Graduate School ETD Form 9
(Revised 01/14)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By PALLABI ROY

Entitled
MOLECULAR CLONING OF THE SOYBEAN PHOTOTROPINS

For the degree of Master of Science

Is approved by the final examining committee:

Dr. JOHN C. WATSON

Dr. STEPHEN K. RANDALL

Dr. GREGORY G. ANDERSON

To the best of my knowledge and as understood by the student in the *Thesis/Dissertation Agreement. Publication Delay, and Certification/Disclaimer (Graduate School Form 32)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Dr. JOHN C. WATSON

Approved by Major Professor(s): _____

Approved by: Dr. SIMON J. ATKINSON 07/08/2014

Head of the Department Graduate Program Date

MOLECULAR CLONING OF THE SOYBEAN PHOTOTROPINS

A Thesis

Submitted to the Faculty

of

Purdue University

by

Pallabi Roy

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

August 2014

Purdue University

Indianapolis, Indiana

This work is dedicated to my parents and my husband

ACKNOWLEDGEMENT

I would like to thank Dr. John Watson for providing me with this wonderful opportunity of working in this lab. The immense support and guidance that he had provided me with are invaluable and commendable. I would also like to express my gratitude towards Dr. Stephen Randall and Dr. Gregory Anderson for their precious suggestions and helpful insights on this work. I appreciate the input of Kay Sherfick and contributions of Benjamin Rains and Elizabeth Lee in this work. I thank the department of Biology for providing me the required equipments for my project.

Finally, I would like to thank my family, especially my parents, Mr. Prabir Kumar Roy and Mrs. Ruma Roy, for their love, support and blessings, and my husband, Shubham Chakravarty for being by my side and encouraging me throughout my work.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

LIST OF ABBREVIATIONS

A_{600} ..... Absorbance at 600 nm

Att ..... Attachment

B, R ..... blue, red light

cDNA ..... complementary DNA

cry ..... Cryptochrome

dUTP ..... deoxyuridine triphosphate

dUMP ..... deoxyuridine monophosphate

EDTA ..... Ethylenediaminetetraaceticacid

EGFP ..... Enhanced Green fluorescent protein

FMN ..... Flavin mononucleotide

GFP ..... Green fluorescent protein

kDa ..... Kilo Dalton

LB ..... Luria-Bertani media

LOV ..... Light, Oxygen, Voltage

mRNA ..... messenger RNA

NLS ..... Nuclear localization signal

NPH1 ..... Non phototrophic hypocotyl 1

NPL1 ..... NPH1-like

ORF ..... Open reading frame

PAS ..... Per, ARNT, Sim

PCR ..... Polymerase chain reaction

PHOT ..... Phototropin

Phy ..... Phytochrome

PIN ..... Pin-formed

PsPK ..... *Pisum sativum* protein kinase

RT-PCR ..... Reverse transcription PCR

Ser ..... Serine

ABCB19 ..... ATP Binding Cassette B19

S.O.C ..... Super Optimal Catabolite

TAE ..... Tris-acetic acid-EDTA

T-DNA ..... Transfer DNA

TE.....Tris EDTA

Ti.....Tumor inducing

T4SS.....Type IV secretion system

UV.....Ultra violet

vir.....Virulence

ABSTRACT

Roy, Pallabi M.S. Purdue University, August 2014. Molecular Cloning of the Soybean Phototropins. Major Professor: John C. Watson

The phototropin photoreceptors are important regulators of plant growth and development and can therefore affect the photosynthetic activity of plants. Phototropin1 and Phototropin2 are versatile protein kinases that become activated when exposed to blue light. Their photobiological actions are best understood in the model plant *Arabidopsis thaliana*, where they are known to trigger several responses to blue light, one of which is phototropism, the bending of plant organs towards light. Additionally, phot1 and phot2 drive stomatal opening, chloroplast arrangement in leaf cells, leaf expansion, and leaf orientation. The phot1-specific response is rapid inhibition of hypocotyl growth, leaf positioning and mRNA stability whereas phot2 mediates the chloroplast avoidance response to high light. These responses impact a plant's ability to capture light for photosynthesis, therefore the phototropins play important roles in optimizing a plant's photosynthetic activity.

Soybean (*Glycine max*) is a very important crop plant in Indiana known for its nutritional versatility and is also utilized for biodiesel production. In spite of soybean being a key crop, there is currently no information about the functionality of soybean phototropins.

