (partial)

S_PBL_c14800	*Vitis vinifera* clone ss0aeb29yh09; BJ553125 *Ipomoea nil*
S_PBL_c16382	XP_002264818.1, PREDICTED: hypothetical protein [*Vitis vinifera*]

The above results (also shown in table 4.6) indicate that organellar sequences appear to be included in the set of 897 initially unannotated sequences from PBL assembly placed in the SPOarrayv1 array; indeed, after further BLAST analysis, two of these sequences were found to have a matching entry in sequences derived from chloroplast (S_PBL_c40771, *Ipomoea batatas* chloroplast atpE/tRNA-Val /tRNA-Met genes) and mithochodria (S_PBL_c22205, P46643.1 AAT1_ARATH Aspartate aminotransferase, mitochondrial).

We found that 69 unannotated transcripts (out of 827) with a notable upregulation in storage roots, with fold change ratio ranging from 3 to 12.6 times (Table 4.6, Appendix F). The present data support that these set of transcripts are new for sweetpotato as specific for induction of storage root formation and certainly need further characterization. Although, a preliminary analysis of these transcripts revealed that they have a matching transcript in a recent study of sweetpotato (Tao et al, 2012), the results of the expression of these transcripts reported by these authors is opposite, suggesting that they are downregulated rather than upregulated. Results of Tao et al, (2012) are based on DGE analysis in contrast to our PBL assembly; time points also differed between the two studies. The five newly annotated transcripts presented above and an additional 69 transcripts based on DGE analysis (N. Firon. personal communication, 2011) suggested that all of them were upregulated in initiating storage roots vs. lignified roots. In contrast, the microarray results validated preferential expression of only 74 transcripts in storage roots. Two of these transcripts were annotated as derived from proteinase inhibitor and expansin (Table 4.6) and have an enhanced expression as supported by their high fold ratio of 7.25 and 3.95, respectively. The fact that a transcript (S_PBL_c14800) storage roots, had a 96% identity

to a nucleotide sequence in *Ipomoea nil* (accession BJ553125) support our notion that this transcripts could be specific for the *Ipomoea* genera. Similarly, S_PBL_c22205 transcript, with fold change ratio of ~2, is another example of sequences that have a matching sequence in the SPTSA assembly (JP105829.1, aspartate aminotransferase) and upregulated in the ITR (initial tuberous roots) and ERT (expanding storage root) stages (Tao et al., 2012). However, its annotation, including S_PBL_c40771, suggest that they are derived from organellar genes; thus contaminating transcripts from non-coding RNA, or inclusive plastid DNA which were co-purified with the nuclear-encoded transcript may be a serious issue for gene expression profiling..

Another, putative novel transcript (gene) associated with storage root formation, identified in this set of probes, is S_PBL_lrc26197 which is upregulated in storage roots (fold change ratio of 3.5) (Table 4.6). Bioinformatic analysis of S_PBL_lrc26197 revealed that it has homologous sequences which were consistently upregulated with DGE analysis from two transcriptomes of sweetpotato: the PBL assembly (N. Firon, personal communication) and the SPTSA assembly (JP125047.1,TSA: *Ipomoea batatas* Contig_20459) (Tao et. al, 2012)]. Indeed, the sequence JP125047.1 that matches with S_PBL_lrc26197, had high abundance in expanding storage roots and initial storage roots (430 and 827 number of transcripts per million of clean tags (TPM) compared to fibrous roots (2.51 TPM) in the study of Tao et al. 2012; however, they were not able to annotate this sequence. Hence, we suggest denominating the putative gene derived from transcript S_PBL_lrc26197 (Figure 4.1) as "*Ipomoea batatas* Inducing Factor 1, IbIF1" or "Inducin". Inducin/IbIF1-like transcript (S_PBL_lrc26197) have been found upregulated in storage roots derived in this work (DGE analysis of PBL assembly), with 727 normalized reads in initiating storage root vs. 74 reads in lignified, non-storage roots. Inducin ORF encodes a peptide of 58 amino acids in length. The closest matching protein sequence to

"Inducin" IbIF1 is accession XP_003611068.1 in grape. XP_003611068.1 has close similarity to At1g51690.1 (protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform) of *Arabidopsis thaliana*. Type 2A serine/threonine protein phosphatases (PP2A) are implicated in the regulation of signal transduction and control of cell metabolism, and their activities are regulated at post-translational levels by regulatory subunits like the one encoded by AT1G51690.1. The notable expression of Inducin/IbIF1 in storage roots as suggested by the SPOArrayv1 array results (3.5X), supported by high normalized counts of reads in storage roots vs. non storage roots in two independent studies, substantiates our view that this gene is worthy of further study.

This study is the first report describing the use of a custom oligonucleotide-based microarray, SPOv1array, in the expression profiling of sweetpotato storage root development. Although, the fold changes detected in this study are slightly low, correlations with previously known genes, such as Expansin, β-amylase and AGPase associated with the onset of storage root formation was found. The present results showed novel genes, which are members of transcription factor family (i.e. BELL, bHLH), aquaporin and calcium sensing and signaling, upregulated in early expanding storage roots and are worth further and detail investigation. The novel genes with enhanced expression in early expanding storage roots genes are implicated in cell development and differentiation, transcription regulators, enzymes and proteins associated with molecular signaling, transport, and cell wall modifications.

4.5 Literature cited

Altschul S., Madden T., Schaffer A., Zhang J.H., Zhang Z., Miller W., Lipman D. (1998) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Faseb Journal 12:A1326-A1326.

Audic S., Claverie J.M. (1997) The significance of digital gene expression profiles. Genome Research 7:986-995.

Baisakh N., RamanaRao M.V., Rajasekaran K., Subudhi P., Galbraith D., Jonda J., Vanier C., Pereira A. (2012) Enhanced salt stress tolerance of rice plants expressing a vacuolar H+-ATPase subunit c1 (*SaVHAc1*) gene from a halophyte grass *Spartina alterniflora* Löisel. Plant Biotechnology Journal 10: 453-464.

Carrera E., Jackson S.D., Prat S. (1999) Feedback control and diurnal regulation of gibberellin 20-oxidase transcript levels in potato. Plant Physiology 119:765-773.

Chou H.-H. (2010) Shared probe design and existing microarray reanalysis using PICKY. Bmc Bioinformatics 11.

Conesa A., Gotz S., Garcia-Gomez J.M., Terol J., Talon M., Robles M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-6.

Desai D.P. (2008) Understanding the genetic basis of storage root formation along with starch and betacarotene biosynthesis and their inter-relation in Sweetpotato (*Ipomoea batatas* LAM.). (Doctoral dissertation), University of Natural Resources and Life Sciences,Vienna. pp. 144.

Du J., Mansfield S.D., Groover A.T. (2009) The Populus homeobox gene ARBORKNOX2 regulates cell differentiation during secondary growth. Plant Journal 60:1000-1014.

Eguchi T., Yoshida S. (2008) Effects of application of sucrose and cytokinin to roots on the formation of tuberous roots in sweetpotato (*Ipomoea batatas* (L.) Lam.). . Plant Root 2:7-13.

Eujayl I., Morris C. (2009) Identification of differentially expressed UniGenes in developing wheat seed using Digital Differential Display. Journal of Cereal Science 49:316-318.

Fukaki H., Tasaka M. (2009) Hormone interactions during lateral root formation. Plant Molecular Biology 69:437-449.

Gou J., Strauss S.H., Tsai C.J., Fang K., Chen Y., Jiang X., Busov V.B. (2010) Gibberellins Regulate Lateral Root Formation in Populus through Interactions with Auxin and Other Hormones. Plant Cell 22:623-639.

Guillaumie S., Goffner D., Barbier O., Martinant J.P., Pichon M., Barriere Y. (2008) Expression of cell wall related genes in basal and ear internodes of silking brown-midrib-3, caffeic acid O-methyltransferase (COMT) downregulated, and normal maize plants. BMC Plant Biol. 8:71.

Guillaumie S., Mzid R., Mechin V., Leon C., Hichri I., Destrac-Irvine A., Trossat-Magnin C., Delrot S., Lauvergeat V. (2010) The grapevine transcription factor WRKY2 influences the lignin pathway and xylem development in tobacco. Plant Molecular Biology 72:215-234.

Hertzberg M., Aspeborg H., Schrader J., Andersson A., Erlandsson R., Blomqvist K., Bhalerao R., Uhlen M., Teeri T.T., Lundeberg J., Sundberg B., Nilsson P., Sandberg G. (2001) A transcriptional roadmap to wood formation. Proceedings of the National Academy of Sciences of the United States of America 98:14732-14737.

Huang X.Q., Madan A. (1999) CAP3: A DNA sequence assembly program. Genome Research 9:868-877.

Ivanchenko M.G., Muday G.K., Dubrovsky J.G. (2008) Ethylene-auxin interactions regulate lateral root initiation and emergence in Arabidopsis thaliana. Plant Journal 55:335-347.

Kim S.H., Song W.K., Kim Y.H., Kwon S.Y., Lee H.S., Lee I.C., Kwak S.S. (2009) Characterization of full-length enriched expressed sequence tags of dehydration-treated white fibrous roots of sweetpotato. Bmb Reports 42:271-276.

Lewis D.R., Negi S., Sukumar P., Muday G.K. (2011) Ethylene inhibits lateral root development, increases IAA transport and expression of PIN3 and PIN7 auxin efflux carriers. Development 138:3485-3495.

Lin Y.-C., Chen H.-M., Chou I.M., Chen A.-N., Chen C.-P., Young G.-H., Lin C.-T., Cheng C.-H., Chang S.-C., Juang R.-H. (2012) Plastidial Starch Phosphorylase in Sweet Potato Roots Is Proteolytically Modified by Protein-Protein Interaction with the 20S Proteasome. PLoS ONE 7:e35336.

McGregor C. (2006) Differential expression and detection of transcripts in sweetpotato (*Ipomoea batatas* (l.) lam.) using cdna microarrays, Louisiana State University, Baton Rouge, USA.

Mele G., Ori N., Sato Y., Hake S. (2003) The knotted1-like homeobox gene BREVIPEDICELLUS regulates cell differentiation by modulating metabolic pathways. Genes & Development 17:2088-2093.

Noh S.A., Lee Y.H., Kim Y.S., Paek K.-H., Shin J.S., Bae J.M. (201) Down-regulation of the IbEXP1 gene enhanced storage root development in sweetpotato. J. Exp. Bot.

Pelloux J., Rusterucci C., Mellerowicz E.J. (2007) New insights into pectin methylesterase structure and function. Trends in Plant Science 12:267-277.

Pernisova M., Klima P., Horak J., Valkova M., Malbeck J., Soucek P., Reichman P., Hoyerova K., Dubova J., Friml J., Zazimalova E., Hejatko J. (2009) Cytokinins modulate auxin-induced organogenesis in plants via regulation of the auxin efflux. Proceedings of the National Academy of Sciences of the United States of America 106:3609-3614.

Raices M., Gargantini P.R., Chinchilla D., Crespi M., Tellez-Inon M.T., Ulloa R.M. (2003) Regulation of CDPK isoforms during tuber development. Plant Molecular Biology 52:1011-24.

Ruzicka K., Ljung K., Vanneste S., Podhorska R., Beeckman T., Friml J., Benkova E. (2007) Ethylene regulates root growth through effects on auxin biosynthesis and transport-dependent auxin distribution. Plant Cell 19:2197-2212.

Ruzicka K., Simaskova M., Duclercq J., Petrasek J., Zazimalova E., Simon S., Friml J., Van Montagu M.C.E., Benkova E. (2009) Cytokinin regulates root meristem activity via modulation of the polar auxin transport. Proceedings of the National Academy of Sciences of the United States of America 106:4284-4289.

Sampedro J., Cosgrove D.J. (2005) The expansin superfamily. Genome Biology 6.

Sarkar D. (2008) The signal transduction pathways controlling in planta tuberization in potato: an emerging synthesis. Plant Cell Reports 27:1-8.

Tanaka M., Kato N., Nakayama H., Nakatani M., Takahata Y. (2008) Expression of class I knotted1-like homeobox genes in the storage roots of sweetpotato (*Ipomoea batatas*). Journal of Plant Physiology 165:1726-1735.

Tao X., Gu Y.-H., Wang H.-Y., Zheng W., Li X., Zhao C.-W., Zhang Y.-Z. (2012) Digital Gene Expression Analysis Based on Integrated De Novo Transcriptome Assembly of Sweet Potato *Ipomoea batatas* (L.) Lam. PLoS ONE 7.

Testone G., Bruno L., Condello E., Chiappetta A., Bruno A., Mele G., Tartarini A., Spano L., Innocenti A.M., Mariotti D., Bitonti M.B., Giannino D. (2008) Peach [Prunus persica (L.) Batsch] KNOPE1, a class 1 KNOX orthologue to Arabidopsis BREVIPEDICELLUS/KNAT1, is misexpressed during hyperplasia of leaf curl disease. J Exp Bot 59:389-402.

Villordon A., LaBonte D., Solis J., Firon N. (2012) Characterization of Lateral Root Development at the Onset of Storage Root Initiation in 'Beauregard' Sweetpotato Adventitious Roots. Hortscience 47:961-968.

Villordon A.Q., La Bonte D.R., Firon N., Kfir Y., Pressman E., Schwartz A. (2009) Characterization of Adventitious Root Development in Sweetpotato. Hortscience 44:651-655.

Vissenberg K., Oyama M., Osato V., Yokoyama R., Verbelen J.P., Nishitani K. (2005) Differential expression of AtXTH17, AtXTH18, AtXTH19 and AtXTH20 genes in Arabidopsis roots. Physiological roles in specification in cell wall construction. Plant and Cell Physiology 46:192-200.

Vom Endt D., Kijne J.W., Memelink J. (2002) Transcription factors controlling plant secondary metabolism: what regulates the regulators? Phytochemistry 61:107-114.

Wang Z., Gerstein M., Snyder M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57-63.

Xie F., Burklew C.E., Yang Y., Liu M., Xiao P., Zhang B., Qiu D. (2012) De novo sequencing and a comprehensive analysis of purple sweet potato (*Ipomoea batatas* L.) transcriptome. Planta 236:101-113.

Yang C.Q., Fang X., Wu X.-M., Mao Y.-B., Wang L.-J., Chen X.-Y. (2012) Transcriptional regulation of plant secondary metabolism. Journal of integrative plant biology 54:703-12.

Yokoyama R., Rose J.K.C., Nishitani K. (2004) A surprising diversity and abundance of xyloglucan endotransglucosylase/hydrolases in rice. Classification and expression analysis. Plant Physiology 134:1088-1099.

You M.K., Hur C.G., Ahn Y.S., Suh M.C., Jeong B.C., Shin J.S., Bae J.M. (2003) Identification of genes possibly related to storage root induction in sweetpotato. Febs Letters 536:101-105.

Zimmermann R., Sakai H., Hochholdinger F. (2010) The Gibberellic Acid Stimulated-Like Gene Family in Maize and Its Role in Lateral Root Development. Plant Physiology 152:356-365.

Chapter 5 Summary and Conclusions

5.1 Summary and conclusions

Understanding gene expression is essential to furthering our knowledge of storage root development in sweetpotato. It also represents a means of identifying critical genes involved in conferring tolerance to important abiotic stresses, e.g., drought, heat, and salinity, and those related to growth. The current research has two goals. First, develop a usable database of genes from transcripts expressed in root tissue. This involved coalescing transcript data from root tissue generated in the current study and extant data from public sources into a meaningful transcriptome. This resource can then be used in breeding and understanding basic growth and developmental processes. A second goal is to use this genome wide resource to identify genes of importance in sweetpotato storage root development under normal growth conditions and drought. Greenhouse studies were also conducted to demonstrate the impact of drought on storage root number and development.

Organizing the tens of thousands of genes expressed into a meaningful transcriptome represented a major component of the study. Sequence data generated by a collaborator using deep sequence technology (454 and Illumina systems) or Next Generation Sequencing and publically available sources was merged into a transcriptome by using a combination of assemblies. We estimated that a *de novo* construction based on combining NGS assemblers and CAP3 is a valid alternative for sweetpotato compared to other software such as CLC Genomics Workbench and Trinity (http://trinityrnaseq.sourceforge.net/). The number of unique transcripts generated in the current study and others ranges from 75K to 101K, and higher than 66K reported in sweetpotato by Schafleitner et al. (2010) and half of 148K by Tao et al., (2012). Our data showed that the number of unique genes expressed in sweetpotato is ~ 48K to 52K. This is a conservative estimate given that it represents transcripts with described sequences in current

databases. This number is similar to the reported 39K protein-coding genes of potato (Potato Genome Sequencing Consortium, 2011) and 21K annotated transcripts from coffee (Mondego et al., 2011); sweetpotato, potato and coffee belong to the Asterid I clade of dicot plant families. Further clues on functional role of the sweetpotato unigenes could be derived from molecular studies on these species.

Microarray gene expression analysis has proven to be a useful tool for providing clues to the mechanisms involved in storage root formation (Desai et al., 2008). We designed a custom array for sweetpotato based on ~ 14K transcripts selected from a GO annotation in the following categories: Transcription Factors, Starch and Sucrose Metabolism, Kinase, Molecular Signaling, and Membrane. Expression profiling in storage roots vs. fibrous roots revealed a set of genes with roles in the control of gene expression, anatomical structure morphogenesis, molecular signaling, and carbohydrate and secondary metabolism. Expression was enhanced in expanding storage root. We confirmed the expression of previously known genes such as transcription factors, and enzymes for other cellular processes (i.e. transport and cell division). Examples are enzymes AGPase, and β-amylase, KNOX proteins, and storage protein sporamin. Next generation sequencing was found to be useful in assessing the dynamics and abundance of transcripts in sweetpotato in multiple tissues as well as to study the mechanism involved in sweetpotato root development. NGS allows the quantification of the expression of genes by what is called Digital Gene Expression analysis. Clues that may arise from Digital Gene Expression analysis can provide insight into the role genes may have in tissues unrelated to the examined plant organ, e.g., the storage root, long-distance signaling. PP2A-like genes encoding serine/Threonine-phosphatases products described in the current study may integrate signals from leaf to roots, similar to events described in potato. Another example is the movement of BEL1-like mRNAs (BELL genes) from the leaf and stem tissues toward the stolon have been

demonstrated to trigger tuberization in potato (Banerjee et al., 2006). Movement of transcripts of one potato BEL1 gene, St BEL5 was correlated with enhanced tuber production and increased cytokinin levels (Banerjee et al., 2006; Chen et al., 2003).

Whether the transcripts of BELL genes reported in this study (chapter2 and 3) act as mobile RNAs is worth further investigation to understand the intricate molecular events during the thickening of adventitious roots. BELL genes appear important. First, expression in the root transcriptome is abundant (we found at least 5 to 7 different transcripts). Secondly, expression studies in drought and salt stressed have a specific role in storage root development given significantly higher expression in different tissues (fibrous roots, storage roots, thick pigmented roots and leaf) in comparison to the controls. Thirdly, BELL genes have been found concomitantly interacting or partnering with KNOX genes in plants such as *Arabidopsis* (Bellaoui et al., 2001), barley (Muller et al., 2001) and potato (Chen et al., 2003). Fourth, three KNOX genes reported in sweetpotato appear to be key for thickening of the adventitious roots into storage roots. The current research showed up-regulation of KNOX genes (Ibkn2 and Ibkn3) in the storage roots. Furthering our contention that BELL and KNOX genes are central to the formation of storage roots in sweetpotato. This represents new and novel data in understanding storage root physiology. Microarray and qRT-PCR validated downregulation of IbBEL1 and IbCRF1 in young storage roots compared to fibrous roots, which is in contrast to the enhanced expression of these transcripts in 2-week-old roots developed under initial drought stress.

Although we have carried out the expression profiling of genes of the lignin metabolic pathway, the results are not conclusive to our surprise (data not shown). The data suggests that these specific genes have to be studied in a more refined time course study at multiple stages of development in contrast to storage roots vs. non-storage roots, as used in the present study.

Based on the global gene expression profiling using the custom oligo microarray we found 783 and 328 Differentially expressed transcripts putatively upregulated and downregulated in storage roots; both qRT-PCR and microarray results support the relevance of calcium as a component for the onset of storage root formation. Calcium is a well-known second messenger involved in both response to biotic and abiotic stresses as well as involvement in developmental and differentiation processes in plants. We had seen an altered expression of 20 genes specifically associated with thickening of adventitious roots under drought and salt stress and enhanced expression of genes specific for abiotic stress response. The reduced number of storage roots observed is likely related.

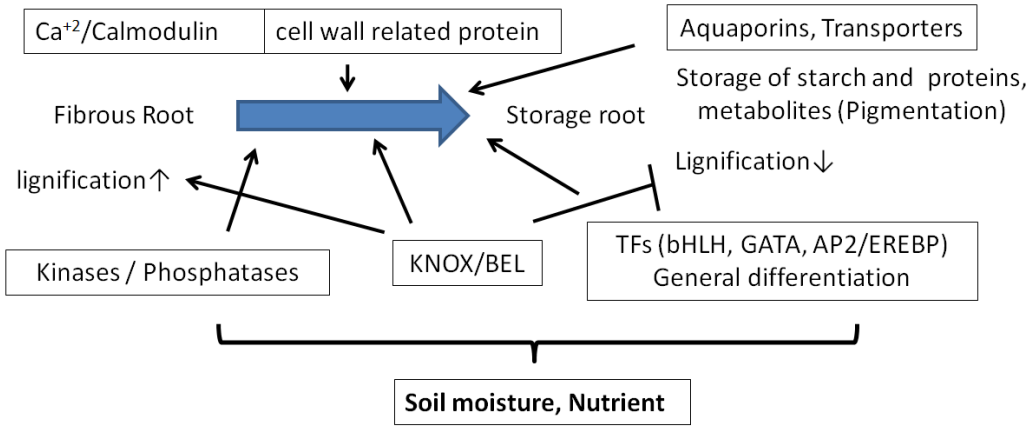


Figure 5.1. Hypothesis of proteins involved in storage root formation. TF=transcription factors. Calcium signaling is an important component in the onset of storage roots. Regulatory transcriptions factor might act as a network that influence both metabolic pathways (enzymes) as well as transport and allocation of secondary metabolites Expression of Knox/BEL genes that act by possibly suppressing the lignification in a similar way to Arabidopsis, peach and poplar. Post-transcriptional modifications are involved in triggering the phase of fibrous to storage root and in response to the environment.

Based on the results from the current study a hypothetical scheme is presented that explains the mechanism of storage root development in sweetpotato (Fig. 5.1). While future studies are necessary to better characterize how these differentially expressed genes contribute to the onset of storage roots, this study is the first step towards an understanding of the complex

mechanism that occurs in sweetpotato. Transcripts of 46 protein transporters are notably downregulated in the storage roots, and most of them represent nitrate transporters and multidrug resistance/P-glycoprotein (MDR/PGP) families of ABC transporters. MDR/PGP families are involved in polar auxin transport and their role in lateral root formation is an example of the cross-talk between auxin and ethylene signaling (Ivanchenko et al., 2008; Lewis et al., 2011; Negi et al., 2010). The fact that a high density of lateral roots had been associated with storage root formation (Villordon et al., 2012), substantiates the roles these transporters might have during the development of sweetpotato roots.

5.2. Future research

Recommendations based on the results of the current work.

1. The CAP3 and MIRA have been integrated into a new assembler, iAssembler (Zheng et al., 2011). It is likely that this software has advantages in developing an updated transcriptome in sweetpotato. The advantage of this software is that it reduces two types of assembly errors: a.) type I error-ESTs derived from alternatively spliced transcripts or paralogs which are incorrectly assembled into one transcript; 2) type II error-ESTs derived from the same transcript and fail to be assembled together (Zheng et al., 2011).

2. Improve trait specific gene association for drought tolerance. Gene expression profiling should be dynamically conducted by sampling at different time points. For example, changes in gene expression in response to an abiotic stress, i.e. drought stress, should be followed in a timely way. The goal of the present study was to assess the long term effect of drought stress imposed on plants developing storage roots and to correlate it with the expression of putative genes associated with the

appearance of storage roots. This study did not focus on incremental time points to more closely relate specific genes in response to salt and/or rought stress.

3. Sectioning of early developing storage roots is difficult, but better characterization of genes associated with the onset of storage root formation could be conducted, and I recommend carrying out this kind research. Although the procedure limits the number of genes one can study, it is worthwhile for those deemed critical.

4. Using sand as a growth substrate facilitated removal of roots for study. However, storage root count may vary if a different substrate is used. Studying the development of storage and non-storage roots under controlled conditions in either greenhouse or growth chambers should be conducted using both sand and soil where sweetpotato is typically grown in.

