July 2013

Identification of Cyclophilin gene family in soybean and characterization of GmCYP1

Hemanta Raj Mainali
The University of Western Ontario

Supervisor
Dr. Sangeeta Dhaubhadel
The University of Western Ontario

Graduate Program in Biology

A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

© Hemanta Raj Mainali 2013

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Agricultural Science Commons, Biochemistry Commons, Bioinformatics Commons, Biology Commons, Biotechnology Commons, Molecular Biology Commons, Plant Biology Commons, and the Plant Breeding and Genetics Commons

Recommended Citation
https://ir.lib.uwo.ca/etd/1362

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca, wlswadmin@uwo.ca.
Identification of *Cyclophilin* gene family in soybean and
characterization of *GmCYP1*

(Thesis format: Monograph)

by

**Hemanta Raj Mainali**

Graduate Program in Biology

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

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

© Hemanta Raj Mainali 2013
Abstract

I identified members of the Cyclophilin (CYP) gene family in soybean (Glycine max) and characterized the GmCYP1, one of the members of soybean CYP. CYPs belong to the immunophilin superfamily with peptidyl-prolyl cis-trans isomerase (PPIase) activity. PPIase catalyzes the interconversion of the cis- and trans-rotamers of the peptidyl-prolyl amide bond of peptides. After extensive data mining, I identified 62 different CYP genes in soybean (GmCYP1 to GmCYP62), of which 8 are multi-domain proteins and 54 are single domain proteins. At least 25% of the GmCYP genes are expressed in soybean. GmCYP1 localizes to the nucleus and the cytoplasm and it interacts with an R1 MYB transcription factor, GmMYB176, in the nucleus. Furthermore, GmCYP1 forms a dimer in planta. A loss-of-function mutation of the Arabidopsis ortholog of GmCYP1 (ROC1, ROC3 and ROC5) showed sensitivity towards salt stress during the seed germination, but did not respond differently than wild type when salt stress was applied to mature plants. This suggests a germination specific role of ROC1, ROC3 and ROC5 in salt stress in Arabidopsis. Moreover, roc3 and roc5 showed early flowering phenotypes under long day condition but not under short day, suggesting a role in photoperiod specific transition from vegetative to reproductive stage in Arabidopsis. Overall, these results propose that GmCYP1 functions in response to salt stress and flowering in soybean.

Key Words
Cyclophilin, soybean, peptidyl-prolyl cis-trans isomerase, transcription factor, salt stress, germination, flowering,
Acknowledgments

This research project would not have been possible without the support of many people. First and foremost, I would like to sincerely thank my supervisor, Dr. Sangeeta Dhaubhadel for her invaluable assistance, support, and guidance throughout the tenure of my work. She always encouraged me to plan, and conduct experiments independently and has been a constant inspiration. Her mentorship has hugely contributed to my academic and personal development.

I would like to thank my co-supervisor Dr. Norman PA Huner and my advisory committee members Dr. Denis Maxwell and Dr. Katherine Dobinson for their helpful advice and suggestions during my graduate studies.

I am grateful to all the previous and current members of Dr. Dhaubhadel Lab, Dr. Xuyan Li, Dr. Shaomin Bian, Ling Chen, Mehran Dastmalchi, Arun Kumaran Anguraj Vadivel, Kishor Duwadi, and Tanya Mahbubani for creating a wonderful working environment, providing support and critical comments. The past two years have been an unparalleled experience and pleasure. I want to acknowledge Dr. Shaomin Bian for teaching me how to work with Arabidopsis in the early days. I would like to thank Alex Molnar for his help in compiling images and Lisa Amyot for confocal training. Lastly, I offer my regards to all of those who supported me in any respect during the completion of my project.

Finally, I want to express my deepest love and gratitude to my beloved parents for their understanding and unconditional love.
# Table of Contents

Abstract ........................................................................................................................................ ii

Acknowledgments ....................................................................................................................... iii

Table of Contents ......................................................................................................................... iv

List of Tables ................................................................................................................................. vii

List of Figures ............................................................................................................................... viii

List of Abbreviations ..................................................................................................................... x

Chapter 1: Introduction ................................................................................................................ 1

1.1 The Peptide bond .................................................................................................................... 1
1.2 Peptidyl prolyl cis/trans isomerase ....................................................................................... 3
1.3 CYPs ...................................................................................................................................... 4
1.4 Human CYP-A ......................................................................................................................... 5
   1.4.1 Structure of hCYP-A ........................................................................................................ 6
1.5 Plant CYPs ............................................................................................................................. 6
1.6 Plant CYPs are involved in stress response ............................................................................. 8
1.7 Arabidopsis and rice CYP gene family .................................................................................... 10
1.8 Soybean CYPs ....................................................................................................................... 11
1.9 Hypothesis ............................................................................................................................... 13
1.10 Thesis objectives ................................................................................................................... 14

Chapter 2: Materials and Methods .......................................................................................... 15

2.1 In silico and phylogenetic analysis ......................................................................................... 15
2.2 Plant materials and growth conditions ................................................................................ 16
2.3 Bacterial strains .................................................................................................................. 17
2.4 Genomic DNA isolation from Arabidopsis ................................................................. 18
2.5 Polymerase Chain Reaction (PCR) condition .............................................................. 18
2.6 Genotyping of T-DNA insertion mutant lines .............................................................. 19
2.7 RNA extraction and Reverse-Transcriptase (RT) PCR analysis ............................ 19
2.8 Cloning .............................................................................................................................. 22
  2.8.1 Cloning of GmCYP1 in Gateway entry vector ...................................................... 22
  2.8.2 Plasmid construction for subcellular localization ............................................. 23
  2.8.3 Cloning for bimolecular fluorescent complementation assay ............................. 28
2.9 Transient expression of protein in N. benthamiana leaves ........................................ 29
2.10 Confocal microscopy .................................................................................................... 30
2.11 Arabidopsis seed germination assay ......................................................................... 30
2.12 Salt stress treatment on plants ................................................................................... 31
2.13 Statistical analysis ........................................................................................................ 31

Chapter 3: Results .................................................................................................................. 32
  3.1 Soybean genome contains 62 putative GmCYPs ...................................................... 32
  3.2 Chromosome location and phylogenetic analysis of soybean CYP gene family ... 35
  3.3 PPIase and CsA binding sites are conserved in GmCYP1 ......................................... 39
  3.4 GmCYP1 interacts with an R1 MYB transcription factor, GmMYB176 ............... 42
  3.5 GmCYP1 forms homodimers in planta ..................................................................... 45
  3.6 GmCYP1 localizes in the nucleus and cytoplasm .................................................... 45
  3.7 GmCYP1 is expressed ubiquitously in all soybean tissues ...................................... 49
  3.8 GmCYP1 orthologs in Arabidopsis have a role in flowering pathway ................. 52
  3.9 GmCYP1 orthologs in Arabidopsis play a role in germination during salt stress ... 59
Chapter 4: Discussion ........................................................................................................................................ 65

4.1 At least 15 GmCYPs are transcribed in soybean ................................................................. 69
4.2 GmCYP1 localizes in nucleus and cytoplasm ............................................................... 70
4.3 GmMYB176 and GmCYP1 interact in planta .............................................................. 71
4.4 GmCYP1 functions in seed development ................................................................... 72
4.5 GmCYP1 has conserved PPIase and CsA binding site .............................................. 73
4.6 Functional characterization of GmCYP1 in Arabidopsis ............................................. 73
   4.6.1 ROC1, ROC3 and ROC5 have a role in seed germination in salt stress .......... 74
   4.6.2 roc3 and roc5 shows early flowering under long day photoperiods ............ 76

Chapter 5: Perspectives and Future Work .................................................................................. 79

References ................................................................................................................................................ 81

Appendix .................................................................................................................................................. 91

Appendix 1 Soybean cyclophilin gene family ....................................................................... 91

CURRICULUM VITAE ......................................................................................................................... 95
List of Tables

Table 2.1 List of primers and their sequences used for genotyping ............................ 20
Table 2.2 List of primers used to make nuclear, and ER localization construct.............. 26
Table 3.1 Soybean CYP gene family. ........................................................................... 33
Table 3.2 Comparison of GmCYP1 with its closely related CYPs from other plant
species, animal, and yeast .......................................................................................... 40
Table 3.3 Arabidopsis orthologs of GmCYP1 and the corresponding T-DNA insertion
mutant lines used in this study .................................................................................. 53
List of Figures

Figure 1.1 Schematic representation of peptide bond and its cis/trans conformation....... 2
Figure 1.2 Structure of human cyclophilin A (hCYP-A)........................................ 7
Figure 2.1 Maps for nuclear localization and ER localization constructs............. 24
Figure 3.1 Schematic representation of different types of domains present in multi-
domain GmCYPs. ........................................................................................................ 34
Figure 3.2 Genomic distributions of GmCYP genes on soybean chromosomes........ 36
Figure 3.3 Phylogenetic tree of GmCYPs constructed using Neighbour-Joining method 38
Figure 3.4 Multiple sequence alignment of deduced amino acid sequence of GmCYP1
with closely related CYPs from other species.......................................................... 41
Figure 3.5 Confocal images showing the interaction of GmCYP1 and GmMYB176 in
planta......................................................................................................................... 43
Figure 3.6 Confocal Image showing GmCYP1 homodimers in planta....................... 46
Figure 3.7 Subcellular localization of GmCYP1......................................................... 47
Figure 3.8 Quantitative expression analysis of GmCYP1 in different soybean tissues .... 50
Figure 3.9 Expression profiles of Arabidopsis cyclophilin ROC1 in different tissues and
developmental stages ............................................................................................... 51
Figure 3.10 T-DNA insertion map and gel images showing homozygous T-DNA
insertion in mutants................................................................................................. 54
Figure 3.11 T-DNA insertion in ROC1, ROC3 and ROC5 genes alter their corresponding
transcript levels........................................................................................................ 56
Figure 3.12 Flowering day phenotype of mutants (roc3 and roc5) under long and short
day light conditions............................................................................................... 57
Figure 3.13 Photograph showing the early flowering phenotype of \textit{roc3} and \textit{roc5} under long day condition................................................................. 58

Figure 3.14 Mutation in GmCYP1 orthologs in \textit{Arabidopsis} impair seed germination in response to salt treatment........................................................................................................ 60

Figure 3.15 Photographs showing the germination of WT and mutant (\textit{roc1}, \textit{roc3}, and \textit{roc5}) seeds in salt stress........................................................................................................ 61

Figure 3.16 Photographs of the WT and mutants (\textit{roc1}, \textit{roc3}, and \textit{roc5}) after salt stress treatment ........................................................................................................ 64
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>μl</td>
<td>microlitre</td>
</tr>
<tr>
<td>μg</td>
<td>microgram</td>
</tr>
<tr>
<td>°C</td>
<td>degree celsius</td>
</tr>
<tr>
<td>ABA</td>
<td>abscisic acid</td>
</tr>
<tr>
<td>BiFC</td>
<td>bimolecular fluorescent complementation</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BP</td>
<td>border primer</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CFP</td>
<td>cyan fluorescent protein</td>
</tr>
<tr>
<td>CLD</td>
<td>cyclophilin-like domain</td>
</tr>
<tr>
<td>CsA</td>
<td>cyclosporine A</td>
</tr>
<tr>
<td>CTAB</td>
<td>cetyl-trimethylammonium bromide</td>
</tr>
<tr>
<td>CYP</td>
<td>cyclophilin</td>
</tr>
<tr>
<td>DAP</td>
<td>days after pollination</td>
</tr>
<tr>
<td>DFCI</td>
<td>Dana-Farber cancer institute</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxyribonucleotide triphosphate</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>ER-LS</td>
<td>ER localization signal</td>
</tr>
</tbody>
</table>
EST  expressed sequence tag
FKBP  FK506 binding protein
FMI  floral meristem identity
GA  gibberellic acid
Kb  kilobase
LP  left primer
Mb  megabase
mm  millimetre
MS  Murashige and Skoog
NLS  nuclear localization signal
nm  nanometre
PPIase  peptidyl prolyl cis/trans isomerase
RNA  ribonucleic acid
RP  right primer
RRM  RNA recognition motif
RT-PCR  reverse transcription-polymerase chain reaction
SDS  sodium dodecyl sulfate
TAE  tris-acetate-EDTA
TBE  tris-borate-EDTA
TC  tentative contig
TTR  tetra-triacopeptide repeats
w/v weight per volume
WD tryptophan-aspartate
WT wild type
Y2H yeast-two-hybrid
YFP yellow fluorescent protein
ZK zinc knuckle
Chapter 1: Introduction

1.1 The Peptide bond

The peptide bond is a covalent bond formed as a result of the condensation reaction between the carboxyl terminal of one amino acid and the amino terminal of another amino acid, with the release of a water molecule (Figure 1.1A). The peptide bond acts as a connector for different amino acids during the synthesis of polypeptides (Nelson and Cox, 2004).

Peptide bond is partial double bond in nature that makes the amide group in planar conformation and hinders the rotation of amide group through the amide bond (Ramachandran and Sasisekharan, 1968). Because of this restriction in bond rotation, peptide bonds exist in two distinct stereo forms: cis- and trans- isomers. Cis isomers have side chains of two amino acids adjacent to each other and are sterically not favored whereas trans isomers have side chains of adjacent amino acids in 180° to each other which makes them thermodynamically and sterically more favorable (Figure 1.1B).

However, for peptide bonds involving a proline, the scenario is different. Because of the peptidyl prolyl imide bond in proline, the difference in free energy between cis and trans isomers is smaller compared to other amino acids (reviewed in Lu et al., 2007). In total, cis conformation occurs in 6.5% of proline containing peptide bonds (X-pro peptide bond) but for the peptide bond without proline, it is only 0.05% (Stewart et al., 1990). This is because the proline amino acid has cyclic side chain with no alpha hydrogen to make hydrogen-bonds in alpha helices and beta
Figure 1.1 Schematic representation of peptide bond and its cis/trans conformation. 

(A) The peptide bond, connecting two amino acids, is highlighted with light purple color. R1 and R2 represent side chains of amino acids. (B) The left side of each structure represents a proline amino acid and the other two amino acids are represented with P₁ and P₂. The interconversion between cis/trans is catalyzed by PPlase. The diagram is adapted from Wang and Heitman (2005).
sheets, which makes it unsuitable for alpha helix region. In contrast, cis rotamer fits well in turn, bend, and coil region during protein folding. This interconversion of cis/trans isomerization in peptidyl prolyl bond is performed by an enzyme called peptidyl prolyl cis/trans isomerase (PPIase).

1.2 Peptidyl prolyl cis/trans isomerase

Peptidyl prolyl cis/trans isomerases (PPIases) are a class of enzymes that help in protein folding by interconverting the cis- and trans-rotamers of the peptidyl-prolyl amide bond of peptides (Figure 1.1). They were isolated for the first time in 1984 from porcine kidney cortex by Fisher et al. (1984).

Protein folding is the physical process by which proteins achieve their functional shape and conformation. During protein synthesis, a linear chain of amino acid is produced from messenger RNA which undergoes folding into a three-dimensional structure. The correct three dimensional structure of the protein is essential for its function. The process of protein folding often starts with translation where the N-terminus of the protein starts folding while the C-terminus is being synthesized. During this process, specialized molecules called chaperones assist in folding the protein in correct three dimensional structure (Lee and Tsai, 2005). Along with that, PPIase assists in protein folding by isomerization of peptidyl prolyl bond. Compared to other steps, isomerization of peptidyl prolyl bond is considered as the rate limiting step in protein folding (Brandts et al., 1975). PPIase activity is especially important during stress conditions where a change in physiological condition denatures the correctly folded protein. It is essential for the protein to be in correct three-dimensional state for its
function, PPIases are involved in the salvage of such misfolded proteins during the stress conditions.

PPIase are broadly classified into three major classes: parvulins, FK506 binding proteins (FKBP) and cyclophilins (CYPs) (Dilworth et al., 2012). Recently, a new class of PPIase has been reported as protein phosphatase 2A phosphatase activator (Jordens et al., 2006). All PPIase classes are found in plants, animals, and microorganisms, and are localized in all major cellular compartments (Pemberton and Kay, 2005). This suggests that they have a critical role in cellular processes from bacteria to higher organisms.

Parvulins are small proteins of 92 amino acids and were reported originally in *Escherichia coli* (Rahfeld et al., 1994). They are found ubiquitously in both prokaryotes and eukaryotes. Unlike other classes of PPIase, parvulins are not characterized by their ability to bind immunosuppressive drugs. In addition, they do not share any sequence identity with other PPIases, but share a common catalytic domain of central β-sheet (reviewed in Lu et al., 2007). The other two classes of PPIases, FKBP and CYPs are collectively called immunophilins as they were originally identified as a receptor for immunosuppressive drugs FK506 and cyclosporine A (CsA), respectively (Harding et al., 1989; Siekierka et al., 1989; Takahashi et al., 1989).

### 1.3 CYPs

CYPs are ubiquitous proteins that are found in organisms ranging from archaea, bacteria to plants and animals (Galat, 1999; Maruyama et al., 2004). As they were originally identified as a receptor for the immunosuppressive drug CsA, they are categorized in immunophilin family of proteins which possesses peptidyl prolyl cis/trans isomerase activity.
CYPs are structurally conserved throughout evolution. Multiple CYPs have been found in the genomes of various prokaryotes, but only a few of them have been studied in detail. The *E. coli* genome encodes two CYPs, a cytosolic form (EcCYP-18) and its periplasmic counterpart (EcCYP-20) (Hayano et al., 1991). In *Saccharomyces cerevisiae*, there are at least 8 different CYPs, *Cpr1* to *Cpr8* (Arevalo-Rodriguez et al., 2004). *Cpr1* has an important role in heat shock survival (reviewed in Arevalo-Rodriguez et al., 2004) and its expression is also induced under salt stress.