Also, being a legume, soybean has many structural and functional features that are not present in *Arabidopsis*. Interestingly, PsPHOT1A (a photoreceptor from garden pea) was found to be a functional phototropin as it was able to complement the *phot1* mutation in *Arabidopsis*. The roles of these proteins in soybean will be elucidated based on the hypothesis that soybean phototropins play essential roles in regulating photosynthetic activity as do the *Arabidopsis* phototropins.

To date, five soybean phototropins, 3 PHOT1s and 2 PHOT2s, are believed to exist. These *GmPHOT* protein coding regions were amplified by RT-PCR and cloned into pCR8/TOPO or pENTR-D/TOPO vectors via TOPO cloning to utilize Gateway cloning technology to create plant transformation constructs subsequently. The cloned *GmPHOT* cDNAs from each of the 5 *GmPHOT*s were sequenced and compared to the *GmPHOT* sequences from the Phytozome database to assess the accuracy of the gene models. The gene models of all the *GmPHOT*s were found to be accurate except that of *GmPHOT1B-2*. The high level of sequence identity between the *GmPHOT*s and *AtPHOT*s and the conservation of LOV domains and catalytic domains indicate structural resemblance between them. This suggests that soybean phototropins should encode active photoreceptors. The cloned protein coding regions from soybean were then recombined into a plant expression vector via Gateway technology, which were then used for transformation of *Agrobacterium tumefaciens*. These plant expression constructs will be utilized in the future to determine the functionality of soybean phototropins in *Arabidopsis*.

CHAPTER 1 INTRODUCTION

Plants, being sessile organisms, require adaptation as a key feature to optimize their functions in fluctuating environmental conditions. Most key events in growth and development throughout the life-cycle of a plant are regulated by light. Seedling development is one of the key processes requiring strict monitoring of light quality and quantity. From gene expression to cellular and subcellular architecture to organ anatomy, the way seedlings develop under illuminated conditions (photomorphogenesis) is discrete from the way they develop in dark (etiolation).While in photomorphogenesis the thrust of plant development is to mainly grow and differentiate in a manner to optimize photosynthesis, etiolation prompts cell elongation in the shoot for the plant to successfully reach light conditions sufficient for photoautotrophic growth.

1.1 Plant photoreceptors

Light is a remarkable developmental cue in plants. For efficient and direct absorption of light, plants have multiple protein molecules known as signaling photoreceptors, or more simply as photoreceptors. These classes of proteins are also central to a plant's ability to monitor the quality, quantity, and direction of incident light. For over a decade now,

many experiments have been performed to characterize several classes of plant photoreceptors that absorb light from red and blue regions of the electromagnetic spectrum.

Phytochromes (phy), represent an important class of photoreceptors that are predominantly responsive to red and far-red light [1, 2]. Five phytochromes have been identified in *Arabidopsis*, designated phyA through phyE. The physiologically inactive P_r form of a phytochrome changes to its active form, P_{fr}, upon absorption of red light [3]. The phytochromes participate in controlling flowering time and also mediate the de-etiolation response in seedlings. There are three other classes of photoreceptors that are activated specifically by UV-A/blue light; the cryptochromes (cry), the phototropins (phot) and Zeitlupe family members [4]. cry1 and cry2, the two cryptochromes in *Arabidopsis* mediate membrane depolarization, inhibition of hypocotyl elongation by blue light and anthocyanin production [5-7]. Cryptochromes work in tandem with phytochromes to synchronize photomorphogenic responses, like regulation of cell elongation and photoperiodic flowering, while phototropins trigger movement responses including the fascinating phototropic curvature that intrigued Darwin more than a century ago. There is an overlap between the absorption region of the blue and red light/far red light receptors and that of the photosynthetic pigments enabling coordinated growth and development of plants.

1.2 Phototropism

Phototropism has been extensively explored. It is a primary response of phototropins to blue light. Phototropism, generally described as the bending of plant organs towards light, is caused by differential growth arising from increased cell elongation on the shaded side relative to the irradiated side. Phototropism, though apparently simple in nature, is actually a highly complicated light-dependent mechanism [8]. For example, different types of bending reactions have been described based upon light intensity and duration [9].