5. Although the usefulness of a custom oligo array for sweetpotato is a good alternative to RNA-seq, the improvement of any custom oligo array could be done by designing and placing multiple probes for each gene apart from the random spotting in the array. A further improvement to the present customized array should be done by including transcripts from all cellular processes.

5.3. Literature cited

Banerjee A.K., Chatterjee M., Yu Y., Suh S.-G., Miller W.A., Hannapel D.J. (2006) Dynamics of a mobile RNA of potato involved in a long-distance signaling pathway. Plant Cell 18:3443-3457.

Bellaoui M., Pidkowich M.S., Samach A., Kushalappa K., Kohalmi S.E., Modrusan Z., Crosby W.L., Haughn G.W. (2001) The Arabidopsis BELL1 and KNOX TALE homeodomain proteins interact through a domain conserved between plants and animals. Plant Cell 13:2455-2470.

Chen H., Rosin F.M., Prat S., Hannapel D.J. (2003) Interacting transcription factors from the three-amino acid loop extension superclass regulate tuber formation. Plant Physiology 132:1391-1404.

Ivanchenko M.G., Muday G.K., Dubrovsky J.G. (2008) Ethylene-auxin interactions regulate lateral root initiation and emergence in Arabidopsis thaliana. Plant Journal 55:335-347.

Jaillon O., Aury J.-M., Noel B., Policriti A., Clepet C., Casagrande A., Choisne N., Aubourg S., Vitulo N., Jubin C., Vezzi A., Legeai F., Hugueney P., Dasilva C., Horner D., Mica E., Jublot D., Poulain J., Bruyere C., Billault A., Segurens B., Gouyvenoux M., Ugarte E., Cattonaro F., Anthouard V., Vico V., Del Fabbro C., Alaux M., Di Gaspero G., Dumas V., Felice N., Paillard S., Juman I., Moroldo M., Scalabrin S., Canaguier A., Le Clainche I., Malacrida G., Durand E., Pesole G., Laucou V., Chatelet P., Merdinoglu D., Delledonne M., Pezzotti M., Lecharny A., Scarpelli C., Artiguenave F., Pe M.E., Valle G., Morgante M., Caboche M., Adam-Blondon A.-F., Weissenbach J., Quetier F., Wincker P., French-Italian P. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463-U5.

Lewis D.R., Negi S., Sukumar P., Muday G.K. (2011) Ethylene inhibits lateral root development, increases IAA transport and expression of PIN3 and PIN7 auxin efflux carriers. Development 138:3485-3495.

Mondego J.M.C., Vidal R.O., Carazzolle M.F., Tokuda E.K., Parizzi L.P., Costa G.G.L., Pereira L.F.P., Andrade A.C., Colombo C.A., Vieira L.G.E., Pereira G.A.G., Brazilian Coffee Genome Project C. (2011) An EST-based analysis identifies new genes and reveals distinctive gene expression features of Coffea arabica and Coffea canephora. Bmc Plant Biology 11.

Muller J., Wang Y.M., Franzen R., Santi L., Salamini F., Rohde W. (2001) In vitro interactions between barley TALE homeodomain proteins suggest a role for protein-protein associations in the regulation of KNOX gene function. Plant Journal 27:13-23.

Negi S., Sukumar P., Liu X., Cohen J.D., Muday G.K. (2010) Genetic dissection of the role of ethylene in regulating auxin-dependent lateral and adventitious root formation in tomato. Plant Journal 61:3-15.

Potato Genome Sequencing Consortium X.X., Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, Wang J, Orjeda G, Guzman F, Torres M, Lozano R, Ponce O, Martinez D, De la Cruz G, Chakrabarti SK, Patil VU, Skryabin KG, Kuznetsov BB, Ravin NV, Kolganova TV, Beletsky AV, Mardanov AV, Di Genova A, Bolser DM, Martin DM, Li G, Yang Y, Kuang H, Hu Q, Xiong X, Bishop GJ, Sagredo B, Mejía N, Zagorski W, Gromadka R, Gawor J, Szczesny P, Huang S, Zhang Z, Liang C, He J, Li Y, He Y, Xu J, Zhang Y, Xie B, Du Y, Qu D, Bonierbale M, Ghislain M, Herrera Mdel R, Giuliano G, Pietrella M, Perrotta G, Facella P, O'Brien K, Feingold SE, Barreiro LE, Massa GA, Diambra L, Whitty BR, Vaillancourt B, Lin H, Massa AN, Geoffroy M, Lundback S, DellaPenna D, Buell CR, Sharma SK, Marshall DF, Waugh R, Bryan GJ, Destefanis M, Nagy I, Milbourne D, Thomson SJ, Fiers M, Jacobs JM, Nielsen KL, Sønderkær M,

Iovene M, Torres GA, Jiang J, Veilleux RE, Bachem CW, de Boer J, Borm T, Kloosterman B, van Eck H, Datema E, Hekkert BL, Goverse A, van Ham RC, Visser RG. (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189-195.

Tao X., Gu Y.-H., Wang H.-Y., Zheng W., Li X., Zhao C.-W., Zhang Y.-Z. (2012) Digital Gene Expression Analysis Based on Integrated De Novo Transcriptome Assembly of Sweet Potato *Ipomoea batatas* (L.) Lam. PLoS ONE 7.

Villordon A., LaBonte D., Solis J., Firon N. (2012) Characterization of Lateral Root Development at the Onset of Storage Root Initiation in 'Beauregard' Sweetpotato Adventitious Roots. Hortscience 47:961-968.

Zheng Y., Zhao L., Gao J., Fei Z. (2011) iAssembler: a package for de novo assembly of Roche-454/Sanger transcriptome sequences. Bmc Bioinformatics 12.

APPENDIX

Appendix A: Sequences similar to Ibkn3 gene and detail of their assembly

AB283029.1 is the accession of Ibkn3 gene (Tanaka et al., 2008). A single contig (Contig5774.1, including 19 singlets) was from *cap3p80v3* assembly and two contigs were found in the *cap3p90* assembly (Contig36256.2 and Contig6970.2 including 10 and 6 sequences, respectively). BM878851.1 (plantlets) and ST_314184.1 (stem) are sequences included in Contig36256.2 and Contig6970.2, respectively, but showed here in the alignment. S_PBL_c31412 is a contig from reads from 454 pyrosequencing (PBL assembly using MIRA).

Contig5774.1	Contig36256.2	Contig6970.2
RT_110215.1	ST_314184.1	RT_110215.1
BM878851.1	RT_054770.1	BM878851.1
RT_523022.1	ST_213376.1	RT_390253.1
RT_390253.1	RT_228288.1	RT_404704.1
RT_404704.1	RT_252477.1	RT_291602.1
ST_314184.1	abyssCtg_107877.1	RT_061754.1
RT_291602.1	abyssCtg_50592.1	
RT_061754.1	RT_248360.1	
RT_054770.1	abyssCtg_219910.1	
ST_213376.1	abyssCtg_344035.1	
RT_228288.1		
RT_252477.1		
abyssCtg_107877.1		
abyssCtg_50592.1		
RT_248360.1		
abyssCtg_219910.1		
abyssCtg_344035.1		
abyssCtg_227780.1		
abyssCtg_441570.1		

>S_PBL_c31412

ggacatgcagtttatggtgatggatggtctgcacccacaaaatgcagcactttacatggatggtcactacatgggagacggtccatatcgcct

aggtccataatccaatccaacataatataaaacgtcgacttccatatcgagtttttttaaggttgtaacttaggacatcatgtatctgcctgagacc

gccttgaaatagacgcgtacactatgccttccggccggcttgccgggaatcttgaacgaggggagctagctaggttaacttctcgtttatcttt

gatctgcaacttgtggtaaccttagcaagcaaagcaagatttctccgaaactatgaaggaaagtagggcgtttttatgtatttgtaacgagtttg

atgtagtaattaagtgcaattgtatatggtgagactgatgctagtaaataatttcagttagcctatatataatattattata

>Contig5774.1 (from cap2p80 assembly)

aaagttggggcgccgccggaggtggtggcgcgtcttgcggcggtgcgtcaggagtttgagtcccggcaacgcgccgccggtctcggtg

ggagagatatttcctccaaggacccagaactcgaccagtttatggaagcgtactatgatatgttagtgaagtaccgagaagaactgaccagg

cctttacaagaagcaatggagttcatgcgacggatcgaatcgcaactaaatatgcttagcaacgccccagtccgggtcttcacttcggatgac

aaatgtgagggtgtcggttcctctgaagacgaccaagataacagcggtggtgaaaccgagcttcccgagattgatcctcgagctgaagacc

gcgagctgaagaaccacttgctgaggaagtacagcggctacctaagcagtctgaagcaagagctttcgaagaaaaagaagaaaggaaaa

ctcccaaaagaagccaggcaaaagctgctcaactggtgggagttgcactacaaatggccttatccctcggaaaccgagaaggtggctttgg

ctgaatcgacggggttggatcagaagcagattaacaactggttcatcaatcaaaggaaacgacactggaagccttctgaggacatgcagttt

atggtgatggatggtctgcacccacaaaatgcagctctttacatggatggtcactacatgggagacggtccatatcgcctaggtccataatcc

aacataatataaaacgtcgacttccttaccgagtttttttaaggtgtaacttaggacatcatgtatctgcctgagaccgccttgaaatagacgcgt

acactatgccttccggccggcttgccgggaatcttgaacgaggggagctagctaggttaacttctcgtttatctttgatctgcaacttgtggtaa

ccttagcaagcaaagcaagatttctccgaaactatgaaggaaagtagggcgtttttatgtatttgtaacgagtttgatgtagtaattaagtgcaat

tgtatatggtgagactgatgctagtaaataatttcagttagcctatatataatattattatgattttagtccttt

>Contig36256.2 (from cap3p90 assembly)

taagtccaggcaaaagctgctcaactggtgggagttgcactacaaatggccttatccctcggaaacggagaaggtggctttggctgaatcg

acggggttggatcagaagcagattaacaactggttcatcaatcaaaggaaacgacactggaagccttctgaggacatgcagtttatggtgat

ggatggtctgcacccacaaaatgcagcactttacatggatggtcactacatgggagacggtccatatcgcctaggtccataatccaatccaa

cataatataaaacgtcgacttccatatcgagtttttttaaggttgtaacttaggacatcatgtatctgcctgagaccgccttgaaatagacgcgta

cactatgccttccggccggcttgccgggaatcttgaacgaggggagctagctaggttaacttctcgtttatctttgatctgcaacttgtggtaac

cttagcaagcaaagcaagatttctccgaaactatgaaggaaagtagggcgtttttatgtatttgtaacgagtttgatgtagtaattaagtgcaatt

gtatatggtgagactgatgctagtaaataatttcagttagcctttatataatattattatgattttagtccttt

>Contig6970.2(from cap3p90 assembly)

aaagttggggcgccgccggaggtggtggcgcgtcttgcggcggtgcgtcaggagtttgagtcccggcaacgcgccgccggtctcggtg

ggagagatatttcctccaaggacccagaactcgaccagtttatggaagcgtactatgatatgttagtgaagtaccgagaagaactgaccagg

cctttacaagaagcaatggagttcatgcgacggatcgaatcgcaactaaatatgcttagcaacgccccagtccgggtcttcacttcggatgac
aaatgtgagggtgtcggttcctctgaagacgaccaagataacagcggtggtgaaaccgagcttcccgagattgatcctcgagctgaagacc
gcgagctgaagaaccacttgctgaggaagtacagcggctacctaagcagtctgaagcaagagctttcgaagaaaaagaagaaaggaaaa
ctcccaaaagaagccaggcaaaagctgctcaactggtgggagttgcactacaaatggccttatccctcggaaaccgagaaggtggctttgg
ctgaatcgacggggttggatcagaagcagattaacaactggttcatcaatcaaaggaaacgacactggaagccttctgaggacatgcagttt
atggtgatggatggtctgcacccacaaaatgcagctctttacatggatggtcactacatgggagacggtccatatcgcctaggtccataatcc
aacataatataaaacgtcgacttccttaccaagttttttaaggtgtaacttaggacatcatgtatctgcctgagaccgccttgaaatatacgcgta
cactatgcctttcggccggcttgccgggaatcttgaacgaggggagctaggctaggttaacttctcgtttatctttgttctgcaacttgtggtaac
cttagcaagcaaagca

Appendix B: Sequences similar to BEL1 gene.

Sequences encoding six BEL1-like products derived from PlantGBD (three BELL transcripts), SPGI (one transcript) and PBL assembly (two transcripts).

>PUT-169a-Ipomoea_batatas-65 PlantGDB-assembled Unique Transcript-fragment derived from Ipomoea_batatas mRNAs Jan_14_2009 (based on GenBank release 169) (in this dissertation referred to as IbBEL2)

gacaaagttggagctcgcgcgcctgcaggtcgacctagtggatccaaagaattcggcacgagggaaattcaaatgaagaaagcaaaacttgttaacatgcttgatgaggtggagcagaggtacagacagtaccatcaccagatgcagatagtgataacatggtttgagcaggctgcaggattggttcagccaagacctacacagctctggcattgcagacgatctcgaagcagttcaggtgcctgaaagacgcgatcttggggcaaattcgcgctgccagcaagagtttgggggaggaagacgggctgggagggaagatcgagggctcgtcgaggctcaaatttgtggacaatcagctcagacagcagagggctttgcagcagttgggaatgatccagcacaatgcttggagaccccagagaggattgccagagcgatctgtttctgtgcttcgcgcctggctgttcgagcatttccttcacccttaccccaaggattcagacaagatcatgctggcaaagcagacaggtcttactaggaaccaggtgtctaactggttcatcaatgctcgtgttcgcctatggaagccgatggtggaggagatgtatatggaggagatcaaggagcaagaacagaatggatcagagaacaaaacaggcaaaggtgaaccacatgaagattcagcttctcaccacgacataagccctggaggcttagatcattaagacaagagttttgcccaaaaccagtccactaccatgcctggagggt

>PUT-169a-Ipomoea_batatas-9796 PlantGDB-assembled Unique Transcript-fragment derived from Ipomoea_batatas mRNAs Jan_14_2009 (based on GenBank release 169) (in this dissertation referred to as IbBEL2)

aaacgcgccaccgagctcaccaccgcagaaagacaggaaattcaaatgaaaaaagccaaacttgttaacatgcttgatgaggtggagcagaggtacagacagtaccatcaccagatgcagatagtgataacatggtttgagcaggctgcagggattggttcagccaagacctacacagctttggcattgcagacgatctcgaagcagttcaggtgcctgaaagacgcgatcttggggcaaattcgcgctgccagcaagagttcgggggaggaagacgg

>PUT-169a-Ipomoea_batatas-10648 PlantGDB-assembled Unique Transcript-fragment derived from Ipomoea_batatas mRNAs Jan_14_2009 (based on GenBank release 169) (in this dissertation referred to as IbBEL6).

atcagtcaacgctttgcgcgcttggcttttcgagcattttctccacccgtatccaagcgatgcagataagcacatgctggctcgacagactggt
ctctccaggagccaggtttcaaattggttcattaatgccagggtccggctgtggaaacccatggtggaagagatgtaccaacaagaagttaa
agtaggcgatgatgacagtgaaaacattatggacagtggagacgatgacaccgtggcacaaacaccaacgcctaacgccgtcagtaattc
gcccaccaccaccgtcaccaccacaccgccaccgtcagtgactgccacgtctgccgatgcagtggtggcagctgcagcaggcaaaagat
ccgaaatcaacgctgcccacgaaagcgacccttcactactcgcaatcaatacccaatacacgtgtcaccgccgtggcacggcggattccg
gcggcggcggcaatgtgtcgctcacgctggggctgcgccacgccggaaatttgcccgaccagaatcatttttcggttaggcatgccatgc
cgggggttcggattactatgttgttttttttgttgctcaaagataccctccaccgatagtcaggagactaatcctccgtcacaacagtcagcata
ctaagtgataagagactggtcaaatggtttattgttatgctgttattatcacacgaaaagcacacacatattcggagaatcatatggtttccatgtt
tatgtcttttgtttaaattatttatttcatctagacatgcaattaattaattaatatatacatagtcaaatgtcaccaatggtagcgttatg

>PUT-169a-Ipomoea_batatas-8950 PlantGDB-assembled Unique Transcript-fragment derived from Ipomoea_batatas mRNAs Jan_14_2009 (based on GenBank release 169) (in this dissertation referred to as IbBEL1).

ggccgggggggaacggaggaatgggtttagcggcggcggcggcgagttcatcgtctctcttctaccctagagaccaaattgaagattgcca
aacagtcccttattcacttctggacagtgaaggtcagaacttgccatataggaacttgatgggtgcacagttgctccatgatttggccggttag
agtatcaaccaattaaacaacttcgatcaaaacctaaaactaggctccaaaaattcattaaagaagttgattttatgctgctctgatctgccttcct
atcatctctgttattatttataacctatacttctttaattatatattactatatgaatggggtataatagtacagtgatgaaagaaaatggggatttaca
tatatacataatacattggggttagatgtatgtatacaatataatatatatttgcttccacggctggc

>Ib28923 (Named as 28923 at SPGI, in this dissertation referred to as IbBEL3).

atccacctgtcatcccttcagatttgctaacctgcgccaccctccaccaccacccacttctctctgatcatcacaaacagaccggtaaagggg
cgagtttgatgatgggggggcagtaacactaacacccccactaatttctccactcccactttgtatatggatccgcaatcttctgttcctctccactt
aaatctaagtaaattaccagaaacccgtttctctacgcctccccacagagcctttaggtttcatgataattctttccacggtggaggtggggag

gttactgtggtgtacaaaccggagcctttgtcagtggctcacgagactactaatattaataatgtnaggggtcaaggtttatctctggctttgtcc

tcccatcacactcaccaacaccaaagcgccagtagtcttcctctggagctcaatctgcaaagatatgacgcctcaatgacctctgggcttctg

gtttctagcattcatggtggtggtgatagtaactgctccacttacaatcatgatggtgacctctcacggagctgttcagttcctgggcccttcaca

ggctatgcttcaatcttgaagggttccagattcttgaaacctgctcagcagttgttggaggaggtytgtgatgtggtcaggggygtttatgctg

agaaactggaagctgattcagctttgatggacccttccatggaaactctgagtgggagttctctggttgatgattcaccaaattgtaatgatgga

ggtgagcacaggaggaagaaatccaggttaattcaatgctagatgaggtttacaggaggtataagcagtactaccaacagctgcaagcagt

tgttgcgtcwtttgaatccgttgctggactaaacaatgcagcgcccttcgcaaacttagcactgaaagccatgtcgaagcacttcagatgctt

gaaagacgccataactgagcagctgcagtttacaggcaaatcgcarggccatatgagytaygagggagaagcctctccragcgctgaaa

atctggggaaaggactttatttccaaagcacagctattcgtaatgctgggataatggaacatcaacctgtctggcgtccccaacgggggttac

cagaacgtgcagtcaccgtgctcagagcatggttgttcgatcactttctccacccttatcccacagacacagacaagctaatgttggctaaac

agacaggtctgtctagaaatcaggtgtctaattggttcatcaatgcaagggttaggctatggaagcctatggtggaagagattcacaatctgg

aaacgcggcaagcgcagaagaattca

>S_PBL_c6271 (in this dissertation referred to as IbBEL4)

gatggtggaagagattgtacctggaggagatcaaggagcaagaaaagagtggaggaggaggagaggacaaaacaagcaaagaggaa

gcagatgaagaaggctcagcatctctgcaacaacacaaaagtccccctggctcagaaaatcaagacaggaatgaaattgacaccccaaaa

caagccaataataataataatccatcaatgtccactgcatcaacatctatgaatgggatacccatgaataccaaccatcctgcaggttttag

cctcataggaccctcagaaatggacagcattactcagggaagtccaaagaagccgcggggctcgggctcaggctcagacatgctgctgc

atagcgccatcgttccatcaatagccatcgatgcagcaaagcccgcamccatgaaatttggcaatgacaggcaaacaagagaaggattcc

ccttggagggatcaacaaacttcatggccgccttcgggtcctacccaatcgaggaaatggggcggttcaccaccgagcagttcccctcgcc

ttattccagcaacgctgtctccctaaccctcggcctgcagcacagcgaaaacctctccgccgccatgtctgctgccacacaccacaatttcct

ccccaaccaagacattcaaatgggaaggggagttgtgattggtgaagcagcaaacgattttgttggcggaatgactaccccgacatctgctc

accccaccagcgtgttcgagaacttcaacattcagaaccggaagaggttcccagcacaattgttgccagactttgtcacctgatcgagtttaa

ggttgtctgcagggtcctctttgtctgatcttcttcacttattttcttcctatttacccaaaacttattaggaagtaaggtaaatttttatttatggtagtg

caaacaaccttgaaaaaattaaaaattaaaagtttttacatagtacaggtctacactttgcagaagaaccactagagttcttatggatagtttagg

cttttagaagatgtggttgagattgta

>S_PBL_c24240(in this dissertation referred to as IbBEL5)

Ccgcaaagagggcttcccgagcggtctgttaatgtcttgagatcgtggcttttcgagcattttcttaacccgtatccgagtgaagcagacaag

gttttgttgtctacgccagactggtctgtctaaaaatcaggtttcaaattggttcattaatgctagggtaaggctgtggaaacccatggtggaag

aaatgtacgaacaagaaactaaagaaggaggggggagaagaaataaacgcacacacaccgatgctgcgtgacgatgataataataaggaa

attactgcaacctcaactacgccgacgacgacaaaaccgacaacaacgttaccgtcggctgcagcgacaaaaagatgccaattcaatgcg

acggaaaacgactcctcacgaaccaacatcatcaataattattatgcgccccagtacgcctcgggaaaccaagtcacgctgattggtaacac

cgccgccgtgccgcccctatctccggtagccggaacgccgcacaactggcctggtcaggcggtggattgtggcggtgggtttgggactcc

ggcgacaggggacgtgtccctaactttgggacttcgccacccagaaaatagtgtaaatgttacaaactttgggggctattaaatatttagtact

aaattaaaacacaaaatggactaaaaaaattatatatctatatatatatagttgtgaattatttgggtttgtaattgactatataac

>Contig22795.1 (in this dissertation referred to as IbBEL7)

Nnnnnnnnnnnnnncttaacaataaactttcaggttctgggtgtttacataaatgcattattcataaagttaggtaaaaaagttcgaatctttttg

tttgattattgaacacctttcatctacaataggtttacaggaggtataagcagtactaccaacagctgcaagcagttgttgcgtcatttgaatccgt

tgctggactaaacaatgcagcgcccttcgcaaacttagcactgaaagccatgtcgaagcacttcagatgcttgaaagacgccataactgagc

agctgcagtttacgggcaagtctcaaggccatatgagttacgagggagaggcctctccaagcgctgataatctggggaaaggactttacttc

caaagcacagctattcgtaatgctgggataatggaacatcaacctgtctggcgtccccaacgggggttaccagaacgtgcagtcactgtgct

cagagcatggttgttcgatcacttcctccacccttatccaacagacacagacaagataatgctggctaaacagacaggtctgtctagaaatca

ggtgtctaactggttcatcaatgacaagggttaggctatggaagcctatggtggaagagattcacaatctggaaacgcggcaagcgcaaaa

gaattcacaaaccaaagaacagaataacaacaatccaagcgagcatttaagcatgagaaattctgtggcctgcgaagatgcatctgcgtcg

atgcaccaaattcaagaactcccgacaaagcgaacgcgaaacgaggccgagaattccatgggaagcgatgaggcgatggatttatcatac

ggcaatttatcgcaccacactcgcctgggaatcggcggcgccaccgccgccggcgggagcagcagcgtctccttaaccctaggccttca

ccaaaacaacgagcttggcttgtcggattctttccggtaacgccgcccagcgcttcgggctggacgcgagcagcgaaggattgtgctgtct

gggttcgaaccgcaaaacacgcaattcggtagagacatcatcgga

Appendix C: Sequences of CBP-like genes.

Contigs (Contig20410.2, Contig13109.2 were assembled with CAP3/ABySS or MIRA, PBLassembly (S_PBL_c1633,S_PBL_c13311) . S_PBL_c13311 have two nucleotidic mismatches differences at the primer location of the reported primers for IbCBP3. Similarly the accession JG699346.1 (IbCBP1-like) and JG699262.1 (IbCBP3-like) are available at GenBank. RT_055908.1, RT_027290.1, RT_483640.1 and JG699346.1 are the accession names of some representatives raw sequences included in Contig13109.2 that contains the sequencing information of that were identified for the design of specific primers for IbCBP1. Similarly RT_038063.1, RT_483640.1 and JG699262.1 are the accession/identifier of representative raw sequences included in Contig20410.2 that have pairs with the specific primers designed for IbCBP3.