The human genome encodes 16 unique CYPs which are categorized into 7 major groups *viz* human CYP A (hCYP-A), hCYP-B, hCYP-C, hCYP-D, hCYP-E, hCYP-40 and hCYP-NK (Galat, 2003). The hCYP-A binds CsA, and forms a ternary complex with calceneurin. CsA-hCYP-A binding to calceneurin inhibits the phosphatase activity of calneceurin. As a result, the transcription factor, nuclear factor of activated T-cells (NFAT), remains inactive in the cytoplasm and the interleukin-2 gene is not transcribed, leading to the inactivation of T-cells (Liu et al., 1991b).

### 1.4 Human CYP-A

Extensive studies have been carried out with hCYP-A, the first CYP to be identified. hCYP-A plays a crucial role in the protein folding process as it possesses both catalytic and chaperone-like activities. For example, hCYP-A promotes the formation and infectivity of human immunodeficiency virus (HIV-1) virions (Braaten and Luban, 2001). hCYP-A is incorporated into HIV-1 virions where it interacts with the HIV-1 Gag protein, a polyprotein precursor of virion structural protein. The presence of four conserved proline residues in HIV-1 is crucial for incorporation of hCYP-A into virions (Franke et al., 1994; Thali et al., 1994). Similarly, other human CYPs are found to be
associated with different signal transduction pathways, cell signaling pathways, regulation of gene expression, and immune response (reviewed in Lu et al., 2007).

1.4.1 Structure of hCYP-A

The hCYP-A consists of eight stranded anti-parallel β-barrel with two α-helices surrounding the barrel (Kallen et al., 1991) (Figure 1.2). Study using site directed mutagenesis of W_{121}A (replacing tryptophan-121 with alanine) did not inhibit phosphatase activity of calcineurin, suggesting its critical role in CsA binding, with no change in PPIase activity (Liu et al., 1991a; Zydowsky et al., 1992). In contrast, mutation in Arg-55, Phe-60 and His-126 inhibited PPIase activity by 99% (Zydowsky et al., 1992). The exact mechanism of bond rotation by CYP during catalysis is not yet known. However, it is believed that the conserved arginine in the hydrophobic active site of hCYP-A makes a hydrogen bond with the peptide nitrogen, resulting in a peptidyl-prolyl bond with more single-bond character.

1.5 Plant CYPs

Compared to human CYPs, very little is known about plant CYPs. The first plant CYPs were identified concurrently from tomato (Lycopersicon esculentum), maize (Zea mays), and oilseed rape (Brassica napus) (Gasser et al., 1990). Recently, with the availability of whole genome sequencing, the identification and characterization of plant CYPs has progressed significantly.
Figure 1.2 Structure of human cyclophilin A (hCYP-A). hCYP-A has eight stranded antiparallel beta barrel with 2 alpha helices surrounding the barrel. The structure was downloaded from Protein Data Bank (http://www.rcsb.org/pdb/images/2cpl_asr_r_500.jpg).
However, compared to other organisms, the total number of plant CYPs in databases is still small, which suggests that many plant CYPs remain to be identified (Opiyo and Moriyama, 2009). So far, Arabidopsis thaliana and rice (Oryza sativa) are the two plant species reported to have highest number of CYPs with 35 AtCYPs (Romano et al., 2004; Trivedi et al., 2012) and 28 OsCYPs (Ahn et al., 2010; Trivedi et al., 2012), respectively.

1.6 Plant CYPs are involved in stress response

Because they are sessile, plants have to be able to endure a wide range of both biotic and abiotic stresses to complete their life cycle. Abiotic stresses arise from excess or deficit of physical or chemical factors including water-logging, drought, temperature changes, excessive soil salinity, and changes in light quality and quantity. Biotic stresses are due to the attack by pathogenic organisms like viruses, bacteria, fungi, and nematodes.

In order to combat biotic and abiotic stressors, plants have developed several sophisticated mechanisms at the cellular and molecular level (reviewed in Atkinson and Urwin, 2012; Fujita et al., 2006). Perception of stress by the stress receptors initiates the signal transduction and activation of hormonal signaling pathways, which then activate the different transcription factors needed for expression of the downstream response genes. There exist different pathways to combat different types of biotic and abiotic stresses. One consequence of many abiotic stress conditions is denaturation of cellular proteins, which can result in cell death. Under such conditions, special molecules called chaperones stabilize aggregated protein molecules inside the cells, and protect them from degradation. The chaperone-like activity of CYPs and their role in the rate limiting step
of protein folding by peptidyl prolyl bond isomerization (Brandts et al., 1975) is
associated with their involvement in different types of stress responses.

Expression of many plant CYPs is induced in response to several different types
of abiotic stresses suggesting their possible function in stress tolerance. For example,
expression of the Arabidopsis CYP, ROTAMASE CYCLOPHILIN 1 (ROC1), increases
upon wounding (Chou and Gasser, 1997). Similarly, transcription of maize and bean
CYPs increase under conditions of heat-shock, wounding, high salinity, or low
temperature (Marivet et al., 1992). Moreover, Solanum commersonii CYP gene
expression is up-regulated by low temperature, abscisic acid, drought, or wounding
(Meza-Zepeda et al., 1998). Pepper CYPs are differentially regulated during abiotic stress
or pathogen infection (Kong et al., 2001).

Support for their role in stress response has also come from overexpression
studies. Expression of Thellungiella halophile CYP, ThCYP1, is induced by salt, heat,
abscisic acid, and hydrogen peroxide and ectopic expression of ThCYP1 in fission yeast
and tobacco cells increases the tolerance to salt stress (Chen et al., 2007). Transgenic
Arabidopsis expressing pigeon-pea CYP (CcCYP1) increases tolerance against drought,
salinity and high temperatures, with enhanced PPIase activity under stressed conditions
(Sekhar et al., 2010). Similarly, overexpression of cotton CYP (GhCYP1) in transgenic
tobacco plants confers tolerance against salt stress and fire-blight disease (Zhu et al.,
2011).
There are 35 CYPs reported in *Arabidopsis*, but only 15 of them are characterized (Gullerova et al., 2006; Li and Luan, 2011; Romano et al., 2004). They are found in cytoplasm (Berardini et al., 2001; Chou and Gasser, 1997; Hayman and Miernyk, 1994), endoplasmic reticulum (ER) (Grebe et al., 2000; Jackson and Soll, 1999), chloroplast (Lippuner et al., 1994; Schubert et al., 2002), and nucleus (Gullerova et al., 2006). Among the identified *AtCYPs*, only some are characterized at the molecular level. Increase in the expression of *ROC1* in response to light is found to be associated with phytochromes and cryptochromes (Chou and Gasser, 1997; Trupkin et al., 2012). *roc1* mutants with a T-DNA insertion in the promoter region display an early flowering phenotype under long day, but not under short day photoperiod (Trupkin et al., 2012). However, loss-of-function mutants of *AtCYP40* reduce the number of juvenile leaves, with no change in inflorescence morphology and flowering time (Berardini et al., 2001). Similarly, gain-of-function mutations in *ROC1* reduce stem elongation and increase shoot branching (Ma et al., 2013). Moreover, *AtCYP59*, a multi-domain CYP with a RNA recognition motif (RRM), regulates transcription and pre-mRNA processing by binding to the C-terminal domain of RNA polymerase II (Gullerova et al., 2006). The *Arabidopsis* plants with a defective *AtCYP20-3* are found to be hypersensitive to oxidative stress conditions created by high light levels, high salt levels, and osmotic shock (Dominguez-Solis et al., 2008). Collectively, these results show the roles of *Arabidopsis* CYPs in different cellular pathways, which necessitate further work to explore the functionality associated with each of the CYPs.
Rice is a model plant representing monocots. The completion of the rice genome sequencing project in 2005 (International Rice Genome Sequencing Project, 2005), and the availability of the sequences in the public domain has facilitated the ability of researchers to carry out genome wide analysis of the gene families. Genome wide analysis of the rice genome revealed 28 OsCYPs distributed on 10 different chromosomes (Ahn et al., 2010; Trivedi et al., 2012). Of the 28 OsCYPs, 15 are predicted to be localized to the cytoplasm, 5 to chloroplast, 3 to mitochondria, and 5 are secretory.

Compared to the Arabidopsis CYPs, little work has been done on the rice CYPs. However, recent studies on rice CYPs show their roles in different types of stresses. OsCYP2 has been reported to have role in different abiotic stress responses (Ruan et al., 2011). The expression of OsCYP2 is up-regulated during salt stress, and its over-expression in rice enhances tolerance towards the salt stress. Similarly, overexpression of the thylakoid-localized OsCYP20-2 in Arabidopsis and tobacco provides increased tolerance towards osmotic stress, and to extremely high light conditions (Kim et al., 2012). Moreover, there are several other OsCYPs found to be up-regulated by abiotic stresses like desiccation and salt stress (Ahn et al., 2010; Trivedi et al., 2012), indicating a critical role of OsCYPs during stress conditions.

1.8 Soybean CYPs

Soybean (Glycine max [L.] Merr) is a legume plant belonging to the Papilionoideae. They are one of the most important crops for their high nutrient content and ability to fix atmospheric nitrogen through symbiosis. Soybean seeds consist of 40% protein, 21% oil and 11% soluble carbohydrates (Hagely et al., 2013). Soybean is an important source of isoflavonoids, a group of legume specific plant natural compounds,
synthesized via the phenylpropanoid pathway. In legumes, isoflavonoids act as signaling molecules for establishing symbiotic nitrogen fixation (Ferguson and Mathesius, 2003) and inhibit pathogen growth by acting as phytoalexins (Aoki et al., 2000; Dixon et al., 2002). Isoflavonoids are considered to have human health benefits. They have beneficial effects against breast cancer, prostate cancer, coronary heart disease, osteoporosis and menopausal symptoms (Messina, 2010).

The genome sequence of soybean (~975 Mb) was completed in 2010 (Schmutz et al., 2010). The soybean genome contains 66,153 protein coding loci located on 20 different chromosomes. Soybean has undergone two whole genome duplication events approximately 59 and 13 million years ago, as a result of which 75% of the genes have multiple copies. A global gene expression analysis of soybean embryos during the development in two cultivars with different isoflavonoid accumulation level showed a critical role of CHALCONE SYNTHASE 7 (CHS7) and CHS8 in isoflavonoid accumulation (Dhaubhadel et al., 2007). CHS8 gene expression is regulated by a MYB transcription factor, GmMYB176 (Yi et al., 2010). GmMYB176 is an R1 MYB transcription factor that recognizes a 23 bp motif containing a TAGT(T/A)(A/T) sequence within the CHS8 promoter. When GmMYB176 was silenced in soybean hairy roots, relative transcript level of CHS8 and total isoflavonoid level decreased (Yi et al., 2010). Interestingly, overexpression of GmMYB176 did not alter either the CHS8 transcript level or the total isoflavonoid content, suggesting the possibility of co-operative and/or combinatorial mechanism requiring an additional co-factor. To find the other co-factor working in combination with GmMYB176, a yeast-two-hybrid (Y2H) was conducted using GmMYB176 as bait to screen a cDNA library constructed from mature
soybean embryos to identify prey proteins. A soybean CYP (GmCYP1, accession # AF456323) was one of the GmMYB176 interacting candidates identified in the Y2H screening (Li and Dhaubhadel, unpublished).

1.9 Hypothesis

As described in section 1.6, plant CYPs are associated with stress response. Sequence alignment of GmCYP1 with cotton CYP, GhCYP1, and pigeon pea CYP, CcCYP1, the expression of which are associated with stress response, shows 91% and 73% identity at amino acid level. Therefore I hypothesize that the GmCYP1 is associated with stress response in soybean.
1.10 Thesis objectives

The availability of the whole genome sequence of soybean (Schmutz et al., 2010) provides an excellent opportunity to conduct genome wide analysis of soybean CYPs and an in-depth study of GmCYP1, the candidate CYP that displayed the protein-protein interaction with GmMYB176 in the Y2H screen. At the beginning of this study, not much was reported on soybean CYPs, although a number of soybean CYP sequences have been deposited in the public databases.

Therefore, the objectives of the present research are:

A) To identify the total number of CYP genes present in soybean genome. This question was addressed by an extensive search for CYP genes in the soybean genome database, *in silico* analysis of the predicted amino acid sequences for their domain information, subcellular localization, and phylogenetic analysis.

B) To characterize GmCYP1 in detail. The subcellular localization of GmCYP1 and its tissue specific expression was determined in soybean.

C) To study the interaction between GmCYP1 and GmMYB176 *in planta*. Because GmCYP1 interacted with GmMYB176 in Y2H assay, it is imperative to confirm their *in planta* interaction.

D) To determine the functional role of the GmCYP1 in *Arabidopsis*. Unlike *Arabidopsis*, a large insertion mutant population is not available for soybean. The functional characterization of GmCYP1 was carried out in *Arabidopsis*, the dicot model plant, by using a reverse genetic approach.
Chapter 2: Materials and Methods

2.1 In silico and phylogenetic analysis

The nucleotide sequence of the *GmCYP1* (accession # AF456323) was used as a query to identify additional *CYPs* present in soybean genome. This was achieved using the BLAST (Altschul et al., 1990) algorithm against the soybean genome database (http://phytozome.net/search.php?show=blast&method=Org_Gmax). The newly identified sequences were subsequently used as queries to find other less similar *GmCYPs*. The chromosomal locations for all *GmCYPs* were obtained from the genome database and used to manually draw the chromosomal map. The molecular weight for each GmCYP was calculated by using ProtParam software (http://web.expasy.org/protparam/). The predicted protein sequences were analysed for the prediction of their sub-cellular localization using TargetP1 (http://www.cbs.dtu.dk/services/TargetP/), and WoLF-PSORT (http://wolfpsort.org/) and domain information was obtained from the soybean genome database. To identify the transcribed *GmCYPs* in soybean, the coding sequence of each *GmCYP* was used as a query to BLAST against the soybean Expressed Sequence Tag (EST) database (http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=soybean). The Tentative Contig (TC) sequences were aligned with the corresponding *GmCYP* sequences to identify the percentage identity and coverage. Similarly, to find the GmCYP orthologs in *Arabidopsis*, the amino acid sequences of GmCYPs were used as queries to BLAST against the *Arabidopsis* protein database (http://www.arabidopsis.org/). The *Arabidopsis CYP* with the highest score was assigned as the ortholog for the corresponding *GmCYP*.
To identify the high scoring CYPs in other eukaryotes, BLASTp was employed using GmCYP1 as a query sequence against the National Centre for Biotechnology Information (NCBI) non-redundant protein database (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins). A multiple sequence alignments of amino acid sequences were performed using ClustalW, and the alignment file was exported into BOXSHADE 3.21 (http://www.ch.embnet.org/software/BOX_form.html). For generating the phylogenetic tree of soybean CYPs, amino acid sequences of all GmCYPs were aligned using ClustalW algorithm. The tree was constructed by the Neighbour-Joining method using MEGA5.1 software with 1000 bootstrap replicates (Tamura et al., 2011). The tree was exported into the Interactive Tree Of Life (http://itol.embl.de) for annotation and manipulation (Letunic and Bork, 2011).

2.2 Plant materials and growth conditions

*Arabidopsis* seeds of selected T-DNA insertion mutant lines were obtained from the Arabidopsis Biological Resource Centre (Ohio State University, USA). For growing in soil, seeds were stratified by suspending in water and keeping at 4°C in the dark for 3 days. The seeds were grown on PRO-MIX BX MYCORRHIZAE™ soil (Rivière-du-Loup, Canada) at 25°C/20°C (day/night), 50% humidity, under long-day conditions (16 h light/8 h dark cycle) with light intensity 100-150 µmol m⁻² s⁻¹. For short-day light regime, 8 h light and 16 h dark was applied with the same temperature and humidity setting as for long-day condition. For comparing phenotypes, the seedlings with 2 true leaves (approximately 7-9 days old) were carefully transferred to new trays without damaging their roots.
For growing seeds under sterile conditions, seeds were surface sterilized, with 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 for at least five times to remove residual bleach. Seeds were sown on 1X Murashige and Skoog (MS) basal salt (PhytoTechnology Laboratories, USA) supplemented with 3% (w/v) sucrose (Sigma-Aldrich, Germany), 0.8% (w/v) phytoagar (PhytoTechnology Laboratories, USA), and appropriate antibiotics.

For *Nicotiana benthamiana*, seeds were sprinkled into wet PRO-MIX BX MYCORRHIZAE™ soil (Rivière-du-Loup, Canada) and pots were covered with a transparent cover to maintain humidity. The growth conditions were as described above for long-day conditions. 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.

### 2.3 Bacterial strains

For cloning purpose, DH5α and DB3.1 strains (Invitrogen, USA) of *E.coli* were used. For transient expression in tobacco epidermal cells, and stable transformation in *Arabidopsis, Agrobacterium tumefaciens* strain GV3101 with rifampicin and gentamycin resistance was used. For all bacterial transformations, electroporation was carried out in Gene Pulser® Cuvette (BioRad Laboratories, Canada) with 0.1 cm electrode gap using MicroPulser™ (BioRad Laboratories, Canada). The electroporation setting used for DH5α and GV3101 was 1.79 kV and 2.18 kV, respectively, for 6 milliseconds.
2.4 Genomic DNA isolation from *Arabidopsis*

Genomic DNA from *Arabidopsis* leaves were extracted by cetyl-trimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987). One or two leaves of *Arabidopsis* were pulverized in a 2 mL microfuge tube with 600 µL extraction buffer (1.4 M NaCl, 2% [w/v] CTAB, 50 mM Tris-HCl pH 8.0, 10 mM EDTA and 0.1 mg/mL RNase [added just prior to use]) and incubated at 65°C for 20 minutes. The sample was allowed to cool down to room temperature, and 600 µL of chloroform was added and vortexed thoroughly before centrifugation for 3 minutes at 15000 g. The aqueous upper phase was transferred into a clean microfuge tube and 0.6 volume of isopropanol was added and mixed well. The sample was then centrifuged for 10 minutes at 15000 g, supernatant removed and the DNA pellets washed with 1 mL 70% (w/v) ethanol. The pellets were air dried and dissolved in 20-50 µL of sterile MilliQ water.