The first detailed studies on phototropism were performed by Darwin in 1880 on canary grass (*Phalaris canariensis*) [10]. Darwin observed that grass coleoptiles exhibit phototropism when exposed to a unilateral light source. Strikingly, he observed that there was no phototropic response when the tip of the coleoptile was covered or removed. However, there was no effect on the response when the lower portion of the coleoptile was covered leading Darwin to propose that the tip of the coleoptile is the vital region for light detection. He also proposed the presence of a transmissible substance that is transmitted from the tip of the coleoptile to the lower regions, causing phototropism. Boysen and Jensen in 1911 (reviewed in [11]) observed the transmission of this substance through a block of gelatin inserted between the tip and the remaining portion of the coleoptile. Insertion of a thin sheet of mica on the irradiated side inhibited the response while no effect was observed upon its insertion on the shaded side. There was an increase in growth on the shaded side relative to the irradiated side of the seedling, triggering the

stem to bend towards the light source. These experiments led to the conclusion that the phototropic curvatures driven by differential growth rates were apparently occurring due to the differential distribution of the transmissible substance, which was later identified as the plant hormone, auxin. The Cholodny-Went hypothesis provides much of the basis for the proposed role of auxin in tropic responses (reviewed in [11]). According to this theory, stimulation by unilateral light produces an auxin gradient between the irradiated side and the shaded side that results in differential growth and phototropic curvature.

Phototropism is divided into two separate responses: first and second positive phototropism. First positive curvature occurs near the coleoptile tip in response to short pulses of low fluence blue light. Exposure of the apical part of the coleoptile to higher fluences leads to the second positive phototropic curvature observed at the base of the coleoptile [12]. The Bunsen-Roscoe reciprocity law, according to which the magnitude of the phototropic response increases with increase in the total dose of energy supplied independent of time, is obeyed by first positive phototropism only [13]. This means that seedlings in response to strong light for short duration of time or to light pulses for a prolonged period of time should produce similar phototropic responses, provided that the fluence admisnistered is same in both cases. Phototropic response is observed both in monocots and dicots substantiating its conserved nature in angiosperms. Typically, blue and UV-A light mediate phototropism. R light absorbed by phytochrome can enhance the B-light induced phototropism in *Arabidopsis* [14]. The B-light elicited phototropism in maize mesocotyls and pea epicotyls is mediated by phytochrome [15]

1.3 Phototropins

The earliest step towards unravelling the intricacies of phototropic signal transduction was put forward by Gallagher and his colleagues [16] when they discovered the rapid B-light dependent phosphorylation of a 120 kDa plasma membrane-associated protein in dark grown epicotyls of pea (*Pisum sativum*). A gradient of phosphorylation of the 120 kDa protein was found to be distributed across oat coleoptiles when irradiated with unilateral blue light [17, 18]. A significant correlation was noted between the phosphorylation of this protein and phototropism with regard to tissue distribution, fluence-response characteristics and recovery in darkness following a light pulse [19]. Similar observations were obtained in maize coleoptiles [20]. In both, pea and maize, rapid phosphorylation has been reported and is likely to be happening early in the phototropic process [19].

It was found in experimental studies with *Arabidopsis* using unilateral blue light as the light source that the nonphototropic hypocotyl mutant (*nph1)* showed disrupted hypocotyl phototropism because of the lack of the 120 kDa protein kinase activity [21]. This work proposed the presence of a genetic locus, *NPH1*, encoding a phosphoprotein that is linked to the photoreception processes mediating all phototropic responses in *Arabidopsis*. The *NPH1* gene was subsequently cloned and sequenced followed by photochemical and biochemical characterization of the *NPH1* flavoprotein kinase. Thus the molecular identity of the photoreceptor for phototropism was resolved [22] [23]. Another slightly smaller phosphoprotein, with 58% identity and 67% similar [24] to *NPH1* was identified while screening for *Arabidopsis* mutants with disrupted chloroplast avoidance response.

This phosphoprotein was named as *NPL1* (*NPH1-like1*) because of its structural similarity with *NPH1*. [25]. *NPL1* is the photoreceptor that regulates chloroplast avoidance response when activated by strong blue light. Due to its role in phototropism, the *NPH1* encoded protein was later renamed phototropin 1 [26] and *NPL1* was subsequently renamed phototropin 2 [24].

Photoreceptor activation occurs maximally at 390 nm [27]. Phot1 has been found to be localized to the plasma membrane by biochemical analysis [16, 28]. When exposed to blue light, a fraction of phot2 was found to be localized to Golgi apparatus [29].