>Contig20410.2 IbCBP3-like (from cap3p90 assembly)

atcaaaccacccatccaacaacataacaacaaacaaaaacctcaaattaaccaaatccgaagaacataatgagtgtggaaatccttgatgct
gctactattcttaacttcgttcaagacgaaaaggtgtttgatgaattcgtgcacgaacgtttcgataacctggatagcaaccacgatggagttct
ctcctatgcggagctgttgagagagctgcggagcctaagagttttggagatgcattttggtgtagatgtggagaccaatccaaccgagcttaa
ccgagtttatgactcaatgtttgttcagttcgatagggactcaagtggggcggtggacgtgaaggagttcaaagaagagatgaagaatatgat
ggtggcgatggcgaatgatattgggttcttgccggttcaaatgctgcttgaggaaaatagtttcttgaagaaagctgttgagaaggagacaac
aaaaataattgctaatgctgcctaattaacaacttatttttttcctctttttttctttttttggtatgttgctttaattttgaaactaaaatcatattttgtattgta
tgtctgaaggacatacatttcttgttatgaagcagtataatttactaatgaaaa

>S_PBL_c1633 IbCBP3-like

atcaaaccacccatccaacaacataacaacaaacaaaaacctcaaattaaccaaatccgaagaatataatgagtgtggaaatccttgatgct
gctactattcttaacttcgttcaagacgaaaaggtgtttgatgaattcgtgcacgaacgtttcgataacctggatagcaaccacgatggagttct
ctcctatgcggagctgttgagagagctgcggagcctaagagttttggagatgcattttggtgtagatgtggagaccaatccaaccgagcttaa
ccgagtttatgactcaatgtttgttcagttcgatagggactcaagtggggcggtggacgtgaaggagttcaaagaagagatgaagaatatgat

ggtggcgatggcgaatgatattgggttcttgccggttcaaatgctgcttgaggaaaatagtttcttgaagaaagctgttgagaaggagacaac

aaaaataattgctaatgctgcctaattaacaacttawttttcctcttttttcttttggtattgttgctttaatttgaaactaaatcatatttgtattgtatgtc

tgaagg

> RT_038063.1 (corrected) IbCBP3

ccaacaacataacaacaaacaaaaacctcaaattaaccaaatccgaagaacataatgagtgtggaaatccttgatggtgctaccattcttaac

ttcgttcaagacgaaaaggtgtttgatgaattcgtgcacgaacgtttcgacaacctggatagcaaccacgatggagttctctcctatgcggag

ctgttgagagagctgcggcgcctaagagttttggagatgcattttggtgtagatgtggagaccaatccaaccgagcttaaccgagtttatgact

caatgtttgttcagttcgatagggactcaagtggggcgatagacgtgaaggagttcaaagaagagatgaagaatatgatggtggcgatggc

gaatgatattgggttcttgccggttcaaatgctgctcgaggaaaatagtttcttgaagaaagctgttgagaaggagacaacaaaaataattgct

aatgctgcctaattaacaacttaa

>Contig13109.2 IbCBP1 (from cap3p90 assembly)

gatcagttcaaacctcccaacaacgtaacaacaaacaaaaacctcaaaccaaaaaatgagcgtggaaatcctcgacggcgctactatcctt

aatttcgtccaagacgaaaaagcctttgatgaattcgtgcacgagcgtttcgacaacctggacagcaaccacgatggagttctctcctatgcg

gagctgctgaaagagttgcgaagtctgagggttttggagatgcatttcggcgtagatgtggagaccgatccggtcgagcttaacttggtttat

gactcgatgtttgttcagtttgatagggactcaagtggggtagtggatgtggaggagttcaaggcagagacgaagaatatgatggtggctat

ggcgaatgacatagggtttttgccagttcaaatgctgcttgaggaacatagcttcttgaagaaagcggttgagaatgagttagcaaaagttgct

aatgctgcataatgattaagttagccgttgagaatgagaacaaattaattagatttttgtgttgcatgtctgaaggaaatttctagtcagggaaag

agaatattgtactattgaatataaaagaaaaaatcttggggtatgcttttgaattatatatatggaaaaacttgtttgagacagtctacaaaat

>RT_027290.1 IbCBP1-like

tggagttctctcctatgcggagctgctgaaagagttgcgaagtctgagggttttggagatgcatttcggcgtagatgtggagaccgatccggt

cgagcttaacttggtttatgactcgatgtttgttcagtttgatagggactcaagtggggtagtggatgtggaggagttcaaggcagagacgaa

gaatatgatggtggctatggcgaatgacatagggtttttgccagttcaaatgctgcttgaggaacatagcttcttgaagaaagcggttgagaat

gagttagcaaaagttgctaatgctgcataatgattaagttagccgttgagaatgagaacaaattaattagatttttgtgttgcatgtctgaaggaa

atttctagtcagggaagagaatattgtactattag

>RT_055908.1 IbCBP1-like

atcagttcaaacctcccaacaacgtaacaacaaacaaaaacctcaaaccaaaaaatgagcgtggaaatcctcgacggcgctactatccttaa

tttcgtccaagacgaaaaagcctttgatgaattcgtgcacgagcgtttcgacaacctggacagcaaccacgatggagttctctcctatgcgga

gctgctgaaagagttgcgaagtctgagggttttggagatgcatttcggcgtagatgtggagaccgatccggtcgagcttaacttggtttatga

ctcgatgtttgttcagtttgatagggactcaagtggggtagtggatgtggaggagttcaaggcagagacgaagaatatgatggtggctatgg

cgaatgacatagggtttttgccagttcaaatgctgcttgaggaacatagcttc

>S_PBL_c13311 IbCBP1-like

caaacctcccaacaacgtaacaacaaacaaaaacctcaaaccaaaaaatgagcgtggaaatcctcgacgctgctactattcttaatttcatcc

aagacgaaaaagccttcgatgaattcgtgcacgagcgattcgacaatctcgacatcaaccacgacggagttctttcctatgcggagcttctga

aagagttgcggagtctgaggattttggagatgcatttcggcgtagatgtggagaccgatccgatcgaacttaacttggtttatgactcgatgttt

gttcagttcgatagggactcaagcggggtagtggatgtggaggagtttaaggcagagacgaagaatatgatggtggctatggcgaatgac

atagggttcttgccagttcaaatgctgcttgaggaacatagcttcttgaagaaagctgttgagaagtgggttagcacaaagttgactaatgctg

cataatgattaagttagccgttgagaatgagaacaaattaattagatttttgtgttgtatgtctgaaggaaatttcttgtcagggaaagagaatatt

gtactattgaatataaaaagaaaaaatcttagggtatkcttttsaattatatatatggaaaaacttgtttgagacagtctacaaaataaaaaaaaaa

aaaanaaaaaaaaaa

Appendix D: CDPK-like sequences

>S_PBL_c3183 IbCDPK1 (1457 nucleotides) gi|967125|gb|AAC49405.1| calcium dependent protein kinase

gagttgtttgaccgcattattcagcgagggcattatagtgagagaaaggctgctgaattgactaagataattgttggagttgttgaggcatgcc
actcccttggggttatgcatagagatctcaaaccagagaatttcttattggttaacaaggatgatgatttctctctcaaagcaattgattttggact
ctcagtctttttcaaaccaggccaaatatttactgatgtcgttggaagcccgtattatgttgctcctgaggtgcttttgaagcattatggtccagaa
gcagatgtttggacagcaggggtcatactctacatattgctaagtggtgtacctccttttgggctgaaacacagcaggggatttttgatgctgt
gttgaaagggcacattgattttgagtcagacccttggcccttaatatcagacagtgcaaaggatttaatcaggaagatgttatgcatgcagccc
tctgagcgtttaactgctcatgaagtattatgtcatccttggatttgtgaaaatggtgttgctcctgatagagcgttagatccagccgtactctctc
gccttaagcagttttctgcaatgaacaagttgaagaagatggccttgcgggtaattgctgaaagcttgtcagaagaggagattgctggtttaag
ggagatgtttaaggccatggatactgataatagtggtgcaattacatttgatgagctacaaagctggtttaaagaaaatatggatctactttraag
gatacagagatacgagaactaatggatgcagctgatgtagataacagtggaactattgattatggagaattcatagcagctactattcacctta
acaaactagaacgggaggaacatcttatggctgcgtttcaatattttgacaaggatggaagtggatatattacagttgatgagctccagcaag
cttgtatagaacataatatcacagatgttctctatgaagatattatccgagaagttgatcaggataatgatggaaggatagattatggggaatttg
ttgccatgatgacaaaaggaaatgcaggtattggaagacgaactatgagaaacagtctgaatatcagcatgagagatgtacctggagctca
ctagctttttgtgagatgcatcatctagaggttgtagagaaatgacagcagatgatgatgagtgattgattgattagatgagaagaaacttgtcc
ttccttccttccttcccaagtactgaaggcagcagcaaaatcagagttgtgagggtgaagcatgcatatgcagtaaatctcttgcttagtgtcatt
gggtgcccacagatagtgttgtaagctttgtagactcaagtgtgacaaaatgatttcagttactggctatggctattatttgttttgctagttattcgt
tggggttgtggttactggaagtgggaggg

>S_PBL_c432 IbCDPK1-like (2289 nucleotides), matches gi|223540210|gb|EEF41784.1| calcium-dependent protein kinase, putative [Ricinus communis] ; calcium-dependent protein kinase isoform 6

ggcagcagcaccctccacctctggttgcctctaattataaataaattcacaactaatttacttcattaattaattgtgaaaatttcaaatatcatccat

ttactcacctcctcctcgccatgcttgtcgtcgtctcgccgtcgtcggtctccagcgcgatcatcacataacagagctctatgctcagggatttg

atccttccataccccgcttaattttcattcttctcatacaaaaccttggatgacgaaagaattgccagttgtcaagttgctgagctgagagggg

gacttgtgattgtttcatgggcaacacatgccgtggatctttcggaagcaagcattttcagggctacacacagcccgaagatcattccatttcc

aactgtaacccttcttccggccgttctaactcttactctgataatttctcctcccctactaggctgagccaagattttcccaaggatacccagccc

aaccccaacgcagaccaacgcatccctatactttatcccaaaaaggaaaacatgtatcgttccaccaataaccaggcctatcacgtcctaggt

cacaagaccgctaacatccgtgatctttacacactgggccgtaagttaggtcaaggtcaatttggtaccacttatctctgcactgagatttccac

tggtgttgactatgcttgtaagtctatttcaaaaagaaagttgatctccaaagaagatgtagaggatgtgcagagggaaattcagataatgcac

catttggctggtcacaagaacatagttactatcaaggggcatatgaagacccttatatgttcatattgtgatggaactctgtagtggtggaga

gttgtttgaccgcattattcagcgagggcattatagtgagagaaaggctgctgaattgactaagataattgttggagttgttgaggcatgccact

cccttggggttatgcatagagatctcaaaccagaaaatttcttattggttaacaaggatgatgatttctcacttaaagcaattgattttggactctc

agttttttcaatccaggccaaattttcactgatgttgttggaagcccatattatgttgctcctgaggtgcttatgaagaattacggtccagaagca

gatgtgtggacagcaggggtcatactttatatactgttaagcggtgtgccgccctttgggctgaaacacagcaagggatattcgatgcagttt

tgaaaggtgacattgattttgagtcagatccttggcccataatttcaaacagtgcaaaggatctaatccggaagatgttgtgcatgcagccttct

gagcgcttaactgctcatgaagtattatgccacccttggatttgtgagaatggtgttgctcctgatagagcattggatccagctgtactttctcgc

cttaaacagttttcagctatgaacaagttgaagaagatggctttacgggtaatagcagaaagcctatcagaagaggagattgctggtttaagg

gagatgtttaaggccatggatactgataatagtggtgcaattacatttgatgaactcaaagctggtttaagaaaatatgggtctactttaaaggat

acagaaatacgagaactaatggatgcggctgatgtggataacagtggaaccatagattatggagaattcgtagcagcgacaattcaccttaa

caaattagaacgtgaagaacatcttgttgcagcatttcaatattttgacaaggatggaagtggttatattacagttgacgagctccagcaagctt

gtgtagaacataacatgacagatgttttctttgaagatataatcagagaagtggatcaagataatgatgggcggattgattatggggagtttgtt

gctatgatgacaaaaggaaatgcaggtattggaagacgaactatgcgaaacagtgtgaatatcagtatgagagatgccccaggagctcatt

agcttttgaaaagacacgtcatctatagggtgtacaggattgatggcagatgatgatgcagcagagcttttgtatgcacgcatgctttcccttga

gttgccatatgcctgcagaaattgttgtaaactttgtagactccatccagtgtatagaaagtaattgtttcagtcttacttttaaggtattaaattgtc

accttctgcagcattgaagatattgacctttaaattattttgactacgctactaactttaaaagtttaaaaaaaaaaaaaaaaaaaaanaaaaaa
aaaaaaa

>Contig9616.1 IbCDPK1 - matches CDPK4_SOLTU Calcium-dependent protein kinase 4 OS=Solanum tuberosum GN=CPK4 PE=2 SV=1 (contains DV037296.1, IbCDPK1, found upregulated in storage root by McGregor(2006)) (from cap3p90 assembly)

ggcagcagcaccctccacctctggttgcctctaattataaataaattcacaactaatttacttcattaattaattgtgaaaatttcaaatatcatccat
ttactcacctcctcctcgccatgcttgtcgtcgtctcgccgtcgtcggtctccagcgcgatcatcacataacagagctctatgctcagggatttg
atccttccataccccgcttaattttcattcttctcatacaaaaccttggatgacgaaagaattgccagttgtcaagttgctgagctgagagggg
gacttgtgattgtttcatgggcaacacatgccgtggatctttcggaagcaagcattttcagggctacacacagcccgaagatcattccatttcc
aactgtaacccttcttccggccgttctaactcttactctgataatttctcctcccctactaggctgagccaagattttcccaaggatacccagccc
aaccccaacgcagaccaacgcatccctatactttatcccaaaaaggaaaacatgtatcgttccaccaataaccaggcctatcacgtcctaggt
cacaagaccgctaacatccgtgatctttacacactgggccgtaagttaggtcaaggtcaatttggtaccacttatctctgcactgagatttccac
tggtgttgactatgcttgtaagtctatttcaaaaagaaagttgatctccaaagaagatgtagaggatgtgcagagggaaattcagataatgcac
catttggctggtcacaagaacatagttactatcaaggggcatatgaagacccttatatgttcatattgtgatggaactctgtagtggtggaga
gttgtttgaccgcattattcagcgagggcattatagtgagagaaaggctgctgaattgactaagataattgttggagttgttgaggcatgccact
cccttggggttatgcatagagatctcaaaccagagaatttcttattggttaacaaggatgatgatttctctctcaaagcaattgattttggactctc
agtctttttcaaaccaggccaaatatttactgatgtcgttggaagcccgtattatgttgctcctgaggtgcttttgaagcattatggtccagaagca
gatgtttggacagcaggggtcatactctacatattgctaagtggtgtacctccttttgggctgaaacacagcaggggatttttgatgctgtgttg
aaagggcacattgattttgagtcagacccttggcccttaatatcagacagtgcaaaggatttaatcaggaagatgttatgcatgcagccctctg
agcgtttaactgctcatgaagtattatgtcatccttggatttgtgaaaatggtgttgctcctgatagagcgttagatccagccgtactctctcgcct
taagcagttttctgcaatgaacaagttgaagaagatggccttgcgggtaattgctgaaagcttgtcagaagaggagattgctggtttaaggga
gatgtttaaggccatggatactgataatagtggtgcaattacatttgatgagctacaaagctggtttaagaaaatatggatctactttggaaggat
acagagatacgagaactaatggatgcagctgatgtagataacagtggaactattgattatggagaa

>Contig6182.2 IbCDPK1- matches calmodulin-domain protein kinase 5 (contains DV037296.1, IbCDPK1, found upregulated in storage root by McGregor (2006)) (from cap2p80 assembly)

agcagcaccctccacctctggttgcctctaattataaataaattcacaactaatttacttcattaattaattgtgaaaatttcaaatatcatccatttac
tcacctcctcctcgccatgcttgtcgtcgtctcgccgtcgtcggtctccagcgcgatcatcacataacagagctctatgctcagggatttgatcc
ttccataccccgcttaattttcattcttctcatacaaaaccttggatgacgaaagaattgccagttgtcaagttgctgagctgagaggggactt
gtgattgtttcatgggcaacacatgccgtggatctttcggaagcaagcattttcagggctacacacagcccgaagatcattccatttccaactg
taacccttcttccggccgttctaactcttactctgataatttctcctcccctactaggctgagccaagattttcccaaggatacccagcccaaccc
caacgcagaccaacgcatccctatactttatcccaaaaaggaaaacatgtatcgttccaccaataaccaggcctatcacgtcctaggtcacaa
gaccgctaacatccgtgatctttacacactgggccgtaagttaggtcaaggtcaatttggtaccacttatctctgcactgagatttccactggtgt
tgactatgcttgtaagtctatttcaaaaagaaagttgatctccaaagaagatgtagaggatgtgcagagggaaattcagataatgcaccatttg
gctggtcacaagaacatagttactatcaaggggcatatgaagacccttatatgttcatattgtgatggaactctgtagtggtggagagttgttt
gaccgcattattcagcgagggcattatagtgagagaaaggctgctgaattgactaagataattgttggagttgttgaggcatgccactcccttg
gggttatgcatagagatctcaaaccagagaatttcttattggttaacaaggatgatgatttctctctcaaagcaattgattttggactctcagtcttt
ttcaaaccaggccaaatatttactgatgtcgttggaagcccgtattatgttgctcctgaggtgcttttgaagcattatggtccagaagcagatgtt
tggacagcaggggtcatactctacatattgctaagtggtgtacctccttttgggctgaaacacagcaggggatttttgatgctgtgttgaaagg
gcacattgattttgagtcagacccttggcccttaatatcagacagtgcaaaggatttaatcaggaagatgttatgcatgcagccctctgagcgtt
taactgctcatgaagtattatgtcatccttggatttgtgaaaatggtgttgctcctgatagagcgttagatccagccgtactctctcgccttaagca
gttttctgcaatgaacaagttgaagaagatggccttgcgggtaattgctgaaagcttgtcagaagaggagattgctggtttaagggagatgttt
aaggccatggatactgataatagtggtgcaattacatttgatgagctacaaagctggtttaagaaaatatggatctactttggaaggatacaga
gatacgagaactaatggatgcagctgatgtagataacagtggaactattgattatggagaa

>Contig26263.2-IbCDPK2 (from cap3p90 assembly)

aagggtacctttgaggatgctctgtatgtacacatagtcatggagctctgcgccggtggggagctttttgataggattgtggagaaggggcat
tatagtgagagggaggctgctaagctgcttaagactattgttggggttgtggaggcttgtcattccttgggggtcatgcatagagatctcaagc
cggagaacttttgtgccttagcactgatgaggatgctactcttaaggccattgattttggcctttctgttttctacaaaccaggtgaaatattttcg

gatgtagttggaagtccttactatgtagcacctgaggttctgcgtaagcagtatggacctgaatcagatgtatggagtgctggaattatcttgta

catactactgagtggagttccgccatttttgggcagaaactgaggtggggatattccgccagatattgaaagagaaactagatcttgaatcaga

gccatggccggggatttcagatagtgccaaggatttgatatgcaaaatgcttgatagggatccagagaagagattgacagcccatgaagtttt

gtgccatccttggattgtggatgacaaaatggcacccgataagcctcttggttctgcagttctttcacgtctgaaacagttctctgcaatgaataa

actcaagaagatggctttacgtgtaattgctgagaggctatcagaagaagaaattggtggtctcaaggagctcttcagaatgatagatacaga

caatagtggaactataacctttgatgagctgaaagaaggtttaagacgagttggatctgaacttatggagtctgagatcaaggatcttatggat

gctgcggatgttgacaacagcgggacaatagactatggtgaatttcttgctgctactgtacacctgaacaagttggaaagagaggaaaatct

actatcggccttctctttctttgacaaagatggtagtggttacataaccattgatgaacttcagcatgcctgcaaagaatttggtctaagcgagct

caatcttgatgaaatgatcaaagaaattgatcaagataatgatgggcaaatagactatggtgaatttgcagcaatgatgaggaaaggcaatgg

aggcggagccgttggaaggaaaaccatgagaaacactttaaatttgggagaggcgctaggacttgtagacaataatgatgaccattgattct

ctaaacaaaacgaattttgatttgtcatctatttccaaccttaatggtacttgtaactcgagggctcacattcattgtgcaactaattgctctctttcc

ctgtttgatgactgcatctgatgtatattctgaatcttgaatgaactaacccctcttccccatattccaattttgtggagatctttt

>Contig34516.2 IbCDPK3 (from cap3p90 assembly)

taggatgcgctgtatgtgcacatagtcatggagctttgcgcgggcggggagttgtttgataggattgtggagaaagggcaatacagcgaga

gagaggctgctaagctaatcaagaccattgttggggtcgtcgaggcttgccactctttgggggtaatgcatagagatctcaagcctgagaac

ttcttgttcctttgctccaacgaggatgctgctctcaaggccactgattttggtctttctgttttctataaacctggggaaacattctctgatgtagtt

ggaagcccctactatgtggcaccagaggttctgtgcaagcattatggacctgaatcagatgtatggagtgctggggttatcttgtacatattact

aagtggtgtccctcctttctgggcagaaactgatgtggggattttccgccagatattgcaagggaaactagatttggaatctgaaccatggcct

ggaatctcagatagtgccaaagatttgattcgtaaaatgcttgatcggaacccgaaaaggaggttgacagcccatgaagtcttatgccatcctt

ggattgtggatgacagtatggcccccgacaagcctcttgactctgcagttctttcacgcttgaaacagttctctgcaatgaataagctcaaaaa

gatggctttacgtgtgatcgctgagaggctatctgaagaggagattggtggtctgaagcagctgttcaaaatgattgatacagacaacagtgg

aactataacctttgatgagctgaaagagggattaagacgagttggatctgaattaatggaatcggagatcaaggatcttatggatgctgcaga

cattgacaacagcgggacaattgactacggagagtttcttgctgctaccgtacacttgaacaagctagaaagagaagaaaatctaatgtctgc

cttctctttctttgatagagatagtagtggttacataaccattgatgagcttcagcaagcgtgcaaagactttggtctaagcgagcttaaccttgat

gaaatgatcagagaaattgatcaagataacgacgggcagatagactacggagaatttgcagcaatgatgaggaatggcaatgggggtgtc

ggaagaagaaccatgcgaaacacactaaatttgagagaagccttgggacaacttgtaatagatgagaagcatgaggaggaggagagtaat

gaatgatagtctctgacaaactaccctaccaaaacaaattttcagatagtttccaagttcatggagttgagactagaaggtttaagctttcctgtt

gagggagggcaggctctcttcattttcttgctttttggtatgggctgctactctcctcttctatatattctcaatcttcaatactttgccatgttccattt

gtatgggaattttgtggatatgattgcctatttgccttgtactcttgtcttattgtttaatgactctccactgatttcagcctaagaaaaaaaaac

Appendix E: List of genes and its probes used as reference in the custom array.