2.5 Polymerase Chain Reaction (PCR) condition

Otherwise stated, for all PCR amplifications, a three step PCR was used: denaturation at 95°C for 1 minute, annealing for 30 seconds (at a temperature dependent upon the primer pairs used), extension at 72°C for 1 minute, with total 35 cycles. Before the start of the amplification, initial denaturation at 95°C for 5 minutes was used for all templates except genomic DNA, for which 10 minutes was used. After the last cycle, a 5 minute extension was carried out at 72°C. For the use of High Fidelity Taq DNA polymerase (Invitrogen, USA), extension was carried out at 68°C, with other amplification conditions same as described above.
2.6 Genotyping of T-DNA insertion mutant lines

To identify homozygous *Arabidopsis* T-DNA insertion mutant lines, genomic DNA was extracted from leaves (three-week-old) and genotyping was carried out using a combination of gene-specific and T-DNA-specific primers. Primer sequences for genotyping are listed in Table 2.1.

For Salk_121820 line (named *roc1*), *roc1*-LP and *roc1*-RP were used as gene-specific primers, and Salk-BP and *roc1*-RP as T-DNA-specific primers. Similarly, for Salk_095698C line (named *roc3*), *roc3*-LP and *roc3*-RP were used as gene-specific primers, Salk-BP in combination with *roc3*-RP as T-DNA-specific primers. For CS455423 line (named *roc5*), *roc5*-LP and *roc5*-RP were used as gene-specific primers, and CS-BP was used in combination with *roc5*-RP as T-DNA-specific primers.

2.7 RNA extraction and Reverse-Transcriptase (RT) PCR analysis

To study the tissue specific expression profile of *GmCYP1*, total RNA was extracted according to Wang and Vodkin (1994) from the following soybean tissues: root, stem, leaf, flower bud, flower, embryo (30, 40, 50, 60, and 70 day after pollination [DAP]), seed coat, and pod wall (30 and 40 DAP). The RNA samples were quantified using a NanoDrop spectrophotometer (Thermo Scientific, USA) and their integrity was checked by separating 200 ng of each sample on 1% (w/v) agarose gel in 1X TAE buffer (40 mM Tris, 40 mM Acetate, 1 mM EDTA, pH 8.2-8.4). One microgram of total RNA from each sample was used for cDNA synthesis using the QuantiTect® Reverse Transcription Kit (Qiagen Inc., USA).
### Table 2.1 List of primers and their sequences used for genotyping

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’- 3’)</th>
<th>Melting temp (Tm) in °C</th>
</tr>
</thead>
<tbody>
<tr>
<td>roc1-LP</td>
<td>ACTGAGATCCGTTCTGTGTTTG</td>
<td>60.1</td>
</tr>
<tr>
<td>roc1-RP</td>
<td>CTCTCTCTGACTCTAGGCCGC</td>
<td>60.8</td>
</tr>
<tr>
<td>Salk-BP</td>
<td>CAACACTCAACCCCTATCTCGGGCT</td>
<td>60.6</td>
</tr>
<tr>
<td>roc3-LP</td>
<td>GTTAGGGGGCTTGAGATAAGAAACCG</td>
<td>59.1</td>
</tr>
<tr>
<td>roc3-RP</td>
<td>CTGAGATCCATTCGTGTTCGCACC</td>
<td>59.9</td>
</tr>
<tr>
<td>roc5-LP</td>
<td>GAATCGGATAACGAGTGTGTTG</td>
<td>55.7</td>
</tr>
<tr>
<td>roc5-RP</td>
<td>GAGTTAGCCATGGAGAGAATCCC</td>
<td>55.1</td>
</tr>
<tr>
<td>GK-LB</td>
<td>ATATTGACCATCATACTCATTGC</td>
<td>50.8</td>
</tr>
<tr>
<td>ROC1-cDNA-F</td>
<td>ACCACAATCGGCAACAACCACAG</td>
<td>60.4</td>
</tr>
<tr>
<td>ROC1-cDNA-R</td>
<td>CATCATCCACAACAAAAAACCCTCTTCTCAGTCTG</td>
<td>60.4</td>
</tr>
<tr>
<td>ROC3-cDNA-F</td>
<td>CCCTAAAGTCTACTTCGACATGACCGTC</td>
<td>59.3</td>
</tr>
<tr>
<td>ROC5-cDNA-F</td>
<td>AAATCCTCTCTATAAGAACCCTCTCCAAAACCSS</td>
<td>62.4</td>
</tr>
<tr>
<td>ROC5-cDNA-R</td>
<td>GCTGACCACAATCAGTGATGGTAACCGAC</td>
<td>62.6</td>
</tr>
<tr>
<td>Actin-F</td>
<td>ATGGCGACCGGCTGAGGATATTCA</td>
<td>58</td>
</tr>
<tr>
<td>Actin-R</td>
<td>GCCTTTGCAATCCACATCTGTGTG</td>
<td>57.2</td>
</tr>
</tbody>
</table>
For quantitative RT-PCR, SsoFastTM EvaGreen® Supermix (Bio-Rad, USA) was used with CFX96 real-time PCR detection system (Bio-Rad Inc. USA). Quantitative analysis of GmCYP1 transcript accumulation was carried out using the following primers: qGmCYP1-F: 5’-CGTTGTCATCGCCAACTGC-3’ and qGmCYP1-R1: 5’-CACAATCCCCTAAACGACGACAC-3’. The amplicon sequences were verified and cloned into pGEMT vector (Promega, USA) according to manufacturer’s instruction.

SOYBEAN UBIQUITIN-3 (SUBI-3) was used as a reference gene for data normalization and to calculate the relative mRNA levels. The data were analyzed using a CFX manager (Bio-Rad Laboratories Inc., USA). The melting curve analysis was performed by using pGEMT-qGmCYP1 plasmid as a template.

To evaluate if a T-DNA insertion disrupted gene expression in the Arabidopsis mutant line, the transcript level of corresponding CYP genes were analysed by RT-PCR. Total RNA was extracted from roc1, roc3 and roc5 and wild type (WT) ecotype-columbia plants using the RNeasy® Plant Mini Kit (Qiagen Inc. USA). RNA was quantified using a NanoDrop spectrophotometer (Thermo Scientific, USA) and 200 ng of RNA was run on 1% (w/v) agarose gel in 1X TAE buffer to check for integrity. To synthesize cDNA, 1 µg of total RNA was reverse transcribed using QuantiTect® Reverse Transcription Kit (Qiagen Inc. USA). ROC1 was amplified by ROC1-cDNA-F and ROC1-cDNA-R primer combination. Similarly, ROC3 was amplified by ROC3-cDNA-F and roc3-RP primers, and ROC5 by ROC5-cDNA-F and ROC5-cDNA-R. Amplification of ACTIN was used as a loading control. The sequences of above mentioned primers are listed in Table 2.1.
2.8 Cloning

2.8.1 Cloning of GmCYP1 in Gateway entry vector

Full length GmCYP1 (519 bp) was amplified from soybean cDNA pooled from mature embryo (60 and 70 DAP) using attB1 and attB2 site containing gateway primers GmCYP1-Gate-F: 5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGCCTAACCCTAAGGTCTTCTTC-3’ and GmCYP1-Gate-R: 5’-GGGGACCACTTTGTACAGAAAGCTGGGTCAGAGGGTTGACCGCAGTTG-3’. The PCR products were run on 1% (w/v) agarose gel in 0.5X TBE buffer and a single band of GmCYP1 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), quantitated using a NanoDrop spectrophotometer (Thermo Scientific, USA), and recombined with the entry vector pDONR-Zeo (Invitrogen, USA) using BP clonase reaction mix (Invitrogen, USA) according to the manufacturer’s instruction. The BP reaction product was transformed into E. coli DH5α by electroporation and plated onto Luria-Bertani (LB) agar plates supplemented with antibiotic zeocin (50 µg/mL). The E. coli colonies containing recombinant plasmids were screened by colony PCR using gene-specific primers. The plasmid DNA was prepared from a positive colony using EZ-10 Spin Column Plasmid DNA Kit (Bio Basic Inc., Canada). The insert in recombinant plasmid pDONR-Zeo-GmCYP1 was verified by sequencing using vector-specific M13 forward and reverse primers.
2.8.2 Plasmid construction for subcellular localization

To create a translational fusion of GmCYP1 with Yellow Fluorescent Protein (YFP) at C-terminal region, the gateway cloning method was used. The recombinant plasmid pDONR-Zeo-GmCYP1 was recombined with the destination vector pEarlyGate101 using LR clonase reaction mix (Invitrogen, USA) to obtain pEG101-GmCYP1. The LR reaction was transformed into *E. coli* DH5α by electroporation, followed by colony PCR screening using the gene-specific primers. The positive *E. coli* DH5α colony containing the recombinant plasmid pEG101-GmCYP1 was grown overnight in 3 mL of LB broth at 37 °C for plasmid DNA isolation using EZ-10 Spin Column Plasmid DNA Kit (Bio Basic Inc., Canada). The plasmid DNA was quantified by a NanoDrop spectrophotometer (Thermo Scientific, USA), then transformed into *A. tumefaciens* GV3101 by electroporation.

For creating a nuclear targeting positive control construct, the NLS sequence (PKKKRKVEDP) of Simian virus 40 (SV40) large T antigen (Goldfarb et al., 1986) was fused with Cyan Fluorescent Protein (CFP). As shown in Figure 2.1A-B, four overlapping forward primers and a reverse primer (listed in Table 2.2) were designed that add NLS sequence to the 5’ region of CFP in four successive PCR amplifications. The vector pEarlyGate102 was used as a template to amplify CFP using NLS-3 and attB2-NLS-CFP-R primers. The CFP amplification product was run on 1% agarose gel to ensure correct amplification. The CFP amplicon was diluted 100 folds and used as a template for second PCR amplification using NLS-2 and attB2-NLS-CFP-R primer combination. Again, the second PCR product was run on the gel, diluted 100 folds and used as a template for next PCR reaction using NLS-1 and attB2-NLS-CFP-R primer set.
Figure 2.1 Maps for nuclear localization and ER localization constructs. (A-B)

Nuclear localization and (C-D) ER localization signal sequences were fused upstream of CFP under the control of 35S promoter. For the ER localization construct, the ER retention signal, KDEL, was fused at the C-terminus of CFP. The localization and retention signal sequences were added by using overlapping forward primers and reverse primers, as shown with purple colored arrows. The CFP sequence is highlighted with cyan color and amino acid sequences are represented with orange color. The intervening sequences between 5’- and 3’- end of the CFP are indicated with multiple dots.
A

35S

NLS

CFP

B

\text{attB1-NLS-1}

\text{NLS-1}

\text{NLS-2}

\text{NLS-3}

\text{ATG GTG CCA AAA AAG AAG AGA AAG GTA GAA GAC CCC GTG AGGGTGAGCAAGGCCGA...}

\text{M V P K K K R K V E D P V}

\text{GGGCCGGGATCCACCGGATCTAGA}

\text{attB2-NLS-CFP-R}

C

35S

ER-LS

CFP

KDEL

D

\text{attB1-ER1-F}

\text{ER1-ER2-F}

\text{ER2-ER3-F}

\text{CCG GGA TCC GTT TTT ATT TTT AAT TTT CTT TCA AAT ACT TCC ACC ATG GGA TTT}

\text{P G S V F I F N F L S N T S T M G F}

\text{ER3-ER4-F}

\text{ER4-CFP-F}

\text{TTT CTC TTT TCA CAA ATG CCC TCA TTT TTT CTG TCG ACA CTT CTC TTG TTA TTT}

\text{F L S Q M P S F F L V S T L L L F}

\text{CTA ATA ATA TCT CAC TCT TCT CAT GCC AGGGTGAGCAAGGCCGAGAGCTGT....}

\text{L I I S H S S H A}

\text{CGGCCGGGATCCACCGGATCTAGA AAA GAT GAA TTA TGA}

\text{K D E L \text{stop}}

\text{CFP-KDEL-R}

\text{attB2-KDEL-R}
Table 2.2 List of primers used for designing nuclear, and ER localization construct.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’ to 3’)</th>
<th>Melting temp (Tm) in °C</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLS-3</td>
<td>AAAAGTGAAGACCCCGTGAGGGTGAGCAAGGGC</td>
<td>67.7</td>
</tr>
<tr>
<td>NLS-2</td>
<td>CAAAAAAGAGAGAAAGGAGTAGAAGACCCCGTGAG</td>
<td>63.5</td>
</tr>
<tr>
<td>NLS-1</td>
<td>AAAGCAGGGCTTCTATGGTGCCAAAAAAAGAAGAAGAGAAAGAACCCCGG</td>
<td>60.4</td>
</tr>
<tr>
<td>attB1-NLS-1</td>
<td>GGGGACAAGTTTGATCAAAAAAGCAAGGCTTCA</td>
<td>58.8</td>
</tr>
<tr>
<td>attB2-NLS-CFP-R</td>
<td>GGGGACCCTTCTATGGTGCCAAAAAAAGCAGGCTTCA</td>
<td>60.2</td>
</tr>
<tr>
<td>ER4-CFP-F</td>
<td>TTCTAAAAATATATCTCTTCACTTCTATGACATG</td>
<td>68.9</td>
</tr>
<tr>
<td>ER3-ER4-F</td>
<td>CCACTATTTTCTTCTTGGACACTATTTCTATCTTCAC</td>
<td>62.7</td>
</tr>
<tr>
<td>ER2-ER3-F</td>
<td>TCCAATTTTTTTCTCTTTTCTTTTTCAAAATG</td>
<td>65.9</td>
</tr>
<tr>
<td>ER1-ER2-F</td>
<td>CCGGGATCCCTTTTTATTTTTTTATTTTTTTTTCTTTAATTCATG</td>
<td>65.1</td>
</tr>
<tr>
<td>CFP-KDEL-R</td>
<td>TCATAATATCATTCTCTTAGTCCCCGGTGGATCC</td>
<td>65.9</td>
</tr>
<tr>
<td>attB1-ER1-F</td>
<td>GGGGACAAGGTATGCAAAAAAAAGCAAGGCTTCA</td>
<td>67.9</td>
</tr>
<tr>
<td>attB2-KDEL-R</td>
<td>GGGGACCCTTCTATGGTGCCAAAAAAAGCAAGGCTTCA</td>
<td>66.8</td>
</tr>
<tr>
<td>attB1-CFP-F</td>
<td>GGGGACAAGGTATGCAAAAAAAAGCAAGGCTTCA</td>
<td>57.9</td>
</tr>
<tr>
<td>attB1-YFP-F</td>
<td>GGGGACAAGGTATGCAAAAAAAAGCAAGGCTTCA</td>
<td>65</td>
</tr>
<tr>
<td>attB2-YFP-R</td>
<td>GGGGACCCTTCTATGGTGCCAAAAAAAGCAAGGCTTCA</td>
<td>63.5</td>
</tr>
</tbody>
</table>
The amplicons produced by NLS-1 and attB2-NLS-CFP-R was then used as a template for the final PCR reaction using the gateway primers attB1-NLS-1 and attB2-NLS-CFP-R. The final PCR product was run on 1% (w/v) agarose gel and DNA was extracted from the gel using EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic, Canada). After quantitation of the gel-extracted NLS-CFP DNA, it was cloned into pDONR-Zeo entry vector, and sequence verified using M13 sequencing primers.

To develop an ER targeting positive control vector, a construct was created, which has an ER localization signal (PGSVFIFNFLSNTSTMGFFLFSQMPSFFLV STLLLFLIISHSSHA) fused to the N-terminus of CFP and ER retention signal (KDEL) to the C-terminus of CFP. The primers used in making the ER localization construct are listed in Table 2.2. As shown in Figure 2.1C-D, similar to the nuclear localization positive control construct, a five-step successive PCR amplification was performed to fuse ER targeting and retention sequences to the CFP. For the first PCR, CFP was amplified from pEarlyGate102 using ER4-CFP-F and CFP-KDEL-R primer combination. The PCR amplification product was run on 1% (w/v) agarose gel. After confirming correct size of the amplicon, the amplicon was diluted 100 times and used as a template for second PCR using ER3-ER4-F and CFP-KDEL-R primer set. The third PCR used 100 times diluted amplicon from second PCR as a template and ER2-ER-3-F and CFP-KDEL-R primer pairs. The amplification product was run on 1% (w/v) agarose gel and diluted 100 times to use as a template for ER1-ER-2-F and CFP-KDEL-R primer pairs. Finally, the 100 times diluted PCR product from ER1-ER2-F and CFP-KDEL-R was used as a template for gateway primers attB1-ER1-F and attB2-KDEL-R. The final PCR product of 1005 bp size was cloned into entry vector pDONR-Zeo, as described before. The recombinant plasmid pDONR-Zeo-ERCFP was recombined with destination vector
pEarlyGate100, and then transformed into *E. coli* DH5α and *A. tumefaciens* GV3101 as described previously.

For construction of CFP control construct, *CFP* fragment was amplified from pEarlyGate102 vector as a template, using gateway primers attB1-CFP-F and attB2-NLS-CFP-R (see Table 2.2 for the primer sequences). The amplified fragment was then cloned into the pDONR-Zeo vector as an entry clone followed by recombination with the destination vector pEarlyGate100 as described above. The recombinant plasmid pEG100-CFP was transformed into *E. coli* DH5α and then into *A. tumefaciens* GV3101. Similarly, to generate the for YFP control vector, *YFP* fragment was amplified from pEarlyGate101 vector using gateway primers attB1-YFP-F and attB2-YFP-R (see Table 2.2 for the primer sequences). The cloning strategies and methods were the same as for CFP control construct.