The signaling processes that take place downstream of the blue light receptor activation are not well understood. Only a few signaling molecules have been identified that form complexes with the phototropins and mediate the blue-light induced phototropic responses. Phot1 and phot2 have different signal transducers to transfer the blue light signal to different phototropic responses. While NPH3 (NONPHOTOTROPHIC HYPOCOTYL3) regulates the phot1 and phot2-mediated phototropism signaling pathway, RPT2 (ROOT PHOTOTROPISM2) regulates the phot1-mediated phototropism and stomatal opening response pathways. RPT2 is not required for the phot2-mediated signaling pathway. The signaling pathway of phot1 and phot2-mediated chloroplast relocation is not regulated by either NPH3 or RPT2 [30]. PKS1 (PHYTOCHROME KINASE SUBSTRATE 1), a plasma membrane-associated protein, has been identified to form a complex with both phot1 and NPH3 in order to mediate the phototropic curvature response [31]. Ca^{2+} has been detected as an important regulator of phototropic stimuli

transduction showing an increased concentration of Ca^{2+} in the wild type *Arabidopsis* when exposed to blue light [32]. *Arabidopsis phot1* mutants severely lacked this response.

Asymmetrical distribution of auxin occurs in response to phototropic stimulation. Genetic studies in *Arabidopsis* have contributed immensely to the identification of several protein families that regulate the transport of auxin [33, 34]. Members of the Pin-formed (PIN) family, named due to pin-like appearance of inflorescence of the *pin1* mutant, are auxin transport proteins that mediate efflux of auxin from cells. The directionality of auxin flow is primarily governed by the polarity of PIN proteins. PIN3 mediates asymmetrical distribution of auxin in both pulse-induced and time-dependent phototropism [35]. Etiolated *Arabidopsis pin3* seedlings have been shown to exhibit delayed phototropism [36]. Other auxin transporters like ABCB19 (auxin efflux transporter ATP BINDING CASSETTE B19) and PIN1 are known to mediate polar auxin transport from the shoot apex to the roots. The polar auxin transport of auxin causes differential cell elongation on the shaded side resulting in phototropic bending.

1.4 Serine/threonine protein kinase and LOV Domains

Arabidopsis contains two phototropins designated as AtPHOT1 and AtPHOT2 [37]. After obtaining PHOT1 and PHOT2 genomic and cDNA sequences, the *Arabidopsis* PHOT1 and PHOT2 was deduced to encode proteins of length 996 and 915 amino acids respectively. The PHOT1 and PHOT2 gene of *Arabidopsis* encodes a 120 kDa and 110 kDa protein kinase activity respectively, which is activated by blue light. Plant

phototropins contain two photosensory LOV1 and LOV2 domains in the amino (N-) terminus and a protein serine/threonine kinase domain in the carboxy (C-) terminus [22]. All of the eleven conserved subdomains that are common to all protein serine/threonine kinases are present in the PHOT1 kinase domain [38].

LOV domains are a subfamily within the PAS superfamily that are highly conserved in proteins that are regulated by light, oxygen and voltage [22]. PAS domains coordinate changes in light, oxygen, redox potential, small ligands and overall energy level of a cell [39]. Domains containing these imperfect repeat sequences were first recognized in the *Drosophila* period clock protein (PER), vertebrate aryl hydrocarbon receptor nuclear translocator (ARNT), and *Drosophila* single-minded protein (SIM). PAS domains allow protein-protein interactions in signaling molecules important for cellular signaling processes. Each LOV domain consists of 110 amino acids. These LOV domains bind FMN (flavin mononucleotide) and absorb blue light. LOV1 and LOV2 domains have almost identical structures with interconnected alpha helices and beta strands. The FMN chromophore is centrally attached with the help of hydrogen bonding and vander Waals forces [40, 41].

When the LOV domains are excited by blue light, a covalent reaction between the cysteine residue of the LOV domain and the fourth carbon of the flavin chromophore occurs, resulting in the formation of a stable adduct. This results in conformational changes within the LOV domains, including unfolding of the Jα helix (an amphipathic conserved α-helix located at the carboxyl-terminus of the LOV2 domain). The LOV2

domain is connected with the kinase domain by the Jα helix. Subsequently, the kinase domain becomes activated, leading to autophosphorylation of the conserved serine residues (Ser^{851} and Ser^{849}) [23, 42, 43].