Accession(Gene)	Gene Annotation	Probeid	Probe Sequence
AB283028.1(Ibkn2)	Class-I knotted1-like homeobox protein IBKN2	Ibkn2_AB283028.1probe1	acggatcgaatcgcagctaaacatgttaagcgacggtccggtccggatcttcaactccga
AB283028.1(Ibkn2)	Class-I knotted1-like homeobox protein IBKN2	Ibkn2_AB283028.1probe2	acccccgagttacgggagaaccggcggcgaaggcaaccacgtggctaccggttttcatct
AB283028.1(Ibkn2)	Class-I knotted1-like homeobox protein IBKN2	Ibkn2_AB283028.1probe3	atctaccgcagccgttcccggcgcggccagtagccgccgctaacgactccatctccatag
AB369253.1(Ibkn4)	Class-I knotted1-like homeobox protein IBKN4	Ibkn4_AB369253.1probe1	actgaagagcggcagcttaagaatacactgttgcgcaagtatggtagccatttgagtagc
DQ195758.1	sporamin B	DQ195758.1_control1	gtcgtcatgccgtcgacgttccagaccttcagattcaacattgcgaccaacaaactctgc
DQ195765.1	sporamin A	DQ195765.1_control1	acgtgaactggggaatccagcacgacagcgcgtccgggcagtatttcctgaaagccggcg
Contig1288.1	sucrose synthase; F8WKW 6_9 GENT UDP-glucose glucosyl-transferase OS=Gardenia jasminoides	Contig12885.1_control1	atggcttctggaagtatgtctcgaagctcgagaggcgcgagactaggcgctacctagaga
Contig2826.1	Probable glutathione S-transferase	Contig28268.1_control1	agaagttccttggaggggagacgttcggacttgcagatttagccgccagtttcatcgcct
D12882.1	beta-amylase gene	D12882.1_control1	tggcgggaagagacggttatcggcctattgccaggatgctggcaaggcaccatgccactc
D12882.1	beta-amylase gene	D12882.1_control1	tggcgggaagagacggttatcggcctattgccaggatgctggcaaggcaccatgccactc
EE875156.1	cysteine proteinase inhibitor(Cystatin)	EE875156.1_control1	agccttgtcgggcgcccaagtgaacgccctaggaaggaaggtaggcgggagaacagagat
EE878453.1	starch phosphorylase (1,4-alpha-glucan phosphorylase)	EE878453.1_control1	cagcagtgatagaacaatccatgaatatgccaaggacatatggaacatccagccagttgt
EE879297.1	UBQ3 (POLYUBIQUITIN 3); protein binding	EE879297.1_control1	aggagggaatccccccataccaacagaggttgatcgttgccggtaagcacctggaggatg
EU784152.1	Escherichia coli hygromycin B phosphotransferase	EU784152.1_control1	acagggtgtcacgttgcaagacctgcctgaaaccgaactgcccgctgttctgcagccggt
Contig35258.1 (IbEF1a)	Elongation factor 1-a(IbEF1a)	IbEF1a_Contig35258.1probe1 _control1	ctctgagtacccaccacttggtcgttttgctgtgagggacatgagacaaacggttgcagt
NM_003140.1	Homo sapiens sex determining region Y	NM_003140.1_control1	Agtgaagcgacccatgaacgcattcatcgtgtggtctcgcgatcagaggcgcaagatggc
Contig14376.1	Q8VXD1_TOBAC Alpha-tubulin OS=Nicotiana tabacum	IbTUBULIN_Contig14376.1probe1	tcgtatgaaattggcacctagccatttttcttttttatttttatgcgcctcatgtttcagc
Contig14376.1	Q8VXD1_TOBAC Alpha-tubulin OS=Nicotiana tabacum	IbTUBULIN_Contig14376.1probe2	Cgccaaggtgcagagggctgtttgcatgatctccaactccaccagtgttgctgaggtgtt

(Appendix E cont.)

Accession(Gene)	Gene Annotation	Probeid	Probe Sequence
Contig14376.1	Q8VXD1_TOBAC Alpha-tubulin OS=Nicotiana tabacum	IbTUBULIN_Contig14376.1 probe3	tgcccaaggacgtgaacgctgctgtggctaccatcaagaccaagcgtaccatccagtttg
EE880927.1(IbBEL1)	AT5G41410.1 POX (plant homeobox) family protein	IbBEL1_EE880927.1probe1	ttgctccatgatttggccggttagagtatcaaccaattaaacaacttcgatcaaaaccta
EE880927.1(IbBEL1)	AT5G41410.1 POX (plant homeobox) family protein	IbBEL1_EE880927.1probe2	cccttattcacttctggacagtgaaggtcagaacttgccatataggaacttgatgggtgc
EE880927.1(IbBEL1)	AT5G41410.1 POX (plant homeobox) family protein	IbBEL1_EE880927.1probe3	ggcggcgagttcatcgtctctcttctaccctagagaccaaattgaagattgccaaacagt
Contig22795.1(IbBEL7)	AT5G02030(BLH9,HB-6) POX (plant homeobox) family protein;BEL1-LIKE HOMEODOMAIN 9	IbBEL3_probe1_Contig22795.1	agggttaggctatggaagcctatggtggaagagattcacaatctggaaacgcggcaagcg
Contig22795.1(IbBEL7)	AT5G02030(BLH9,HB-6) POX (plant homeobox) family protein;BEL1-LIKE HOMEODOMAIN 9	IbBEL3_probe1_Contig22795.1 _control1	agggttaggctatggaagcctatggtggaagagattcacaatctggaaacgcggcaagcg
Contig22795.1(IbBEL7)	AT5G02030(BLH9,HB-6) POX (plant homeobox) family protein;BEL1-LIKE HOMEODOMAIN 9	IbBEL3_probe1_Contig22795.1 _ control2	agggttaggctatggaagcctatggtggaagagattcacaatctggaaacgcggcaagcg
Contig22795.1(IbBEL7)	AT5G02030(BLH9,HB-6) POX (plant homeobox) family protein;BEL1-LIKE HOMEODOMAIN 9	IbBEL3_probe1_Contig22795.1 _control3	agggttaggctatggaagcctatggtggaagagattcacaatctggaaacgcggcaagcg
Contig22795.1(IbBEL7)	AT5G02030(BLH9,HB-6) POX (plant homeobox) family protein;BEL1-LIKE HOMEODOMAIN 9	IbBEL3_probe2_Contig22795.1	agcagcgtctccttaaccctaggccttcaccaaaacaacgagcttggcttgtcggattct
EE881365.1(IbBEL6)	AT2G23760(BLH4) BEL1-like homeodomain 4;BLH4 (SAWTOOTH 2); DNA binding	IbBEL6_EE881365.1probe1	ggcggcaatgtgtcgctcacgctggggctgcgccacgccggaaatttgcccgaccagaat
EE881365.1(IbBEL6)	AT2G23760(BLH4) BEL1-like homeodomain 4;BLH4 (SAWTOOTH 2); DNA binding	IbBEL6_EE881365.1probe2	gcccacgaaagcgacccttcactactcgcaatcaatacccaatacacgtgtcaccgccgt
EE881365.1(IbBEL6)	AT2G23760(BLH4) BEL1-like homeodomain 4;BLH4 (SAWTOOTH 2); DNA binding	IbBEL6_EE881365.1probe3	atcaacgctgcccacgaaagcgacccttcactactcgcaatcaatacccaatacacgtgt

(Appendix E cont.)

Accession(Gene)	Gene Annotation	Probeid	Probe Sequence
Contig39444.1(IbHLH1)	AT1G72210 (bHLH096) basic helix-loop-helix (bHLH) DNA-binding superfamily protein	IbBHLH_Contig39444.1probe1	gcggcgctccaagcgagttgtcttccggatatcaactggcggctgcaggagcggcggaca
Contig39444.1(IbHLH1)	AT1G72210 (bHLH096) basic helix-loop-helix (bHLH) DNA-binding superfamily protein	IbBHLH_Contig39444.1probe2	cgcggagttcttcaacatcccacagtactccacgggcattacagggaatgttgaaagcct
Contig39444.1(IbHLH1)	AT1G72210 (bHLH096) basic helix-loop-helix (bHLH) DNA-binding superfamily protein	IbBHLH_Contig39444.1probe3	aagtggcatctgtggatgaaatagcagctgcagtgaatcagattttaggtaggattcacc
S_PBL_c31897(IbBME)	AT3G54810.1(BME3,BME3-ZF) Plant-specific GATA-type zinc finger transcription factor family protein	IbBME_S_PBL_c31897probe2	acggctattcccagagtaccgaccagcagctagtccgacgtttgttccaacgctacactc
Contig23250.2	AT2G33990.1 iqd9 (IQEdomain 9); calmodulin_binding family protein	IbCBP2_Contig23250.2probe1	accccgcagactatcttatgaatcgggtgtaaaagtagttcttaataaagggaatgagaa
Contig23250.2	AT2G33990.1 iqd9 (IQEdomain 9); calmodulin_binding family protein	IbCBP2_Contig23250.2probe2	ccagcccgatatatggatcgagtaacctcgaactgggaaaagcaaactggggctggagct
Contig22153.1	AT4G09570.1 calcium-dependent protein kinase 4	IbCDPK2_Contig22153.1probe1	cacattcattgtgcaactaattgctctctttccctgtttgatgactgcatctgatgtata
S_PBL_c1655 (IbCKR)	Cyclin-dependent kinases regulatory subunit, putative [Ricinus communis]	IbCKR_S_PBL_c1655probe1	aaaatgggtcagatccagtattccgagaagtacttcgatgatacctacgagtacaggcat
S_PBL_c1655 (IbCKR)	Cyclin-dependent kinases regulatory subunit, putative [Ricinus communis]	IbCKR_S_PBL_c1655probe2	tactatcagactatcagctacttcaccggtcccgatttctctcgctgcgaaatctgcaac
S_PBL_c3693(IbCRF1)	AT4G11140(CRF1) cytokinin response factor 1	IbCRF1_S_PBL_c3693probe1	caccttctattttcgggcgtaaatgcgtaatgcgaccgtgacgtggcacttttggcggca
S_PBL_c5300(IbCRF2)	AT4G27950.1(CRF4) cytokinin response factor 4	IbCRF4_S_PBL_c5300probe1	ggacgtgaacgtcgacgattatttccaagactgcggcgacttttccggcatcgatgccgt
S_PBL_c5300(IbCRF2)	AT4G27950.1(CRF4) cytokinin response factor 4	IbCRF4_S_PBL_c5300probe2	gagattctcgctaacgctacttccaactccggctacgactccgccgatgaatcccggaat
S_PBL_c5300(IbCRF2)	AT4G27950.1(CRF4) cytokinin response factor 4	IbCRF4_S_PBL_c5300probe3	ccccggagattctcgctaacgctacttccaactccggctacgactccgccgatgaatccc
AB478416.1	*Ipomoea batatas* CycD3;1 mRNA for cyclin D3, complete cds	IbCYCD3;2_AB478416.1probe1	atgcatatttcacctctgatagctccaatgattcttgggcagtttgtttctccccccaca

(Appendix E cont.)

Accession(Gene)	Gene Annotation	Probeid	Probe Sequence
AB478416.1	*Ipomoea batatas* CycD3;1 mRNA for cyclin D3, complete cds	IbCYCD3;2_AB478416.1probe2	aacggcgtaattgatgcatatttcacctctgatagctccaatgattcttgggcagtttgt
Contig3374.1(IbHB1)	IbHB1_Contig3374.1probe1	IbHB1_Contig3374.1probe1	agacgccggcggcgcagccattttacatgcagatttcagccgctacgccgctcaccatgt
Contig401.1(IbHB2)	AT2G46680(ATHB7,HB-7) homeobox 7	IbHB2_Contig401.1probe1	tgttgcaatagtggcctccagtctccactacctattctatctgcatcgatcatagtccat
Contig401.1(IbHB2)	AT2G46680(ATHB7,HB-7) homeobox 7	IbHB2_Contig401.1probe2	agatccaagtcgaagcagatcgagcatgattacaggaagctgaagatgagttatgatgct
S_PBL_c15425(IbIAA)	AT4G29080 (PAP2, IAA27) phytochrome-associated protein 2; Auxin-responsive Aux/IAA family protein	IbIAA_S_PBL_c15425probe1	agactcttgcaagcgtctaaggatcatgaagagttcagatgctgttggtctagctcccag
S_PBL_c15425(IbIAA)	AT4G29080 (PAP2, IAA27) phytochrome-associated protein 2; Auxin-responsive Aux/IAA family protein	IbIAA_S_PBL_c15425probe2	ctccgggtgttcctatccgggatggattgagtgaaagtagattgatggatcttcatggtt
S_PBL_c11096(IbMAP65)	AT1G14690.2 microtubule-associated protein 65-7	IbMAP65_S_PBL_c11096probe1	ttgcaatatcatcagtagcttagttactgtagtagagaatacgtatttgaacaagggact
S_PBL_c11096(IbMAP65)	AT1G14690.2 microtubule-associated protein 65-7	IbMAP65_S_PBL_c11096probe2	aggtggtgggtccccgaacaaatgggagtggcaatggaaccgccaacaggaggttgtctc
S_PBL_c11096(IbMAP65)	AT1G14690.2 microtubule-associated protein 65-7	IbMAP65_S_PBL_c11096probe3	aacaaagagcaagatggagtgtttggttctacaccaagccccgctcgacaactgggcaca
S_PBL_c19033(IbNAM)	Q8LRL4_PETHY Nam-like protein 11 Petunia hybrida NH11	IbNAM_S_PBL_c19033probe1	aagctacattggcttttatggcagcgaagacgcatcagggcaaggttatgaccagagcat
Contig2072.2	AT2G12400.1 unknown protein;	IbTAP_Contig2072.2probe1	tggcttgtgcgttacacccgggcgactaacaccagatttctacagccagatgatggcctc
Contig2072.2	AT2G12400.1 unknown protein;	IbTAP_Contig2072.2probe2	agcttgcagtcccggtgaagtggatttcaacaacgcaacacaggtgtggagtaagtatgt
EF192419.1 (IbTIP)	*tonoplast instrinsic protein* (*TIP*)	IbTIP_EF192419.1probe3	ttctgcggcaatagcccacgcttttgcactcttcgtcgcagtttccgtcggcgccgacat
S_PBL_c3428(IbXX)	XP_002510084.1 conserved hypothetical protein [Ricinus communis]	IbXX_S_PBL_c3428probe3	agcctcggttgaactttcaagtagtagacctgcggttctgtaattagtaatggaattaaa
S_PBL_c3428(IbXX)	XP_002510084.1 conserved hypothetical protein [Ricinus communis]	IbXX_S_PBL_c3428probe3	agcctcggttgaactttcaagtagtagacctgcggttctgtaattagtaatggaattaaa

Appendix F: List of significant upregulated genes in storage roots from the microarray results (FDR<0.05) and fold change ratio >1.48. FC=Fold change, P=biological process, F=molecular function, C=cell component, and NA=no data available.

Unigene	FC	Annotation	Probeid	GO group
S_PBL_c3576	12.61	NoMatchingHitfound	S_PBL_c3576	NA
S_PBL_c2430	7.61	B9NDK1 B9NDK1_POPTR Predicted protein OS=Populus trichocarpa GN=POPTRDRAFT_789085	S_PBL_c2430	NA
Contig33633.1	7.25	Q93X99_IPOBA Proteinase inhibitor	CUST_76135_PI427086615	F
Contig14347.1	6.55	B9R8R1_RICCO Glucan endo-1,3-beta-glucosidase, putative	CUST_45040_PI427086615	P
Contig056161.3	6.34	A9ZSX7_9ASTE Alcohol dehydrogenase (Fragment)	CUST_22924_PI427086615	P
Contig13858.1	6.06	B9RJV7_RICCO Aspartic proteinase nepenthesin-1, putative	CUST_44268_PI427086615	F
Contig28401.1	5.92	B9SSC2_RICCO DNA binding protein, putative	CUST_67623_PI427086615	P
Contig34683.1	5.79	OLP1_SOLLC Osmotin-like protein	CUST_77751_PI427086615	P
Contig16599.1	5.76	O80431_TOBAC Endo-xyloglucan transferase (EXGT)	CUST_48523_PI427086615	P
S_PBL_c3396	5.65	NoMatchingHitfound	S_PBL_c3396	NA
Contig35533.1	5.23	B3IWI0_ORYSJ QLTG-3-1 protein; Low-temperature germinability 3-1	CUST_79137_PI427086615	P
Contig35525.1	5.18	B9S3J8_RICCO Repetitive proline-rich cell wall protein 2, putative	CUST_79121_PI427086615	P
Contig045390.3	5.1	U497E_VITVI UPF0497 membrane protein 14	CUST_6323_PI427086615	C
Contig9244.1	5.03	B7P071_IPOBA Expansin	CUST_37159_PI427086615	P
Contig23824.1	4.99	AMYB_IPOBA Beta-amylase	CUST_60183_PI427086615	C
Contig14498.1	4.86	B9SSC2_RICCO DNA binding protein, putative	CUST_45261_PI427086615	F
Contig16253.1	4.84	B9RYW1_RICCO Glucan endo-1,3-beta-glucosidase, putative	CUST_47947_PI427086615	F
Contig046841.3	4.72	PRP3_SOYBN Repetitive proline-rich cell wall protein 3	CUST_8445_PI427086615	NA
Contig8425.1	4.71	PER4_VITVI Peroxidase 4	CUST_35950_PI427086615	P
Contig15245.1	4.65	Q6RX69_PETHY Expansin-1	CUST_46413_PI427086615	P
Contig16795.1	4.52	Q6R5L6_CAPAN Sadtomato protein (Fragment)	CUST_48846_PI427086615	P
Contig8829.1	4.44	D7TNT3_VITVI Non-specific lipid-transfer protein (Fragment)	CUST_36542_PI427086615	NA
Contig18155.1	4.39	B9T3K3_RICCO Protein P21, putative	CUST_51006_PI427086615	P
Contig13241.1	4.37	Q2VT58_CAPAN Proline-rich protein	CUST_43270_PI427086615	P
DQ195758.1	4.34	*Ipomoea batatas* isolate pTrip1Ex2-1 sporamin B precursor, mRNA, complete cds	DQ195758.1_control1	C
Contig35601.1	4.3	Q8W112_ARATH Beta-D-glucan exohydrolase-like protein	CUST_79231_PI427086615	NA
Contig7506.1	4.28	O24329_RICCO Putative uncharacterized protein	CUST_34536_PI427086615	NA
D12882.1	4.21	IPBAMYB *Ipomoea batatas* beta-amylase gene, complete cds	D12882.1_control1	C
Contig16338.1	4.2	Q9FIQ5_ARATH At5g46700	CUST_48091_PI427086615	NA
S_PBL_c19740	4.16	NoMatchingHitfound	S_PBL_c19740	NA

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig5724.1	4.05	Q8W1A6_PETHY Aquaporin-like protein	CUST_31637_PI427086615	P
BU691724.1	4.03	AMYB_IPOBA Beta-amylase	CUST_94023_PI427086615	C
S_PBL_c17231	4	NoMatchingHitfound	S_PBL_c17231	NA
S_PBL_lrc49266	3.95	DQ515800.1 *Ipomoea batatas* expansin mRNA, complete cds	S_PBL_lrc49266	NA
Contig4105.1	3.89	Q8W2N1_TOBAC Cytochrome P450-dependent fatty acid hydroxylase	CUST_29097_PI427086615	P
Contig19260.1	3.85	B9T1E3_RICCO Glucan endo-1,3-beta-glucosidase, putative	CUST_52686_PI427086615	P
S_PBL_c40350	3.76	NoMatchingHitfound	S_PBL_c40350	NA
Contig9836.1	3.71	A8QXP6_IPOBA Class-I knotted1-like homeobox protein IBKN2	CUST_38048_PI427086615	F
RT_091599.1	3.7	A2Q374_MEDTR Gibberellin regulated protein	CUST_82993_PI427086615	F
Contig15133.1	3.67	B3F8H6_NICLS Sucrose sythase	CUST_46239_PI427086615	P
contig16960	3.66	Q42936_NICPL Pectinesterase (Fragment)	contig16960	C
DQ195765.1	3.6	*Ipomoea batatas* isolate pTrip1Ex2-7 sporamin A precursor, mRNA, complete cds	DQ195765.1_control1	C
Contig2866.1	3.51	B3SHI0_IPOBA Anionic peroxidase swpa7	CUST_27186_PI427086615	P
S_PBL_lrc26197	3.5	XP_003611068.1 hypothetical protein MTR_5g010060 [Medicago truncatula]	S_PBL_lrc26197	NA
Contig14099.1	3.49	Q9FHM9_ARATH AT5g51550/K17N15_10	CUST_44677_PI427086615	NA
Contig18258.1	3.42	Q6RJY7_CAPAN Elicitor-inducible protein EIG-J7	CUST_51170_PI427086615	F
Contig18649.1	3.4	Q8VWL8_SOLLC Beta-mannosidase	CUST_51746_PI427086615	F
Contig30442.1	3.4	Q6K7G5_ORYSJ Os02g0779200 protein	CUST_70953_PI427086615	P
Contig33311.1	3.38	A8QXP6_IPOBA Class-I knotted1-like homeobox protein IBKN2	CUST_75588_PI427086615	C
Contig6072.1	3.38	A9PFC6_POPTR Putative uncharacterized protein	CUST_32201_PI427086615	NA
S_PBL_lrc55259	3.37	NoMatchingHitfound	S_PBL_lrc55259	NA
RT_336384.1	3.29	B9R8R1_RICCO Glucan endo-1,3-beta-glucosidase, putative	CUST_89027_PI427086615	P
Contig11521.1	3.24	XTH1_SOLLC Probable xyloglucan endotransglucosylase/hydrolase 1	CUST_40601_PI427086615	P
Contig6959.1	3.23	PER15_IPOBA Peroxidase 15	CUST_33667_PI427086615	P
Contig23250.1	3.22	Q2HPE5_CAPAN CAPIP2, involved in disease resistance and drought and salt stress tolerance.	CUST_59160_PI427086615	NA
S_PBL_c10818	3.2	NoMatchingHitfound	S_PBL_c10818	NA
Contig13544.1	3.17	B9SUI3_RICCO Putative uncharacterized protein	CUST_43764_PI427086615	NA
Contig28990.1	3.16	PHSL_IPOBA Alpha-1,4 glucan phosphorylase L isozyme, chloroplastic/amyloplastic	CUST_68590_PI427086615	C
Contig25254.1	3.13	Q9LNJ3_ARATH At1g01300	CUST_62566_PI427086615	F
Contig35839.1	3.09	Q93ZJ4_ARATH At1g76160/T23E18_10	CUST_79620_PI427086615	P
Contig14486.1	3.07	Q9MB62_IPONI Phytocyanin-related protein	CUST_45239_PI427086615	C
Contig18425.1	3.07	Q0H8W0_TOBAC Endo-beta-1,4-glucanase	CUST_51397_PI427086615	P

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig16638.1	3.06	Q9ZQS8_IPOBA Fravanone 3-hydroxyrase	CUST_48589_PI427086615	P
Contig8256.1	3.06	Q9SML5_CAPAN Knolle	CUST_35693_PI427086615	P
Contig050184.3	3.04	U497E_VITVI UPF0497 membrane protein 14	Contig050184.3	C
Contig050955.3	3.04	Q84UC3_SOLTU Sucrose synthase 2	CUST_14388_PI427086615	P
S_PBL_c46290	3.04	NoMatchingHitfound	S_PBL_c46290	NA
Contig1100.1	3.03	B6TKU2_MAIZE Protein kinase	CUST_24543_PI427086615	NA
Contig35404.1	3.03	B9RHG2_RICCO Serine-threonine protein kinase, plant-type, putative	CUST_78935_PI427086615	P
Contig35705.1	3.03	Q9SFB1_ARATH F17A17.37 protein	CUST_79393_PI427086615	NA
Contig12083.1	2.97	C8YZA8_CAPAN UPA15	CUST_41446_PI427086615	F
Contig18487.1	2.97	VATG2_TOBAC V-type proton ATPase subunit G 2	CUST_51527_PI427086615	C
Contig9874.1	2.97	BH096_ARATH Transcription factor bHLH96	CUST_38111_PI427086615	F
S_PBL_c12071	2.96	NoMatchingHitfound	S_PBL_c12071	NA
Contig24793.1	2.93	Q56R06_SOLLC Putative permease I	CUST_61794_PI427086615	P
Contig7372.1	2.92	O82151_TOBAC Beta-D-glucan exohydrolase	CUST_34327_PI427086615	F
S_PBL_c51732	2.92	NoMatchingHitfound	S_PBL_c51732	NA
Contig14935.1	2.91	B9SWY1_RICCO Pyruvate decarboxylase, putative	CUST_45908_PI427086615	P
Contig17527.1	2.89	GUN8_ARATH Endoglucanase 8	CUST_50001_PI427086615	P
Contig32758.1	2.89	Q07446_SOLLC Peroxidase	CUST_74751_PI427086615	P
RT_327455.1	2.86	C7C5S8_NICBE PME inhibitor	CUST_88753_PI427086615	C
Contig13410.1	2.84	GUN24_ARATH Endoglucanase 24	CUST_43543_PI427086615	P
S_PBL_c10460	2.82	P83241 IP23_CAPAN Proteinase inhibitor PSI-1.2 OS=Capsicum annuum PE=1	S_PBL_c10460	F
S_PBL_c37694	2.82	NoMatchingHitfound	S_PBL_c37694	NA
Contig5948.1	2.81	Q93XQ1_NICAL Cellulose synthase catalytic subunit	CUST_32000_PI427086615	P
Contig7758.1	2.79	Q93ZJ4_ARATH At1g76160/T23E18_10	CUST_34941_PI427086615	P
Contig10714.1	2.78	B9T6G8_RICCO Nodulation receptor kinase, putative	CUST_39372_PI427086615	C
Contig35329.1	2.78	B9RDW4_RICCO Polygalacturonase, putative	CUST_78773_PI427086615	P
Contig18822.1	2.77	Q6AVU2_ORYSJ Endoplasmic oxidoreductin 1, putative, expressed	CUST_51974_PI427086615	P
Contig4462.1	2.76	Q42935_NICPL Pectinesterase (Fragment)	CUST_29659_PI427086615	C
Contig13915.1	2.75	Q9FSH4_SOLLC B2-type cyclin dependent kinase	CUST_44361_PI427086615	P
Contig20514.1	2.75	B9GT61_POPTR Predicted protein	Contig20514.1	NA
RT_354503.1	2.75	B9RJV7_RICCO Aspartic proteinase nepenthesin-1, putative	CUST_89523_PI427086615	F
Contig10072.1	2.74	GP1_SOLLC Polygalacturonase-1 non-catalytic subunit beta	CUST_38396_PI427086615	F
Contig10378.1	2.74	FH1_ORYSJ Formin-like protein 1	CUST_38886_PI427086615	C
Contig23513.1	2.74	C4B8D7_TOBAC AtEB1c-like protein	CUST_59646_PI427086615	NA

