January 2015

Soybean isoflavonoid biosynthesis: constituents and circumstance at the transcriptomic and molecular levels

Mehran Dastmalchi
The University of Western Ontario

Supervisor
Sangeeta Dhaubhadel
The University of Western Ontario

Graduate Program in Biology
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

SOYBEAN ISOFLAVONOID BIOSYNTHESIS: CONSTITUENTS AND CIRCUMSTANCE AT THE TRANSCRIPTOMIC AND MOLECULAR LEVELS

(Thesis format: Integrated Article)

by

Mehran Dastmalchi

Graduate Program in Biology

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario

ABSTRACT

Isoflavonoids are specialized metabolites, almost exclusive to the legume family of plants. They are actors in symbiosis with nitrogen-fixing bacteria and in plant stress response. Isoflavonoids are noted for their human health benefits. Isoflavonoid content in legumes has proven to be a complex trait. The goal of the present research is to determine the mechanisms underlying isoflavonoid biosynthesis in soybean.

The first approach was to unravel the genetic factors of isoflavonoid biosynthesis. A branch-point enzyme of the phenylpropanoid pathway, chalcone isomerase (CHI), catalyzes the reaction producing flavanones, the nucleus for many downstream metabolites such as isoflavonoids. I identified twelve soybean *CHI* genes, including five new members. The evolutionary history of the family shows that the enzymatic fold evolved from being catalytically inactive to being a chalcone-to-flavanone isomerase. Four CHIs in soybean were identified with the latter functionality. The CHIs showed differential temporal and spatial expression, pointing to the potential function of CHI1A in soybean seed isoflavonoid production. On the molecular level of organization, the long-postulated model of a subcellular isoflavonoid enzyme complex forming at the surface of the endoplasmic reticulum (ER) was substantiated in the present study. Here, I identified key players in the 'metabolon', established *in planta* subcellular localization, and investigated protein-protein interaction. The results suggest that isoflavone synthase (IFS), a cytochrome P450, is a nucleating metabolic center at the surface of the ER, interacting with upstream pathway enzymes. Finally, a transcriptomic study was undertaken to find genetic elements linked with isoflavonoid content variation in four soybean cultivars. The results suggest that competing branches of the phenylpropanoid pathway are combinatorially regulated to coordinate flux into isoflavonoid biosynthesis. The candidate genes encode enzymes in the overlapping pathways, several transcription factors, metabolic transporters and more.

Study of the *CHI* gene family and isoflavonoid biosynthesis has provided us with new insights into production and regulation of this important plant natural product. This knowledge can facilitate the manipulation of metabolic content and composition in legumes, and introduction of the isoflavonoid pathway into non-legume crops.

KEYWORDS

Isoflavonoids; soybean; specialized metabolism; chalcone isomerase; isoflavone synthase; biosynthesis; metabolon; phylogeny; cloning; subcellular localization; protein-protein interaction; transcriptomics.

CO-AUTHORSHIP STATEMENT

The following thesis contains material from a manuscript.

My supervisors Dr. Mark Bernards and Dr. Sangeeta Dhaubhadel provided insight and strategic direction for the projects and also edited the thesis.

Chapter 2: Authors' contributions

MD designed the research, performed the experiments, analyzed the data and drafted the manuscript. SD conceived the study and participated in its design. MD and SD edited the manuscript. A version of the chapter was published by the journal *Planta*.

To my Mother

ACKNOWLEDGEMENTS

One man can row across a lake in a dinghy, but it takes a whole fleet to man a ship from the eastern shore of the Atlantic – that's where I come from – to the other. So, here are my thanks to the crew above and aboard.

I dedicate this thesis to my mother, who has always been my lighthouse in fair and foul weather. Thanks as well to my father – for the eyebrows and affinity for literature. To my brothers: Keyvan, for encouraging me on in this terrifying world of science, and Bahram, for introducing me to everyone from *Bataille* to *Sebald*. A good part of this journey, arguably the steepest incline of the present dramatic narrative, has been made possible by my sweetheart and lovely partner in life, Tanya, the Yang to my fragmentary Ying.

I also want to thank my adopted family: the rolling numbers that have come through during my time at Agriculture Canada. Eridan Pereira, my colleague, roommate, and brother; Reza, Mandana, Behnaz, Jaya, Preetam, Jerry, Alex, Danny, Nikita and Korey; everyone in the Dhaubhadel lab, Arun, Hemanta, Kishor, Caroline, Jaeju, Pravesh, Arjun, Xuyan and Shaomin – thank you for lighting up my days. Thank you Ling for our fiery debates, and for teaching me all the fundamentals of lab-work. Thank you Lisa for training me on confocal microscopy, and your companionship through dire straits! Thanks to Bryan, my Creative Director and Able Seaman in Uxbridge, Ontario, Canada.

I must thank the individuals who have been essential to the present research: my advisory committee, Dr. Rima Menassa and Dr. Susanne Kohalmi for their invaluable input and guidance; Dr. Ryan Austin and Patrick Chapman for their contribution to the RNA sequencing project; Alex Molnar for his help with the figures in this thesis and beyond. I also have a great debt of gratitude to Dr. Mark Bernards for his insight throughout my years at Western and his advice at crucial turning points in my research.

Most importantly, I must thank my supervisor and mentor, Dr. Sangeeta Dhaubhadel for giving me the opportunity to join her lab, and for her guidance, support and constant pursuit of excellence in science. Thank you for always leaving your door open to me for advice on research and more.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

2-HID	2-HydroxyIsoflavanone Dehydratase
ABC	ATP Binding Cassette
ADT	Arogenate Dehydratase
At	*Arabidopsis thaliana*
bHLH	Basic Helix-loop-helix
BiFC	Bi-molecular Fluorescence Complementation
BLAST	Basic Local Alignment Search Tool
BLASTN	Nucleotide BLAST
BWA	Burrows-Wheeler Aligner
C4H	Cinnamate-4-Hydroxylase
CFP	Cyan Fluorescent Protein
CHI	Chalcone Isomerase
CHIL	CHI-like
CHR	Chalcone Reductase
CHS	Chalcone Synthase
Co-IP	Co-Immunoprecipitation
CRAN	Comprehensive R Archive Network
DAP	Days After Pollination

DE	Differentially Expressed
DESeq	Differential Expression Sequencing
DFR	Dihydroflavonol 4-Reductase
DNA	Deoxyribonucleic Acid
ER	Endoplasmic Reticulum
F3'5'H	Flavonoid 3',5'-Hydroxylase
F3'H	Flavonoid 3'-Hydroxylase
F3H	Flavanone 3-Hydroxylase
F6H	Flavonoid 6-Hydroxylase
FAP	Fatty Acid-binding Protein
FLIM	Fluorescence-Lifetime Imaging Microscopy
FLS1	Flavonol Synthase
FRET	Förster Resonance Energy Transfer
GFP	Green Fluorescent Protein
Gm	*Glycine max*
GO	Gene Ontology Database
GUS	β-glucuronidase
H3K4	Histone 3 Lysine 4
HAD	Histone Deacetylase

HIV	Human Immunodeficiency Virus
HPLC	High-Performance Liquid Chromatography
HRP	Horseradish Peroxidase
I2'H	Isoflavone 2-Hydroxylase
IFR	Isoflavone Reductase
IFS	Isoflavone Synthase
IOMT	Isoflavone O-methyltransferase
LB	Luria-Bertani Medium
LC-MS/MS	Liquid Chromatography-tandem Mass Spectrometry
LRR	Leucine-rich Repeat
M-A plot	M (log ratios) and A (mean average) plot
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionization Time-of-Flight
MATE	Multi-drug and Toxin Extrusion
MS	Mass Spectrometry
MYB	Myeloblastosis
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NB	Nucleotide-binding
NCBI	National Centre for Biotechnology Information
Nod	Nodulation Factor

NTC	No Template Control
P450	Cytochrome P450
PAL	Phenylalanine Ammonia-lyase
PDT	Prephenate Dehydratase
PVDF	Polyvinylidene Difluoride
qPCR	Quantitative Polymerase Chain Reaction
RNA	Ribonucleic Acid
RT-PCR	Reverse Transcriptase-Polymerase Chain Reaction
SCN	Soybean Cyst Nematode
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
TAIR	The Arabidopsis Information Resource
TT12	Transparent Testa 12
UGT	UDP glucuronosyltransferases
UPLC	Ultra-Performance Liquid Chromatography
UV	Ultra-violet
YC	C-Terminal Fragment of YFP
YFP	Yellow Fluorescent Protein
YN	N-Terminal Fragment of YFP

CHAPTER ONE – GENERAL INTRODUCTION

1.1 Soybean: a global commodity with specialized metabolites

Soybean (*Glycine max* [L.] Merr.) is an important legume crop, with its global commodity value at 60 billion dollars, behind rice and wheat (FAO 2014). Soybean seeds have very high protein (40%) and oil (20%) content (Hill and Breidenbach 1974; Ohlrogge and Kuo 1984), making it a lucrative crop, for human and animal nutrition, as well as for industrial feedstock and biofuels. Soy products are becoming an increasingly important component of the North American diet, as a source of high-value proteins, essential fatty acids, vitamins and minerals, with low fat and carbohydrate content (Messina 2010). The benefits of human soy-food consumption have been extensively studied, and include reducing the risk of a number of chronic illnesses, such as cardiovascular disease, hypercholesterolemia, hormone-dependent cancers and osteoporosis (Cederroth and Nef 2009; Dixon 2004; Dixon and Ferreria 2002; Lamartiniere 2000; Sarkar and Li 2003; Miadokova 2009). These health benefits are primarily conferred by the rich repertoire of micronutrients and plant specialized metabolites in soybean, including isoflavonoids (Dixon 2004; Dixon and Ferreria 2002). Soybean seeds accumulate the largest amount of isoflavonoids, as compared to other plant organs, and among legumes (Dhaubhadel et al. 2003). The advent of high-throughput 'omics' strategies has provided us with a wealth of knowledge, and yet the variability and complexity of isoflavonoid production and regulation of their biosynthesis have proven to be elusive.

1.2 Isoflavonoids are legume-specific specialized metabolites

Isoflavonoids are plant specialized metabolites, almost exclusive to the Fabaceae that play a role in symbiosis with nitrogen-fixing bacteria in the soil, and a role in plant stress and disease response. Isoflavonoids are noted for their human health benefits. These benefits

are accessible to humans, as soybean seeds are the primary dietary source of isoflavonoids. The genomics, transcriptomics and metabolomics of the isoflavonoid pathway have been extensively studied (Jones and Vodkin 2013; Dhaubhadel et al. 2007; Le et al. 2007; Farag et al. 2008; Lin et al. 2014); however, knowledge of the functional and regulatory elements involved in the pathway has been a more gradual process (Agrawal et al. 2008; Hajduch et al. 2005).

Knowledge of the molecular nature and the complex flux between overlapping phenylpropanoid pathways has raised the importance of characterizing the large gene families and identifying the key players in isoflavonoid biosynthesis. In plants, there is a disjunction between transcript, metabolite and protein abundances. Soybean is a paleopolyploid with multi-gene families encoding enzymes, transcription factors and structural elements of metabolic pathways (Schmutz et al. 2010). Each member corresponds to an isoform often specific to certain spatio-temporal dimensions and/or responsive to environmental interactions.

Soybean isoflavonoid production is highly variable and multi-genic, coordinated at the whole-plant, organ, tissue, cellular, and molecular levels. Establishing and understanding the proteomes at these different levels under the spectra of environmental conditions, would allow us to find important proteins, functionally characterize them, and build a model of the different elements driving isoflavonoid production. The importance of this approach is highlighted by the frequency of isoflavonoid pathway enzymes that have the ability to bind to different ‘metabolons’ (multi-enzyme complexes) (Winkel 2004b), leading us to a conjectured model of combinatorial substrate specificity at the molecular level. Due to these challenges, knowledge of the molecular and genetic mechanisms that regulate isoflavonoid biosynthesis is tantamount to our understanding of the pathway.

1.2.1 Roles in human health and nutrition

Isoflavonoids are one of the key plant specialized metabolites that are biologically available and active in effecting human health benefits. Clinical and epidemiological studies have shown that isoflavonoids in particular, and soy foods in general, contribute to human health by reducing the risk of chronic illnesses and ailments (Dixon 2004;

Dixon and Ferreria 2002; Duffy et al. 2007; Lamartiniere et al. 1995; Messina and Barnes 1991). The structure of isoflavonoids, specifically genistein, bears similarity to endogenous estrogenic hormone, 17β-estradiol. Many of the physiological and chemo-preventive benefits of isoflavonoids are linked to this structural similarity (Dixon and Ferreria 2002). Competition of phytoestrogens with 17β-estradiol, in binding to intra-nuclear estrogen receptors, modulates the activity of endogenous hormones. Therefore, dietary isoflavonoid intake can have a) an anti-estrogenic effect in individuals with high estrogen levels, or b) an estrogenic effect in individuals with low estrogen levels (reviewed in (Sarkar and Li 2003)). Tissue-type and estrogen receptor composition can affect the estrogenic activity of isoflavones in the cell (Doerge and Sheehan 2002).

Studies on the long-term effects of dietary isoflavonoid intake have shown a correlative reduction in the risk of breast and prostrate cancer (Duffy et al. 2007; Lamartiniere 2000; Xu et al. 2009). Soy consumption in early childhood is linked with a 28-60% reduction in the risk of developing breast cancer (Shu et al. 2001). Dietary genistein has been shown to inhibit the cascade of prostate cancer cell metastasis, blocking mitogen-activated protein kinase p38 (Xu et al. 2009). More recently, genistein has been shown to interfere with a tyrosine kinase involved in the actin dynamics of HIV infection (Guo et al. 2013).

Isoflavonoids are also noted for their role in climacteric medicine and moderating the endocrine and somatic changes occurring in women at menopause. Intake of isoflavonoids has been shown to reduce osteoporosis (Potter et al. 1998), and ease hot flashes (Howes et al. 2006); and has been linked with reduced risk of cardiovascular disease and high blood cholesterol levels (Strom et al. 2001; Walker et al. 2001).

Plant-based products in human diet that contain pseudo-hormonal compounds are controversial. Despite, the wealth of clinical and epidemiological data showing no linkage between dietary isoflavonoids and hormonal abnormality, phytoestrogens remain problematic. In particular, the use of soy-based infant formula and isoflavonoid supplements among hormonally susceptible groups has been controversial due to its structural similarity to17β-estradiol (Chen and Rogan 2004; Mendez et al. 2002). The dosage of isoflavonoids accumulated in infants, through soy-formula, is much higher than

adult consumption, considering their small body mass. There have been no direct links reported, to date, between soy-formula and adverse developmental or health effects.

Overall, isoflavonoids have been shown to have a wide range of human health benefits, owing, in part, to their capacity to bind to estrogen receptors and modulate estrogenic activity. The controversy attached to this functionality creates a challenge for the soy industry to adapt to consumer requirements. This research must focus on the natural variation in soybean seed isoflavonoid content and the underlying molecular, genetic mechanisms.

1.2.2 Roles in plant-environment interactions

Isoflavonoids are of dual importance in plant interactions with their environment: as phytoalexins in basal and innate responses to pathogens and stress, and as signals in symbiosis with nitrogen-fixing bacteria.

Phytoalexins are low-molecular weight, host-induced antimicrobial substances that are synthesized and released in response to pathogen attack or stress. Phytoalexins constitute a broad-spectrum plant defense against both prokaryotic and eukaryotic microorganisms (Paxton 1981). Soybean phytoalexins include both the simple isoflavones and their more complex derivatives. Several studies have reported the rapid increase in isoflavonoid levels upon pathogen attack in soybean (Graham 1991; Graham 1995; Lozovaya et al. 2004). The synthesis of isoflavonoids, particularly the pterocarpan glyceollins or active aglycones such as daidzein, formononetin, and genistein [reviewed in (Jeandet et al. 2013)], during defense response is controlled both by the host (Hsieh and Graham 2001) and the pathogen (Rivera-Vargas et al. 1993). Synthesis and release of phytoalexins can limit pathogen colonization, induce toxicity, and increase plant resistance. The synthesis and function of glyceollin have been extensively studied. This pterocarpan phytoalexin inhibits the growth of a wide range of soybean pathogens such as *Phytophthora sojae*, *Sclerotinia sclerotiorum,* and *Macrophomina phaseolina* (Lygin et al. 2010). This has been supported experimentally, in soybean, by silencing isoflavone synthase (IFS) and chalcone synthase (CHS) expression, which reduced the levels of isoflavone daidzein, a glyceollin precursor, and decreased levels of glyceollin, leading to increased

susceptibility to pathogens including the root rot oomycete *P. sojae* (Subramanian et al. 2005; Graham et al. 2007) and the fungal pathogen *Fusarium virguliforme* (Lygin et al. 2010).

The rapid accumulation of glyceollin and other phytoalexins is imperative to limiting colonization by pathogens. The speed of synthesis is important, as pathogens might be capable of degrading phytoalexins, avoiding or limiting the efficacy of these plant defense compounds. Higher ability to degrade or modify phytoalexins, such as glyceollin, into innocuous compounds in *Diaprothe phaseolorum* var. *meridionales* and *Rhizoctonia solani,* results in faster growth (Lygin et al. 2010). The challenge for soybean breeders is to enhance the production and composition of glyceollins in response to pathogenic stimuli. This would improve soybean resistance to a broad range of plant pathogen infections.

Another form of plant-environment interaction mediated by isoflavonoids is the symbiosis between legumes and nitrogen-fixing bacteria. Atmospheric nitrogen represents a vast inaccessible resource for plants. Legumes have developed symbioses with nitrogen-fixing rhizobia to circumvent this problem (Mulligan and Long 1985). Isoflavonoids are key factors in this interaction as signaling molecules that induce rhizobial nodulation genes, leading to bacterial attachment to root hair cells of the legume plant (Phillip 1992). The subsequent cascade of events includes invagination of the root hair cell by a bacterial infection thread and the concurrent development of the root nodule, a specialized organ that allows for an exchange of fixed nitrogen in the form of ammonia from the rhizobia (Spaink 2000). Treatment of soybean roots with *Bradyrhizobium japonicum* induces the expression of the key isoflavonoid gene *IFS* encoding the enzyme catalyzing the first step of the pathway (Subramanian et al. 2006). Evidence for the involvement of isoflavonoids in root nodulation came from RNA interference (RNAi) silencing of *IFS* genes in soybean hairy root composite plants, which led to severely reduced nodulation (Subramanian et al. 2006).

The legume-rhizobial symbiotic relationship mediated by isoflavonoids allows the plants to acquire nitrogen indirectly from the atmosphere, and to reduce the economic and environmental burden imposed by the use of nitrogenous fertilizers.

1.3 Isoflavonoid synthesis and accumulation in soybean tissues

1.3.1 Structure and diversity

The isoflavone aglycones are the most bioactive forms of isoflavonoids with phytoestrogenic capability in animals including humans (Folman and Pope 1969; Kudou et al. 1991). Soybean seeds contain nine different isoflavonoids (Figure 1.1): the three core isoflavone aglycones daidzein, genistein, and glycitein, their corresponding 7-O-β-D glycosides (daidzin, genistin, and glycitin), and 6″-O-malonyl-7-O-β-D glycosides (malonyldaidzin, malonylgenistin, and malonylglycitin) (Kudou et al. 1991). The malonyl glycosides are thermally unstable and, as a result of processing, are converted into their corresponding acetylglycosides (6″-O-acetyldaidzin, 6″-O-acetylgenistin, and 6″-O-acetylglycitin) commonly found in soy food products, such as tofu and soy sauce (Griffith and Collison 2001). Glycosylation and malonylation of soybean isoflavones allows their compartmentalization into vacuoles for storage owing to their stability and solubility. The amount of each isoflavone varies considerably in soybean seeds, but in the majority of cases they are found in the ratio of 4:5:1 (daidzein: genistein: glycitein) (Wang and Murphy 1994). The majority of soybean seed isoflavonoids accumulate in the form of their malonyl derivatives (Figure 1.2). Plant vacuoles appear to have an unknown mechanism that confers stability to the malonyl isoflavone conjugates *in planta*.

1.3.2 Overview of biosynthesis

Isoflavonoids are synthesized through a legume-specific branch of the phenylpropanoid pathway (Figure 1.3). Biosynthesis begins with the recruitment of phenylalanine from the shikimate pathway. Once phenylalanine has been recruited to the phenylpropanoid pathway it is deaminated into cinnamate by the enzyme phenylalanine ammonia-lyase (PAL). The emanating branches produce a plethora of specialized metabolites, including flavonoids, anthocyanins, stilbenoids, lignin and isoflavonoids (Figure 1.3).

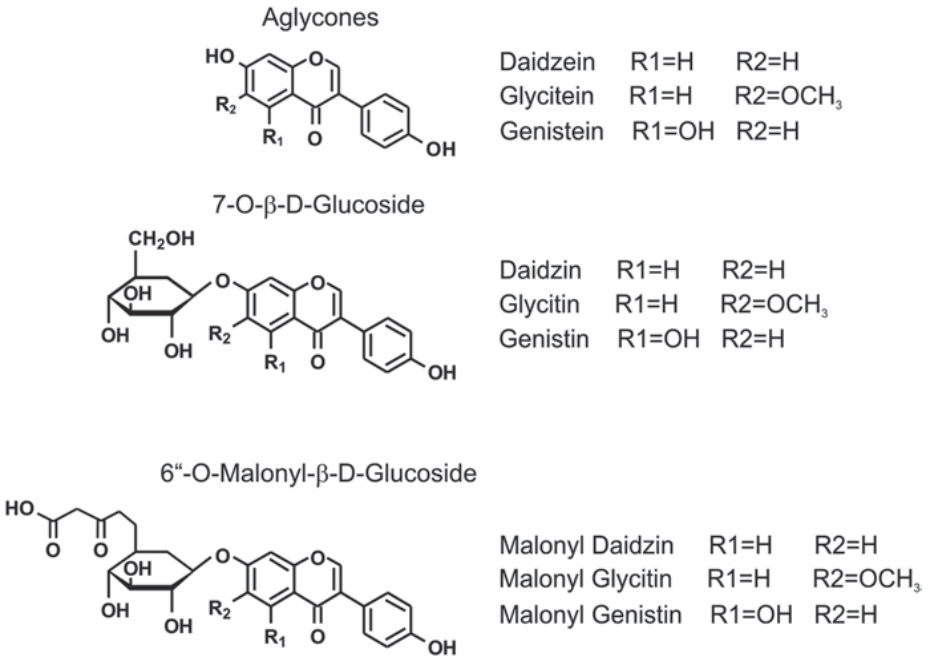


Figure 1.1 Chemical structure of nine major isoflavonoids identified in soybean tissues; adapted from: (Dastmalchi and Dhaubhadel 2014)

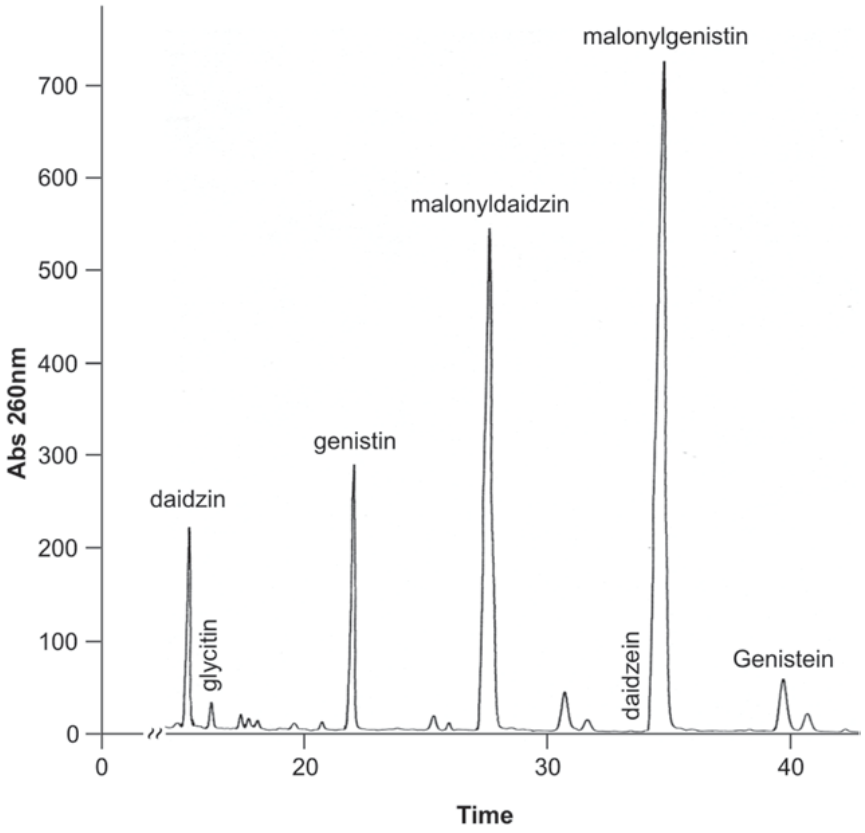


Figure 1.2 Content and composition of soybean seed isoflavonoids. A representative HPLC chromatogram displaying soybean seed isoflavonoids; majority of isoflavonoids accumulate in the form of their malonyl derivatives; adapted from: (Dastmalchi and Dhaubhadel 2014)

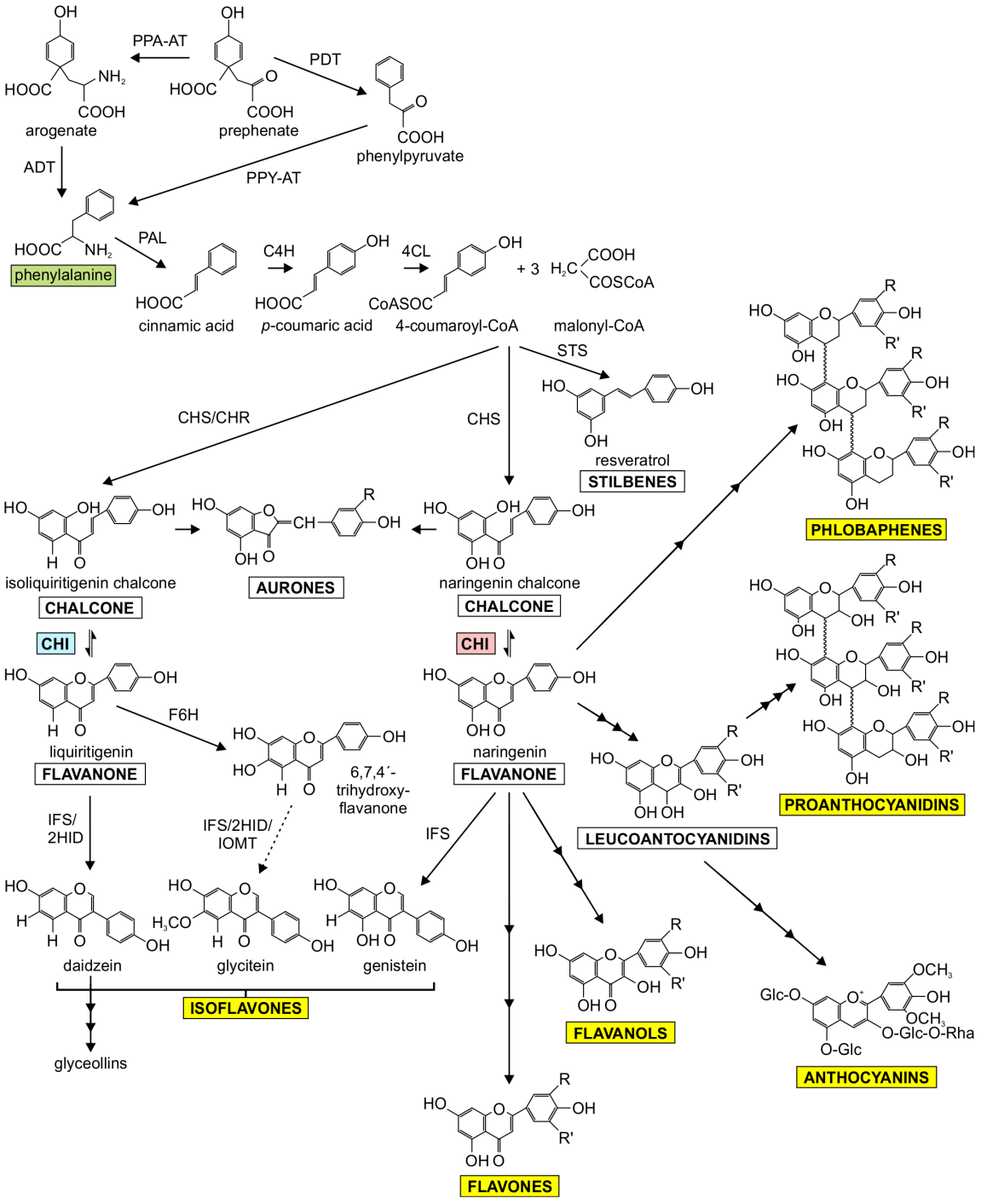

PPA-AT
PDT
arogenate
prephenate
phenylpyruvate
ADT
PPY-AT
phenylalanine
PAL
C4H
4CL
+ 3
cinnamic acid
p-coumaric acid
4-coumaroyl-CoA
malonyl-CoA
STS
CHS/CHR
CHS
resveratrol
STILBENES
isoliquiritigenin chalcone
CHALCONE
AURONES
naringenin chalcone
CHALCONE
PHLOBAPHENES
CHI
CHI
liquiritigenin
FLAVANONE
F6H
6,7,4´-trihydroxy-flavanone
naringenin
FLAVANONE
PROANTHOCYANIDINS
LEUCOANTOCYANIDINS
IFS/2HID
IFS/2HID/IOMT
IFS
daidzein
glycitein
genistein
ISOFLAVONES
glyceollins
FLAVANOLS
ANTHOCYANINS
FLAVONES

Figure 1.3 Phenylalanine channeled into the general phenylpropanoid pathway from the shikimate pathway, producing isoflavones and other specialized metabolites. The dotted arrows represent speculative steps and multiple arrows indicate two or more steps in the pathway. Only the enzymes involved in the isoflavonoid biosynthesis are shown. Chalcone isomerase (CHI) is highlighted in blue (type II) and red (type I); metabolites derived directly or indirectly from flavanones are highlighted in yellow. PPA-AT, prephenate aminotransferase; PDT, prephenate dehydratase; PPY-AT, phenylpyruvate aminotransferase; ADT, arogenate dehydratase; PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; CHR, chalcone reductase; CHS, chalcone synthase; IFS, isoflavone synthase; 2HID, 2-hydroxyisoflavanone dehydratase; IOMT, isoflavone O-methyltransferase; F6H, flavonoid 6-hydroxylase.

The first committed step in the flavonoid and isoflavonoid pathways is the formation of a chalcone, by the plant-specific polyketide synthase, CHS, via the condensation of three malonyl-CoA molecules and *p*-coumaroyl-CoA. All flavonoids and isoflavonoids are built from this chalcone nucleus. The soybean genome contains nine *CHS* genes (*GmCHS1–GmCHS9*) (Dhaubhadel et al. 2007). Members of the *GmCHS* gene family are expressed differentially during plant development and/or in response to environmental cues. GmCHS7 and GmCHS8 are critical for isoflavonoid biosynthesis and accumulation in soybean (Yi et al. 2010; Dhaubhadel et al. 2007).

Chalcone synthase is capable of producing tetrahydroxy chalcone (naringenin chalcone) or through coupled activity with the legume-specific chalcone reductase (CHR), trihydroxy chalcone (isoliquiritigenin chalcone). The product of the CHS/CHR coaction, isoliquiritigenin chalcone, allows legumes to produce novel-derived compounds, such as daidzein and glyceollins. The latter is an important phytoalexin, increasing soybean resistance to pathogen infection (Graham et al. 1990). The chalcone nucleus is further converted by chalcone isomerase (CHI) into a flavanone, with naringenin being a substrate for flavonoid and isoflavonoid metabolism, and isoliquiritigenin going into the legume-specific isoflavonoid branch (Ralston et al. 2005).

A pair of enzymes, 2-hydroxyisoflavanone synthase (IFS) and 2-hydroxyisoflavanone dehydratase (2-HID), catalyze sequential reactions converting the flavanones into isoflavones (Akashi et al. 1999; Jung et al. 2000; Shimamura et al. 2007). Two *IFS* genes, *GmIFS1* and *GmIFS2*, have been found in soybean, both encoding enzymes with the catalytic ability to mediate the production of isoflavones (Jung et al. 2000). GmIFS1 and GmIFS2 differ by 14 amino acid residues and show differential tissue-specific expression in soybean (Dhaubhadel et al. 2003). IFS has been proven necessary and sufficient for introducing isoflavonoid production in non-legumes, albeit at lower levels (Liu et al. 2002; Sreevidya et al. 2006; Liu et al. 2007; Shih et al. 2008).

Soybean isoflavone aglycones can undergo conjugation, resulting in their corresponding glycoside and malonyl-glycoside derivatives. Conjugation with a glucose or malonyl-glucose molecule confers the aglycones with enhanced water solubility and reduced

chemical reactivity, thereby altering the physiological activity and allowing their proper storage in cell vacuoles (Jones and Vogt 2001).

Enzymes of the isoflavonoid pathway, as with those of the phenylpropanoid pathway at large, are part of multi-gene families, owing to the polyploid nature of soybean. Therefore, there is added complexity in trying to differentiate between isoforms, to determine specificity to biosynthetic pathways, and/or tissues.

1.3.3 Accumulation of isoflavonoids in soybean

Isoflavonoids accumulate in all organs of soybean plants, but their levels vary depending on the tissue type and developmental stage. The highest concentrations of isoflavonoids are found in mature seeds and leaves (Dhaubhadel et al. 2003). Soybean embryos express all the key biosynthetic genes involved in isoflavonoid biosynthesis, demonstrating the possibility for *de novo* metabolite biosynthesis (Dhaubhadel et al. 2003). However, there is strong evidence indicating that many plant natural products are transported from their site of synthesis to the site of accumulation. For example, glucosinolates in *Brassica* species are transported into seeds from their production sites (Du and Halkier 1998; Gijzen et al. 1989). Similarly, nicotine is synthesized in roots and gets transported to leaves in tobacco (Shitan et al. 2009). Therefore, it was not surprising that soybean seeds accumulate the largest amount of isoflavonoids, since the developing seed serves as a sink organ for many products synthesized in other plant tissues.

In the case of soybean isoflavonoids, the following indirect evidence suggests the long-distance transport of isoflavonoids within the plants: (a) embryos excised from soybean seeds accumulated isoflavonoids from a synthetic medium, (b) a maternal effect on seed isoflavonoid level was observed in the reciprocal crosses between two soybean cultivars with different seed isoflavonoid levels, and (c) [^{14}C] malonylgenistin and malonyldaidzin accumulated in mature embryos and leaves when fed through the stem (Dhaubhadel et al. 2003). However, the long-distance transport of isoflavone glucosides has not been demonstrated directly yet. Taken together, total isoflavonoid accumulation in soybean seed is a result of both *de novo* synthesis within the seed and transport from maternal tissues.

1.4 Chalcone isomerase: the 'perfect' enzymatic fold

A key branch-point enzyme of the phenylpropanoid pathway, CHI, catalyzes the reaction producing flavanones, the skeletal backbone for many downstream metabolites. This reaction is conducted by the cyclization of either naringenin chalcone to naringenin (common to all vascular plants) or isoliquiritigenin chalcone to liquiritigenin (legume-specific) (Figure 1.3). Naringenin is channelled into the synthesis of flavonoids such as flavones, flavonols, condensed tannins and anthocyanins, as well as the isoflavone, genistein. Liquiritigenin, however, leads only into the legume-specific isoflavonoid pathway and is converted by the action of IFS and other enzymes into glycitein and daidzein, the latter being the precursor to glyceollins.

The CHI fold has been dubbed a 'perfect enzyme', catalyzing an intramolecular and stereospecific cyclization that approaches the diffusion-controlled limit (Jez and Noel 2002). This fold is speculated to have evolved adaptively from fungal and bacterial CHI progenitors that lack catalytic and structural elements of *bona fide* CHI enzymes (Jez et al. 2000; Gensheimer and Mushegian 2004). Vascular plants, and legumes in particular, have a spectrum of genes annotated as encoding for CHI or CHI-like enzymes, which are indicative of the molecular evolution of this enzyme. These enzymes, fall into four types (I-IV) based on their catalytic ability.

Evolution from the fungal precursors to photosynthetic organisms appears to coincide with the evolution of a CHI fold occupied by fatty-acids, leading to 'type III' CHIs. These have been studied in Arabidopsis, as the fatty acid binding proteins (FAPs) and have been shown to affect fatty acid biosynthesis in plant cells and storage in the developing embryo (Ngaki et al. 2012). The secondary structure elements surrounding the fatty acid pocket underwent another shift in the evolutionary timeline, giving rise to CHI-like (CHIL) enzymes. This isoform again lacks chalcone-to-flavanone catalytic ability, but has more sterically restricted pockets surrounding the ligand-binding clefts, and has several polar residues that in conjunction interact with the substrates through hydrogen bonds (Ngaki et al. 2012). This forms an enzyme structure that can be broadly super-imposed on the *bona fide* CHI structure, however, without conservation of the active site residues. The CHIL

group makes up the ‘type IV’ CHIs; in contrast with the ‘type III’ CHIs they are found only in land plants and their function is, as yet, unknown. Work in a number of flowering plants indicates a role in enhancing flavonoid production (Morita et al. 2014).

The next proposed evolutionary step, would be the development of ‘type I’ CHIs, common to all vascular plants, and containing all the necessary active site residues for the conversion of naringenin chalcone to naringenin. The speculated final step that would lead to a putatively isoflavonoid-specific, leguminous ‘type II’ CHI, with the additional capacity to catalyze the conversion of isoliquiritigenin chalcone to liquiritigenin, remains unknown. ‘Type II’ CHIs, studied in legumes, appear to have a substitution at one of four critical active site residues, which is speculated to increase the catalytic ability to convert the additional chalcone substrate, isoliquiritigenin chalcone (Ralston et al. 2005). This would make the enzymatic fold more amenable to isoflavonoid biosynthesis.

The current evolutionary landscape has come about from an ancestral CHI-like homolog that was recruited into fatty-acid synthesis, later morphed into assuming polar substrates through secondary structural changes, and finally developed into a CHI fold that catalyzes the ‘almost perfect’ cyclization of chalcones to chiral flavanones.

1.5 The isoflavonoid metabolon

1.5.1 Metabolon: a multi-enzyme complex

Plant production of specialized metabolites is differentiated and partitioned at many levels: structurally, at the tissue, organ, cellular, subcellular (organelle) and molecular levels; temporally, at different stages of plant development; and in response to various environmental stimuli. This extent of specialization speaks to the complexity of multi-cellular, sessile organisms such as plants.

First defined by Paul Srere in 1985 (Srere 1985), the concept of the ‘metabolon’ extends our idea of compartmentalization to a molecular level, wherein enzymes and other elements, structural or regulatory, aggregate sequentially, with greater metabolic

efficiency and/or output. Such an enzyme complex is advantageous in a subcellular context, as it a) creates pockets of high concentration of intermediates, overcoming kinetic constraints, b) decreases cytoplasmic diffusion of intermediates or enzymes and protects them from extraneous reactions c) coordinates flux of intermediates or enzymes that might overlap with other pathways (Ralston and Yu 2006). The latter point is particularly pertinent, when considering the complexity and inherently shared nature of plant metabolite biosynthesis.

Multi-enzyme complexes associated with structural elements have been reported for primary and specialized metabolism in prokaryotic and eukaryotic cells (Srere 1987). The tryptophan synthase complex is a very simple early 'metabolon', identified in the bacteria *Salmonella typhimurium*, where an intermediate indole is channelled from the active site of the alpha to the beta subunit, through a 25 Å hydrophobic tunnel (Anderson et al. 1991; Anderson et al. 1995). More elaborate and multi-factored enzyme complexes have been demonstrated for the Calvin-Benson cycle (Graciet et al. 2004); the dhurrin biosynthetic pathway (Jensen et al. 2011; Nielsen et al. 2008); and more recently in the sporopollenin pathway in *Arabidopsis* tapetum cells (Lallemand et al. 2013). The metabolons in a majority of cases are characterized by the dynamic interactions between soluble enzymes, structural elements, regulatory and intermediary proteins, and membrane-bound proteins. Growing experimental evidence has described the endoplasmic reticulum (ER) as a platform for specialized plant natural product synthesis, with nucleating cytochrome P450 (P450) enzymes, anchoring these metabolons to specific regions (Winkel 2004a).

1.5.2 Cytochrome P450: a metabolon anchor

Cytochrome P450s (P450) are a superfamily of heme-containing enzymes involved in the metabolism of endogenous, and in some cases xenobiotic compounds (Neve and Ingelman-Sundberg 2008). Briefly, the common mode of action for cytochrome P450s is a monooxygenase reaction, dependent on the transfer of electrons from NADPH. This involves the insertion of one atom of oxygen into the aliphatic position of an organic substrate (RH), while the other oxygen atom is reduced to water (Nelson 2009).

P450s are mainly localized to the ER, where in some cases they are thought to anchor metabolons. Some families of P450s have been shown to localize to other subcellular locations: mitochondria (Neve and Ingelman-Sundberg 1999), plasma membrane (Robin et al. 2000), and lysosomes (Ronis et al. 1991; Loeper et al. 1993). In plants the majority of P450s investigated with relation to metabolic pathways are integral membrane proteins anchored in the ER by a stretch of hydrophobic amino acids that form an α-helix at the N-terminus of the protein. The P450 thusly lodged into the ER membrane reveals its cytosolic, catalytic face to interacting proteins (Neve and Ingelman-Sundberg 2008). The molecular architecture of P450s provides a metabolic center for organizing larger macromolecular complexes. Therefore, it is appropriate to imagine the surface of the ER as being studded by P450 anchors that aggregate metabolically relevant enzymes into specialized regio-specific complexes.

1.5.3 Phenylpropanoid enzyme complexes

The phenylpropanoid pathway is a large and diverse biosynthetic network that produces a wide array of specialized metabolites. There are several P450s involved in the different branches of this pathway, including the second enzyme in the pathway, cinnamate-4-hydroxylase (CYP73A, C4H) that forms *p*-coumarate, and the legume-specific isoflavone synthase (CYP93C, IFS). The idea that enzymes in the phenylpropanoid pathway might channel metabolites through close interaction was advanced by Helen A. Stafford in her pioneering work on 4-hydroxycinnamate hydroxylase in sorghum (Stafford 1974).

Extensive work since has shown that several of the enzymes in the phenylalanine-derived pathways interact, and there has been some evidence to corroborate P450s as nucleating factors for their respective metabolons. Hrazdina and Wagner (Wagner and Hrazdina 1984; Hrazdina and Wagner 1985) were the first to provide substantial evidence for enzyme complexes in phenylpropanoid metabolism, with the channelling of intermediates between phenylalanine ammonia-lyase (PAL) and C4H, the first and second enzymes in the pathway, respectively. Numerous other studies have added to the concept of the P450 anchor, with cytoplasmic flavonoid enzymes associating as well via protein-protein interactions. These interactions often extend beyond consecutive enzymes in the pathway, describing a model of cohesive interactions tethered to the ER. The flavonoid metabolon

has been the focus for the majority of these studies, uncovering interactions between key enzymes in the pathway, including CHS, CHI, flavanone 3-hydroxylase (F3H), dihydroflavonol reductase (DFR) and flavonol synthase (FLS1) (Hrazdina and Wagner 1985; Burbulis and Winkel-Shirley 1999; Saslowsky and Winkel-Shirley 2001; Winkel-Shirley 2001, 2002; Saslowsky et al. 2005).

The model of a legume-specific isoflavonoid metabolon (Figure 1.4), anchored by IFS, has been surprisingly less researched. In fact, the isoflavonoid pathway is a great example of extensive overlap and competition for metabolites and enzymes, including CHI and its products, the flavanones. Therefore, an isoflavonoid metabolon that exists transiently or aggregates around an existing core, which includes the P450, IFS, would share a number of upstream enzymes and/or metabolites with the different branches of the general phenylpropanoid pathway. Investigating the protein dynamics (localization and interaction) within a proposed metabolon consisting of isoflavonoid pathway enzymes and other structural and accessory elements would further our understanding of biosynthesis. The context of large gene families encoding enzymes such as CHI creates an additional layer of complexity that can be unravelled with the study of enzymatic isoforms to determine involvement in the biosynthetic machinery.

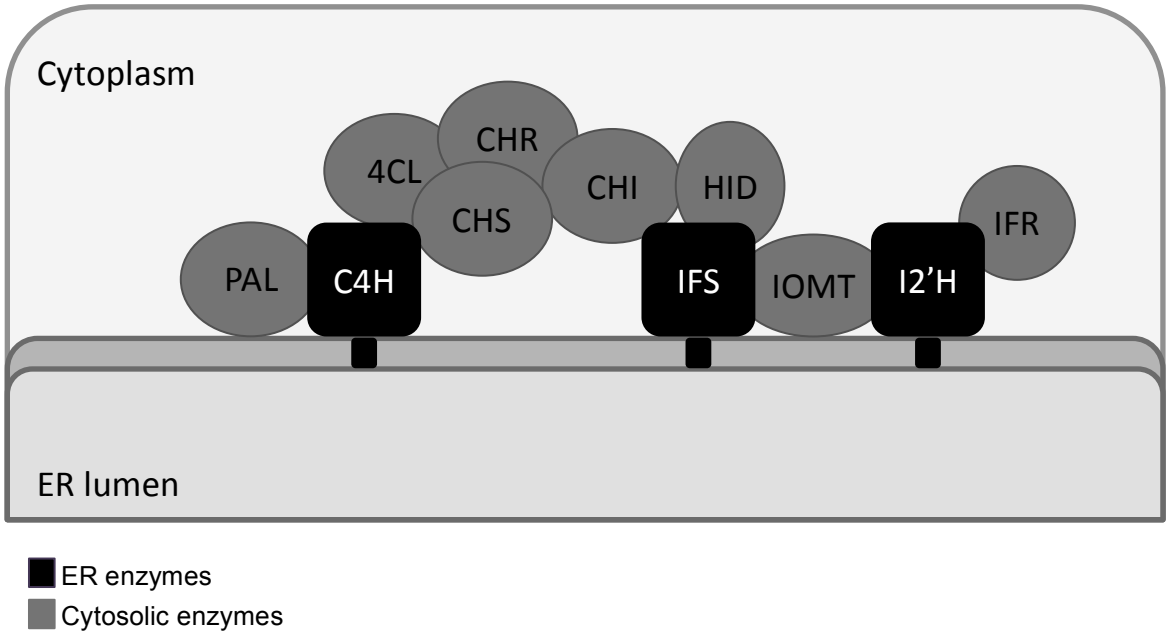


Figure 1.4 Proposed model of isoflavonoid biosynthesis; modified from (Ralston and Yu 2006); PAL (phenylalanine ammonia lyase); C4H (cinnamate 4-hydroxylase); 4CL (4-coumarate-CoA ligase); CHS (chalcone synthase); CHR (chalcone reductase); CHI (chalcone isomerase); IFS (isoflavone synthase); HID (2-hydroxyisoflavanone dehydratase); IOMT (isoflavone O-methyltransferase); I2'H (isoflavone 2'-hydroxylase); IFR (isoflavone reductase)

1.6 Objectives of study

The overall objective of this study is to investigate the molecular and genetic mechanisms of isoflavonoid biosynthesis in soybean, with a particular focus on the roots, due to its physiological and agronomic importance. As the genome of soybean has been sequenced and extensively annotated it is possible to gain a better understanding of large gene families encoding key pathway enzymes. The genes encoding the isoflavonoid pathway have been partially characterized, but research has not encompassed the complexity of gene families. Additionally, the interactions of enzymes in the pathway have not been investigated to substantiate the proposed metabolon. Finally, our scope has been limited to the genes encoding the major enzymes in the pathway and the advent of next-generation sequencing promises to widen our lenses significantly.

Therefore, the objectives of this study are as follows:

A. **To identify the soybean *CHI* gene family and to determine the evolutionary, genetic and subcellular characteristics of its members**.

   This will include a study of the evolutionary history of the chalcone isomerase enzymatic fold by phylogenetic analysis across species in the plant and fungal kingdoms; a complete transcript profile of the 'core' gene family members at organ and tissue levels; an investigation into CHI subcellular localization to gain insight at the molecular level.

B. **To investigate the molecular basis for an isoflavonoid metabolon in soybean.**

   A study will be conducted of the subcellular localization of key members of the isoflavonoid pathway. Protein-protein interactions, within the metabolon, will be analyzed by Bi-molecular Fluorescence Complementation (BiFC) assays. The scope of metabolon studies will be extended by a Co-IP and LC-MS/MS approach to identifying interacting partners of the P450 'anchor'.

C. **To study the genetic basis of isoflavonoid content variation between cultivars.** Cultivars containing high and low levels of isoflavonoids in the roots will be selected based on HPLC analysis and the respective transcriptomes will be analyzed by RNA sequencing, as a means of identifying genetic factors of metabolite variation.