### 2.8.3 Cloning for bimolecular fluorescent complementation (BiFC) assay

For the BiFC assay, the recombinant plasmid pDONR-Zeo-GmCYP1 was recombined separately with pEarlyGate201-YN or pEarlyGate202-YC developed by Lu et al., (2010) to obtain pEG201-GmCYP1-YN and pEG202-GmCYP-YC. The recombinant plasmids were transformed into *E. coli* DH5α, screened by colony PCR, and then transformed into *A. tumefaciens* GV3101. BiFC constructs for GmMYB176 (pEG201-GmMYB176-YN and pEG202-GmMYB176-YC) were available in the lab (Li et al., 2012).
2.9 Transient expression of protein in *N. benthamiana* leaves

The subcellular localization of GmCYP1 was studied by infiltrating *A. tumefaciens* GV3101 carrying pEG101-GmCYP1 in *N. benthamiana* leaves as described by Sparkes et al., (2006). Briefly, a single colony of *A. tumefaciens* GV3101 was picked and inoculated into infiltration culture medium (LB broth containing 10 mM 2-N-morpholino-ethanesulfonic acid [MES] pH 5.6, and 100 µM acitosyringone) 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 OD$_{600}$ reached 0.5-0.8. The culture was centrifuged in 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 then 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 acitosyringone) to a final OD$_{600}$=1 and incubated at room temperature for 2-3 hours with gentle agitation to activate the virulence gene required for transformation.

The leaves of 4-6 weeks old *N. benthamiana* were infiltrated by placing the tip of the syringe, without needle, against underside of the leaf by depressing the syringe plunger gently while giving support on the upper side of the leaf with finger. For co-infiltration of two constructs, equal volumes of both construct-bearing strains in Gamborg’s solution were mixed together and then infiltrated into the leaves. The infiltrated leaves were labelled and plants were returned to the growth room at normal growth condition as described in section 2.2. The protein expression was visualized by confocal microscopy after 72 hours for sub-cellular localization, and 48 hours for BiFC assay.
2.10 Confocal microscopy

Epidermal cell layers of *N. benthamiana* leaves were visualized using Leica TCS SP2 inverted confocal microscope. For YFP visualization, excitation wavelength of 514 nm was used and emissions were collected between 525-545 nm. For visualization of CFP, an excitation wavelength of 434 nm was used and emissions were collected between 460-490 nm. For co-localization, ‘Sequential Scan Tool’ was utilized which records the image in sequential order instead of acquiring them simultaneously in different channels.

2.11 *Arabidopsis* seed germination assay

For the seed germination assay, both WT and mutant (*roc1*, *roc3* and *roc5*) seeds were surface sterilized as described in section 2.2. The 1X MS plates supplemented with 0, 50, 100, or 150 mM NaCl were divided into four sections by marking the plate with permanent marker. The sterile seeds of WT and mutants (*roc1*, *roc3*, and *roc5*) were plated on to separate sections of the plate. The plates were kept in the dark at 4°C for 3 days for stratification, and then transferred to a growth chamber at 23±2 °C under long day conditions. The germinating seeds were observed under a dissecting microscope Nikon SMZ1500 (Nikon, Japan) every day for 7 days, and the number of germinated seeds were recorded on each day. The germination was defined as a 0.5 mm protrusion of the radicle. The data were expressed as percentage germination by dividing with total number of seeds plated for each genotype on each plate. The close-up images of germinated seeds were captured by a Nikon DXM1200 (Nikon, Japan) digital camera integrated into the microscope by using ACT-1 image software (Nikon, Japan) with a resolution of 3840 x 3072.
2.12 Salt stress treatment on plants

To study the possible role of GmCYP1 orthologs, ROC1, ROC3 and ROC5 in salt stress in Arabidopsis, 18 day old WT and mutant (roc1, roc3 and roc5) plants grown under the long day condition were watered for 12 days at 4 day intervals, with increasing concentrations (100, 200 and 300 mM) of NaCl as described in Chen et al., (2012). Plants were allowed to recover for one week and survivors were recognized by those having green inflorescence. A similar experiment was carried out on agar plates as well, where 10 day-old plants grown in MS basal medium were transferred to MS media supplemented with 0, 50, 100 or 150 mM NaCl. Surviving plants were counted after 9 days of stress treatment.

2.13 Statistical analysis

All statistical analyses were performed using Microsoft Excel spreadsheet software. Significant differences between the means in bolting and flowering day calculation were studied by student’s t-test at 99% confidence level.
Chapter 3: Results

The results of a recent study in Dr. Dhaubhadel’s laboratory suggested the requirement for a combinatorial action of GmMYB176 with other factor(s) for CHS8 gene expression and isoflavonoid synthesis in soybean (Yi et al., 2010). In an effort to identify the proteins that interact with GmMYB176, a Y2H assay was performed using GmMYB176 as bait protein and cDNA library constructed from soybean embryo (50-60 DAP) as prey proteins. Of 659 yeast colonies screened, 43 of them were CYP (accession number AF456323) and was named as GmCYP1 (Li and Dhaubhadel, unpublished).

3.1 Soybean genome contains 62 putative GmCYPs

To identify potential CYP genes in soybean, BLAST was performed against the soybean genome database using the nucleotide sequence of GmCYP1 as a query. A total of 62 different soybean CYPs, located on 18 different chromosomes, were identified and were named GmCYP1 to GmCYP62 (Table 3.1, Appendix 1). Of the 62 GmCYPs, 54 of them have only one domain, called a cyclophilin-like-domain (CLD), which is responsible for the cis/trans isomerization of peptidyl prolyl peptide bond. The remaining 8 GmCYPs have other domains in addition to the CLD. As shown in Figure 3.1, GmCYP8, GmCYP9, GmCYP16, and GmCYP17 contained two Tetra-Triacopeptide Repeat (TTR) at the C-terminal end, GmCYP20 and GmCYP35 contained three Tryptophan-Aspartate (WD) repeats at the N-terminal end, and GmCYP56 and GmCYP59 contained an RNA Recognition Motif (RRM) and a Zinc Knuckle (ZK) at their C-terminal end. The molecular masses of GmCYPs ranged from 13.234 kDa (GmCYP5) to 99.69 kDa (GmCYP54).
<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Transcript size (bp)</th>
<th>Protein size (AA)</th>
<th>Predicted subcellular Location</th>
<th>Domain info</th>
</tr>
</thead>
<tbody>
<tr>
<td>GmCYP1</td>
<td>973</td>
<td>172</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP2</td>
<td>1069</td>
<td>172</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP3</td>
<td>786</td>
<td>172</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP4</td>
<td>735</td>
<td>172</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP5</td>
<td>354</td>
<td>117</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP6</td>
<td>398</td>
<td>131</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP7</td>
<td>1138</td>
<td>175</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP8</td>
<td>1577</td>
<td>360</td>
<td>-</td>
<td>MD</td>
</tr>
<tr>
<td>GmCYP9</td>
<td>1241</td>
<td>360</td>
<td>-</td>
<td>MD</td>
</tr>
<tr>
<td>GmCYP10</td>
<td>1260</td>
<td>253</td>
<td>Chloroplast</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP11</td>
<td>890</td>
<td>175</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP12</td>
<td>768</td>
<td>255</td>
<td>Chloroplast</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP13</td>
<td>793</td>
<td>164</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP14</td>
<td>1078</td>
<td>260</td>
<td>Chloroplast</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP15</td>
<td>798</td>
<td>165</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP16</td>
<td>1371</td>
<td>361</td>
<td>-</td>
<td>MD</td>
</tr>
<tr>
<td>GmCYP17</td>
<td>1720</td>
<td>361</td>
<td>-</td>
<td>MD</td>
</tr>
<tr>
<td>GmCYP18</td>
<td>2034</td>
<td>597</td>
<td>Nucleus</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP19</td>
<td>2192</td>
<td>597</td>
<td>Nucleus</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP20</td>
<td>2518</td>
<td>616</td>
<td>-</td>
<td>MD</td>
</tr>
<tr>
<td>GmCYP21</td>
<td>876</td>
<td>194</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP22</td>
<td>902</td>
<td>179</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP23</td>
<td>1268</td>
<td>265</td>
<td>Chloroplast</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP24</td>
<td>681</td>
<td>226</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP25</td>
<td>1046</td>
<td>235</td>
<td>Secretory</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP26</td>
<td>1630</td>
<td>238</td>
<td>Secretory</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP27</td>
<td>2597</td>
<td>659</td>
<td>Nucleus</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP28</td>
<td>1191</td>
<td>263</td>
<td>Chloroplast</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP29</td>
<td>1152</td>
<td>326</td>
<td>Secretory</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP30</td>
<td>1411</td>
<td>327</td>
<td>Secretory</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP31</td>
<td>1497</td>
<td>337</td>
<td>Mitochondria #</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP32</td>
<td>1446</td>
<td>337</td>
<td>Mitochondria #</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP33</td>
<td>984</td>
<td>328</td>
<td>Secretory</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP34</td>
<td>1048</td>
<td>204</td>
<td>Secretory</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP35</td>
<td>2351</td>
<td>616</td>
<td>-</td>
<td>MD</td>
</tr>
<tr>
<td>GmCYP36</td>
<td>2514</td>
<td>668</td>
<td>Nucleus</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP37</td>
<td>1972</td>
<td>493</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP38</td>
<td>600</td>
<td>199</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP39</td>
<td>793</td>
<td>232</td>
<td>Secretory</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP40</td>
<td>1178</td>
<td>236</td>
<td>Secretory</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP41</td>
<td>1204</td>
<td>236</td>
<td>Secretory</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP42</td>
<td>1042</td>
<td>165</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP43</td>
<td>2975</td>
<td>870</td>
<td>Nucleus</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP44</td>
<td>728</td>
<td>241</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP45</td>
<td>1042</td>
<td>226</td>
<td>Secretory</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP46</td>
<td>3097</td>
<td>805</td>
<td>Nucleus</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP47</td>
<td>1257</td>
<td>387</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP48</td>
<td>1584</td>
<td>439</td>
<td>Chloroplast</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP49</td>
<td>921</td>
<td>243</td>
<td>Mitochondria</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP50</td>
<td>705</td>
<td>234</td>
<td>Chloroplast</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP51</td>
<td>1392</td>
<td>232</td>
<td>Mitochondria</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP52</td>
<td>1592</td>
<td>439</td>
<td>Chloroplast</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP53</td>
<td>1968</td>
<td>445</td>
<td>Chloroplast</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP54</td>
<td>2999</td>
<td>886</td>
<td>Nucleus</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP55</td>
<td>1010</td>
<td>286</td>
<td>Chloroplast</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP56</td>
<td>2333</td>
<td>633</td>
<td>Nucleus</td>
<td>MD</td>
</tr>
<tr>
<td>GmCYP57</td>
<td>531</td>
<td>176</td>
<td>-</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP58</td>
<td>1608</td>
<td>350</td>
<td>Chloroplast</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP59</td>
<td>1830</td>
<td>562</td>
<td>Nucleus</td>
<td>MD</td>
</tr>
<tr>
<td>GmCYP60</td>
<td>1860</td>
<td>445</td>
<td>Chloroplast</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP61</td>
<td>684</td>
<td>227</td>
<td>Mitochondria</td>
<td>SD</td>
</tr>
<tr>
<td>GmCYP62</td>
<td>1489</td>
<td>292</td>
<td>-</td>
<td>SD</td>
</tr>
</tbody>
</table>

AA, amino acid; SD, single-domain; MD, multi-domain; (#), prediction with less confidence; (-), no signal peptide detected
Figure 3.1 Schematic representation of different types of domains present in multi-domain GmCYPs. GmCYPs having a specific domain types are listed together in right side of the diagram. The diagram is not drawn to scale. CLD, cyclophilin-like-domain; TTR, tetra-triacopeptide repeat; WD, tryptophan-aspartate repeats; RRM, RNA recognition motif; ZK, zinc knuckle.
Of the 62 GmCYPs, it was ascertained that 12 contain a chloroplast transit peptide (cTP), 10 contain a signal peptide (S), 5 contain a mitochondrial targeting peptide (mTP), 9 contain a nuclear localization signal, and the remaining 26 do not have any targeting peptide for localization to specific cell organelle and are considered cytosolic. The detailed information on each of the GmCYPs listed in Table 3.1 is in Appendix 1.

As of May 2013, the DFCI soybean gene index contains 1,354,268 ESTs representing 73,178 TC sequences. Screening of the soybean EST database for CYPs revealed that 15 GmCYP genes were represented with 99-100% identity over 100% coverage which implies that at least 25% of the GmCYPs are transcribed in various soybean tissues (Appendix 1).

The full length GmCYP1 transcript (973 bp) was predicted to encode a single domain protein of 172 amino acid residues with a calculated molecular mass of 18.22 kDa and a pI of 8.69. The coding region of GmCYP1 (519 bp) has 100% sequence identity with TC464530 and TC433886 over 100% coverage in alignment. GmCYP1 contains only one exon and is located on the long arm (q arm) of chromosome 11, approximately 16 Mb from the centromere (Figure 3.2).

### 3.2 Chromosome location and phylogenetic analysis of soybean CYP gene family

For visualization of genome organization and distribution of GmCYPs at different chromosomes, a chromosome map was manually constructed. The 62 GmCYPs were found on 18 different chromosomes of the soybean genome. A diagrammatic representation of the chromosomal distribution of soybean CYP genes is
Figure 3.2 Genomic distributions of *GmCYP* genes on soybean chromosomes. Chromosomal locations of *GmCYPs* are indicated based on the location of the genes, length of chromosome and position of centromere. The chromosomes are drawn to scale and chromosome numbers are shown under each chromosome. The *GmCYPs* that are clustered together and speculated to have undergone segmental duplication are indicated by boxes of the same color. The chromosomes are drawn to scale.
depicted in Figure 3.2. Certain chromosomes, such as chromosome 11 and 19, had a relatively dense occurrence of CYP and contain maximum number of CYPs (6 each), whereas only one CYP (GmCYP54) was present on chromosome 14. There were no CYPs on chromosome 8 or 16. GmCYP1 was located on chromosome 11 (Figure 3.2).

To explore the evolutionary relationship among soybean CYPs, a phylogenetic analysis was performed using the predicted amino acid sequences (Figure 3.3). The corresponding genes, whose putative amino acid sequences were used to make phylogenetic tree, are enlisted in Appendix 1. A phylogenetic tree was generated by the Neighbour-Joining method with bootstrap trial 1000. As observed for many other genes in soybean, most of the GmCYPs clustered together in a pair, reflecting the ancient genome duplication event (Li and Dhaubhadel, 2011; Schmutz et al., 2010). Of the 62 GmCYPs, 54 were found to be clustered in pairs in the phylogenetic tree. Among the remaining 8, 7 GmCYPs branched-off from the terminal branch of another pair of GmCYPs. GmCYP12 was the only soybean CYP that did not pair up with any other CYP. Moreover, each of the multi-domain GmCYPs paired with another multi-domain GmCYPs. Interestingly, the GmCYPs predicted to be targeted to the same subcellular compartment formed separate clades. For example, GmCYPs with chloroplast transit peptide (GmCYP10, GmCYP23, GmCYP14 and GmCYP28) formed a distinct clade on the tree. Another four chloroplast-localizing GmCYPs (GmCYP48, GmCYP52, GmCYP53, and GmCYP60) also formed a distinct clade, but in a different location of the tree. Similarly, the GmCYPs with nuclear localization signal (GmCYP27, GmCYP36, GmCYP54, GmCYP43, and GmCYP46) also formed a separate clade on the tree (Figure 3.3).
Figure 3.3 Phylogenetic tree of GmCYPs constructed using Neighbour-Joining method. The phylogenetic tree was generated by MEGA5.1 software (Tamura et al., 2011) using putative amino acid sequence of GmCYPs, listed in Appendix 1, and tree was annotated using Interactive Tree of Life (Letunic and Bork, 2011). The numbers next to the branch shows the 1000 bootstrap replicates expressed in percentage. The solid line represents the real branch length and dotted lines added later for better visualization. The putative chloroplast, mitochondria, ER, nucleus, and cytosol localizing GmCYPs are indicated with green, red, blue, pink, and grey color, respectively. The multi-domain CYPs are underlined.
By comparing the location of GmCYPs on the chromosome map and to phylogenetic position, one third of GmCYPs showed an interesting pattern. The GmCYPs that cluster together on a chromosome can be paired-up with clustered genes from another chromosome. For example, GmCYP36, and GmCYP13, which are clustered in the sub-telomere region of chromosome 3, are most similar to GmCYP27, and GmCYP15, respectively, which are clustered together in sub-telomere region on chromosome 9. Similarly, GmCYP4, GmCYP39, and GmCYP46 on chromosome 4 paired up with GmCYP3, GmCYP38, and GmCYP43, respectively, from chromosome 6. GmCYPs from chromosome 11, GmCYP18 and GmCYP32 paired up with GmCYP19 and GmCYP31, respectively, from chromosome 1 whereas GmCYP1 and GmCYP41, from chromosome 11, paired up with GmCYP2 and GmCYP40, respectively, from chromosome 12.

### 3.3 PPIase and CsA binding sites are conserved in GmCYP1

In order to find sequences closely related to GmCYP1, protein-protein BLAST (BLASTp) was employed using GmCYP1 as a query against NCBI non-redundant protein database. Table 3.2 shows the list of 10 high-scoring and previously characterized CYPs. An alignment of the deduced sequence of GmCYP1 with previously characterised CYPs from several different plants, human and yeast revealed three general features (Table 3.2 and Figure 3.4). First, three amino acid residues that critically affect PPIase activity (R55, F60 and H126) (Zydowsky et al., 1992) were conserved. Second, the tryptophan residue (W121) implicated in substrate CsA binding (Liu et al., 1991a; Zydowsky et al., 1992) was present in all the CYPs studied, including GmCYP1. Third, a conserve region of 7 amino acids (RSGKPLH) typical of the single domain plant CYP is present in GmCYP1.
Table 3.2 Comparison of GmCYP1 with closely related CYPs from other plant species, animal, and yeast.