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig5202.1	2.74	B9SRH5_RICCO Transcription factor, putative	CUST_30797_PI427086615	P
Contig12173.1	2.73	B9R9R5_RICCO Pectinesterase	CUST_41600_PI427086615	C
Contig26524.1	2.72	B9RYW8_RICCO Cucumisin, putative	CUST_64550_PI427086615	P
Contig11515.1	2.71	B9SH92_RICCO Polygalacturonase, putative	CUST_40591_PI427086615	F
Contig1311.1	2.65	Q769D9_IPOBA Dihydroflavonol 4-reductase	CUST_24826_PI427086615	F
Contig7096.1	2.65	B9SMX0_RICCO DNA binding protein, putative	CUST_33884_PI427086615	F
Contig26258.1	2.6	B9RLK7_RICCO Alpha-1,4-glucan-protein synthase [UDP-forming], putative	CUST_64185_PI427086615	P
Contig3387.1	2.58	B9S0F2_RICCO Xyloglucan endotransglucosylase/hydrolase protein 14, putative	CUST_27997_PI427086615	C
S_PBL_c41294	2.58	NoMatchingHitfound	S_PBL_c41294	NA
Contig39795.1	2.57	B9H8X0_POPTR 1-aminocyclopropane-1-carboxylate	CUST_80488_PI427086615	C
Contig12230.1	2.56	B9S8L8_RICCO Multicopper oxidase, putative	CUST_41686_PI427086615	P
Contig10632.1	2.55	B9GXZ8_POPTR Pectinesterase	CUST_39225_PI427086615	P
S_PBL_lrc46977	2.54	NoMatchingHitfound	S_PBL_lrc46977	NA
AB283028.1	2.53	*Ipomoea batatas* Ibkn2 mRNA for class-I knotted1-like homeobox protein IBKN2	Ibkn2_AB283028.1probe1	C
EE878453.1	2.53	starch phosphorylase	EE878453.1_control1	F
Contig5213.1	2.52	B9SH75_RICCO Xyloglucan:xyloglucosyl transferase, putative	CUST_30816_PI427086615	P
Contig5920.1	2.52	RBOHB_SOLTU Respiratory burst oxidase homolog protein B	CUST_31958_PI427086615	P
RT_344595.1	2.52	B9MTB7_POPTR Aquaporin, MIP family, PIP subfamily	CUST_89245_PI427086615	P
Contig12977.1	2.51	B9RYW1_RICCO Glucan endo-1,3-beta-glucosidase, putative	CUST_42841_PI427086615	F
Contig27409.1	2.5	Q0E7D4_COFAR Sucrose synthase	CUST_66015_PI427086615	P
Contig4451.1	2.5	Q6SA75_TOBAC ANT-like protein	CUST_29641_PI427086615	F
S_PBL_c11782	2.5	NoMatchingHitfound	S_PBL_c11782	NA
Contig043927.3	2.48	B9T135_RICCO Lipid binding protein, putative	CUST_4216_PI427086615	F
Contig17893.1	2.48	B9RXP7_RICCO Beta-glucosidase, putative	CUST_50579_PI427086615	P
Contig13041.1	2.45	B9RGK6_RICCO Alpha-galactosidase/alpha-n-acetylgalactosaminidase, putative	CUST_42941_PI427086615	P
DV037125.1	2.45	AMYB_IPOBA Beta-amylase	CUST_93874_PI427086615	C
Contig18195.1	2.44	B9S3D6_RICCO C-4 methyl sterol oxidase, putative	CUST_51070_PI427086615	P
Contig18948.1	2.42	Q6Z576_ORYSJ Os08g0411300 protein	CUST_52241_PI427086615	NA
Contig10290.1	2.41	A4UV42_SOLTU Putative HVA22 protein	CUST_38762_PI427086615	P
Contig3639.1	2.41	B9RKA0_RICCO 1-aminocyclopropane-1-carboxylate oxidase, putative	CUST_28394_PI427086615	C
Contig8631.1	2.41	B9RUF4_RICCO Peroxidase 31, putative	CUST_36247_PI427086615	P

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig051468.3	2.4	B9RB66_RICCO DNA binding protein, putative	CUST_15185_PI427086615	NA
Contig6966.1	2.4	Q948Z3_SOLTU Putative peroxidase	CUST_33675_PI427086615	P
Contig35011.1	2.39	B9R6Y4_RICCO Serine carboxypeptidase, putative	CUST_78313_PI427086615	F
Contig13772.1	2.38	FLA2_ARATH Fasciclin-like arabinogalactan protein 2	CUST_44129_PI427086615	C
Contig045839.3	2.37	Q7XYR7_GOSHI Class III peroxidase	CUST_6996_PI427086615	P
Contig1967.1	2.37	B9SKK1_RICCO Uclacyanin-2, putative	CUST_25819_PI427086615	C
Contig2313.1	2.37	B9RP67_RICCO Cellulose synthase A catalytic subunit 6 [UDP-forming], putative	CUST_26347_PI427086615	P
Contig28882.1	2.37	Q9SXH2_IPOBA 24 kDa vacuolar protein VP24	CUST_68453_PI427086615	NA
S_PBL_c19369	2.37	NoMatchingHitfound	S_PBL_c19369	NA
S_PBL_c46119	2.36	NoMatchingHitfound	S_PBL_c46119	NA
Contig18655.1	2.35	Q5F1U6_SOLTU UTP:alpha-D-glucose-1-phosphate uridylyltransferase	CUST_51758_PI427086615	P
contig28314	2.35	B9T7D5_RICCO Zeamatin, putative	contig28314	P
RT_329245.1	2.34	Q8W2N1_TOBAC Cytochrome P450-dependent fatty acid hydroxylase	CUST_88825_PI427086615	P
RT_378266.1	2.34	Q42863_IPOBA Starch phosphorylase	CUST_90003_PI427086615	F
Contig16471.1	2.33	Q6TF29_SOLCH Rapid alkalinization factor 1	CUST_48301_PI427086615	NA
Contig6736.1	2.33	Q2XTD7_SOLTU Polygalacturonase-like protein-like	CUST_33300_PI427086615	C
Contig11525.1	2.32	Q84UY3_PETHY Alcohol dehydrogenase 2	CUST_40607_PI427086615	P
Contig24107.1	2.32	PFPB_SOLTU Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit beta	CUST_60588_PI427086615	F
Contig6672.1	2.32	Q9LLT0_SOLLC Beta-galactosidase	CUST_33198_PI427086615	C
Contig048543.3	2.31	Q6YBV2_POPTM Cellulose synthase	CUST_10896_PI427086615	P
Contig9184.1	2.31	B9SRH5_RICCO Transcription factor, putative	CUST_37071_PI427086615	NA
Contig11302.1	2.3	B9SKH8_RICCO Dynamin, putative	CUST_40278_PI427086615	C
Contig25858.1	2.3	D2DGW4_SOYBN Aux/IAA protein	CUST_63543_PI427086615	P
Contig5502.1	2.3	B9RXP7_RICCO Beta-glucosidase, putative	CUST_31292_PI427086615	F
Contig11652.1	2.29	CALR_NICPL Calreticulin	CUST_40814_PI427086615	P
Contig17898.1	2.29	B9T4B7_RICCO Squamosa promoter-binding protein, putative	CUST_50589_PI427086615	F
Contig3996.1	2.29	B9S9N2_RICCO Xyloglucan endotransglucosylase/hydrolase protein 9, putative	CUST_28932_PI427086615	P
Contig9303.1	2.28	A3KCF5_IPOBA Glucose-1-phosphate adenylyltransferase	CUST_37242_PI427086615	P
Contig049930.3	2.27	C0IRG2_ACTER Xyloglucan endotransglucosylase/hydrolase 3	CUST_12897_PI427086615	C
Contig19078.1	2.27	GDL8_ARATH GDSL esterase/lipase At1g28590	CUST_52416_PI427086615	C
Contig21485.1	2.27	A4GU25_SOYBN Pectate lyase (Fragment)	CUST_56316_PI427086615	P
Contig7428.1	2.27	XTH33_ARATH Probable xyloglucan endotransglucosylase/hydrolase protein 33	CUST_34409_PI427086615	P

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig8646.1	2.27	C8YZA8_CAPAN UPA15	CUST_36266_PI427086615	P
Contig11910.1	2.26	A3KCF5_IPOBA Glucose-1-phosphate adenylyltransferase	CUST_41190_PI427086615	P
Contig13144.1	2.26	B9IDR8_POPTR Pectinesterase	CUST_43124_PI427086615	C
Contig26653.1	2.26	Q9SXH2_IPOBA 24 kDa vacuolar protein VP24	CUST_64727_PI427086615	NA
Contig11408.1	2.25	B9SCH2_RICCO ATP-binding cassette transporter, putative	CUST_40448_PI427086615	P
Contig14737.1	2.25	A8CWX0_SOLTU Apyrase 3	CUST_45618_PI427086615	P
S_PBL_c11712	2.25	NoMatchingHitfound	S_PBL_c11712	NA
Contig10092.1	2.24	B9SKV5_RICCO Cinnamoyl-CoA reductase, putative	CUST_38431_PI427086615	F
Contig20041.1	2.23	Q6IVK6_TOBAC Putative UDP-glucose dehydrogenase 2	CUST_53982_PI427086615	P
Contig20777.1	2.23	B9H498_POPTR White-brown-complex ABC transporter family	CUST_55101_PI427086615	F
Contig11508.1	2.22	Q0IY20_ORYSJ Os10g0378400 protein	CUST_40579_PI427086615	NA
Contig3702.1	2.21	Q6RHX7_SOLLC Xyloglucan endotransglucosylase-hydrolase XTH9	CUST_28490_PI427086615	P
Contig048991.3	2.2	B9RT55_RICCO Clasp, putative	CUST_11509_PI427086615	NA
Contig16970.1	2.2	B9H2G7_POPTR Nucleobase ascorbate transporter	CUST_49124_PI427086615	P
Contig17113.1	2.2	B9H7N5_POPTR Acyl-[acyl-carrier-protein] desaturase	CUST_49345_PI427086615	P
Contig24587.1	2.2	B9T755_RICCO Peptide transporter, putative	CUST_61385_PI427086615	P
Contig13711.1	2.19	B9RFA3_RICCO Respiratory burst oxidase, putative	CUST_44031_PI427086615	P
Contig5388.1	2.19	B9TA90_RICCO Thermostable beta-glucosidase B, putative	CUST_31114_PI427086615	F
S_PBL_c2144	2.19	NoMatchingHitfound	S_PBL_c2144	NA
Contig050186.3	2.17	B9T1Y2_RICCO ATP binding protein, putative	CUST_13299_PI427086615	C
Contig15686.1	2.17	B9RZX1_RICCO Polygalacturonase, putative	CUST_47071_PI427086615	P
Contig16899.1	2.17	Q3EAA4_ARATH At4g05520	CUST_49016_PI427086615	P
Contig041611.3	2.16	Q3SA35_FAGSY Putative respiratory burst oxidase (Fragment)	CUST_891_PI427086615	P
Contig11635.1	2.16	B9S8A9_RICCO Purine permease, putative	CUST_40769_PI427086615	P
Contig5206.1	2.16	B9R873_RICCO Endo-1,4-beta-glucanase, putative	CUST_30801_PI427086615	P
Contig12278.1	2.15	Q93XQ1_NICAL Cellulose synthase catalytic subunit	CUST_41772_PI427086615	P
Contig20515.1	2.15	Q9M4D2_CAPAN Chloroplast ferredoxin-NADP+ oxidoreductase	CUST_54745_PI427086615	P
Contig35483.1	2.15	B9SB23_RICCO Protein binding protein, putative	CUST_79053_PI427086615	F
Contig37761.1	2.15	B9RF23_RICCO DNA binding protein, putative	CUST_80078_PI427086615	F
RT_285034.1	2.15	E0CQN5_VITVI Acyl-[acyl-carrier-protein] desaturase (Fragment)	CUST_87450_PI427086615	P
Contig050495.3	2.14	B9RYW1_RICCO Glucan endo-1,3-beta-glucosidase, putative	CUST_13741_PI427086615	C
Contig11918.1	2.14	B9SER1_RICCO Oligopeptide transporter, putative	CUST_41202_PI427086615	P
Contig13299.1	2.14	B9RXQ4_RICCO Pectinesterase	CUST_43355_PI427086615	C
Contig9166.1	2.14	B9GSC5_POPTR Pectinesterase	CUST_37043_PI427086615	C
Contig9453.1	2.14	A2PZD5_IPONI Caffeoyl-CoA O-methyltransferase	CUST_37464_PI427086615	F

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig3355.1	2.13	PHSL_IPOBA Alpha-1,4 glucan phosphorylase L isozyme, chloroplastic/amyloplastic	CUST_27945_PI427086615	F
Contig8787.1	2.13	B9S8A9_RICCO Purine permease, putative	CUST_36481_PI427086615	P
Contig8821.1	2.13	Q94FM5_TOBAC Elicitor-inducible cytochrome P450	CUST_36529_PI427086615	P
Contig20150.1	2.12	D7TK45_VITVI Whole genome shotgun sequence of line PN40024	CUST_54140_PI427086615	NA
Contig6287.1	2.12	Q9M2R0_ARATH Anthranilate phosphoribosyltransferase-like protein	CUST_32537_PI427086615	NA
Contig6755.1	2.12	B9SD00_RICCO Aspartic proteinase nepenthesin-2, putative	CUST_33332_PI427086615	P
Contig6942.1	2.12	B9SC23_RICCO Endosomal P24A protein, putative	CUST_33642_PI427086615	C
RT_336882.1	2.12	Q10CK6_ORYSJ UDP-glucose 6-dehydrogenase, putative, expressed	CUST_89043_PI427086615	P
Contig041472.3	2.11	B9IDR8_POPTR Pectinesterase	CUST_693_PI427086615	C
Contig045030.3	2.11	B9SRH5_RICCO Transcription factor, putative	CUST_5807_PI427086615	F
Contig34935.1	2.11	Q940P5_ARATH At2g20740/F5H14.29	CUST_78198_PI427086615	NA
contig15122	2.1	B8R4D5_NICBE Subtilisin-like protein (Fragment)	contig15122	C
Contig20948.1	2.1	C6ZRV2_SOYBN Leucine-rich repeat receptor-like protein kinase	CUST_55450_PI427086615	F
Contig3720.1	2.1	B9SV36_RICCO Bel1 homeotic protein, putative	CUST_28513_PI427086615	C
Contig043907.3	2.09	E6NU40_9ROSI JHL23J11.5 protein	CUST_4185_PI427086615	NA
Contig10623.1	2.09	Q2XTC1_SOLTU Actin-like	CUST_39213_PI427086615	P
Contig22621.1	2.09	B9T6I0_RICCO Nuclear transcription factor, X-box binding, putative	CUST_58144_PI427086615	C
Contig23807.1	2.09	B9SB09_RICCO Sphingolipid fatty acid alpha hydroxylase	CUST_60161_PI427086615	P
Contig33052.1	2.09	B9T0A9_RICCO Adenosine kinase, putative	CUST_75248_PI427086615	C
S_PBL_c22205	2.09	P46643.1 AAT1_ARATH Aspartate aminotransferase, mitochondrial	S_PBL_c22205	C
S_PBL_lrc50519	2.09	NoMatchingHitfound	S_PBL_lrc50519	NA
Contig34639.1	2.08	Q9LWB0_SOLLC Homeodomain protein	CUST_77694_PI427086615	P
Contig6038.1	2.08	B9RFA3_RICCO Respiratory burst oxidase, putative	CUST_32148_PI427086615	P
Contig7350.1	2.08	Q9ZP35_TOBAC Alpha-expansin (Fragment)	CUST_34295_PI427086615	P
CB330750.1	2.07	AT2G33810.1 squamosa promoter binding protein-like 3	IbSPL3_CB330750.1probe1	F
Contig30233.1	2.07	B9T0A9_RICCO Adenosine kinase, putative	CUST_70653_PI427086615	C
S_PBL_lrc55132	2.07	NoMatchingHitfound	S_PBL_lrc55132	NA
Contig35728.1	2.06	A5AKL7_VITVI Putative uncharacterized protein	CUST_79431_PI427086615	C
Contig047139.3	2.05	Q501C9_ARATH At2g38960	CUST_8891_PI427086615	C
Contig705.1	2.05	Q8RVF8_TOBAC Thioredoxin peroxidase	CUST_23950_PI427086615	P
Contig7689.1	2.03	B9T6Z7_RICCO Pectate lyase, putative	CUST_34824_PI427086615	C
Contig9255.1	2.03	Q94ET1_IPOBA MADS-box protein	CUST_37180_PI427086615	NA
Contig046040.3	2.02	Q9M7B5_9ROSI Putative aquaporin PIP2-1	CUST_7285_PI427086615	C

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
RT_109343.1	2.02	B9GXG3_POPTR High mobility group family	CUST_83513_PI427086615	F
RT_209301.1	2.02	O82151_TOBAC Beta-D-glucan exohydrolase	CUST_85653_PI427086615	NA
Contig13201.1	2.01	Q2LAI9_SOLLC Auxin response factor 4	CUST_43212_PI427086615	P
Contig17916.1	2.01	B9R912_RICCO DNA binding protein, putative	CUST_50613_PI427086615	P
Contig3964.1	2	B9RBS2_RICCO Leucine rich repeat receptor kinase, putative	CUST_28883_PI427086615	NA
Contig39981.1	2	B9RJR6_RICCO DNA binding protein, putative	CUST_80515_PI427086615	P
RT_187492.1	2	B9RXQ4_RICCO Pectinesterase	CUST_85436_PI427086615	C
Contig12341.1	1.99	Q9XFL2_TOBAC Secretory peroxidase	CUST_41874_PI427086615	C
Contig17772.1	1.99	O82151_TOBAC Beta-D-glucan exohydrolase	CUST_50385_PI427086615	NA
Contig34887.1	1.99	B9RDN5_RICCO Carbonyl reductase, putative	CUST_78123_PI427086615	P
Contig9948.1	1.99	Q8H9C6_SOLTU Pyruvate decarboxylase (Fragment)	CUST_38219_PI427086615	P
Contig044258.3	1.98	B9SRH5_RICCO Transcription factor, putative	CUST_4681_PI427086615	P
Contig047941.3	1.98	RBOHB_SOLTU Respiratory burst oxidase homolog protein B	CUST_10043_PI427086615	P
Contig048665.3	1.98	FH1_ARATH Formin-like protein 1	CUST_11074_PI427086615	P
Contig14452.1	1.98	Q9FLN2_ARATH Emb|CAB62355.1	CUST_45196_PI427086615	NA
Contig15885.1	1.98	D7T4Z4_VITVI Whole genome shotgun sequence of line PN40024, scaffold_118.assembly12x (Fragment)	CUST_47359_PI427086615	NA
Contig23509.1	1.98	Q8S3K6_TOBAC Caffeic acid O-methyltransferase II	CUST_59640_PI427086615	F
Contig8806.1	1.98	B9SIT3_RICCO L-lactate dehydrogenase	CUST_36506_PI427086615	P
S_PBL_c16613	1.98	NoMatchingHitfound	S_PBL_c16613	NA
S_PBL_c19033	1.98	Q8LRL4_PETHY Nam-like protein 11 OS=Petunia hybrida GN=NH11 PE=2 SV=1	IbNAM_S_PBL_c19033probe1	F
S_PBL_c7862	1.98	NoMatchingHitfound	S_PBL_c7862	NA
Contig045814.3	1.97	D7MQE2_ARALY Cinnamoyl-CoA reductase family	CUST_6953_PI427086615	F
Contig047266.3	1.97	Q564D7_SOLLC Pectinesterase	CUST_9066_PI427086615	C
Contig2591.1	1.97	B9R7T0_RICCO Glucan endo-1,3-beta-glucosidase, putative	CUST_26775_PI427086615	F
Contig29952.1	1.97	B9RCG6_RICCO Polygalacturonase, putative	CUST_70141_PI427086615	P
Contig35560.1	1.97	O82151_TOBAC Beta-D-glucan exohydrolase	CUST_79173_PI427086615	F
Contig049707.3	1.96	B9SG57_RICCO Multidrug resistance protein mdtG, putative (Fragment)	CUST_12582_PI427086615	N
Contig12431.1	1.96	D2WL28_IPOBA Cinnamate 4-hydroxylase	CUST_42008_PI427086615	P
Contig17611.1	1.96	Q8GT43_TOBAC Putative rac protein (Fragment)	CUST_50127_PI427086615	F
Contig25696.1	1.96	Q8L6U7_COFAR Putative cyclin dependent kinase	CUST_63309_PI427086615	P
Contig28550.1	1.96	B9RN85_RICCO Boron transporter, putative	CUST_67822_PI427086615	F
Contig10778.1	1.95	VPS2B_ARATH Isoform 2 of Vacuolar protein sorting-associated protein 2 homolog 2	CUST_39457_PI427086615	F

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig17680.1	1.95	B9RYE4_RICCO Myosin heavy chain, fast skeletal muscle, embryonic, putative	CUST_50246_PI427086615	C
Contig3360.1	1.95	D3YE76_SOLTU Expansin (Fragment)	CUST_27956_PI427086615	P
Contig3821.1	1.95	B9RFW6_RICCO Serine-threonine protein kinase, plant-type, putative	CUST_28655_PI427086615	P
S_PBL_c22019	1.95	NoMatchingHitfound	S_PBL_c22019	NA
Contig1142.1	1.94	D7US50_PETHY Chalcone isomerase	CUST_24603_PI427086615	NA
Contig13166.1	1.94	B9SYB2_RICCO Asymmetric leaves1 and rough sheath, putative	Contig13166.1	P
S_PBL_c11096	1.94	Q9FEV7_TOBAC Microtubule-associated protein MAP65-1c OS=Nicotiana tabacum GN=map65-1c PE=2 SV=1	IbMAP65_S_PBL_c11096probe1	F
Contig049447.3	1.93	E7BTN4_VITVI Pectate lyase-like protein 3	CUST_12177_PI427086615	C
Contig052231.3	1.93	A2TEI4_9ROSI Xyloglucan endotransglycosylase/hydrolase XTH-38	CUST_21317_PI427086615	P
Contig13061.1	1.93	B9RGI3_RICCO DNA-binding protein MNB1B, putative	CUST_42980_PI427086615	F
Contig16110.1	1.93	B9SQL9_RICCO Putative uncharacterized protein	CUST_47723_PI427086615	NA
Contig049137.3	1.92	PHSL2_SOLTU Alpha-1,4 glucan phosphorylase L-2 isozyme, chloroplastic/amyloplastic	CUST_11721_PI427086615	F
Contig041975.3	1.91	B9RKA0_RICCO 1-aminocyclopropane-1-carboxylate oxidase, putative	CUST_1424_PI427086615	P
Contig22493.1	1.91	A5AKC2_VITVI Putative uncharacterized protein	CUST_57952_PI427086615	F
Contig25374.1	1.91	Q8VWL8_SOLLC Beta-mannosidase	CUST_63222_PI427086615	P
Contig8907.1	1.91	D7UBE2_VITVI Whole genome shotgun sequence of line PN40024, scaffold_171.assembly12x (Fragment)	CUST_36656_PI427086615	NA
S_PBL_c8626	1.91	NoMatchingHitfound	S_PBL_c8626	NA
Contig15431.1	1.9	C4P7W4_VITVI CBL-interacting protein kinase 15	CUST_46683_PI427086615	P
Contig28599.1	1.9	Q9M5A8_SOLLC Chaperonin 21	CUST_68051_PI427086615	C
Contig8031.1	1.9	Q8RVF8_TOBAC Thioredoxin peroxidase	CUST_35364_PI427086615	P
RT_053291.1	1.9	Q9FIQ5_ARATH At5g46700	CUST_81857_PI427086615	NA
Contig1449.1	1.89	Q9M9U3_ARATH At1g18720/F6A14_17	CUST_25003_PI427086615	C
Contig19808.1	1.89	B9SAK5_RICCO Amino acid binding protein, putative	CUST_53537_PI427086615	P
Contig20538.1	1.89	D7T946_VITVI Whole genome shotgun sequence of line PN40024, scaffold_11.assembly12x (Fragment)	CUST_54782_PI427086615	NA
Contig24364.1	1.89	Q9LU40_ARATH Emb|CAB10440.1	CUST_61079_PI427086615	C
Contig35669.1	1.89	B6T0P4_MAIZE Histone H4	CUST_79331_PI427086615	P
Contig9277.1	1.89	Q6RX67_PETHY Expansin-3	CUST_37210_PI427086615	P
RT_056763.1	1.89	B9IDR8_POPTR Pectinesterase	CUST_81958_PI427086615	C
RT_290879.1	1.89	B9S0F2_RICCO Xyloglucan endotransglucosylase/hydrolase protein 14, putative	CUST_87655_PI427086615	P
RT_443260.1	1.89	B9SN34_RICCO ATP binding protein, putative	CUST_90851_PI427086615	F