Earlier studies prior to the biochemical characterization of phototropin had shown that phototropins undergo blue-light activated phosphorylation. Experiments showing phot1 gene expression in insect cells established that it was an autophosphorylation reaction and that the phototropin molecule itself was the photoreceptor for the reaction [23]. Thus phototropic responses are a consequence of the autophosphorylation of phototropin. If the cysteine of the LOV domain is mutated to either alanine or serine, autophosphorylation is inhibited. In the dark, the LOV domains bind to the kinase domain, inactivating it [27, 44]. This means activation of the kinase domain is essential for phototropin signaling. LOV domains become activated (illuminated state) or repressed (dark state) depending upon the light conditions. LOV2 works as the “molecular light switch” that controls the activation or repression of the C-terminal kinase domain [44].

The light detection capability for regulating phototropin activity is different for LOV1 and LOV2. The quantum efficiency required for forming the cysteinyl adduct in both phot1 and phot2 is greater in LOV2 in comparison to LOV1 [45]. It has been shown from mutagenesis studies that the photoreactivity of LOV1 is not sufficient to promote phototropism and chloroplast avoidance responses in either phot1 or phot2. Autophosphorylation of phot1 is dependent upon LOV2 photoreactivity [46]. Likewise, chloroplast avoidance response in phot2 is solely mediated by the photoreactivity caused

by LOV2 [47]. The exact functions of LOV1 domain is not known so far. A hypotheses have been framed based on X-ray scattering analysis, suggesting a role of LOV1 in dimerization of the photoreceptor or prolonging the activation duration of the photoreceptor [47].

The light-dependent photochemical reaction in LOV2 [46] and the activity of the kinase domain [29] are two attributes which are essential for a physiological response as evident from functional analyses of mutated phototropins. The kinase domain has been determined to be responsible for the localization of phototropins to plasma membrane or Golgi apparatus [29]. Casein is the only known substrate for the phot2 kinase domain detected from in vitro studies. The serine residues on the LOV1 domain of phot1 are phosphorylated by its kinase domain in vivo [44, 48]. Another substrate for phot1, auxin efflux transporter ATP BINDING CASSETTE B19 (ABCB19) protein, has been detected within the hypocotyl apex region of *Arabidopsis* [49]. This suggests that proteins apart from phototropin itself can also be phosphorylated by phototropin.

1.5 Photobiology of phototropins

Phototropins are detected not only in higher plants but also in green algae (*Chlamydomonous reinhardtii*) [4]. PHOT1 and PHOT2 have some overlapping functions but have some distinct functions, too. Both PHOT1 and PHOT2 together drive phototropism, stomatal opening, chloroplast arrangement in leaf cells, leaf expansion, and leaf orientation. PHOT1 specifically inhibits the growth of hypocotyls in young seedlings, whereas PHOT2 mediates the chloroplast avoidance response towards high

light (Table 1). Collectively, these responses impact a plant's ability to capture light for photosynthesis. The *phot1* mutants exhibit normal hypocotyl curvatures in response to unilateral blue light from fluence rates of 1 to 100 µmole m^{-1} s^{-1}. But the *phot1* mutants do not exhibit phototropism at fluence rates below 1 µmole m^{-1} s^{-1}[25]. The *phot2* mutants exhibit normal hypocotyl phototropism in response to unilateral blue light from fluence rates of 0.1 to 100 µmole m^{-1} s^{-1}[42]. Hypocotyl phototropism was found to be impaired only in the *phot1 phot2* double mutants at high light intensities.

High light intensities generate reactive oxygen species (ROS) in the chloroplasts of plants, which causes deleterious effects like photodamage resulting in necrosis and bleaching. In order to avoid photodamage, chloroplasts are arranged in a manner that enables them to receive the desired amount of light required for photosynthesis. Chloroplasts accumulate at the periclinal side of cells when the intensity of light is low. This helps maximize the amount of light absorbed for photosynthesis. This response is called chloroplast accumulation response. PHOT1 mediates this response under low light intensities. It was found in experimental studies that *phot1* mutants showed normal accumulation movement even at low light intensities indicating that PHOT1 is not the only player in chloroplast accumulation response. At higher light intensities, the chloroplasts tend to move away from the periclinal side to avoid the excess amount of light entering the chloroplast apparatus. PHOT2 mediates chloroplast avoidance in response to high light intensities. The *phot1 phot2* double mutants did not exhibit either the chloroplast accumulation or the avoidance response [42]. These results indicate that both PHOT1 and PHOT2 contribute to chloroplast accumulation response but only