1.7 Literature Cited

Agrawal GK, Hajduch M, Graham K, Thelen JJ (2008) In-Depth Investigation of the Soybean Seed-Filling Proteome and Comparison with a Parallel Study of Rapeseed. Plant Physiol 148 (1):504-518

Akashi T, Aoki T, Ayabe S-i (1999) Cloning and Functional Expression of a Cytochrome P450 cDNA Encoding 2-Hydroxyisoflavanone Synthase Involved in Biosynthesis of the Isoflavonoid Skeleton in Licorice. Plant Physiol 121 (3):821-828

Anderson KS, Kim AY, Quillen JM, Sayers E, Yang XJ, Miles EW (1995) Kinetic characterization of channel impaired mutants of tryptophan synthase. J Biol Chem 270 (50):29936-29944

Anderson KS, Miles EW, Johnson KA (1991) Serine modulates substrate channeling in tryptophan synthase. A novel intersubunit triggering mechanism. J Biol Chem 266 (13):8020-8033

Burbulis IE, Winkel-Shirley B (1999) Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway. Proc Natl Acad Sci U S A 96 (22):12929-12934

Cederroth CR, Nef S (2009) Soy, phytoestrogens and metabolism: A review. Mol Cell Endocrinol 304 (1-2):30 - 42

Chen AM, Rogan WJ (2004) Isoflavones in soy infant formula: a review of evidence for endocrine and other activity infants. Annu Rev Nutr 24:33-54

Dastmalchi M, Dhaubhadel S (2014) Soybean Seed Isoflavonoids: Biosynthesis and Regulation. In: Jetter R (ed) Phytochemicals – Biosynthesis, Function and Application, vol 44. Recent Advances in Phytochemistry. Springer International Publishing, pp 1-21

Dhaubhadel S, Gijzen M, Moy P, Farhangkhoee M (2007) Transcriptome analysis reveals a critical role of *CHS7* and *CHS8* genes for isoflavonoid synthesis in soybean seeds. Plant Physiol 143 (1):326-338

Dhaubhadel S, McGarvey BD, Williams R, Gijzen M (2003) Isoflavonoid biosynthesis and accumulation in developing soybean seeds. Plant Mol Biol 53 (6):733-743

Dixon RA (2004) Phytooestrogen. Annu Rev Plant Biol 55:225-261

Dixon RA, Ferreria D (2002) Genistein. Phytochemistry 60:205-211

Doerge DR, Sheehan DM (2002) Goitrogenic and estrogenic activity of soy isoflavones. Environ Health Perspect 110 (Supplement 3):349–353

Du L, Halkier BA (1998) Biosynthesis of glucosinolates in the developing silique walls. Phytochemistry 48:1145-1150

Duffy C, Perez K, Partridge A (2007) Implications of Phytoestrogen Intake for Breast Cancer. CA: A Cancer Journal for Clinicians 57:260-277

FAOSTAT (2014) Food and Agriculture Organization of the UN. http://faostat.fao.org/default.aspx.

Farag MA, Huhman DV, Dixon RA, Sumner LW (2008) Metabolomics reveals novel pathways and differential mechanistic and elicitor-specific responses in phenylpropanoid and isoflavonoid biosynthesis in Medicago truncatula cell cultures. Plant Physiol 146 (2):387-402

Folman Y, Pope GS (1969) Effect of norethisterone acetate, dimethylstilboestrol, genistein and coumestrol on uptake of [3H]oestradiol by uterus, vagina and skeletal muscle of immature mice. J Endocrinol 44:213-218

Gensheimer M, Mushegian A (2004) Chalcone isomerase family and fold: no longer unique to plants. Protein Sci 13:540-544

Gijzen M, McGregor I, Seguin-Swart G (1989) Glucosinolate uptake by developing rapeseed embyos. Plant Physiol 89:260-263

Graciet E, Lebreton S, Gontero B (2004) Emergence of new regulatory mechanisms in the Benson-Calvin pathway via protein-protein interactions: a glyceraldehyde-3-phosphate dehydrogenase/CP12/phosphoribulokinase complex. J Exp Bot 55 (400):1245-1254

Graham TL (1991) Flavonoid and Isoflavonoid Distribution in Developing Soybean Seedling Tissues and in Seed and Root Exudates. Plant Physiol 95 (2):594-603

Graham TL (1995) Cellular biochemistry of phenylpropanoid responses of soybean to infection by *Phytophthora sojae*. In: Daniel M, Purkayanstha RP (eds) Handbook of phytoalexin metabolism and action. Marcel Dekker, New York, pp 85-116

Graham TL, Graham MY, Subramanian S, Yu O (2007) RNAi Silencing of Genes for Elicitation or Biosynthesis of 5-Deoxyisoflavonoids Suppresses Race-Specific Resistance and Hypersensitive Cell Death in Phytophthora sojae Infected Tissues. Plant Physiol 144:728-740

Graham TL, Kim JE, Graham MY (1990) Role of Constitutive Isoflavone Conjugates in the Accumulation of Glyceollin in Soybean Infected with Phytophthora-Megasperma. Molecular Plant Microbe Interactions 3 (3):157-166

Griffith AP, Collison MW (2001) Improved methods for the extraction and analysis of isoflavones from soy-containing foods and nutritional supplements by reversed-phase high-performance liquid chromatography and liquid chromatography-mass spectrometry. J Chromatogr A 913 (1-2):397-413

Guo J, Xu X, Rasheed TK, Yoder A, Yu D, Liang H, Yi F, Hawley T, Jin T, Ling B, Wu Y (2013) Genistein interferes with SDF-1- and HIV-mediated actin dynamics and inhibits HIV infection of resting CD4 T cells. Retrovirology 10:62

Hajduch M, Ganapathy A, Stein JW, Thelen JJ (2005) A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. Plant Physiol 137 (4):1397-1419

Hill JE, Breidenbach RW (1974) Proteins of soybean seeds: I. Isolation and characterization of the major components. Plant Physiol 53 (5):742-746

Howes LG, Howes JB, Knight DC (2006) Isoflavone therapy for menopausal flushes: a sytematic review and meta-analysis. Maturitas 55:203-211

Hrazdina G, Wagner GJ (1985) Metabolic pathways as enzyme complexes: evidence for the synthesis of phenylpropanoids and flavonoids on membrane associated enzyme complexes. Arch Biochem Biophys 237 (1):88-100

Hsieh MC, Graham TL (2001) Partial purification and characterization of a soybean beta-glucosidase with high specific activity towards isoflavone conjugates. Phytochemistry 58 (7):995-1005

Jeandet P, Clement C, Courot E, Cordelier S (2013) Modulation of phytoalexin biosynthesis in engineered plants for disease resistance. Int J Mol Sci 14 (7):14136-14170

Jensen K, Osmani SA, Hamann T, Naur P, Moller BL (2011) Homology modeling of the three membrane proteins of the dhurrin metabolon: catalytic sites, membrane

surface association and protein-protein interactions. Phytochemistry 72 (17):2113-2123

Jez JM, Bowman ME, Noel JP (2000) Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. Nat Struc Mol Biol 7:786-791

Jez JM, Noel JP (2002) Reaction mechanism of chalcone isomerase. J Biol Chem 277 (2):1361-1369

Jones P, Vogt T (2001) Glycosyltransferases in secondary plant metabolism: transquilizers and stimulant controllers. Planta 213:164-174

Jones SI, Vodkin LO (2013) Using RNA-Seq to profile soybean seed development from fertilization to maturity. PLoS One 8 (3):e59270

Jung W, Yu O, Lau SMC, O'Keefe DP, Odell J, Fader G, McGonigle B (2000) Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes Nat Biotechnol 18 (5):559-559

Kudou S, Fleury Y, Welt D, Magnolato D, Uchida T, Kitamura K (1991) Malonyl isoflavone glycosides in soybean seeds (*Glycine max* Merill). Agric Biol Chem 55:2227-2233

Lallemand B, Erhardt M, Heitz T, Legrand M (2013) Sporopollenin biosynthetic enzymes interact and constitute a metabolon localized to the endoplasmic reticulum of tapetum cells. Plant Physiol 162 (2):616-625

Lamartiniere CA (2000) Protection against breast cancer with genistein: a component of soy. Am J Clin Nutr 71 (6):1705S-1707S

Lamartiniere CA, Moore JB, Brown NM, Thompson R, Hardin MJ, Barnes S (1995) Genistein suppresses mammary cancer in rats. Carcinogenesis 16:2833-2840

Le BH, Wagmaister JA, Kawashima T, Bui AQ, Harada JJ, Goldberg RB (2007) Using genomics to study legume seed development. Plant Physiol 144 (2):562-574

Lin H, Rao J, Shi J, Hu C, Cheng F, Wilson ZA, Zhang D, Quan S (2014) Seed metabolomic study reveals significant metabolite variations and correlations among different soybean cultivars. J Integr Plant Biol

Liu CJ, Blount JW, Steele CL, Dixon RA (2002) Bottlenecks for metabolic engineering of isoflavone glycoconjugates in Arabidopsis. Proc Natl Acad Sci USA 99 (22):14578-14583

Liu RR, Hu YL, Li HL, Lin ZP (2007) Production of soybean isoflavone genistein in non-legume plants via genetically modified secondary metabolism pathway. Metab Eng 9 (1):1-7

Loeper J, Descatoire V, Maurice M, Beaune P, Belghiti J, Houssin D, Ballet F, Feldmann G, Guengerich FP, Pessayre D (1993) Cytochromes P-450 in human hepatocyte plasma membrane: recognition by several autoantibodies. Gastroenterology 104 (1):203-216

Lozovaya VV, Lygin AV, Zernova OV, Li S, Hartman GL, Widholm JM (2004) Isoflavonoid accumulation in soybean hairy roots upon treatment with *Fusarium solani*. Plant Physiol Biochem 42 (7-8):671-679

Lygin AV, Hill CB, Zernova OV, Crull L, Widholm JM, Hartman GL, Lozovaya VV (2010) Response of soybean pathogens to glyceollin. Phytopathology 100 (9):897-903

Mendez MA, Anthony MS, Arab L (2002) Soy-based formulae and infant growth and development: A review. J Nutr 132 (8):2127-2130

Messina M (2010) Insights gained from 20 years of soy research. J Nutr 140 (12):2289S-2295S

Messina M, Barnes S (1991) The role of soy products in reducing cancer risk. J Natl Cancer Inst: JNCI 83:541-546

Miadokova E (2009) Isoflavonoids - an overview of their biological activities and potential health benefits. Interdiscip Toxicol 2 (4):211-218

Morita Y, Takagi K, Fukuchi-Mizutani M, Ishiguro K, Tanaka Y, Nitasaka E, Nakayama M, Saito N, Kagami T, Hoshino A, Iida S (2014) A chalcone isomerase-like protein enhances flavonoid production and flower pigmentation. Plant J 78:294-304

Mulligan JT, Long SR (1985) Induction of Rhizobium meliloti nodC expression by plant exudate requires nodD. Proc Natl Acad Sci USA 82:6609-6613

Nelson DR (2009) The cytochrome p450 homepage. Human genomics 4 (1):59-65

Neve EP, Ingelman-Sundberg M (1999) A soluble NH(2)-terminally truncated catalytically active form of rat cytochrome P450 2E1 targeted to liver mitochondria(1). FEBS Lett 460 (2):309-314

Neve EP, Ingelman-Sundberg M (2008) Intracellular transport and localization of microsomal cytochrome P450. Anal Bioanal Chem 392 (6):1075-1084

Ngaki MN, Louie GV, Philippe RN, Manning G, Pojer F, Bowman ME, Li L, Larsen E, Wurtele ES, Noel JP (2012) Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. Nature 485 (7399):530-533

Nielsen KA, Tattersall DB, Jones PR, Moller BL (2008) Metabolon formation in dhurrin biosynthesis. Phytochemistry 69 (1):88-98

Ohlrogge JB, Kuo TM (1984) Control of Lipid Synthesis during Soybean Seed Development: Enzymic and Immunochemical Assay of Acyl Carrier Protein. Plant Physiol 74 (3):622-625

Paxton JD (1981) Phytoalexins- A Working Redefinition. Phytopathol Z 101 (2):106-109

Phillip DA (1992) Flavonoids:plant signals to soil microbes. In: Stafford HA, Ibrahim RK (eds) Recent advances in phytochemistry:phenolic metabolism in plants, vol 26. Plenum Press, New York, pp 201-223

Potter SM, Baum JA, Teng H, Stillman RJ, Shay NF, Erdman JW (1998) Soy protein and isoflavones: their effects on blood lipids and bone density in postmenopausal women. Am J Clin Nutr 68 (6):1375S-1379S

Ralston L, Subramanian S, Matsuno M, Yu O (2005) Partial reconstruction of flavonoid and isoflavonoid biosynthesis in yeast using soybean type I and type II chalcone isomerases. Plant Physiol 137 (4):1375-1388

Ralston L, Yu O (2006) Metabolons involving plant cytochrome P450s. Phytochem Rev 5 (2):459-472

Rivera-Vargas LI, Schmitthenner AF, Graham TL (1993) Soybean flavonoid effects on and metabolism by *Phytophthora sojae*. Phytochemistry 32 (4):851-857

Robin MA, Descatoire V, Le Roy M, Berson A, Lebreton FP, Maratrat M, Ballet F, Loeper J, Pessayre D (2000) Vesicular transport of newly synthesized cytochromes P4501A to the outside of rat hepatocyte plasma membranes. The Journal of pharmacology and experimental therapeutics 294 (3):1063-1069

Ronis MJ, Johansson I, Hultenby K, Lagercrantz J, Glaumann H, Ingelman-Sundberg M (1991) Acetone-regulated synthesis and degradation of cytochrome P450E1 and

cytochrome P4502B1 in rat liver [corrected]. European journal of biochemistry / FEBS 198 (2):383-389

Sarkar FH, Li Y (2003) Soy isoflavones and cancer prevention. Cancer Invest 21 (5):744-757

Saslowsky D, Winkel-Shirley B (2001) Localization of flavonoid enzymes in Arabidopsis roots. Plant J 27 (1):37-48

Saslowsky DE, Warek U, Winkel BS (2005) Nuclear localization of flavonoid enzymes in Arabidopsis. J Biol Chem 280 (25):23735-23740

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463 (7278):178-183

Shih C, Chen Y, Wang M, Chu IK, Lo C (2008) Accumulation of isoflavone genistin in transgenic tomato plants overexpressing a soybean isoflavone synthase gene. J Agric Food Chem 56 (14):5655-5661

Shimamura M, Akashi T, Sakurai N, Suzuki H, Saito K, Shibata D, Ayabe SI, Aoki T (2007) 2-Hydroxyisoflavanone dehydratase is a critical determinant of isoflavone productivity in hairy root cultures of Lotus japonicus. Plant and Cell Physiology

Shitan N, Morita M, Yazaki K (2009) Identification of a nicotine transporter in leaf vacuoles of *Nicotiana tabacum*. Plant Signaling & Behavior 4:530-532

Shu XO, Jin F, Dai Q, Wen W, Potter JD, Kushi LH, Ruan Z, Gao YT, Zheng W (2001) Soyfood intake during adolescence and subsequent risk of breast cancer among Chinese women. Cancer Epidemiol Biomarkers Prev 10:483-488

Spaink HP (2000) Root nodulation and infection factors produced by rhizobial bacteria. Annu Rev Microbiol 54:257-288

Sreevidya VS, Rao CS, Sullia SB, Ladha JK, Reddy PM (2006) Metabolic engineering of rice with soybean isoflavone synthase for promoting nodulation gene expression in rhizobia. J Exp Bot 57 (9):1957-1969

Srere PA (1985) The metabolon. Trends Biochem Sci 10 (3):109-110

Srere PA (1987) Complexes of sequential metabolic enzymes. Annu Rev Biochem 56:89-124

Stafford HA (1974) Activation of 4-hydroxycinnamate hydroxylase in extracts from sorghum. Plant Physiol 54 (5):686-689

Strom BL, Schinnar R, Ziegle EE, Barhard KT, Sammel MD, Macones GA, Stallings VA, Drulis JM, Nelson SE, Hanson SA (2001) Exposure to soy-based formula in infancy and endochinological and reproductive outcomes in young adulthood. J Am Med Assoc 286:807-814

Subramanian S, Graham MA, Yu O, Graham TL (2005) RNA interference of soybean isoflavone synthase genes leads to silencing in tissues distal to the transformation site and to enhanced susceptibility to *Phytophthora sojae*. Plant Physiol 137:1345-1353

Subramanian S, Stacey G, Yu O (2006) Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. Plant J 48 (2):261-273

Wagner GJ, Hrazdina G (1984) Endoplasmic reticulum as a site of phenylpropanoid and flavonoid metabolism in hippeastrum. Plant Physiol 74 (4):901-906

Walker HA, Dean TS, Sanders TAB, Jackson G, Ritter JM, Chowienczyk PJ (2001) The Phytoestrogen Genistein Produces Acute Nitric Oxide-Dependent Dilation of Human Forearm Vasculature With Similar Potency to 17 beta-Estradiol. Circulation 103 (2):258-262

Wang H, Murphy PA (1994) Isoflavone Composition of American and Japanese Soybeans in Iowa: Effects of Variety, Crop Year, and Location. J Agric Food Chem 42:1674 - 1677

Winkel BS (2004a) Metabolic channeling in plants. Annu Rev Plant Biol 55:85-107

Winkel BSJ (2004b) Metabolic channeling in plants. Annu Rev Plant Biol 55 (1):85-107

Winkel-Shirley B (2001) It takes a garden. How work on diverse plant species has contributed to an understanding of flavonoid metabolism. Plant Physiol 127 (4):1399-1404

Winkel-Shirley B (2002) Biosynthesis of flavonoids and effects of stress. Curr Opin Plant Biol 5 (3):218-223

Xu L, Ding Y, Catalona WJ, Yang XJ, Anderson WF, Jovanovic b, Wellman K, Killmer J, Huang X (2009) MEK4 function, genistein treatment, and invasion of human prostate cancer cells. The Journal of the National Cancer Institute 101:1141-1155

Yi J, Derynck MR, Chen L, Dhaubhadel S (2010) Differential expression of *CHS7* and *CHS8* genes in soybean. Planta 231:741-753

CHAPTER TWO – SOYBEAN CHALCONE ISOMERASE: EVOLUTION OF THE FOLD, AND THE DIFFERENTIAL EXPRESSION AND LOCALIZATION OF THE GENE FAMILY

2.1 Introduction

The phenylpropanoid pathway produces a diverse array of plant natural products. A key branch-point enzyme, chalcone isomerase (CHI), catalyzes the reaction producing flavanones, the backbone for many downstream metabolites such as flavonoids (all vascular plants) and isoflavonoids (Fabaceae) that contribute to plant signalling, attraction, protection and development (Phillip 1992; Stafford 1997; Aoki et al. 2000; Winkel-Shirley 2001).

Chalcone isomerase (CHI) catalyzes the cyclization of chalcones to flavanones, a common substrate for many downstream metabolites (Figure 1.3). The flavanones produced in legumes include naringenin and liquiritigenin. The latter is almost exclusive to the action of CHIs in legume species and is channelled into the isoflavonoid pathway for the production of the isoflavones daidzein and glycitein. Naringenin is common to the flavonoid and isoflavonoid pathways, producing the isoflavone genistein in the latter.

The CHI fold is speculated to have evolved from fungal and bacterial progenitors that lack catalytic and structural elements of functional CHIs. In the plant kingdom CHIs appear to have developed into fatty-acid binding proteins (FAPs). Further evolution lead to increased steric restriction surrounding the ligand-binding clefts and the addition of several polar residues to the fold. This allowed the chalcone substrate to bind to the enzymatic cleft with the formation of hydrogen bonds and undergo the conversion to flavanones (Ngaki et al. 2012).

A speculated distinction between CHIs in legumes and other vascular plants has been their respective capabilities in converting chalcones to flavanones. In all vascular plants studied CHIs can convert naringenin chalcone to naringenin. It is speculated that certain substitutions of active site residues in legume-specific type ‘II’ CHIs enable the additional

capability to efficiently convert isoliquiritigenin chalcone to liquiritigenin (Ralston et al. 2005). The ability to produce liquiritigenin is important in the production of the isoflavone aglycones, daidzein and glycitein. Whether or not this distinction is truly present *in planta* has not been determined.

The current model of phenylpropanoid biosynthesis is of flavonoid and isoflavonoid branches forming protein complexes, or 'metabolons', at the surface of the endoplasmic reticulum (ER), anchored by cytochrome P450s (Figure 1.4) (Ralston and Yu 2006). The isoflavonoid metabolon is speculated to be one of soluble cytoplasmic enzymes amalgamating at the surface of the ER, anchored by IFS (Ralston and Yu 2006); producing isoflavone aglycones (daidzein, genistein and glycitein), which are conjugated into their corresponding 7-O-glycosides and malonyl-glycosides and transported to and stored in the vacuole (Kudou et al. 1991). This model of cytoplasmic phenylpropanoid synthesis has been questioned, with recent reports pointing to the presence of flavonoids and corresponding enzymes in the nucleus, in a diverse range of species (Yu et al. 2008; Saslowsky et al. 2005; Dhaubhadel et al. 2008). Therefore, I investigated the *in planta* subcellular localization of soybean CHIs (GmCHIs), to characterize leguminous enzymes and explore the possibility of differential targeting of the biosynthetic machinery.

The *CHI* family in soybean constitutes four subfamilies, with only the first two subfamilies (including *GmCHI1A*, *GmCHI1B1*, *GmCHI1B2* and *GmCHI2*) having the catalytic activity to convert chalcone to flavanone. The most recent study on *GmCHIs* enumerated seven members in the family (Ralston et al. 2005; Chen et al. 2011). I identified five additional genes in the *GmCHI* family, members of the 'type III' subgroup. In addition to providing an updated look at the *GmCHI* gene family, I present the evolutionary history of the GmCHI fold, the molecular and spatio-temporal specialization of the CHI superfamily in soybean, and their subcellular localization. Analysis of *GmCHI* gene expression suggested the potential functionality of *GmCHI1A* and *GmCHI1B2* in isoflavonoid production in the leaf and seeds, whereas promoter analysis of the *GmCHIs* highlighted a more detailed picture of sub-organ expression patterns.

2.2 Materials and methods

2.2.1 Plant materials and growth conditions

Soybean (*Glycine max* L. Merr.) cv. Harosoy63 was grown in the field at the Southern Crop Protection and Food Research Centre, London, Ontario, during 2011 and 2012. Tissues studied were harvested as appropriate during different development stages as indicated in section 2.2.4.

Nicotiana benthamiana and *Arabidopsis thaliana* ecotype Col-0 were grown on Pro-Mix BX Mycorrhizae^{TM} soil (Rivière-du-Loup, Canada) in a growth chamber with a 16 h light cycle at 25°C, and an 8 h dark cycle at 20°C, with 60-70% relative humidity, with light intensity 100-150 μmol $m^{-2}s^{-1}$.

Arabidopsis seeds were stratified by planting in soil, and incubation at 4°C for 3 days. For growth of *Arabidopsis* transgenic seeds under sterile and selective media, seeds were surface-sterilized, with a solution of 20% (v/v) bleach (Lavo Inc., Canada) and 0.1% (w/v) sodium dodecyl sulphate (Sigma-Aldrich, Germany), for 10 minutes with gentle shaking every two minutes. Seeds were washed with sterile distilled water five times (or until water was completely transparent) to remove residual bleach. Seeds were sown on 1X Murashige and Skoog (MS) basal salt (*Phyto*Technology Laboratories, USA) supplemented with 3% (w/v) sucrose (Sigma-Aldrich, Germany), 0.8% (w/v) phytoagar (*Phyto*Technology Laboratories, USA), and hygromycin (50 μg/ml).

Nicotiana benthamiana seeds were sprinkled onto wet soil. After two weeks, seedlings were transferred to new pots and watered regularly. The nutrient mixture of nitrogen, phosphorous, and potassium (20-20-20) was applied once a week. Leaves were infiltrated at the second or third trifoliate stage of true leaf development.

2.2.2 Identification of the soybean *CHI* gene family and *in silico* analysis

Accessions from Chen et al., (2009) were used as a query to search *GmCHI* genes by BLASTN analysis (Altschul et al. 1990). Acquired nucleotide sequences were entered into the Soybean Genome Database (www.phytozome.net) (Schmutz et al. 2010) and

queried against the new soybean genome assembly (Wm82.a2.v1) and homologous sequences in the genome were analyzed for sequence similarity and coverage. To cover all gene family members, including distant, ancestral types, a key word search using "chalcone isomerase" was used against the *G. max* Wm82.a2.v1 target in all the protein databases available (Pfam, Panther, KOG, EC, GO, KEGG Orthology, Cluster KEGG Orthology). Predicted amino acid sequences from Phytozome were used to generate an amino acid sequence alignment of the *GmCHI* gene family using Clustal W (McWilliam et al. 2013). Predicted subcellular localization of the proteins encoded by *GmCHI* genes was determined using SignalP 4.1 and TargetP 1.1 (Emanuelsson et al. 2000). Protein molecular weight was predicted using the molecular weight calculator "Bioinformatics Organization" (Stothard 2000).

2.2.3 Phylogenetic analysis

To find homologous sequences across plant divisions, the Phytozome evolutionary species tree was chosen to pick model organisms, from the chlorophyta: *Chlamydomonas reinhardtii*; the bryophyte: *Physcomitrella patens*; the lycopodiophyta: *Selaginella moellendorffii*; non-leguminous angiosperms, *Populus trichocarpa*, *Prunus persica*, *Arabidopsis thaliana, Solanum lycopersicum, Zea mays* and *Oryza sativa*; and finally legume (Fabaceae) relatives *Medicago truncatula* and *Phaseolus vulgaris*. The keyword "chalcone isomerase" was then queried against these databases as before. *GmCHI* sequences and literature resources were also used to BLAST for homologous sequences in the above species, and the fungal species, *Saccharomyces cerevisiae* and *Fusarium graminearum*.

Multiple sequence alignments of the deduced amino acids of CHIs or CHI-like protein regions across the species were obtained using Clustal W (McWilliam et al. 2013). Alignments were imported into Mega5.1 for phylogenetic analysis (Tamura et al. 2011). The evolutionary history was inferred using the Maximum Likelihood method based on the JTT matrix-based model (Tamura et al. 2011). The percentage of trees in which the associated taxa clustered together is indicated next to the branches. The initial trees for the heuristic search were obtained by applying the Neighbour-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was

used to model evolutionary rate differences among sites. The rate variation model allowed for some sites to be evolutionarily invariable. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 85 amino acid sequences. Alignment gaps, missing data and ambiguous bases were allowed at any position.

2.2.4 RNA isolation and qPCR

Total RNA was isolated following the procedure of Wang and Vodkin (1994), from soybean (cv. Harosoy63) tissues (roots, stems, leaves, flower buds, flowers, embryos [30, 40, 50, 60 and 70 DAP], seed coats and pod walls). Total RNA was quantified using a NanoDrop 1000 spectrophotometer (ThermoScientific, USA). Concentration and integrity were confirmed by gel electrophoresis of 0.2 µg RNA on an agarose gel (1% w/v) in 0.5X TAE buffer, and stained with RedSafe nucleic acid staining solution (iNtRON Biotechnology, Korea). Contaminating DNA in RNA samples was removed using the TURBO DNA-free™ Kit (Life Technologies, USA). The DNase-treated RNA (1 µg) was used in a reverse transcription reaction using the ThermoScript™ RT-PCR System (Invitrogen, USA).

For qPCR SsoFastTM EvaGreen® Supermix (Bio-Rad, USA) was used with CFX96 real-time PCR detection system (Bio-Rad, USA). Primer sequences for qPCR are listed in Table 2.1. The amplicons were verified and cloned into pGEMT vector (Promega, USA) according to the manufacturer's instructions and the sequences verified. All qPCR reactions were done in technical triplicates, with a no template control (NTC), and fold expression was normalized to the reference genes *CONS4* and *CONS6* (Libault et al. 2008). The data were analyzed using CFX manager (Bio-Rad, USA).

2.2.5 RNAseq expression profile

A publicly available database containing the Illumina high-throughput transcriptome sequencing of soybean at:
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29163 was used to mine the expression profiles of *CHI* gene family members and other isoflavonoid-specific genes. The RNAseq data were downloaded from the NCBI database

(http://www.ncbi.nlm.nih.gov/) with accession numbers SRX062325-SRX062334. The relative expression was normalized across the libraries corresponding to each tissue. The heatmap for *GmCHI* transcripts and other key isoflavonoid biosynthetic genes was generated in R using the heatmap.2 function from the gplots CRAN library. The following Glyma accession numbers were queried against the database to generate the heatmap: Glyma20g38560, *GmCHI1A*; Glyma20g38570, *GmCHI1B1*; Glyma10g43850, *GmCHI1B2*; Glyma20g38580, *GmCHI2*; Glyma13g33730, *GmCHI3A1*; Glyma15g39050, *GmCHI3A2*; Glyma03g31100, *GmCHI3B1*; Glyma19g33940, *GmCHI3B2*; Glyma06g14820, *GmCHI4A*; Glyma04g40030, *GmCHI4B*; Glyma01g43880, *GmCHS7*; Glyma11g01350, *GmCHS8*; Glyma14g00870, *GmCHR*; Glyma13g24200, *GmIFS2*.

2.2.6 Cloning and plasmid construction

For the subcellular localization study, *GmCHIs* were amplified by PCR using attB1 and attB2 site-containing gateway primers (Table 2.1). The PCR products were run on a 1% (w/v) agarose gel in 0.5X TBE buffer and a single band corresponding to *GmCHI* amplicons was excised from the gel. The DNA was extracted from the gel using EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic Inc., Canada), quantified using a NanoDrop spectrophotometer (ThermoScientific, USA), and recombined with the entry vector pDONR-Zeo (Invitrogen, USA) using BP clonase reaction mix (Invitrogen, USA) according to gateway cloning technology. The BP reaction product was transformed into *E. coli* DH5α by electroporation and grown on LB medium supplemented with zeocin (50 µg/mL). Colonies were screened by PCR, using gene-specific primers, to select colonies carrying the recombinant plasmids. Plasmid DNA was prepared from positive colonies using EZ-10 Spin Column Plasmid DNA Kit (Bio Basic Inc., Canada). The inserted sequences in recombinant plasmids pDONR-Zeo-*GmCHI* were verified by sequencing, using vector-specific M13 forward and reverse primers, and gene-specific primers.

Recombinant plasmids, containing the 'entry clone', were used in an LR recombination reaction with the destination vector pEarleyGate101 (Invitrogen, USA). The LR reaction was transformed into *E. coli* DH5α, grown on LB medium supplemented with kanamycin (50 µg/mL), recombinants were screened by PCR and plasmid DNA was prepared from

positive colonies, as described above for the BP reaction. The plasmid DNA carrying the 'destination clone', pEG101-*GmCHI*, was sequenced to confirm sequence identity, and subsequently transformed into *Agrobacterium tumefaciens* strain GV3101 and *Agrobacterium rhizogenes* strain K599 via electroporation. All *E. coli* and *Agrobacterium* cultures were grown at 37°C and 28°C, respectively.

To determine the subcellular localization of GmCHIs in soybean, the constructs in *A. rhizogenes* strain K599 were transformed into soybean cotyledons to generate transgenic hairy roots. Six day-old soybean cotyledons were harvested and transformed by *A. rhizogenes* strains carrying the constructs using the hairy root transformation system (Subramanian et al. 2004). Transgenic hairy roots were selected 20-30 days post-infiltration, using a Leica MZ FL III fluorescence stereomicroscope, and an YFP filter (excitation 510/20 nm; barrier filter 560/40 nm).

Floral dip transformation method was used to transform *GmCHI* promoter fragments fused with GFP into wild-type *Arabidopsis* Col-0 (Zhang et al. 2006). Plants grown in the chamber (as described in section 2.2.1) were selected for floral dip after 4 weeks, when they had approximately 20-30 inflorescences. Prior to transformation any siliques, already formed, were clipped off the plant.

Agrobacterium colonies containing the destination clone were selected and grown at 28°C for 2 days, in LB medium containing kanamycin (50 μg/mL). This feeder culture was used to inoculate 500 mL of liquid LB medium containing the same concentration of kanamycin. The culture was incubated again until the cells reached the stationary phase (OD 1.5–2.0). *Agrobacterium* cells were collected by centrifugation at 4,000 g for 10 min at room temperature, and gently resuspended into one volume of freshly made 5% wt/vol sucrose solution with a stirring bar. Silwet L-77 was added to a concentration of 0.02% vol/vol. The *Agrobacterium* cell suspension was then transferred to a 500 mL beaker and the *Arabidopsis* plants were inverted and their aerial parts were dipped for 1 min.

Table 2.1 Sequence of oligonucleotides used in present chapter: primers for amplification and gene profiling, subcellular localization of full-length genes, and promoter analysis

Gene	Primer name	Sequence	Amplicon size	Application
GmCHI1A	CHI1A-F	5'-TGGGGTGTTAAATAGAAAAGAGGAGTT-3'	170	qPCR
	CHI1A-R	5'-GAACTTTATGAACTTCCCCTCAATCG-3'		
GmCHI1B1	CHI1B1-F	5'-GAACTTTATGAACTTCCCCTCAATCG-3'	165	
	CHI1B1-R	5'-AACTGGTTTGCTCGTTTATTAAATGGG-3'		
GmCHI1B2	CHI1B2-F	5'-CCACTTTGTTTGAATTTGTACCCC-3'	214	
	CHI1B2-R	5'-TTCCTATTTATGTGCTTTTACCTATAACATG-3'		
GmCHI2	CHI2-F	5'-CCCCTTACTGTCCATGTCTAAGTACTG-3'	193	
	CHI2-R	5'-ACCCATGAAAATGACCAAGACCT-3'		
GmCHI3A1	CHI3A1-F	5'-GCTTAGTGCTAGTAACACAGGCGG-3'	166	
	CHI3A1-R	5'-AGGAATAGGATTGTGTATGCTTATGCC-3'		
GmCHI3A2	CHI3A2-F	5'-TACATCTATATTTTCTTGAAATGGTTCAGTCTC-3'	239	
	CHI3A2-R	5'-GAACTAACACTGATAAAGTGTTTCCTT-3'		
GmCHI4A	CHI4A-F	5'-GAGAATGCCAATGTGGTAGAGGCCA-3'	155	
	CHI4A-R	5'-AAACACAGACCCCAAGATTATTACACC-3'		
GmCHI4B	CHI4B-F	5'-GAGAATGCCAATGTGGTAGAGGCAG-3'	160	
	CHI4B-R	5'-AACACAGATCCCCAAGATTACACAATA-3'		

Table 2.1 Continued

Subcellular localization		
Gene	**Sequence**	**Amplicon size**
GmCHI1A	5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCAACGATCAGCGCGGTTCAG-3’ 5’-GGGGACCACTTTGTACAAGAAAGCTGGGTCGACTATAATGCCGTGGCTCAATAC-3’	715
GmCHI1B1	5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCCACACCAGCATCCATCACT-3’ 5’-GGGGACCACTTTGTACAAGAAAGCTGGGTCGTTTTCAATGTTGGGATTGGCCTC-3’	739
GmCHI1B2	5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCCACACCAGCATCCATCACC-3’ 5’-GGGGACCACTTTGTACAAGAAAGCTGGGTCGTTTTCATTGTTGGGATTGGCCTC-3’	739
GmCHI2	5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCAATGGCATTTCCGTCCGTA-3’ 5’-GGGGACCACTTTGTACAAGAAAGCTGGGTCGTTATGAGATTGAGGGTCACAAAC-3’	745
GmCHI3A1	5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGCCTGGAGCTGTAGCAGCTTCC-3’ 5’-GGGGACCACTTTGTACAAGAAAGCTGGGTCGGGTTTCAGTACTTTTGACAAGCT-3’	907
GmCHI3A2	5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGCCTGGAGCTGTAGCAGCTTCC-3’ 5’-GGGGACCACTTTGTACAAGAAAGCTGGGTCGGGTTTGAGTACTTTTGACAAGCC-3’	907
GmCHI4A&B	5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGCTACTGAAGAGGTTTTGGTTG-3’ 5’-GGGGACCACTTTGTACAAGAAAGCTGGGTCCTTGGACAACTCCTGCGAGAAAGT-3’	687

Table 2.1 Continued

	Promoter Analysis	
CHI1Apro: GUS	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGATTTGTTGACCTGCTATGGC-3'	1519
	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCTCTCAAACTCCTCTTTTCTATTTAACACC-3'	
CHI1B1pro: GUS	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGCAACAGGTGCGTTTGTC-3'	1490
	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCAATAAGTAGTACACCACTAGGACTGTGTC-3'	
CHI1B2pro: GUS	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGACTTAGGATATGTAAGGCAGGTAGG-3'	1560
	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCAATGTGTCAAATGGAGGACAGTG-3'	
CHI2pro: GUS	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGCTCCCAATTCCAACAACCATTCTAG-3'	1352
	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCTGTTATTGAATGTGATGTGAATTTGAGAAACG-3'	
CHI3A1pro: GUS	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAACTTACACAGGAATTAGTCTCATAACAAG-3'	1809
	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCAAGGAATAGGATTGTGTATGCTTATGCC-3'	
CHI3A2pro: GUS	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCAGTTTTCATCAAGCAATCATTCTATCC-3'	1122
	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCTTTAAGGACTAGGATTGTGTATGCTCAG-3'	
CHI4Apro: GUS	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCAGTTGCACCAACATCAATGTC-3'	1637
	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCGATGGAGATTGAGATAGAGACTGG-3'	
CHI4Bpro: GUS	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCCTGAAAGAGCGGATTCCTCTAC-3'	1928
	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCGATGGAGATGGAGATGGAGAC-3'	

For the promoter-GUS assay, promoter fragments (approximately 1.1-1.8 kb upstream of the translation start site) of *GmCHI* gene family members were amplified using the primer sets listed in the Table 2.1, and cloned into the pDONR-Zeo (Invitrogen, USA) vector as described previously. The entry clones were recombined upstream of a *GUS* reporter gene in pMDC162 vector (Curtis and Grossniklaus 2003), in a LR recombination reaction, transformed into *E. coli* DH5α, and screened using kanamycin (50 μg/mL) supplemented LB medium. The destination clones, pMDC-*PromoterGmCHI*, were transformed into *A. tumefaciens* strain GV3101 via electroporation.

2.2.7 Plant transformation

To monitor the transient expressions of fusion proteins, the constructs in *A. tumefaciens* strain GV3101 were transformed into *N. benthamiana* leaf epidermal cells by infiltration (Sparkes et al. 2006). Briefly, a single colony of *A. tumefaciens* GV3101 was used to inoculate 'infiltration culture' medium (LB broth containing 10 mM 2-*N*morpholino-ethanesulfonic acid [MES] pH 5.6, and 100 μM acetosyringone) supplemented with kanamycin (50 μg/mL), rifampicin (10 μg/mL), and gentamycin (50 μg/mL) and grown at 28°C with shaking (225 rpm) until the OD600 reached 0.5-0.8. The culture was centrifuged in a microfuge tube at 3000 rpm for 30 minutes at room temperature and the pellet was washed with infiltration culture medium to remove the residual antibiotics. The pellet was re-suspended in Gamborg's solution (3.2 g/L Gamborg's B5 and vitamins, 20 g/L sucrose, 10 mM MES pH5.6, and 200 μM acetosyringone) to a final OD600=1 and incubated at room temperature for 2-3 hours with gentle agitation to activate the virulence gene required for transformation. To verify subcellular localization pattern of each GmCHI, the pEG101-GmCHI constructs were also co-transformed, in a 1:1 mixture, with organelle markers translationally fused with CFP (Nelson et al. 2007).

The leaves of *N. benthamiana* plants at the second or third trifoliate leaf stage (4-6 week-old) were transformed using a 1 mL syringe. Bacteria were slowly injected into the abaxial side of the leaf. The infiltrated leaves were labelled and plants were returned to the growth room at normal growth condition as described in section 2.2.1. Protein expression was visualized by confocal microscopy after 48 hours.

The dipped plants were placed inside plastic bags to maintain high humidity during an overnight incubation. Subsequently, the plastic was removed and the plants were transferred to a growth chamber and maintained until senescence. Seeds were harvested and transformants selected on MS medium supplemented with Hygromycin (50 μg/mL), as described in section 2.2.1.

2.2.8 Confocal microscopy

Epidermal cell layers of *N. benthamiana* leaves and transgenic soybean hairy roots were visualized using a Leica TCS SP2 inverted confocal microscope. For confocal microscopy, a 63X water-immersion objective was used at excitation wavelengths of 514 and 458 nm, and emission spectra of 530-560 nm and 470-500 nm, for YFP and CFP, respectively. For co-localization the 'Sequential Scan Tool' was utilized which records fluorescence in a sequential fashion instead of acquiring emissions simultaneously in different channels.

2.2.9 Histochemical GUS assay

Histochemical GUS staining was performed on T2 transgenic *Arabidopsis* tissues (Stangeland and Salehian 2002). Tissues were incubated for 16 h at 37°C in 15% methanol containing 0.5 mg/mL X-gluc (5-Bromo-4-chloro-3-indolyl-β-D-glucuronide), except for two week-old seedlings, for which incubation time was reduced to 4 h at 37°C. Tissues were de-stained with 95% ethanol in three half-hour rinses and photographed through a LeicaM2 FLIIITM microscope using a QImaging Retiga 2000R camera.

2.3 Results

2.3.1 Soybean *CHI*s: a gene family of twelve

To identify all the members of the soybean *CHI* family, previously identified soybean *CHI* gene sequences (Chen et al. 2011; Ralston et al. 2005) were used as a query for BLASTN search against the soybean genome database (www.phytozome.net). An

additional search, using 'chalcone isomerase' as a key word was performed. These searches identified twelve *GmCHIs* across ten chromosomes; seven *GmCHIs* were previously reported (Ralston et al. 2005), while five 'type III' *GmCHIs* (*GmCHI3A2*, *GmCHI3B1*, *GmCHI3B2*, *GmCHI3C1*, *GmCHI3C2*) are novel. The predicted molecular masses of GmCHIs range from 23-25 kDa except for the 'type III' GmCHIs, which vary from 29-46 kDa. GmCHI3A and GmCHI3B subfamilies possess signal peptides at the N-terminal region that indicate putative chloroplast subcellular localization. Detailed information on each *GmCHI*, including gene location, length of coding sequence, deduced molecular weight of proteins, and subcellular location are shown in Table 2.2.

The CHI1 subfamily: GmCHI1A, GmCHI1B1, and GmCHI1B2, contain all the critical active site residues that define 'type II legume-specific CHIs', while GmCHI2 falls under the 'type I' category, with the substitution of threonine 190 for serine (Figure 2.1). The CHI3 (type III) and CHI4 (type IV) subfamilies do not retain the catalytic core of the CHI fold, and bear substitutions at nearly all of the critical sites including the widely conserved Asn 113 (Figure 2.1). Pairwise amino acid sequence identities between the gene family members range from 17 to 96% (Table 2.3).

2.3.2 Phylogenetic analysis of the CHI fold; from homologs in fungal soybean pathogens to soybean CHIs

A phylogenetic analysis of GmCHIs was performed and compared with CHIs across different species and kingdoms from ancestral homologues in fungi that lack catalytic ability, to CHIs that can bind and catalyze both trihydroxychalcone (isoliquiritigenin) and tetrahydroxychalcone (naringenin) (type II). The fungal outgroup, which includes *Saccharomyces cerevisiae* (baker's yeast) and *Fusarium graminearum* (a soybean pathogen), represents a clade of ancestral CHI homologues that were found by 'Protein BLAST' of GmCHI1A amino acid sequence against the protein database. The fungal clade serves to root the tree and describe a perceived 'start-point' for the evolution of the CHI-fold.

As shown in Figure 2.2, the phylogenetic tree consisted of two main branches. The smaller branch included CHI3B and CHI3C subfamilies (type III). *Arabidopsis thaliana*

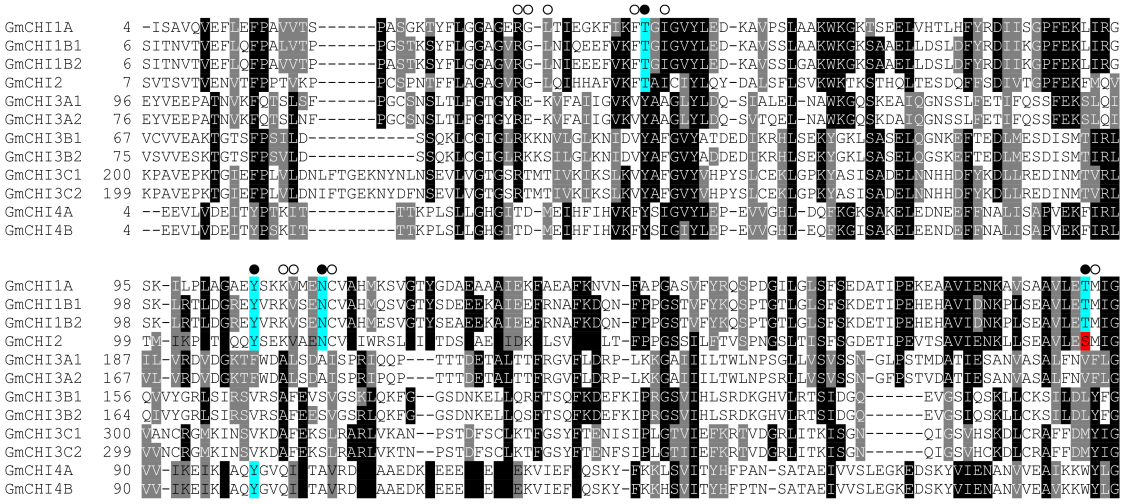


Figure 2.1 Multiple sequence alignment of deduced amino acid sequences of the soybean *CHI* gene family. Black shade indicates identical residues; grey shade indicates similar residues. Empty circles and black filled circles indicate residues that line the active site and critical active site residues, respectively. Thr48, Tyr106, Asn113 and Thr190 are shown in blue shade, and substitution at the fourth critical active site residue (T190S) is shown in red shade. Hyphen indicates a gap.