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig10338.1	1.88	B9SKV5_RICCO Cinnamoyl-CoA reductase, putative	CUST_38830_PI427086615	F
Contig11725.1	1.88	B9RCM6_RICCO Vacuolar ATP synthase subunit E, putative	CUST_40918_PI427086615	C
Contig17688.1	1.88	Q9LEB0_TOBAC Pectinesterase	CUST_50261_PI427086615	P
Contig19201.1	1.88	MTP4_ORYSJ Metal tolerance protein 4	CUST_52606_PI427086615	P
Contig2848.1	1.88	B9IAJ5_POPTR Glycosyl hydrolase family 9	CUST_27155_PI427086615	F
Contig32137.1	1.88	Y1143_ARATH Probable LRR receptor-like serine/threonine-protein kinase At1g14390	CUST_73787_PI427086615	F
Contig33465.1	1.88	C6TL96_SOYBN Adenylyl-sulfate kinase	CUST_75910_PI427086615	F
Contig3893.1	1.88	Q04126_TOBAC 23-kDa ploypeptide of photosystem II oxygen-evolving complex	CUST_28761_PI427086615	F
RT_256767.1	1.88	Q93XQ1_NICAL Cellulose synthase catalytic subunit	CUST_86642_PI427086615	P
Contig046337.3	1.87	Q4QZT3_COFCA Sucrose synthase	CUST_7702_PI427086615	P
Contig046603.3	1.87	D7KMB1_ARALY Secretory carrier membrane protein family protein	CUST_8098_PI427086615	P
Contig050556.3	1.87	B9RLP4_RICCO Hydrolase, hydrolyzing O-glycosyl compounds, putative	Contig050556.3	P
Contig18460.1	1.87	Q2XTD7_SOLTU Polygalacturonase-like protein-like	CUST_51486_PI427086615	F
Contig22237.1	1.87	Q9LU40_ARATH Emb|CAB10440.1	CUST_57505_PI427086615	NA
Contig29764.1	1.87	B9SC72_RICCO DNA binding protein, putative	CUST_69890_PI427086615	C
RT_187222.1	1.87	Q40473_TOBAC PS60 protein	CUST_85431_PI427086615	P
S_PBL_c11803	1.87	NoMatchingHitfound	S_PBL_c11803	NA
S_PBL_c16189	1.87	NoMatchingHitfound	S_PBL_c16189	NA
S_PBL_c16116	1.86	NoMatchingHitfound	S_PBL_c16116	NA
contig04573	1.85	C4P7W4_VITVI CBL-interacting protein kinase 15	contig04573	P
Contig055065.3	1.85	B9S6S4_RICCO DNA-damage-inducible protein f, putative	CUST_19501_PI427086615	P
Contig10096.1	1.85	B9SE98_RICCO Molybdopterin cofactor synthesis protein A, putative	CUST_38440_PI427086615	P
Contig13908.1	1.85	B3XWM9_NICBE Lectin receptor kinase-like protein	Contig13908.1	F
Contig17316.1	1.85	B9SCU1_RICCO Glucan endo-1,3-beta-glucosidase, putative	CUST_49638_PI427086615	F
contig27450	1.85	Q564D7_SOLLC Pectinesterase	contig27450	C
Contig12624.1	1.84	D7SJ88_VITVI Whole genome shotgun sequence of line PN40024, scaffold_0.assembly12x (Fragment)	CUST_42324_PI427086615	C
Contig18922.1	1.84	A3F771_IPONI Auxin response factor 8	CUST_52201_PI427086615	F
contig30208	1.84	B9SKH8_RICCO Dynamin, putative	contig30208	C
Contig5377.1	1.84	B9SIM4_RICCO Poly-A binding protein, putative	CUST_31095_PI427086615	P
Contig6886.1	1.84	Q9SMR9_ARATH Putative uncharacterized protein AT4g39840	CUST_33546_PI427086615	NA
Contig9184.1	1.84	B9SRH5_RICCO Transcription factor, putative	CUST_37072_PI427086615	NA
Contig047383.3	1.83	A7L745_PHAVU Family 1 glycosyltransferase	CUST_9219_PI427086615	F

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig050592.3	1.83	Q6K8D6_ORYSJ Os02g0760200 protein	CUST_13874_PI427086615	NA
Contig10398.1	1.83	B9SB09_RICCO Sphingolipid fatty acid alpha hydroxylase	CUST_38916_PI427086615	P
Contig12102.1	1.83	Q8LBL4_ARATH Putative thaumatin-like protein	CUST_41477_PI427086615	P
Contig13526.1	1.83	Q9SXX5_NICPA Fructose-bisphosphate aldolase	CUST_43732_PI427086615	P
Contig2072.2	1.83	AT2G12400.1 similar to unknown protein [Arabidopsis thaliana] (TAIR:AT2G25270.1); similar to hypothetical protein OsJ	IbTAP_Contig2072.2probe1	NA
Contig23452.1	1.83	B9RRX5_RICCO Serine/threonine-protein kinase, putative	CUST_59552_PI427086615	P
Contig4385.1	1.83	B9RXQ9_RICCO Xyloglucan endotransglucosylase/hydrolase protein A, putative	CUST_29502_PI427086615	P
Contig7722.1	1.83	DPEP_SOLTU 4-alpha-glucanotransferase, chloroplastic/amyloplastic	CUST_34878_PI427086615	P
RT_018988.1	1.83	B9I1J7_POPTR Nucleobase ascorbate transporter	CUST_81015_PI427086615	P
RT_039965.1	1.83	C4B8D7_TOBAC AtEB1c-like protein	CUST_81496_PI427086615	F
Contig047331.3	1.82	B9S8L8_RICCO Multicopper oxidase, putative	CUST_9145_PI427086615	P
Contig052615.3	1.82	B9GMS0_POPTR Predicted protein	CUST_16750_PI427086615	NA
Contig13505.1	1.82	B9S164_RICCO WRKY transcription factor, putative	CUST_43695_PI427086615	P
Contig15638.1	1.82	B9RLS7_RICCO Heparanase-2, putative	CUST_47003_PI427086615	C
Contig16007.1	1.82	B9R9I9_RICCO Peroxidase 66, putative	CUST_47541_PI427086615	C
Contig19412.1	1.82	GAE3_ARATH UDP-glucuronate 4-epimerase 3	CUST_52990_PI427086615	P
Contig31816.1	1.82	A9PF21_POPTR 2-dehydro-3-deoxyphosphoheptonate aldolase/ 3-deoxy-d-arabino-heptulosonate 7-phosphate synthetase	CUST_73202_PI427086615	C
Contig043084.3	1.81	Q8S9H4_SOLLC Ethylene response factor 1	CUST_3001_PI427086615	P
Contig14376.1	1.81	Q8VXD1_TOBAC Alpha-tubulin	IbTUBULIN_Contig14376.1probe1	C
Contig23828.1	1.81	B9SYJ3_RICCO Acetylglucosaminyltransferase, putative	CUST_60190_PI427086615	NA
Contig395.1	1.81	Q53U40_SOLLC Similar to ATP synthase subunit H protein	CUST_23502_PI427086615	P
Contig40418.1	1.81	B9SMN9_RICCO Transcription factor, putative	CUST_80591_PI427086615	NA
Contig8370.1	1.81	B9SEN5_RICCO Early nodulin, putative	CUST_35857_PI427086615	P
Contig9468.1	1.81	A9P888_POPTR Putative uncharacterized protein	CUST_37481_PI427086615	NA
S_PBL_c44093	1.81	NoMatchingHitfound	S_PBL_c44093	NA
Contig042016.3	1.8	Q9FZ86_ARATH At1g18650	CUST_1471_PI427086615	P
Contig049478.3	1.8	B9RBK2_RICCO Blue copper protein, putative	Contig049478.3	C
Contig1695.1	1.8	CHLP_TOBAC Geranylgeranyl diphosphate reductase, chloroplastic	CUST_25409_PI427086615	P
Contig35465.1	1.8	Q2VY18_SOLLC CONSTANS interacting protein 2a	CUST_79028_PI427086615	P
Contig3835.1	1.8	KPYA_TOBAC Pyruvate kinase isozyme A, chloroplastic	CUST_28679_PI427086615	P
Contig8023.1	1.8	RBOHB_SOLTU Respiratory burst oxidase homolog protein B	CUST_35352_PI427086615	P
Contig9298.1	1.8	B9GTR1_POPTR Fasciclin-like arabinogalactan protein 12.2	CUST_37234_PI427086615	C

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
RT_252469.1	1.8	Q93XK9_SOLLC Vacuolar-type H+-pyrophosphatase (Fragment)	CUST_86522_PI427086615	P
S_PBL_c30281	1.8	NoMatchingHitfound	S_PBL_c30281	NA
S_PBL_c4790	1.8	NoMatchingHitfound	S_PBL_c4790	NA
Contig041325.3	1.79	C8YZA8_CAPAN UPA15	CUST_479_PI427086615	C
Contig052081.3	1.79	B9REN2_RICCO Non-specific lipid-transfer protein	CUST_16107_PI427086615	NA
Contig15683.1	1.79	B9SNY7_RICCO Peptidase, putative	CUST_47067_PI427086615	P
Contig16303.1	1.79	PER5_VITVI Peroxidase 5	CUST_48036_PI427086615	P
Contig26212.1	1.79	B9RBW5_RICCO Triose phosphate/phosphate translocator, chloroplast, putative	CUST_64137_PI427086615	F
Contig989.1	1.79	B9S0D7_RICCO Tonoplast intrinsic protein, putative	CUST_24398_PI427086615	C
Contig9975.1	1.79	Q49N13_SOYBN Putative receptor-like protein kinase 3	CUST_38260_PI427086615	F
RT_207869.1	1.79	Q8LCM9_ARATH Putative uncharacterized protein	CUST_85647_PI427086615	NA
S_PBL_c25632	1.79	gar2_schpo ame: full=protein gar2	S_PBL_c25632	P
S_PBL_c40771	1.79	emb|AJ429702.1| *Ipomoea batatas* chloroplast atpE gene (partial)	S_PBL_c40771	C
Contig14957.1	1.78	O63067_SOYBN Aspartokinase-homoserine dehydrogenase	CUST_45944_PI427086615	P
Contig15917.1	1.78	B9T7X6_RICCO Transcription factor, putative	CUST_47411_PI427086615	F
Contig19145.1	1.78	B9RKN2_RICCO Sugar transporter, putative	CUST_52513_PI427086615	C
contig22472	1.78	Q9SDP1_ALLCE S-adenosylhomocysteine hydrolase (Fragment)	contig22472	F
DV037740.1	1.78	AT4G37750.1 Integrase-type DNA-binding superfamily protein	CUST_93495_PI427086615	P
Contig4474.1	1.77	B9RKC2_RICCO O-methyltransferase, putative	CUST_29678_PI427086615	F
Contig11559.1	1.76	Q8GZV0_TOBAC Obtusifoliol-14-demethylase	CUST_40656_PI427086615	C
Contig11909.1	1.76	C7FE10_PANGI Polygalacturonase inhibiting protein	CUST_41187_PI427086615	P
Contig1207.1	1.76	FLS_PETHY Flavonol synthase/flavanone 3-hydroxylase	CUST_24689_PI427086615	P
Contig14333.1	1.76	Q9FYW7_SOLLC BAC19.9	CUST_45022_PI427086615	NA
Contig15330.1	1.76	Q6RHX7_SOLLC Xyloglucan endotransglucosylase-hydrolase XTH9	CUST_46537_PI427086615	P
Contig18307.1	1.76	B9SRH5_RICCO Transcription factor, putative	CUST_51233_PI427086615	P
Contig2677.1	1.76	B9S1M1_RICCO Neutral alpha-glucosidase ab, putative	CUST_26888_PI427086615	P
Contig8077.1	1.76	B9IQS4_POPTR Fasciclin-like arabinogalactan protein 9.2	CUST_35427_PI427086615	C
S_PBL_c15391	1.76	NoMatchingHitfound	S_PBL_c15391	NA
Contig11204.1	1.75	Q93XQ1_NICAL Cellulose synthase catalytic subunit	CUST_40135_PI427086615	P
Contig21177.1	1.75	D2U578_SOLLC Cell cycle switch 52B	CUST_55793_PI427086615	P
Contig22795.1	1.75	B9SV36_RICCO Bel1 homeotic protein, putative	IbBEL3_probe2_Contig22795.1	P
Contig23841.1	1.75	Q9SXH2_IPOBA 24 kDa vacuolar protein VP24	CUST_60802_PI427086615	NA
contig24479	1.75	Q9SLF1_ARATH At2g16660/T24I21.7	contig24479	NA
Contig24720.1	1.75	B9RC93_RICCO Receptor kinase, putative	CUST_61698_PI427086615	F
contig33776	1.75	B9S327_RICCO Serine/threonine-protein kinase PBS1, putative	contig33776	P

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig34542.1	1.75	O81536_SOLLC Annexin p34	CUST_77555_PI427086615	F
contig35948	1.75	B9S1V5_RICCO CDK, putative	contig35948	P
Contig9278.1	1.75	B9HEV9_POPTR Predicted protein (Fragment)	CUST_37212_PI427086615	NA
RT_083745.1	1.75	B9RJ77_RICCO LysM domain GPI-anchored protein 1, putative	CUST_82737_PI427086615	C
Contig12064.1	1.74	B9RR23_RICCO Pectinesterase	CUST_41411_PI427086615	C
Contig18798.1	1.74	Q18PQ3_IPOBA Starch branching enzyme I	CUST_51947_PI427086615	C
Contig2072.2	1.74	AT2G12400.1 similar to unknown protein [Arabidopsis thaliana] (TAIR:AT2G25270.1); similar to hypothetical protein OsJ	IbTAP_Contig2072.2probe2	NA
Contig2604.1	1.74	B9SAX4_RICCO Cellulose synthase A catalytic subunit 3 [UDP-forming], putative	CUST_26791_PI427086615	P
Contig27885.1	1.74	Q9SQ62_IPOBA Anionic peroxidase	CUST_66806_PI427086615	P
Contig4681.1	1.74	Q9SLF1_ARATH At2g16660/T24I21.7	CUST_29981_PI427086615	NA
RT_351811.1	1.74	B9SWB6_RICCO Structural constituent of cell wall, putative	CUST_89437_PI427086615	NA
Contig10178.1	1.73	Q2QP56_ORYSJ Respiratory burst oxidase, putative, expressed	CUST_38567_PI427086615	P
Contig20331.1	1.73	B9N5N9_POPTR Uridine kinase	CUST_54389_PI427086615	C
contig22135	1.73	Q940P5_ARATH At2g20740/F5H14.29	contig22135	P
contig22458	1.73	A9P7U5_POPTR Pyruvate kinase	contig22458	P
Contig8249.1	1.73	B5LAU0_CAPAN Putative cinnamoyl-CoA reductase	CUST_35672_PI427086615	F
S_PBL_c11962	1.73	NoMatchingHitfound	S_PBL_c11962	NA
S_PBL_c27656	1.73	NoMatchingHitfound	S_PBL_c27656	NA
Contig042503.3	1.72	B9SRR7_RICCO Nodulation receptor kinase, putative	CUST_2167_PI427086615	C
Contig043916.3	1.72	B9SEN5_RICCO Early nodulin, putative	CUST_4202_PI427086615	P
Contig054388.3	1.72	TL29_ARATH Thylakoid lumenal 29 kDa protein, chloroplastic	CUST_18743_PI427086615	P
Contig11583.1	1.72	B9REF8_RICCO Beta-glucosidase, putative	CUST_40693_PI427086615	P
Contig14569.1	1.72	D7U0D4_VITVI Whole genome shotgun sequence of line PN40024, scaffold_2.assembly12x (Fragment)	CUST_45389_PI427086615	P
Contig24990.1	1.72	U603_ARATH UPF0603 protein At1g54780, chloroplastic	CUST_62057_PI427086615	P
Contig5661.1	1.72	B9S729_RICCO Patellin-5, putative	CUST_31537_PI427086615	P
Contig6397.1	1.72	A9XG40_TOBAC Subtilisin-like protease	CUST_32756_PI427086615	C
Contig043402.3	1.71	B9SL24_RICCO Patellin-6, putative	CUST_3475_PI427086615	P
Contig052664.3	1.71	A8QXP6_IPOBA Class-I knotted1-like homeobox protein IBKN2	CUST_16815_PI427086615	C
Contig12171.1	1.71	B9SH66_RICCO Glucan endo-1,3-beta-glucosidase, putative	CUST_41596_PI427086615	C
Contig15558.1	1.71	B9S0A9_RICCO Leucoanthocyanidin dioxygenase, putative	CUST_46889_PI427086615	P
contig17274	1.71	B9R869_RICCO Myosin heavy chain, striated muscle, putative	contig17274	F

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig1947.1	1.71	B9RJR6_RICCO DNA binding protein, putative	CUST_25783_PI427086615	P
Contig33726.1	1.71	Q9LEC9_SOLTU Alpha-glucosidase	CUST_76263_PI427086615	C
Contig33955.1	1.71	B9SC77_RICCO Cinnamoyl-CoA reductase, putative	CUST_76705_PI427086615	F
Contig34246.1	1.71	B5LAU5_CAPAN DH putative beta-hydroxyacyl-ACP dehydratase	CUST_77105_PI427086615	P
RT_145708.1	1.71	Q0IY20_ORYSJ Os10g0378400 protein	CUST_84545_PI427086615	NA
RT_290881.1	1.71	A5B2M5_VITVI Putative uncharacterized protein	CUST_87659_PI427086615	C
Contig15033.1	1.7	Q93XQ1_NICAL Cellulose synthase catalytic subunit	CUST_46059_PI427086615	P
Contig16480.1	1.7	HEM6_TOBAC Coproporphyrinogen-III oxidase, chloroplastic	CUST_48315_PI427086615	C
Contig22662.1	1.7	Q9SWV1_SOLLC ER33 protein (Fragment)	CUST_58208_PI427086615	P
Contig3731.1	1.7	PGMP_SOLTU Phosphoglucomutase, chloroplastic	CUST_28528_PI427086615	P
Contig4194.1	1.7	B9SSC2_RICCO DNA binding protein, putative	CUST_29220_PI427086615	P
Contig6850.1	1.7	O04897_SOLLC Fructokinase	CUST_33487_PI427086615	F
Contig6973.1	1.7	B9RR97_RICCO Xylem serine proteinase 1, putative	CUST_33683_PI427086615	C
Contig86.1	1.7	Q672Q6_SOLLC Photosystem II oxygen-evolving complex protein 3	CUST_23071_PI427086615	P
Contig9110.1	1.7	Y1699_ARATH Probable LRR receptor-like serine/threonine-protein kinase At1g69990	CUST_36958_PI427086615	P
Contig9838.1	1.7	Q8LJR6_SOYBN GTP-binding protein	CUST_38051_PI427086615	C
EE876438.1	1.7	B9RYY6_RICCO Auxin-induced protein 5NG4, putative	CUST_93273_PI427086615	C
Contig042152.3	1.69	Q94FM5_TOBAC Elicitor-inducible cytochrome P450	CUST_1667_PI427086615	P
Contig050485.3	1.69	B9RYW1_RICCO Glucan endo-1,3-beta-glucosidase, putative	Contig050485.3	F
Contig055164.3	1.69	B9RXP7_RICCO Beta-glucosidase, putative	CUST_19615_PI427086615	F
Contig12629.1	1.69	RBOHB_SOLTU Respiratory burst oxidase homolog protein B	CUST_42329_PI427086615	P
Contig1591.1	1.69	E0CPP6_VITVI Whole genome shotgun sequence of line PN40024, scaffold_1.assembly12x (Fragment)	CUST_25254_PI427086615	F
Contig18351.1	1.69	B9SPE8_RICCO 3-beta-hydroxy-delta5-steroid dehydrogenase, putative	CUST_51296_PI427086615	C
Contig21784.1	1.69	B9S4H4_RICCO Receptor protein kinase CLAVATA1, putative	CUST_56735_PI427086615	C
Contig26876.1	1.69	Q9SLN8_TOBAC Allyl alcohol dehydrogenase	CUST_65200_PI427086615	P
Contig35206.1	1.69	D5FNB4_CAPAN Mitochondrial ATP synthase 6kDa subunit	CUST_78602_PI427086615	NA
Contig5399.1	1.69	B9SU04_RICCO Glucan endo-1,3-beta-glucosidase, putative	CUST_31126_PI427086615	P
RT_012435.1	1.69	Q8VYM8_ARATH Putative senescence-associated protein 5	CUST_80889_PI427086615	NA
S_PBL_c3428	1.69	B9R6Z1_RICCO Putative uncharacterized protein OS=Ricinus communis GN=RCOM_1587270 PE=4 SV=1	IbXX_S_PBL_c3428probe3	F
Contig049940.3	1.68	B9ICT1_POPTR Pectinesterase	CUST_12911_PI427086615	C
Contig16573.1	1.68	B9TA90_RICCO Thermostable beta-glucosidase B, putative	CUST_48475_PI427086615	F
Contig34671.1	1.68	B9RJW6_RICCO DNA binding protein, putative	CUST_77735_PI427086615	P