<table>
<thead>
<tr>
<th>Name</th>
<th>Species</th>
<th>Accession #</th>
<th>AA Identity (%)</th>
<th>AA Similarity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GhCYP1</td>
<td><em>Gossypium hirsutum</em></td>
<td>ACT63839.1</td>
<td>91</td>
<td>96</td>
</tr>
<tr>
<td>ROC3</td>
<td><em>Arabidopsis thaliana</em></td>
<td>NP_179251.1</td>
<td>88</td>
<td>94</td>
</tr>
<tr>
<td>ROC6</td>
<td><em>Arabidopsis thaliana</em></td>
<td>NP_179709.1</td>
<td>84</td>
<td>92</td>
</tr>
<tr>
<td>ROC1</td>
<td><em>Arabidopsis thaliana</em></td>
<td>AAB71402.1</td>
<td>84</td>
<td>92</td>
</tr>
<tr>
<td>BnCYP</td>
<td><em>Brassica napus</em></td>
<td>AAA62706.1</td>
<td>83</td>
<td>91</td>
</tr>
<tr>
<td>ROC5</td>
<td><em>Arabidopsis thaliana</em></td>
<td>NP_195213.1</td>
<td>79</td>
<td>87</td>
</tr>
<tr>
<td>CcCYP1</td>
<td><em>Cajanus cajan</em></td>
<td>ADB04247.1</td>
<td>73</td>
<td>81</td>
</tr>
<tr>
<td>hCYP-D</td>
<td><em>Homo sapiens</em></td>
<td>pdb</td>
<td>2Z6W</td>
<td>A</td>
</tr>
<tr>
<td>hCYP-A</td>
<td><em>Homo sapiens</em></td>
<td>NP_066953.1</td>
<td>69</td>
<td>79</td>
</tr>
<tr>
<td>Cpr1</td>
<td><em>Saccharomyces cerevisiae</em></td>
<td>NP_010439.1</td>
<td>63</td>
<td>74</td>
</tr>
</tbody>
</table>

AA, Amino acid.
Figure 3.4 Multiple sequence alignment of deduced amino acid sequence of
GmCYP1 with closely related CYPs from other species. BLASTp was used to find
closely related CYPs from other species. Amino acid sequences of ROC3, BnCYP,
GhCYP1, CcCYP1, ROC6, ROC1, ROC5, Cpr1, hCYP-D, hCYP-A, and GmCYP1 (See
Table 3.2 for accession numbers) were aligned by ClustalW and shading was performed
using BOXSHADE 3.21. Identical and conserved amino acids are represented with dark
and light grey backgrounds, respectively. Amino acids involved in PPlase activity (R_{55},
F_{60} and H_{126}) and CsA binding (W_{121}) are marked with purple and red asterisks,
respectively. A conserved region for single domain plant CYPs is shown with a green bar
on the top. The position of amino acid indicated for PPlase and CsA binding sites are
with reference to human CYP A (hCypA).
3.4 GmCYP1 interacts with an R1 MYB transcription factor, GmMYB176

To further confirm the Y2H results of protein-protein interaction between GmCYP1 and GmMYB176, their interaction in planta was studied by bimolecular fluorescent complementation (BiFC) assay. The basic principle of BiFC is that the candidate proteins are fused with either the N-terminal half (YN) or C-terminal half (YC) of YFP and are transiently co-expressed in *N. benthamaina* leaves. As a result of protein-protein interaction, the N-terminal and the C-terminal half of YFP come in close proximity to each other resulting in fluorescence (Ohad et al., 2007).

A translational fusion of GmCYP1 was created with YN or YC of YFP and co-expressed in tobacco epidermal cells in the following combinations: (A) GmCYP1-YN and GmMYB176-YC, (B) GmMYB176-YN and GmCYP1-YC. Protein expression was monitored by confocal microscopy. The negative controls used for the experiment were: co-expression of (C) GmCYP1-YN and YC only, (D) GmCYP1-YC and YN only, (E) GmCYP1-YN only, and (F) GmCYP1-YC only. As shown in Figure 3.5A, GmCYP1 interacts with GmMYB176 in planta. The interaction between GmCYP1 and GmMYB176 was strong in the nucleus, however, only a weak interaction was observed in the cytoplasm. Similar results were obtained for the reciprocal combination (Figure 3.5B). No YFP signal was detected with control constructs (Figure 3.5C-F), or for GmMYB176-YN and YC only, and GmMYB176-YC and YN only (Li et al., 2012).
Figure 3.5 Confocal images showing the interaction of GmCYP1 and GmMYB176 in planta. *A. tumefaciens* GV3101 carrying (A) GmCYP1-YN and GmMYB176-YC, (B) GmMYB176-YN and GmCYP1-YC, (C) GmCYP1-YN and YC only, (D) GmCYP1-YC and YN only, (E) GmCYP1-YN only, and (F) GmCYP1-YC only were infiltrated into *N. benthamiana* leaves and imaged after 48 hours using confocal microscopy. Interaction between GmCYP1 and GmMYB176 is visualized by strong yellow fluorescence in the nucleus and weak in cytoplasm, indicated with red and white arrows, respectively (A and B). No fluorescence was detected in control constructs (C-F).
3.5 GmCYP1 forms homodimers in planta

When *A. tumefaciens* GV3101 carrying the BiFC plasmids containing GmCYP1-YN, and GmCYP1-YC were co-infiltrated into *N. benthamiana* leaves and visualized by confocal microscope, strong yellow fluorescence was observed in the nucleus and the cytoplasm suggesting that GmCYP1 forms homodimers *in planta* (Figure 3.6).

3.6 GmCYP1 localizes in the nucleus and cytoplasm

To study the subcellular localization of GmCYP1, a translation fusion of GmCYP1 with YFP was created under the control of 35S promoter and transiently expressed in *N. benthamiana* leaves as described above in sections 2.8 and 2.9. Although there was no nuclear localization sequence detected in GmCYP1 (Table 3.1), confocal imaging of the GmCYP1-YFP infiltrated leaves showed nuclear and cytoplasmic localization of GmCYP1 (Figure 3.7). The control constructs YFP only, and CFP only, were localized to both the nucleus and cytoplasm, as shown in Figure 3.7A-B. As shown by several studies earlier (English and Voeltz, 2013; Nziengui et al., 2007) as well as in the present study, the ER targeted CFP displayed a net-like structure (Figure 3.7C). The nuclear localization construct (NLS-CFP) also showed its correct nuclear localization, which can be seen in the overlapped image of NLS-CFP and bright field (Figure 3.7 D-F). To confirm the nuclear localization of GmCYP1, it was co-expressed with NLS-CFP. The co-expression of GmCYP1-YFP and NLS-CFP showed overlap of CFP signal of NLS-CFP with YFP signal of GmCYP1 (Figure 3.7 G-I), confirming the nuclear localization of GmCYP1.
Figure 3.6 Confocal Image showing GmCYP1 homodimers in planta. (A) *A. tumefaciens* GV3101 carrying plasmids with GmCYP1-YN and GmCYP1-YC were co-infiltrated into the leaves of *N. benthamiana* and visualised after 48 hours by confocal microscopy. The dimer formation in the nucleus and cytoplasm is indicated with white and red arrows, respectively. (B) The bright field image of (A).
**Figure 3.7 Subcellular localization of GmCYP1.** *A. tumefaciens* GV3101 carrying the plasmids with following constructs were infiltrated into *N. benthamiana* leaves and visualized by confocal microscopy after 72 hours. (A) YFP only control, (B) CFP only control, (C) ER-CFP control, (D-F) NLS-CFP control, (G-I) co-expression of GmCYP1-YFP and NLS-CFP, and (J-L) co-expression of GmCYP1-YFP and ER-CFP.
There were several net-like structures seen in the confocal images of GmCYP1-YFP which resembled to proteins that localize to the ER. However, co-expression of ER-CFP with GmCYP1-YFP did not demonstrate perfect overlapping signals of the CFP in ER-CFP with GmCYP1-YFP (Figure 3.7J-L). This suggests that GmCYP1 is not an ER protein. The epidermal cells of *N. benthamiana* have large vacuole, leaving very small space in between the tonoplast and plasma membrane. The smaller cytosolic region makes cytosolic protein look like an ER protein.

**3.7 *GmCYP1* is expressed ubiquitously in all soybean tissues**

To study temporal and spatial expression of *GmCYP1* in soybean, a detailed transcript analysis, using real-time-quantitative PCR with gene-specific primers, was performed. Total RNA isolated from different tissues at several different developing stages from the soybean cultivar Harosoy63 was used in the analysis. As shown in Figure 3.8, *GmCYP1* was expressed in all the tissues, albeit at varying levels. The *GmCYP1* transcript accumulation was higher in embryos compared to other tissues. The level of *GmCYP1* transcript increased in embryos during the development, showing highest levels (three-fold increase) in the embryos at 60 and 70 DAP compared to that in the embryo at 30, 40 or 50 DAP. A 23 and 18 fold less *GmCYP1* transcript was found in seed coat compared to the 60 and 70 DAP embryos, respectively.

Interestingly, the *Arabidopsis* ortholog of *GmCYP1*, *ROC1* (AGI ID: At4g38740) was also found to have higher transcript accumulation in seed compared to that of other tissues (Figure 3.9). The normalized mean expression data of *ROC1* was compiled from AtGenExpress Visualization Tool (http://jsp.weigelworld.org/expviz/expviz.jsp) and are represented in graphical form. As shown in Figure 3.9A, *ROC1* transcript accumulation is
Figure 3.8 Quantitative expression analysis of *GmCYP1* in different soybean tissues.

Total RNA extracted from root, stem, leaf, flower bud, flower, embryo (30, 40, 50, 60, and 70 DAP), seed coat and pod wall (30 and 40 DAP) were used for quantitative RT-PCR analysis of *GmCYP1*. Two biological replicates and three technical replicates for each biological replicate were carried out. The standard error of the mean is represented with an error bar. The data were normalized against the *SUBI*-3 gene.
Figure 3.9 Expression profiles of *Arabidopsis* cyclophilin *ROC1* in different tissues and developmental stages. The mean normalized expression values of *ROC1* were obtained from AtGenExpress Visualization Tool (http://jsp.weigelworld.org/expviz/expviz.jsp). Error bars indicate the standard deviation of the mean. The stages of seed developments are: stage 3, mid globular to early heart embryos; stage 4, early to late heart embryos; stage 5, late heart to mid torpedo embryos; stage 6, mid to late torpedo embryos; stage 7, late torpedo to early walking-stick embryos; stage 8, walking-stick to early curled cotyledons embryos; stage 9, curled cotyledons to early green cotyledon embryos; and stage 10, green cotyledon embryos.
highest in seeds compared to other tissues. During different stages of seed development, the expression of ROC1 increased gradually from mid globular stage (stage 3) through early curled cotyledon embryo (stage 8) and then from early green cotyledon (stage 9) to green cotyledon embryos (stage 10) (Figure 3.9B).

3.8 GmCYP1 orthologs in Arabidopsis function in transition from vegetative to reproductive phase

As T-DNA insertion mutant lines for GmCYP1 are not available in soybean, in the present study I used Arabidopsis as a model system for the functional characterization of GmCYP1, assuming that this gene’s function is conserved between these two plant species. A search for the GmCYP1 ortholog in Arabidopsis identified 4 genes, ROC3 (AGI ID: At2g16600), ROC6 (AGI ID: At2g21130), ROC1 (AGI ID: At4g38740), and ROC5 (AGI ID: At4g34870) with 87.6, 84.1, 83.7 and 79.0 % amino acid identity, respectively, with GmCYP1. Since there was no T-DNA insertion mutant line available for ROC6 in the Arabidopsis Biological Resource Centre, the other three genes were further studied. The information for the T-DNA insertion mutant lines for each of these genes are provided in Table 3.3.

To identify homozygous T-DNA insertion mutants for each of the lines (roc1, roc3 and roc5), seeds were grown in soil and DNA was extracted from the leaves. PCR was carried out using the gene-specific and the T-DNA-specific primer pairs, as described in section 2.6. As expected, the homozygous T-DNA insertion lines gave PCR amplicons with T-DNA-specific primer pairs only, while WT plants produced PCR amplicons with the gene-specific primer pairs only (Figure 3.10). The heterozygous lines
Table 3.3 *Arabidopsis* orthologs of *GmCYP1* and the corresponding T-DNA insertion mutant lines used in this study.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>AGI ID</th>
<th>% identity with GmCYP1 (aa level)</th>
<th>ABRC stock name</th>
<th>T-DNA insertion site</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>ROC3</em></td>
<td>AT2G16600</td>
<td>87.6</td>
<td>Salk_095698C</td>
<td>Exon</td>
</tr>
<tr>
<td><em>ROC1</em></td>
<td>AT4G38740</td>
<td>84.1</td>
<td>Salk_121820</td>
<td>Promoter</td>
</tr>
<tr>
<td><em>ROC5</em></td>
<td>AT4G34870</td>
<td>79.0</td>
<td>CS455423</td>
<td>Exon</td>
</tr>
</tbody>
</table>

*aa*- amino acid
Figure 3.10 T-DNA insertion map and gel images showing homozygous T-DNA insertion in mutants. Genotyping was carried out to identify homozygous T-DNA insertion mutant plants for each of the lines. The primer binding sites are indicated by arrowhead with corresponding primer name. LP and RP combination acts as gene-specific primer pairs, and BP and RP acts as T-DNA-specific primer pairs. A single band with T-DNA-specific primers and gene-specific primers indicate the homozygous mutant and WT, respectively. The T-DNA insertion map is not drawn to scale.
produced PCR amplicons with both gene-specific and T-DNA specific primer pairs (data not shown).

To determine if the T-DNA insertion disrupted the gene expression in the mutant lines, the transcript level of the corresponding gene in the \textit{roc1}, \textit{roc3} and \textit{roc5} lines was compared with that of WT. As shown in Figure 3.11, T-DNA insertion in the promoter region of \textit{roc1} decreased the expression of \textit{ROC1} whereas \textit{ROC3} and \textit{ROC5} transcript was completely abolished in \textit{roc3} and \textit{roc5}, respectively.

To study the phenotypic differences between WT and mutants (\textit{roc3} and \textit{roc5}), each of the mutants were grown together with WT under the same growth conditions. The early flowering phenotype of \textit{roc1} under long day condition was reported previously (Trupkin et al., 2012). Similar to \textit{roc1}, both \textit{roc3} and \textit{roc5} showed early bolting and early flowering compared to WT when grown under long day conditions (Figure 3.12A-B and 3.13). Under short day conditions, there was no significant difference in bolting and flowering time between WT and mutants (\textit{roc3} and \textit{roc5}) (Figure 3.12 C). Under long day conditions, \textit{roc3} and \textit{roc5} bolted on day 25.33±1.50 and 24.94±1.25, respectively whereas WT bolted on day 31.18±1.78. Similarly, \textit{roc3} and \textit{roc5} started flowering earlier, on day 28.61±1.29 and 28.65 ±1.27, respectively compared to WT (day 34.65±1.54).

Flowering time of \textit{Arabidopsis} can also be expressed in term of total number of rosette leaves when the first flower opens. The earlier a plant starts to flower, the fewer the number of rosette leaves develop. At the time of first flower opening, both \textit{roc3} and \textit{roc5} showed a fewer number of rosette leaves (11.89±1.32 for \textit{roc3} and 12.53±1.01 for
Figure 3.11 T-DNA insertion in *ROC1*, *ROC3* and *ROC5* genes alter their corresponding transcript levels. Total RNA from homozygous T-DNA insertion mutants of each line and WT was extracted and reverse transcribed into cDNA. RT-PCR was carried out using setS of gene-specific primer pairs for each genotype, and was compared to that of with WT. *ACTIN* was used as a loading control.
Figure 3.12 Flowering day phenotype of mutants (*roc3* and *roc5*) under long and short day light conditions. Bolting and flowering time of *roc3*, *roc5* and WT in (A-B) long day and (C) short day conditions. Eighteen plants from each genotype under each light regime were recorded for their bolting and flowering time. Flowering time under the long day condition is also expressed in term of total number of rosette leaves during the emergence of first flower for *roc3*, *roc5* and WT. Error bars represent the standard error of the mean of three biological replicates. Student’s t-test was performed at p<0.01 and statistically significant difference in bolting and flowering time of mutants with WT are indicated with two asterisks (**).
Figure 3.13 Photograph showing the early flowering phenotype of *roc3* and *roc5* under long day condition. (A) The 28 day-old plants from each of the genotypes were taken out from the tray and transferred into new pots for photography. B-D show the total number of rosette leaves at the time of flowering for *roc3, roc5*, and WT, respectively.
roc5) compared to the WT (18.41 ±1.54) grown under the long day conditions. Figures 3.12 and 3.13 clearly show the early transition from vegetative to reproductive stage in roc3 and roc5 compared to the WT. Bolting day was considered as the emergence of bolt of 1 cm length and flowering day was considered as the day of opening of the first flower. A total of 18 plants from each genotype were used for data collection for each biological replicate. Each experiment was done 3 times.