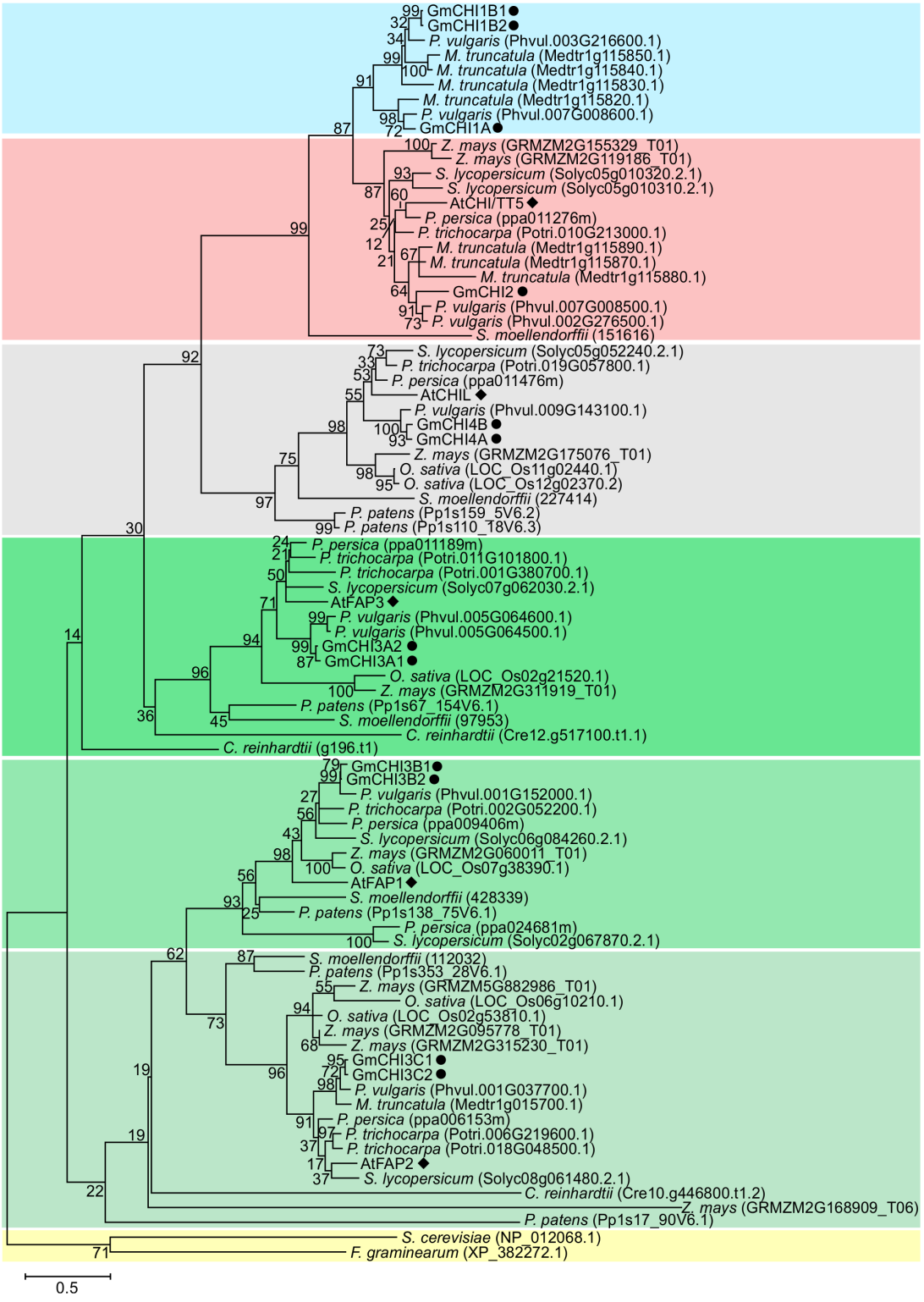

GmCHI1B1
GmCHI1B2
P. vulgaris (Phvul.003G216600.1)
M. truncatula (Medtr1g115850.1)
M. truncatula (Medtr1g115840.1)
M. truncatula (Medtr1g115830.1)
M. truncatula (Medtr1g115820.1)
P. vulgaris (Phvul.007G008600.1)
GmCHI1A
Z. mays (GRMZM2G155329_T01)
Z. mays (GRMZM2G119186_T01)
S. lycopersicum (Solyc05g010320.2.1)
S. lycopersicum (Solyc05g010310.2.1)
AtCHI/TT5
P. persica (ppa011276m)
P. trichocarpa (Potri.010G213000.1)
M. truncatula (Medtr1g115890.1)
M. truncatula (Medtr1g115870.1)
M. truncatula (Medtr1g115880.1)
GmCHI2
P. vulgaris (Phvul.007G008500.1)
P. vulgaris (Phvul.002G276500.1)
S. moellendorffii (151616)
S. lycopersicum (Solyc05g052240.2.1)
P. trichocarpa (Potri.019G057800.1)
P. persica (ppa011476m)
AtCHIL
P. vulgaris (Phvul.009G143100.1)
GmCHI4B
GmCHI4A
Z. mays (GRMZM2G175076_T01)
O. sativa (LOC_Os11g02440.1)
O. sativa (LOC_Os12g02370.2)
S. moellendorffii (227414)
P. patens (Pp1s159_5V6.2)
P. patens (Pp1s110_18V6.3)
P. persica (ppa011189m)
P. trichocarpa (Potri.011G101800.1)
P. trichocarpa (Potri.001G380700.1)
S. lycopersicum (Solyc07g062030.2.1)
AtFAP3
P. vulgaris (Phvul.005G064600.1)
P. vulgaris (Phvul.005G064500.1)
GmCHI3A2
GmCHI3A1
O. sativa (LOC_Os02g21520.1)
Z. mays (GRMZM2G311919_T01)
P. patens (Pp1s67_154V6.1)
S. moellendorffii (97953)
C. reinhardtii (Cre12.g517100.t1.1)
C. reinhardtii (g196.t1)
GmCHI3B1
GmCHI3B2
P. vulgaris (Phvul.001G152000.1)
P. trichocarpa (Potri.002G052200.1)
P. persica (ppa009406m)
S. lycopersicum (Solyc06g084260.2.1)
Z. mays (GRMZM2G060011_T01)
O. sativa (LOC_Os07g38390.1)
AtFAP1
S. moellendorffii (428339)
P. patens (Pp1s138_75V6.1)
P. persica (ppa024681m)
S. lycopersicum (Solyc02g067870.2.1)
S. moellendorffii (112032)
P. patens (Pp1s353_28V6.1)
Z. mays (GRMZM5G882986_T01)
O. sativa (LOC_Os06g10210.1)
O. sativa (LOC_Os02g53810.1)
Z. mays (GRMZM2G095778_T01)
Z. mays (GRMZM2G315230_T01)
GmCHI3C1
GmCHI3C2
P. vulgaris (Phvul.001G037700.1)
M. truncatula (Medtr1g015700.1)
P. persica (ppa006153m)
P. trichocarpa (Potri.006G219600.1)
P. trichocarpa (Potri.018G048500.1)
AtFAP2
S. lycopersicum (Solyc08g061480.2.1)
C. reinhardtii (Cre10.g446800.t1.2)
Z. mays (GRMZM2G168909_T06)
P. patens (Pp1s17_90V6.1)
S. cerevisiae (NP_012068.1)
F. graminearum (XP_382272.1)
0.5
Type II CHI1 Subfamily
Type I - AtCHI/TT5 CHI2 Subfamily
Type IV - AtCHIL CHI4 Subfamily
Type III - AtFAP3 CHI3A Subfamily
Type III - AtFAP1 CHI3B Subfamily
Type III - AtFAP2 CHI3C Subfamily
fungus outgroup

Figure 2.2 Molecular phylogenetic analysis of the deduced amino acids of CHIs or CHI-like proteins. The locus numbers are indicated in parentheses. The protein sequences from different families: (Fabaceae) *Glycine max, Medicago truncatula* and *Phaseolus vulgaris*; (Salicaceae) *Populus trichocarpa*; (Rosaceae) *Prunus persica*; (Brassicaceae) *Arabidopsis thaliana*; (Solanaceae) *Solanum lycopersicum*; (Poaceae) *Zea mays*, *Oryza sativa*; (Selaginellales) *Selaginella moellendorffii*; (Funariaceae) *Physcomitrella patens*; (Chlamydomonadaceae) *Chlamydomonas reinhardtii*; (Saccharomycetaceae) *Saccharomyces cerevisiae*; (Nectriaceae) *Fusarium graminearum*, were aligned using ClustalW2 and the phylogenetic tree was created by the Maximum Likelihood method using Mega5.1. Bootstrap values are shown as percentages next to branch-points. Branch lengths are measured as substitutions per site. The labels denote the soybean CHI subfamily divisions (1, 2, 3 and 4), the catalytic functional "types" (I, II, III and IV), the *Arabidopsis* FAPs, CHIL and CHI members, and the "Fungal outgroup". • and ♦ indicate soybean and *Arabidopsis* members, respectively.

Table 2.2 Molecular and genetic characteristics of the soybean *CHI* gene family

Gene Name	Locus Name	Gene Location	CDS	Predicted Protein Molecular Weight (kDa)	Subcellular Localization	Splice variant(s)	CHI type / FAP type
GmCHI1A	Glyma.20G241500	Chr20: 47259311 - 47261877	657	23.27	Nucleo-cytoplasmic	2	Type II
GmCHI1B1	Glyma.20G241600	Chr20: 47264269 - 47266795	681	25.03	Nucleo-cytoplasmic	1	Type II
GmCHI1B2	Glyma.10G292200	Chr10: 51006927 - 51009538	681	24.97	Nucleo-cytoplasmic	1	Type II
GmCHI2	Glyma.20G241700	Chr20: 47267106 - 47268550	687	24.89	Nucleo-cytoplasmic	1	Type I
GmCHI3A1	Glyma.13G262500	Chr13: 36603843 - 36608523	906	32.49	Chloroplast / Plastid	1	Type III / FAP3-like
GmCHI3A2	Glyma.15G242900	Chr15: 46403868 - 46408623	846	30.26	Chloroplast / Plastid	2	Type III / FAP3-like
GmCHI3B1	Glyma.03G154600	Chr03: 36971024 - 36974353	804	29.66	Chloroplast / Plastid*	2	Type III / FAP1-like
GmCHI3B2	Glyma.19G156900	Chr19: 41733799 - 41736875	828	30.38	Chloroplast / Plastid*	1	Type III / FAP1-like
GmCHI3C1	Glyma.14G098100	Chr14: 9259148 - 9266365	1245	45.99	Cytoplasmic*	2	Type III / FAP2-like
GmCHI3C2	Glyma.17G226600	Chr17: 38105929 - 38111302	1242	46.19	Cytoplasmic*	1	Type III / FAP2-like
GmCHI4A	Glyma.06G143000	Chr06: 11642031 - 11644022	629	23.50	Nucleo-cytoplasmic	3	Type IV
GmCHI4B	Glyma.04G222400	Chr04: 49304081 - 49306117	629	23.57	Nucleo-cytoplasmic	1	Type IV

* Predicted subcellular localization based on the prediction software TargetP 1.1, and not determined experimentally as other members of the family. CHI, chalcone isomerase; CDS, coding DNA sequence; FAP, fatty acid-binding protein; kDa, kilodalton

Table 2.3 Pairwise comparison of coding DNA and deduced amino acid sequence comparison of the soybean *CHI* family[a]

Nucleotide												
	GmCHI1A	GmCHI1B1	GmCHI1B2	GmCHI2	GmCHI3A1	GmCHI3A2	GmCHI3B1	GmCHI3B2	GmCHI3C1	GmCHI3C2	GmCHI4A	GmCHI4B
GmCHI1A		72.26	72.21	61.26	41.83	41.03	49.43	49.18	44.09	44.85	45.47	44.47
GmCHI1B1	66.97		95.74	63.96	45.06	43.83	50.47	50.08	44.59	44.89	44.36	43.58
GmCHI1B2	67.89	96.46		63.96	45.52	44.44	50.16	49.76	44.30	44.74	43.77	43.00
GmCHI2	49.08	51.33	50.88		42.02	41.56	49.37	48.35	45.35	44.17	42.20	41.43
GmCHI3A1	21.63	24.17	23.22	24.53		94.44	45.17	46.04	42.32	42.22	59.22	58.90
GmCHI3A2	21.36	23.92	23.44	25.24	92.17		44.59	44.60	41.50	41.52	58.25	57.77
GmCHI3B1	18.61	21.89	21.39	19.31	18.25	17.92		94.90	60.32	60.82	44.12	43.22
GmCHI3B2	19.70	22.89	22.89	20.30	18.46	18.15	92.88		59.42	59.90	42.68	41.77
GmCHI3C1	16.91	19.05	20.00	16.59	18.58	20.29	34.08	34.55		94.35	41.76	40.46
GmCHI3C2	16.91	19.05	20.00	16.59	18.24	20.29	34.96	34.67	92.96		41.76	40.78
GmCHI4A	31.25	31.25	31.25	29.33	23.15	24.38	18.97	18.46	18.50	18.50		96.66
GmCHI4B	31.25	30.29	30.29	30.29	22.66	23.88	18.97	18.46	18.50	18.50	95.22	
Amino Acid												

[a] Sequence identity (%) between twelve chalcone isomerase family members at the nucleotide (top triangle) and amino acid (bottom triangle) levels; nucleotide sequences correspond to the coding region sequences for each gene, acquired from the Joint Genome Institute assembly of *Glycine max* genome (Wm82.a2.v1); amino acid sequences are deduced from the coding sequences; nucleotide and amino acid sequences were aligned by ClustalW and percentage identities were calculated from the subsequent multiple sequence alignment.

has five genes encoding CHI, of which three are Fatty Acid-binding Proteins (FAPs) (Ngaki et al. 2012). AtFAP1 and 2 each form a broad clade, which includes members from the tracheophyta (vascular plants), lycophyta and bryophyta, and as a distant ancestor to the clade, a sequence from the single-celled algae, *Chlamydomonas reinhardtii.* Phylogenetic analysis revealed that GmCHI3C subfamily falls into a clade with AtFAP2 from *A. thaliana*, while GmCHI3B groups with AtFAP1. These two clades are ancestral to the AtFAP3 clade, which appears to be the next stepping-stone towards CHI-like sequences. AtFAP3 (At1g53520) clusters tightly together with GmCHI3A subfamily. The majority of the CHIs and CHI homologues fall into the other main branch of the tree that is comprised of 'CHI Subfamily 4 – Type IV', 'CHI Subfamily 2 – Type I' and 'CHI Subfamily 1 – Type II' (Figure 2.2). The subfamily classifications, as before, are derived from the soybean genome database annotations that have been made based on characteristic functional elements. The 'CHI subfamily 4 clade' importantly includes *A. thaliana* CHIL (At5g05270), which has substitutions at several catalytic residues, and therefore is not a *bona fide* CHI. This situation is mirrored in soybean with GmCHI4A and GmCHI4B that only retain one of four core residues (Figure 2.1). Soybean GmCHI4 subfamily and AtCHIL are tightly clustered with a confidence of 98%.

The 'CHI subfamily 2 – type I' clade includes a number of GmCHI homologs from inside and outside the Fabaceae, signifying broad representation across vascular plants. The clade includes AtCHI/TT5, and GmCHI2 that has a substitution at one of its core catalytic residues (T190S) (Figure 2.1), clustering together the known 'type 1' CHIs. An ancestral outgroup to this entire clade is a solitary member from the lycophyta, *S. moellendorffii*.The last branch in the tree breaks off with a bootstrap confidence of 91%: the 'CHI subfamily 1 – type II' clade. It is important to note that despite the comprehensive selection of CHIs from a wide range of species, only CHIs from the Fabaceae cluster in this clade, including *M. truncatula*, *P. vulgaris*, and soybean (Figure 2.2). GmCHI1A, GmCHI1B1 and GmCHI1B2, retain all the catalytic active site residues of the CHI fold associated with the ability to catalyze both tri- and tetra- hydroxychalcone (Figure 2.1).

2.3.3 Expression analysis of *GmCHI* gene family

To determine the tissue-specific expression patterns of the *GmCHI* gene family, qRT-PCR was performed. Total RNA was extracted from various tissues at different development stages and was used for qPCR analysis of eight core *GmCHI* genes (*CHI1*, *CHI2*, *CHI3A* and *CHI4* subfamilies) (Figure 2.3). *GmCHI1A* was expressed ubiquitously throughout all tissues under study, with a pattern of increased expression with embryo development, reached a maximum at 70 DAP. *GmCHI1B1* was expressed in the root and the flower bud, and was absent from the rest of the tissues whereas *GmCHI1B2* was expressed ubiquitously with highest expression in the flower bud, and the latter stages of embryo development. *GmCHI2* was strongly expressed in the seed coat, leaf, flower bud and flower. Transcript accumulation of the genes in subfamily 3 of the *GmCHIs* were very different, with *GmCHI3A2* (not shown) being entirely undetected from the samples, and *GmCHI3A1* being strongly expressed in the embryos during the mid-stages of development (30, 40 and 50 DAP). *GmCHI4A* was ubiquitously expressed across all tissues with increased expression as embryos approached seed maturity, as was *GmCHI4B*; however the latter, was most strongly expressed at 70 DAP of embryo maturity.

The relative expression patterns of all twelve *GmCHI* genes, along with some of the key isoflavonoid biosynthetic genes were studied using publicly available data, mined from the RNA-sequencing of the soybean transcriptome. As shown in Figure 2.4, transcript levels of *GmCHI1A, GmCHI1B1, GmCHI1B2, GmCHR, GmCHS7*, and *GmCHS8* were higher in roots, compared to other soybean tissues. In the leaves, 'type III' subfamilies, *GmCHI3A* and *GmCH3B*, were heavily expressed. Expression of *GmCHI3C* genes was not detected in the tissue under study, and was not included. During the early developmental phases of the seeds *GmCHI2*, *GmCHI4A* and *GmCHI3B2* were expressed, while *GmCHI3A1* and *GmCHI3B1* transcripts accumulated specifically at early maturation. To complement and add further detail to my analysis of the expression patterns of *GmCHIs*, the promoter fragments of eight core members of the gene family

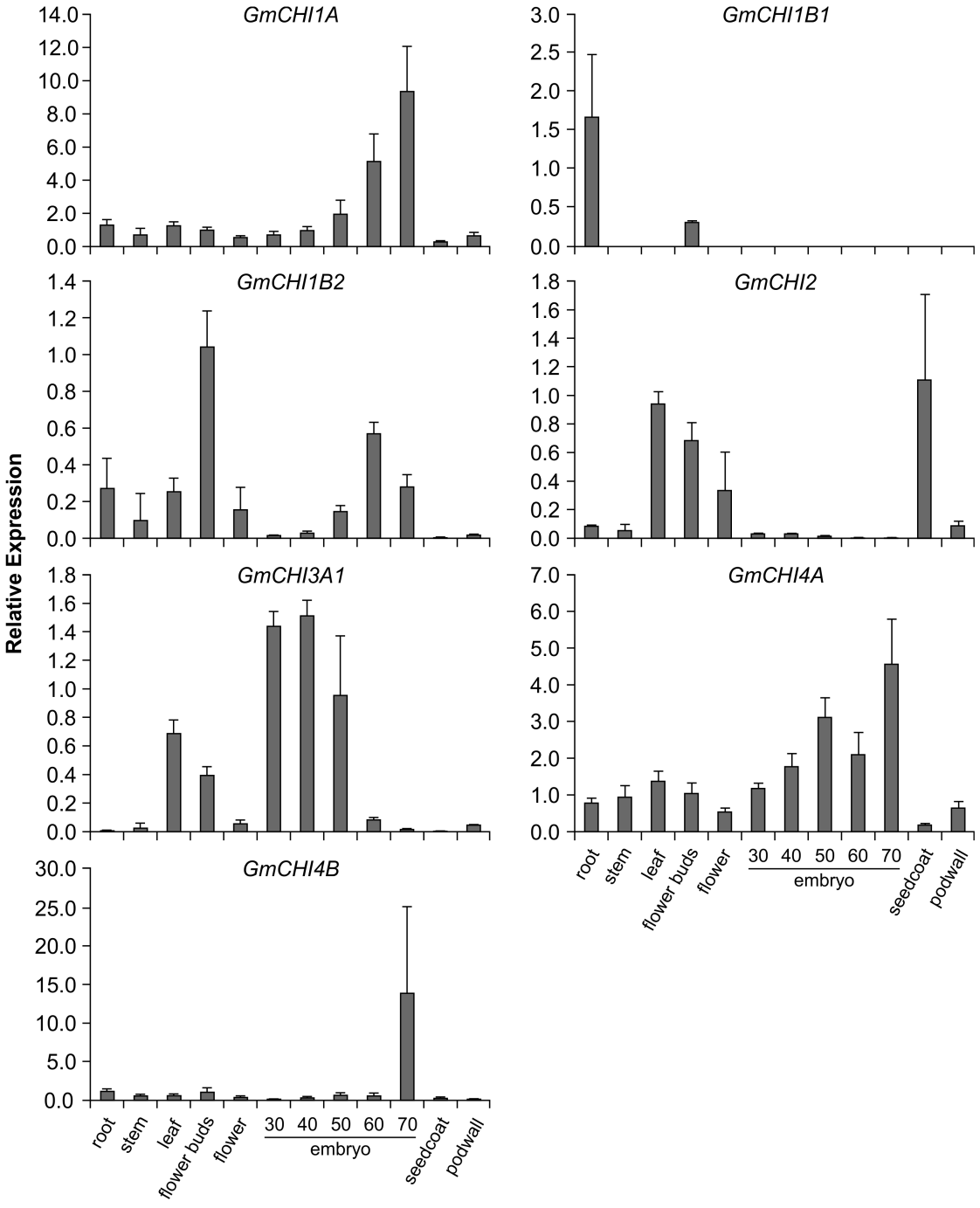


Figure 2.3 Expression analyses of seven *GmCHI* genes across twelve different tissues. Total RNA was extracted from soybean root, stem, leaf, flower bud, flower, embryo (30, 40, 50, 60, and 70 DAP), seed coat and pod wall, and used in quantitative RT-PCR analysis. Error bars indicate SEM of two biological and three technical replicates. Values were normalized against the reference genes *CONS4* and *CONS6*.

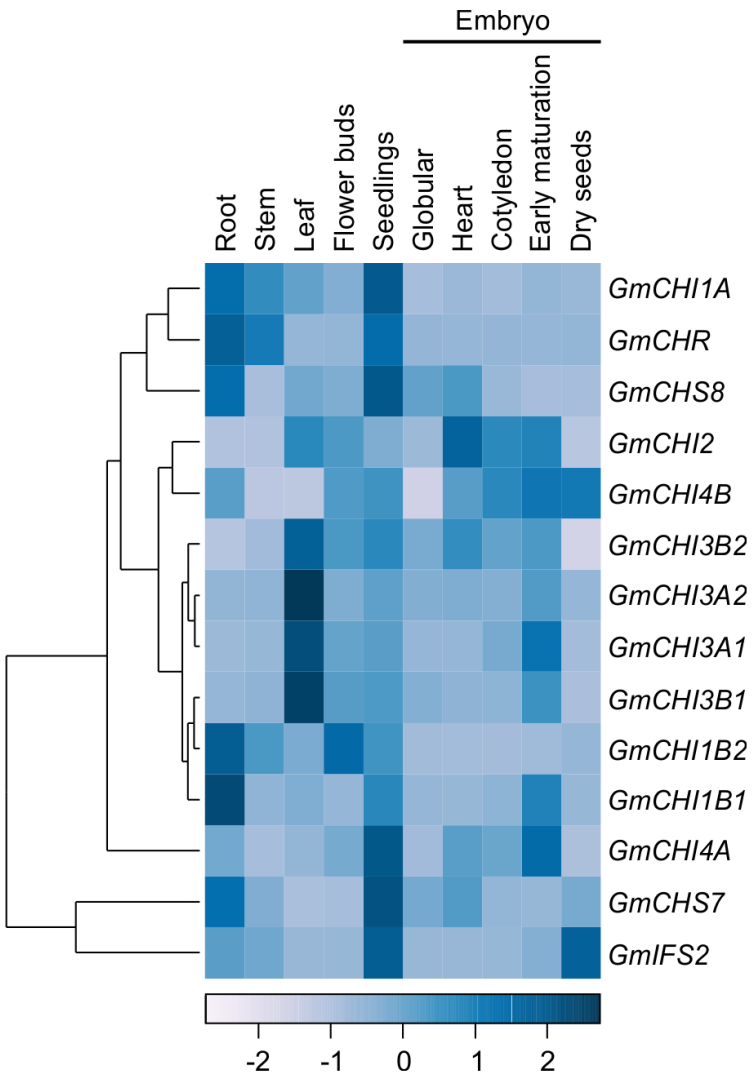


Figure 2.4 Heatmap of expression profile of the *GmCHI* gene family and other key isoflavonoid-specific genes. The genome-wide RNAseq data of soybean were obtained and data normalized as indicated in 'Materials and methods'. The transcript accumulation of *GmCHI*s, *GmCHR*, *GmCHS7*, *GmCHS8* and *GmIFS2,* across ten tissues, including five stages of embryo development is shown relative to the respective control indicated by a gradient from white or light blue (low) to dark blue (high).

were amplified and cloned upstream of a *GUS* reporter gene and transformed into *A. thaliana* (Figures 2.5, 2.6). Histochemical staining of 2-week-old T2 generation of *A. thaliana* seedlings showed GUS activity driven strongly by all eight, core *GmCHI* promoters (Figure 2.5) covering major structures, including the leaves and the roots, but not the hypocotyl; the shoot apical meristem and the cotyledons showed little to no GmCHI1B1:GUS activity (Figure 2.6a).

In the roots, GUS activity was driven strongly by all eight, core *GmCHI* promoters, particularly in the procambium, with a vein of blue running along the structure, ending in an unstained root apical meristem (Figure 2.6b). The tips of lateral roots, as well, did not have GUS activity. All the *GmCHI* promoters, excluding *CHI1B1*, exhibited strong GUS activity in 4- to 6-week rosette leaves (Figure 2.5). The latter had punctate points of expression at the lobes of the leaf and in the trichomes (Figure 2.6c). *GmCHI2* promoter was the only other member of the family that had distinctive expression in the trichomes and in the protodermal cells bordering the trichome. *GmCHI4A* promoter activity was absent from the midrib and petiole of the leaf, as compared to very strong activity in the rest of the leaf. The anther, containing the microsporangia, and the style containing the ovules, were not stained for GUS activity in any of the *GmCHI*-promoters. GUS activity was present in the stigma, at the upper part of the gynoecium, with *GmCHI1A*, *GmCHI1B2*, *GmCHI2*, *GmCHI4A* and *GmCHI4B* promoters. The filament was stained in all promoters except *GmCHI1B1*. The calyx was stained in all *GmCHI* promoters, except for *GmCHI3A1*, with only minor staining in a venation pattern for *GmCHI1B2* and *GmCHI3A2*. The corolla was stained in *GmCHI1A* and *GmCHI2*. Finally, the peduncle and the pedicel of *GmCHI1A*, *GmCHI1B2*, *GmCHI2*, *GmCHI4A* and *GmCHI4B* were stained for promoter-GUS activity (Figure 2.6d). In 7-week, early siliques, silique walls were either partially (*GmCHI1B1*, *GmCHI1B2*, *GmCHI2*, *GmCHI4A* and *GmCHI4B*) or strongly (*GmCHI1A*, *GmCHI3A1* and *GmCHI3A2*) stained in all GmCHI promoters (Figure 2.6e). Moderate GUS activity was observed in seed coats at the early stage with *GmCHI1A*, *GmCHI2* and *GmCHI4A* promoters. In 10-week-old mature siliques (prior to senescence and dehiscing), the valves, septa and funiculi were stained in *GmCHI1A*, GmCHI1B1, *GmCHI1B2*, *GmCHI2*, *GmCHI3A2* and *GmCHI4B* promoters (Figure 2.6f).

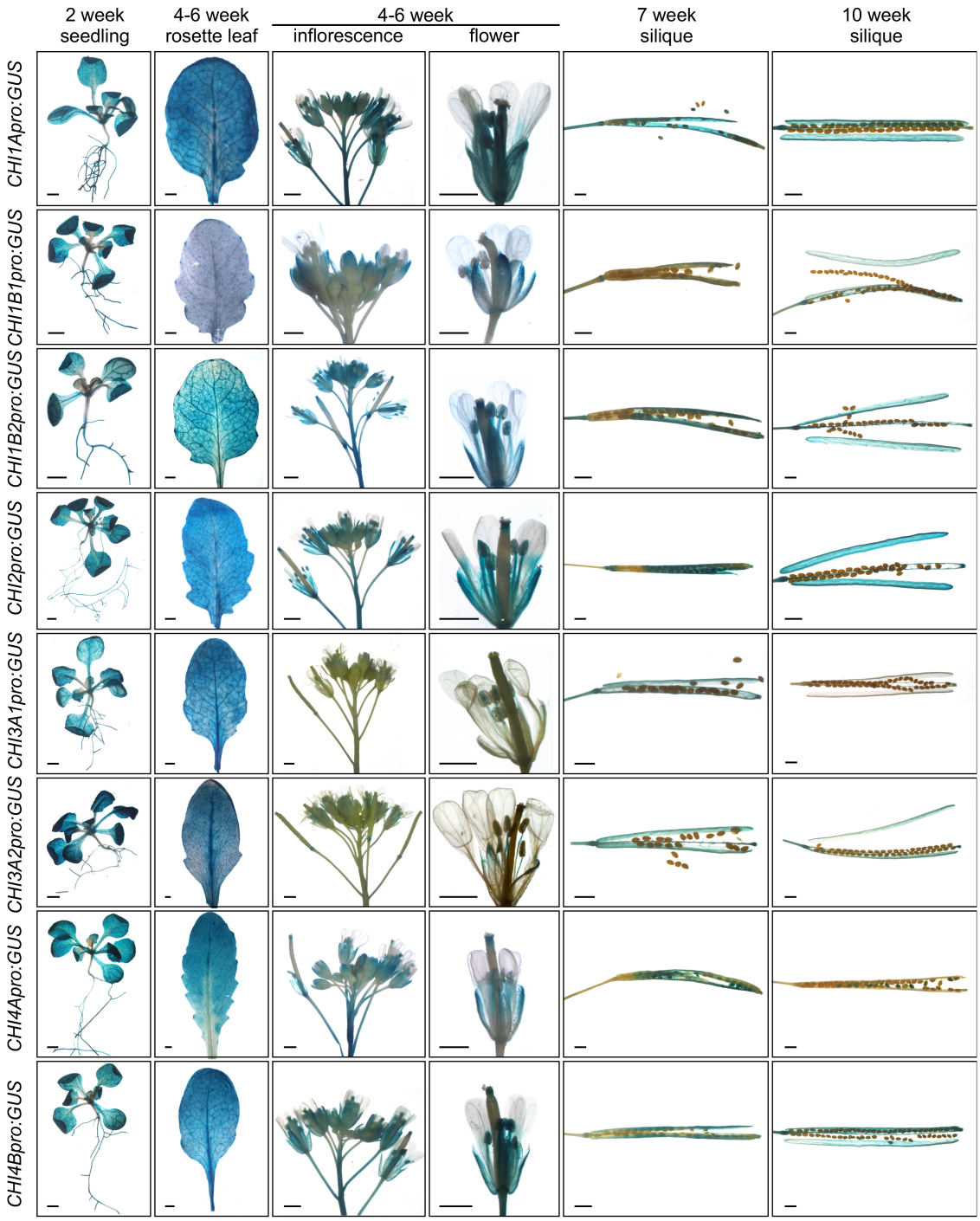

2 week seedling
4-6 week rosette leaf
4-6 week
inflorescence
flower
7 week silique
10 week silique
CHI1Apro:GUS
CHI1B1pro:GUS
CHI1B2pro:GUS
CHI2pro:GUS
CHI3A1pro:GUS
CHI3A2pro:GUS
CHI4Apro:GUS
CHI4Bpro:GUS

Figure 2.5 Soybean *CHI* promoter-GUS expression in *Arabidopsis* tissue. Constructs containing *GmCHI* promoters (core eight CHIs), driving GUS gene expression, were transformed into *Arabidopsis*. T_2 generation plants were selected for histochemical staining. Tissues used include 2-week-old whole seedling, 4- to 6-week-old rosette leaf, 6-week-old inflorescence and flower, and 7- to 10-week-old siliques (early to mature siliques). Scale bar indicates 1000 μm.

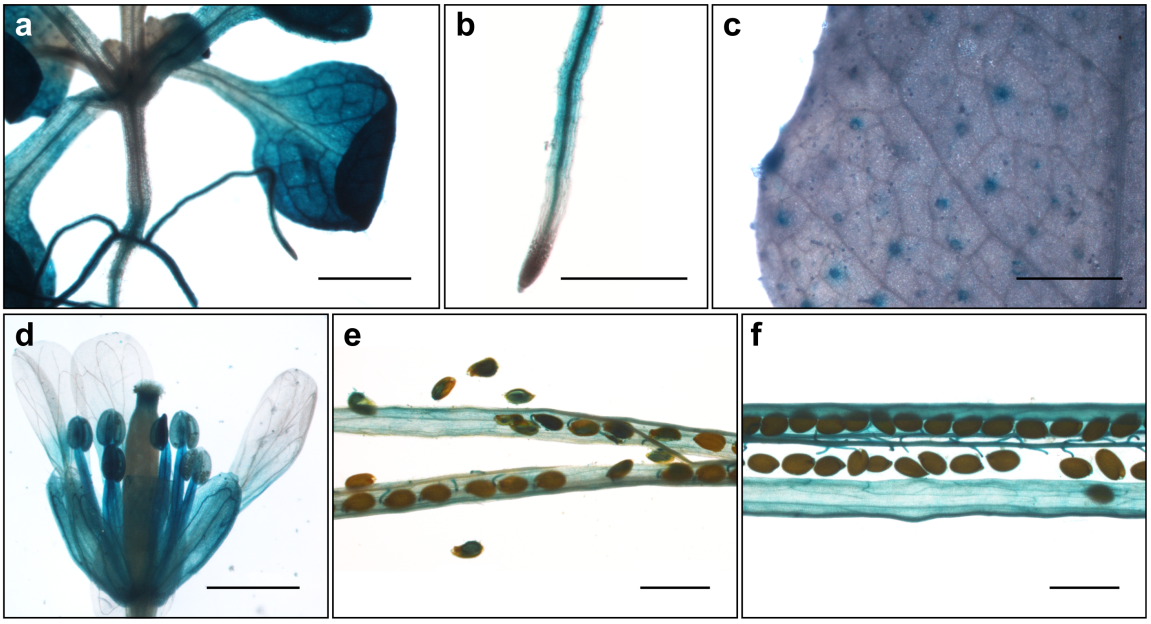


Figure 2.6 Promoters of the soybean *CHI* gene family lead to differential GUS activity: (a) *GmCHI1B1pro::GUS* absent in hypocotyl and cotyledons of 2-week seedlings; (b) *GmCHI1Apro::GUS* absent in root apical meristem; (c) *GmCHI1B1pro::GUS* in 4 week rosette leaf restricted to trichomes and surrounding protodermal cells; (d) *GmCHI2pro::GUS* in filament, stigma, calyx and corolla; (e) *GmCHI1Apro::GUS* in seed coats of early 7-week siliques; (f) *GmCHI4Bpro::GUS* in valves, septa and funiculi of 10-week mature siliques. Scale bar indicates 1000 µm.

In case of mature seed coats, only GmCHI4A promoter was able to activate GUS expression.

2.3.4 GmCHI subcellular localization

To understand the potential functional divisions in the *GmCHI* family, subcellular localization was investigated by creating translational fusions of eight full-length *GmCHIs* upstream of the reporter gene *YFP*. The constructs were transiently expressed in *N. benthamiana* leaf epidermal cells (Figure 2.7). In the absence of localizing signals GmCHIs from subfamilies 1, 2 and 4 localized to the cytoplasm and nucleus. Nuclear localization of the GmCHIs was confirmed by co-expression with a nuclear localization signal-bearing CFP fusion (Figure 2.8). GmCHI3 subfamily members localized to the chloroplast, as predicted by *in silico* analysis. Chloroplast localization was confirmed through co-localization with chloroplast auto-fluorescence and co-expression with plastid organelle markers (Figure 2.8e, f). GmCHI-YFP fusions were also transformed into soybean, via the hairy root transformation method. Expression of the GmCHIs, in soybean hairy roots confirmed the results in *N. benthamiana leaves*, with subfamilies 1, 2 and 4 showing nuclear-cytoplasmic localization, and subfamily 3, exhibiting root plastid localization (Figure 2.9).

2.4 Discussion

2.4.1 Soybean *CHI* gene family has twelve members

CHIs are a unique group of enzymes that catalyze an intramolecular, stereospecific cyclization, where the same substrate molecule contains both the nucleophile and the double bond (Bednar and Hadcock 1988). CHIs in legumes have broader substrate specificity than that of other vascular plants. Previously, a total of seven *GmCHIs* belonging to four different subfamilies were reported in soybean (Ralston et al. 2005). A genome-wide search for *GmCHI* genes identified twelve members, categorized into four subfamilies, located across ten chromosomes. Five of these genes are newly identified,

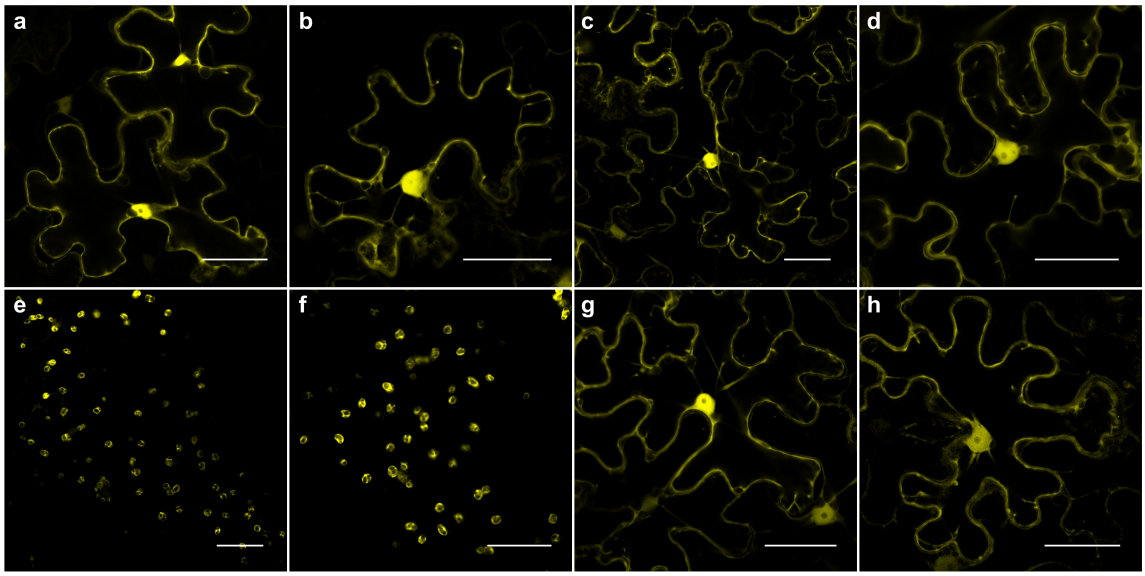


Figure 2.7 Subcellular localization of the GmCHI family. The eight core *GmCHI* genes were translationally fused upstream of the reporter gene *YFP*, transformed into *N. benthamiana* by *A. tumefaciens* mediated transformation and visualized in the leaf epithelial cells by confocal laser-scanning microscopy. (a) GmCHI1A, (b) GmCHI1B1, (c) GmCHI1B2, (d) GmCHI2, (e) GmCHI3A1, (f) GmCHI3A2, (g) GmCHI4A and (h) GmCHI4B. Scale bar indicates 40 µm.

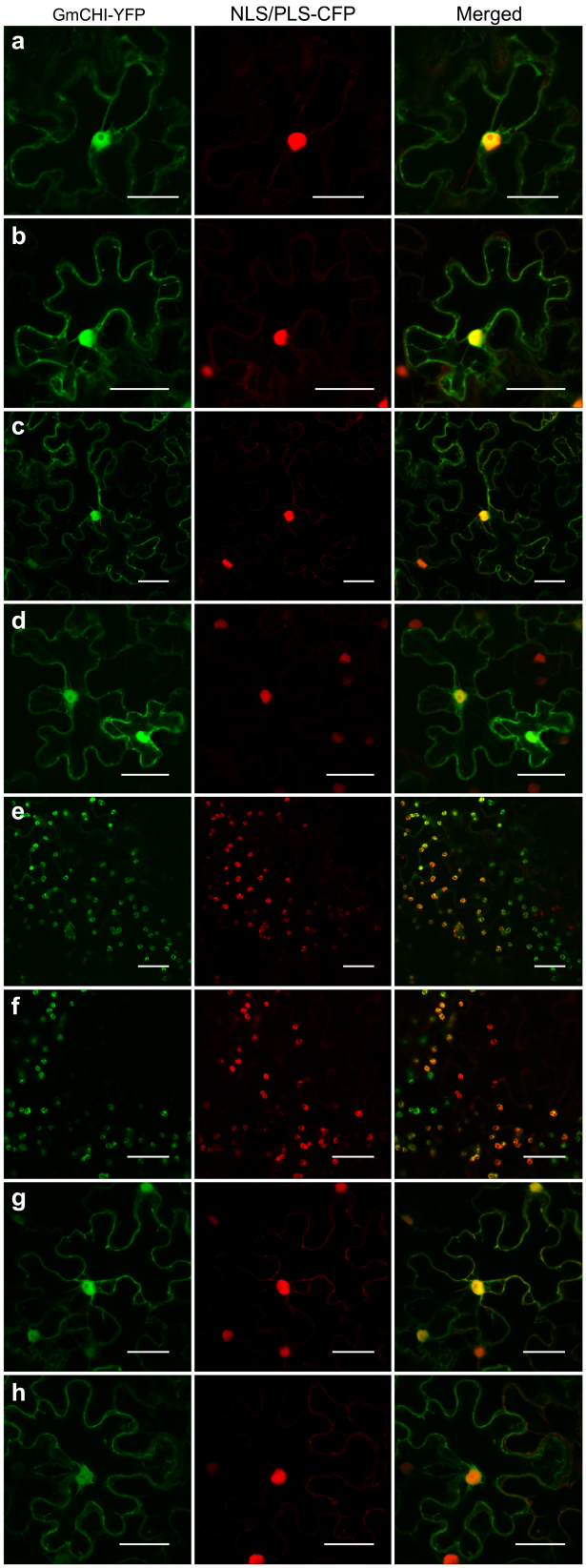

GmCHI-YFP
NLS/PLS-CFP
Merged
a
b
c
d
e
f
g
h

Figure 2.8 Co-localization of GmCHI-YFP and organelle markers fused with CFP: (a) GmCHI1A and NLS-CFP, (b) GmCHI1B1 and NLS-CFP, (c) GmCHI1B2 and NLS-CFP, (d) GmCHI2 and NLS-CFP, (e) GmCHI3A1 and PLS-CFP, (f) GmCHI3A2 and PLS-CFP, (g) GmCHI4A and NLS-CFP, and (h) GmCHI4B and NLS-CFP. YFP, yellow fluorescent protein; CFP, cyan fluorescent protein; NLS, nuclear localization signal; PLS, plastid localization signal. Scale bar indicates 40 μm

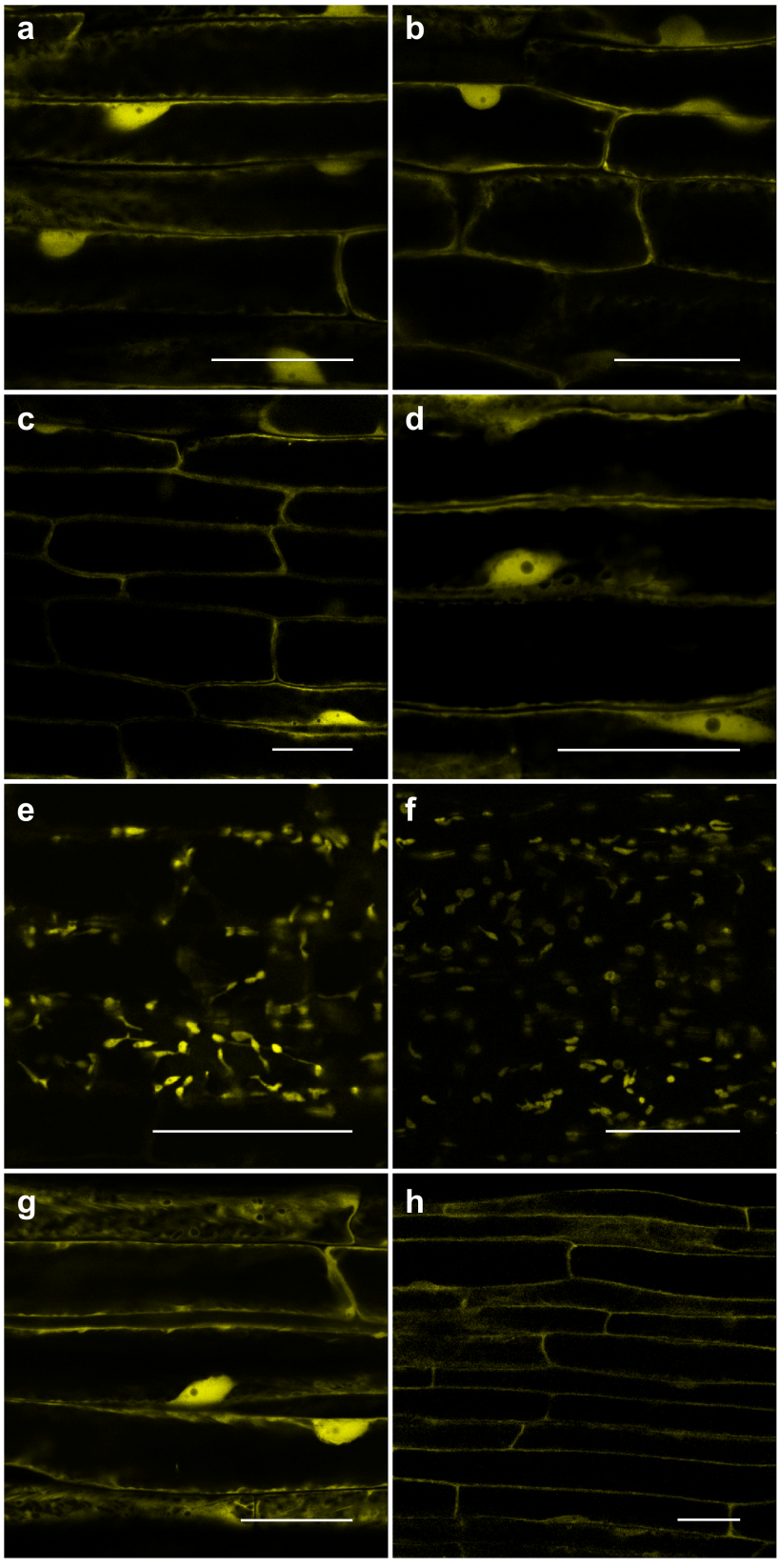

a
b
c
d
e
f
g
h

Figure 2.9 Subcellular localization of GmCHIs in soybean hairy roots. Constructs containing translational fusion of GmCHI-YFP was transformed into *A. rhizogenes* strain K599, and used in the hairy root transformation of soybean cotyledons. Confocal microscopy was used to detect expression after the emergence of hairy roots. (a) GmCHI1A, (b) GmCHI1B1, (c) GmCHI1B2, (d) GmCHI2, (e) GmCHI3A1, (f) GmCHI3A2, (g) GmCHI4A, and (h) GmCHI4B. Scale bar indicates 40 μm.

including *GmCHI3A2*, *GmCHI3B1*, *GmCHI3B2*, *GmCHI3C1* and *GmCHI3C2*. A phylogenetic analysis of the *GmCHI* family, along with homologous sequences across plant divisions, and a fungal outgroup, allowed us to make the current classifications, and deduce the evolutionary history of soybean *CHIs* (Figure 2.2).

2.4.2 Evolution of the CHI fold and the legume-specific branch

As indicated in the phylogenetic tree (Figure 2.3), GmCHIs are divided primarily into two clades. The first includes GmCHI3B and GmCHI3C subfamilies in the ancestral clade, clustering with AtFAP1 and AtFAP2, respectively. This represents the more basal branch of the FAP family found across the plant kingdom, homologous to fungal CHI-fold proteins. They do not contain conserved residues of the active site or of the fold. The second large clade encompasses the rest of the GmCHI subfamilies: GmCHI3A, GmCHI4, GmCHI2 and GmCHI1 subfamilies. This 'core' clade of eight soybean genes was chosen for further study and characterization.

The GmCHI3A subfamily clusters with AtFAP3 in the 'core' clade and appear to be true 'type III' CHIs with subcellular localization to the chloroplast suggesting that they may have similar function as AtFAPs, and can be named as GmFAPs. A functional analysis of type III GmCHIs will demonstrate if they possess FAP like role, and belong to GmFAPs instead of GmCHIs group. A confidence level of 92% substantiates the next bifurcation in the evolutionary road of CHI that leads to one clade representing the functionally active CHI 'type I and II' and the other representing the functionally unknown 'type IV' CHIs. The proximity of these clades in contrast to the 'type III' CHIs suggests that these CHI-like isoforms (type IV) are the sterically restricted precursors to the *bona fide* CHI (Ngaki et al. 2012). There are representatives in this clade spanning the divisions of the plant kingdom, including, GmCHI4A and GmCHI4B, which cluster together with the AtCHIL. Distant species in the plant kingdom, from the single-celled green algae, *C. reinhardtii* to the Fabaceae, have members in the 'type III and IV' clades. The evolution of the CHI fold from a FAP to a CHI-like fold appears to have started early in the development of the plant kingdom, in the algal progenitors, approaching its culmination with the evolution of early vascular plants such as, *S. moellendorffii* and its ancestral 'type I' CHI.

The division of 'type I' CHIs (GmCHI2) and 'type II' (GmCHI1A, GmCHI1B1 and GmCHI1B2) demarcates the CHIs common to all vascular plants and the CHIs arising in the Fabaceae, respectively (Ralston et al. 2005). There are no representatives from the non-leguminous sample in the 'type II' clade. A look at the amino acid alignment of GmCHIs (Figure 2.1) displays the fact that GmCHI2 has a T190S substitution at one of four critical active site residues. This has been speculated to be the integral change differentiating between the two types of CHIs and is associated with the appearance of isoflavonoid biosynthesis in legumes (Livingstone et al. 2010). Whether or not this substitution affects the isoform specialization into flavonoid or isoflavonoid biosynthesis can only be determined *in planta*, as catalytic advantages can be overlooked in the context of metabolon synthesis.

2.4.3 Tissue and organ expression patterns of GmCHI genes

Expression analysis of the gene family was conducted by juxtaposing qPCR data (Figure 2.3), with RNA sequencing data. The RNAseq data was mined and compiled into a heatmap of genes of interest, including *GmCHIs* (except for unexpressed *GmCHI3C1* and *GmCHI3C2*) and isoflavonoid-related genes: *GmCHS7*, *GmCHS8*, *GmCHR* and *GmIFS2* (Figure 2.4). The purpose of including the latter was for correlation of *GmCHI* transcript accumulation, with that of genes associated with isoflavonoid production. *GmCHS7* and *GmCHS8* code for chalcone synthase isoforms involved in seed isoflavonoid synthesis (Dhaubhadel et al. 2007); *GmCHR* has a role in isoflavonoid production and soybean resistance to root and stem rot pathogen, *Phytophthora sojae* (Subramanian et al. 2005); *GmIFS2* was chosen in preference to *GmIFS1*, as the former is not ubiquitously present, and its expression is induced after inoculation with *P. sojae* (Dhaubhadel et al. 2003). Isoflavonoid content is highest in soybean leaves and seeds; however, as seen with low *IFS* expression in the former (Dhaubhadel et al. 2003) (Figure 2.4), gene expression does not always correlate with isoflavonoid accumulation. Metabolite accumulation can be affected by rates of synthesis, turnover, transport and conjugation (Winkel 2004). Supra-molecular complexes like the isoflavonoid metabolon are thought to be in dynamic flux, with protein association and dissociation governed by expression of their constituents and environmental conditions (Bassard et al. 2012a; Bassard et al. 2012b). Therefore, high

transcript levels of a *GmCHI* in conjunction with *GmCHS7*, *GmCHS8*, *GmCHR* or *GmIFS2*, and high isoflavonoid content, may indicate *de novo* isoflavonoid biosynthesis in the respective organ. This line of reasoning is confounded by the fact that, *GmCHI* is a key checkpoint enzyme for many branches of the general phenylpropanoid pathway.