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig39444.1	1.68	C0JP34_LOTJA Putative basic helix-loop-helix protein BHLH26 (Fragment)	IbBHLH_Contig39444.1probe1	P
Contig5760.1	1.68	Q6ZEZ1_ORYSJ Putative ribose-5-phosphate isomerase	CUST_31697_PI427086615	P
RT_357758.1	1.68	Q40473_TOBAC PS60 protein	CUST_89609_PI427086615	C
S_PBL_c16992	1.68	NoMatchingHitfound	S_PBL_c16992	NA
S_PBL_lrc42816	1.68	NoMatchingHitfound	S_PBL_lrc42816	NA
Contig047825.3	1.67	Q9FIQ5_ARATH At5g46700	CUST_21097_PI427086615	NA
Contig049275.3	1.67	Q43797_TOBAC Inorganic pyrophosphatase	CUST_11927_PI427086615	P
Contig049331.3	1.67	Q5WA73_ORYSJ Os06g0163300 protein	CUST_12003_PI427086615	NA
Contig20711.1	1.67	B9RCG6_RICCO Polygalacturonase, putative	CUST_55013_PI427086615	F
Contig30548.1	1.67	Q2HZF5_VITVI Aquaporin PIP2	CUST_71219_PI427086615	P
Contig550.1	1.67	Q5JBR2_IPOBA Anionic peroxidase swpb2	CUST_23746_PI427086615	P
Contig5625.1	1.67	A9PDK5_POPTR Predicted protein	CUST_31486_PI427086615	NA
Contig6831.1	1.67	Q9SKC9_ARATH At2g02050	CUST_33460_PI427086615	P
Contig6974.1	1.67	B6TNE4_MAIZE GDU1	CUST_33685_PI427086615	NA
DV038147.1	1.67	Q68HC9_SOLTU Glucose-6-phosphate isomerase	CUST_93693_PI427086615	P
RT_065253.1	1.67	Q6K8D6_ORYSJ Os02g0760200 protein	CUST_82237_PI427086615	NA
RT_290349.1	1.67	E1VD15_SOLTU Sterol reductase	CUST_87633_PI427086615	F
S_PBL_c12609	1.67	NoMatchingHitfound	S_PBL_c12609	NA
Contig049931.3	1.66	B9I1A6_POPTR Fasciclin-like arabinogalactan protein (Fragment)	CUST_12899_PI427086615	C
Contig050015.3	1.66	B9RBE5_RICCO Glucan endo-1,3-beta-glucosidase, putative	Contig050015.3	P
Contig051577.3	1.66	B9RT19_RICCO Serine/threonine protein kinase, putative	CUST_21253_PI427086615	C
Contig053451.3	1.66	Q539E7_VITVI Plastid hexose transporter	CUST_17713_PI427086615	F
Contig10350.1	1.66	B9T1Y2_RICCO ATP binding protein, putative	CUST_38846_PI427086615	C
Contig10466.1	1.66	B9RKF7_RICCO Glucan endo-1,3-beta-glucosidase, putative	CUST_39007_PI427086615	P
Contig11578.1	1.66	D3YM76_9SOLA Squalene epoxidase	CUST_40685_PI427086615	C
contig14135	1.66	D7MLU4_ARALY Phosphofructokinase family protein	contig14135	C
Contig325.1	1.66	A9YF26_IPOTF Putative anthocyanin transcriptional regulator	CUST_23411_PI427086615	P
Contig3304.1	1.66	C1KA92_POPTR Peroxidase	CUST_27874_PI427086615	P
Contig4098.1	1.66	Q9M1E7_ARATH At3g45600	CUST_29086_PI427086615	C
Contig515.1	1.66	Q6QDC7_IPONI Aquaporin-like protein	CUST_23694_PI427086615	P
Contig7355.1	1.66	FLA1_ARATH Fasciclin-like arabinogalactan protein 1	CUST_34303_PI427086615	C
EE883668.1	1.66	B9SL24_RICCO Patellin-6, putative	CUST_93341_PI427086615	P
RT_049925.1	1.66	B9RYW1_RICCO Glucan endo-1,3-beta-glucosidase, putative	CUST_81755_PI427086615	F
RT_327471.1	1.66	B9SCH2_RICCO ATP-binding cassette transporter, putative	CUST_88755_PI427086615	P

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
S_PBL_c16753	1.66	NoMatchingHitfound	S_PBL_c16753	NA
S_PBL_c20441	1.66	NoMatchingHitfound	S_PBL_c20441	NA
S_PBL_c35024	1.66	NoMatchingHitfound	S_PBL_c35024	NA
Contig045965.3	1.65	E1VD15_SOLTU Sterol reductase	CUST_7182_PI427086615	P
Contig051213.3	1.65	O82151_TOBAC Beta-D-glucan exohydrolase	CUST_14787_PI427086615	F
Contig056065.3	1.65	PGMC_PEA Phosphoglucomutase, cytoplasmic	CUST_20559_PI427086615	P
Contig11872.1	1.65	A2Q440_MEDTR Harpin-induced 1	CUST_41129_PI427086615	NA
Contig11879.1	1.65	B9SH79_RICCO Nucleoside transporter, putative	CUST_41140_PI427086615	P
Contig1537.1	1.65	SODCP_PETHY Superoxide dismutase [Cu-Zn], chloroplastic	CUST_25177_PI427086615	P
Contig16216.1	1.65	Q69UK6_ORYSJ Putative C2 domain-containing protein	CUST_47879_PI427086615	F
Contig1938.1	1.65	B9RQC7_RICCO Phosphofructokinase, putative	CUST_25767_PI427086615	C
Contig20780.1	1.65	B9SSC2_RICCO DNA binding protein, putative	CUST_55107_PI427086615	P
Contig21622.1	1.65	ATPO_IPOBA ATP synthase subunit O, mitochondrial	CUST_56513_PI427086615	P
Contig21663.1	1.65	Q84QE4_TOBAC Putative chloroplast thiazole biosynthetic protein	CUST_56570_PI427086615	P
Contig6985.1	1.65	Q9FLN3_ARATH At5g40960	CUST_33702_PI427086615	NA
S_PBL_c10170	1.65	NoMatchingHitfound	S_PBL_c10170	NA
S_PBL_c12903	1.65	NoMatchingHitfound	S_PBL_c12903	NA
S_PBL_c28833	1.65	NoMatchingHitfound	S_PBL_c28833	NA
AB478416.1	1.64	*Ipomoea batatas* CycD3;1 mRNA for cyclin D3, complete cds	IbCYCD3;2_AB478416.1probe1	F
Contig043786.3	1.64	Q10J83_ORYSJ JmjC domain containing protein, expressed	CUST_4009_PI427086615	C
Contig046650.3	1.64	Q40336_MEDSA Proline-rich cell wall protein	CUST_8173_PI427086615	NA
Contig15375.1	1.64	LAX4_MEDTR Auxin transporter-like protein 4	CUST_46600_PI427086615	P
Contig16955.1	1.64	D7TGK6_VITVI Pectinesterase (Fragment)	CUST_49097_PI427086615	C
Contig17141.1	1.64	IQD32_ARATH Protein IQ-DOMAIN 32	CUST_49391_PI427086615	C
Contig19985.1	1.64	Q40541_TOBAC Protein kinase	CUST_54456_PI427086615	NA
Contig28191.1	1.64	B9S292_RICCO Homeobox protein, putative	CUST_67339_PI427086615	P
Contig40350.1	1.64	B9RYU9_RICCO Endo-1,4-beta-glucanase, putative	CUST_80579_PI427086615	P
Contig7716.1	1.64	Q9AVP4_TOBAC BY-2 kinesin-like protein 10	CUST_34868_PI427086615	NA
RT_084621.1	1.64	Q9C540_ARATH At1g26100	CUST_82769_PI427086615	P
RT_373485.1	1.64	Q9MB62_IPONI Phytocyanin-related protein	CUST_89937_PI427086615	C
Contig050777.3	1.63	Q7XAE3_PETIN Putative fructokinase 2	CUST_14132_PI427086615	F
Contig16197.1	1.63	B9RD90_RICCO Pectinesterase	CUST_47851_PI427086615	C
Contig20229.1	1.63	Q9LWS9_ORYSJ Os06g0114700 protein	CUST_54241_PI427086615	NA
Contig31356.1	1.63	B9R6R8_RICCO Xylulose kinase, putative	CUST_72464_PI427086615	P
Contig35753.1	1.63	B6T440_MAIZE Adenosylhomocysteinase	CUST_79470_PI427086615	F

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
S_PBL_c28918	1.63	NoMatchingHitfound	S_PBL_c28918	NA
S_PBL_c39395	1.63	NoMatchingHitfound	S_PBL_c39395	NA
Contig042438.3	1.62	Q9SLF1_ARATH At2g16660/T24I21.7	CUST_2068_PI427086615	NA
Contig050817.3	1.62	A6N4B9_MANIN Endo-beta-1,4-glucanase	CUST_14187_PI427086615	F
Contig20747.1	1.62	B9T001_RICCO WD-repeat protein, putative	CUST_55060_PI427086615	P
Contig33488.1	1.62	Q9ZP33_SOLLC Expansin	CUST_75937_PI427086615	P
Contig35715.1	1.62	PGMC_SOLTU Phosphoglucomutase, cytoplasmic	CUST_79410_PI427086615	P
Contig9035.1	1.62	Q8W2N2_TOBAC Cytochrome P450-dependent fatty acid hydroxylase	CUST_36847_PI427086615	P
RT_291306.1	1.62	Q2QUH9_ORYSJ DHHC zinc finger domain containing protein, expressed	CUST_87665_PI427086615	NA
S_PBL_c26605	1.62	NoMatchingHitfound	S_PBL_c26605	NA
Contig10157.1	1.61	BI1L_ARATH BI1-like protein	CUST_38536_PI427086615	C
Contig10853.1	1.61	Q8W1A8_PETHY Aquaporin-like protein	CUST_39576_PI427086615	P
Contig13656.1	1.61	Q06XL8_VITVI KUP2	CUST_43939_PI427086615	C
Contig20007.1	1.61	B9RXP7_RICCO Beta-glucosidase, putative	CUST_54467_PI427086615	F
Contig2796.1	1.61	Q5JBR3_IPOBA Anionic peroxidase swpb1	CUST_27074_PI427086615	P
Contig30054.1	1.61	B9H1J6_POPTR Predicted protein	CUST_70408_PI427086615	F
Contig34182.1	1.61	B9RN88_RICCO ADP-ribosylation factor, arf, putative	CUST_77017_PI427086615	C
Contig9846.1	1.61	B9SGR1_RICCO ATP binding protein, putative	CUST_38064_PI427086615	C
RT_045065.1	1.61	B9R6Y4_RICCO Serine carboxypeptidase, putative	CUST_81629_PI427086615	F
S_PBL_c14800	1.61	vitis vinifera clone ss0aeb29yh09; BJ553125 Ipomoea nil	S_PBL_c14800	NA
Contig1459.1	1.6	Q9LFS3_ARATH AT5g16010/F1N13_150	CUST_25020_PI427086615	P
Contig16778.1	1.6	D7T766_VITVI Pyruvate kinase (Fragment)	CUST_48819_PI427086615	P
Contig16851.1	1.6	A3F771_IPONI Auxin response factor 8	CUST_48939_PI427086615	P
Contig27281.1	1.6	B9T173_RICCO Receptor protein kinase CLAVATA1, putative	CUST_65855_PI427086615	F
Contig28842.1	1.6	BGL46_ARATH Beta-glucosidase 46	CUST_68388_PI427086615	P
Contig30150.1	1.6	Q6EQH5_ORYSJ Membrane protein-like	CUST_70538_PI427086615	NA
Contig35581.1	1.6	B9R869_RICCO Myosin heavy chain, striated muscle, putative	CUST_79212_PI427086615	F
Contig35667.1	1.6	Q9SEE6_SOLTU Pectinesterase	CUST_79327_PI427086615	C
Contig3999.1	1.6	B7X6S6_TOBAC Secretory carrier-associated membrane protein 2	CUST_28938_PI427086615	C
Contig4138.1	1.6	B9S075_RICCO Endo-1,4-beta-glucanase, putative	CUST_29147_PI427086615	P
Contig4865.1	1.6	Q4PS96_SOLLC Plastidic hexokinase	CUST_30261_PI427086615	C
Contig5077.1	1.6	B9S3V0_RICCO Protein binding protein, putative	CUST_30604_PI427086615	C
Contig5488.1	1.6	Q9LM03_SOLTU Methionine synthase	CUST_31265_PI427086615	C
Contig9863.1	1.6	Q9FYW7_SOLLC BAC19.9	CUST_38092_PI427086615	NA
RT_129785.1	1.6	AP4E_ARATH AP-4 complex subunit epsilon	CUST_84129_PI427086615	F

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
RT_234832.1	1.6	B5LAW1_CAPAN Putative aminotransferase	CUST_86095_PI427086615	F
S_PBL_c17568	1.6	NoMatchingHitfound	S_PBL_c17568	NA
S_PBL_c51385	1.6	NoMatchingHitfound	S_PBL_c51385	NA
Contig041366.3	1.59	B9SHB4_RICCO Polygalacturonase-1 non-catalytic subunit beta, putative	CUST_533_PI427086615	P
Contig043865.3	1.59	A0S5Z4_SESIN Peroxidase	CUST_4131_PI427086615	C
Contig10060.1	1.59	C0Z328_ARATH AT5G25460 protein	CUST_38379_PI427086615	NA
Contig12957.1	1.59	B9S479_RICCO Nodulation receptor kinase, putative	CUST_42813_PI427086615	C
Contig19004.1	1.59	Q9SDN5_TOBAC FH protein NFH2	CUST_52322_PI427086615	C
Contig32814.1	1.59	B6STB2_MAIZE Cytochrome c oxidase polypeptide VIb	CUST_74823_PI427086615	C
Contig34615.1	1.59	Q8L416_ORYSJ Os01g0908400 protein	CUST_77660_PI427086615	C
Contig6557.1	1.59	Q9LYJ5_ARATH Polygalacturonase-like protein	CUST_33019_PI427086615	P
Contig7736.1	1.59	B9RMR5_RICCO Photosystem II core complex proteins psbY, chloroplast	CUST_34903_PI427086615	C
RT_281296.1	1.59	B9RVT1_RICCO Receptor protein kinase CLAVATA1, putative	CUST_87339_PI427086615	F
RT_283793.1	1.59	A0MNL4_POPTR CBL-interacting protein kinase 6	CUST_87413_PI427086615	P
Contig047443.3	1.58	Q8LL10_PETHY Hairy meristem	CUST_9315_PI427086615	P
Contig048862.3	1.58	U503A_ARATH UPF0503 protein At3g09070, chloroplastic	CUST_11329_PI427086615	C
Contig11146.1	1.58	Q93XQ1_NICAL Cellulose synthase catalytic subunit	CUST_40017_PI427086615	P
Contig13594.1	1.58	Q84QE7_TOBAC Putative photosystem I subunit III	CUST_43859_PI427086615	C
Contig14060.1	1.58	Q8W1A8_PETHY Aquaporin-like protein	CUST_44596_PI427086615	P
Contig1975.1	1.58	Q7Y078_SOLTU Sucrose synthase 4	CUST_25858_PI427086615	P
Contig35069.1	1.58	SPXM3_ARATH SPX domain-containing membrane protein At4g22990	CUST_78400_PI427086615	C
Contig5276.1	1.58	B9INU8_POPTR Equilibrative nucleoside transporter (Fragment)	CUST_30904_PI427086615	P
Contig532.1	1.58	Q2XTB4_SOLTU Putative uncharacterized protein	CUST_23720_PI427086615	NA
Contig5687.1	1.58	B9RG92_RICCO Aspartic proteinase nepenthesin-1, putative	CUST_31578_PI427086615	F
Contig5859.1	1.58	U497F_RICCO UPF0497 membrane protein 15	Contig5859.1	C
Contig8765.1	1.58	Q96569_SOLLC L-lactate dehydrogenase	CUST_36445_PI427086615	P
EE876093.1	1.58	AT2G02850.1 plantacyanin	IbARPN_EE876093.1probe2	C
RT_309836.1	1.58	B9T0A1_RICCO Peroxidase 17, putative	CUST_88227_PI427086615	P
S_PBL_c26896	1.58	NoMatchingHitfound	S_PBL_c26896	NA
Contig043308.3	1.57	Q60CX0_SOLTU Putative FAD binding domain containing protein, identical	CUST_3330_PI427086615	NA
Contig045876.3	1.57	Q8LRL4_PETHY Nam-like protein 11	CUST_7052_PI427086615	P
Contig050436.3	1.57	D2D2Z8_GOSHI Phosphoglycerate kinase	CUST_13658_PI427086615	P
Contig051263.3	1.57	B9SMT9_RICCO Aquaporin NIP1.1, putative	CUST_21243_PI427086615	P

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig051531.3	1.57	Q1I0X5_CAPAN Pyruvate kinase	CUST_15274_PI427086615	F
Contig1039.1	1.57	Q9SXH2_IPOBA 24 kDa vacuolar protein VP24	CUST_24467_PI427086615	NA
Contig1217.1	1.57	Q84MD9_ARATH At3g08610	CUST_24706_PI427086615	C
Contig13497.1	1.57	B5M9J3_TOBAC Sucrose transporter	CUST_43685_PI427086615	P
Contig14657.1	1.57	P93390_TOBAC Phosphate/phosphoenolpyruvate translocator	CUST_45510_PI427086615	C
Contig16252.1	1.57	A9P7U5_POPTR Pyruvate kinase	CUST_47946_PI427086615	P
Contig18949.1	1.57	A7UE73_SOLTU LRR receptor-like kinase	CUST_52244_PI427086615	F
Contig22092.1	1.57	B9S377_RICCO Ceramidase, putative	CUST_57308_PI427086615	C
Contig23842.1	1.57	B9R873_RICCO Endo-1,4-beta-glucanase, putative	CUST_60208_PI427086615	P
Contig34888.1	1.57	VATL_TOBAC V-type proton ATPase 16 kDa proteolipid subunit	CUST_78125_PI427086615	F
Contig35817.1	1.57	PGLR_VITVI Probable polygalacturonase	CUST_79575_PI427086615	P
Contig652.1	1.57	B9GF99_POPTR Heavy metal ATPase (Fragment)	CUST_23890_PI427086615	P
RT_157972.1	1.57	B9SL14_RICCO Systemin receptor SR160, putative	CUST_84847_PI427086615	C
RT_232288.1	1.57	B9SMR1_RICCO Dimethylaniline monooxygenase, putative	CUST_86023_PI427086615	NA
Contig048113.3	1.56	Q94EN5_CAMSI Beta-1,3-glucanase	CUST_10295_PI427086615	P
Contig050464.3	1.56	B8Y8A0_GOSHI Blue copper-like protein	CUST_13700_PI427086615	C
Contig10557.1	1.56	GATL7_ARATH Probable galacturonosyltransferase-like 7	CUST_39129_PI427086615	P
Contig10798.1	1.56	B9R7K1_RICCO Plastoquinol-plastocyanin reductase, putative	CUST_39488_PI427086615	NA
Contig12237.1	1.56	B9SUV9_RICCO Catalytic, putative	CUST_41697_PI427086615	NA
Contig12654.1	1.56	B9T2V6_RICCO Multicopper oxidase, putative	CUST_42365_PI427086615	P
Contig15820.1	1.56	B9ICT1_POPTR Pectinesterase	CUST_47267_PI427086615	F
Contig15838.1	1.56	Q9LUI6_ARATH Kinase-like protein	CUST_47295_PI427086615	P
Contig15882.1	1.56	B9T7E3_RICCO Acyltransferase, putative	CUST_47356_PI427086615	P
Contig1779.1	1.56	Q6L460_SOLDE ATP synthase D chain, mitochondrial, putative	CUST_25530_PI427086615	P
Contig18308.1	1.56	Q948N7_IPOBA Starch branching enzyme II	CUST_51236_PI427086615	C
Contig18806.1	1.56	B9RYW8_RICCO Cucumisin, putative	CUST_52152_PI427086615	P
Contig246.1	1.56	D7UDD0_VITVI Whole genome shotgun sequence of line PN40024, scaffold_122.assembly12x (Fragment)	CUST_23305_PI427086615	NA
Contig29050.1	1.56	B9RND0_RICCO Kinase, putative	CUST_68676_PI427086615	P
Contig29781.1	1.56	B9SRA9_RICCO Auxin:hydrogen symporter, putative	CUST_69911_PI427086615	C
Contig32737.1	1.56	B9S4A6_RICCO DNA binding protein, putative	CUST_74727_PI427086615	NA
Contig33030.1	1.56	E0XN34_SOLLC Alpha-mannosidase	CUST_75215_PI427086615	F
Contig34928.1	1.56	Q2PYX3_SOLTU Fructose-bisphosphate aldolase	CUST_78194_PI427086615	P
Contig35170.1	1.56	B9T812_RICCO Endoplasmic oxidoreductin-1, putative	CUST_78552_PI427086615	P
Contig35183.1	1.56	Q2RAK2_ORYSJ Pyruvate kinase	CUST_78573_PI427086615	F

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig3941.1	1.56	Q9SU16_ARATH At4g12700	CUST_28844_PI427086615	NA
Contig647.1	1.56	NDK2_TOBAC Nucleoside diphosphate kinase 2, chloroplastic	CUST_23884_PI427086615	P
Contig8003.1	1.56	D5FI33_IPOBA MADS-box protein	CUST_35319_PI427086615	F
RT_014667.1	1.56	B0EW04_SOYBN Trihelix transcription factor	CUST_80940_PI427086615	F
RT_397671.1	1.56	BXL4_ARATH Beta-D-xylosidase 4	CUST_90295_PI427086615	C
S_PBL_c16382	1.56	ref|XP_002264818.1| PREDICTED: hypothetical protein [Vitis vinifera]	S_PBL_c16382	NA
Contig048622.3	1.55	A0AAU0_POPTR WOX4 protein (Fragment)	CUST_11011_PI427086615	P
Contig14263.1	1.55	B9S3L7_RICCO Histidine triad (Hit) protein, putative	CUST_44912_PI427086615	P
Contig26601.1	1.55	B9SCL8_RICCO Cyclin-dependent kinases regulatory subunit, putative	CUST_64658_PI427086615	C
Contig33148.1	1.55	Q56X96_ARATH Cyclopropyl isomerase	CUST_75364_PI427086615	P
Contig3413.1	1.55	A4GWX5_SOLCH Ovule receptor-like kinase 28	CUST_28038_PI427086615	F
Contig38162.1	1.55	PERN_IPOBA Neutral peroxidase	CUST_80158_PI427086615	P
Contig042083.3	1.54	Q96569_SOLLC L-lactate dehydrogenase	CUST_1569_PI427086615	P
Contig055423.3	1.54	Q9XFL2_TOBAC Secretory peroxidase	CUST_19895_PI427086615	P
Contig12028.1	1.54	Q8L5U4_ARATH Putative rubisco subunit binding-protein alpha subunit	CUST_41349_PI427086615	P
Contig12727.1	1.54	B9H1J6_POPTR Predicted protein	Contig12727.1	NA
Contig14063.1	1.54	C7C5S8_NICBE PME inhibitor	CUST_44601_PI427086615	F
Contig1697.1	1.54	B9SYV2_RICCO Cysteine proteinase inhibitor, putative	CUST_25412_PI427086615	F
Contig17730.1	1.54	Q7X7J4_ORYSJ OSJNBb0070J16.15 protein	CUST_50324_PI427086615	P
Contig18231.1	1.54	Q653E2_ORYSJ Os09g0560300 protein	CUST_51125_PI427086615	NA
Contig23499.1	1.54	Q9FLN3_ARATH At5g40960	Contig23499.1	NA
Contig23980.1	1.54	B9SZH5_RICCO Auxin-induced protein 5NG4, putative	CUST_60408_PI427086615	P
Contig25622.1	1.54	B9RHH4_RICCO Ribonucleoprotein, chloroplast, putative	CUST_63100_PI427086615	C
Contig28660.1	1.54	B9S327_RICCO Serine/threonine-protein kinase PBS1, putative	CUST_68141_PI427086615	P
Contig29502.1	1.54	B5LAW1_CAPAN Putative aminotransferase	CUST_69427_PI427086615	C
Contig31173.1	1.54	B9SSL3_RICCO Aspartic proteinase nepenthesin-1, putative	CUST_72209_PI427086615	NA
Contig4126.1	1.54	D7T4Z4_VITVI Whole genome shotgun sequence of line PN40024, scaffold_118.assembly12x (Fragment)	CUST_29129_PI427086615	NA
Contig462.1	1.54	B9SN27_RICCO Nucleic acid binding protein, putative	CUST_23590_PI427086615	P
Contig6686.1	1.54	B9STR3_RICCO Endosomal P24A protein, putative	CUST_33223_PI427086615	P
Contig6879.1	1.54	Q8GYU5_ARATH At1g61240	CUST_33536_PI427086615	NA
Contig048259.3	1.53	Q84UC3_SOLTU Sucrose synthase 2	CUST_10492_PI427086615	P
Contig051925.3	1.53	Q9LV36_ARATH Emb|CAB36779.1	CUST_15870_PI427086615	NA