3.9 GmCYP1 orthologs in Arabidopsis plays a role in seed germination during salt stress

To investigate if a mutation in GmCYP1 orthologs in Arabidopsis affects seed germination under normal and stress conditions, seeds from roc1, roc3, roc5, and WT were germinated on medium supplemented with 0, 50, 100 or 150 mM NaCl, and seed germination was recorded every day for up to seven days. There was no significant difference in seed germination rate between WT and mutants under the no salt condition (Figure 3.14A). However, with successive increases in salt concentration (50 mM, 100 mM or 150 mM), both the rate and the efficiency of seed germination were decreased in all of the mutants compared to that of the WT (Figure 3.14B-D). For medium with 50 mM NaCl, more than 95% of the WT seeds germinated by day two, whereas it was less than 60% for the mutants, and took an additional two days to reach 95% (Figure 3.14B). At 100 mM NaCl, 95% of the WT seeds germinated by day three, but less than 10% had for each of the mutants, and final germination percentage reached only 70% on the 7th day post-seeding (Figure 3.14C). Similarly, at 150 mM NaCl, 65% of the WT seeds were germinated by day three, whereas less than 1% mutant seeds were germinated
Figure 3.14 Mutation in GmCYP1 orthologs in *Arabidopsis* impair seed germination in response to salt treatment. The WT and mutants (*roc1*, *roc3*, and *roc5*) were germinated on MS medium supplemented with different salt concentrations. Germination was defined as 0.5 mm protrusion of the radicle. Germinated seeds were counted every day, up to seven days after stratification, and data expressed in percentage germination. (A), (B), (C), and (D) show the rate of seed germination in MS medium supplemented with 0 mM, 50 mM, 100 mM, or 150 mM NaCl, respectively. Error bars represent the standard error of the mean for three biological replicates.
Figure 3.15 Photographs showing the germination of WT and mutant (*roc1, roc3, and roc5*) seeds in salt stress. WT and mutant seeds were germinated on MS media supplemented with the following salt concentrations (A) control (0 mM NaCl); (B) 50 mM NaCl; (C) 100 mM NaCl; (D) 150 mM NaCl. The germination rate was scored each for each genotype everyday up-to seven days. All the photographs were taken on the 7th day post-seeding. For 100 and 150 mM NaCl plates, photos to the left and right of the plates are close-up shots of the germinated seeds to show the cotyledons development on the 7th day.
(Figure 3.14D). On day seven, more than 95% of WT seeds were germinated in 150 mM NaCl condition, but it was less than 29% for the mutants.

Interestingly, at 100 mM NaCl, the germinated seeds of mutants were not able to produce green cotyledons whereas all of the germinated WT seeds developed green cotyledons (Figure 3.15). For medium with no salt and 50 mM salt, by the end of 7th day, all of the mutants were able to develop cotyledons similar to those of the WT, and there were no morphological differences between WT and the mutants (Figure 3.15A-B). On the 100 mM NaCl plates, even though 70% of the mutant seeds had germinated, none of them developed green cotyledons by the end of 7th day, whereas most of the germinated WT were able to develop green cotyledon (Figure 3.15 C). On 150 mM NaCl plate, neither WT nor mutant seeds were able to develop green cotyledon (Figure 3.15D).

Although all three mutants were more sensitive to salt during their germination than was WT, they responded consistently when salt stress was applied to mature plants (Figure 3.16). No morphological differences were noticed between the soil grown mutants and WT plants during the salt stress treatment or during the recovery period of one week (Figure 3.16A-C). A similar experiment was carried out on plates as described in “Materials and Methods” section. The results were similar to soil-grown plants with no morphological differences between WT and the mutant plants during the course of stress treatment (Figure 3.16D-G).
Figure 3.16 Photographs of the WT and mutants (roc1, roc3, and roc5) after salt stress treatment. Salt stress was given to WT and mutant (18 days-old soil-grown or 10 day-old plate-grown) plants. Phenotypic response towards the salt stress by the all three mutants was found to be same as that of the WT.
Chapter 4: Discussion

CYPs are ubiquitous proteins that are found in organisms ranging from archaea, bacteria to plants and animals (Galat, 1999; Maruyama et al., 2004). CYPs have peptidyl prolyl cis/trans isomerase activity and are structurally conserved throughout evolution. Different organisms have different numbers of CYPs. E. coli, S.cerevisiae, and human contain 2, 8, and 16 CYP genes, respectively (Arevalo-Rodriguez et al., 2004; Galat, 2003; Hayano et al., 1991). However, plant genomes contain relatively larger numbers of CYPs. For example, Arabidopsis and rice genome have 35 and 28 CYPs, respectively (Ahn et al., 2010; Romano et al., 2004; Trivedi et al., 2012). The present research focusses on a genome wide scan and systematic characterization of the soybean CYP family and the detailed characterization of GmCYP1.

An extensive data mining of the soybean genome identified 62 putative CYP genes (Table 3.1, Appendix 1), making soybean CYPs the largest CYP family reported to date. Soybean is a palaeopolyploid with a genome size of approximately 950 Mb. It has undergone at least two whole genome duplications, approximately 59 and 13 million years ago (Schmutz et al., 2010). Because of the genome duplication events, nearly 75% of the soybean genes are present in multiple copies. Gene duplication events result in two copies of each gene which undergo shuffling and rearrangement creating the potential for new diversity. There are four possible fates of duplicated genes (reviewed in Charon et al., 2012; Moore and Purugganan, 2005). First, one copy of the gene may be deleted during the course of evolution. Such deletion results in loss of functional redundancy among the duplicate genes. Second, both copies of the genes may be retained and share
the ancestral function, but gradually develop partially different functions (sub-functionalization). Third, one copy of the gene may acquire new function(s) during the course of evolution (neo-functionalization). Finally, there may be an intermediate stage between sub-and neo-functionalization. All of these outcomes depend on the role of the specific gene in plant growth and development. Only those genes that are associated with critical function for normal plant growth and development are retained, while others may be deleted. The higher number of CYPs found in the soybean genome likely reflects the important role of \textit{GmCYPs} in soybean during normal growth and development, as well as in response to environmental stimuli.

Phylogenetic analysis of soybean CYPs illustrated that most of the GmCYPs are clustered in pairs (except for GmCYP57, GmCYP54, GmCYP38, GmCYP39, GmCYP42, GmCYP37, GmCYP12, and GmCYP33) (Figure 3.3). As explained earlier, this trend supports the ancient genome duplication events as well as retention of duplicated genes during the course of evolution. Interestingly, the chromosome map shown in Figure 3.2 suggested the possibility of inter-chromosomal rearrangements between different soybean chromosomes after genome duplication. Phylogenetic analysis and genomic distribution of \textit{GmCYP} genes together (Figures 3.2 and 3.3) revealed some interesting findings. Several \textit{GmCYP} genes that were located together on the same chromosome had their closest family members also grouped together and located on other chromosome, providing evidence for segmental duplication of chromosomal regions containing those \textit{GmCYP} genes. Segmental duplication plays a vital role in generating members of a gene family during evolution (Cannon et al., 2004).
Among 62 GmCYPs, 54 of them contain single domain and 8 have multi-domains. Of 8 multi-domain GmCYPs, GmCYP8, GmCYP9, GmCYP16, and GmCYP17 have an N-terminal CYP-like domain (CLD), and two tetratricopeptide repeats (TPR) at the C-terminus (Figure 3.1). The TPR motif is degenerate in nature and consists of a 34 amino-acid repeat usually arranged in tandem arrays (Blatch and Lassle, 1999). They are found in multiple copies in different proteins where they mediate protein-protein interactions and often help in assembly of multi-protein complexes. The *Arabidopsis* ortholog of GmCYP8, GmCYP9, GmCYP16, and GmCYP17 is AtCYP40 (AGI ID: At2g15790) and is involved in microRNA-mediated gene regulation (Earley and Poethig, 2011). Loss-of-function mutation of *AtCYP40* showed a precocious phase change with reduced number of juvenile leaves, without altering the flowering time (Berardini et al., 2001). This vegetative phase change was associated with reduced activity of ARGONAUTE1 (AGO1) protein, a component of RNA-induced silencing complex (RISC), which elevated the expression of *miRNA-156*-regulated members of *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) gene family (Smith et al., 2009).

Moreover, the conserved amino acids in TPR domain of AtCYP40 was required for interaction between AtCYP40 and cytoplasmic Hsp90 proteins (Earley and Poethig, 2011). Previous studies have demonstrated the requirement of Hsp90 for loading of small RNA into RISCs (Iki et al., 2010), and the interaction between Hsp90 and AtCYP40 was essential for the function of AtCYP40 *in planta* (Earley and Poethig, 2011), suggesting the critical role of AtCYP40 TPR domain in microRNA-mediated gene regulation. The possible association of the TPR motif of GmCYP8, GmCYP9, GmCYP16 and/or GmCYP17 with microRNA-mediated gene regulation can be speculated.
Two soybean CYPs, GmCYP20 and GmCYP35, have WD domains along with CLD. Both GmCYP20 and GmCYP35 have three WD repeats at their N-terminal end and a CLD at the C-terminal end (Figure 3.1). WD domains are approximately 40 amino acids long, usually end with tryptophan-aspartate (WD) dipeptide, and are ubiquitously found in eukaryotes (Neer et al., 1994). The tandemly repeated WD domains make a series of four stranded anti-parallel β-sheets, which fold together to form a typical solenoid protein domain called β-propeller (van Nocker and Ludwig, 2003). The WD repeat-containing proteins are involved in a wide variety of cellular functions by providing binding sites for two or more proteins, or by fostering transient interactions with other proteins (Stirnimann et al., 2010; van Nocker and Ludwig, 2003). The Arabidopsis WD repeat-containing proteins are involved in proteasome-mediated protein degradation, where the WD domain of the adapter protein interacts with the E3 ligase component and facilitates substrate assembly for degradation (Zhang et al., 2008). The Arabidopsis CYP, AtCYP71 (AGI ID: At3g44600), contains 2 WD repeats (Romano et al., 2004). AtCYP71 physically interacts with FAS1 subunit of CHROMATIN ASSEMBLY FACTOR-1 via the WD domain and functions in chromatin remodelling (Li and Luan, 2011). The very high sequence similarity between AtCYP71 with GmCYP20 (87%) and GmCYP35 (83%) suggests that these two GmCYPs may play similar roles in soybean.

Further, sequence analysis identified 2 soybean CYPs, GmCYP56 and GmCYP59, with an RNA recognition motif (RRM) and zinc knuckle (ZK) at the C-terminus along with CLD at the N-terminal end (Figure 3.1). The RRM is a small RNA binding motif of 90 amino acids and is conserved in wide variety of organisms
(Bandziulás et al., 1989). As transcription of a gene by RNA polymerase II and processing of the nascent mRNA occurs simultaneously in a complex (Zorio and Bentley, 2004), different proteins involved in nascent mRNA processing interact each other via RRM. AtCYP59 (AGI ID: At1g53720) is the Arabidopsis orthologs of GmCY56 (80%) and GmCYP59 (66%) (Appendix 1) and has an RRM motif along with a ZK domain (Romano et al., 2004). AtCYP59 localizes in the nucleus and regulates transcription by interacting with the C-terminal domain of the largest subunit of RNA polymerase II (Gullevrova et al., 2006). AtCYP59 also interacts with the conserved sequence of unprocessed mRNA (G[U/C]N[G/A]CC[A/G]), and binding with this specific RNA sequence inhibits the PPIase activity in vitro (Bannikova et al., 2013). Based on the functional association of AtCYP59 in transcription regulation, I speculate that the two multi-domain soybean CYPs, GmCYP56 and GmCYP59, possibly play a role in regulation of transcription in soybean via their RRM.

4.1 At least 15 GmCYPs are transcribed in soybean

Fifteen of 62 GmCYPs have 100% sequence identity with TC sequences in the soybean EST database, with 100% query coverage (Appendix 1). This level of identity suggests that at least 25% of GmCYP genes are transcribed in different soybean tissues during normal growth and development or in response to stress. Additionally, 33 GmCYPs displayed greater than 95% sequence identity with TC sequences in the soybean EST database, however, with less than 100% query coverage. The differences observed here could be due to the cultivar specific differences in the two databases (EST and whole genome databases). The whole genome sequencing of soybean utilized the cultivar Williams 82 whereas the DFCI soybean gene index was created from the EST data.
obtained from several cDNA libraries from different soybean cultivars. Taking into consideration of this fact, the number of transcribed \textit{GmCYPs} in soybean can be expected to be more than 15.

4.2 \textbf{GmCYP1 localizes in nucleus and cytoplasm}

A translational fusion of GmCYP1 with YFP illustrated its nuclear cytoplasmic localization in \textit{N. benthamiana} leaves (Figure 3.7). It was interesting to observe the nuclear localization of GmCYP1 despite the fact that it does not contain a known NLS. Molecules which are smaller in size than 20-40 kDa such as ions, water, proteins can pass through the nuclear pore complex by diffusion (Fried and Kutay, 2003), whereas movement of larger molecules (70 kDa or higher) entails an active transport system (D'Angelo et al., 2009), mediated by transport receptors and signal peptides (Harel and Forbes, 2004). The size of GmCYP1-YFP is 45.22 kDa, which is not considerably larger than the size of molecules that can pass through the nuclear pore complex by diffusion. Therefore, it is not clear whether GmCYP1-YFP localization in the nucleus was due to passive diffusion or to active transport. To confirm its nuclear localization, two YFP can be translationally fused with GmCYP1, which makes the total protein size 72.22 kDa (Lee et al., 2012). The mechanism of nuclear import/export of large proteins lacking NLS is by a piggy-back mechanism where cargo proteins temporarily bind with other transport proteins, and the complex translocates into or out from nucleus with the help of importins (Harel and Forbes, 2004). In summary, nuclear localization of GmCYP1, without having known NLS, indicates a possible role of GmCYP1 in regulation of gene expression.
4.3 GmMYB176 and GmCYP1 interact in planta

Previous work in Dr. Dhaubhadel’s laboratory demonstrated the interaction between GmCYP1 and GmMYB176 in Y2H assay (Li and Dhaubhadel, unpublished). However, this result could not be verified in the targeted Y2H assay between GmCYP1 and GmMYB176, which led to the hypothesis that there may be an indirect interaction between GmCYP1 and GmMYB176 via involvement of other proteins. To confirm that, a BiFC assay was carried out where both GmCYP1 and GmMYB176 proteins were translationally fused with either the N- or C- terminal half of YFP. The results revealed a strong interaction between GmCYP1 and GmMYB176 in the nucleus, and an extremely weak interaction in the cytoplasm (Figure 3.5).

It has been confirmed that GmMYB176 interacts with 14-3-3 proteins that affect localization of GmMYB176 (Li et al., 2012). Yaffe et al., (1997) reported two consensus 14-3-3 binding motifs, RSXpSXP (mode I) and RXY/FXpSXP (mode 2). However, some 14-3-3 binding proteins do not contain either of the above mentioned consensus binding sites (Aitken, 2006). In silico analysis revealed a putative phosphorylation site at serine/threonine residue (pST) within the GmCYP1 sequence ENFVKKHTGPGILSM (97-112), where T$_{105}$ is potentially phosphorylated. The interaction between GmCYP1 and GmMYB176 in Y2H and BiFC assays could be explained by GmCYP1 being a true client of a 14-3-3 protein. It is possible that 14-3-3 acts as a scaffold to facilitate binding of these two proteins. However, this speculation needs to be verified experimentally. Also, it is not clear whether the interaction between GmCYP1 and GmMYB176 leads to CHS8 gene expression, and subsequent isoflavonoid biosynthesis in soybean, or to some other, as-yet unknown function.
When the plasmids containing GmCYP1-YN and GmCYP1-YC were transiently co-expressed in *N. benthamiana* leaves and visualized by confocal microscopy, they showed a strong fluorescent signal in both the nucleus and the cytoplasm, suggesting that GmCYP1 monomers interact with each other and form a dimer *in planta*. There are no published literature on homo-dimerization of plant CYPs; however, recombinant hCYP-A has been reported to form monomers, dimers, and trimers in *E. coli* (Zhang et al., 2011). Even though the mechanism of GmCYP1 dimerization is not known yet, it can be speculated that if GmCYP1 is a client protein of 14-3-3 protein, binding of 14-3-3 with GmCYP1 could bring two GmCYP1 monomers together to produce fluorescence in the BiFC assay.

### 4.4 *GmCYPI* functions in seed development

Quantitative analysis of *GmCYPI* transcript accumulation in different soybean tissues showed expression of *GmCYPI* in all soybean tissues, but the level varied depending on tissue types (Figure 3.8). *GmCYPI* expression level were higher in embryos compared to that in other tissues, and reached its maximum when the embryos approached maturity (60 and 70 DAP). Interestingly, *ROC1* (AGI ID: At4g38740), the *Arabidopsis* ortholog of *GmCYPI*, was also found to have higher expression levels in seed tissues compared to that in other tissues (Figure 3.9). Moreover, the trend of a gradual increase in expression levels of *GmCYPI* and *ROC1* in different developmental stages of soybean, and *Arabidopsis* seeds, respectively, are similar. This similarity in expression pattern suggests a possibly conserved role of *GmCYPI* and *ROC1* in seed development.
4.5 GmCYP1 has conserved PPIase and CsA binding site

Sequence alignment with other previously characterized CYPs showed the conserved PPIase and CsA binding sites in GmCYP1 (Figure 3.4). Additionally, a signature sequence for a single domain plant CYPs, which consists of a seven amino acid stretch (RSGKPLH), was conserved in GmCYP1. This conserved sequence is thought to be associated with plant-specific function of CYPs. Two of the plant CYPs used in the alignment have a stress-specific function. Overexpression of cotton CYP (GhCYP1) in Arabidopsis or tobacco confers increased tolerance to biotic and abiotic stress (Zhu et al., 2011). Similarly, overexpression of pigeon pea CYP (ccCYP1) in Arabidopsis provides increased tolerance against multiple abiotic stresses, with increased PPIase activity (Sekhar et al., 2010). The high sequence identity of GmCYP1 with GhCYP1 and CcCYP1 (Table 3.2), suggests their possible function in stress response.