Transcript analysis of the eight 'core' *GmCHI* genes showed seven were actively transcribed in soybean cultivar Harosoy63. It is possible that *GmCHI3A2* could be a cultivar-specific isoform. *GmCHI1A*, *GmCHI1B2*, *GmCHI4A* and *GmCHI4B* were expressed ubiquitously, albeit at varying levels as expected in a diploidized tetraploid like soybean (Schmutz et al. 2010). Differences in expression patterns among these genes may reflect isoform specificity. In leaves, a site of high isoflavonoid accumulation, catalytically active *GmCHI2* is expressed in conjunction with *GmCHS8*, but *GmCHR* and *GmIFS2* are present at a lower level. In soybean seeds, isoflavonoid accumulation begins in the early maturing embryo, and reaches an apex from 50 DAP till full maturation (Dhaubhadel et al. 2003). The qPCR results showed a trend of steady expression increase for *GmCHI1A* and *GmCHI4A*, starting at 30 DAP and culminating at 70 DAP. In mature seeds, where most of the isoflavonoids are accumulated, only *GmCHI1A* from the catalytically active *GmCHIs* is expressed concurrently with high levels of *GmIFS2* and *GmCHS7* and *GmCHS8* (Dhaubhadel et al. 2007). This juxtaposition of qPCR and transcriptome data suggests that *GmCHI1A* could be the isoform specifically involved in isoflavonoid biosynthesis in seeds.

The seeds are a site of fatty acid accumulation, particularly in the early stages of maturation (Ruuska et al. 2002; Dhaubhadel et al. 2007; Jones et al. 2010), which coincides with expression in subfamilies *GmCHI3A* and *GmCHI3B*, shown in qPCR data (for the former) and RNAseq analysis (both), linking this set of genes with the *AtFAPs*.

Soybean GmCHI4 encode 'type IV' CHIs that lack the conserved residues for chalcone to flavanone conversion, and are very similar to AtCHIL. Little information has been available as to their putative function, until a recent study where 'type IV' CHIs in petunia and torenia were discovered to play a role as enhancers of flavonoid production and floral pigmentation (Morita et al. 2014). The study also demonstrated the co-regulated expression of 'type IV' *CHIs* with 'type I' *CHIs*. *GmCHI4A* appears to have an

expression pattern that is roughly similar to that of *GmCHI1A,* increasing with embryonic development from 30 to 70 DAP. *GmCHI4B* expression was ubiquitous; with an abrupt positive fold change at 70 DAP, where it is expressed most strongly.

The findings of the *GmCHI* promoter analysis complement our knowledge of *GmCHI* transcript accumulation, with detailed information about sub-organ expression patterns. A caveat is that *Arabidopsis*, as a non-legume, might not contain the complete set of legume-specific gene regulators. It does, however, remain a consistent and rapid model system for this work. The histochemical staining of 2-week *Arabidopsis* seedlings showed GUS activity driven strongly across the tissue by all eight *GmCHI* promoters (Figure 2.5) except in the hypocotyl, root transition and meristematic zones (Figures 2.6a, b). Saslowsky et al., (2005) did not find fluorescence in the hypocotyl or the root transition zones after staining 4-day-old *Arabidopsis* seedlings with flavonoid-specific dye DPBA. This could explain the lack of *GmCHI* promoter-driven *GUS* expression in the hypocotyl regions of the seedlings and the root oscillatory transition zone (Figures 2.6a, b). The reason for a lack of *GUS* expression in the root apical meristem could lie in the nature of meristematic cells, their constant division and lack of differentiation or metabolite accumulation (Figure 2.6b).

Flavonoids contribute a large and diverse metabolic component to plant physiology in the flower, with roles in seed dormancy, pigmentation of flowers and fruits, protection against UV-B damage and reduction in herbivory (Phillip 1992; Stafford 1997; Aoki et al. 2000; Winkel-Shirley 2001). CHI converts naringenin chalcone to naringenin, which can then be recruited into the anthocyanin pathway (Figure 1.3), producing metabolites important in floral pigmentation, attraction, pollination and seed dispersal (Winkel-Shirley 2001). Therefore, naringenin substrates are in heavy competitive demand in plant flowers, which explains the high levels of *GmCHI1A* and *GmCHI2* expression in this tissue. This explanation also extends to high levels of GUS-activity in the stamen filament, which was recorded for all *GmCHIs* except *GmCHI1B1*. Previous work in Kiwifruit showed that ovary and stigma pigmentation are independent of the filament, which could explain the lack of *GmCHI* promoter presence in the former (Fraser et al.

2013). Of course, accumulation of anthocyanins in the respective tissues must be distinguished from *de novo* synthesis.

Promoter activity of *GmCHI1B1* was manifested as a unique punctate pattern of leaf *GUS* expression. Closer examination reveals that expression includes the trichomes and the protodermal cells bordering the trichome (Figure 2.6c). The accumulation and importance of flavonoids in the trichomes of various plant species is well documented (Shirley et al. 1995; Valkama et al. 2004; Cui et al. 2011). In cultivated tomato (*Solanum lycopersicum*), a mutant lacking an isoform of CHI (SlCHI1) leads to a failure to accumulate flavonoids and terpenoids, drawing a link between the two distinct pathways, and results in susceptibility to herbivorous beetles (Kang et al. 2014). Therefore, GmCHI1B1 might be a soybean isoform of CHI tasked with flavonoid (in non-legumes) and/or isoflavonoid production (in legumes) in the leaf trichomes, with a potentially critical role in plant defense and stress response. The leaf, along with the mature seed, in soybean is the major organ for isoflavonoid accumulation (Dhaubhadel et al. 2003). So, expression of *GmCHI1B1*, solely in leaf trichomes, might be an example of isoform specificity in a very important isoflavonoid-accumulating organ. From the promoter-GUS study I learnt that *GmCHI1A*, *GmCHI2* and *GmCHI4A* are expressed early in seed maturation, while only *GmCHI4A* is expressed in the seed coat at the 10-week stage, before senescence and dehiscing. This is contradicted by qPCR data that shows *GmCHI2* being most strongly expressed in the seed coat. The expression pattern of *GmCHI2* overlaps with that of *GmIFS*. Both *GmIFS1* and *GmIFS2* are expressed in the seed coat and embryo as determined by Dhaubhadel et al. (2003).

2.4.4 Core GmCHIs nucleo-cytoplasmic; ‘GmFAPs’ plastidic

Findings that GmCHI subfamilies 1 and 2 localize to the nucleus and cytoplasm, taken in conjunction with the co-localization of other isoflavonoid biosynthetic and conjugating enzymes (Saslowsky et al. 2005; Dhaubhadel et al. 2008), raises the possibility of a concerted function for this machinery outside of the cytoplasm and the conjectured ER-tethered metabolon.

The transport of these functionally active CHIs across the nucleus is neither unprecedented, AtCHI localizes to the nucleus (Saslowsky et al. 2005), nor is it dependent necessarily on active sequestration. The intracellular transport of GmCHI-YFP fusions across the nuclear pore complex (Figure 2.7), in the absence of a known nuclear localization signal, is possible through diffusion, as they range between 50-60 kDa in size. The nuclear pore complex has been shown experimentally to be a selective barrier against the diffusion of molecules larger than 60 kDa (Tamura and Hara-Nishimura 2013; Weis 2003). I have confirmed that the YFP fluorescence perceived in *N. benthamiana* leaf epithelial cells resulted from the intact GmCHI-YFP fusion protein (approximately 50 kDa) and not from cleaved YFP (27 kDa) (Figure 2.10). Presence of an approximately 50 kDa band for each GmCHI-YFP shows that the protein was extracted from *N. benthamiana* as a whole fusion. In the case of GmCHI1B1, GmCHI1B2 and GmCHI4B, the protein was not expressed at a high enough level, due to technical difficulties, to be perceived by western blot analysis.

GmCHI subfamily 1 and 2 exhibit similar subcellular localization: therefore, if either family includes the putative 'type II' CHI(s), they are not differentially localized in correspondence with their role in isoflavonoid biosynthesis. Association and interaction studies between the CHI enzymes in subfamilies 1 and 2 and other isoflavonoid-specific enzymes might give us a clearer picture of the metabolon, and how the enzymes and their respective substrates are allocated to different pathways. It is possible that co-localization with other enzymes involved in the machinery might lead to differential subcellular localization or aggregation of the isoflavonoid-specific enzyme to a nucleating anchor in a sub-compartment of the cell, whether it is at the surface of the ER or in the nucleus.

The possibility of *in situ* isoflavonoid biosynthesis in the nucleus is improbable, due to the anchoring of GmIFS, the key pathway enzyme, to the ER. The inherent nature of P450 enzymes involved in phenylpropanoid production precludes the possibility of nuclear transport and synthesis, unless they are enzymatically cleaved post-translationally, transported by a tertiary protein, or as in the case of glucokinase, in yeast, assembling an NLS signal through interaction with a regulatory protein (Bosco et al. 2000). There has been no evidence to suggest such a process to date.

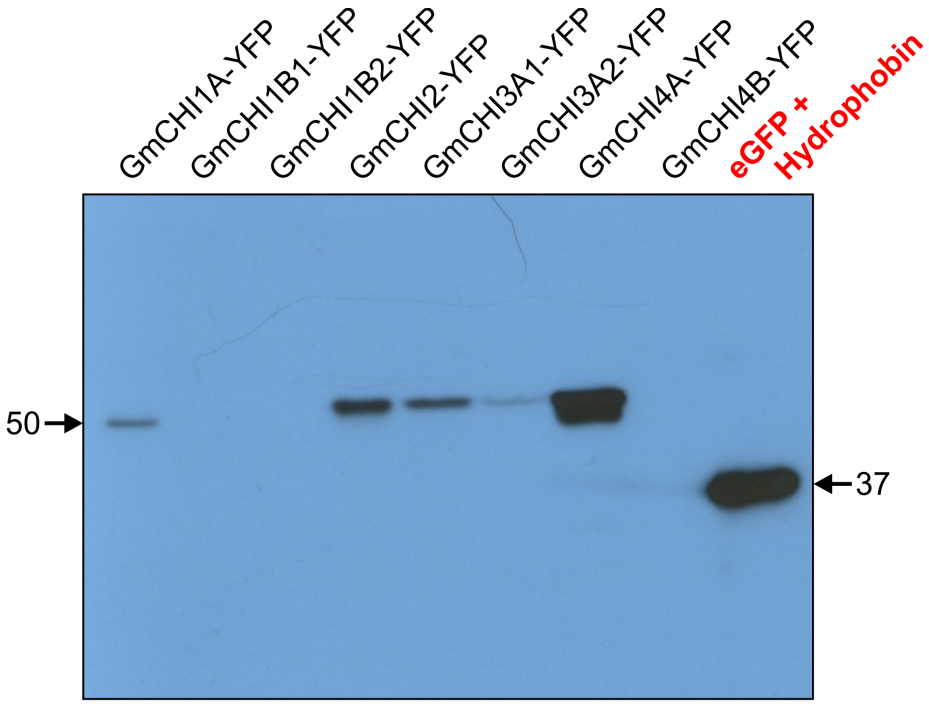


Figure 2.10 Western blot analysis of GmCHI-YFP proteins. Constructs containing translational fusions of GmCHI-YFP were transiently expressed in *N. benthamiana* leaves**.** Total soluble proteins were separated by SDS-PAGE and transferred into PVDF membrane by electroblotting. Fusion proteins were detected by sequential incubation of the blot with anti-GFP antibody and anti-rabbit IgG conjugated with horseradish peroxidase, followed by chemiluminescent reaction. eGFP fused with hydrophobin (~37 kDa) used as a control. All GmCHI-YFP (~50 kDa) appear as a single band indicating no cleavage of YFP tag; GmCHI1B1, GmCHI1B2 and GmCHI4B did not have adequate expression in the biological replicates shown in the present gel.

Therefore, it can be assumed that GmCHIs localized to the nucleus must be involved in a non-synthetic, alternative function, potentially in conjunction with other flavonoid and/or isoflavonoid machinery. Subfamily 3A of the CHIs that localized to the chloroplast in *N. benthamiana* and the plastids in soybean hairy roots, do not retain the critical active site residues of the CHI fold. Based on the phylogenetic analyses, the members of this family, GmCHI3A1 and GmCHI3A2, cluster together with AtFAP3. The FAPs appear to have a CHI-like fold, initially arising in mosses, providing the skeleton for flavonoid and isoflavonoid metabolism. The catalytic cleft of the FAPs in *Arabidopsis* is occupied by fatty acids (Ngaki et al. 2012). These catalytic residues appear to be conserved in the soybean GmCHI3 subfamily; therefore, I speculate that they are of similar function. The *Arabidopsis* FAPs also have an N-terminal chloroplast-transit sequence. They are expressed maximally in seeds, and their knockout plants show elevated α-linolenic acid levels, associated with fatty acid biosynthesis (Ngaki et al. 2012). In the transcript analysis, *GmCHI3A1* expression was present in the early stages of embryo development (30-50 DAP) coinciding with fatty acid storage. Therefore, based on sequence comparison, phylogenetic clustering, transcript analysis, and subcellular localization it can be deduced that the *GmCHI3* subfamily share a similar evolutionary lineage as the FAPs, and that *GmCHI3A1* is the member that appears to be involved in fatty acid biosynthesis.

2.5 Conclusion

Overall, I have identified twelve GmCHIs in soybean; four members, GmCHI1A, GmCHI1B1, GmCHI1B2 and GmCHI2 are *bona fide* GmCHIs; GmCHI3A1 and GmCHI3A2 are likely FAP3 isoforms that localize to the chloroplast; GmCHI3B1 and GmCHI3B2 are likely FAP1 enzymes that are predicted to localize to the chloroplast; GmCHI3C1 and GmCHI3C2 are predicted cytoplasmic FAP2; GmCHI4A and GmCHI4B are CHI-like isoforms that have not been functionally studied in soybean. Further investigation is required, *in planta*, into the conjunction of the 'type I and II' CHIs, with isoflavonoid structural enzymes to identify isoflavonoid-specific member(s). Based on phylogenetic, transcript and promoter analysis it can be speculated that GmCHI1A is involved in seed isoflavonoid production. My present findings suggest that

there is GmCHI isoform specificity at a sub-organ, tissue-specific level that changes temporally. Furthermore, the conundrum of GmCHIs in the nucleus can shed light on alternative function that might arise from their differential localization, although this must coincide with other biosynthetic machinery.

2.6 Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410

Aoki T, Akashi T, Ayabe S (2000) Flavonoids of leguminous plants: structure, biological activity, and biosynthesis. J Plant Res 113:475-488

Bassard JE, Mutterer J, Duval F, Werck-Reichhart D (2012a) A novel method for monitoring the localization of cytochromes P450 and other endoplasmic reticulum membrane associated proteins: a tool for investigating the formation of metabolons. FEBS J 279 (9):1576-1583

Bassard JE, Richert L, Geerinck J, Renault H, Duval F, Ullmann P, Schmitt M, Meyer E, Mutterer J, Boerjan W, De Jaeger G, Mely Y, Goossens A, Werck-Reichhart D (2012b) Protein-protein and protein-membrane associations in the lignin pathway. Plant Cell 24 (11):4465-4482

Bednar RA, Hadcock JR (1988) Purification and characterization of chalcone isomerase from soybeans. J Biol Chem 263 (20):9582-9588

Bosco D, Meda P, Iynedjian PB (2000) Glucokinase and glucokinase regulatory protein: mutual dependence for nuclear localization. Biochem J 348 Pt 1:215-222

Chen H, Seguin P, Jabaji S, Liu W (2011) Spatial distribution of isoflavones and isoflavone-related gene expression in high- and low-isoflavone soybean cultivars. Can J Plant Sci 91 (4):697-705

Cui H, Zhang ST, Yang HJ, Ji H, Wang XJ (2011) Gene expression profile analysis of tobacco leaf trichomes. BMC Plant Biol 11:76

Curtis MD, Grossniklaus U (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol 133 (2):462-469

Dhaubhadel S, Farhangkhoee M, Chapman R (2008) Identification and characterisation of isoflavonoid specific glycosyltransferase and malonyltransferase from soybean seeds. J Exp Bot 59 (4):981-994

Dhaubhadel S, Gijzen M, Moy P, Farhangkhoee M (2007) Transcriptome analysis reveals a critical role of *CHS7* and *CHS8* genes for isoflavonoid synthesis in soybean seeds. Plant Physiol 143 (1):326-338

Dhaubhadel S, McGarvey BD, Williams R, Gijzen M (2003) Isoflavonoid biosynthesis and accumulation in developing soybean seeds. Plant Mol Biol 53 (6):733-743

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300 (4):1005-1016

Fraser LG, Seal AG, Montefiori M, McGhie TK, Tsang GK, Datson PM, Hilario E, Marsh HE, Dunn JK, Hellens RP, Davies KM, McNeilage MA, De Silva HN, Allan AC (2013) An R2R3 MYB transcription factor determines red petal colour in an Actinidia (kiwifruit) hybrid population. BMC Genomics 14:28

Jones S, Gonzalez D, Vodkin L (2010) Flux of transcript patterns during soybean seed development. BMC Genomics 11 (1):136

Kang JH, McRoberts J, Shi F, Moreno JE, Jones AD, Howe GA (2014) The flavonoid biosynthetic enzyme chalcone isomerase modulates terpenoid production in glandular trichomes of tomato. Plant Physiol 164 (3):1161-1174

Kudou S, Fleury Y, Welt D, Magnolato D, Uchida T, Kitamura K (1991) Malonyl isoflavone glycosides in soybean seeds (*Glycine max* Merill). Agric Biol Chem 55:2227-2233

Libault M, Thibivilliers S, Bilgin DD, Radwan O, Benitez M, Clough SJ, Stacey G (2008) Identification of Four Soybean Reference Genes for Gene Expression Normalization. Plant Gen 1 (1):44-54

Livingstone JM, Seguin P, Strömvik MV (2010) An *in silico* study of the genes for the isoflavonoid pathway enzymes in soybean reveals novel expressed homologues. Can J Plant Sci 90 (4):453-469

McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R (2013) Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res 41 (Web Server issue):W597-600

Morita Y, Takagi K, Fukuchi-Mizutani M, Ishiguro K, Tanaka Y, Nitasaka E, Nakayama M, Saito N, Kagami T, Hoshino A, Iida S (2014) A chalcone isomerase-like protein enhances flavonoid production and flower pigmentation. Plant J 78:294-304

Nelson DE, Repetti PP, Adams TR, Creelman RA, Wu J, Warner DC, Anstrom DC, Bensen RJ, Castiglioni PP, Donnarummo MG, Hinchey BS, Kumimoto RW, Maszle DR, Canales RD, Krolikowski KA, Dotson SB, Gutterson N, Ratcliffe OJ, Heard JE (2007) Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proc Natl Acad Sci USA 104 (42):16450-16455

Ngaki MN, Louie GV, Philippe RN, Manning G, Pojer F, Bowman ME, Li L, Larsen E, Wurtele ES, Noel JP (2012) Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. Nature 485 (7399):530-533

Phillip DA (1992) Flavonoids:plant signals to soil microbes. In: Stafford HA, Ibrahim RK (eds) Recent advances in phytochemistry:phenolic metabolism in plants, vol 26. Plenum Press, New York, pp 201-223

Ralston L, Subramanian S, Matsuno M, Yu O (2005) Partial reconstruction of flavonoid and isoflavonoid biosynthesis in yeast using soybean type I and type II chalcone isomerases. Plant Physiol 137 (4):1375-1388

Ralston L, Yu O (2006) Metabolons involving plant cytochrome P450s. Phytochem Rev 5 (2):459-472

Ruuska SA, Girke T, Benning C, Ohlrogge JB (2002) Contrapuntal networks of gene expression during Arabidopsis seed filling. Plant Cell 14:1191-1206

Saslowsky DE, Warek U, Winkel BS (2005) Nuclear localization of flavonoid enzymes in Arabidopsis. J Biol Chem 280 (25):23735-23740

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463 (7278):178-183

Shirley BW, Kubasek WL, Storz G, Bruggemann E, Koornneef M, Ausubel FM, Goodman HM (1995) Analysis of Arabidopsis mutants deficient in flavonoid biosynthesis. Plant J 8 (5):659-671

Sparkes IA, Runions J, Kearns A, Hawes C (2006) Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat Protoc 1 (4):2019-2025

Stafford HA (1997) Roles of flavonoids in symbiotic and defense functions in legume roots. Botanical Review 63 (1):27-39

Stangeland B, Salehian Z (2002) An improved clearing method for GUS assay in *Arabidopsis* endosperm and seeds. Plant Mol Biol Rep 20:107-114

Stothard P (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. BioTechniques 28 (6):1102-1104

Subramanian S, Graham MA, Yu O, Graham TL (2005) RNA interference of soybean isoflavone synthase genes leads to silencing in tissues distal to the transformation site and to enhanced susceptibility to *Phytophthora sojae*. Plant Physiol 137:1345-1353

Subramanian S, Hu X, Lu GH, Odell JT, Yu O (2004) The promoters of two isoflavone synthase genes respond differentially to nodulation and defense signals in transgenic soybean roots. Plant Mol Biol 54 (5):623-639

Tamura K, Hara-Nishimura I (2013) The molecular architecture of the plant nuclear pore complex. J Exp Bot 64 (4):823-832

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 28 (10):2731-2739

Valkama E, Salminen JP, Koricheva J, Pihlaja K (2004) Changes in leaf trichomes and epicuticular flavonoids during leaf development in three birch taxa. Ann Bot 94 (2):233-242

Wang CS, Vodkin LO (1994) Extraction of RNA from tissues containing high levels of procyanidins that bind RNA. Plant Mol Biol 12:132-145

Weis K (2003) Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. Cell 112 (4):441-451

Winkel BSJ (2004) Metabolic channeling in plants. Annu Rev Plant Biol 55 (1):85-107

Winkel-Shirley B (2001) It takes a garden. How work on diverse plant species has contributed to an understanding of flavonoid metabolism. Plant Physiol 127 (4):1399-1404

Yu XH, Chen MH, Liu CJ (2008) Nucleocytoplasmic-localized acyltransferases catalyze the malonylation of 7-O-glycosidic (iso)flavones in Medicago truncatula. Plant J 55 (3):382-396

Zhang X, Henriques R, Lin SS, Niu QW, Chua NH (2006) Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. Nat Protoc 1 (2):641-646

CHAPTER THREE – TWIN ANCHORS OF THE ISOFLAVONOID METABOLON

3.1 Introduction

The phenylpropanoid pathway is a large and diverse biosynthetic network that produces a wide array of specialized metabolites, with competition between overlapping pathways dictating metabolic flux. One of the most well-characterized plant metabolons is that of flavonoid biosynthesis. As a focus of interaction and channelling studies, key enzymes in the pathway including CHS, CHI, F3H, DFR and FLS1 have been shown to interact (Burbulis and Winkel-Shirley 1999; Crosby et al. 2011; Hrazdina and Wagner 1985; Saslowsky and Winkel-Shirley 2001; Saslowsky et al. 2005; Winkel-Shirley 2002). However, there has been no direct evidence that this complex of soluble enzymes interacts with the P450s of that pathway, flavonoid 3′-hydroxylase (F3′H) or flavonoid 3′, 5′-hydroxylase (CYP75A, F3′5′H) (Ralston and Yu 2006).

Surprisingly, there has been very scant evidence for the speculated formation of an isoflavonoid metabolon, which would be in competition with the flavonoid pathway for common substrates and enzymes. The only prior finding was the indirect association of IOMT with IFS. Liu and Dixon (Liu and Dixon 2001) found that alfalfa IOMT translocated from the cytoplasm to the ER upon elicitor treatment, colocalizing with IFS. Data suggests that close association between IFS and IOMT is integral for the rapid production of the conjugated isoflavonoid daidzin.

A key branch-point enzyme of the phenylpropanoid pathway, chalcone isomerase (CHI), catalyzes the reaction producing flavanones, the skeletal backbone for many downstream metabolites. The reaction involves the cyclization of either a) a 4,2′,4′,6′-tetrahydroxychalcone (naringenin chalcone) to naringenin; or b) a 4,2′,4′-trihydroxychalcone (isoliquiritigenin chalcone) to liquiritigenin. Naringenin can then be channelled into an isoflavone genistein, or a plethora of different flavonoids. Liquiritigenin, however, leads into the legume-specific isoflavonoid pathway and is

converted by isoflavone synthase (IFS) into glycitein and daidzein, the latter being the precursor to glyceollins, an important group of phytoalexins. Therefore, the logic behind the presence of an isoflavonoid metabolon is apparent, with the competition between overlapping phenylpropanoid pathways.

The model of an isoflavonoid metabolon is predicated on the anchoring of the complex by P450s involved in the pathway, including the second enzyme in the phenylpropanoid pathway, cinnamate-4-hydroxylase (CYP73A, C4H) that forms *p*-coumarate, and the legume-specific isoflavone synthase (CYP93C, IFS) (Figure 1.3). These enzymes would form multiple metabolic centers for the aggregation of cytosolic reaction partners via protein-protein interaction and the sequential conversion of intermediates (Wagner and Hrazdina 1984; Hrazdina and Wagner 1985), finally producing isoflavones and their derived conjugates.

As soybean is a paleopolyploid with multi-gene families encoding the enzymes of the pathway, certain isoforms, through evolution (e.g. neo-functionalization) might have become preferentially attached to a certain metabolon, given certain tissue or environmental conditions. The identification of specific members of a proposed isoflavonoid metabolon, and the interactions that substantiate such a complex are investigated in this chapter.

In this chapter I present the localization of key enzymes in the pathway, GmCHS, GmCHR, GmCHI and GmIFS. Subsequently interaction between GmIFS and the aforementioned enzymes is displayed by *in planta* Bimolecular Fluorescence Complementation (BiFC) assays. Further, to expand the range of candidates for interaction with the putative nucleating factor of the metabolon, GmIFS, I explored a Co-Immunopreciptation (Co-IP) and LC-MS/MS approach. My findings show, for the first time that isoflavone synthase 2 (GmIFS2) interacts with upstream enzymes chalcone isomerase 1B1 (GmCHI1B1), cinnamate-4-hydroxylase (GmC4H) and aldehyde dehydratase (GmADT). Further, I present evidence of an isoflavonoid metabolon that includes tandem P450 enzymes, GmIFS2 and GmC4H, interaction with the adjacent

GmCHI1B1, as well as an interaction further upstream with the shikimate enzyme GmADT.

3.2 Materials and Methods

3.2.1 Plant materials and growth conditions

Refer to chapter 2 section 2.2.1 for *N. benthamiana* and soybean growth conditions. Soybean cv. Harosoy63 cotyledons were harvested from seedlings and used for hairy root transformation at six days of growth.

3.2.2 Cloning, plasmid construction and confocal microscopy

Genes of interest were amplified by PCR using gene specific primers (Table 3.1) and specific bands corresponding to the full-length gene were cloned into the pDONR-Zeo entry vector (Invitrogen, USA) and subsequently cloned into destination vectors as described in section 2.2.6.

For subcellular localization of GmCHS7, GmCHS8, GmCHR14, GmCHI2, GmIFS1 and GmIFS2, the entry clones were recombined, into the destination vector pEarelyGate101 to create a translational fusion of the gene of interest with the downstream reporter, yellow fluorescent protein (YFP) (Invitrogen, USA). Cloned vectors were transformed by electroporation into *E. coli* strain DH5α. Recombinant plasmids were purified from the *E. coli*, sequenced to confirm sequence integrity and transformed by electroporation into *A. tumefaciens* strain GV3101. Transient expression in *N. benthamiana* and YFP fluorescence were assayed after 48-72 hours, using a Leica TCS SP2 inverted confocal microscope, with a 63x water immersion objective, an excitation wavelength of 514 nm, and emissions collected between 530 and 560 nm. The experiment was repeated at least three times for each gene. The technical details are elaborated in sections 2.2.7 and 2.2.8.

Table 3.1 Sequence of oligonucleotides used chapter 3 for cloning of isoflavonoid pathway genes

Gene	Primer name	Sequence	Amplicon size (bp)
GmCHS7	CHS7-F	5'-GGGG ACA AGT TTG TAC AAA AAA GCA GGC TTC ATG GTT AGC GTA GCT GAG ATC AGG CAG GC-3'	1170
	CHS7-R	5'-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC GAT GGC CAC ACT ATG CAA AAC AAC AGT TTC-3'	
GmCHS8	CHS8-F	5'-GGGG ACA AGT TTG TAC AAA AAA GCA GGC TTC ATG GTG AGC GTA GCT GAG ATC CGC C-3'	1170
	CHS8-R	5'-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC GAT GGC CAC ACT GCG CAG AAC AAC AGT TTC-3'	
GmCHR	CHR-F	5'-GGGG ACA AGT TTG TAC AAA AAA GCA GGC TTC CTT GTC AAC CCT TTG AGA GTT AGA ATG-3'	1002
	CHR-R	5'-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC TAT TTG ATC ATC CCA GAG ATC AGC-3'	
GmIFS1	IFS1-F	5'-GGGG ACA AGT TTG TAC AAA AAA GCA GGC TTC CAA CAA CAG TTC TTG CAC TGA GG-3'	1596
	IFS-R	5'-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC AGA AAG GAG TTT AGA TGC AAC GCC -3'	
GmIFS2	IFS2-F	5'-GGGG ACA AGT TTG TAC AAA AAA GCA GGC TTC CAT AAG TAG AGT ATG GTG GAG TAG ATG C-3'	1693

3.2.3 Bi-molecular Fluorescence Complementation (BiFC) assay

Protein-protein interactions and their subcellular localization were studied using Bimolecular Fluorescence Complementation (BiFC) assays. This technique functions through the use of split fluorescent protein (eYFP) segments that are brought together to form a functional fluorophore as a result of protein-protein interaction. Genes of interest (*GmCHS7, GmCHS8, GmCHIR, GmCHI2, GmIFS1* and *GmIFS2*) cloned in pDONR-ZEO vector were recombined separately into destination vectors pEarleyGate201-YN and pEarleyGate202-YC (Lu et al. 2010), using Gateway technology (Invitrogen, USA). The YN and YC vectors incorporate the N-terminal (1-174 aa) and C-terminal (175-239 aa) segments of eYFP, downstream of the gene of interest, respectively. Cloned vectors were transformed as above into *E. coli*; and recombinant plasmids sequenced, and transformed into *A. tumefaciens* strain GV3101. *Nicotiana benthamiana* leaves were co-infiltrated with 1:1 combinations of *A. tumefaciens* carrying fusions containing the YN and YC fragments (Sparkes et al. 2006). Transient expression and YFP fluorescence were assayed after 48 hours using confocal microscopy, as with the subcellular localization.

3.2.4 Hairy root transformation; protein extraction and co-immunoprecipitation

Isoflavone synthase 2 (*GmIFS2*) cloned into pEarleyGate101 and an empty pEarleyGate104 vector, containing YFP, were transformed into *A. rhizogenes* strain K599 via electroporation. Six day-old soybean cotyledons were harvested and transformed by *A. rhizogenes* containing the constructs using the hairy root transformation system (Subramanian et al. 2004). Transgenic hairy roots were selected 20-30 days post-infiltration, using a Leica MZ FL III fluorescence stereomicroscope, and an YFP filter (excitation 510/20 nm; barrier filter 560/40 nm). Roots were harvested and frozen with liquid nitrogen.

Total protein was extracted from 1 g of frozen root tissue using 1 mL of extraction buffer A [50 mM HEPES pH 7.5, 150 mM NaCl, 0.5% (v/v) Triton X-100, 0.1% (v/v) Tween-20, 0.2% (v/v) sodium orthovanadate, 200 mM and 2X protease inhibitor cocktail (Roche Applied Science, QE, Canada)]. After mixing well and 15 min of incubation on ice, the protein extract was diluted 1 : 1with extraction buffer B (extraction buffer without Triton

X-100 or Tween-20) and centrifuged for 30 min at 10,000 g at 4°C to sediment debris. Supernatant was transferred to a new tube for Co-IP. The above is a description of 'moderate' stringency extraction, while for 'high' stringency extraction 2 mL of Buffer A was used and no dilution with buffer B.

Co-immunoprecipitation was conducted using μMAC epitope tag protein isolation kit (Miltenyi Biotech Inc. CA, USA). For this, crude protein extract was incubated on ice with anti-GFP micro beads for 30 min. The samples were passed through a μcolumn provided in the kit, and the column was washed ten times with extraction Buffer A. Finally, proteins bound to the column were eluted with 0.1M Na_2CO_3 (pH 11.3) in 1M MES (pH 3.0). The eluate was quantified using a Bradford assay.

3.2.5 SDS-PAGE and silver staining

Proteins were separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) on 10% acrylamide gels, with a run time 1 hour and 30 min at 100 V to 150 V. Gels were treated with 50% methanol, 12% acetic acid and 0.05% formaldehyde for 1 hour. Washed three times with 50% ethanol and treated with 0.02% sodium thiosulphate for 1 min. Incubated with 0.2% Silver nitrate and 0.075% formaldehyde for 1 h. Gels were developed with 6% sodium carbonate and 0.075% formaldehyde. Colour development was stopped using 0.5% glycine.

3.2.6 Western blot

Proteins were separated on an SDS-PAGE gel (10%) and electro-blotted onto a PVDF membrane (Bio-Rad, ON, Canada). The fusion protein of GmIFS2-YFP and that of YFP-only were identified by sequential incubation of the blot with anti-GFP monoclonal antibody (Clontech, CA, USA) at a dilution of 1:3000 followed by HRP-conjugated goat anti-mouse secondary antibody (Pierce, IL, USA) at a dilution of 1:5000. The HRP detection was performed by using Super Signal West Femto (ThermoScientific, USA).

3.2.7 Liquid chromatography tandem mass spectrometry (LC-MS/MS)

Protein bands from SDS-PAGE silver-stained gels were spot-picked using the Ettan Spot-Picker (Amersham Biosciences, NJ, USA). Samples were subjected to in-gel digestion with trypsin. In-gel digestion was performed by the MALDI Mass Spectrometry Facility, Western University, using a MassPREP automated digester station (PerkinElmer). The in-gel digested samples were analysed by LC-MS/MS. NanoAcquity UPLC® columns (Waters, USA) were used for LC and MS/MS was performed using the Orbitrap Elite™ Hybrid Ion Trap-Orbitrap Mass Spectrometer (ThermoScientific, USA), in the Biological Mass Spectrometry Lab, Schulich School of Medicine and Dentistry, Western University.

3.3 Results

3.3.1 Subcellular localization of isoflavonoid pathway enzymes

The P450 enzymes GmIFS1 and GmIFS2 are both functionally active and play a role in isoflavonoid biosynthesis in soybean (Jung et al. 2000; Subramanian et al. 2005). Cytochrome P450 enzymes are mainly localized to the endoplasmic reticulum (ER), where they could serve as a nucleating factor for a putative enzyme complex. To investigate the possibility of interaction between isoflavonoid pathway enzymes, subcellular localization of the selected enzymes was first established. This was accomplished by creating translational fusions of isoflavonoid biosynthesis genes: *GmCHS7, GmCHS8, GmCHR14, GmCHI2, GmIFS1* and *GmIFS2,* upstream of the reporter gene *YFP* and transiently expressed, by a constitutive 35S promoter, in *N. benthamiana* leaf epidermal cells (Figure 3.1a-h).

Two soybean genes encoding CHS in soybean have been shown to be involved in isoflavonoid metabolism, *GmCHS7* and *GmCHS8* (Dhaubhadel et al. 2007; Yi et al. 2010). Here, their localization is shown. Prediction of the subcellular transport of these proteins, by *in silico* research, indicated cytoplasmic localization, with no presence of transit, nuclear or other signal peptides in the deduced amino acid sequence. The translational fusions of GmCHS7 and GmCHS8 showed differential subcellular

localization in *N. benthamiana* leaf epidermal cells: GmCHS7 was localized to the cytoplasm, while GmCHS8 was localized to the nucleus and cytoplasm (Figure 3.1 a, b).

The soybean gene encoding *GmCHR14* (Glyma.14G005700) was chosen as a representative *CHR* due to the suggested role of this gene in isoflavonoid biosynthesis (Graham et al. 2007). Subcellular localization of GmCHR14 *in silico,* and confirmed *in planta* was nucleo-cytoplasmic (Sepiol and Dhaubhadel, unpublished result) (Figure 3.1 c). The localization of all four functional GmCHIs (GmCHI1A, GmCHI1B1, GmCHI1B2 and GmCHI2) to the cytoplasm and the nucleus was shown in chapter 2 (Figure 2.8 a-d); and repeated here as well is that of GmCHI2 (Figure 3.1 d).

GmIFS1 and GmIFS2 localize to the ER in *N. benthamiana* epidermal leaves (Figure 3.1 e-h). GmIFS2 (CYP93C1) localization to the ER has previously been established by subcellular localization with fluorescent protein fusion (Ralston and Yu 2006). My findings reaffirm the properties of both isoforms as integral ER membrane P450 enzymes. Localization to the ER is apparent in the reticulate pattern of fluorescence (Figure 3.1 f, h), covering parts of the cell, around the nuclear membrane, but absent inside the nucleus (Figure 3.1 e-h). The deduced amino acid sequences of GmIFS1 and GmIFS2 lack KDEL/HDEL sequences (ER retention signal), but have a 23bp signal peptide at the N-terminus, which consists of largely hydrophobic sequences (SignalP, www.cbs.dtu.dk/services/SignalP/). The localization and sequence of GmIFS1 and GmIFS2 conforms to the general hypothesis of P450 enzymes that are inserted into the ER membrane co-translationally by their N-terminal hydrophobic domains. There was no visible difference in the localization patterns of the two isoforms GmIFS1 and GmIFS2 in *N. benthamiana*.

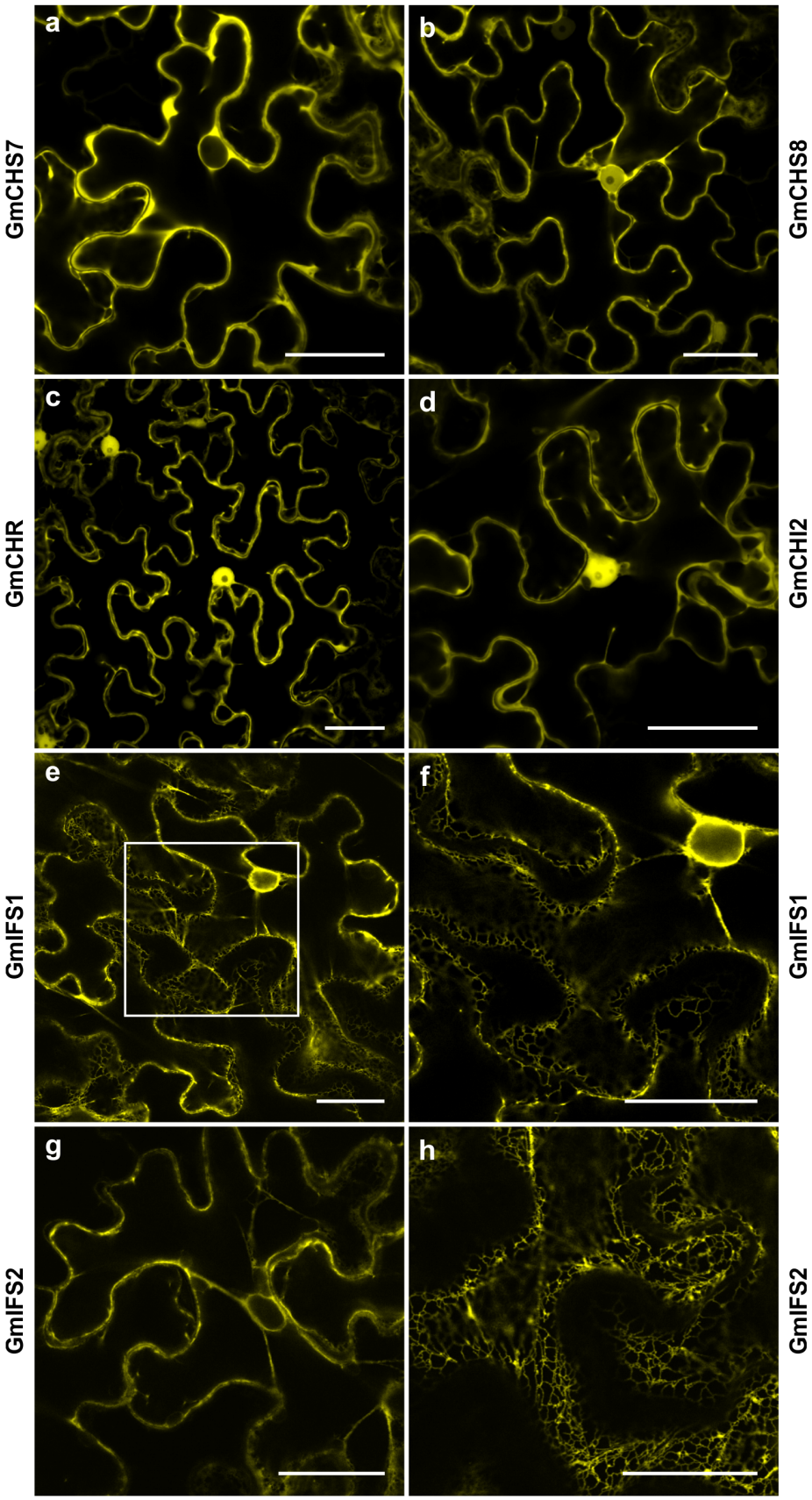

a
b
c
d
e
f
g
h
GmCHS7
GmCHS8
GmCHR
GmCHI2
GmIFS1
GmIFS1
GmIFS2
GmIFS2

Figure 3.1 Subcellular localization of the isoflavonoid pathway enzymes; (a) GmCHS7 localizes to the cytoplasm; (b) GmCHS8 localizes to the nucleus and cytoplasm; (c) GmCHR14 localizes to the nucleus and cytoplasm (Sepiol and Dhaubhadel, unpublished); (d) GmCHI2 localizes to the nucleus and cytoplasm; (e) GmIFS1 localizes to the endoplasmic reticulum; (f) magnified image of reticulate structure of GmIFS1 localization from 3.1e; (g) GmIFS1 localizes to the endoplasmic reticulum; (h) reticulate structure of GmIFS2 localization. Scale bar indicates 40 μm.

3.3.2 Interactions between core isoflavonoid pathway enzymes

To assess whether IFS interacts with upstream enzymes in the isoflavonoid pathway to form a potential enzyme complex, I investigated interactions between GmIFS1/GmIFS2 and proteins of interest: GmCHS7, GmCHS8, GmCHR14 and GmCHI2, by BiFC. The principle of BiFC is that interacting proteins fused with either N-terminal (YN) or C-terminal (YC) halves of YFP and transiently expressed in *N. benthamiana* will bring the two halves in close proximity to each other, resulting in fluorescence (Ohad et al. 2007). Protein expression was monitored by confocal microscopy. Fluorescence denotes either direct or indirect interaction. Therefore, consecutive and non-consecutive enzymes in a cohesive metabolon should produce fluorescence when assayed by BiFC. Additionally, this method determines the subcellular location of interaction. Interactions were assayed with combinations being tested in both directions, e.g. GmIFS1-YN with GmCHS7-YC; GmIFS1-YC with GmCHS7-YN. Each construct was also assayed with an empty vector control to record any background fluorescence from interactions between a specific gene and YFP fragments (YN or YC).

GmIFS1 and GmIFS2 showed positive interaction with GmCHS7, GmCHS8, and GmCHI2 (Figure 3.2.1-4). GmCHR14 interaction was assayed with all the proteins of interest, but only detected with GmIFS2 (Figure 3.2.3). All interactions with GmIFS1/GmIFS2 had a reticulate ER pattern of localization. Despite the cytoplasmic (GmCHS7) and nucleo-cytoplasmic (GmCHR14, GmCHS8 and GmCHIs) localization of the pathway enzymes, their interactions with GmIFS1/GmIFS2 appear to be restricted to the surface of the ER and are absent from other organelles (Figure 3.2.1 a-l). The intensity of interaction was very strong between GmIFS2 and GmCHR14 (Figure 3.2.3), and between GmIFS1/GmIFS2 and GmCHI2 (Figure 3.2.2); interaction of GmIFS1/GmIFS2 with GmCHS7/GmCHS8 was weaker (Figure 3.2.1 a-l). These differences in fluorescence intensity could indicate differences in proximity of adjacent and non-adjacent enzymes, but cannot be quantitatively measured with this assay. Evidence for the interaction between both isoforms of GmCHS and all functional GmCHIs was found (shown for GmCHI2 only) (Figure 3.2.4). However, as GmCHS7 and GmCHS8 have differential localization, the same applied to their

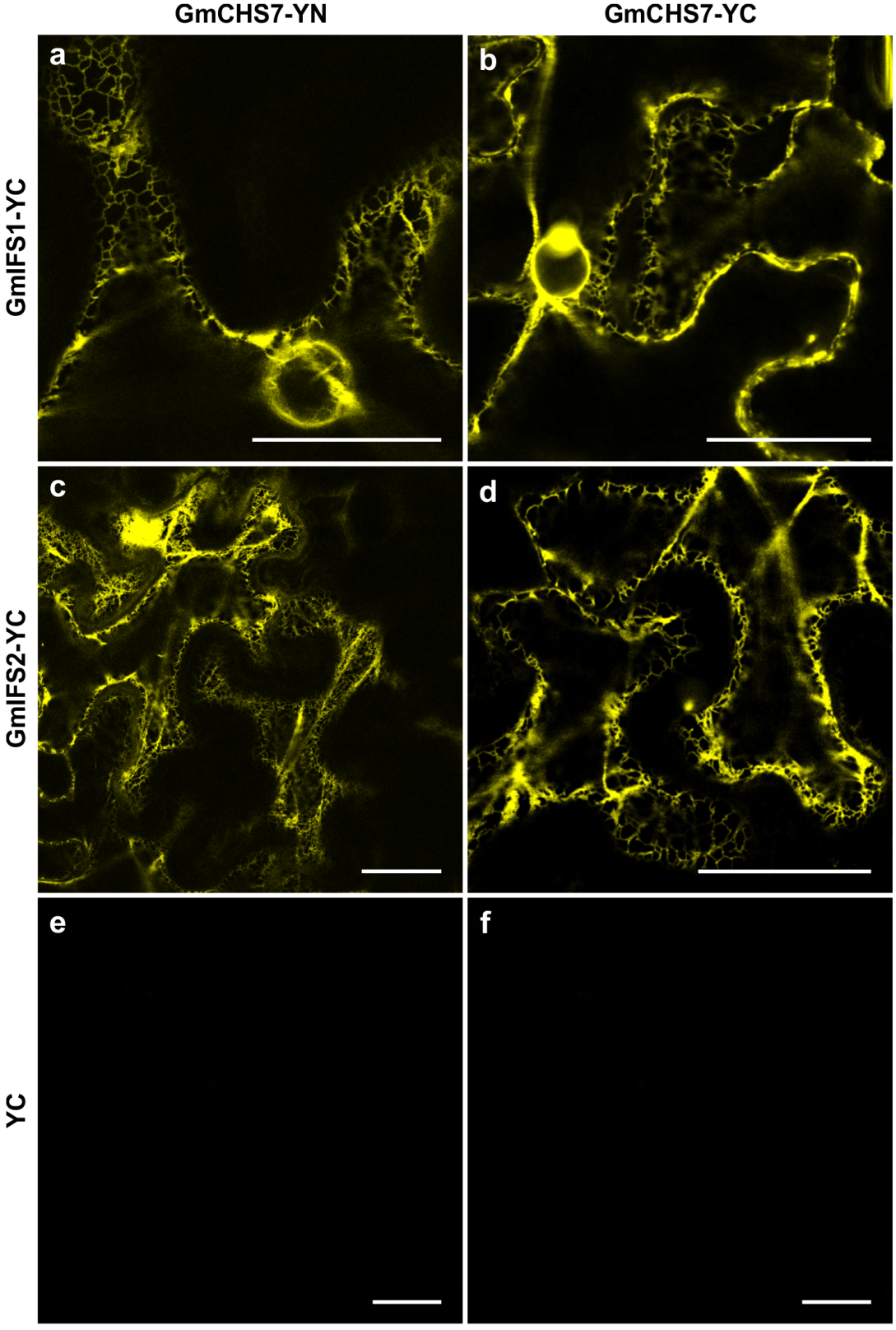

GmCHS7-YN
GmCHS7-YC
GmIFS1-YC
GmIFS1-YN
GmIFS2-YC
GmIFS2-YN
YC
YN
a
b
c
d
e
f

Figure 3.2 Interaction of isoflavonoid pathway enzymes by BiFC: bi-directional interaction between proteins assayed by the co-expression of translational fusions with N-terminal (YN) and C-terminal (YC) fragments of YFP. Proximity of the two fragments forms a whole YFP protein. Fluorescence indicates the presence and localization of the interaction.

Figure 3.2.1 a-f Interaction of GmCHS7 with GmIFS1/GmIFS2 is localized to the ER: (a) GmCHS7-YN with IFS1-YC, (b) GmCHS7-YC with IFS1-YN, (c) GmCHS7-YN with IFS2-YC, (d) GmCHS7-YC with IFS2-YN, (e) GmCHS7-YN with YC, (f) GmCHS7-YC with YN. Scale bar indicates 40 μm.