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig13628.1	1.53	D7TGK6_VITVI Pectinesterase (Fragment)	CUST_43903_PI427086615	C
Contig1542.1	1.53	B9RG74_RICCO NADH dehydrogenase, putative	CUST_25188_PI427086615	P
Contig15944.1	1.53	Q96569_SOLLC L-lactate dehydrogenase	CUST_47447_PI427086615	P
Contig16578.1	1.53	Q0J5I8_ORYSJ Os08g0431500 protein (Fragment)	CUST_48486_PI427086615	C
Contig22107.1	1.53	Q6L4K0_SOLDE Putative DNA-binding protein	CUST_57331_PI427086615	P
Contig26724.1	1.53	B9SM94_RICCO Sur2 hydroxylase/desaturase, putative	CUST_64978_PI427086615	P
Contig29495.1	1.53	Q1PCD2_SOLLC Glucose-6-phosphate isomerase	CUST_69417_PI427086615	P
Contig33353.1	1.53	Q66GR8_ARATH At1g03470	CUST_75648_PI427086615	NA
Contig34941.1	1.53	B9SEN5_RICCO Early nodulin, putative	CUST_78205_PI427086615	P
Contig34960.1	1.53	B9RZ87_RICCO Rnf5, putative	CUST_78230_PI427086615	P
Contig35532.1	1.53	B3F8H6_NICLS Sucrose sythase	CUST_79135_PI427086615	P
Contig35885.1	1.53	Q9SDN6_TOBAC FH protein NFH1	CUST_79738_PI427086615	C
Contig6216.1	1.53	B9GSE6_POPTR Glycosyltransferase, CAZy family GT8	CUST_32428_PI427086615	P
Contig6485.1	1.53	B9SKY3_RICCO ATP binding protein, putative	CUST_32904_PI427086615	F
Contig6727.1	1.53	B9S6V4_RICCO DNA binding protein, putative	CUST_33286_PI427086615	P
Contig7343.1	1.53	B9GYJ7_POPTR 10-formyltetrahydrofolate synthetase	CUST_34286_PI427086615	P
Contig7775.1	1.53	Q94EF5_ORYSJ Os01g0769200 protein	CUST_34977_PI427086615	NA
Contig8825.1	1.53	Q5WA73_ORYSJ Os06g0163300 protein	CUST_36535_PI427086615	NA
Contig946.1	1.53	B9SJ12_RICCO DUF26 domain-containing protein 1, putative	CUST_24294_PI427086615	F
RT_054279.1	1.53	D0U6M6_ORYSJ Stress-induced protein kinase	CUST_81887_PI427086615	P
Contig045618.3	1.52	Q76MS5_SOLLC LEXYL1 protein	CUST_6659_PI427086615	C
Contig15892.1	1.52	Q952R1_SOLLC Succinate dehydrogenase subunit 3	CUST_47370_PI427086615	P
Contig16120.1	1.52	B9SR17_RICCO Dtdp-glucose 4-6-dehydratase, putative	CUST_47739_PI427086615	C
Contig17247.1	1.52	CHS6_IPOBA Chalcone synthase DII	CUST_49537_PI427086615	F
Contig17483.1	1.52	B9T1D1_RICCO Alanine-glyoxylate aminotransferase, putative	CUST_49933_PI427086615	F
Contig22882.1	1.52	B9S2A4_RICCO Serine/threonine-protein kinase PBS1, putative	CUST_58623_PI427086615	P
Contig25053.1	1.52	B9RP67_RICCO Cellulose synthase A catalytic subunit 6 [UDP-forming], putative	CUST_62151_PI427086615	P
Contig2854.1	1.52	B9R9V0_RICCO Receptor protein kinase CLAVATA1, putative	CUST_27167_PI427086615	F
Contig4118.1	1.52	B9SEH3_RICCO Cytochrome B561, putative	CUST_29117_PI427086615	P
Contig4720.1	1.52	B6TEY9_MAIZE Blue copper protein	CUST_30037_PI427086615	C
Contig7429.1	1.52	PFPA_SOLTU Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha	CUST_34411_PI427086615	F
Contig8342.1	1.52	CH10_ARATH 10 kDa chaperonin	CUST_35823_PI427086615	F

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig8447.1	1.52	B9DHB2_ARATH AT4G35920 protein	CUST_35985_PI427086615	C
Contig9505.1	1.52	Q2QLQ4_ORYSJ Erg28 like protein, expressed	CUST_37539_PI427086615	C
RT_332947.1	1.52	B9STF6_RICCO Kinetochore protein nuf2, putative	CUST_88945_PI427086615	NA
RT_372264.1	1.52	B6U8L8_MAIZE Heparanase-like protein 3	CUST_89921_PI427086615	C
S_PBL_c16213	1.52	NoMatchingHitfound	S_PBL_c16213	NA
S_PBL_c23348	1.52	NoMatchingHitfound	S_PBL_c23348	NA
S_PBL_c48361	1.52	NoMatchingHitfound	S_PBL_c48361	NA
S_PBL_c9152	1.52	NoMatchingHitfound	S_PBL_c9152	NA
Contig048513.3	1.51	B0I545_ZINEL Cellulose synthase Z632	CUST_10860_PI427086615	P
Contig050527.3	1.51	Q6J8X1_9ROSI Cellulose synthase	CUST_13780_PI427086615	P
Contig13220.1	1.51	Q5F2L4_CAPCH Putative receptor associated protein (Fragment)	CUST_43242_PI427086615	C
Contig1543.1	1.51	Q8VYV1_ARATH AT5g08050/F13G24_250	CUST_25189_PI427086615	NA
Contig16613.1	1.51	Q653Y0_ORYSJ Os06g0681200 protein	CUST_48545_PI427086615	C
contig18912	1.51	Q9FYW7_SOLLC BAC19.9	contig18912	NA
Contig19200.1	1.51	C6ZRW9_SOYBN Receptor-like protein kinase	CUST_52604_PI427086615	F
Contig2309.1	1.51	B9SVF6_RICCO Acyltransferase, putative	CUST_26341_PI427086615	P
Contig26628.1	1.51	B9MVJ1_POPTR Lysine/histidine transporter	CUST_64691_PI427086615	P
Contig29234.1	1.51	Q6IVK6_TOBAC Putative UDP-glucose dehydrogenase 2	CUST_69036_PI427086615	P
Contig30885.1	1.51	RHN1_NICPL Ras-related protein RHN1	CUST_71721_PI427086615	P
Contig31906.1	1.51	B9RZC1_RICCO Dolichyl-diphosphooligosaccharide--protein glycosyltransferase, putative	CUST_73333_PI427086615	P
Contig32689.1	1.51	HOX11_ORYSJ Homeobox-leucine zipper protein HOX11	CUST_74668_PI427086615	P
Contig33350.1	1.51	B9T3A9_RICCO Ptm1, putative	CUST_75644_PI427086615	F
Contig36530.1	1.51	B9RLW7_RICCO NAC domain-containing protein 21/22, putative	CUST_79894_PI427086615	P
Contig3989.1	1.51	Q9FYF8_ARATH At1g67350	CUST_28924_PI427086615	NA
Contig6708.1	1.51	B9RC93_RICCO Receptor kinase, putative	CUST_33253_PI427086615	P
RT_219609.1	1.51	B6TJP1_MAIZE Ras-related protein RIC2	CUST_85741_PI427086615	C
S_PBL_c27014	1.51	NoMatchingHitfound	S_PBL_c27014	NA
S_PBL_c3892	1.51	NoMatchingHitfound	S_PBL_c3892	NA
S_PBL_lrc27472	1.51	NoMatchingHitfound	S_PBL_lrc27472	NA
Contig045157.3	1.5	B9SIH7_RICCO Transcription factor, putative	CUST_5984_PI427086615	P
Contig050379.3	1.5	Q8L7E9_ARATH AT4G35920 protein	CUST_13583_PI427086615	C
Contig053096.3	1.5	B9RW41_RICCO Protein phosphatase-2c, putative	CUST_17325_PI427086615	P
Contig10678.1	1.5	B2ZGR8_SOYBN NAC domain protein (Fragment)	CUST_39307_PI427086615	P
Contig10995.1	1.5	MAF1_SOLLC MFP1 attachment factor 1	CUST_39791_PI427086615	F

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig16304.1	1.5	Q8W1A5_PETHY Aquaporin-like protein	CUST_48037_PI427086615	C
Contig16308.1	1.5	B6T0P4_MAIZE Histone H4	CUST_48046_PI427086615	P
Contig16971.1	1.5	B9SAI9_RICCO Glucose-6-phosphate/phosphate translocator 1, chloroplast, putative	CUST_49126_PI427086615	C
Contig17163.1	1.5	E2IFI7_NICBE Plastid NEP interaction protein	Contig17163.1	C
Contig19954.1	1.5	D0QU16_SOLLC Peptide N-glycanase	CUST_53868_PI427086615	F
Contig20043.1	1.5	A9CPA7_SOYBN Protein disulfide isomerase family	CUST_53986_PI427086615	P
Contig23245.1	1.5	B9HFW6_POPTR Cytochrome P450	CUST_59279_PI427086615	P
Contig25586.1	1.5	B9T595_RICCO NADH-ubiquinone oxidoreductase 12 kDa subunit, mitochondrial, putative	CUST_63048_PI427086615	P
Contig28335.1	1.5	A5BJJ8_VITVI CBL-interacting protein kinase 07	CUST_67532_PI427086615	P
Contig32838.1	1.5	B9RYW8_RICCO Cucumisin, putative	CUST_74858_PI427086615	P
Contig3374.1	1.5	B9R6T5_RICCO Homeobox protein, putative	IbHB1_Contig3374.1probe1	P
Contig35364.1	1.5	Q94ET1_IPOBA MADS-box protein	CUST_78832_PI427086615	P
Contig370.1	1.5	B9SWB6_RICCO Structural constituent of cell wall, putative	CUST_23467_PI427086615	NA
Contig4100.1	1.5	Q40473_TOBAC PS60 protein	CUST_29089_PI427086615	P
Contig4374.1	1.5	O63067_SOYBN Aspartokinase-homoserine dehydrogenase	CUST_29480_PI427086615	P
Contig6601.1	1.5	B9SIM4_RICCO Poly-A binding protein, putative	CUST_33087_PI427086615	P
Contig71.1	1.5	B9S2L1_RICCO Thioredoxin m(Mitochondrial)-type, putative	CUST_23047_PI427086615	P
Contig7687.1	1.5	C7C5S8_NICBE PME inhibitor	CUST_34819_PI427086615	C
Contig9796.1	1.5	B9SBN5_RICCO Rubisco subunit binding-protein beta subunit, rubb, putative	CUST_37994_PI427086615	C
RT_006906.1	1.5	B9ZZE4_IPONI Glyceraldehyde-3-phosphate dehydrogenase	CUST_80775_PI427086615	P
RT_052282.1	1.5	B9RS47_RICCO Aspartate aminotransferase	CUST_81829_PI427086615	F
Contig045371.3	1.49	Q40515_TOBAC A-type cyclin	CUST_6293_PI427086615	F
Contig046159.3	1.49	B9SWA9_RICCO Non-specific lipid-transfer protein	Contig046159.3	C
Contig048227.3	1.49	Q5JBR3_IPOBA Anionic peroxidase swpb1	CUST_10454_PI427086615	P
Contig051697.3	1.49	B9RHY4_RICCO Phosphoglycerate kinase	CUST_15506_PI427086615	P
Contig16635.1	1.49	A5BTB0_VITVI Pyruvate kinase	CUST_48583_PI427086615	F
Contig17128.1	1.49	Q9MAZ1_SOYBN Nonclathrin coat protein zeta2-COP	CUST_49372_PI427086615	C
Contig18848.1	1.49	Q6R608_SOLTU 4-alpha-glucanotransferase	CUST_52008_PI427086615	P
Contig24099.1	1.49	GDL40_ARATH GDSL esterase/lipase At2g27360	CUST_60576_PI427086615	C
Contig26664.1	1.49	B9HQA7_POPTR Beta-galactosidase	CUST_65536_PI427086615	F
Contig27116.1	1.49	UGAL1_ARATH UDP-galactose transporter 1	CUST_65503_PI427086615	C
Contig28438.1	1.49	D4AHS6_IPOBA Granule-bound starch synthase I	CUST_67680_PI427086615	C

(Appendix F cont.)

Unigene	FC	Annotation	Probeid	GO group
Contig32837.1	1.49	PPR28_ARATH Pentatricopeptide repeat-containing protein At1g09900	CUST_74855_PI427086615	C
Contig34699.1	1.49	B9REF8_RICCO Beta-glucosidase, putative	CUST_77771_PI427086615	P
Contig32711.1	1.49	B9RKT6_RICCO Aldehyde dehydrogenase, putative	CUST_74696_PI427086615	P
Contig3960.1	1.49	ATP5E_IPOBA ATP synthase subunit epsilon, mitochondrial	CUST_28878_PI427086615	F
Contig401.1	1.49	D7P230_NICBE HB1	IbHB2_Contig401.1probe1	P
Contig7217.1	1.49	UGPI3_ARATH Uncharacterized GPI-anchored protein At5g19250	CUST_34085_PI427086615	C
Contig7960.1	1.49	B9RG41_RICCO NADH dehydrogenase, putative	CUST_35258_PI427086615	C
Contig9886.1	1.49	O24329_RICCO Putative uncharacterized protein	CUST_38129_PI427086615	NA
RT_041135.1	1.49	B9RS47_RICCO Aspartate aminotransferase	CUST_81529_PI427086615	F
RT_285216.1	1.49	B9SMY1_RICCO Mads box protein, putative	CUST_87453_PI427086615	P
RT_356926.1	1.49	B9RUC1_RICCO Ring finger protein, putative	CUST_89580_PI427086615	F
RT_387100.1	1.49	B9T491_RICCO Dopamine beta-monooxygenase, putative	CUST_90149_PI427086615	NA
S_PBL_c278	1.49	NoMatchingHitfound	S_PBL_c278	NA
Contig047674.3	1.48	B9SB23_RICCO Protein binding protein, putative	CUST_9654_PI427086615	P
Contig15320.1	1.48	Q8H8T0_ORYSJ Alpha-1,4-glucan protein synthase	CUST_46526_PI427086615	P
Contig17292.1	1.48	B9RYC5_RICCO Serine-threonine protein kinase, plant-type, putative	CUST_49607_PI427086615	F
Contig3969.1	1.48	B9S162_RICCO Purine transporter, putative	CUST_28894_PI427086615	P
Contig721.1	1.48	B9RJ32_RICCO Vacuolar ATP synthase proteolipid subunit 1, 2, 3, putative	CUST_23974_PI427086615	F
S_PBL_c7639	1.61	NoMatchingHitfound	S_PBL_c7639	NA
Contig28605.1	1.52	B9RTL8_RICCO Protein binding protein, putative	CUST_68061_PI427086615	NA
S_PBL_c20147	4.17	NoMatchingHitfound	S_PBL_c20147	NA
S_PBL_c14309	1.55	NoMatchingHitfound	S_PBL_c14309	NA
S_PBL_c31709	1.55	NoMatchingHitfound	S_PBL_c31709	NA
S_PBL_c32247	1.64	NoMatchingHitfound	S_PBL_c32247	NA
S_PBL_lrc27581	2.58	NoMatchingHitfound	S_PBL_lrc27581	NA
S_PBL_c34448	2.45	NoMatchingHitfound	S_PBL_c34448	NA
S_PBL_c8171	2.35	NoMatchingHitfound	S_PBL_c8171	NA
S_PBL_c1516	2.05	NoMatchingHitfound	S_PBL_c1516	NA
S_PBL_c54882	1.89	NoMatchingHitfound	S_PBL_c54882	NA
S_PBL_c43019	1.72	NoMatchingHitfound	S_PBL_c43019	NA
S_PBL_c15482	1.58	NoMatchingHitfound	S_PBL_c15482	NA
S_PBL_c16662	1.57	NoMatchingHitfound	S_PBL_c16662	NA
Contig12013.1	1.53	Q6Z608_ORYSJ Os08g0261100 protein	CUST_41326_PI427086615	NA

Appendix G: List of significant downregulated genes in storage roots from the microarray results (FDR<0.05) and fold change ratio <-1.58. FC=fold change.

Probeid	FC	Annotation	Unigene	GO term	GO category
CUST_11464_PI427086615	-2.04	D7M111_ARALY Invertase/pectin methylesterase inhibitor family protein	Contig048963.3	GO:0005618	cell wall
CUST_16127_PI427086615	-1.91	C9WF05_GOSHI Class III peroxidase	Contig052093.3	GO:0006950	response to stress
CUST_25807_PI427086615	-2.21	O23190_ARATH MAP3K-like protein kinase	Contig1960.1	GO:0007165	signal transduction
CUST_31615_PI427086615	-2.08	B9S775_RICCO Peroxidase 10, putative	Contig5710.1	GO:0006950	response to stress
CUST_37581_PI427086615	-2.37	A1E0X8_SOLTU NAC domain protein NAC2	Contig9539.1	GO:0030528	transcription regulator activity
CUST_40954_PI427086615	-1.55	B8LFG9_IPOBA Tonoplast intrinsic protein	Contig11751.1	GO:0005215	transporter activity
CUST_44967_PI427086615	-2.25	B3SHI0_IPOBA Anionic peroxidase swpa7	Contig14301.1	GO:0006950	response to stress
CUST_54963_PI427086615	-2.11	D1M7W9_SOLLC Heat stress transcription factor A3	Contig20679.1	GO:0030528	transcription regulator activity
contig03166	-1.68	Q40090_IPOBA SPF1 protein	Contig050116.3	GO:0030528	transcription regulator activity
CUST_13187_PI427086615	-1.69	Q40090_IPOBA SPF1 protein	Contig050116.3	GO:0030528	transcription regulator activity
CUST_13188_PI427086615	-1.83	Q40090_IPOBA SPF1 protein	Contig050116.3	GO:0030528	transcription regulator activity
contig10788	-1.94	B9RC22_RICCO Multidrug resistance pump, putative	Contig13553.1	GO:0030528	transcription regulator activity
contig24525	-1.58	B9SAP4_RICCO Multidrug resistance-associated protein 2, 6 (Mrp2, 6), abc-transoprter, putative	Contig20352.1	GO:0005215	transporter activity
contig36064	-1.80	PDR1_TOBAC Pleiotropic drug resistance protein 1	RT_381219.1	GO:0005215	transporter activity
CUST_13887_PI427086615	-1.69	B9I191_POPTR Multidrug resistance protein ABC transporter family	Contig050604.3	GO:0005215	transporter activity
CUST_14631_PI427086615	-1.82	B9SX20_RICCO NAC domain-containing protein, putative	Contig051116.3	GO:0005215	transporter activity
CUST_36993_PI427086615	-1.85	Q76CU1_TOBAC PDR-type ABC transporter 2 (Fragment)	Contig9131.1	GO:0005215	transporter activity
CUST_37299_PI427086615	-2.98	B9SCG8_RICCO Nitrate transporter, putative	Contig9348.1	GO:0005215	transporter activity
CUST_41967_PI427086615	-1.80	PDR1_TOBAC Pleiotropic drug resistance protein 1	Contig12402.1	GO:0005215	transporter activity
CUST_43782_PI427086615	-1.97	B9RC22_RICCO Multidrug resistance pump, putative	Contig13553.1	GO:0005215	transporter activity
CUST_4579_PI427086615	-1.69	Q8GU64_ORYSJ MRP-like ABC transporter	Contig044185.3	GO:0005215	transporter activity
CUST_46118_PI427086615	-2.26	D7M111_ARALY Invertase/pectin methylesterase inhibitor family protein	Contig15056.1	GO:0005618	transporter activity
CUST_48781_PI427086615	-2.88	Q84MZ8_TOBAC High affinity nitrate transporter protein	Contig16753.1	GO:0005215	transporter activity
CUST_49343_PI427086615	-1.53	B9RQF2_RICCO ATP-binding cassette transporter, putative	Contig17112.1	GO:0005215	transporter activity
CUST_54420_PI427086615	-1.72	B9I9S5_POPTR Multidrug resistance protein ABC transporter family	Contig20352.1	GO:0005215	transporter activity
CUST_23161_PI427086615	-1.52	Q0WSR2_ARATH Putative peroxidase	Contig144.1	GO:0006950	response to stress
CUST_42875_PI427086615	-1.93	B9T8I2_RICCO Peroxidase N, putative	Contig12997.1	GO:0006950	response to stress
CUST_11378_PI427086615	-1.59	B6V6Z9_POPEU GAST-like protein	Contig048896.3	GO:0006950	response to stress
CUST_25604_PI427086615	-1.60	B9R887_RICCO Gibberellin-regulated protein 1, putative	Contig1829.1	GO:0009725 GO:0005554	response to hormone stimulus molecular function unknown

(Appendix G cont.)

Probeid	FC	Annotation	Unigene	GO term	GO category
CUST_39798_PI427086615	-1.68	B9HPF8_POPTR GRAS family transcription factor	Contig11000.1	GO:0030528	transcription regulator activity
CUST_42099_PI427086615	-1.62	B9HZY3_POPTR NAC domain protein, IPR003441	Contig12488.1	GO:0030528	transcription regulator activity
CUST_4519_PI427086615	-1.92	D5LHU3_IPONI Gibberellin 2-oxidase 2	Contig044148.3	GO:0009685	gibberellin metabolic process
CUST_14117_PI427086615	-1.58	C7ENF8_IPOBA WRKY1	Contig050769.3	GO:0030528	transcription regulator activity
CUST_26266_PI427086615	-1.53	Q9FXS1_TOBAC WRKY transcription factor NtEIG-D48	Contig2257.1	GO:0030528	transcription regulator activity
CUST_26353_PI427086615	-1.74	D3GDP7_SOLLC JA-induced WRKY protein	Contig2317.1	GO:0030528	transcription regulator activity
CUST_26489_PI427086615	-1.94	C7E5X8_CAPAN Transcription factor WRKY	Contig2412.1	GO:0030528	transcription regulator activity
CUST_36533_PI427086615	-1.86	B6U788_MAIZE WRKY39v2-superfamily of TFs having WRKY and zinc finger domains	Contig8824.1	GO:0030528	transcription regulator activity
CUST_3846_PI427086615	-2.17	C7E5X8_CAPAN Transcription factor WRKY	Contig043659.3	GO:0030528	transcription regulator activity
CUST_47820_PI427086615	-1.70	B9S164_RICCO WRKY transcription factor, putative	Contig16173.1	GO:0030528	transcription regulator activity
CUST_49258_PI427086615	-1.53	B9N4E6_POPTR ABC transporter family protein (Fragment)	Contig17052.1	GO:0005215	transporter activity
CUST_30987_PI427086615	-1.90	Q5GA67_SOLLC BHLH transcriptional regulator	Contig5325.1	GO:0030528	transcription regulator activity
CUST_34219_PI427086615	-1.51	D7P233_TOBAC MYC1a transcription factor	Contig7297.1	GO:0030528	transcription regulator activity
CUST_38571_PI427086615	-1.86	C3W4Q3_VITVI R2R3 transcription factor MYB108-like protein 1	Contig10181.1	GO:0030528	transcription regulator activity
CUST_4141_PI427086615	-1.81	D7P234_TOBAC MYC1b transcription factor	Contig043877.3	GO:0030528	transcription regulator activity
CUST_4142_PI427086615	-1.83	D7P234_TOBAC MYC1b transcription factor	Contig043877.3	GO:0030528	transcription regulator activity
CUST_48150_PI427086615	-1.59	D7P233_TOBAC MYC1a transcription factor	Contig16379.1	GO:0030528	transcription regulator activity
CUST_54086_PI427086615	-1.54	B9SYQ1_RICCO R2r3-myb transcription factor, putative	Contig20112.1	GO:0030528	transcription regulator activity
CUST_24041_PI427086615	-1.93	ERF2_TOBAC Ethylene-responsive transcription factor 2	Contig769.1	GO:0030528	transcription regulator activity
CUST_26785_PI427086615	-2.03	Q53JG2_ORYSJ AP2 domain containing protein, expressed	Contig2597.1	GO:0030528	transcription regulator activity
CUST_26786_PI427086615	-2.11	Q53JG2_ORYSJ AP2 domain containing protein, expressed	Contig2597.1	GO:0030528	transcription regulator activity
CUST_28007_PI427086615	-1.71	A3F770_IPONI APETALA2-like protein	Contig3392.1	GO:0030528	transcription regulator activity
CUST_32801_PI427086615	-1.73	A9PL97_POPTR AP2/ERF domain-containing transcription factor	Contig6424.1	GO:0030528	transcription regulator activity
CUST_33975_PI427086615	-1.89	B9RG00_RICCO AP2 domain transcription factor RAP2.3, putative	Contig7149.1	GO:0030528	transcription regulator activity
CUST_35057_PI427086615	-2.11	A9PL97_POPTR AP2/ERF domain-containing transcription factor	Contig7824.1	GO:0030528	transcription regulator activity

Vita

Julio Solis-Sarmiento was born to Benedicta Sarmiento and Bernardino Solis in Abancay City of Apurimac, Peru. He pursued his high school education in the same city and finished his bachelor's degree at Universidad Nacional Mayor de San Marcos (UNMSM), Lima, Peru. He then started working at International Potato Center (CIP) in molecular genetics of an Andean crop, *Lepidium meyenii*, potato, and sweetpotato. He pursued his master's degree in biochemistry and molecular biology at Universidad Peruana Cayetano Heredia. He also took courses at the I Master in Bioinformatic at the Universidad Internacional de Andalucia. While working at CIP, he lectured in molecular biology, genetics and bioinformatic at the School of Biological Sciences at UNMSM. He also organized and lectured in two Bioinformatic courses in Peru in 2008 and 2009. In 2008, he was admitted to Louisiana State University to earn a Ph. D. degree in the School of Plant, Environmental and Soil Sciences under the guidance of Dr. Don Labonte.