4.6 Functional characterization of GmCYP1 in Arabidopsis

To further characterize GmCYP1, Arabidopsis was used as a model plant. Due to unavailability of mutant lines for GmCYP1 (http://www.soybase.org/mutants/index.php), Arabidopsis orthologs of GmCYP1 were chosen for further characterization, assuming that they have similar role. Three Arabidopsis orthologs of GmCYP1 were identified as ROC1, ROC3 and ROC5 (Table 3.3) and their corresponding T-DNA insertion mutant lines were studied. Genotyping was carried out to identify homozygous T-DNA insertion mutant plants for each of the lines followed by transcript analysis. Transcription of ROC3 and ROC5 genes was completely disrupted in their respective homozygous T-DNA insertion mutant lines (Figure 3.11). However, reduced expression level of ROC1 transcript was detected in the roc1 mutant. The T-DNA insertion in roc3 and roc5
mutants was in exon region, whereas the insertion was in promoter region in \textit{roc1} line, and did not completely abolish \textit{ROC1} expression (Figure 3.10 and 3.11).

4.6.1 \textbf{ROC1, ROC3 and ROC5 have a role in seed germination in salt stress}

To study the possible role of \textit{GmCYP1} orthologs in \textit{Arabidopsis} towards the stress, salt stress was applied during the seed germination, and to the mature plants in separate experiments. All of the mutants (\textit{roc1}, \textit{roc3}, and \textit{roc5}) were sensitive towards salt during seed germination and their germination rate and efficiency decreased with the increase in salt concentration (Figure 3.14 and 3.15). Interestingly, when salt stress was applied to mature plants, the WT and the mutants did not respond differentially (Figure 3.16). These results clearly demonstrated a specific role of \textit{GmCYP1} orthologs, \textit{ROC1}, \textit{ROC3}, and \textit{ROC5}, in seed germination during salt stress.

Salt stress affects the physiology of a plant, resulting in improper seed germination and seedling development by damaging membrane organization, generating reactive oxygen species and causing metabolic toxicity due to excess ions (Parida and Das, 2005). The salt stress is caused by elevated concentration of Na\(^+\) and Cl\(^-\) ions inside the cell, that results into ionic toxicity and osmotic stress (Daszkowska-Golec, 2011). The increase in concentration of Na\(^+\) ions inside the cells creates ionic toxicity which are either exported out from plasma membrane by Na\(^+\)/K\(^+\) pump and Na\(^+\)/H\(^+\) antiporters or kept inside vacuole by tonoplast antiporters, Na\(^+\)/H\(^+\) exchanger (Blumwald et al., 2000).

Seed germination involves complex metabolic events which start with uptake of water by dry seeds, a process called imbibition. The water uptake by seeds is triphasic (Daszkowska-Golec, 2011). During phase I, uptake of water by dry seeds goes rapidly,
which makes the cells hydrated. The second phase, called the plateau phase, brings testa rupture. The third phase can be visualized by rupture of the endosperm and protrusion of the radicle. The amount of different phytohormones is perfectly regulated during different phases of seed germination. The transition from phase II to phase III involves the effect of hormonal as well as environmental change that lead to changes in the gene expression involved in the signal transduction of plant hormones (Daszkowska-Golec, 2011). The co-operation between different phytohormones and environmental stimuli has a crucial role for successful and timely germination of the seed. It is well established that the phytohormones abscisic acid (ABA) concentration decreases and gibberellic acid (GA) concentration increases in an antagonistic way during normal germination. However, in order to cope with stress and to break the arrest of seed germination in stress conditions, different metabolic pathways involving phytohormone biosynthesis and signal transduction pathways, chromatic modifications and microRNA post-transcriptional modification are accompanied (Daszkowska-Golec, 2011).

Recent findings on Arabidopsis seed germination under salt stress show the cross-talk between components of ABA and GA signaling pathway (Yuan et al., 2011). There are three transcription factors, ABI3, ABI4 and ABI5 that specifically expression of regulate ABA-related genes during seed germination (Finkelstein and Lynch, 2000; Lopez-Molina et al., 2002). In normal conditions, proteasome-mediated degradation of ABI3 and ABI5 facilitates the germination by decreasing the ABA content in germinating seeds. Similarly, GA promotes the proteasome-mediated degradation of RGL2, a negative regulator of GA signaling, in order to increase the GA content to facilitate the germination. In salt stress condition, these two pathways communicate each other to
facilitate the *Arabidopsis* seed germination (Yuan et al., 2011). The delay in germination of WT seeds in salt stress condition is correlated with increased transcript levels of *ABI3*, *ABI5*, and *RGL2*. I speculate that PPIase activity of ROC1, ROC3, and ROC5 might be associated with proper folding or activation of proteins required for the salt stress specific cross-talk between ABI3, ABI5, and RGL2 on delayed germination of *roc1*, *roc3*, and *roc5* in salt stress condition.

**4.6.2 roc3 and roc5 shows early flowering phenotype under long day photoperiods**

To study phenotypic differences elicited by disruption of *GmCYP1* orthologs (*roc1*, *roc3* and *roc5*) in *Arabidopsis*, both WT and mutants were grown as described in section 2.2. Previously, it was reported that *roc1* shows an early flowering phenotype under long day conditions, but not under short days (Trupkin et al., 2012). Consistent with this phenotype, two other mutants, *roc3* and *roc5*, also showed early flowering phenotype under long day conditions, but not under short days (Figure 3.12 and 3.13). However, it is not known whether or not the *GmCYP1* can complement the early flowering phenotype of *roc1*, *roc3*, and *roc5*.

Flowering is a major developmental transition from vegetative to reproductive stage. The correct timing of flowering determines the reproductive success of plants. Plants have evolved a complex network of pathways that respond to different environmental and endogenous cues to maximize reproductive success (Simpson and Dean, 2002). In *Arabidopsis*, there are five flowering pathways known yet (Henderson and Dean, 2004; Simpson and Dean, 2002; Srikanth and Schmid, 2011). They are
photoperiodic pathway, gibberellin pathway, autonomous pathway, vernalization pathway, and endogenous (or ageing) pathway. All of these pathways activate the set of downstream floral pathway integrator genes, which activate the downstream floral meristem identity (FMI) genes. FMI genes encode proteins that promote the transition from vegetative to reproductive stage.

Both roc3 and roc5 show early flowering under long day condition but not under short day, a photoperiod-specific flowering time phenotype (Figure 3.12 and 3.13). Photoperiod pathway is dependent on day length, where many angiosperms start flowering with change in day and night lengths. In Arabidopsis, a long day photoperiod promotes flowering by activation of CONSTANS (CO) transcription factor (Putterill et al., 1995). CO transcript levels are controlled by the plant circadian clock and show rhythmic, diurnal expression (Suarez-Lopez et al., 2001). CO activates the floral integrator gene, FLOWERING LOCUS T (FT) (Samach et al., 2000) which further activates the downstream FMI genes. Another protein with key role in the autonomous pathway called FLOWERING LOCUS C (FLC) also plays central role in the photoperiod pathway by suppressing the activation of FT (Helliwell et al., 2006). There exists an antagonistic activity between FLC and CO in controlling flowering time in photoperiodic pathway. It can be speculated that ROC1, ROC3 or ROC5 either down-regulate CO or up-regulate FLC in order to maintain the normal flowering time in WT Arabidopsis plants. The mechanism of early flowering in roc1, roc3, and roc5 can be further studied by at least monitoring the transcript level of CO, FLC and FT genes in both WT and mutants.
In summary, the present study demonstrated that soybean genome contains the largest \textit{CYP} gene family identified so far, with 62 unique members. GmCYP1 localized to the nucleus and cytoplasm, and its transcripts accumulated at higher level in the seed tissue. GmCYP1 interacted with GmMYB176 \textit{in planta} and is the first plant CYP that has been shown to form a dimer \textit{in vivo}. GmCYP1 orthologs in \textit{Arabidopsis} were shown to function in the transition from vegetative to reproductive phase, and during the seed germination upon salt stress.
Identification of the soybean CYP gene family is the first step in exploring the potentially important CYPs that are involved in different pathways of plant growth and development. As many of the plant CYPs are involved in different types of stress responses and/or other cellular pathways, characterization of GmCYPs will be helpful in finding the candidate CYPs that can be used for biotechnological application.

Regarding the subcellular localization of GmCYP1, where translational fusion of GmCYP1 with YFP was used, the nuclear localization of GmCYP1-YFP could be either by passive diffusion or by active transport. In order to confirm this, two YFP proteins could be fused with GmCYP1 in order to generate a protein of 72.22 kDa, which could not pass through the nuclear pore complex by passive diffusion (D'Angelo et al., 2009; Lee et al., 2012).

It is reported that 14-3-3 proteins interact with GmMYB176 and affect the GmMYB176 localization and isoflavonoid biosynthesis in soybean (Li et al., 2012). It is speculated in this thesis that in planta interaction between GmMYB176 and GmCYP1 is possibly facilitated via 14-3-3 proteins, as there is a putative 14-3-3 binding site in GmCYP1. The interaction between 14-3-3 and GmCYP1 can be tested by Y2H and/or BiFC assay. To further confirm this, deletion of the putative 14-3-3 binding site in GmCYP1 can be performed, followed by BiFC assay between the mutant GmCYP1 and GmMYB176. Additionally, this experiment will provide insights on in planta dimer formation of GmCYP1.
As shown in Figure 3.12 and 3.13, Arabidopsis orthologs of GmCYP1 are involved in the transition from the vegetative to reproductive stage. The functional role of GmCYP1 in the flowering pathway can be further studied by the development of roc1, roc3 and roc5 complementation lines expressing GmCYP1. It is speculated in this thesis that both ROC3 and ROC5 are involved in the photoperiod pathway. Experimental evidence can be gathered by studying transcript levels of flowering marker genes like CO, FLC and FT. Similarly, all of the mutants, roc1, roc3, and roc5, were sensitive to salt during seed germination. The potential role of GmCYP1 in seed germination can be studied by generating the complementation lines expressing GmCYP1. It is not known how ROC1, ROC3 and ROC5 are involved in seed germination under salt stress.

Generation of double and/or triple mutants may give a clue into whether they work via independent pathways or through the same pathway. Experimental evidence towards possible role of GmCYP1 orthologs in ABA/GA signaling crosstalk can be generated by studying the transcript level of ABI3, ABI4, ABI5, and RGL2 (Yuan et al., 2011). Last but not the least; GmCYP1 should be overexpressed in Arabidopsis WT to study its role in other biotic and abiotic stresses.
References


Brandts JF, Halvorson HR and Brennan M (1975) Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry* **14**:4953-4963.


Li X, Chen L and Dhaubhadel S (2012) 14-3-3 proteins regulate the intracellular localization of the transcriptional activator GmMYB176 and affect isoflavonoid synthesis in soybean. *Plant J* 71:239-250.


## Appendix

### Appendix 1 Soybean *cyclophilin* gene family.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Locus Name</th>
<th>Gene Location</th>
<th>Strand Information</th>
<th>Transcript Size (bp)</th>
<th>Protein Size (amino acid)</th>
<th>Protein Mol Weight (kD)</th>
<th>Predicted Sub-cellular Location</th>
<th>Domain Information</th>
<th>Corresponding TC number (% Identity)</th>
<th>Best Arabidopsis TAIR10 hit</th>
</tr>
</thead>
<tbody>
<tr>
<td>GmCYP1</td>
<td>Glyma11g10480</td>
<td>Gm11: 7494447 – 7495419</td>
<td>+</td>
<td>973</td>
<td>172</td>
<td>18.22</td>
<td>-</td>
<td>SD</td>
<td>TC464530 (100)</td>
<td>AT2G16600.1</td>
</tr>
<tr>
<td>GmCYP2</td>
<td>Glyma12g02790</td>
<td>Gm12: 1802211 - 1803279</td>
<td>+</td>
<td>1069</td>
<td>172</td>
<td>18.12</td>
<td>-</td>
<td>SD</td>
<td>TC422035 (100)</td>
<td>AT2G16600.1</td>
</tr>
<tr>
<td>GmCYP3</td>
<td>Glyma06g00740</td>
<td>Gm06: 426934 - 427719</td>
<td>-</td>
<td>786</td>
<td>172</td>
<td>18.193</td>
<td>-</td>
<td>SD</td>
<td>TC428777 (100)</td>
<td>AT2G16600.1</td>
</tr>
<tr>
<td>GmCYP4</td>
<td>Glyma04g00700</td>
<td>Gm04: 400295 - 401029</td>
<td>-</td>
<td>735</td>
<td>172</td>
<td>18.197</td>
<td>-</td>
<td>SD</td>
<td>TC428777 (94)</td>
<td>AT2G16600.1</td>
</tr>
<tr>
<td>GmCYP5</td>
<td>Glyma02g02580</td>
<td>Gm02: 1907237 - 1907629</td>
<td>-</td>
<td>354</td>
<td>117</td>
<td>13.234</td>
<td>-</td>
<td>SD</td>
<td>TC464530 (84)</td>
<td>AT2G16600.1</td>
</tr>
<tr>
<td>GmCYP6</td>
<td>Glyma15g38980</td>
<td>Gm15: 45488205 - 45488602</td>
<td>-</td>
<td>398</td>
<td>131</td>
<td>14.808</td>
<td>-</td>
<td>SD</td>
<td>TC422035 (80)</td>
<td>AT2G16600.1</td>
</tr>
<tr>
<td>GmCYP7</td>
<td>Glyma03g41210</td>
<td>Gm03: 46735997 - 46737827</td>
<td>-</td>
<td>1138</td>
<td>175</td>
<td>18.884</td>
<td>-</td>
<td>SD</td>
<td>TC429055 (100)</td>
<td>AT3G56070.1</td>
</tr>
<tr>
<td>GmCYP8</td>
<td>Glyma05g10100</td>
<td>Gm05: 10132037 - 10135654</td>
<td>+</td>
<td>1577</td>
<td>360</td>
<td>40.015</td>
<td>-</td>
<td>MD</td>
<td>TC443482 (100)</td>
<td>AT2G15790.1</td>
</tr>
<tr>
<td>GmCYP9</td>
<td>Glyma17g20430</td>
<td>Gm17: 19440629 - 19445366</td>
<td>-</td>
<td>1241</td>
<td>360</td>
<td>40.12</td>
<td>-</td>
<td>MD</td>
<td>TC443482 (95)*</td>
<td>AT2G15790.1</td>
</tr>
<tr>
<td>GmCYP10</td>
<td>Glyma20g39340</td>
<td>Gm20: 46685758 - 46688282</td>
<td>-</td>
<td>1260</td>
<td>253</td>
<td>27.103</td>
<td>Chloroplast</td>
<td>SD</td>
<td>TC462161 (100)</td>
<td>AT5G13120.1</td>
</tr>
<tr>
<td>GmCYP11</td>
<td>Glyma19g43830</td>
<td>Gm19: 49391997 - 49393557</td>
<td>-</td>
<td>890</td>
<td>175</td>
<td>18.955</td>
<td>-</td>
<td>SD</td>
<td>TC422384 (100)</td>
<td>AT3G56070.1</td>
</tr>
<tr>
<td>GmCYP12</td>
<td>Glyma03g38450</td>
<td>Gm03: 44797568 - 44800364</td>
<td>-</td>
<td>768</td>
<td>255</td>
<td>28.755</td>
<td>Chloroplast</td>
<td>SD</td>
<td>TC428054 (82)</td>
<td>AT3G62030.3</td>
</tr>
<tr>
<td>GmCYP13</td>
<td>Glyma03g35620</td>
<td>Gm03: 42773255 - 42775949</td>
<td>-</td>
<td>793</td>
<td>164</td>
<td>18.153</td>
<td>-</td>
<td>SD</td>
<td>TC425919 (100)</td>
<td>AT2G36130.1</td>
</tr>
<tr>
<td>GmCYP14</td>
<td>Glyma19g41050</td>
<td>Gm19: 47359304 - 47361799</td>
<td>-</td>
<td>1078</td>
<td>260</td>
<td>28.32</td>
<td>Chloroplast</td>
<td>SD</td>
<td>TC428054 (100)</td>
<td>AT3G62030.3</td>
</tr>
</tbody>
</table>
### Appendix 1, continued