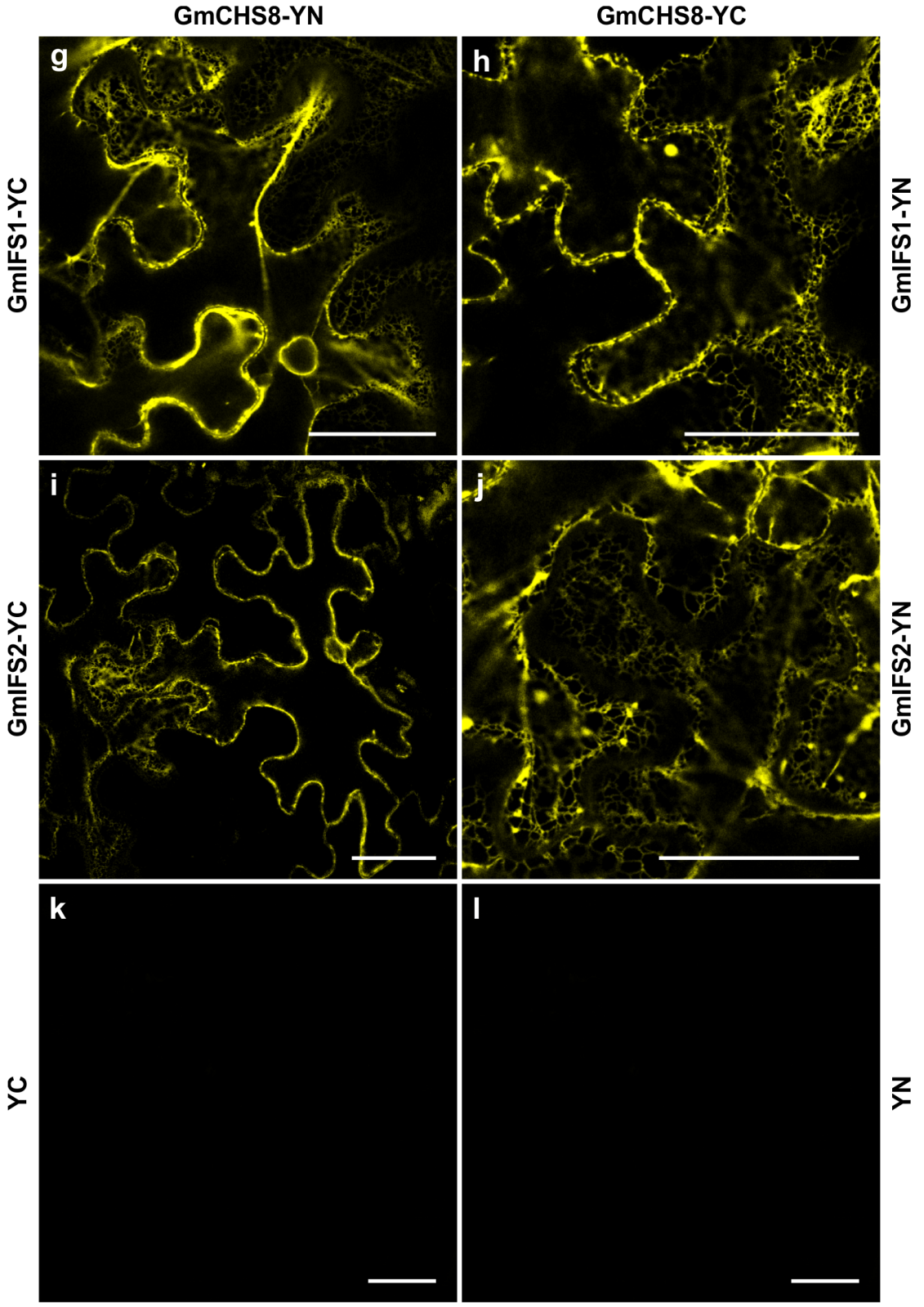

GmCHS8-YN
GmCHS8-YC
GmIFS1-YC
GmIFS1-YN
GmIFS2-YC
GmIFS2-YN
YC
YN
g
h
i
j
k
l

Figure 3.2.1 g-h Interaction of GmCHS8 with GmIFS1/GmIFS2 is localized to the ER: (g) GmCHS8-YN with IFS1-YC, (h) GmCHS8-YC with IFS1-YN, (i) GmCHS8-YN with IFS2-YC, (j) GmCHS8-YC with IFS2-YN, (k) GmCHS8-YN with YC, (l) GmCHS8-YC with YN. Scale bar indicates 40 μm.

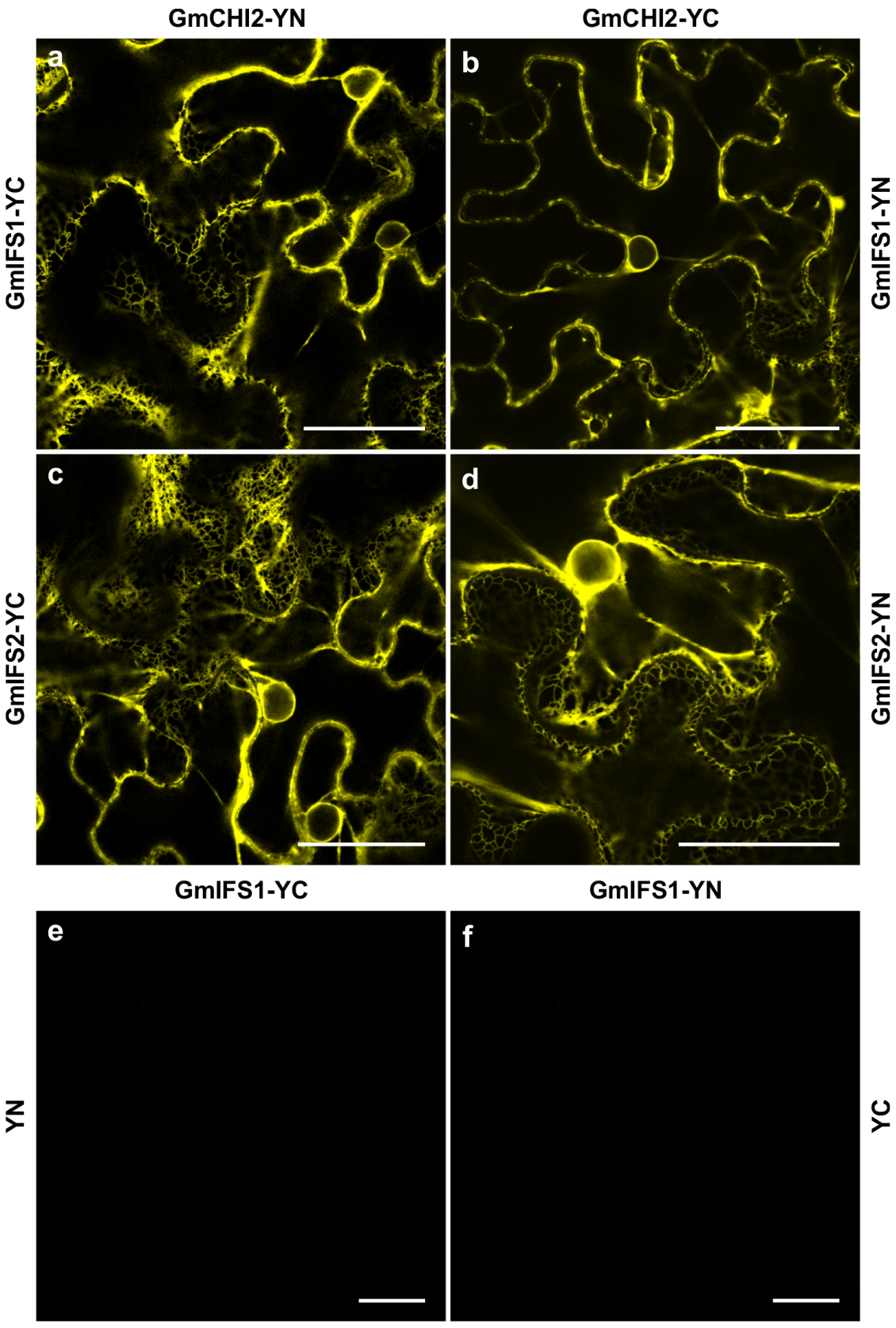

GmCHI2-YN
GmCHI2-YC
a
b
GmIFS1-YC
GmIFS1-YN
c
d
GmIFS2-YC
GmIFS2-YN
GmIFS1-YC
GmIFS1-YN
e
f
YN
YC

Figure 3.2.2 Interaction of GmCHI2 with GmIFS1/GmIFS2 is localized to the ER: (a) GmCHI2-YN with GmIFS1-YC, (b) GmCHI2-YC with GmIFS1-YN, (c) GmCHI2-YN with GmIFS2-YC, (d) GmCHI2-YC with GmIFS2-YN, (e) GmIFS1-YC with YN, (f) GmIFS2-YN with YC. Scale bar indicates 40 μm.

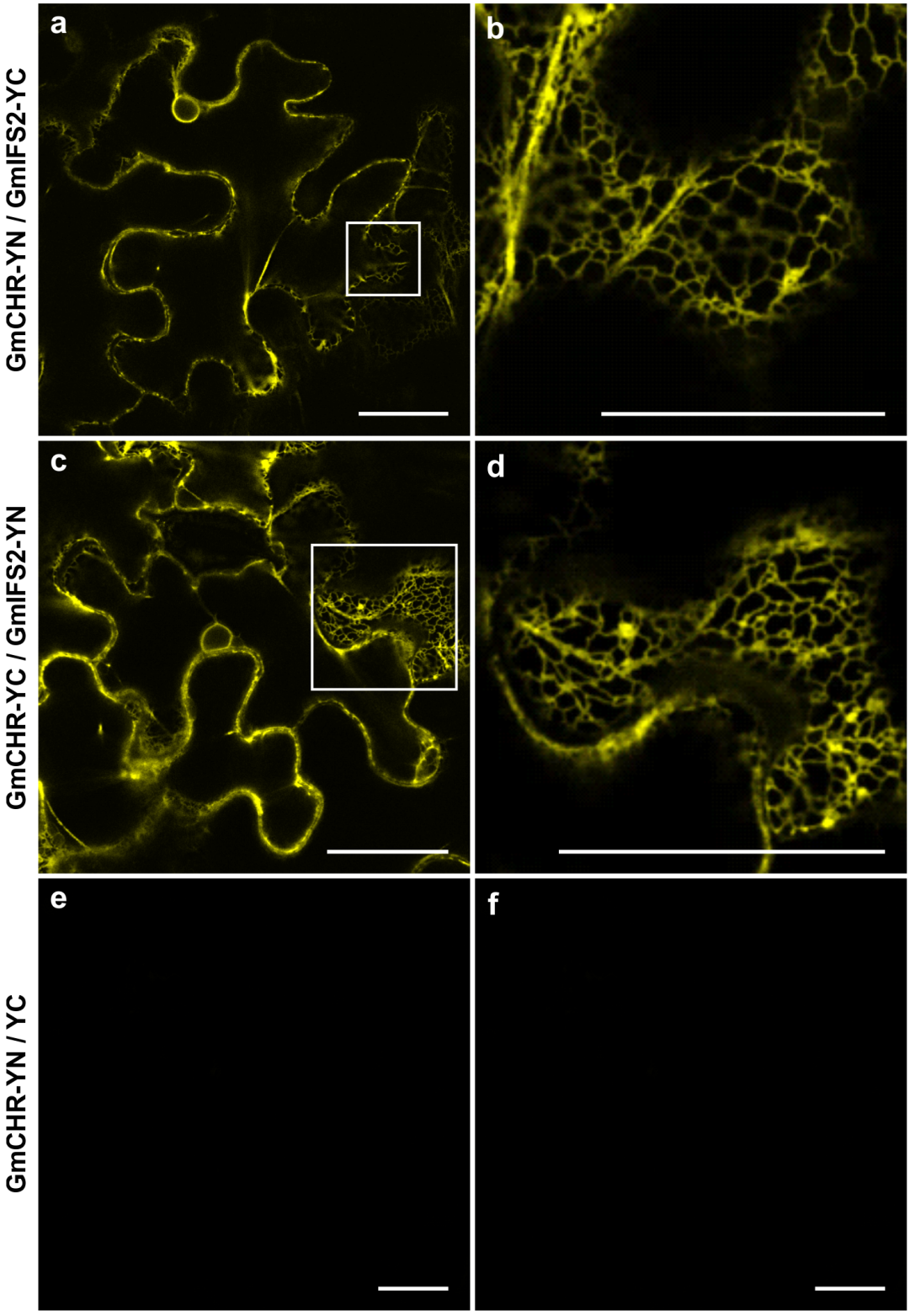

a
b
c
d
e
f
GmCHR-YN / GmIFS2-YC
GmCHR-YC / GmIFS2-YN
GmCHR-YN / YC
GmCHR-YC / YN

Figure 3.2.3 Interaction of GmCHR14 with GmIFS2 is localized to the ER: (a, b) GmCHR14-YN with GmIFS2-YC, (b) magnified image of reticulate ER localization, (c, d) GmCHR14-YC with GmIFS2-YN, (d) magnified image of reticulate ER localization, (e) GmCHR14-YN with YC, (f) GmCHR14-YC with YN. Scale bar indicates 40 μm.

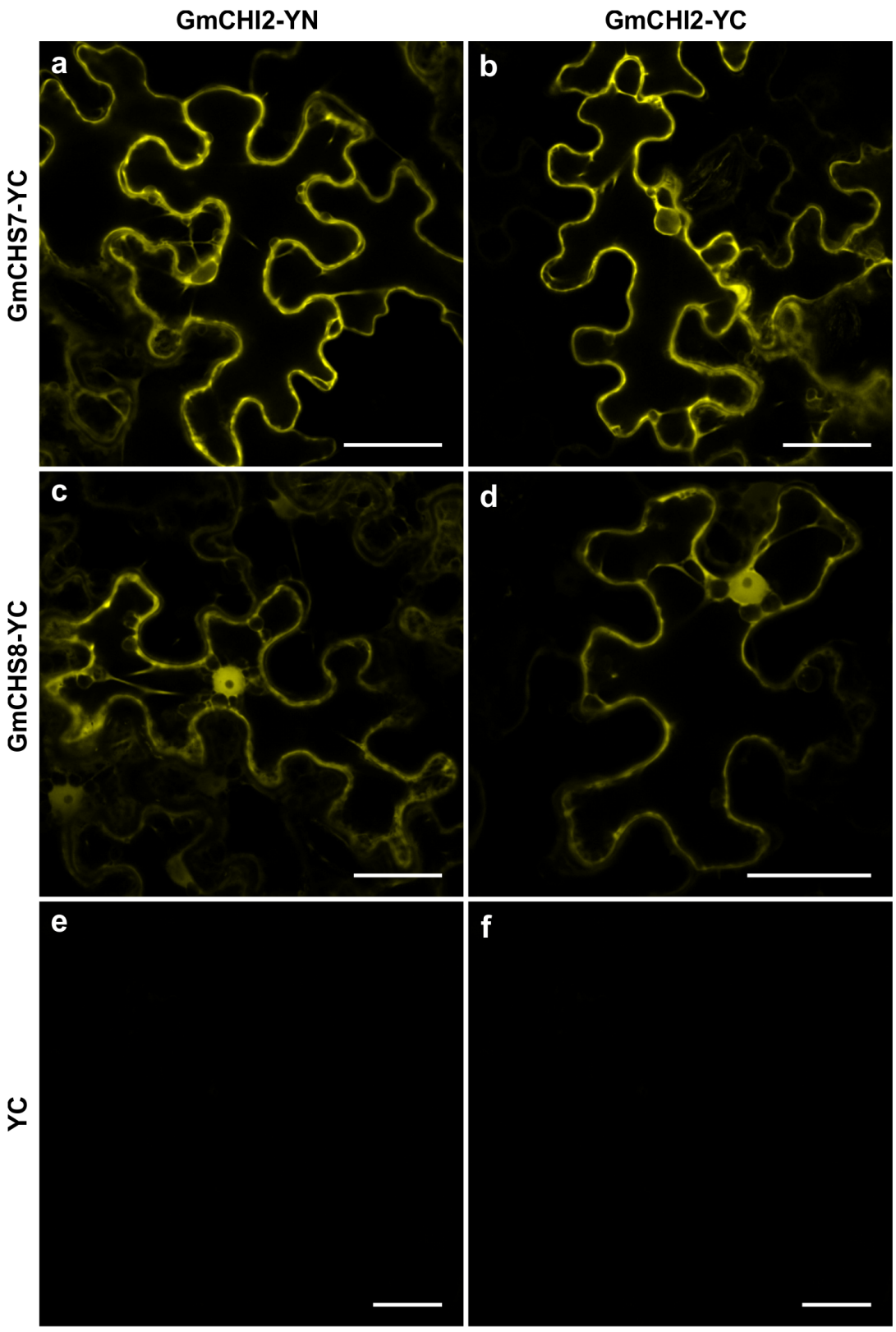

GmCHI2-YN
GmCHI2-YC
GmCHS7-YC
GmCHS7-YN
GmCHS8-YC
GmCHS8-YN
YC
YN
a
b
c
d
e
f

Figure 3.2.4 Interaction of GmCHI2 with GmCHS7 is localized to the cytoplasm: (a) GmCHI2-YN with GmCHS7-YC, (b) GmCHI2-YC with GmCHS7-YN; **interaction of GmCHI2 with GmCHS8 is localized to the nucleus and cytoplasm:** (c) GmCHI2-YN with GmCHS8-YC, (d) GmCHI2-YC with GmCHS8-YN; (e) GmCHI2-YN with YC, (f) GmCHI2-YC with YN. Scale bar indicates 40 μm.

interactions: GmCHS7 interacted with GmCHIs in the cytoplasm; GmCHS8 interacted with GmCHIs in the cytoplasm and the nucleus.

There was no discrepancy between reciprocal combinations in all experiments; similar interaction and location were recorded. No interaction was seen between constructs containing genes of interests and empty vector controls. These BiFC experiments provide evidence for the protein-protein interactions of GmIFS isoforms with GmCHI2 and GmCHS7/GmCHS8, and GmIFS2 interaction with GmCHR14.

3.3.3 Co-IP based identification of GmIFS2 interacting candidates

The two GmIFS isoforms are differentially expressed and induced; GmIFS1 is ubiquitously expressed across soybean tissues, while GmIFS2 appears to be expressed in correlation with accumulation of isoflavonoid content in developing embryos. GmIFS2 is also inducible by infection with *Phytophthora sojae* (Dhaubhadel et al. 2003). Based on this premise, and the additional evidence of GmIFS2 interaction with the isoflavonoid-specific GmCHR14 (Figure 3.2.3), GmIFS2 was chosen as the bait protein for Co-IP based investigation of the isoflavonoid metabolon.

In order to identify GmIFS2 interaction partners, the full-length gene fused upstream of YFP was expressed in soybean hairy roots using the hairy root transformation system. Hairy roots expressing YFP (pEarleyGate104, vector only) were used as a control; hereafter, referred to as 'YFP-only' roots. Three biological replicates were prepared for each construct from pooled hairy roots. Total protein was extracted according to the protocol outlined in section 3.2.5, high detergent stringency used for 'GmIFS2-YFP replicate 1' and moderate detergent stringency for 'GmIFS2-YFP replicates 2 and 3'. Moderate detergent stringency was applied for protein extraction of all three YFP-only replicates.

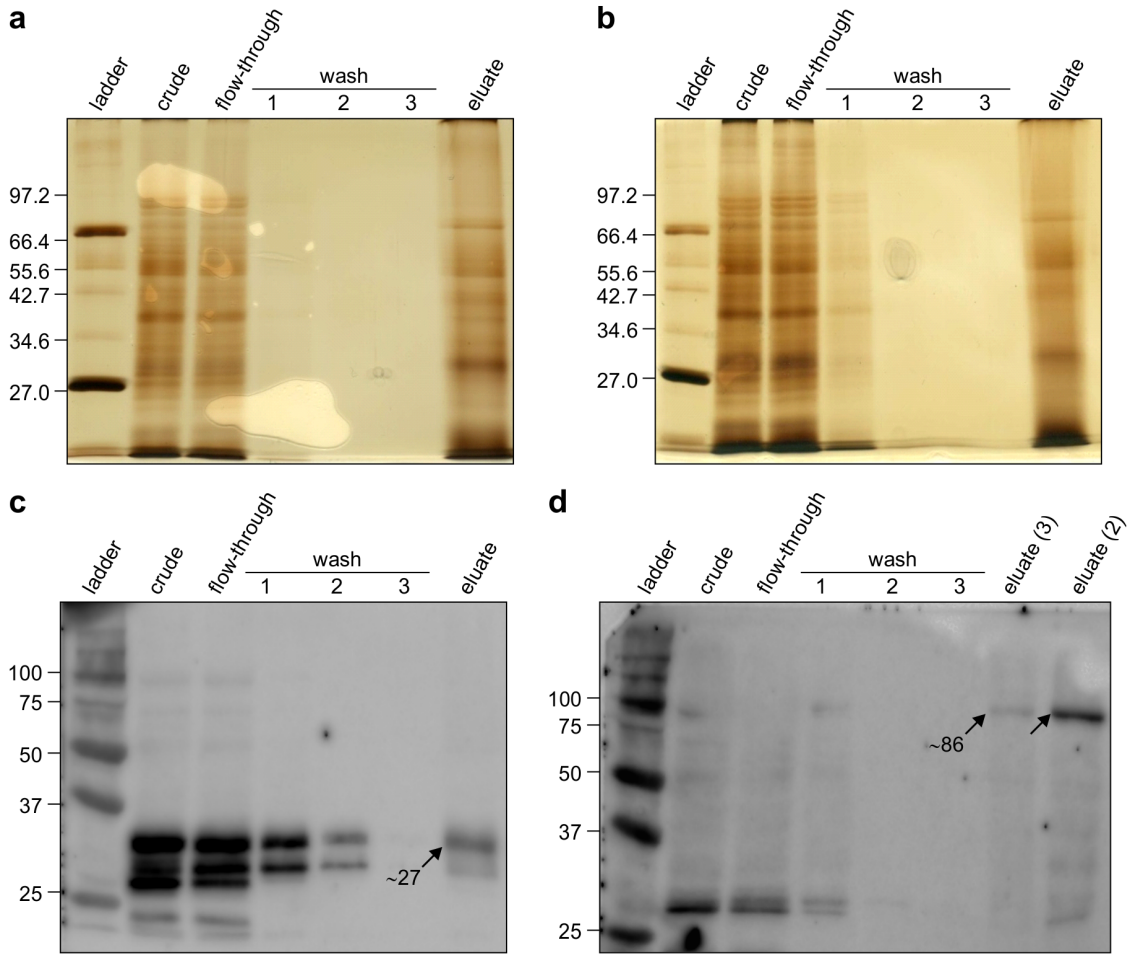


Figure 3.3 Protein extraction and co-immunoprecipitation: from soybean hairy roots containing (a, c) YFP, pEarelyGate104 and (b, d) GmIFS2-YFP, pEarleyGate101-IFS2; (a, b) proteins separated on an SDS-PAGE gel by electrophoresis and silver-stained; (c, d) western blot of YFP and GmIFS2-YFP (eluates from replicates 2 and 3), identified by sequential incubation of the blot with anti-GFP monoclonal antibody followed by HRP-conjugated goat anti-mouse secondary antibody. Crude: crude protein extract from soybean hairy roots; flow-through: crude extract incubated with anti-GFP microbeads and applied to µcolumn, with the flow-through collected; wash: sequential wash steps with buffer A; eluate: final step of co-immunoprecipitation, elution of bound proteins from the column.

Figure 3.3 a, b shows representative silver-stained SDS-PAGE gels of the protein extraction and Co-IP process for GmIFS2-YFP and YFP-only. Comparison between the protein profile of the crude and the eluate shows the preferential selection of interacting proteins. Western blot analysis against YFP using anti-GFP antibody show the presence of the bait fusion (GmIFS2-YFP, ~86 kDa) and control (YFP-only, 27 kDa) proteins, respectively (Figure 3.3 c, d).

The protein mixtures eluted from each sample were spot-picked, trypsin in-gel digested, and identified by LC-MS/MS. A total of 502 proteins were identified from the overlap of the three GmIFS2-YFP replicates (Figure 3.4.1 a); 316 proteins were identified from the overlap of the three YFP-only replicates (Figure 3.4.1 b). The proteins from these respective overlaps were compared against each other, resulting in a third Venn diagram (Figure 3.4.1 c). Proteins present in both GmIFS2-YFP and YFP-only lists, represent abundant proteins that are possibly interacting with YFP and are not specific to the GmIFS2 interaction. From the total 502 IFS2-YFP proteins, 330 proteins are specific to GmIFS2 and 172 proteins are shared with YFP.

The set of 330 proteins were then compiled into a list based on *Arabidopsis* homolog TAIR identifiers (The Arabidopsis Information Resource, http://arabidopsis.org) and subjected to GO (Gene Ontology Database, http://geneontology.org) functional annotation. The resulting graphs (Figure 3.5 a-c) represent the molecular and cellular function, and cellular compartmentalization of this protein list. The list of 330 putative GmIFS2-interacting proteins includes 10% ER resident proteins and 42% responsive to stress.

Interesting candidates for interaction with the isoflavonoid metabolon were selected after a quality check of the respective mass spectra, and only if they were identified by more than two unique peptides with significant PEAKS score. Selection was also limited to potential enzymes involved in the shikimate or phenylpropanoid pathway. The four selected GmIFS2-interacting proteins include: two isoforms of arogenate dehydratase from subfamily 6 (GmADT6), cinnamate 4-hydroxylase (GmC4H), and chalcone isomerase (GmCHI1B1) and the bait itself (Table 3.2).

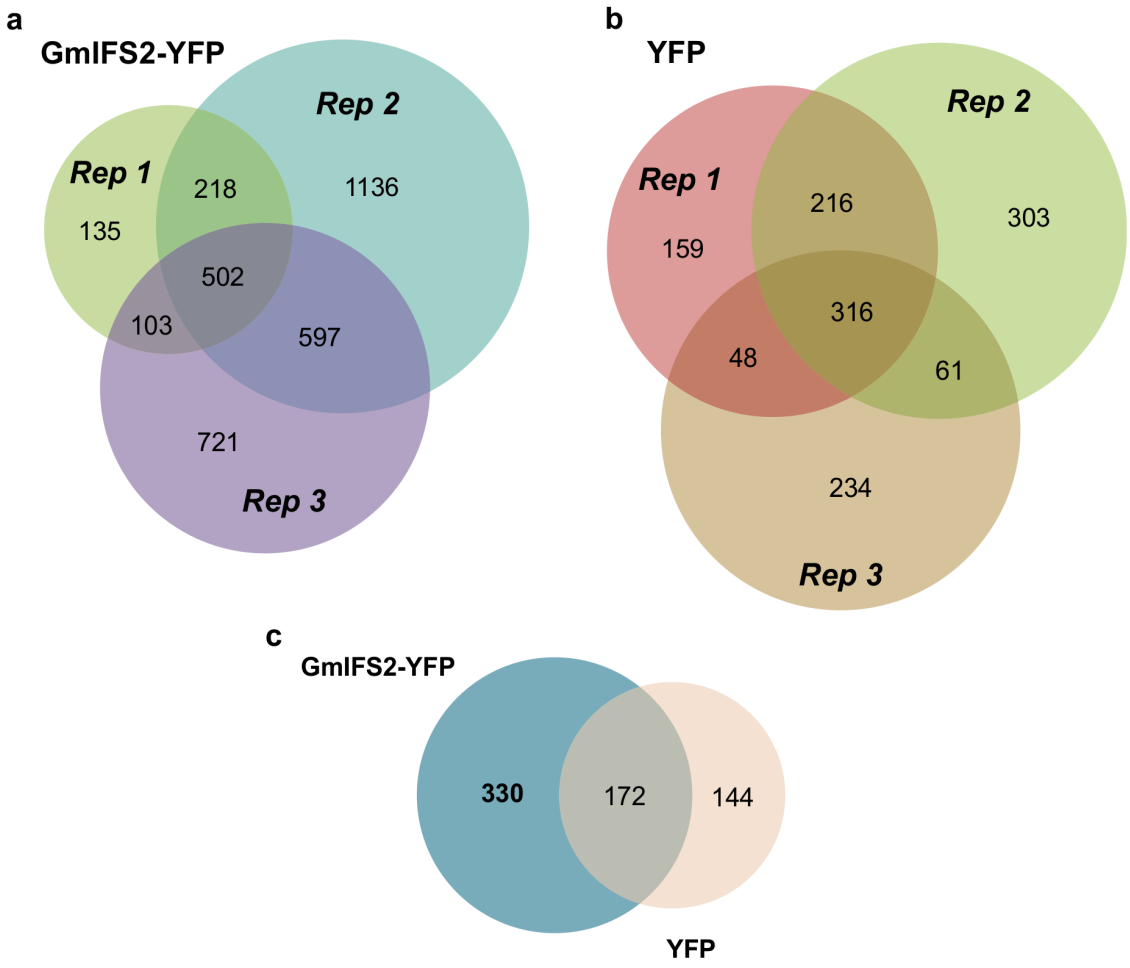


Figure 3.4 Overlap of interacting candidate proteins identified by LC-MS/MS analysis; biological replicates 1, 2 and 3 of (a) GmIFS2-YFP interacting proteins; (b) YFP interacting proteins; (c) Overlap of GmIFS2-YFP interacting proteins with YFP interacting proteins; 330 candidate proteins putatively interacting only with GmIFS2

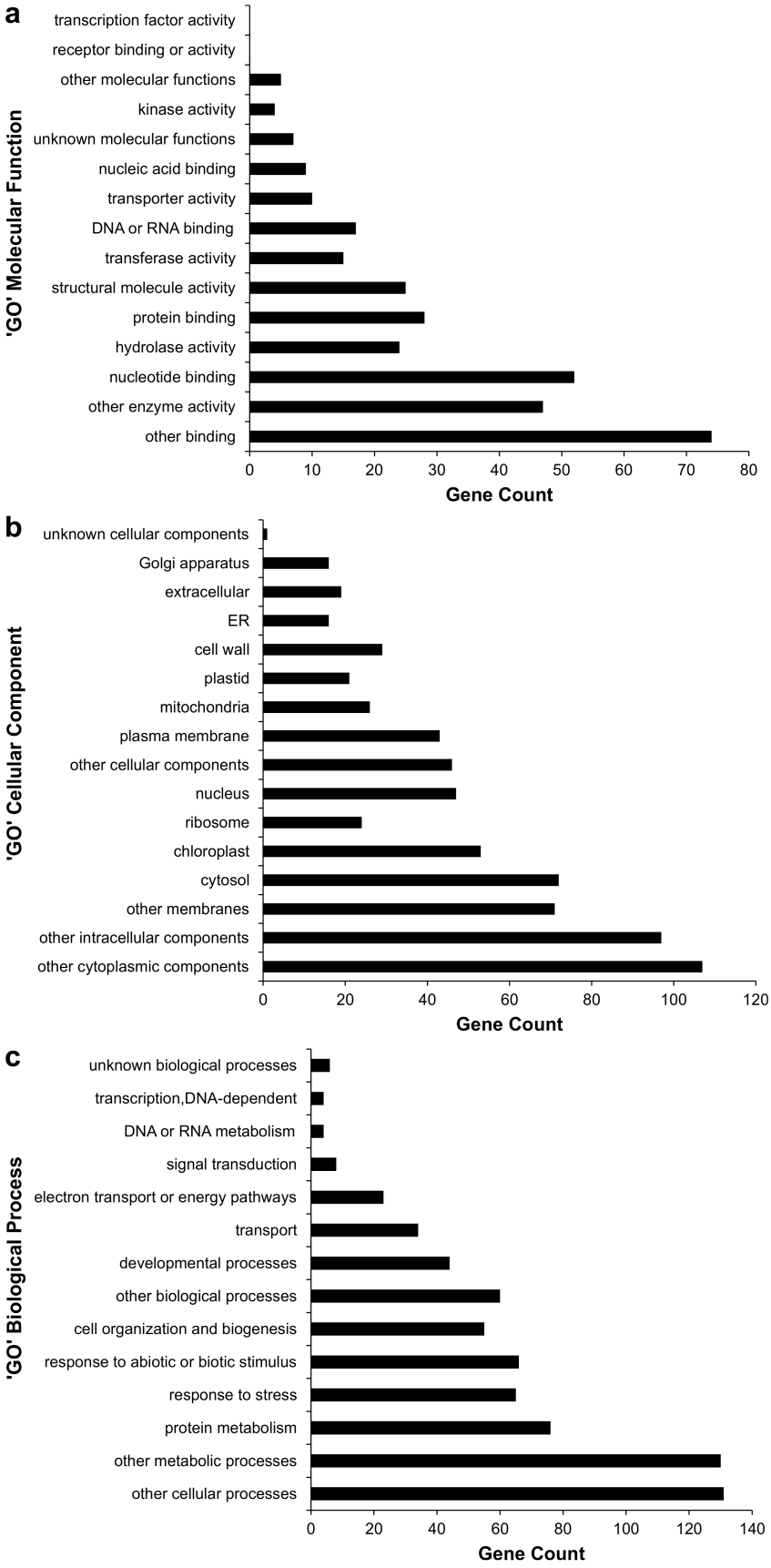

a
'GO' Molecular Function
transcription factor activity
receptor binding or activity
other molecular functions
kinase activity
unknown molecular functions
nucleic acid binding
transporter activity
DNA or RNA binding
transferase activity
structural molecule activity
protein binding
hydrolase activity
nucleotide binding
other enzyme activity
other binding
0 10 20 30 40 50 60 70 80
Gene Count
b
'GO' Cellular Component
unknown cellular components
Golgi apparatus
extracellular
ER
cell wall
plastid
mitochondria
plasma membrane
other cellular components
nucleus
ribosome
chloroplast
cytosol
other membranes
other intracellular components
other cytoplasmic components
0 20 40 60 80 100 120
Gene Count
c
'GO' Biological Process
unknown biological processes
transcription,DNA-dependent
DNA or RNA metabolism
signal transduction
electron transport or energy pathways
transport
developmental processes
other biological processes
cell organization and biogenesis
response to abiotic or biotic stimulus
response to stress
protein metabolism
other metabolic processes
other cellular processes
0 20 40 60 80 100 120 140
Gene Count

Figure 3.5 ‘**GO’ annotations of 330 candidate GmIFS2-interacting proteins** (a) Molecular Function; (b) Cellular Component; (c) Biological function annotation; using the Gene Ontology Database annotations of TAIR identifiers that correspond to the 330 candidate interacting proteins.

Table 3.2 Shikimate and phenylpropanoid pathway enzymes that co-immunoprecipitate with GmIFS2 (in all three replicates)

Gene description	Glyma identifier	No. of unique peptides	Identified peptides	PEAKS score (%)	PTM	Predicted localization
Isoflavone Synthase 2 (GmIFS2)	Glyma.13G173500.1.p	4 (rep1) 6 (rep2) 1 (rep3)	R.FLETGAEGEAGPLDLR.G R.LVDEVDTQNLPYIR.A K.LIM*NDLLNATTVNK.L R.IDDILNK.F R.FQTSAIR.R K.FDPVVER.V K.AREEVYSVVGK.D R.HLPNPPSPK.P	99.0 99.9 99.9 98.2 97.1 99.5 93.2 79.0	Y *oxidation	ER
Arogenate Dehydratase 6 (GmADT6)	Glyma.12G181800.1.p Glyma.13G319000.1.p	2 (rep1) 2 (rep2) 1 (rep3)	R.LVDDANVGTAK.H R.DTAAIASAR.A	99.3 92.8	N	Cytoplasm
Cinnamate 4-Hydroxylase (GmC4H)	Glyma.14G205200.1.p	5 (rep1) 1 (rep2) 2 (rep3)	N.NPDAAVSGTVIR.R K.LGGYDIPAESK.I R.FESEEDPIFQR.L K.NNPDAAVSGTVIR.R K.LPYLQAVVK.E K.DYFVDER.K R.NLVVVSSPELAK.E	99.9 99.9 99.9 99.7 71.9 92.4 96.6	N	ER
Chalcone Isomerase (GmCHI1B1)	Glyma.20G241600.1.p	3 (rep1) 2 (rep2) 2 (rep3)	R.GLNIQEEFVK.F K.QSPTGTLGLSFSK.D K.ELEANPNIEN	99.9 99.7 93.0	N	Cytoplasm

*Indicates predicted oxidation of methionine residue based on LC-MS/MS analysis

3.4 Discussion

3.4.1 Phenylpropanoid metabolism and the 'metabolon' as an explanation

Phenylpropanoid metabolism produces a largely unique array of plant specialized metabolites, including lignins, sinapate esters, stilbenes, aurones, flavonoids and isoflavonoids. In 1974, Helen Stafford (Stafford 1974) proposed a model that would functionally describe the *in planta* delineation of such a ramose biosynthetic pathway, the 'metabolon': an enzyme complex forming at the molecular level to channel metabolites and facilitate production of specialized metabolites. In recent years, there has been gathering information supporting the existence of such a metabolon, in lignin, anthocyanin and flavonoid biosynthesis (Winkel 2004; Ralston and Yu 2006). Information on an isoflavonoid metabolon has been lacking despite the overwhelming logic behind its existence, given the extent of overlap of metabolites and enzymes with the flavonoid pathway.

3.4.2 Localization of IFS to the ER; soluble enzymes in the cytoplasm or nucleus

The basis for an isoflavonoid metabolon would be IFS functioning as an ER anchor. I have demonstrated that GmIFS1 and GmIFS2 are localized to the ER membrane, the latter has previously been shown by (Ralston and Yu 2006), displaying a reticulate pattern of fluorescence in *N. benthamiana* leaf epithelial cells. The soluble enzymes of the pathway CHS, CHR and CHI, are localized to the nucleus and cytoplasm, with the exception of GmCHS7 (in contrast to GmCHS8), which is strictly cytoplasmic. This dual localization to the nucleus as well as the cytoplasm has previously been described in *Arabidopsis* for CHS and CHI (Saslowsky et al. 2005). Furthermore, nuclear localization has also been shown for isoflavonoid glycosyl- and malonyl- transferases (Dhaubhadel et al. 2008). Here, it was determined that the legume-specific enzyme, GmCHR14 (Sepiol and Dhaubhadel, unpublished), and isoforms of GmCHS and GmCHI associated with isoflavonoid production, have shown similar nucleo-cytoplasmic localization.

The mechanism of transport and the function of these metabolic enzymes in the nucleus require further investigation. Nuclear transport in the absence of a known nuclear

localization signal is possible through diffusion for proteins smaller than 60 kDa (Tamura and Hara-Nishimura 2013). This would apply to the fusion proteins of GmCHIs and GmCHR14 with YFP (50-60 kDa). Whether diffusion is the means of transport in the nucleus will still have to be confirmed. The functionality of such proteins within the nucleus could be synthetic or regulatory (Saslowsky et al. 2005).

The fusion of GmCHS8 and YFP, which is approximately 70 kDa, exceeds the size threshold for entry into the nucleus by diffusion, and could require assisted or active transport; such mechanisms would explain the differential localization of the two GmCHS isoforms and the exclusion of GmCHS7 from the nucleus. *Arabidopsis* CHS possesses a four peptide sequence resembling a classic NLS signal on the surface of the tertiary protein conformation (Saslowsky et al. 2005). *In silico* analysis of the deduced amino acid sequence of GmCHS7 and GmCHS8 did not reveal any NLS signals. Nuclear localization of soybean GmCHS enzymes, despite the absence of a conventional NLS sequence, might be mediated by an accessory protein. Transport and function of GmCHS8, GmCHR14 and GmCHIs must be further investigated at a molecular level to determine these possibilities.

3.4.3 Isoflavone synthase interacts with upstream enzymes in the isoflavonoid pathway

The *in planta* protein-protein interaction study demonstrated that the soluble enzymes of phenylpropanoid metabolism, GmCHS7, GmCHS8, GmCHR14 and GmCHI2, regardless of individual localization, interact with GmIFS1/GmIFS2 at the surface of the ER. This is the first evidence for interaction of GmIFS1/GmIFS2 with upstream enzymes. Association between alfalfa IFS and the subsequent IOMT is the only interaction previously shown (Liu and Dixon 2001). The fact that GmIFS1/GmIFS2 interactions were localized to the ER substantiates the model of IFS as a metabolic center (Galili et al. 1998), nucleating (Winkel 2004; Jorgensen et al. 2005) free, soluble enzymes that are inherently either cytoplasmic or nucleo-cytoplasmic.

Protein-protein interaction of GmIFS1/GmIFS2 with non-adjacent enzymes such as GmCHS7 and GmCHS8, suggests a cohesive model of the metabolon. The interactions of

the ER anchor, GmIFS1/GmIFS2, extend upstream through the pathway, whether through direct or indirect association. GmCHS7 and GmCHS8 have previously been shown to play a role in isoflavonoid production in soybean seed and roots (Dhaubhadel et al. 2007). The GmCHS isoform specifically involved in the isoflavonoid metabolon, in a given tissue, would have to be determined by more direct *in planta* techniques such as immunoprecipitation.

GmIFS1/GmIFS2 interaction with GmCHI2 isoforms produced stronger fluorescence intensity than GmCHS7/GmCHS8, which might be indicative of the physical proximity of the former. Such distinctions can be better quantified by Förster resonance energy transfer / fluorescence-lifetime imaging microscopy (FRET / FLIM) analysis. Recently, such technology was applied to demonstrate the interactions of CHS with flavonol synthase and diflavonol reductase in *Arabidopsis* protoplast cells (Crosby et al. 2011).

3.4.4 Specific interaction between GmIFS2 and GmCHR14

The interaction of isoflavonoid pathway enzymes GmIFS2 and GmCHR14 at the surface of the ER, and the specificity of the isoforms involved is a significant finding in supporting the isoflavonoid metabolon. GmIFS2 is inducible (UV or pathogenic stress), with an expression profile largely correlating with isoflavonoid accumulation (Dhaubhadel et al. 2003). Interestingly, GmCHR14 only interacted with GmIFS2 *in planta* (Figure 3.2.3 a-d). CHR coaction with CHS produces the legume-specific isoliquiritigenin, which is converted downstream to daidzein by IFS, and subsequently metabolized into phytoalexins (compounds that inhibit pathogen infection) such as glyceollins (Shimada et al. 2003). Therefore the interaction of GmCHR14 and GmIFS2 could be integral to the ‘inducible’ isoflavonoid metabolon that plays a role in abiotic and biotic stress response. Further proteomic studies under pathogen inoculation are warranted to study induction of the metabolon.

Jung *et al*., (Jung et al. 2000) heterologously expressed CHR and IFS in tobacco plants, which lead to the accumulation of genistein and daidzein. The latter cannot be produced in non-legumes in the absence of CHR, which as previously mentioned, synthesizes the upstream substrate isoliquiritigenin together with CHS.

Interestingly, there was no interaction between GmCHS7/GmCHS8 and GmCHR14. This absence could be due to the transience of such associations or due to the indirect nature of channeling between these, presumably, consecutive enzymes. The complete catalytic process involving CHS and CHR includes three steps: decarboxylation condensation, thioester cyclization, and aromatization (Havsteen 2002; Katsuyama et al. 2007; Zernova et al. 2009). The joint action of CHR and CHS has not been demonstrated experimentally; therefore, the identity of the intermediate substrate of CHR and whether or not it requires active sequestration or channeling is not known (Yu et al. 2000; Yu and McGonigle 2005; Yu et al. 2003; Liu et al. 2007).

3.4.5 Differential GmCHI2 interactions with CHS7 and CHS8

The differentially localized GmCHS isoforms, GmCHS7 (cytoplasmic) and GmCHS8 (nucleo-cytoplasmic), interacted with GmCHI2 in their respective subcellular domains. The association with GmCHI2, which itself is localized to the nucleus and cytoplasm, does not change GmCHS localization. Evidence for the presence of interaction, irrespective of the location, between *Arabidopsis* homologs of CHS and CHI, was first shown by Burbulis and Winkel-Shirley (Burbulis and Winkel-Shirley 1999). The requirement for such interaction is not to increase catalytic efficiency; the conversion of chalcones into flavanones approaches the diffusion limit with CHI catalysis (Jez and Noel 2002). However, without metabolite channeling between CHS and CHI, chalcones released into solution will undergo non-enzymatic cyclization into flavanones, yielding a mix of biologically inactive and active isomers (Jez and Noel 2002). As the catalytic capacity of CHI is particularly strong, channeling might modulate conversion, reserving a pool of chalcones for flux into the many branches of the phenylpropanoid pathway. Additionally, close association between GmCHS, GmCHI and GmIFS, ensures that the moderately lipophilic chalcones and flavanones are not sequestered into cellular membranes (Jez et al. 2002; Jez and Noel 2002).

3.4.6 The wider network of the isoflavonoid metabolon

To further extend the search for isoflavonoid metabolon elements beyond the few adjacent enzymes in the pathway, I followed a Co-IP based approach, resulting in a list of

330 proteins putatively interacting with GmIFS2 in soybean hairy roots. As the metabolic center of the metabolon, nucleating various enzymes, the stress-inducible GmIFS2 was the perfect bait protein for such a study. The list does not include proteins that also interacted with the YFP control, a measure that reduces non-specific interacting proteins. Metabolic enzymes interacting with GmIFS2 include GmC4H, GmCHI1B1, and two isoforms from the GmADT6 family. Evidence for the interaction of GmIFS2 with GmCHIs was shown *in planta* by BiFC assays (Figure 3.2.2). The Co-IP based approach substantiates the interaction of GmIFS2 with GmCHI1B1 in all three replicates, a novel finding in terms of IFS interaction with upstream enzymes. The specific identification of GmCHI1B1, from four catalytically active GmCHIs, is functionally important, denoting the possibility of this isoform to be preferentially involved in soybean root isoflavonoid production.

The association of GmADT6 with GmIFS2, the former being a shikimate pathway enzyme, could also be very significant. The coding sequence of the two *GmADT6* genes identified, Glyma.12G181800 and Glyma.13G319000, were 93.96% identical. They both encode ADT enzymes classified as being 'ADT6-like'. They share a high degree of homology (72.79-73.9%) with the *Arabidopsis* gene (AT1G08250.1), which encodes the plastid-localized AtADT6 (Cho et al. 2007; Corea et al. 2012). Arogenate dehydratase is the last enzyme in the shikimate pathway converting arogenate to phenylalanine or prephenate to phenylpyruvate. Phenylalanine produced by GmADT6 can then be channelled into the phenylpropanoid pathway, which would explain a downstream interaction with GmIFS2, to drive the flux of shared metabolites into the isoflavonoid branch. This interaction extends the reach of a possible isoflavonoid metabolon beyond our expectations for protein complexes producing specialized metabolites. It is possible that certain isoform(s) of GmADT6, two having been identified in the Co-IP approach, are specialized to preferentially participate in the isoflavonoid branch of the phenylpropanoid pathway.

Another novel finding of the Co-IP results was the interaction of GmC4H with GmIFS2. The nature of this interaction, direct or indirect (within a complex), cannot be confirmed within the present study. This is the first discovery of an interaction between two P450

enzymes. It has been conjectured for a long time that phenylpropanoid metabolism involves the channelling of metabolites from one enzyme center to the next (Ralston and Yu 2006). I have shown evidence of interaction between two such centers, both P450 enzymes, anchored in the ER, describing a 'tandem' P450 complex.

Achnine *et al*., showed that overexpression of C4H results in the ER localization of both PAL1 and PAL2 in tobacco, with PAL1 having greater affinity for membrane association. FRET studies, based on dual-labeling immuno-fluorescence, showed that PAL and C4H are located within 100 Å of each other on the membrane (Achnine et al. 2004). Therefore, evidence has been strong for C4H as an ER-localized anchor for a metabolon at the entry-point for the phenylpropanoid pathway. The Co-IP based approach to finding the GmIFS2 interactome showed that an isoflavonoid metabolon is in interaction all the way upstream to the GmC4H anchor, and further to GmADT6, possibly directing flux from the entry-point, downstream to isoflavone biosynthesis. This corroborates the speculative model (Figure 1.4), of P450 anchors on the surface of the ER aggregating metabolically relevant enzymes into specialized regio-specific complexes.

One of the guiding definitions of the metabolon as envisioned by Paul Srere (Srere 1985) was the structural elements, in the form of cytoskeletal (Chuong et al. 2004) or accessory proteins (Graciet et al. 2004; Winkel 2004) that would form the physical architecture of this molecular level of compartmentalization. To this end, the discovery of at least 25 proteins functionally annotated for structural molecule activity provides a new avenue for metabolon research. In fact, attempts to synthetically regenerate plant specialized metabolites in various hosts have been plagued by the lack of such structural elements (Wilson et al. 2014). Finding specific proteins involved in compartmentalizing and aiding biosynthesis could be important in synthetic biology.

Isoflavone aglycones produced by IFS are very rapidly conjugated and stored in the cellular compartments such as the vacuole and released upon demand for stress response or signalling for symbiosis with nitrogen-fixing rhizobia (Dhaubhadel et al. 2008). Interestingly, there were 15 proteins annotated for transferase activity interacting with GmIFS2. Given the multitude of metabolic pathways converging on the surface of the

ER, a subset of these transferases might be involved in isoflavone conjugation. Subsequent to conjugation, such metabolites must be shuttled into the vacuole. Ten transporter enzymes were also identified in the search, which might be involved in the subsequent transport into cellular compartments for storage, through the secretory pathway (golgi) or into the vacuole. Several of these proteins have ATP-binding activity. In this context, it is also important to note that there are 16 proteins associated with the ER and the golgi, respectively, in the list of GmIFS2-interacting candidates. No proteins annotated for transcription factor or receptor binding activity were identified, which reaffirms the specificity of my isolation technique.

As mentioned previously, GmIFS2 was chosen as the bait protein of choice, due to its induced expression in response to inoculation of soybean by *P. sojae* (Dhaubhadel et al. 2003). Isoflavonoids, the active aglycone, daidzein, and their complex derivatives, glyceollins, are phytoalexins. Silencing of key isoflavonoid pathway genes (e.g. IFS) reduces isoflavonoid content and plant resistance (Graham et al. 2007). Rapid induction of isoflavonoid biosynthesis is one of the major reasons supporting the existence of a metabolon. Proteomic research into the profiles of soybean roots before and after inoculation with a wide range of pathogens has added to our understanding of this plant response. For example, infection by Soybean Cyst Nematode (SCN), leads to a 3-fold increase of isoflavone reductase, which converts vestitone to 2'-hydroxyformononetin, a prominent phytoalexin (Afzal et al. 2009). Infection of soybean root hairs by *Bradyrhizobium japonicum* causes upregulation of GmPAL and GmCHI, both of which are implied in the isoflavonoid metabolon, the former through interaction with the GmC4H anchor and the latter with the GmIFS anchor (Nguyen et al. 2012; Wan et al. 2005). Abiotic elements such as flooding can also increase abundance of phenylpropanoid enzymes, CHI and isoflavonoid malonyl-transferase in the roots (Khatoon et al. 2012; Komatsu et al. 2013). This body of information underlines the importance of investigating the interacting partner of GmIFS2 in soybean roots. Of the proteins enumerated in the Co-IP results, 42% were responsive to stress. These candidates must be closely examined and might shed light on the ‘inducible metabolon’ and how the plant rapidly responds to infection with phytoalexin production.

3.4.7 Importance of findings for metabolic engineering of isoflavonoids

The present findings can shed light on the required metabolic and structural elements for isoflavonoid production in legumes or the introduction into non-legumes. Efforts to engineer isoflavonoid biosynthesis in non-legumes has shown that the pool of flavanone substrates is restricted in accessibility and tightly channelled in endogenous systems toward flavonoid production; introduction of IFS and induction by UV or pathogenic stress, while increasing isoflavonoid production, results in a concomitant reduction in flavonoids (Yu and McGonigle 2005; Yu et al. 2003; Liu et al. 2002). Production of isoflavonoids in non-legumes such as *Arabidopsis* can be enhanced up to 30-fold by the introduction of IFS and legume-specific CHIs, in a mutant background that is blocked in flavonol and anthocyanin production (Liu et al. 2002). The evidence suggests that the bottleneck for heterologous isoflavonoid production is competition for flavanone substrates with overlapping pathways.

The discussion of bottlenecks and competition between concurrent branches of the phenylpropanoid pathway underlines the importance of molecular and proteomic determination of isoform specificity within metabolons. In legumes, with the co-existing and overlapping flavonoid and isoflavonoid pathways, the multi-gene families that encode shared enzymes could be a mechanism for specialization. ‘Type I’ CHIs, such as soybean GmCHI2 might be preferentially involved in flavonoid metabolism, while the legume-specific ‘type II’ CHIs that have the additional capacity to convert isoliquiritigenin chalcone, might be involved in isoflavonoid metabolism (Ralston et al. 2005). As there are three type II CHIs in soybean, the specific interaction of GmCHI1B1 with GmIFS2 in the Co-IP results suggests that this isoform is the chalcone isomerase member of the isoflavonoid metabolon in soybean roots.

3.5 Conclusion

The major outcome of the present research is the identification of the elements and the circumstances of the isoflavonoid metabolon, which have not been substantiated until now. This includes the interaction of GmIFS with GmCHS, GmCHR and GmCHI at the surface of the ER; GmIFS2 interaction with GmADT6, GmC4H and GmCH1B1 (Co-IP). This is the first evidence to corroborate the isoflavonoid metabolon, and the interaction of IFS with upstream enzymes reaching up to the shikimate pathway, suggesting a regulation of phenylalanine flux into isoflavonoid biosynthesis.

3.6 Literature Cited

Achnine L, Blancaflor EB, Rasmussen S, Dixon RA (2004) Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. Plant Cell 16 (11):3098-3109

Afzal AJ, Natarajan A, Saini N, Iqbal MJ, Geisler M, El Shemy HA, Mungur R, Willmitzer L, Lightfoot DA (2009) The nematode resistance allele at the rhg1 locus alters the proteome and primary metabolism of soybean roots. Plant Physiol 151 (3):1264-1280

Burbulis IE, Winkel-Shirley B (1999) Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway. Proc Natl Acad Sci U S A 96 (22):12929-12934

Cho MH, Corea OR, Yang H, Bedgar DL, Laskar DD, Anterola AM, Moog-Anterola FA, Hood RL, Kohalmi SE, Bernards MA, Kang C, Davin LB, Lewis NG (2007) Phenylalanine biosynthesis in Arabidopsis thaliana. Identification and characterization of arogenate dehydratases. J Biol Chem 282 (42):30827-30835

Chuong SD, Good AG, Taylor GJ, Freeman MC, Moorhead GB, Muench DG (2004) Large-scale identification of tubulin-binding proteins provides insight on subcellular trafficking, metabolic channeling, and signaling in plant cells. Mol Cell Proteomics 3 (10):970-983

Corea OR, Bedgar DL, Davin LB, Lewis NG (2012) The arogenate dehydratase gene family: towards understanding differential regulation of carbon flux through phenylalanine into primary versus secondary metabolic pathways. Phytochemistry 82:22-37

Crosby KC, Pietraszewska-Bogiel A, Gadella TW, Jr., Winkel BS (2011) Forster resonance energy transfer demonstrates a flavonoid metabolon in living plant cells that displays competitive interactions between enzymes. FEBS Lett 585 (14):2193-2198

Dhaubhadel S, Farhangkhoee M, Chapman R (2008) Identification and characterisation of isoflavonoid specific glycosyltransferase and malonyltransferase from soybean seeds. J Exp Bot 59 (4):981-994

Dhaubhadel S, Gijzen M, Moy P, Farhangkhoee M (2007) Transcriptome analysis reveals a critical role of *CHS7* and *CHS8* genes for isoflavonoid synthesis in soybean seeds. Plant Physiol 143 (1):326-338

Dhaubhadel S, McGarvey BD, Williams R, Gijzen M (2003) Isoflavonoid biosynthesis and accumulation in developing soybean seeds. Plant Mol Biol 53 (6):733-743

Galili G, Sengupta-Gopalan C, Ceriotti A (1998) The endoplasmic reticulum of plant cells and its role in protein maturation and biogenesis of oil bodies. Plant Mol Biol 38 (1-2):1-29

Graciet E, Lebreton S, Gontero B (2004) Emergence of new regulatory mechanisms in the Benson-Calvin pathway via protein-protein interactions: a glyceraldehyde-3-phosphate dehydrogenase/CP12/phosphoribulokinase complex. J Exp Bot 55 (400):1245-1254

Graham TL, Graham MY, Subramanian S, Yu O (2007) RNAi Silencing of Genes for Elicitation or Biosynthesis of 5-Deoxyisoflavonoids Suppresses Race-Specific

Resistance and Hypersensitive Cell Death in Phytophthora sojae Infected Tissues. Plant Physiol 144:728-740

Havsteen BH (2002) The biochemistry and medical significance of the flavonoids. Pharmacol Ther 96 (2-3):67-202

Hrazdina G, Wagner GJ (1985) Metabolic pathways as enzyme complexes: evidence for the synthesis of phenylpropanoids and flavonoids on membrane associated enzyme complexes. Arch Biochem Biophys 237 (1):88-100

Jez JM, Bowman ME, Noel JP (2002) Role of hydrogen bonds in the reaction mechanism of chalcone isomerase. Biochemistry 41 (16):5168-5176

Jez JM, Noel JP (2002) Reaction mechanism of chalcone isomerase. J Biol Chem 277 (2):1361-1369

Jorgensen K, Rasmussen AV, Morant M, Nielsen AH, Bjarnholt N, Zagrobelny M, Bak S, Moller BL (2005) Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. Curr Opin Plant Biol 8 (3):280-291

Jung W, Yu O, Lau SMC, O'Keefe DP, Odell J, Fader G, McGonigle B (2000) Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes Nat Biotechnol 18 (5):559-559

Katsuyama Y, Funa N, Miyahisa I, Horinouchi S (2007) Synthesis of unnatural flavonoids and stilbenes by exploiting the plant biosynthetic pathway in Escherichia coli. Chem Biol 14 (6):613-621

Khatoon A, Rehman S, Salavati A, Komatsu S (2012) A comparative proteomics analysis in roots of soybean to compatible symbiotic bacteria under flooding stress. Amino acids 43 (6):2513-2525

Komatsu S, Makino T, Yasue H (2013) Proteomic and biochemical analyses of the cotyledon and root of flooding-stressed soybean plants. PLoS One 8 (6):e65301

Liu CJ, Blount JW, Steele CL, Dixon RA (2002) Bottlenecks for metabolic engineering of isoflavone glycoconjugates in Arabidopsis. Proc Natl Acad Sci USA 99 (22):14578-14583

Liu CJ, Dixon RA (2001) Elicitor-induced association of isoflavone O-methyltransferase with endomembranes prevents the formation and 7-O-methylation of daidzein during isoflavonoid phytoalexin biosynthesis. Plant Cell 13 (12):2643-2658

Liu RR, Hu YL, Li HL, Lin ZP (2007) Production of soybean isoflavone genistein in non-legume plants via genetically modified secondary metabolism pathway. Metab Eng 9 (1):1-7

Lu Q, Tang X, Tian G, Wang F, Liu K, Nguyen N, Kohalmi SE, Keller WA, Tsang EWT, Harada JJ, Rothstein SJ, Cui Y (2010) Arabidopsis homolog of the yeast TREX-2 mRNA export complex: components and anchoring nucleoporin. Plant J 61 (2):259-270

Nguyen TH, Brechenmacher L, Aldrich JT, Clauss TR, Gritsenko MA, Hixson KK, Libault M, Tanaka K, Yang F, Yao Q, Pasa-Tolic L, Xu D, Nguyen HT, Stacey G (2012) Quantitative phosphoproteomic analysis of soybean root hairs inoculated with Bradyrhizobium japonicum. Mol Cell Proteomics 11 (11):1140-1155

Ohad N, Shichrur K, Yalovsky S (2007) The analysis of protein-protein interactions in plants by bimolecular fluorescence complementation. Plant Physiol 145:1090-1099

Ralston L, Subramanian S, Matsuno M, Yu O (2005) Partial reconstruction of flavonoid and isoflavonoid biosynthesis in yeast using soybean type I and type II chalcone isomerases. Plant Physiol 137 (4):1375-1388

Ralston L, Yu O (2006) Metabolons involving plant cytochrome P450s. Phytochem Rev 5 (2):459-472

Saslowsky D, Winkel-Shirley B (2001) Localization of flavonoid enzymes in Arabidopsis roots. Plant J 27 (1):37-48

Saslowsky DE, Warek U, Winkel BS (2005) Nuclear localization of flavonoid enzymes in Arabidopsis. J Biol Chem 280 (25):23735-23740

Shimada N, Aoki T, Sato S, Nakamura Y, Tabata S, Ayabe S (2003) A cluster of genes encodes the two types of chalcone isomerase involved in the biosynthesis of general flavonoids and legume-specific 5-deoxy(iso)flavonoids in Lotus japonicus. Plant Physiol 131 (3):941-951

Sparkes IA, Runions J, Kearns A, Hawes C (2006) Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat Protoc 1 (4):2019-2025

Srere PA (1985) The metabolon. Trends Biochem Sci 10 (3):109-110

Stafford HA (1974) Activation of 4-hydroxycinnamate hydroxylase in extracts from sorghum. Plant Physiol 54 (5):686-689

Subramanian S, Graham MA, Yu O, Graham TL (2005) RNA interference of soybean isoflavone synthase genes leads to silencing in tissues distal to the transformation site and to enhanced susceptibility to *Phytophthora sojae*. Plant Physiol 137:1345-1353

Subramanian S, Hu X, Lu GH, Odell JT, Yu O (2004) The promoters of two isoflavone synthase genes respond differentially to nodulation and defense signals in transgenic soybean roots. Plant Mol Biol 54 (5):623-639

Tamura K, Hara-Nishimura I (2013) The molecular architecture of the plant nuclear pore complex. J Exp Bot 64 (4):823-832

Wagner GJ, Hrazdina G (1984) Endoplasmic reticulum as a site of phenylpropanoid and flavonoid metabolism in hippeastrum. Plant Physiol 74 (4):901-906

Wan J, Torres M, Ganapathy A, Thelen J, DaGue BB, Mooney B, Xu D, Stacey G (2005) Proteomic analysis of soybean root hairs after infection by Bradyrhizobium japonicum. Mol Plant Microbe Interact 18 (5):458-467

Wilson SA, Cummings EM, Roberts SC (2014) Multi-scale engineering of plant cell cultures for promotion of specialized metabolism. Curr Opin Biotechnol 29C:163-170

Winkel BS (2004) Metabolic channeling in plants. Annu Rev Plant Biol 55:85-107

Winkel-Shirley B (2002) Biosynthesis of flavonoids and effects of stress. Curr Opin Plant Biol 5 (3):218-223

Yi J, Derynck MR, Chen L, Dhaubhadel S (2010) Differential expression of *CHS7* and *CHS8* genes in soybean. Planta 231:741-753

Yu O, Jung W, Shi J, Croes RA, Fader GM, McGonigle B, Odell JT (2000) Production of the Isoflavones Genistein and Daidzein in Non-Legume Dicot and Monocot Tissues. Plant Physiol 124 (2):781-794

Yu O, McGonigle B (2005) Metabolic engineering of isoflavone biosynthesis. Adv Agron 86:147 - 190

Yu O, Shi J, Hession AO, Maxwell AA, McGonigle B, Odell JT (2003) Metabolic engineering to increase isoflavone biosynthesis in soybean seeds. Phytochemistry 63:753-763

Zernova OV, Lygin AV, Widholm JM, Lozovaya VV (2009) Modification of isoflavones in soybean seeds via expression of multiple phenolic biosynthetic genes. Plant Physiol Biochem 47 (9):769-777

CHAPTER FOUR – TRANSCRIPTOMIC INSIGHTS INTO ISOFLAVONOID BIOSYNTHESIS IN SOYBEAN ROOTS

4.1 Introduction

Soybean is a paleopolyploid with two duplication events, with estimated times of approximately 14 and 44 million years ago (Schmutz et al. 2010). The majority of phenylpropanoid and specialized metabolism enzymes that have been characterized belong to multi-gene families (Blanc and Wolfe 2004; Shoemaker et al. 2006; Livingstone et al. 2010). The large number of new duplications and iterations of core metabolic genes have been conjectured to be partially responsible for the array of plant specialized metabolites (Dixon et al. 2002).

Duplication of genes can lead to neo-functionalization, sub-functionalization or the loss of genetic function. Gene duplication creates genetic redundancy, relieving selection pressure, allowing secondary copies to accumulate mutations that can lead to functional novelty over generations. A second possible outcome of gene duplication, sub-functionalization, is the divergence of both paralogs leading to distribution of the original functionality between the pair. A third and more common outcome following gene duplication is successive mutations in one or more copy, leading to loss of function.

These concepts are applied to multi-gene families encoding enzymatic proteins or proteins involved in plant metabolism. As has been shown with CHS in the literature (Dhaubhadel et al. 2007; Yi et al. 2010) and in chapter 2 with CHI, multi-gene families that encode specialized metabolism enzymes often display a range of catalytic ability and function. *GmCHS7* and *GmCHS8* were shown to have a role in isoflavonoid biosynthesis and their differential localization was shown in chapter 3 (section 3.3.1). The evolution of the CHI fold from protist and fungal homologs, to fatty acid binding proteins in photosynthetic organisms, to chalcone-to-flavanone catalytic ability, is another example of this evolution. Neo-functionalization could explain the legume-specific CHI ability, or enhanced ability, to catalyze isoliquiritigenin.

The release of a high-quality draft of the soybean genome, the newest assembly of which (v2.0) includes a predicted 56,044 protein-coding loci and 88,647 transcripts (Schmutz et al. 2010), and the advent of next-generation sequencing technology provides us with the tools to investigate isoflavonoid biosynthesis from a transcriptomic perspective. For this purpose the root was chosen as the organ of study; due to the dual importance of isoflavonoids in the root as a) signalling molecules for *nod* gene induction in nitrogen-fixing symbiotic bacteria (Subramanian et al. 2005), and b) as phytoalexins and anti-microbial agents warding off pathogen infection (Phillips and Kapulnik 1995; Subramanian et al. 2005).

In the attempt to gain insight into gene function within large gene families, and to extend the knowledge of isoflavonoid-related genes beyond a few well-studied pathway enzymes a transcriptomic study was undertaken. Four cultivars were chosen for the study: two with high root isoflavonoid content, Conrad and AC Colombe; two with low isoflavonoid content, AC Glengarry and Pagoda (Figure 4.1, Dhaubhadel, unpublished). The high isoflavonoid cultivars have been characterized as having increased resistance to *Phytophthora sojae* (Poysa, personal communication). Mulitple sources have cited increased resistance in Conrad soybeans to a broad range of pathogen infection (*Fusarium graminearum* and *P. sojae*) (Wang et al. 2012; Li et al. 2010). Therefore, the exploration of transcriptomic mechanisms underlying isoflavonoid content can have an important agronomic impact in improving pathogen resistance in soybean cultivars.

Next-generation Illumina sequencing technology has relatively little variation between technical replicates and is ideal for identifying differentially expressed genes (Marioni et al. 2008). In this chapter next-generation RNA sequencing was used to find candidate transcripts, differentially expressed in high root isoflavonoid cultivars, to dissect this trait further. This requires four pair-wise comparisons between high and low isoflavonoid cultivars to find differentially expressed (DE) genes, and an overlap analysis of the resulting lists of transcripts to find candidates that are consistently up or downregulated in the high isoflavonoid cultivars. Close attention was paid to the expression of phenylpropanoid pathway genes that might have an impact on isoflavonoid content in soybean roots.

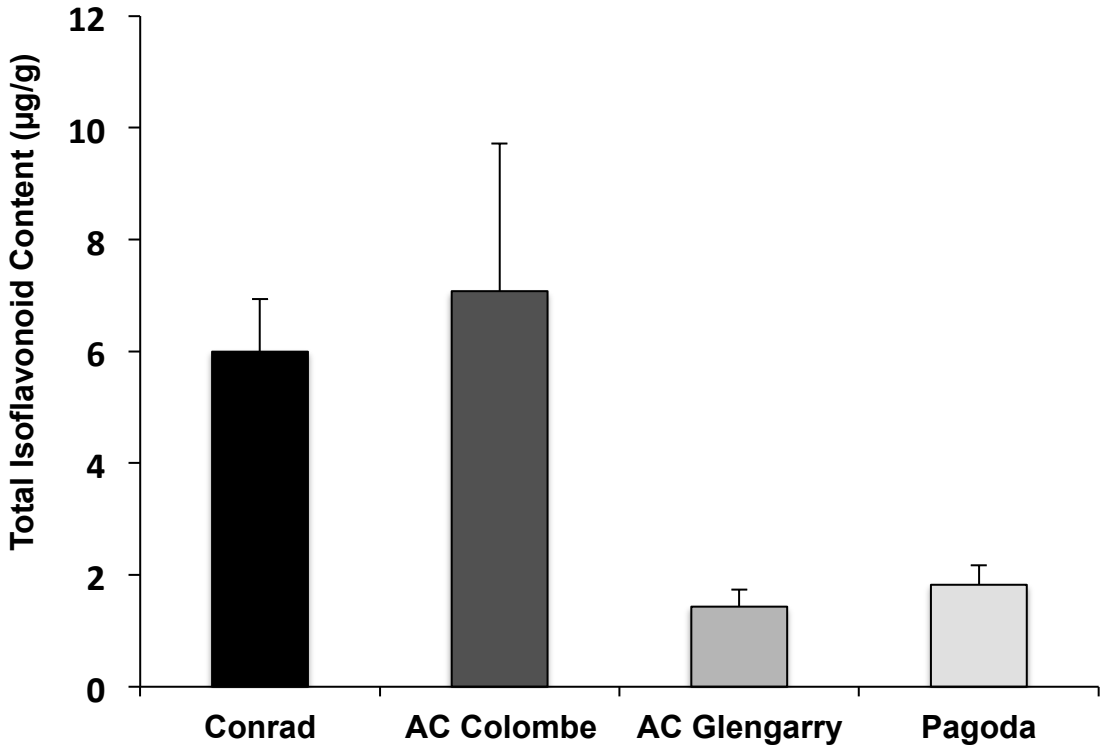


Figure 4.1 Isoflavonoid content of the roots of four soybean cultivars. Total isoflavonoid content (daidzein, genistein, glycitein, daidzin, genistin and glycitin), as measured by HPLC, of the roots of soybean cultivars: Conrad, AC Colombe, AC Glengarry and Pagoda (Dhaubhadel, unpublished).

4.2 Materials and Methods

4.2.1 Plant Material and RNA Extraction

Soybean seeds from the cultivars, Conrad, AC Colombe, AC Glengarry and Pagoda, were planted in Pro-Mix BX Mycorrhizae^{TM} soil (Rivière-du-Loup, Canada) in a growth chamber with a 16 h light cycle at 25°C, and an 8 h dark cycle at 20°C, with 60-70% relative humidity, with light intensity 100-150 μmol $m^{-2}s^{-1}$.

Two week-old soybean roots were harvested, flash-frozen with liquid nitrogen and ground to a fine powder. Three biological replicates were prepared for each of the four cultivars. RNA was extracted from the twelve samples according to the instructions in the RNeasy Mini Kit (Qiagen, Germany). RNA integrity and concentration were recorded using the Agilent 2100 (Bioanalyzer, USA). Integrity for samples ranged from 8.80-10.0, with 8 being the minimum required for high-quality RNA sequencing. A total of 4 μg of RNA, at a concentration of approximately 100 ng/μL was submitted per replicate.

4.2.2 High-throughput RNA sequencing and alignment

The root tissue mRNA from the four cultivars was sequenced with an Illumina Hiseq2000 at the DNA Technologies Unit of the Plant Biotechnology Institute (Saskatoon, SK, Canada). Three replicates from each of the cultivars were sequenced using 100bp paired-end reads, resulting in twelve RNA libraries. Adaptor sequences were removed using a custom Perl script. Reads from each library were mapped against the *Glycine max* transcriptome v2.0, Wm82.a2.v1 (Schmutz et al. 2010).

4.2.3 Quality assurance of sequencing data

The total reads from each RNA sequencing library (biological replicate from each soybean cultivar) were subjected to 3′ end trimming, with a base quality > 30 (Li and Durbin 2009). Alignment was performed using the paired-end module of Burrows-Wheeler Aligner (BWA), a program that aligns relatively short nucleotide sequences against a long reference sequence such as the soybean genome. Contaminating features such as PCR duplicates and reads with low mapping quality (mapping quality ≥ 20) were

removed, and the uniquely mapped reads per transcript were determined using Samtools (Li et al. 2009).

4.2.4 Differential gene expression analysis

Alignments of RNA sequencing libraries to the soybean transcriptome were imported into R and assessed for differential gene expression using DESeq (Anders and Huber 2010), and mined for relevant expression information regarding genes of interest. Before calculating differential expression an additional filter was developed to remove the bottom 10% of transcripts based on total read counts.

Transcripts were normalized by counts across biological replicates and cultivars. Dispersion estimates were calculated for each gene using DESeq functions to estimate transcript expression variation across replicates for a given cultivar. Dispersion is the square of the coefficient of biological variation. This gives a visual estimation of the 'tightness' of distribution across the biological replicates (Anders and Huber 2010).

The mean of the counts divided by the size factors corresponds to the mean for each condition (e.g. Mean A and Mean B). Thereby the fold change is the ratio of Mean B over Mean A, and the log2 fold change is used to determine up- or down-regulation. For each gene the DEseq function calculates a chi-square p-value by comparing the mean of condition A against condition B, where the null hypothesis is that Mean A and Mean B are equal.

M-A plots displaying log2 fold change as a function of average expression for each gene model, was used to illustrate the distribution of significant transcripts using those variables between cultivars. This would describe the magnitude of fold change required in terms of expression of transcripts to be considered significant, i.e. lower expressed transcripts usually require a much higher fold change to be considered significant. Similarly, histograms of Benjamini-Hochberg adjusted p-values for all genes against the frequency of gene expression were plotted for pair-wise comparisons between cultivars, e.g. Conrad compared against AC Glengarry replicates. Thusly, the histograms display

significant differential expression between paired cultivars. Genes with at least one read mapping to them were chosen for the p-value histograms.

Genes were also ranked by total read count into percentiles and plotted against the log p-value. This allowed the visual depiction of significant or 'baseline' ($p < 0.001$), highly significant ($p < 1e\text{-}15$), and extremely significant differential expressed genes ($p < 1e\text{-}55$) in the pair-wise comparison between cultivars.

Differential expression analyses between high and low root isoflavonoid cultivars were used to generate lists of genes upregulated and downregulated (p-value < 0.05) in the high isoflavonoid cultivars. Thereby, genes with significant differential expression, in the high isoflavonoid cultivars, are up or downregulated if their fold change is above or below one, respectively. Four analyses were done, comparing: (a) AC Glengarry with AC Colombe (b) AC Glengarry with Conrad (c) Pagoda with AC Colombe (d) Pagoda with Conrad. This resulted in four lists for both up and downregulated genes.

The lists generated were then subjected to an overlap study using the online Venn diagram software, Venny (Oliveros 2007). The overlap of the four pair-wise differential expression studies, represented gene models consistently up or downregulated in high root isoflavonoid cultivars.

4.2.5 Gene model annotation and processing

Genes up or downregulated in all four of the differential expression analyses between high and low isoflavonoid cultivars were annotated using the soybean transcriptome annotation (Soybean genome assembly version 2.0 in the Phytozome database, Wm82.a2.v1). The 'GO' annotations (Gene Ontology Database) of TAIR identifiers that correspond to the upregulated genes were used to generate graphs representing molecular function, cellular component, and biological function. Gene models were inspected manually using the annotations, and compiled into families for consideration based on function, and potential involvement in the phenylpropanoid pathway.

Heatmaps were generated for significantly upregulated genes in pair-wise comparisons between low and high root isoflavonoid cultivars, and for select differentially expressed

genes putatively involved in the isoflavonoid pathway. Heatmaps were generated using the R programming package DEseq, and the heatmap.2 function in the gplots CRAN library. Raw gene expression data were normalized for replicate and library differences in read counts and coverage by variance stabilized transformation, using the DESeq package, prior to heatmap generation.

4.3 Results

4.3.1 Data quality and coverage of soybean transcriptome in four cultivars

The root transcriptomes of four soybean cultivars were studied, encompassing high, Conrad and AC Colombe, and low, AC Glengarry and Pagoda, root isoflavonoid content. Root mRNA from three replicates for each of the four cultivars was sequenced, using 100bp paired-end reads, by Illumina HiSeq (PBI, Canada). The resulting millions of reads from the twelve RNA libraries were first filtered and subsequently mapped against the soybean transcriptome v2.0 (Wm82.a2.v1) (Schmutz et al. 2010). Mapping and quality information including total reads per RNA sequencing library, base quality > 30 (Q30%), mapped reads, uniquely mapped reads and gene coverage (%) are summarized in Table 4.1. Coverage of the 56,044 protein-coding loci ranged from 62.74% to 65.09%. Biological variation for all the gene models in each cultivar was displayed using dispersion graphs (Figure 4.2 a-d). Dispersion plots have, overall ‘tight’ dispersion, indicative of good quality RNA sequencing libraries for the three biological replicates per cultivar.

4.3.2 Differential expression analyses between high and low root isoflavonoid cultivars

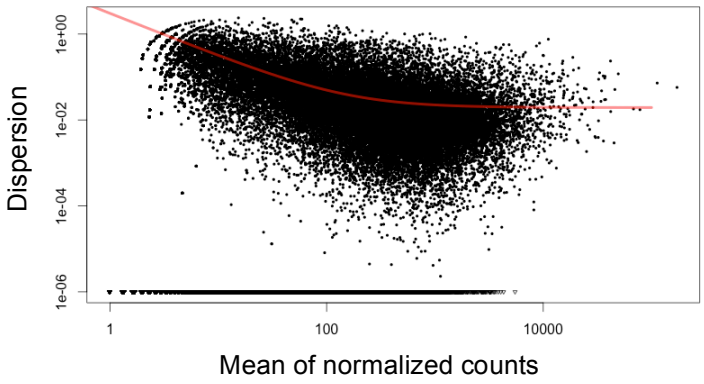
Alignments of RNA sequencing libraries to the soybean transcriptome were imported into R and assessed for differential gene expression using DESeq (Anders and Huber 2010). The appropriate functions to display significance of differentially expressed genes in pair-wise comparisons between cultivars were plotted. The M-A plot displaying log2 fold change as a function of gene expression strength displayed typical form.

Table 4.1 RNA sequencing quality and coverage of soybean transcriptome. Coverage of the soybean transcriptome (Wm82.a2.v1) in soybean roots of Conrad, AC Colombe, AC Glengarry and Pagoda biological replicates ranged from 62.74% to 65.09%.

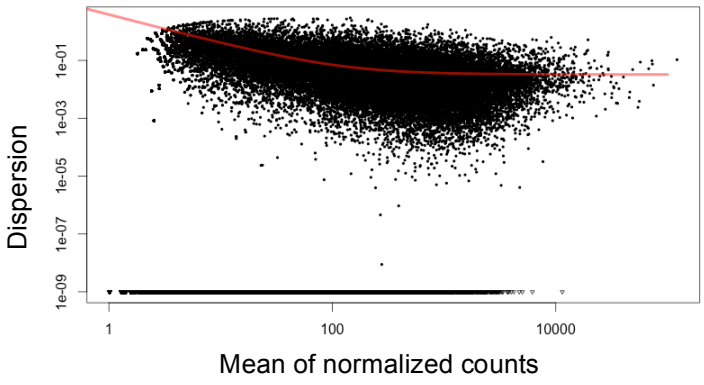
	Replicate	Total Reads	Q30 Bases (%)	Mapped Reads	Uniquely Mapped	Gene Coverage (%)
Conrad	1	43642500	92.91	30762485	24744536	64.28
	2	51423168	92.58	36648290	29163761	64.34
	3	41426398	92.67	29497436	23948533	63.96
AC Colombe	1	64237226	92.88	43062343	33956444	65.09
	2	41552636	92.65	29416157	23831591	63.96
	3	50795644	92.41	33359756	27313027	64.41
AC Glengarry	1	33146096	93.52	23522302	19393241	63.34
	2	49338208	93.54	34540611	27773084	64.27
	3	71279056	93.47	49674464	38962472	64.44
Pagoda	1	19820794	93.53	14005991	11582472	62.74
	2	42622034	93.45	30545146	24479762	63.26
	3	40106364	93.67	27711116	22598173	63.64

Library replicate: the soybean cultivar replicate sequenced; total reads: the total number of sequence reads obtained; Q30 bases (%): the percentage of bases with > 30 quality after 3′ end trimming; mapped reads: the number of reads that mapped to one of the soybean transcript models; uniquely mapped reads: number of mapped reads mapping uniquely to a transcript model with a mapping quality ≥ 20; genes hit: number of gene models with one or more reads mapped to its transcript as a percentage of the total number of gene models (56,044 protein-coding loci).

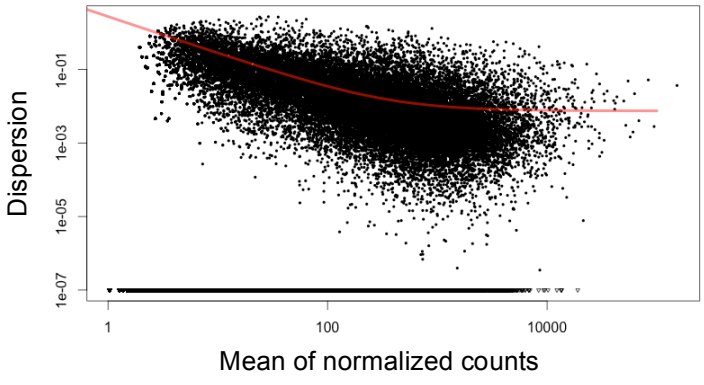
a


Dispersion
1e+00
1e-02
1e-04
1e-06
1
100
10000
Mean of normalized counts

b


Dispersion
1e-01
1e-03
1e-05
1e-07
1e-09
1
100
10000
Mean of normalized counts

c


Dispersion
1e-01
1e-03
1e-05
1e-07
1
100
10000
Mean of normalized counts

d

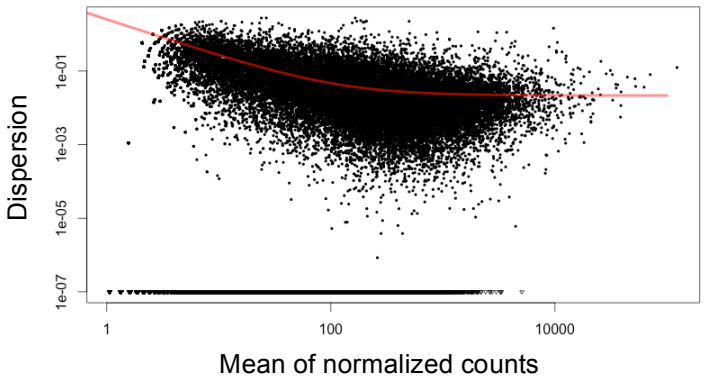

Dispersion
1e-01
1e-03
1e-05
1e-07
1
100
10000
Mean of normalized counts

Figure 4.2 Estimated dispersion for all genes for each cultivar. Dispersion values plotted as a function of average normalized expression for reach gene model. Biological variation displayed for (a) Conrad (b) AC Colombe (c) AC Glengarry and (d) Pagoda biological replicates.

The magnitude of log2 fold change decreased on average with increased normalized expression, while significant differentially expressed genes were interspersed throughout (shown in red) (Figure 4.3 a-d). The adjusted p-value distribution for all genes assessed for differential expression between high and low cultivars displayed a high frequency of highly significant genes (Figure 4.4 a-d). The enrichment of low p values represents differentially expressed genes. Genes that are not differentially expressed are spread uniformly toward a p-value of one. Enrichment of p values at and around one represents genes that have low read counts. Finally, Figure 4.5 (a-d) shows the p-value distribution in terms of baseline ($p < 0.001$), high ($p < 1e\text{-}15$), and extreme ($p < 1e\text{-}55$) significance levels. An abundance of gene models being significantly, and highly significantly, differentially expressed in each of the comparisons is representative of the genetic differences between the cultivars.

Four lists were generated from pair-wise differential expression analyses of low against high root isoflavonoid cultivars: AC Glengarry with AC Colombe; AC Glengarry with Conrad; Pagoda with AC Colombe; Pagoda with Conrad. These four lists were filtered to include genes significantly differentially expressed (p-value < 0.05). Genes up or downregulated in high isoflavonoid cultivars had a fold change above or below one, respectively.

The lists generated were assessed to find gene models that were consistently up or downregulated in all four differential expression analyses. The overlap studies are illustrated in two, four-way Venn diagram for up and downregulation (Figure 4.6). The numbers in the ovals represent the number of gene models in the four lists, and their corresponding overlap is also enumerated. The core overlap of 153 and 113 is the number of genes consistently up and downregulated, respectively. These two lists of genes can be studied for their correlation with isoflavonoid biosynthesis and a putative role in accumulation. The list of 153 upregulated and 113 downregulated genes will hereafter be referred to as ‘high isoflavonoid’ and ‘low isoflavonoid’ genes. This does not denote a role in the pathway, but an association with the isoflavonoid content of the cultivars in which they are up or downregulated.

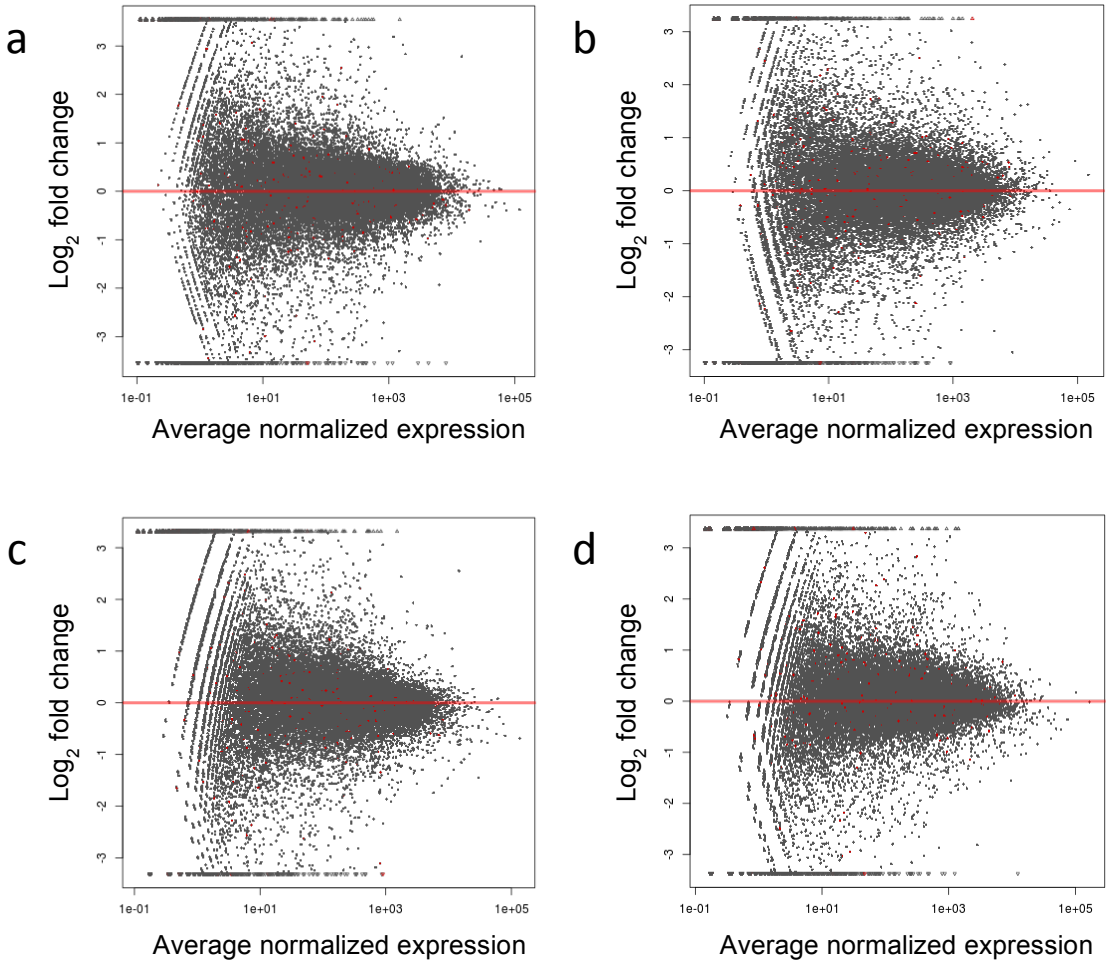

a
b
c
d
Log2 fold change
Average normalized expression
1e-01
1e+01
1e+03
1e+05
3
2
1
0
-1
-2
-3

Figure 4.3 M-A plot of log2 fold change as a function of average expression of gene. Significantly differentially expressed genes (FDR ≤ 0.001) are represented in red. Plots displayed for pair-wise comparisons (a) AC Glengarry with AC Colombe (b) AC Glengarry with Conrad (c) Pagoda with AC Colombe (d) Pagoda with Conrad.

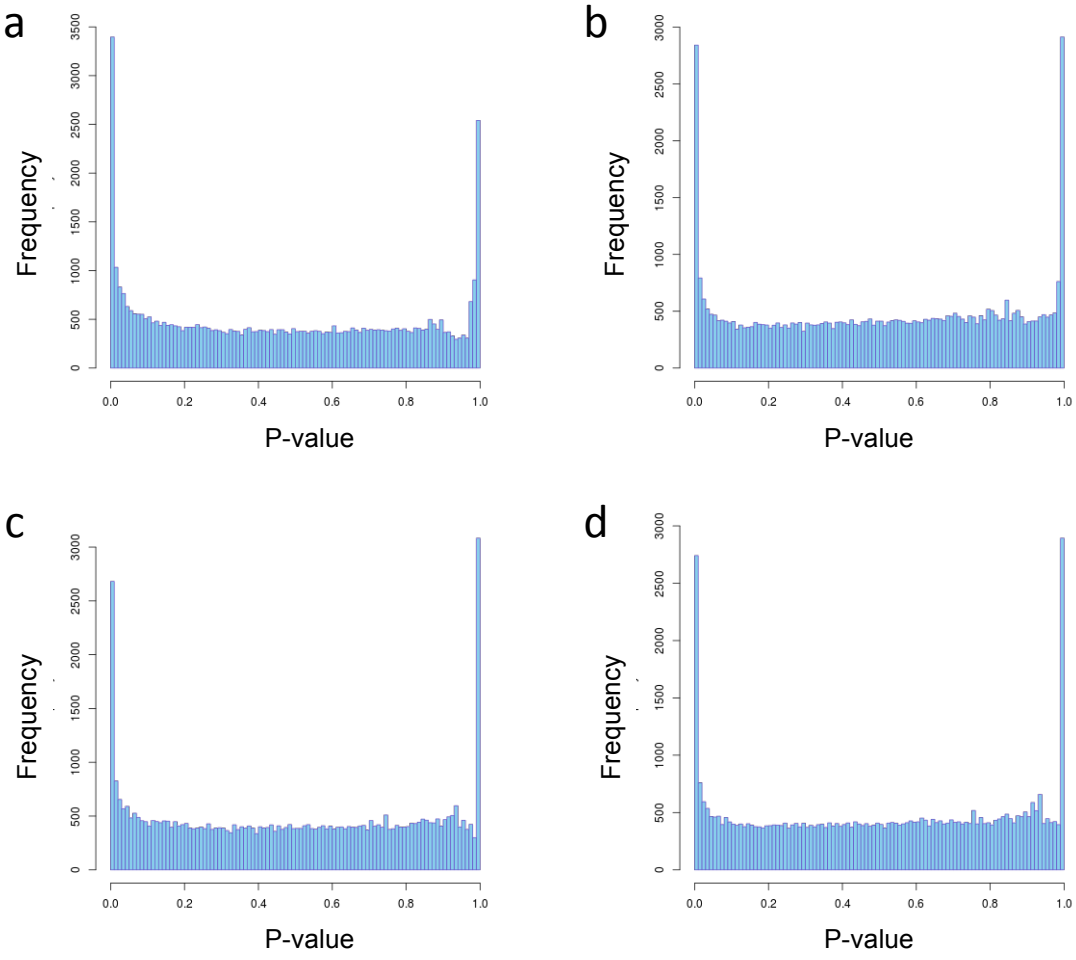

a
Frequency
0
500
1000
1500
2000
2500
3000
3500
0.0
0.2
0.4
0.6
0.8
1.0
P-value
b
Frequency
0
500
1000
1500
2000
2500
3000
0.0
0.2
0.4
0.6
0.8
1.0
P-value
c
Frequency
0
500
1000
1500
2000
2500
3000
0.0
0.2
0.4
0.6
0.8
1.0
P-value
d
Frequency
0
500
1000
1500
2000
2500
3000
0.0
0.2
0.4
0.6
0.8
1.0
P-value

Figure 4.4 Gene frequency as a function of Benjamini-Hochberg adjusted p-values. Histogram of p-values for all genes assessed for significant differential expression between two cultivars (a) AC Glengarry with AC Colombe (b) AC Glengarry with Conrad (c) Pagoda with AC Colombe (d) Pagoda with Conrad.

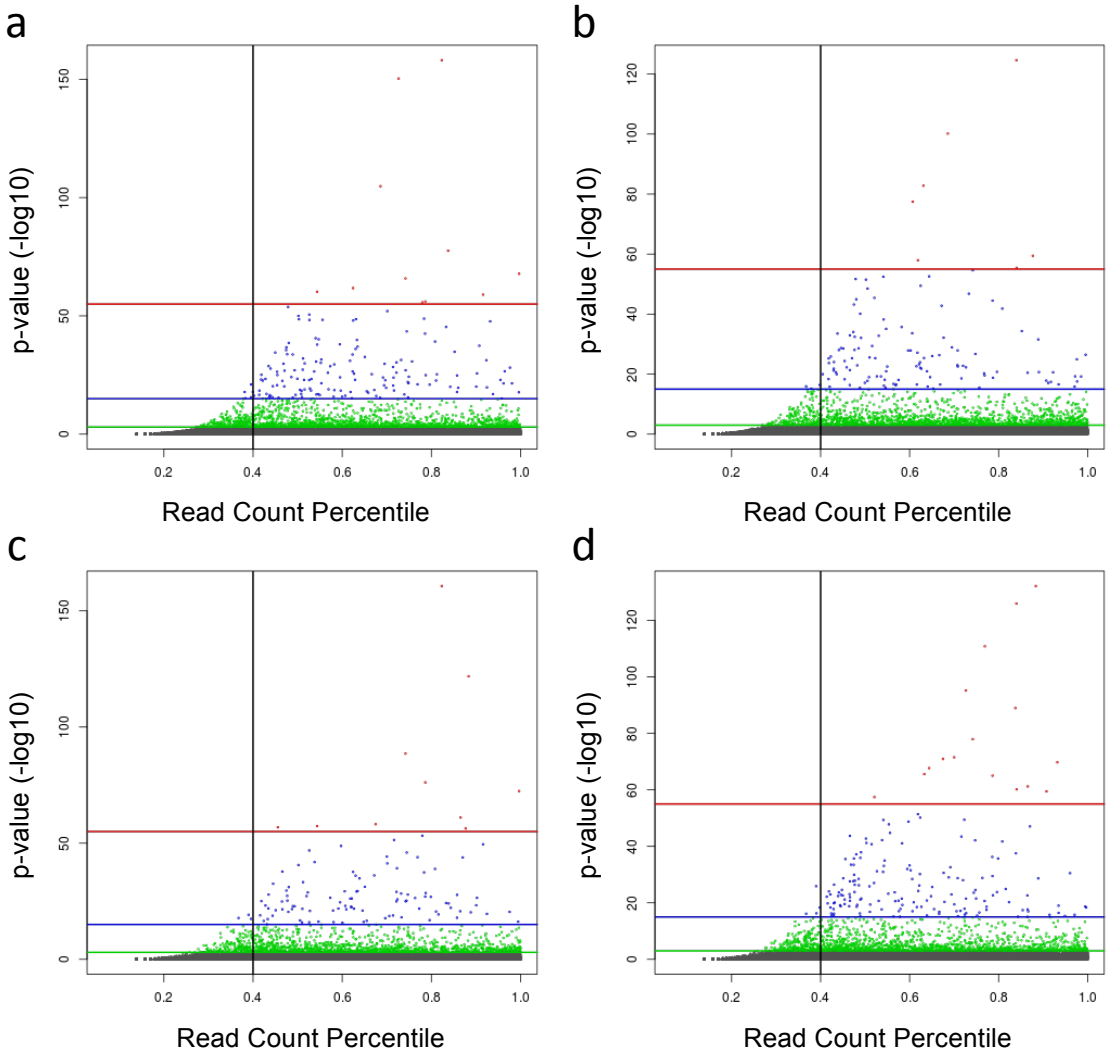
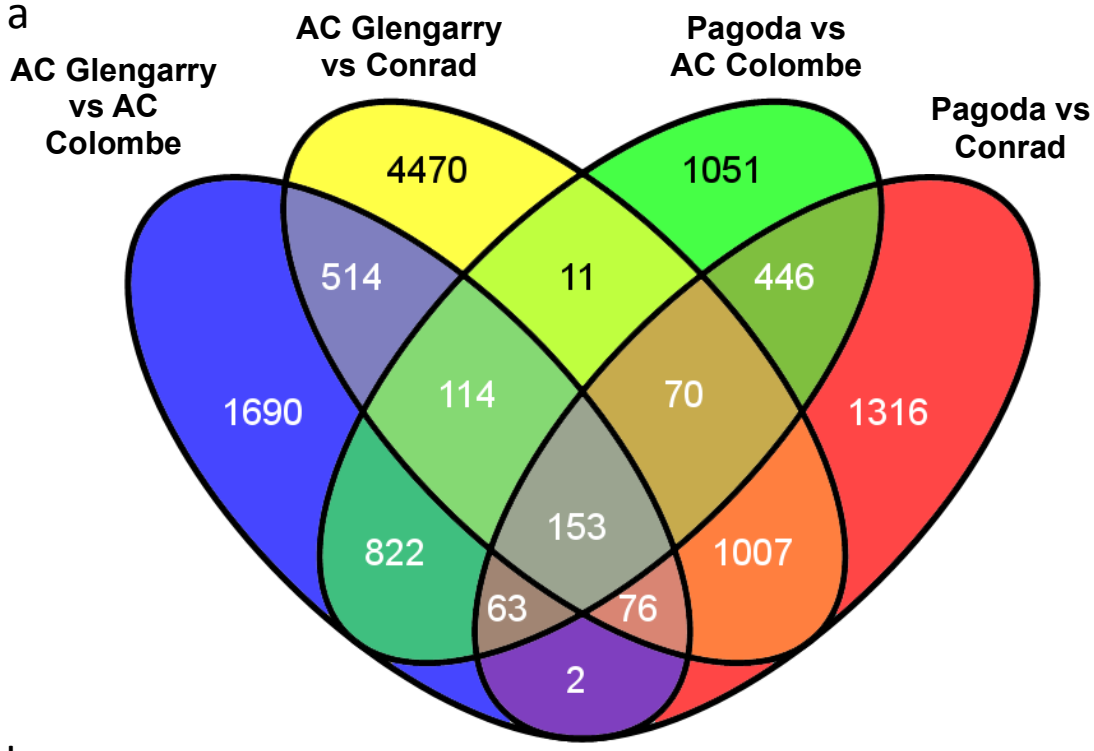


Figure 4.5 Genes ranked by total read count against –log10 of the p-value. Significance scores divided into three levels: baseline ($p < 0.001$), high ($p < 1e\text{-}15$), and extreme ($p < 1e\text{-}55$) significance. Average gene counts divided into percentiles. Four plots representing differential expression between cultivars (a) AC Glengarry with AC Colombe (b) AC Glengarry with Conrad (c) Pagoda with AC Colombe (d) Pagoda with Conrad.

a


AC Glengarry vs AC Colombe
AC Glengarry vs Conrad
Pagoda vs AC Colombe
Pagoda vs Conrad
4470
1051
514
11
446
1690
114
70
1316
153
822
1007
63
76
2

b

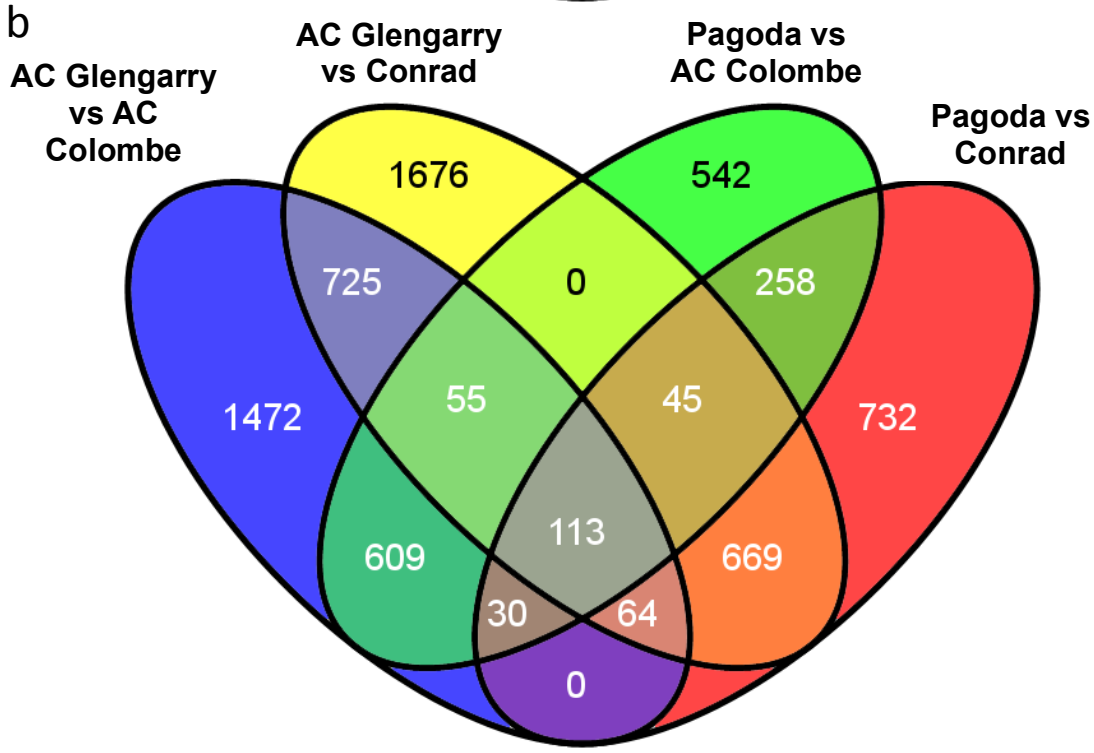

AC Glengarry vs AC Colombe
AC Glengarry vs Conrad
Pagoda vs AC Colombe
Pagoda vs Conrad
1676
542
725
0
258
1472
55
45
732
113
609
669
30
64
0

Figure 4.6 Overlap study of upregulated and downregulated genes in high root isoflavonoid cultivars. Venn diagrams depicting the number of (a) upregulated and (b) downregulated genes and their overlap in four pair-wise differential expression studies: (list 1) AC Glengarry with AC Colombe (list 2) AC Glengarry with Conrad (list 3) Pagoda with AC Colombe (list 4) Pagoda with Conrad. The core of 153 and 113 represent the number of genes consistently up- and down-regulated in all four comparisons, respectively.