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Locus Name</th>
<th>Gene Location</th>
<th>Strand Information</th>
<th>Transcript Size (bp)</th>
<th>Protein Size (amino acid)</th>
<th>Protein Mol Weight (kD)</th>
<th>Predicted Sub-cellular Location</th>
<th>Domain Information</th>
<th>Corresponding TC number (% Identity)</th>
<th>Best Arabidopsis TAIR10 hit</th>
</tr>
</thead>
<tbody>
<tr>
<td>GmCYP15</td>
<td>Glyma19g38260</td>
<td>Gm19: 45219846 - 45222425</td>
<td>-</td>
<td>798</td>
<td>165</td>
<td>18.303</td>
<td>-</td>
<td>SD</td>
<td>TC430194 (99)</td>
<td>AT2G36130.1</td>
</tr>
<tr>
<td>GmCYP16</td>
<td>Glyma07g33200</td>
<td>Gm07: 38125958 - 38130437</td>
<td>+</td>
<td>1371</td>
<td>361</td>
<td>40.247</td>
<td>-</td>
<td>MD</td>
<td>TC460136 (97)*</td>
<td>AT2G15790.1</td>
</tr>
<tr>
<td>GmCYP17</td>
<td>Glyma02g15250</td>
<td>Gm02: 13761866 - 13766845</td>
<td>-</td>
<td>1720</td>
<td>361</td>
<td>40.205</td>
<td>-</td>
<td>MD</td>
<td>TC460136 (100)</td>
<td>AT2G15790.1</td>
</tr>
<tr>
<td>GmCYP18</td>
<td>Glyma11g05130</td>
<td>Gm11: 3560583 - 3565430</td>
<td>-</td>
<td>2034</td>
<td>597</td>
<td>65.371</td>
<td>Nucleus</td>
<td>SD</td>
<td>TC421742 (96)*</td>
<td>AT5G67530.1</td>
</tr>
<tr>
<td>GmCYP19</td>
<td>Glyma01g40170</td>
<td>Gm01: 51956948 - 51961655</td>
<td>+</td>
<td>2192</td>
<td>597</td>
<td>65.205</td>
<td>Nucleus</td>
<td>SD</td>
<td>TC421742 (97)*</td>
<td>AT5G67530.1</td>
</tr>
<tr>
<td>GmCYP20</td>
<td>Glyma13g39500</td>
<td>Gm13: 40090324 - 40096203</td>
<td>-</td>
<td>2518</td>
<td>616</td>
<td>69.361</td>
<td>-</td>
<td>MD</td>
<td>TC458065 (99)*</td>
<td>AT3G44600.1</td>
</tr>
<tr>
<td>GmCYP21</td>
<td>Glyma09g11960</td>
<td>Gm09: 12253982 - 12259079</td>
<td>+</td>
<td>876</td>
<td>194</td>
<td>20.965</td>
<td>-</td>
<td>SD</td>
<td>TC422604 (99)*</td>
<td>AT2G38730.1</td>
</tr>
<tr>
<td>GmCYP22</td>
<td>Glyma13g26260</td>
<td>Gm13: 29469229 - 29475935</td>
<td>+</td>
<td>902</td>
<td>179</td>
<td>19.367</td>
<td>-</td>
<td>SD</td>
<td>TC422604 (99)*</td>
<td>AT2G38730.1</td>
</tr>
<tr>
<td>GmCYP23</td>
<td>Glyma10g44560</td>
<td>Gm10: 50887100 - 50889483</td>
<td>-</td>
<td>1268</td>
<td>265</td>
<td>28.424</td>
<td>Chloroplast</td>
<td>SD</td>
<td>TC465254 (100)*</td>
<td>AT5G13120.1</td>
</tr>
<tr>
<td>GmCYP24</td>
<td>Glyma18g07030</td>
<td>Gm18: 5730116 - 5732839</td>
<td>+</td>
<td>681</td>
<td>226</td>
<td>24.348</td>
<td>-</td>
<td>SD</td>
<td>TC451996 (100)*</td>
<td>AT2G29960.1</td>
</tr>
<tr>
<td>GmCYP25</td>
<td>Glyma18g03120</td>
<td>Gm18: 2059860 - 2062743</td>
<td>+</td>
<td>1046</td>
<td>235</td>
<td>25.72</td>
<td>Secretory</td>
<td>SD</td>
<td>TC459271 (99)*</td>
<td>AT3G55920.1</td>
</tr>
<tr>
<td>GmCYP26</td>
<td>Glyma11g35280</td>
<td>Gm11: 36969574 - 36973449</td>
<td>-</td>
<td>1630</td>
<td>238</td>
<td>25.831</td>
<td>Secretory</td>
<td>SD</td>
<td>TC470206 (100)</td>
<td>AT3G55920.1</td>
</tr>
<tr>
<td>GmCYP27</td>
<td>Glyma19g34290</td>
<td>Gm19: 41904818 - 41910607</td>
<td>+</td>
<td>2597</td>
<td>659</td>
<td>73.522</td>
<td>Nucleus</td>
<td>SD</td>
<td>TC461722 (99)*</td>
<td>AT3G63400.1</td>
</tr>
<tr>
<td>GmCYP28</td>
<td>Glyma10g27990</td>
<td>Gm10: 36797008 - 36800396</td>
<td>-</td>
<td>1191</td>
<td>263</td>
<td>29.017</td>
<td>Chloroplast</td>
<td>SD</td>
<td>TC424043 (99)*</td>
<td>AT3G62030.3</td>
</tr>
<tr>
<td>GmCYP29</td>
<td>Glyma18g48350</td>
<td>Gm18: 57781870 - 57786498</td>
<td>-</td>
<td>1152</td>
<td>326</td>
<td>36.367</td>
<td>Secretory</td>
<td>SD</td>
<td>TC459175 (96)*</td>
<td>AT4G17070.1</td>
</tr>
<tr>
<td>GmCYP30</td>
<td>Glyma09g38040</td>
<td>Gm09: 43528564 - 43534075</td>
<td>+</td>
<td>1411</td>
<td>327</td>
<td>36.694</td>
<td>Secretory</td>
<td>SD</td>
<td>TC459175 (96)*</td>
<td>AT4G17070.1</td>
</tr>
<tr>
<td>GmCYP31</td>
<td>Glyma01g41270</td>
<td>Gm01: 52827230 - 52832563</td>
<td>+</td>
<td>1497</td>
<td>337</td>
<td>37.089</td>
<td>Mitochondria #</td>
<td>SD</td>
<td>TC428737 (96)*</td>
<td>AT4G17070.1</td>
</tr>
<tr>
<td>GmCYP32</td>
<td>Glyma11g04140</td>
<td>Gm11: 2753246 - 27582111</td>
<td>-</td>
<td>1446</td>
<td>337</td>
<td>37.009</td>
<td>Mitochondria #</td>
<td>SD</td>
<td>TC428737 (99)*</td>
<td>AT4G17070.1</td>
</tr>
</tbody>
</table>
## Appendix 1, continued

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Locus Name</th>
<th>Gene Location</th>
<th>Strand information</th>
<th>Transcript size (bp)</th>
<th>Protein size (amino acid)</th>
<th>Protein Mol Weight (kD)</th>
<th>Predicted Sub-cellular Location</th>
<th>Domain information</th>
<th>Corresponding TC number (% Identity)</th>
<th>Best Arabidopsis TAIR10 hit</th>
</tr>
</thead>
<tbody>
<tr>
<td>GmCYP33</td>
<td>Glyma17g15850</td>
<td>Gm17: 12578011 - 12581421</td>
<td>-</td>
<td>984</td>
<td>328</td>
<td>36.229</td>
<td>Secretry</td>
<td>SD</td>
<td>TC433463 (85)</td>
<td>AT4G17070.1</td>
</tr>
<tr>
<td>GmCYP34</td>
<td>Glyma11g27000</td>
<td>Gm11: 26579046 - 26585230</td>
<td>-</td>
<td>1048</td>
<td>204</td>
<td>21.944</td>
<td>Secretry</td>
<td>SD</td>
<td>TC423512 (100)</td>
<td>AT2G29960.1</td>
</tr>
<tr>
<td>GmCYP35</td>
<td>Glyma12g30780</td>
<td>Gm12: 34416244 - 34421927</td>
<td>+</td>
<td>2351</td>
<td>616</td>
<td>69.326</td>
<td>-</td>
<td>MD</td>
<td>TC458065 (99)*</td>
<td>AT3G44600.1</td>
</tr>
<tr>
<td>GmCYP36</td>
<td>Glyma03g31440</td>
<td>Gm03: 39350303 - 39356504</td>
<td>+</td>
<td>2514</td>
<td>668</td>
<td>74.261</td>
<td>Nucleus</td>
<td>SD</td>
<td>TC461219 (97)*</td>
<td>AT3G63400.1</td>
</tr>
<tr>
<td>GmCYP37</td>
<td>Glyma17g33430</td>
<td>Gm17: 37316059 - 37322598</td>
<td>-</td>
<td>1972</td>
<td>493</td>
<td>55.642</td>
<td>-</td>
<td>SD</td>
<td>TC443202 (100)*</td>
<td>AT4G33060.1</td>
</tr>
<tr>
<td>GmCYP38</td>
<td>Glyma06g00650</td>
<td>Gm06: 363388 - 364747</td>
<td>-</td>
<td>600</td>
<td>199</td>
<td>21.845</td>
<td>-</td>
<td>SD</td>
<td>TC446906 (84)</td>
<td>AT4G34960.1</td>
</tr>
<tr>
<td>GmCYP39</td>
<td>Glyma04g00580</td>
<td>Gm04: 335346 - 339587</td>
<td>-</td>
<td>793</td>
<td>232</td>
<td>25.520</td>
<td>Secretry</td>
<td>SD</td>
<td>EV278562 (100)*</td>
<td>AT4G34960.1</td>
</tr>
<tr>
<td>GmCYP40</td>
<td>Glyma12g03540</td>
<td>Gm12: 2349948 - 2355775</td>
<td>+</td>
<td>1178</td>
<td>236</td>
<td>26.182</td>
<td>Secretry</td>
<td>SD</td>
<td>TC446906 (100)*</td>
<td>AT4G34960.1</td>
</tr>
<tr>
<td>GmCYP41</td>
<td>Glyma11g11370</td>
<td>Gm11: 8098156 - 8103774</td>
<td>+</td>
<td>1204</td>
<td>236</td>
<td>26.151</td>
<td>Secretry</td>
<td>SD</td>
<td>TC428887 (100)</td>
<td>AT4G34960.1</td>
</tr>
<tr>
<td>GmCYP42</td>
<td>Glyma01g34630</td>
<td>Gm01: 47060025 - 47069273</td>
<td>-</td>
<td>1042</td>
<td>165</td>
<td>18.082</td>
<td>-</td>
<td>SD</td>
<td>TC429931 (100)*</td>
<td>AT1G01940.1</td>
</tr>
<tr>
<td>GmCYP43</td>
<td>Glyma06g07380</td>
<td>Gm06: 5375484 - 53811118</td>
<td>-</td>
<td>2975</td>
<td>870</td>
<td>98.357</td>
<td>Nucleus</td>
<td>SD</td>
<td>TC450817 (89)</td>
<td>AT3G63400.1</td>
</tr>
<tr>
<td>GmCYP44</td>
<td>Glyma01g05320</td>
<td>Gm01: 5005455 - 5010050</td>
<td>-</td>
<td>728</td>
<td>241</td>
<td>27.29</td>
<td>-</td>
<td>SD</td>
<td>TC479526 (92)</td>
<td>AT1G26940.1</td>
</tr>
<tr>
<td>GmCYP45</td>
<td>Glyma02g11730</td>
<td>Gm02: 9920653 - 9927355</td>
<td>+</td>
<td>1042</td>
<td>226</td>
<td>25.613</td>
<td>Secretry</td>
<td>SD</td>
<td>TC479526 (100)</td>
<td>AT1G26940.1</td>
</tr>
<tr>
<td>GmCYP46</td>
<td>Glyma04g07300</td>
<td>Gm04: 5667892 - 5674523</td>
<td>-</td>
<td>3097</td>
<td>805</td>
<td>90.834</td>
<td>Nucleus</td>
<td>SD</td>
<td>TC450817 (84)</td>
<td>AT3G63400.1</td>
</tr>
<tr>
<td>GmCYP47</td>
<td>Glyma05g07830</td>
<td>Gm05: 7783152 - 7784581</td>
<td>-</td>
<td>1257</td>
<td>387</td>
<td>42.016</td>
<td>-</td>
<td>SD</td>
<td>TC435070 (99)*</td>
<td>AT1G74070.1</td>
</tr>
<tr>
<td>GmCYP48</td>
<td>Glyma06g38470</td>
<td>Gm06: 41409385 - 41415353</td>
<td>-</td>
<td>1584</td>
<td>439</td>
<td>47.320</td>
<td>Chloroplast</td>
<td>SD</td>
<td>TC429384 (97)*</td>
<td>AT3G15520.1</td>
</tr>
<tr>
<td>GmCYP49</td>
<td>Glyma07g19450</td>
<td>Gm07: 19399481 - 19403975</td>
<td>+</td>
<td>921</td>
<td>243</td>
<td>27.729</td>
<td>Mitochondria</td>
<td>SD</td>
<td>TC452908 (81)</td>
<td>AT3G66654.1</td>
</tr>
<tr>
<td>GmCYP50</td>
<td>Glyma09g12590</td>
<td>Gm09: 13634138 - 13634895</td>
<td>-</td>
<td>705</td>
<td>234</td>
<td>25.476</td>
<td>Chloroplast</td>
<td>SD</td>
<td>TC493586 (95)*</td>
<td>AT5G35100.1</td>
</tr>
</tbody>
</table>
Appendix 1, continued

<table>
<thead>
<tr>
<th>Gene  Name</th>
<th>Locus Name</th>
<th>Gene Location</th>
<th>Strand Information</th>
<th>Transcript Size (bp)</th>
<th>Protein Size (amino acid)</th>
<th>Protein Mol Weight (kD)</th>
<th>Predicted Sub-cellular Location</th>
<th>Domain Information</th>
<th>Corresponding TC number (% Identity)</th>
<th>Best Arabidopsis TAIR10 hit</th>
</tr>
</thead>
<tbody>
<tr>
<td>GmCYP51</td>
<td>Glyma10g35030</td>
<td>Gm10: 43184361 - 43192370</td>
<td>+</td>
<td>1392</td>
<td>232</td>
<td>26.174</td>
<td>Mitochondria</td>
<td>SD</td>
<td>TC482811 (99)*</td>
<td>AT3G66654.1</td>
</tr>
<tr>
<td>GmCYP52</td>
<td>Glyma12g22700</td>
<td>Gm12: 24729849 - 24736879</td>
<td>+</td>
<td>1592</td>
<td>439</td>
<td>47.508</td>
<td>Chloroplast</td>
<td>SD</td>
<td>TC429384 (94)</td>
<td>AT3G15520.1</td>
</tr>
<tr>
<td>GmCYP53</td>
<td>Glyma13g23850</td>
<td>Gm13: 27156006 - 27164451</td>
<td>-</td>
<td>1968</td>
<td>445</td>
<td>48.730</td>
<td>Chloroplast</td>
<td>SD</td>
<td>TC431861 (96)*</td>
<td>AT3G01480.1</td>
</tr>
<tr>
<td>GmCYP54</td>
<td>Glyma14g14970</td>
<td>Gm14: 15684452 - 15696919</td>
<td>-</td>
<td>2999</td>
<td>886</td>
<td>99.69</td>
<td>Nucleus</td>
<td>SD</td>
<td>TC471889(97)*</td>
<td>AT3G63400.1</td>
</tr>
<tr>
<td>GmCYP55</td>
<td>Glyma15g24560</td>
<td>Gm15: 25322384 - 25325646</td>
<td>+</td>
<td>1010</td>
<td>286</td>
<td>31.234</td>
<td>Chloroplast</td>
<td>SD</td>
<td>TC461630 (99)*</td>
<td>AT5G35100.1</td>
</tr>
<tr>
<td>GmCYP56</td>
<td>Glyma15g28380</td>
<td>Gm15: 31058468 - 31086596</td>
<td>-</td>
<td>2332</td>
<td>633</td>
<td>74.013</td>
<td>Nucleus</td>
<td>MD</td>
<td>TC481910 (98)*</td>
<td>AT1G53720.1</td>
</tr>
<tr>
<td>GmCYP57</td>
<td>Glyma15g37190</td>
<td>Gm15: 42829057 - 42837754</td>
<td>+</td>
<td>531</td>
<td>176</td>
<td>19.419</td>
<td>-</td>
<td>SD</td>
<td>TC422604 (95)*</td>
<td>AT2G38730.1</td>
</tr>
<tr>
<td>GmCYP58</td>
<td>Glyma17g13180</td>
<td>Gm17: 10089614 - 10091393</td>
<td>+</td>
<td>1608</td>
<td>350</td>
<td>38.183</td>
<td>Chloroplast</td>
<td>SD</td>
<td>TC435070 (94)</td>
<td>AT1G74070.1</td>
</tr>
<tr>
<td>GmCYP59</td>
<td>Glyma19g00460</td>
<td>Gm19: 247582 - 263127</td>
<td>-</td>
<td>1830</td>
<td>562</td>
<td>65.707</td>
<td>Nucleus</td>
<td>MD</td>
<td>TC481910(76)</td>
<td>AT1G53720.1</td>
</tr>
<tr>
<td>GmCYP60</td>
<td>Glyma19g01260</td>
<td>Gm19: 870876 – 876246</td>
<td>-</td>
<td>1860</td>
<td>445</td>
<td>48.731</td>
<td>Chloroplast</td>
<td>SD</td>
<td>TC431861 (99)*</td>
<td>AT3G01480.1</td>
</tr>
<tr>
<td>GmCYP61</td>
<td>Glyma20g00820</td>
<td>Gm20: 537200 - 540414</td>
<td>+</td>
<td>684</td>
<td>227</td>
<td>25.696</td>
<td>Mitochondria</td>
<td>SD</td>
<td>TC452908 (86)</td>
<td>AT3G66654.1</td>
</tr>
<tr>
<td>GmCYP62</td>
<td>Glyma20g32530</td>
<td>Gm20: 41142287 - 41147574</td>
<td>-</td>
<td>1489</td>
<td>292</td>
<td>33.172</td>
<td>-</td>
<td>SD</td>
<td>TC452908 (99)*</td>
<td>AT3G66654.1</td>
</tr>
</tbody>
</table>

SD, single domain; MD, multi-domain; (#), prediction with less confidence; (-), no signal peptide detected; (*), contigs with >95% sequence identity but with query coverage <100%; Underline, contigs with 99-100% sequence identity with 100% sequence coverage.
CURRICULUM VITAE

Hemanta Raj Mainali

Education

The University of Western Ontario
London, Ontario, Canada
2011 – 2013, M. Sc. in Biology

Kathmandu University
Dhulikhel, Kavre, Nepal

Awards and Scholarships

2011 – 2013 Western Graduate Research Scholarship
2010 Best Poster Award, 2nd International Symposium on
Biotechnology, Kathmandu, Nepal
2009 Vice Chancellor Gold Medal, Kathmandu University, Nepal

Work Experience

2011 – 2013 MSc. Graduate Student, University of Western Ontario
2011 – 2013 Graduate Teaching Assistant, University of Western Ontario
2010-2011 Teaching Assistant, Kathmandu University, Nepal
2009 Volunteer, Centre for Molecular Dynamics, and Intrepid Nepal
**Poster Presentation**