Genes that fall outside of this core overlap would include genes up or downregulated in one or more of the comparisons but not all four. These genes might represent cultivar-specific differences in the root transcriptome that are not consistently differentially expressed, within the parameters of the study.

4.3.3 Functional and structural annotation of differentially expressed genes

The set of 153 genes upregulated in high isoflavonoid cultivars were then compiled into a list based on *Arabidopsis* homolog TAIR identifiers (The Arabidopsis Information Resource, http://arabidopsis.org) and subjected to GO (Gene Ontology Database, http://geneontology.org) functional annotation. The resulting graphs (Figure 4.7 a-c) represent the molecular and cellular function, and cellular compartmentalization of this protein list.

The majority (40) of 'high isoflavonoid' genes were predicted to localize to the nucleus. Thirteen genes were annotated as transcription factors and five as nucleic acid binding proteins, with the sum of DNA-dependent transcriptional elements coming to 18. Two genes were annotated for a role in the regulation of chromatin structure: a SET7/9 family, histone H3K4-specific methyltransferase protein and a histone deacetylase 8 (*GmHDA8*) family protein. Only 5 'high isoflavonoid' genes were associated with the ER and golgi, as part of the secretory pathway.

Among the 'high isoflavonoid' genes, 30 were annotated as being responsive to stress. These genes included members of the dirigent family of proteins. Several nucleotide binding disease resistance proteins were identified, containing C-terminal leucine-rich repeat (LRR) domains fused to central nucleotide-binding (NB) domain (NB-LRR proteins). Fourteen genes upregulated in high isoflavonoid cultivars were annotated for transferase activity. Furthermore, 8 'high isoflavonoid' genes with putative transporter activity were identified, including a member of the multi-drug and toxin extrusion transporters (MATE) family.

The lists of differentially expressed genes were manually inspected for genes putatively involved in the isoflavonoid pathway, based on functional annotation from the soybean

database. Flavonoid 6-hydroxylase (*GmF6H*) was upregulated, while isoflavone 2'-hydroxylase (*GmI2'H*), flavonoid 3'-hydroxylase (*GmF3'H*), flavonoid 3', 5'-hydroxylase (*GmF3'5'H*), and dihydroflavonol 4-reductase (*GmDFR*) were downregulated in high root isoflavonoid cultivars.

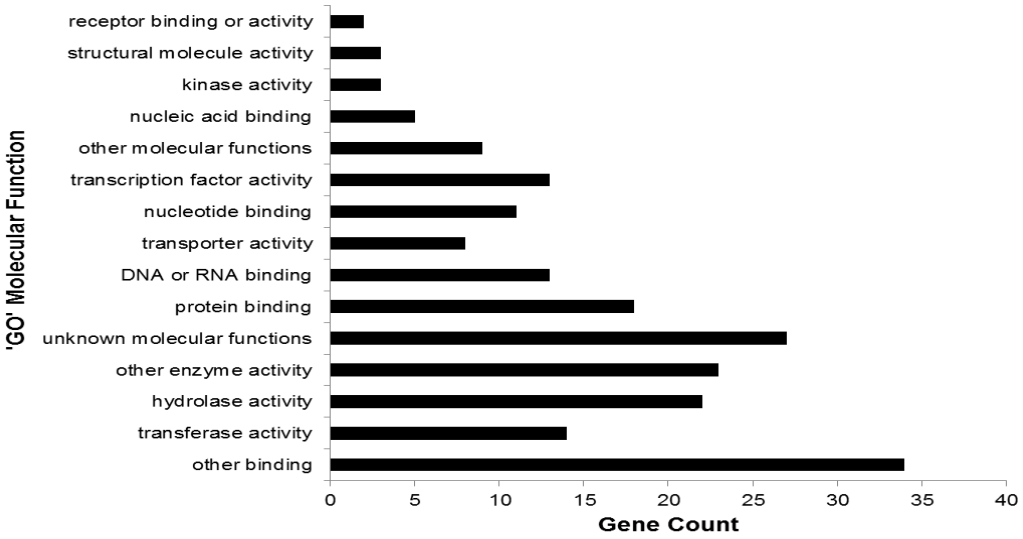
Heatmaps were generated using variance-stabilized transformation of raw read counts for the set of 153 'high isoflavonoid' genes, in pair-wise comparisons between low and high root isoflavonoid cultivars (Figure 4.8 a-d). The heatmaps display the higher expression (p-value < 0.05) of the 'high isoflavonoid' genes in the cultivars, AC Colombe and Conrad, as compared with AC Glengarry and Pagoda. Heatmaps were also generated to show the differential expression of the phenylpropanoid genes mentioned above, in the high and low isoflavonoid cultivars.

4.4 Discussion

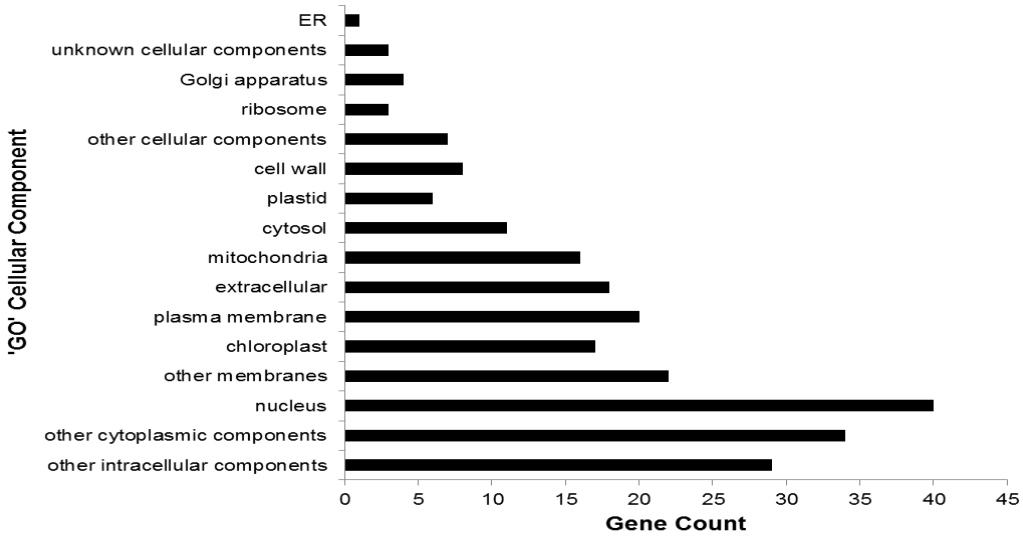
4.4.1 Transcriptomic and metabolomic variation across soybean cultivars

Variation in metabolomes across cultivars of soybean, or any other plant organism, arises from the different genetic, epigenetic and environmental context of its selection and propagation. Soybean metabolomes are particularly diverse and varied: 169 metabolites were found in a study of the seed metabolomes of 29 common soybean cultivars, with 104 of the metabolites showing significant variability (Lin et al. 2014). Therefore, a transcriptomic study of soybean cultivars based on the isoflavonoid composition (six major compounds) will be limited by the scope of metabolites studied, and the multiplicity of mechanisms that might result in high or low isoflavonoid content. Given these limitations, a transcriptomic study can provide interesting genetic candidates that are differentially expressed between the cultivars. Further functional work would be required to substantiate or reject the link between gene expression and isoflavonoid biosynthesis and/or accumulation.

a


receptor binding or activity
structural molecule activity
kinase activity
nucleic acid binding
other molecular functions
transcription factor activity
nucleotide binding
transporter activity
DNA or RNA binding
protein binding
unknown molecular functions
other enzyme activity
hydrolase activity
transferase activity
other binding
'GO' Molecular Function
0 5 10 15 20 25 30 35 40
Gene Count

b


ER
unknown cellular components
Golgi apparatus
ribosome
other cellular components
cell wall
plastid
cytosol
mitochondria
extracellular
plasma membrane
chloroplast
other membranes
nucleus
other cytoplasmic components
other intracellular components
'GO' Cellular Component
0 5 10 15 20 25 30 35 40 45
Gene Count

c

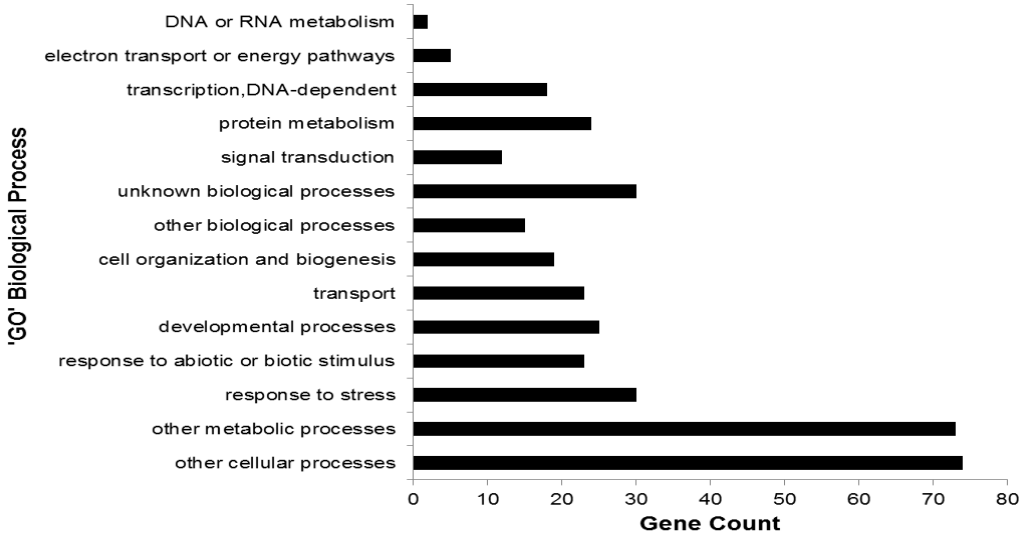

DNA or RNA metabolism
electron transport or energy pathways
transcription,DNA-dependent
protein metabolism
signal transduction
unknown biological processes
other biological processes
cell organization and biogenesis
transport
developmental processes
response to abiotic or biotic stimulus
response to stress
other metabolic processes
other cellular processes
'GO' Biological Process
0 10 20 30 40 50 60 70 80
Gene Count

Figure 4.7 ‘GO’ annotations of 153 genes upregulated in high root isoflavonoid cultivars (a) Molecular Function; (b) Cellular Component; (c) Biological function annotation; using the Gene Ontology Database annotations of TAIR identifiers that correspond to the 153 upregulated genes

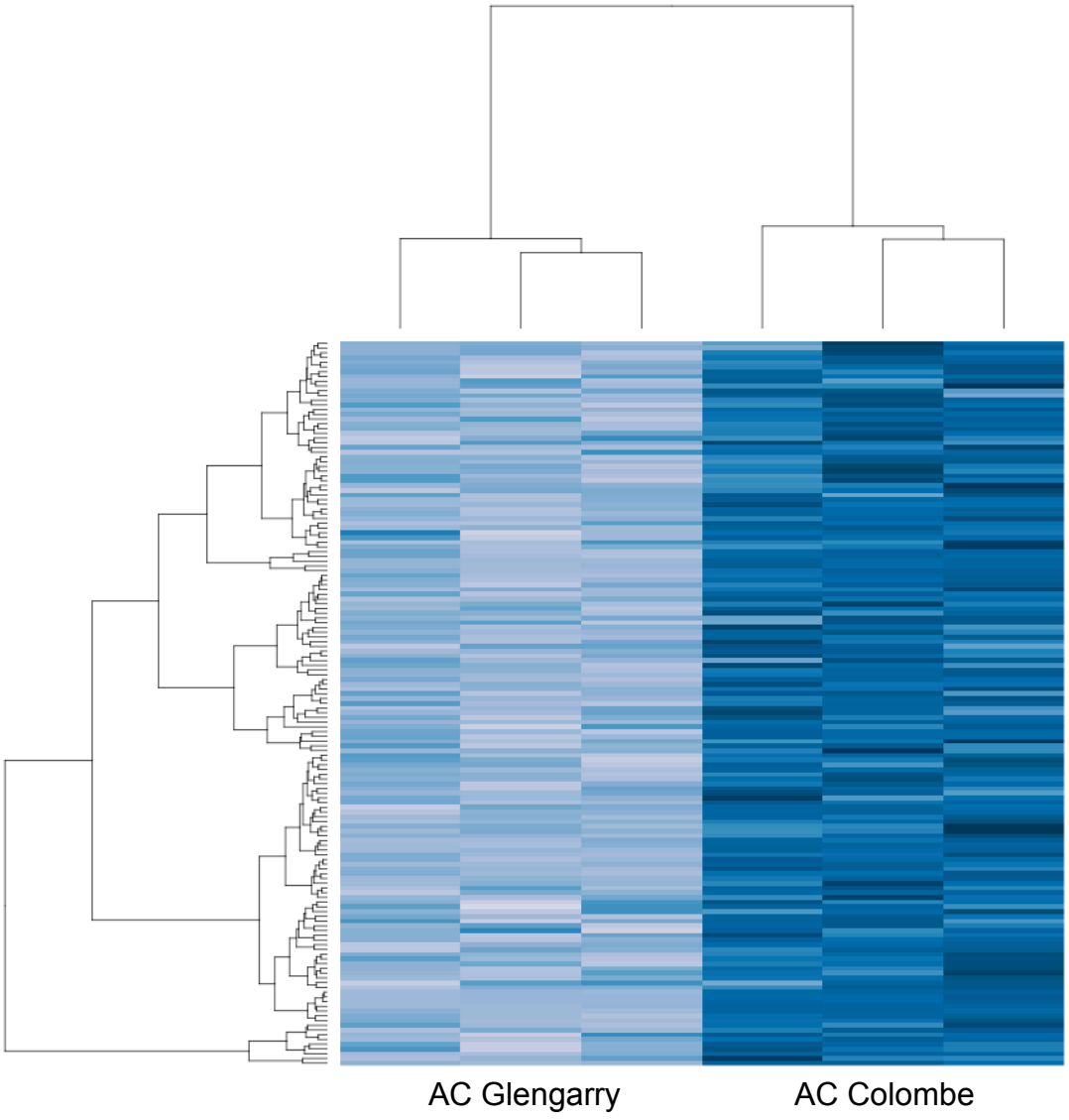


Figure 4.8 a

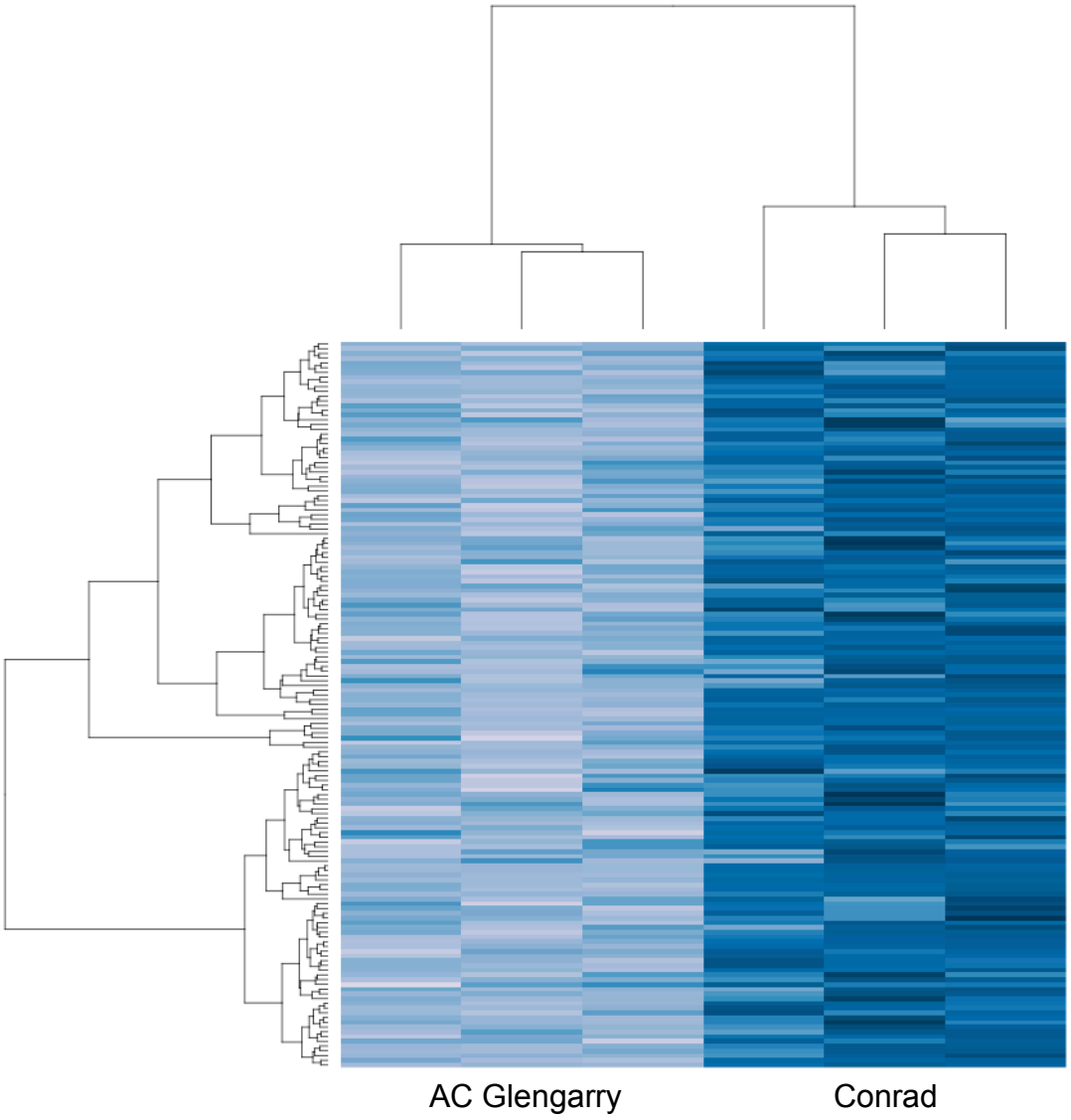


Figure 4.8 b

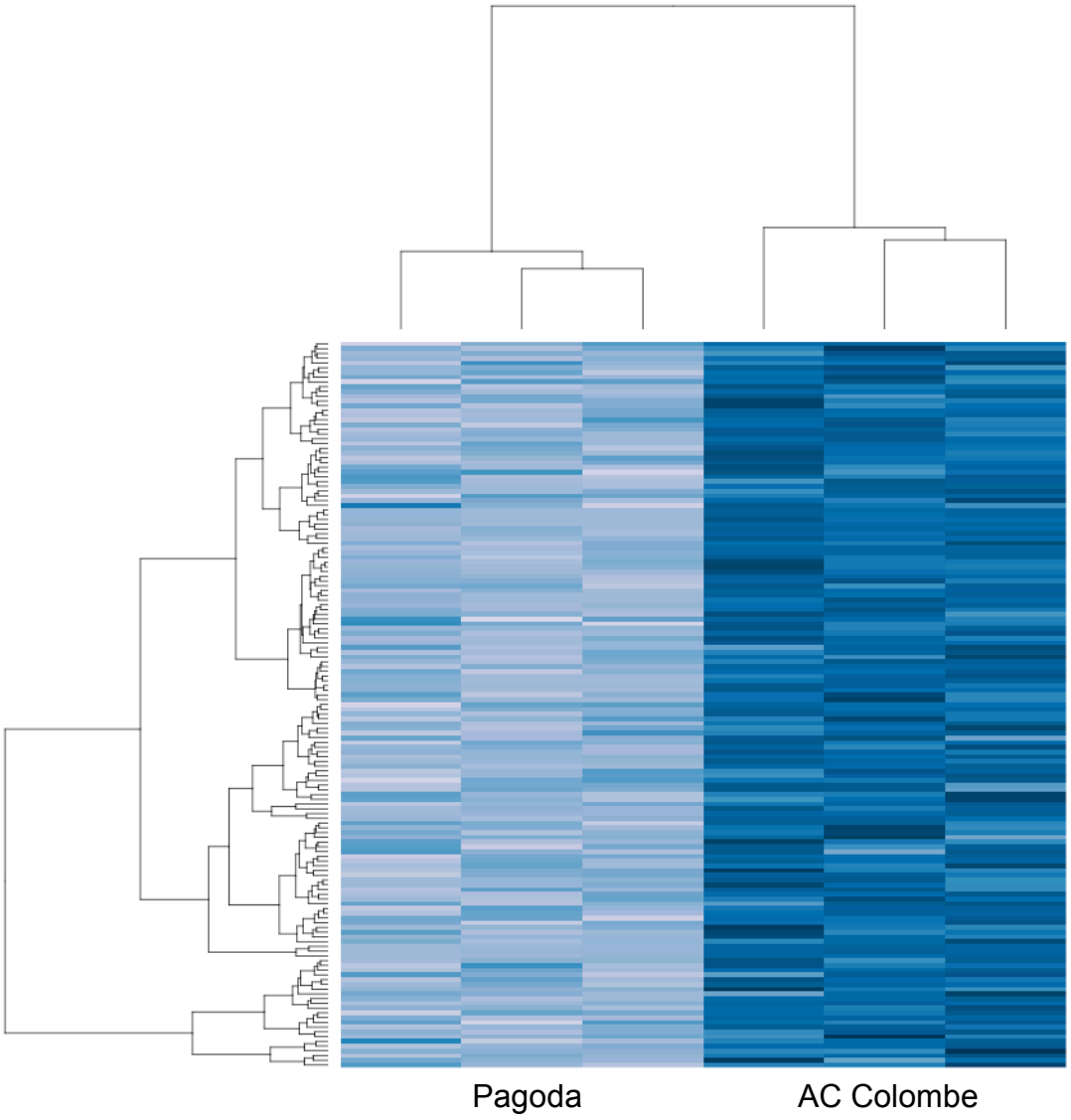


Figure 4.8 c

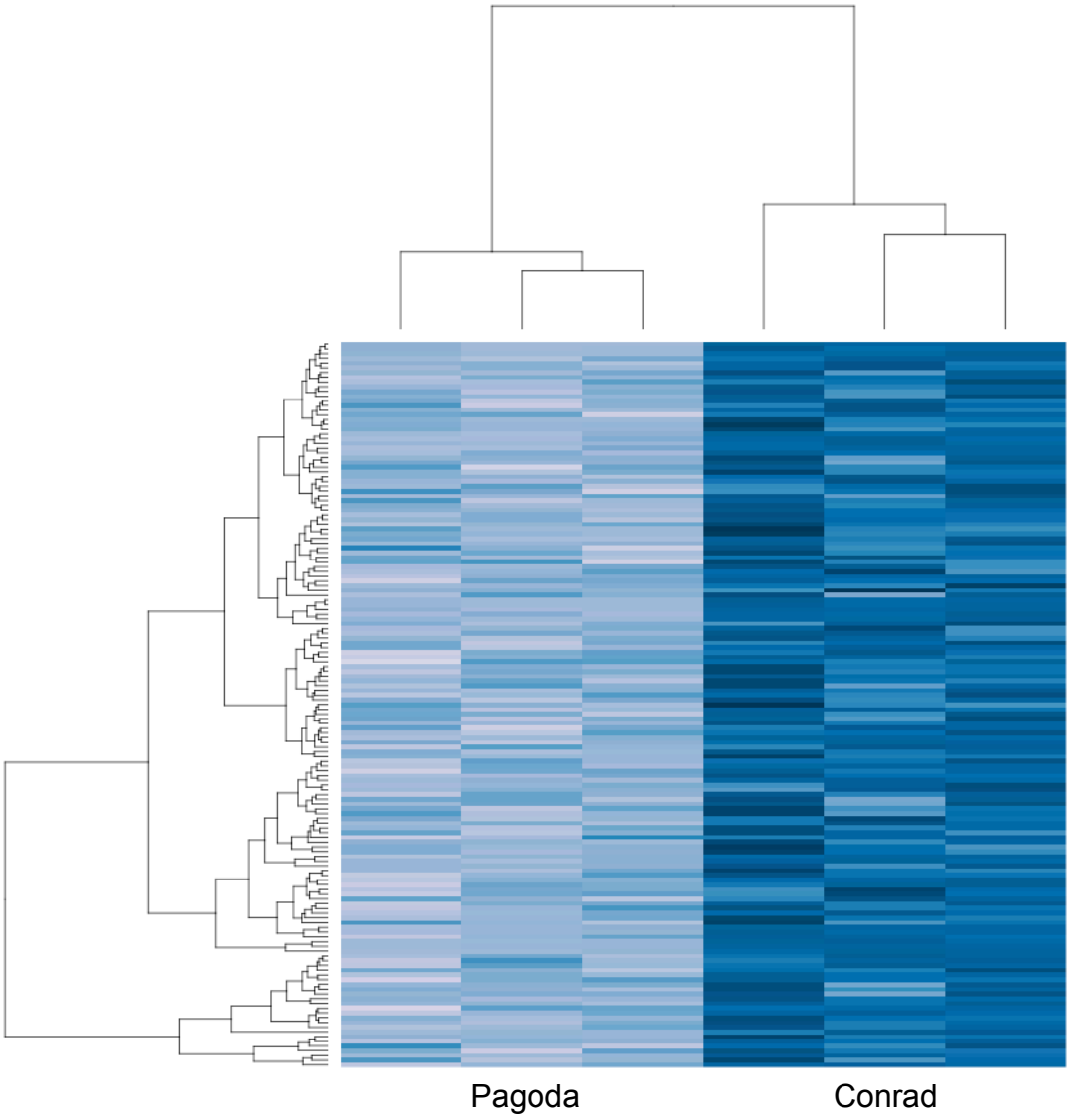
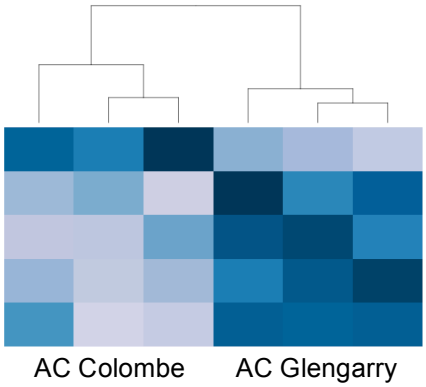


Figure 4.8 d

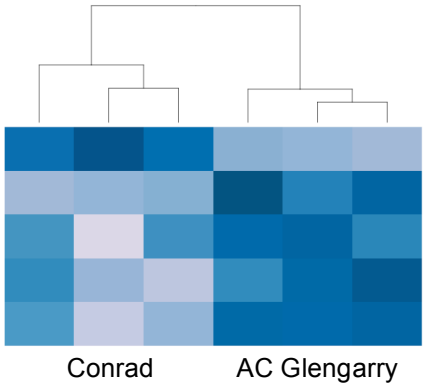
Figure 4.8 Heatmaps of genes upregulated in high root isoflavonoid cultivars. Read counts for gene models were normalized across cultivars. Four heatmaps representing genes upregulated in high isoflavonoid cultivars as compared with low root isoflavonoid cultivars were generated (three columns per cultivar representing biological replicates): (a) AC Glengarry with AC Colombe (b) AC Glengarry with Conrad (c) Pagoda with AC Colombe (d) Pagoda with Conrad.

a



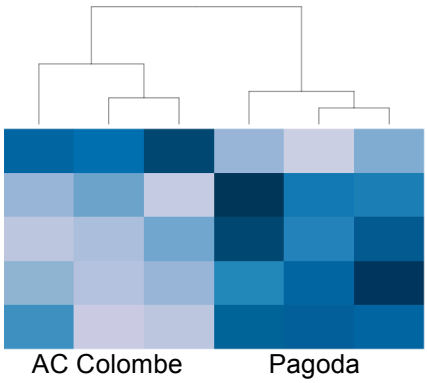
GmF6H - Glyma.08G326900.1
GmI2'H - Glyma.09G049300.1
GmF3'H - Glyma.17G134200.1
GmF3'5'H - Glyma.13G072100.1
GmDFR - Glyma.17G173200.1

b



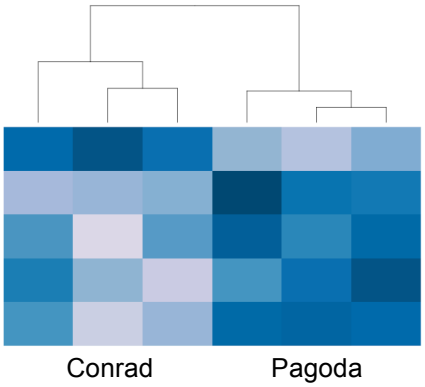
GmF6H - Glyma.08G326900.1
GmI2'H - Glyma.09G049300.1
GmF3'H - Glyma.17G134200.1
GmF3'5'H - Glyma.13G072100.1
GmDFR - Glyma.17G173200.1

c



GmF6H - Glyma.08G326900.1
GmI2'H - Glyma.09G049300.1
GmF3'H - Glyma.17G134200.1
GmF3'5'H - Glyma.13G072100.1
GmDFR - Glyma.17G173200.1

d



GmF6H - Glyma.08G326900.1
GmI2'H - Glyma.09G049300.1
GmF3'H - Glyma.17G134200.1
GmF3'5'H - Glyma.13G072100.1
GmDFR - Glyma.17G173200.1

Figure 4.9 Heatmaps of significantly differentially expressed phenylpropanoid genes. Lists of genes up- or down-regulated consistently were mined for gene models annotated with putative function in the phenylpropanoid pathway. *GmF6H* was upregulated, while *GmI2'H*, *GmF3'H*, *GmF3'5'H*, *GmDFR* were downregulated in high root isoflavonoid cultivars

The soybean cultivars chosen for this study, high (a) Conrad and AC Colombe, and low (b) AC Glengarry and Pagoda, root isoflavonoid cultivars, represent a metabolomic spectrum, resulting from differing biosynthesis and/or accumulation. While, the different steps of the isoflavonoid biosynthesis have been partially identified and characterized, past research has suggested that overall content is a coefficient of the metabolic flux of overlapping phenylpropanoid pathways. In addition, storage, accessory and structural elements could play an important role in metabolic content and composition. Therefore, the results of an overarching transcriptomic study of isoflavonoid content in the root, allow us to move beyond the few genes encoding enzymes in the pathway, to a macro-genetic perspective.

4.4.2 Differentially expressed phenylpropanoid genes

Differentially expressed phenylpropanoid genes might be indicative of up or down regulation at key enzymatic points in the overlapping metabolic pathways that direct flux into the isoflavonoid branch in root tissue. Attempts at reconstituting the isoflavonoid pathway in heterologous systems have underlined the importance of competition for flavanone substrates (Yu et al. 2000; Park et al. 2011; Falcone Ferreyra et al. 2012), particularly naringenin. This has been described as a 'bottleneck' (Liu et al. 2002); biosynthesis of isoflavonoids increases in mutant backgrounds with reduced or absent flow of substrates into flavonoid metabolism (Liu et al. 2007; Rasmussen and Jones 2013; Dhaubhadel et al. 2008). Another mechanism affecting the level of isoflavone aglycones and their glucoside conjugates is metabolism into downstream metabolites such as phytoalexins and signalling molecules (Dhaubhadel et al. 2008).

Interestingly, differential expression analyses identified four phenylpropanoid-related genes that were downregulated (*GmI2'H*, *GmF3'H*, *GmF3'5'H* and *GmDFR*), and one upregulated gene (*GmF6H*) in high root isoflavonoid cultivars. *GmI2'H* (Glyma.09G049300.1) is a putative isoflavonoid pathway gene annotated as a cytochrome P450, family 81 (CYP81D3), involved in the NADPH-dependent conversion of isoflavone into 2'-hydroxyisoflavone (Akashi et al. 1998; Liu et al. 2003). Substrates for GmI2'H include the isoflavones, daidzein and genistein, and the isoflavone derivative,

formononetin. Therefore, *GmI2'H* encodes a P450 capable of channelling isoflavones into downstream metabolites, and eventually pterocarpan phytoalexins (Shimada et al. 2000). Downregulation of *GmI2'H* in soybean roots might lead to decreased flux of isoflavone aglycones into subsequent pathways, thereby accounting for higher levels of daidzein and genistein in Conrad and AC Colombe.

Further phenylpropanoid genes that were downregulated include *GmF3'H*, *GmF3'5'H*, and *GmDFR*. All three are involved in the processing of naringenin to flavonoid metabolism. *GmF3'H* (Glyma.17G134200.1) and *GmF3'5'H* (Glyma.13G072100.1) encode putative cytochrome P450s that determine the hydroxylation pattern of the B-ring of flavonoids (Schwinn et al. 2014; Hagmann et al. 1983; Schoenbohm et al. 2000; Seitz et al. 2006). Together these P450s, F3'H and F3'5'H, are responsible for the production of dihydroflavonol, which is subsequently converted to leucoanthocyanidin by the action of DFR, the third flavonoid gene downregulated in high isoflavonoid cultivars (*GmDFR,* Glyma.17G173200.1) (Tanaka et al. 1995). The combined downregulation of these flavonoid pathway genes could play a role in the flux of flavanone substrates (naringenin) into the competing isoflavonoid biosynthesis branch and the increased production of genistein. As mentioned previously, flux of naringenin has been shown to be a bottleneck in attempts to increase or introduce isoflavonoid biosynthesis (Liu et al. 2007; Liu et al. 2002). The downregulation of the above genes is a significant finding in the search for possible mechanisms mediating metabolite flow.

4.4.3 Flavonoid 6-Hydroxylase 3: a determinant of glycitein content?

Flavonoid 6-Hydroxylase 3 (*GmF6H3*) (Glyma.08G326900.1) transcript was found at a higher level in high isoflavonoid cultivars, and encodes for a CYP450 enzyme that catalyzes the A-ring hydroxylation of liquiritigenin, synthesizing 6,7,4′-trihydroxyflavanone (Latunde-Dada et al. 2001). GmF6H3 is associated with two other highly similar isoforms GmF6H1 and GmF6H2 of the CYP71D9 family (Artigot et al. 2013). The intermediate produced in this reaction is further catalyzed by IFS (2-HIS) to produce 2,6,7,4′-trihydroxyflavanone, which undergoes a dehydration (by 2-hydroxyisoflavone dehydratase, 2-HID), and methylation (6-isoflavone-O-

methyltransferase 6-IOMT) event to produce glycitein (Latunde-Dada et al. 2001). Glycitein, genistein and daidzein are the three isoflavone aglycones produced in soybean, which appear predominantly in conjugated forms, as β-glycosides and their malonyl derivatives.

GmF6H was originally identified as a P450 that responds to elicitor-induction by chitosan and laminarin, both of which are compounds that originate from cell walls of fungi or marine microorganisms and are associated with defense reactions in higher plants (Latunde-Dada et al. 2001). It has been reported that GmF6H3 is the isoform responsible for glycitein content of soybean seeds, which is almost-exclusively limited to the hypocotyls (embryo-axis), as it was the sole member expressed in the tissue of question (Artigot et al. 2013). Expression of *GmF6H3* was concurrent with isoflavonoid accumulation in the hypocotyl (25-40 DAP); it was also absent from the seeds of soybean cultivars that are null-mutants for glycitein accumulation.

Strikingly, GmF6H3 unlike the other members of the GmF6H family lacks the characteristic N-terminal hydrophobic transmembrane domain that anchors P450 enzymes into the ER (Artigot et al. 2013). The N-terminus of GmF6H3 lacks the helix domain, and is reduced to four apolar amino acids, but is followed by the proline-rich and heme-binding domains, the latter being critical for P450 activity. Analysis of the deduced amino acid sequence of GmF6H3, using Target P 1.1, did not reveal any signal sequences that would target the protein to the ER or any other organelle.

The activity of GmF6H3 would suggest a close catalytic association with upstream CHI and downstream IFS, 2-HID and IOMT. Production of the 6, 7, 4′-trihydroxyflavanone intermediate would require the rapid subsequent catalysis into an isoflavanone and then isoflavone (Figure 1.3). Cytoplasmic localization of GmF6H3, as a free, cytoplasmic P450, would putatively require interaction with proteins in the isoflavonoid pathway to integrate its functionality in a metabolon. Such interactions might be of a more transient nature that is not readily perceived by the Co-IP approach, but must be further investigated.

The upregulation of *GmF6H3* in Conrad and AC Colombe, the two high isoflavonoid cultivars, as compared with the low isoflavonoid cultivars, suggests a role for this P450 enzyme in isoflavonoid content and composition. *GmF6H3* has been shown to be induced upon elicitor treatment, is expressed in glycitein-rich tissues and is absent in glycitein null-mutants (Artigot et al. 2013). This body of evidence places greater significance with regard to GmF6H3 as a possible regulator of isoflavonoid content in soybean roots.

4.4.4 Stress response in high isoflavonoid cultivars

Isoflavonoid biosynthesis is inducible by certain abiotic and biotic stimuli, as discussed in section 1.2.2. Therefore, it would be interesting to further investigate the 30 'high isoflavonoid' genes that are annotated as being responsive to stress. Subsets of these genes are annotated as being involved in disease resistance such as, dirigent proteins, which dictate the stereochemistry of other proteins (Shi et al. 2012), and several nucleotide binding disease resistance proteins. The latter included NB-LRR proteins that are associated with *R* gene function (Cesari et al. 2014; Martin et al. 2011). Several of these genes were annotated as being root hair specific. As root hairs are often the site for pathogen entry, the convergence of tissue-specialization and function could be indicative of a role for these genes in inhibiting pathogen infection or colonization.

As mentioned in the introduction to this chapter Conrad and AC Colombe, are also favoured by their increased resistance to *P. sojae* (Poysa, personal communication) and other soybean pathogens (Wang et al. 2012; Li et al. 2010). Therefore, the differential expression of genes involved in stress response, localized to the root hairs could be suggestive of a mechanism underlying the improved resistance to pathogen infection.

4.4.5 Transcriptional regulation: transcription factors and chromatin regulators

Subcellular localization of the 'high isoflavonoid' genes to the nucleus highlights the potential role of the cognate proteins in transcriptional regulation. The list of 40 nuclear-localized genes includes 18 transcriptional elements. Among the transcription factors upregulated in high isoflavonoid cultivars, are basic helix-loop-helix (bHLH), MADS-box and WD40 superfamily of proteins. The concurrently high transcript abundance of

these class of transcription factors could indicate physical interaction and regulatory synergy, as has been shown between particular sub-classes of MYB and bHLH transcription factors (Feller et al. 2011). Combinatorial plant gene regulation might be a factor in the coordinated redirection of flux in specialized metabolism, leading to isoflavonoid content variation in soybean cultivars.

Another function for nuclear-localized proteins encoded by upregulated 'high isoflavonoid' genes could be the regulation of chromatin structure. Two genes annotated for such function were identified: a SET7/9 family, histone H3K4-specific methyltransferase protein and a histone deacetylase 8 (*GmHDA8*) family protein. The former, histone-modifier protein, cannot be immediately associated with activation or repression of transcriptional regulation, as methylation of H3K4 would have to be regarded in the larger landscape and context of methylations (Qian et al. 2014).

GmHDA8, on the other hand, could be associated with the downregulation of certain genes, explaining the transcriptomic changes between cultivars. Histone tails are inherently positively charged due to the amine groups present on their lysine and arginine residues. Thusly, the positive charges on the histone tail bind with negative charges on the phosphate groups of the DNA backbone. Acetylation of the histone tail, catalyzed by histone acetyltransferases, neutralizes the charge, unravelling the histone-DNA association, and increasing accessibility to transcriptional elements. Deacetylation conferred by elements such as GmHDA8 would increase DNA-histone affinity and condense the overall structure, rendering it inaccessible to transcription machinery (Ma et al. 2013). Hypoacyetylated chromatin is silent, or has reduced gene expression. The functionality of histone acetylation can be specialized to a certain plant organ such as the root, and might be temporally determined. The presence or absence of acetylated histone tails allows for the plasticity of the transcriptome that could be regulated to improve resistance to stress or symbiosis in specialized tissues (Boycheva et al. 2014; Fang et al. 2014). Therefore, GmHDA8 might be a histone-modifying element responsible for some of the overall changes in the root transcriptome that coincide between the high isoflavonoid cultivars, Conrad and AC Colombe.

4.4.6 Vacuolar sequestration: a putative regulatory mechanism for isoflavonoid accumulation

As explained in section 3.4.6, isoflavone aglycones produced by the enzyme complex in the cytoplasm, tethered to the ER, are conjugated by transferases and transported rapidly into cellular compartments such as the vacuole. From this storage point, isoflavonoids can be sequestered outside of the metabolic flux and released upon demand for stress response or as signalling molecules for symbiosis with nitrogen-fixing rhizobia (Dhaubhadel et al. 2008). Interestingly, there were 14 genes annotated for transferase activity that were upregulated in high isoflavonoid cultivars. Due to the complexity of the plant metabolome and the general transferase activity, a subset of these transferases might be involved in isoflavone conjugation. Furthermore, 8 genes with putative transporter activity were identified to be consistently upregulated in high isoflavonoid cultivars. In this context, only 5 'high isoflavonoid' genes were associated with ER or golgi localization, which narrows down a preliminary search for potential transporters into the secretory pathway. The underrepresentation of vacuolar transporters could be due to the fact that transporters involved in the channelling of specialized metabolites can have broad substrate acceptance within a class of compounds (Frank et al. 2011; Marinova et al. 2007; Sugiyama et al. 2007). Therefore, there would be no requirement for the upregulation of transporters specific to isoflavonoid compounds.

The majority of phenylpropanoid metabolites, anthocyanins, proanthocyanidins and isoflavones are stored in the central vacuole of the cell; yet, the molecular basis of transport is still poorly understood. The glycosyl- and malonyl-glycosylation of isoflavone aglycones is necessary for their appropriate channelling into the vacuole for storage (Dhaubhadel et al. 2008). UDP glucuronosyltransferases (UGT) and MATE transporters are important families of proteins in the glycosylation and the subsequent transport of isoflavones (Zhao et al. 2011). Interestingly, one of the candidate upregulated genes identified in the differential expression analysis of soybean cultivars, was an uncharacterized member of the MATE family (Glyma.10G267800), and hereafter referred to as *GmMATE10*.

GmMATE10 was consistently upregulated in high isoflavonoid cultivars, suggesting a possible role in the sequestration of isoflavonoids. The predicted localization of this protein was to the secretory pathway, based on the presence of signal peptide. This could imply localization to the ER, golgi bodies or the vacuole, with the last option being highly probable, given its functional annotation as a MATE transporter. The deduced amino acid sequence of this gene had an amino acid sequence similarity of 32.47% with the TT12 homolog in *Arabidopsis,* based on Clustal W (McWilliam et al. 2013), the latter being a known proanthocyanidin vacuolar transporter (Zhao et al. 2011).

MATE transporters in *Medicago truncatula*, MATE1 and MATE2, are involved in transport of phenylpropanoids. MATE1 is a tonoplast epicatechin 3'-*O*-glucoside (E3'G) transporter (Zhao and Dixon 2009). MATE2 is a flavonoid transporter involved in vacuolar sequestration of anthocyanins and other flavonoids in flowers and leaves. MATE2 transporter prefers malonylated flavonoid glucosides and is co-expressed with three genes encoding malonyltransferases (Zhao et al. 2011). The discovery of an isoflavonoid-specific MATE transporter, in turn, could be used as a tool to investigate co-expressed transferases, depicting a more complete image of the conjugation and transport of isoflavonoids.

Furthermore, up-regulation of *GmMATE10* could indicate a mechanism for the vacuolar transport and sequestration of isoflavonoids and/or precursors, in competition with parallel pathways for metabolic flux. This could describe a regulatory mechanism determining higher isoflavonoid biosynthesis.

The study of plant specialized metabolites is progressing from an extensive study of the biosynthesis to an understanding of the inter- and sub-cellular modes of transport and storage. Bottlenecks for production and accumulation in native or heterologous systems will depend on sequestration of key intermediates and storage of the specialized product (Zhao and Dixon 2009). Elucidating the mechanism of isoflavonoid accumulation, whether it is by ATP binding cassette (ABC) or MATE transporters (such as GmMATE10), will be important in metabolic engineering of the isoflavonoid pathway.

4.5 Conclusion

The present transcriptomic study was designed to allow us to move beyond the few genes encoding enzymes in the pathway, to a bird's-eye view of the genetic actors involved in isoflavonoid content variation. Differential expression analyses identified possible candidates: three genes involved in the processing of naringenin away from isoflavonoid toward flavonoid biosynthesis, were downregulated (*GmF3'H*, *GmF3'5'H* and *GmDFR*); one gene that could putatively metabolize isoflavone aglycones toward phytoalexin production was also downregulated (*GmI2'H*); and the isoflavonoid gene *GmF6H3*, which encodes a putative P450 that provides the intermediate for the isoflavone glycitein was upregulated in high isoflavonoid cultivars. These cultivars, Conrad and AC Colombe, are also favourably placed in resistance to pathogens, which was reflected in the upregulation of a series of stress response genes. Furthermore, potential transcription factors involved in such large-scale transcriptomic changes, and transporters mediating the flow and storage of isoflavonoids metabolites have been identified. This represents a wealth of genetic data that can be harvested to pursue the molecular and functional basis of isoflavonoid content variation.

4.6 Literature Cited

Akashi T, Aoki T, Ayabe S (1998) CYP81E1, a cytochrome P450 cDNA of licorice (Glycyrrhiza echinata L.), encodes isoflavone 2'-hydroxylase. Biochem Biophys Res Commun 251 (1):67-70

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11 (10):R106

Artigot MP, Baes M, Dayde J, Berger M (2013) Expression of flavonoid 6-hydroxylase candidate genes in normal and mutant soybean genotypes for glycitein content. Mol Biol Rep 40 (7):4361-4369

Blanc G, Wolfe KH (2004) Widespread Paleopolyploidy in Model Plant Species Inferred from Age Distributions of Duplicate Genes. Plant Cell 16:1667-1678

Boycheva I, Vassileva V, Iantcheva A (2014) Histone acetyltransferases in plant development and plasticity. Current genomics 15 (1):28-37

Cesari S, Kanzaki H, Fujiwara T, Bernoux M, Chalvon V, Kawano Y, Shimamoto K, Dodds P, Terauchi R, Kroj T (2014) The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. Embo J 33 (17):1941-1959

Dhaubhadel S, Farhangkhoee M, Chapman R (2008) Identification and characterisation of isoflavonoid specific glycosyltransferase and malonyltransferase from soybean seeds. J Exp Bot 59 (4):981-994

Dhaubhadel S, Gijzen M, Moy P, Farhangkhoee M (2007) Transcriptome analysis reveals a critical role of *CHS7* and *CHS8* genes for isoflavonoid synthesis in soybean seeds. Plant Physiol 143 (1):326-338

Dixon RA, Achnine L, Kota P, Liu CJ, Reddy MSS, Wang L (2002) The phenylpropanoid pathway and plant defence - A genomics perspective. Mol Plant Pathol 3 (5):371-390

Falcone Ferreyra ML, Rius S, Casati P (2012) Flavonoids: Biosynthesis, Biological functions and Biotechnological applications. Frontiers in Plant Science 3

Fang H, Liu X, Thorn G, Duan J, Tian L (2014) Expression analysis of histone acetyltransferases in rice under drought stress. Biochem Biophys Res Commun 443 (2):400-405

Feller A, Machemer K, Braun EL, Grotewold E (2011) Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. The Plant Journal 66 (1):94-116

Frank S, Keck M, Sagasser M, Niehaus K, Weisshaar B, Stracke R (2011) Two differentially expressed MATE factor genes from apple complement the Arabidopsis transparent testa12 mutant. Plant Biol (Stuttg) 13 (1):42-50

Hagmann ML, Heller W, Grisebach H (1983) Induction and characterization of a microsomal flavonoid 3'-hydroxylase from parsley cell cultures. European journal of biochemistry / FEBS 134 (3):547-554

Latunde-Dada AO, Cabello-Hurtado F, Czittrich N, Didierjean L, Schopfer C, Hertkorn N, Werck-Reichhart D, Ebel J (2001) Flavonoid 6-hydroxylase from soybean (Glycine max L.), a novel plant P-450 monooxygenase. J Biol Chem 276 (3):1688-1695

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25 (14):1754-1760

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25 (16):2078-2079

Li X, Han Y, Teng W, Zhang S, Yu K, Poysa V, Anderson T, Ding J, Li W (2010) Pyramided QTL underlying tolerance to Phytophthora root rot in mega-environments from soybean cultivars 'Conrad' and 'Hefeng 25'. Theor Appl Genet 121:651-658

Lin H, Rao J, Shi J, Hu C, Cheng F, Wilson ZA, Zhang D, Quan S (2014) Seed metabolomic study reveals significant metabolite variations and correlations among different soybean cultivars. J Integr Plant Biol

Liu CJ, Blount JW, Steele CL, Dixon RA (2002) Bottlenecks for metabolic engineering of isoflavone glycoconjugates in Arabidopsis. Proc Natl Acad Sci USA 99 (22):14578-14583

Liu CJ, Huhman D, Sumner LW, Dixon RA (2003) Regiospecific hydroxylation of isoflavones by cytochrome p450 81E enzymes from Medicago truncatula. Plant J 36 (4):471-484

Liu RR, Hu YL, Li HL, Lin ZP (2007) Production of soybean isoflavone genistein in non-legume plants via genetically modified secondary metabolism pathway. Metab Eng 9 (1):1-7

Livingstone JM, Seguin P, Strömvik MV (2010) An *in silico* study of the genes for the isoflavonoid pathway enzymes in soybean reveals novel expressed homologues. Can J Plant Sci 90 (4):453-469

Ma X, Lv S, Zhang C, Yang C (2013) Histone deacetylases and their functions in plants. Plant Cell Rep 32 (4):465-478

Marinova K, Pourcel L, Weder B, Schwarz M, Barron D, Routaboul JM, Debeaujon I, Klein M (2007) The Arabidopsis MATE transporter TT12 acts as a vacuolar flavonoid/H+ -antiporter active in proanthocyanidin-accumulating cells of the seed coat. Plant Cell 19 (6):2023-2038

Marioni JC, Mason CE, Mane SM, AStephens M, Gilad Y (2008) RNA-seq:an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 18:1509-1517

Martin T, Biruma M, Fridborg I, Okori P, Dixelius C (2011) A highly conserved NB-LRR encoding gene cluster effective against Setosphaeria turcica in sorghum. BMC Plant Biol 11:151

McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R (2013) Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res 41 (Web Server issue):W597-600

Oliveros JC (2007) VENNY. An interactive tool for comparing lists with Venn Diagrams.

Park N, Xu H, Li X, Kim S-J, Park S (2011) Enhancement of flavone levels through overexpression of chalcone isomerase in hairy root cultures of *Scutellaria baicalensis*. Functional & Integrative Genomics 11 (3):491-496

Phillips DA, Kapulnik Y (1995) Plant isoflavonoids, pathogens and symbionts. Trends Microbiol 3 (2):58-64

Qian Y, Xi Y, Cheng B, Zhu S, Kan X (2014) Identification and characterization of the SET domain gene family in maize. Mol Biol Rep 41 (3):1341-1354

Rasmussen S, Jones C (2013) Potential for producing increased levels of isoflavones in transgenic plants. CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources 8

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463 (7278):178-183

Schoenbohm C, Martens S, Eder C, Forkmann G, Weisshaar B (2000) Identification of the Arabidopsis thaliana flavonoid 3'-hydroxylase gene and functional expression of the encoded P450 enzyme. Biol Chem 381 (8):749-753

Schwinn K, Miosic S, Davies K, Thill J, Gotame TP, Stich K, Halbwirth H (2014) The B-ring hydroxylation pattern of anthocyanins can be determined through activity of the flavonoid 3'-hydroxylase on leucoanthocyanidins. Planta

Seitz C, Eder C, Deiml B, Kellner S, Martens S, Forkmann G (2006) Cloning, functional identification and sequence analysis of flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase cDNAs reveals independent evolution of flavonoid 3',5'-hydroxylase in the Asteraceae family. Plant Mol Biol 61 (3):365-381

Shi H, Liu Z, Zhu L, Zhang C, Chen Y, Zhou Y, Li F, Li X (2012) Overexpression of cotton (Gossypium hirsutum) dirigent1 gene enhances lignification that blocks the spread of Verticillium dahliae. Acta biochimica et biophysica Sinica 44 (7):555-564

Shimada N, Akashi T, Aoki T, Ayabe S (2000) Induction of isoflavonoid pathway in the model legume Lotus japonicus: molecular characterization of enzymes involved in phytoalexin biosynthesis. Plant science : an international journal of experimental plant biology 160 (1):37-47

Shoemaker RC, Schlueter J, Doyle JJ (2006) Paleopolyploidy and gene duplication in soybean and other legumes. Curr Opin Plant Biol 9 (2):104 - 109

Subramanian S, Graham MA, Yu O, Graham TL (2005) RNA interference of soybean isoflavone synthase genes leads to silencing in tissues distal to the transformation site and to enhanced susceptibility to *Phytophthora sojae*. Plant Physiol 137:1345-1353

Sugiyama A, Shitan N, Yazaki K (2007) Involvement of a soybean ATP-binding cassette-type transporter in the secretion of genistein, a signal flavonoid in legume-Rhizobium symbiosis. Plant Physiol 144 (4):2000-2008

Tanaka Y, Fukui Y, Fukuchi-Mizutani M, Holton TA, Higgins E, Kusumi T (1995) Molecular cloning and characterization of Rosa hybrida dihydroflavonol 4-reductase gene. Plant Cell Physiol 36 (6):1023-1031

Wang H, Wijeratne A, Wijeratne S, Lee S, Taylor CG, St Martin SK, McHale L, Dorrance AE (2012) Dissection of two soybean QTL conferring partial resistance to Phytophthora sojae through sequence and gene expression analysis. BMC Genomics 13:428

Yi J, Derynck MR, Chen L, Dhaubhadel S (2010) Differential expression of *CHS7* and *CHS8* genes in soybean. Planta 231:741-753

Yu O, Jung W, Shi J, Croes RA, Fader GM, McGonigle B, Odell JT (2000) Production of the Isoflavones Genistein and Daidzein in Non-Legume Dicot and Monocot Tissues. Plant Physiol 124 (2):781-794

Zhao J, Dixon RA (2009) MATE transporters facilitate vacuolar uptake of epicatechin 3'-O-glucoside for proanthocyanidin biosynthesis in Medicago truncatula and Arabidopsis. Plant Cell 21 (8):2323-2340

Zhao J, Huhman D, Shadle G, He XZ, Sumner LW, Tang Y, Dixon RA (2011) MATE2 mediates vacuolar sequestration of flavonoid glycosides and glycoside malonates in Medicago truncatula. Plant Cell 23 (4):1536-1555

CHAPTER FIVE – GENERAL DISCUSSION

5.1 A substantiated model – the isoflavonoid metabolon

Plant production of specialized metabolites is differentiated and partitioned at many levels, with the metabolon representing the final, molecular level of organization. Here I have provided evidence for the existence of an isoflavonoid metabolon, with tandem P450 anchors, C4H and IFS, nucleating cytoplasmic enzymes to assemble at the surface of the ER through protein-protein interactions (Figure 5.1).

Prior to this research, evidence for the formation of an isoflavonoid metabolon was speculative, with the co-localization of IFS and IOMT in alfalfa being the only indication of such a complex (Liu and Dixon 2001). The interactions of GmIFS1 and GmIFS2 with cytoplasmic enzymes upstream in the isoflavonoid pathway, including GmCHS7, GmCHS8 and GmCHI2, and the isoform-specific interactions of GmIFS2 with GmCHR14, GmADT6, GmC4H and GmCH1B1, have provided evidence of the constituents and circumstances of the isoflavonoid metabolon. The present research extends our expectations of the reach of a metabolon, and suggests a mechanism for the regulation of phenylalanine flux into isoflavonoid biosynthesis.

5.2 Isoform specificity: spatio-temporal optimization of metabolons

The constituents of a proposed isoflavonoid metabolon might differ on the basis of tissue, development and external stimuli. Previous literature has shown that members of the gene families in the isoflavonoid pathway encoding: GmIFS (Dhaubhadel et al. 2003), GmCHS (Dhaubhadel et al. 2007), and in this study, GmCHI, are differentially expressed in various organs and tissues. In fact, this level of organization is very intricate, as was observed in the striking expression of *GmCHI1B1:GUS* in leaf trichomes. Therefore, as isoflavonoids have been detected in a variety of organs and tissues, most strongly in the

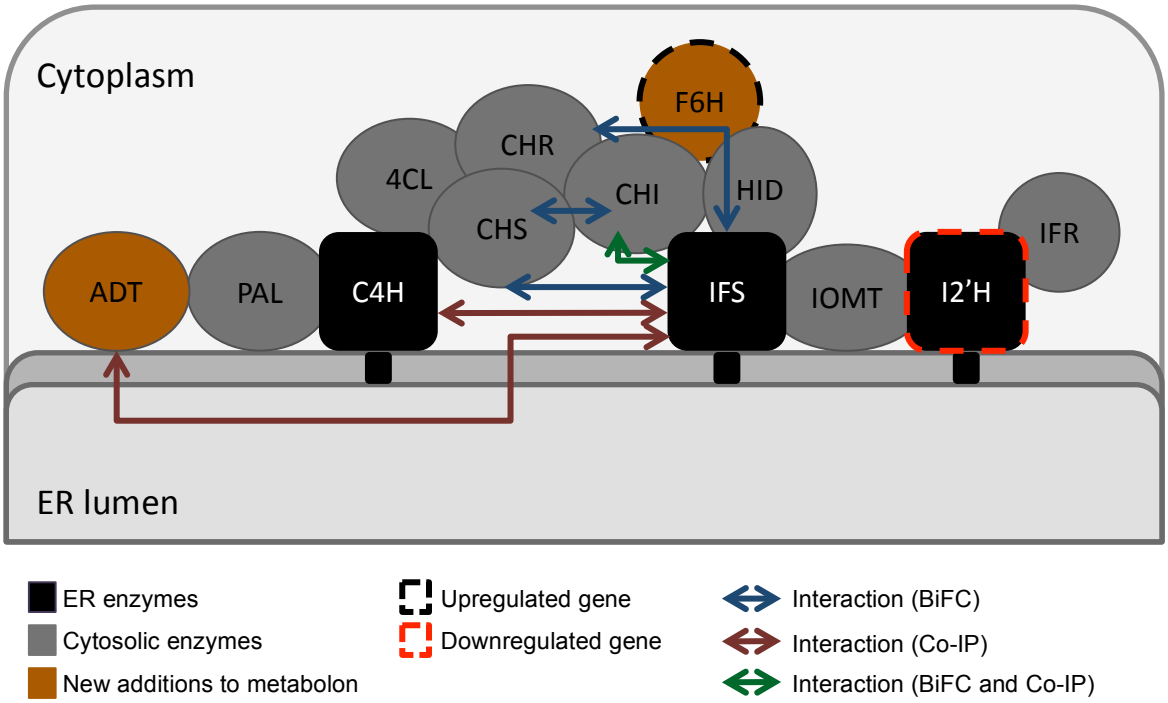


Figure 5.1 An updated, generalized model of the isoflavonoid metabolon. Protein-protein interactions confirmed by BiFC and Co-IP studies are shown; up and downregulation of putative members of the metabolon are highlighted; new additions to the metabolon are also noted. ADT (arogenate dehydratase); PAL (phenylalanine ammonia lyase); C4H (cinnamate 4-hydroxylase); 4CL (4-coumarate-CoA ligase); CHS (chalcone synthase); CHR (chalcone reductase); CHI (chalcone isomerase); F6H (flavone 6-hydroxylase); IFS (isoflavone synthase); HID (2-hydroxyisoflavanone dehydratase); IOMT (isoflavone O-methyltransferase); I2'H (isoflavone 2'-hydroxylase); IFR (isoflavone reductase)

leaves and the seeds, and with great physiological importance in the roots, it can be assumed that the metabolons of these organs might vary.

The Co-IP based approach of finding interacting partners with GmIFS2 allowed the specific identification of GmCHI1B1, from four catalytically active GmCHIs. This finding suggests that GmCHI1B1 would be the member of the gene family involved in an isoflavonoid metabolon in soybean roots.

In soybean seeds, the spatio-temporal organization of isoflavonoid production might shift toward the recruitment of GmCHI1A. Isoflavonoid accumulation follows embryo maturation reaching an apex from 50 DAP till full maturation (Dhaubhadel et al. 2003). This trend was followed closing by GmCHI1A and GmCHI4A (Figure 2.3). GmCHI1A is a 'type II' legume-specific CHI that could convert both isoliquiritigenin and naringenin chalcone to their corresponding flavanones; GmCHI4A is a 'type IV' CHI that has not been functionally characterized, but has been linked with enhancing (iso)flavonoid production in other organisms (Morita et al. 2014). In mature seeds, where most of the isoflavonoids are accumulated, only *GmCHI1A* was expressed concurrently with high levels of *GmIFS2*, *GmCHS7* and *GmCHS8* (Figure 2.4). This describes an expression pattern that would indicate the possibility of GmCHI1A being a partner in the isoflavonoid metabolon of soybean seeds, alongside GmIFS2 and other pathway enzymes.

5.3 The 'inducible' metabolon

As mentioned previously, one of the factors that may influence the specific constituents of the isoflavonoid metabolon, in a given tissue, is environmental stimulus. Isoflavonoid biosynthesis is particularly associated with stress and pathogen response. Isoflavonoids, the active aglycones, daidzein, and their complex derivatives, glyceollins, are phytoalexins. Rapid induction of isoflavonoid biosynthesis is one of the major advantages supporting the existence of a metabolon. Infection by soybean cyst nematode (SCN), a

pathogen (Afzal et al. 2009), and *B. japonicum*, a symbiotic bacterium (Nguyen et al. 2012), both effect the induction of isoflavone-related enzymes, IFR, PAL, and CHI. Abiotic stress such as flooding also increases abundance of phenylpropanoid enzymes, CHI and IOMT in the roots (Khatoon et al. 2012; Komatsu et al. 2013). These functional studies are indicative of the possibility that the constituents of the isoflavonoid metabolon could change in soybean tissues exposed to biotic or abiotic stress. Rearrangement or substitution of elements could channel metabolites into downstream pathways leading to phytoalexin production.

As *GmIFS2* has been shown to be inducible by *P. sojae* infection (Dhaubhadel et al. 2003), the interacting partners of GmIFS2, from the Co-IP and LC-MS/MS study, could be functionally studied to identify new candidates in the 'inducible' metabolon. Beyond the pathway enzymes identified (GmADT6, GmC4H and GmCH1B1), 42% of the proteins enumerated in the Co-IP results were responsive to stress. These candidates could function as structural, accessory or biosynthetic elements in the soybean root metabolon assembled to respond rapidly to stress.

Furthermore, the information from the transcriptomic study of isoflavonoid content can also help furnish the image of stress resistance and induction in soybean roots. The 30 'high isoflavonoid' genes that were annotated as being responsive to stress, including dirigent proteins and several associated with *R* gene function, would be ideal candidates to examine for association with isoflavonoid production or the metabolon. This is particularly pertinent, as high root isoflavonoid cultivars are also characterized with higher resistance to pathogens (Wang et al. 2012; Li et al. 2010). The candidates from both the Co-IP and transcriptomics studies would be ideal starting points for investigation into broader elements involved in a proposed 'inducible' metabolon.

5.4 Metabolic engineering of plant specialized metabolites

Efforts to engineer isoflavonoid biosynthesis in non-legumes have described the accessibility of flavanone substrates as the bottleneck. Endogenous metabolic systems in

non-legumes and leguminous plants or varieties with absent/low isoflavonoid production, have tightly regulated channeling of metabolites toward flavonoid pathways (Liu et al. 2002; Sreevidya et al. 2006; Winkel 2004; Yu and McGonigle 2005; Yu et al. 2003).

The findings from the metabolon and transcriptomic studies can be applied to the pursuit of metabolic engineering of isoflavonoids. The former has helped us identify the metabolic and structural constituents of the production machinery, while the latter has informed us on the large-scale nature of differential expression that is required to increase isoflavonoid production.

The metabolon study highlighted the importance of recruiting upstream enzymes such as GmADT6 to the isoflavonoid complex at the surface of the ER. Such information can be used in the progression of synthetic reconstitution of isoflavonoid biosynthesis. The addition of specific isoforms of upstream phenylpropanoid enzymes to IFS and isoflavonoid-specific enzymes could improve production efficiency in heterologous systems. Moving forward, knowledge of the elements and interactions within the metabolon can provide the platform for rapid discovery, synthesis and modification of isoflavonoids.

The differential expression study of high and low root isoflavonoid cultivars identified possible candidates that would regulate the metabolic flux into the isoflavonoid pathway in preference to other branches of phenylpropanoid metabolism. The three genes, *GmF3'H*, *GmF3'5'H* and *GmDFR*, encoding enzymes in competition with IFS for flavanone substrates were downregulated, while the isoflavonoid gene *GmF6H3* that provides the intermediate for glycitein was upregulated in high isoflavonoid cultivars. This describes one method of mediating metabolite flow away from flavonoid toward isoflavonoid biosynthesis. The other method is the stabilization and sequestration of isoflavone metabolites in their aglycone or conjugated forms, by downregulation of genes metabolizing isoflavones to phytoalexins (*GmI2'H*), and upregulation of putative isoflavone transporters (*GmMATE10*).

The conjunction of knowledge from the present study on a) the metabolic enzyme isoforms linked with isoflavonoid production; b) the constituents and circumstances of

the isoflavonoid metabolon depending on time and tissue; c) the overall transcriptomic changes mediating metabolic flux toward the isoflavonoid pathway, can help us understand and manipulate the metabolic content of soybean and provide the platform for introduction into heterologous systems.

5.5 Literature Cited

Afzal AJ, Natarajan A, Saini N, Iqbal MJ, Geisler M, El Shemy HA, Mungur R, Willmitzer L, Lightfoot DA (2009) The nematode resistance allele at the rhg1 locus alters the proteome and primary metabolism of soybean roots. Plant Physiol 151 (3):1264-1280

Dhaubhadel S, Gijzen M, Moy P, Farhangkhoee M (2007) Transcriptome analysis reveals a critical role of *CHS7* and *CHS8* genes for isoflavonoid synthesis in soybean seeds. Plant Physiol 143 (1):326-338

Dhaubhadel S, McGarvey BD, Williams R, Gijzen M (2003) Isoflavonoid biosynthesis and accumulation in developing soybean seeds. Plant Mol Biol 53 (6):733-743

Khatoon A, Rehman S, Salavati A, Komatsu S (2012) A comparative proteomics analysis in roots of soybean to compatible symbiotic bacteria under flooding stress. Amino acids 43 (6):2513-2525

Komatsu S, Makino T, Yasue H (2013) Proteomic and biochemical analyses of the cotyledon and root of flooding-stressed soybean plants. PLoS One 8 (6):e65301

Li X, Han Y, Teng W, Zhang S, Yu K, Poysa V, Anderson T, Ding J, Li W (2010) Pyramided QTL underlying tolerance to Phytophthora root rot in mega-environments from soybean cultivars 'Conrad' and 'Hefeng 25'. Theor Appl Genet 121:651-658

Liu CJ, Blount JW, Steele CL, Dixon RA (2002) Bottlenecks for metabolic engineering of isoflavone glycoconjugates in Arabidopsis. Proc Natl Acad Sci USA 99 (22):14578-14583

Liu CJ, Dixon RA (2001) Elicitor-induced association of isoflavone O-methyltransferase with endomembranes prevents the formation and 7-O-methylation of daidzein during isoflavonoid phytoalexin biosynthesis. Plant Cell 13 (12):2643-2658

Morita Y, Takagi K, Fukuchi-Mizutani M, Ishiguro K, Tanaka Y, Nitasaka E, Nakayama M, Saito N, Kagami T, Hoshino A, Iida S (2014) A chalcone isomerase-like protein enhances flavonoid production and flower pigmentation. Plant J 78:294-304

Nguyen TH, Brechenmacher L, Aldrich JT, Clauss TR, Gritsenko MA, Hixson KK, Libault M, Tanaka K, Yang F, Yao Q, Pasa-Tolic L, Xu D, Nguyen HT, Stacey G (2012) Quantitative phosphoproteomic analysis of soybean root hairs inoculated with Bradyrhizobium japonicum. Mol Cell Proteomics 11 (11):1140-1155

Sreevidya VS, Rao CS, Sullia SB, Ladha JK, Reddy PM (2006) Metabolic engineering of rice with soybean isoflavone synthase for promoting nodulation gene expression in rhizobia. J Exp Bot 57 (9):1957-1969

Wang H, Wijeratne A, Wijeratne S, Lee S, Taylor CG, St Martin SK, McHale L, Dorrance AE (2012) Dissection of two soybean QTL conferring partial resistance to Phytophthora sojae through sequence and gene expression analysis. BMC Genomics 13:428

Winkel BS (2004) Metabolic channeling in plants. Annu Rev Plant Biol 55:85-107

Yu O, McGonigle B (2005) Metabolic engineering of isoflavone biosynthesis. Adv Agron 86:147 - 190

Yu O, Shi J, Hession AO, Maxwell AA, McGonigle B, Odell JT (2003) Metabolic engineering to increase isoflavone biosynthesis in soybean seeds. Phytochemistry 63:753-763

APPENDICES

Appendix A: Complete list of proteins that co-immunoprecipitate with GmIFS2 (in all three replicates). Number of unique peptides, identified by LC-MS/MS analysis, associated with Glyma identifiers indicated for each replicate.

Gene description	Glyma identifier	No. of unique peptides (per biol. repl.)		
		1	2	3
Ribosomal protein S5/Elongation factor G/III/V family protein	Glyma.15G257100.1.p	23	24	4
TCP-1/cpn60 Chaperonin family protein	Glyma.11G185500.1.p	8	2	12
TCP-1/cpn60 Chaperonin family protein	Glyma.12G087300.1.p	8	2	12
TCP-1/cpn60 Chaperonin family protein	Glyma.08G364000.1.p	8	7	6
TCP-1/cpn60 Chaperonin family protein	Glyma.18G298100.1.p	8	7	6
ARP protein (REF)	Glyma.02G133500.1.p	11	2	5
Lipoxygenase 1	Glyma.07G007000.1.p	7	6	5
Chaperonin-60alpha	Glyma.12G078100.1.p	11	4	3
P-loop containing nucleoside triphosphate hydrolases superfamily protein	Glyma.08G335300.1.p	2	9	6
26S Proteasome regulatory subunit S2 1A	Glyma.12G117600.1.p	4	11	2
26S Proteasome regulatory subunit S2 1A	Glyma.06G289400.1.p	4	11	1
Heat shock protein 60	Glyma.10G193200.1.p	10	1	5
TCP-1/cpn60 Chaperonin family protein	Glyma.11G244700.1.p	8	3	5
TCP-1/cpn60 Chaperonin family protein	Glyma.18G012400.1.p	8	3	5
Heat shock protein 60	Glyma.20G197100.1.p	10	1	5
2-Oxoglutarate dehydrogenase, E1 component	Glyma.02G094300.1.p	7	4	4
TCP-1/cpn60 Chaperonin family protein	Glyma.02G273200.1.p	7	1	7
ATP-citrate lyase B-1	Glyma.09G035600.1.p	7	1	7
Glycosyl hydrolase family protein	Glyma.10G106600.1.p	7	4	4
Lipoxygenase 1	Glyma.13G347700.1.p	2	5	8
TCP-1/cpn60 Chaperonin family protein	Glyma.14G043400.1.p	7	1	7
ATP-citrate lyase B-1	Glyma.15G140300.1.p	7	1	7
Vacuolar ATP synthase subunit D (VATD)	Glyma.07G023400.1.p	5	2	7
3-Methylcrotonyl-CoA carboxylase	Glyma.07G213500.1.p	7	4	3
Ribosomal protein S5/Elongation factor G/III/V family protein	Glyma.08G170000.2.p	4	6	4

Vacuolar ATP synthase subunit D (VATD)	Glyma.08G218500.1.p	5	2	7
Potassium channel beta subunit 1	Glyma.20G077200.1.p	4	4	6
Translation initiation factor 3B1	Glyma.03G002400.1.p	3	2	8
CLPC homologue 1	Glyma.04G200400.1.p	4	2	7
CLPC homologue 1	Glyma.06G165200.1.p	4	2	7
51 kDa subunit of complex I	Glyma.02G003300.1.p	4	2	6
Protein of unknown function DUF2359, transmembrane	Glyma.05G139900.1.p	6	2	4
TUDOR-SN protein 1	Glyma.06G162500.1.p	5	6	1
Translation initiation factor 3B1	Glyma.07G107700.1.p	3	2	7
Protein of unknown function DUF2359, transmembrane	Glyma.08G095200.1.p	6	2	4
51 kDa subunit of complex I	Glyma.10G005500.1.p	4	2	6
Transducin/WD40 repeat-like superfamily protein	Glyma.01G029600.1.p	2	3	6
HAD superfamily, subfamily IIIB acid phosphatase	Glyma.07G014500.1.p	4	2	5
TCP-1/cpn60 Chaperonin family protein	Glyma.09G158200.1.p	5	4	2
Coatomer, beta-subunit	Glyma.13G360100.1.p	3	4	4
Sec23/Sec24 protein transport family protein	Glyma.19G226200.1.p	4	4	3
Ribosomal protein S4 (RPS4A) family protein	Glyma.02G047800.1.p	4	3	3
Eukaryotic translation initiation factor 3E	Glyma.02G187100.1.p	2	6	2
Ribosomal protein S4 (RPS4A) family protein	Glyma.02G263200.1.p	4	3	3
N-terminal nucleophile aminohydrolases (Ntn hydrolases) superfamily protein	Glyma.04G077400.1.p	2	5	3
TUDOR-SN protein 1	Glyma.04G203000.1.p	4	4	2
Acyl-CoA oxidase 1	Glyma.05G062200.1.p	5	4	1
N-terminal nucleophile aminohydrolases (Ntn hydrolases) superfamily protein	Glyma.06G078500.1.p	2	5	3
ADP/ATP Carrier 3	Glyma.06G290600.1.p	6	2	2
Phospholipase D alpha 1	Glyma.07G031100.1.p	2	1	7
Sucrose synthase 4	Glyma.09G073600.2.p	2	2	6
Annexin 5	Glyma.09G171600.1.p	2	1	7
Eukaryotic translation initiation factor 3E	Glyma.10G106100.1.p	2	6	2
Pyruvate decarboxylase-2	Glyma.13G231700.1.p	2	2	6
Cytosolic NADP+-dependent isocitrate dehydrogenase	Glyma.14G211000.1.p	4	3	3
Ribosomal protein S4 (RPS4A) family protein	Glyma.16G127500.1.p	4	3	3
Aldehyde dehydrogenase 2B4	Glyma.18G137300.1.p	3	1	6
Protein prenylyltransferase superfamily protein	Glyma.19G243300.1.p	3	5	2

ATP synthase alpha/beta family protein	Glyma.20G123600.1.p	3	3	4
NmrA-like negative transcriptional regulator family protein	Glyma.01G172600.1.p	4	4	1
Alpha/beta-Hydrolases superfamily protein	Glyma.01G239500.1.p	4	1	4
Proteasome alpha subunit D2	Glyma.02G046400.1.p	2	4	3
Ribosomal protein L2 family	Glyma.02G049300.1.p	4	3	2
TCP-1/cpn60 Chaperonin family protein	Glyma.02G071900.1.p	4	4	1
heat shock cognate protein 70-1	Glyma.03G171100.1.p	2	2	5
26S proteasome regulatory complex, non-ATPase subcomplex, Rpn2/Psmd1 subunit	Glyma.05G241500.1.p	2	3	4
Purin-rich alpha 1	Glyma.07G225200.1.p	3	2	4
TCP-1/cpn60 chaperonin family protein	Glyma.08G175900.1.p	2	3	4
UDP-Glycosyltransferase superfamily protein	Glyma.08G181000.1.p	4	2	3
Annexin 8	Glyma.11G153800.1.p	3	2	4
TCP-1/cpn60 chaperonin family protein	Glyma.15G250500.1.p	2	3	4
Proteasome alpha subunit D2	Glyma.16G126200.1.p	2	4	3
Ribosomal protein L2 family	Glyma.16G129700.1.p	4	3	2
TCP-1/cpn60 chaperonin family protein	Glyma.16G153200.1.p	5	2	2
2-Oxoglutarate dehydrogenase, E1 component	Glyma.18G287400.1.p	3	2	4
Heat shock cognate protein 70-1	Glyma.19G172200.1.p	2	2	5
Purin-rich alpha 1	Glyma.20G025100.1.p	3	2	4
Pyruvate kinase family protein	Glyma.20G189300.1.p	2	1	6
26S Proteasome, regulatory subunit Rpn7;Proteasome component (PCI) domain	Glyma.05G044500.1.p	2	1	5
26S Proteasome, regulatory subunit Rpn7;Proteasome component (PCI) domain	Glyma.05G044900.1.p	2	1	5
Coatomer, beta-subunit	Glyma.07G034700.1.p	2	2	4
Phosphofructokinase 3	Glyma.07G126400.1.p	3	1	4
Mitochondrial HSO70 2	Glyma.07G184300.1.p	4	3	1
Cytochrome P450, family 93, subfamily D, polypeptide 1	Glyma.07G202300.1.p	3	4	1
S-Adenosyl-L-methionine-dependent methyltransferases superfamily protein	Glyma.07G214700.1.p	2	2	4
26S Proteasome regulatory complex, non-ATPase subcomplex, Rpn2/Psmd1 subunit	Glyma.08G048800.1.p	2	2	4
MLP-like protein 43	Glyma.08G230100.1.p	3	4	1
Ribosomal protein L13 family protein	Glyma.09G024000.1.p	4	1	3
Sterol methyltransferase 1	Glyma.10G090500.1.p	5	2	1
MOS4-Associated complex 3B	Glyma.10G133700.1.p	3	1	4
Voltage dependent anion channel 2	Glyma.11G048200.1.p	4	2	2

Cytochrome P450, family 93, subfamily D, polypeptide 1	Glyma.13G173500.1.p	3	4	1
Cinnamate-4-hydroxylase	Glyma.14G205200.1.p	5	1	2
Sterol methyltransferase 1	Glyma.15G276300.1.p	5	2	1
SKU5 Similar 4	Glyma.17G012300.1.p	3	2	3
Malate dehydrogenase	Glyma.17G100600.1.p	6	1	1
MOS4-Associated complex 3B	Glyma.20G084500.1.p	3	1	4
Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein	Glyma.02G002800.1.p	2	2	3
Ribophorin II (RPN2) family protein	Glyma.03G164400.1.p	2	2	3
Phosphofructokinase 3	Glyma.04G086900.1.p	3	1	3
Phosphoenolpyruvate carboxykinase 1	Glyma.04G089700.1.p	3	2	2
Phosphoglucomutase/phosphomannomutase family protein	Glyma.05G237000.1.p	4	1	2
Phosphofructokinase 3	Glyma.06G088600.1.p	3	1	3
Nascent polypeptide-associated complex subunit alpha-like protein 2	Glyma.06G137700.1.p	2	3	2
Thiamine pyrophosphate dependent pyruvate decarboxylase family protein	Glyma.07G153100.1.p	2	2	3
HAD superfamily, subfamily IIIB acid phosphatase	Glyma.08G200100.1.p	2	1	4
Phospholipase D alpha 1	Glyma.08G211700.1.p	2	3	2
Copper amine oxidase family protein	Glyma.08G282800.1.p	3	2	2
Phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent	Glyma.09G270500.1.p	2	1	4
Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein	Glyma.10G006200.1.p	2	2	3
Lysyl-tRNA synthetase 1	Glyma.10G069800.1.p	2	1	4
Ribosomal protein 1	Glyma.10G223400.1.p	3	3	1
Tubulin beta 8	Glyma.10G235100.1.p	3	1	3
Dolichyl-diphosphooligosaccharide-protein glycosyltransferase 48kDa subunit family protein	Glyma.11G120300.1.p	3	1	3
Ribosomal protein L22p/L17e family protein	Glyma.11G249700.1.p	4	1	2
NAD(P)-binding Rossmann-fold superfamily protein	Glyma.12G019800.1.p	2	2	3
Nucleolin like 2	Glyma.12G027400.1.p	3	2	2
Plasma membrane intrinsic protein 2	Glyma.12G172500.1.p	3	2	2
Ribosomal protein L4/L1 family	Glyma.13G136400.1.p	4	2	1
Plasma membrane intrinsic protein 2	Glyma.13G325900.1.p	3	2	2
Glucose-6-phosphate dehydrogenase 6	Glyma.16G063200.1.p	4	1	2
Selenium-binding protein 1	Glyma.18G002800.1.p	4	1	2
Ribosomal protein L22p/L17e family protein	Glyma.18G007500.1.p	4	1	2
Eukaryotic translation initiation factor 2 subunit 1	Glyma.18G024800.1.p	3	3	1

NADH-ubiquinone dehydrogenase, mitochondrial, putative	Glyma.18G041700.1.p	2	3	2
Glutamate decarboxylase 4	Glyma.18G043600.1.p	3	1	3
Phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent	Glyma.18G219100.1.p	2	1	4
D-3-Phosphoglycerate dehydrogenase	Glyma.20G129400.1.p	4	2	1
Tubulin beta 8	Glyma.20G159200.1.p	3	1	3
Ribosomal protein 1	Glyma.20G168100.1.p	3	3	1
Chalcone-flavanone isomerase family protein	Glyma.20G241600.1.p	3	2	2
2-Oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	Glyma.01G162700.3.p	2	3	1
Uncharacterized	Glyma.02G021600.1.p	3	1	2
Heat shock protein 89.1	Glyma.02G305600.1.p	2	1	3
Ribophorin I	Glyma.04G015300.1.p	3	1	2
Ribosomal L5P family protein	Glyma.04G111500.1.p	3	2	1
Acetyl Co-enzyme a carboxylase biotin carboxylase subunit	Glyma.05G221100.1.p	2	2	2
Ribophorin I	Glyma.06G015400.1.p	3	1	2
Gamma carbonic anhydrase 2	Glyma.06G030700.1.p	2	1	3
Eukaryotic translation initiation factor 4A1	Glyma.06G053200.1.p	3	1	2
3-Hydroxyacyl-CoA dehydrogenase family protein	Glyma.06G254700.1.p	2	2	2
Ribosomal L5P family protein	Glyma.06G274200.1.p	3	2	1
Ribosomal L5P family protein	Glyma.06G317400.2.p	3	2	1
Ribosomal L5P family protein	Glyma.06G323000.1.p	3	2	1
Eif4a-2	Glyma.07G007400.1.p	3	1	2
Laccase/Diphenol oxidase family protein	Glyma.07G142600.1.p	2	1	3
Multifunctional protein 2	Glyma.07G246300.1.p	2	2	2
Heat shock protein 70 (Hsp 70) family protein	Glyma.08G025700.1.p	2	2	2
Ribosomal protein L6 family protein	Glyma.08G141700.1.p	3	2	1
Eif4a-2	Glyma.08G190100.1.p	3	1	2
NADH-ubiquinone oxidoreductase-related	Glyma.09G129000.1.p	2	3	1
Villin 2	Glyma.10G099200.1.p	2	2	2
Membrane steroid binding protein 1	Glyma.10G237000.1.p	2	3	1
Peroxidase superfamily protein	Glyma.11G049600.1.p	3	1	2
Sulfite reductase	Glyma.11G093200.1.p	2	1	3
ADP/ATP carrier 3	Glyma.12G116300.1.p	2	2	2
Ribosomal L5P family protein	Glyma.12G131600.1.p	3	2	1

Plasma membrane intrinsic protein 1;4	Glyma.14G061500.1.p	3	2	1
Ribosomal L5P family protein	Glyma.15G227600.1.p	3	2	1
TCP-1/cpn60 chaperonin family protein	Glyma.16G208700.1.p	2	2	2
Class II aminoacyl-tRNA and biotin synthetases superfamily protein	Glyma.17G223600.1.p	4	1	1
Catalase 2	Glyma.17G261700.1.p	3	1	2
Ribosomal protein S6e	Glyma.18G151800.1.p	2	3	1
Laccase/Diphenol oxidase family protein	Glyma.18G193400.1.p	2	1	3
Phosphofructokinase family protein	Glyma.20G007400.1.p	2	2	2
2-Oxoglutarate dehydrogenase, E1 component	Glyma.20G099200.1.p	2	1	3
Pyruvate kinase family protein	Glyma.20G211200.1.p	2	2	2
Protein phosphatase 2A-2	Glyma.01G038800.1.p	2	1	2
Metallopeptidase M24 family protein	Glyma.01G119400.1.p	2	2	1
Protein phosphatase 2A-2	Glyma.02G025900.1.p	2	1	2
Ribosomal protein 5B	Glyma.02G167900.1.p	2	1	2
Ribosomal protein L16p/L10e family protein	Glyma.02G220000.1.p	2	2	1
Uncharacterised protein family (UPF0172)	Glyma.02G301900.1.p	2	2	1
Tubulin alpha-2 chain	Glyma.04G088500.1.p	2	2	1
Ribosomal protein L16p/L10e family protein	Glyma.04G224000.1.p	2	2	1
HSP20-like chaperones superfamily protein	Glyma.04G236300.1.p	3	1	1
Tubulin alpha-4 chain	Glyma.05G110200.1.p	2	2	1
Heat shock protein 70 (Hsp 70) family protein	Glyma.05G219400.1.p	3	1	1
Receptor for activated C kinase 1C	Glyma.05G243800.1.p	2	2	1
Tubulin alpha-2 chain	Glyma.06G090500.1.p	2	2	1
HSP20-like chaperones superfamily protein	Glyma.06G128000.1.p	3	1	1
Ribosomal protein L16p/L10e family protein	Glyma.06G140900.1.p	2	2	1
Translation elongation factor EF1B/ribosomal protein S6 family protein	Glyma.06G170900.1.p	3	1	1
SecY protein transport family protein	Glyma.06G227200.1.p	2	2	1
Ribosomal protein L5 B	Glyma.07G059900.1.p	2	1	2
Receptor for activated C kinase 1C	Glyma.08G051600.1.p	2	2	1
Ribosomal protein L30/L7 family protein	Glyma.08G222100.1.p	2	2	1
Protein phosphatase 2A-2	Glyma.08G293400.1.p	2	1	2
RAB GTPase homolog B1C	Glyma.09G016700.1.p	2	1	2
Ribosomal protein 5B	Glyma.09G103400.1.p	2	1	2

Thioredoxin family protein	Glyma.10G014700.1.p	2	1	2
Tubulin alpha-2 chain	Glyma.10G255500.1.p	2	2	1
SecY protein transport family protein	Glyma.12G156800.1.p	2	2	1
Ribosomal protein 5B	Glyma.12G201900.1.p	2	1	2
SecY protein transport family protein	Glyma.12G230900.1.p	2	2	1
SecY protein transport family protein	Glyma.13G268600.1.p	2	2	1
Ribosomal protein 5B	Glyma.13G300300.1.p	2	1	2
Ribosomal protein L16p/L10e family protein	Glyma.14G187800.1.p	2	2	1
Phosphofructokinase 3	Glyma.15G020900.1.p	2	1	2
RAB GTPase homolog B1C	Glyma.15G122300.1.p	2	1	2
Plasma-membrane associated cation-binding protein 1	Glyma.15G154700.1.p	2	2	1
Tubulin alpha-4 chain	Glyma.17G156900.1.p	2	2	1
Eukaryotic translation initiation factor 2	Glyma.18G029500.1.p	2	2	1
Ribosomal protein L16p/L10e family protein	Glyma.18G050600.1.p	2	2	1
Heat shock protein 81.4	Glyma.18G074100.1.p	2	2	1
Protein phosphatase 2A-2	Glyma.18G129700.1.p	2	1	2
Pyruvate decarboxylase-2	Glyma.18G204200.1.p	2	2	1
Casein lytic proteinase B3	Glyma.19G027900.1.p	2	1	2
Tubulin alpha-4 chain	Glyma.20G136000.1.p	2	2	1
Plasma membrane intrinsic protein 1;4	Glyma.02G255000.1.p	2	1	1
Dihydrolipoamide succinyltransferase	Glyma.02G292500.1.p	2	1	1
Ribosomal protein S8e family protein	Glyma.03G086400.1.p	2	1	1
Catalase 2	Glyma.04G017500.1.p	2	1	1
RAN GTPase 3	Glyma.04G069300.1.p	2	1	1
Tubulin alpha-5	Glyma.05G157300.1.p	2	1	1
T-complex protein 1 alpha subunit	Glyma.05G165500.1.p	2	1	1
Ribosomal protein L6 family protein	Glyma.05G184100.1.p	2	1	1
Catalase 2	Glyma.06G017900.1.p	2	1	1
PDI-like 5-4	Glyma.06G270200.1.p	2	1	1
Coatomer, beta subunit	Glyma.07G004100.1.p	2	1	1
Tubulin alpha-5	Glyma.08G115100.1.p	2	1	1
Mitochondrial substrate carrier family protein	Glyma.08G115600.1.p	2	1	1
T-complex protein 1 alpha subunit	Glyma.08G123100.1.p	2	1	1

General regulatory factor 7	Glyma.08G363800.1.p	2	1	1
Proteasome component (PCI) domain protein	Glyma.10G140700.1.p	2	1	1
Ribosomal protein S13/S18 family	Glyma.10G221600.1.p	2	1	1
Ribosomal protein S13/S18 family	Glyma.10G224200.1.p	2	1	1
NmrA-like negative transcriptional regulator family protein	Glyma.11G070600.1.p	2	1	1
Translocon-associated protein (TRAP), alpha subunit	Glyma.11G098300.1.p	2	1	1
Sulfite reductase	Glyma.12G019400.1.p	2	1	1
Translocon-associated protein (TRAP), alpha subunit	Glyma.12G024500.1.p	2	1	1
PDI-like 5-4	Glyma.12G134500.1.p	2	1	1
Arogenate dehydratase 6	Glyma.12G181800.1.p	2	1	1
Ribosomal protein L1p/L10e family	Glyma.12G184900.1.p	2	1	1
R-protein L3 B	Glyma.12G240600.1.p	2	1	1
Proteasome component (PCI) domain protein	Glyma.13G055200.1.p	2	1	1
Eukaryotic translation initiation factor 3 subunit 7 (eIF-3)	Glyma.13G238200.1.p	2	1	1
Arogenate dehydratase 6	Glyma.13G319000.1.p	2	1	1
Dihydrolipoamide succinyltransferase	Glyma.14G022100.1.p	2	1	1
Eukaryotic aspartyl protease family protein	Glyma.17G016600.1.p	2	1	1
General regulatory factor 7	Glyma.18G298300.1.p	2	1	1
Proteasome component (PCI) domain protein	Glyma.20G089400.1.p	2	1	1
S18 ribosomal protein	Glyma.20G167600.1.p	2	1	1
Ribosomal protein S13/S18 family	Glyma.20G169900.1.p	2	1	1

CURRICULUM VITAE

Mehran Dastmalchi

Education

2010-2014 **Ph.D. Biology** (Cell and Molecular Biology)
University of Western Ontario
London, Ontario, Canada

2005-2009 **H.B.Sc. Biology**
University of Toronto
Toronto, Ontario, Canada

Major Awards and Scholarships

2014 University of Western Ontario, Biology Department, London, ON, Canada
Western Graduate Research Scholarship

2014 Plant Metabolism, Banff, AB, Canada
Best Oral Presentation Award

2014 University of Western Ontario, Biology Department, London, ON, Canada
Travel Award

2013 Phytochemical Society of North America, Corvallis, OR, USA
Travel Award

2012 Canadian Society of Plant Biology, Edmonton, AB, Canada
George H. Duff Student & Post-Doc Travel Bursary

Research Experience

2010-2014 Agriculture and Agri-Food Canada, London, ON, Canada
(Dhaubhadel Lab)
Graduate Research Student

2008-2009 University of Toronto, Toronto, ON, Canada
(McCourt Lab)
Undergraduate Research Volunteer

Teaching Experience

2014 Third year undergraduate "Advanced Genetics" *Teaching Assistant*

2014-2013 Second year undergraduate "Plants as a Human Resource" *Teaching Assistant*

2013-2012	Second year undergraduate "Plants as a Human Resource" *Teaching Assistant*
2012-2013	Fourth year undergraduate thesis student *PhD Supervisor*
2011-2012	Third year undergraduate "Cell Biology" *Teaching Assistant*
2011-2012	Second year undergraduate "Scientific Writing" *Teaching Assistant*
2010-2011	First year undergraduate "Biology" *Lab Instructor*

Refereed Publications

Dastmalchi M and Dhaubhadel S (2014) Soybean seed isoflavonoids: biosynthesis and regulation, *Recent Advances in Phytochemistry* 44, Ed. Reinhard Jetter

Dastmalchi M and Dhaubhadel S (2014) Soybean chalcone isomerase: evolution of the fold, and the differential expression and localization of the gene family, *Planta* [Epub ahead of print]

Accepted: **Dastmalchi M** and Dhaubhadel S (2014) Proteomic insights into synthesis of isoflavonoids in soybean seeds, *Proteomics*

Oral Presentations

Dastmalchi M and Dhaubhadel S (2014) Soybean isoflavonoid biosynthesis – constituents and circumstance in the metabolon, Plant Metabolism, Banff, AB, Canada (Awarded 'Best Oral Presentation')

Dastmalchi M and Dhaubhadel S (2013) *Chalcone isomerase* and the soybean isoflavonoid biosynthesis machinery – a genetic and functional characterization, Phytochemical Society of North America, Corvallis, OR, USA

Dastmalchi M and Dhaubhadel S (2012) Soybean *chalcone isomerase* gene family: characterization and role in isoflavonoid biosynthesis, Phytochemical Society of North America, London, ON, Canada

Dastmalchi M and Dhaubhadel S (2012) Soybean *chalcone isomerase* gene family and their role in isoflavonoid biosynthesis, Canadian Society of Plant Biologists, Edmonton, AB, Canada

Dastmalchi M and Dhaubhadel S (2011) Identification and characterization of members of the *chalcone isomerase* gene family involved in isoflavonoid biosynthesis of *Glycine max,* Canadian Plant Genomics Workshop, Niagara, ON, Canada

Dastmalchi M and Dhaubhadel S (2010) Identification and characterization of members of the *chalcone isomerase* gene family involved in isoflavonoid biosynthesis of *Glycine max,* Canadian Society of Plant Physiologists, Eastern Regional Meeting, St. Catharines, ON, Canada

Poster Presentations

Dastmalchi M and Dhaubhadel S (2014) Soybean *chalcone isomerase*: evolutionary history of the fold as told by sequence, expression and localization of the gene family, International Conference on Legume Genetics and Genomics, Saskatoon, SK, Canada

Dastmalchi M and Dhaubhadel S (2011) identification and characterization of members of the *chalcone isomerase* gene family involved in isoflavonoid biosynthesis of *Glycine max,* Canadian Plant Genomics Workshop, Niagara, ON, Canada