PHOT2 controls chloroplast avoidance response. It has been shown experimentally that *phot1* and *phot2* single mutants exhibit normal stomatal opening response but the response is abolished in *phot1 phot2* double mutants [50]. This suggests that PHOT1 and PHOT2 regulates stomatal opening in a redundant fashion in response to blue light in plants. Both PHOT1 and PHOT2 contribute to the leaf flattening response. The *Arabidopsis* mutant plants lacking both PHOT1 and PHOT2 gene were detected with leaf curling phenotype [51]

In our lab, a set of protein kinases and their genes from garden pea was studied in order to understand their role in photoregulated development in young seedlings. A suite of partial cDNA clones that encode distinct forms of protein serine/threonine kinase domains were isolated and named as PsPK1 to PsPK5. The *Arabidopsis* PHOT1 gene encodes a blue light-induced protein kinase that is closely related to PsPK4 and PsPK5, suggesting that PsPK4 and PsPK5 genes will encode an active photoreceptors. The *Arabidopsis* phot1 mutants do not exhibit phototropism due to the absence of the 120 kDa phosphoprotein.

The PsPHOT1A (PsPK4) protein coding region was cloned into a plant transformation vector containing a 35S promoter (a strong viral promoter obtained from Cauliflower Mosaic Virus required to drive the transgene expression). It was then introduced into *Arabidopsis phot1* mutants, via *Agrobacterium*-mediated plant transformation using the floral dip method [52] for testing whether PsPHOT1A can complement the null mutation in *Arabidopsis*. Phototropic curvature was detected in all the transformants in response to

unilateral blue light and the response was similar to the wild type control plants. This experiment concluded that PsPHOT1A was a functional phototropin and thus encodes an active photoreceptor.

1.6 Expression profile of phototropins

Abundant expression of phot1 is observed in etiolated seedlings while its expression decreases when irradiated with blue light. Rice (*Oryza sativa*) has been shown to contain two phototropins: OsPHOT1 which has 61% protein sequence identity with AtPHOT1 and OsPHOT2 which has 62% protein sequence identity with AtPHOT2. These two homologues of rice show differential expression levels when irradiated with white light with an upregulation of OsPHOT2 transcript levels while downregulation in case of OsPHOT1. There is a high level of OsPHOT1 transcript observed in coleoptiles while OsPHOT2 was strongly expressed in mature leaves [53].

Maize (*Zea mays*) has two phototropins, ZmPHOT1 and ZmPHOT2. ZmPHOT1 is abundantly expressed in etiolated coleoptiles, with its expression rapidly declining on exposure to pulses of low fluence blue light and continuous blue light of high fluence rate. Both ZmPHOT1 and ZmPHOT2 are weakly expressed in young leaves and there is no specific expression in the coleoptile tip region [54]. When dark grown pea seedlings were illuminated with continuous white light, red light or blue light, a significant decrease (10 fold) in the mRNA transcript levels of PsPHOT1B was observed, the decrease being largest in case of blue light [55]. Phytochromes, phyA and phyB, have been detected to be involved in the reduction of the PsPHOT1B transcript levels in

response to red light. High levels of PsPHOT1A transcripts have been detected in the leaves and stems of light grown mature plants [56].

1.7 Soybean (*Glycine max*)

Soybean is an important crop plant in Indiana important for its seed protein and oil content. At present, there is no knowledge about the function of phototropins in soybean. Soybean, a prominent member of the legume family, is structurally and functionally quite different from *Arabidopsis*. The photobiological actions of the phototropins are best understood in the model plant *Arabidopsis*, because of the availability of loss-of-function mutants in the *PHOT1* and *PHOT2* genes. Since, most of our knowledge of the phototropins comes from research on *Arabidopsis*, we want to extend this knowledge to soybean phototropins and in the process, hoping to unravel other unexplored aspects of phototropin function.

The soybean genome, approximately 1.1 gigabases in size, was sequenced by whole genome shotgun sequencing [57]. Interrogation of the soybean genome sequence with the pea PsPHOT1A and PsPHOT1B, and *Arabidopsis* phot1 and phot2 revealed 5 genes predicted to encode full length phototropins. These polypeptide sequences were compared to the *Arabidopsis* phototropins and the published legume sequences, showing that 3 of the soybean phototropins belong to phot1-type and 2 of them belong to phot2-type (Table 2) (Watson unpublished). The genomic sequences of *GmPHOT1A*, *GmPHOT1B-1*, *GmPHOT1B-2*, *GmPHOT2A* and *GmPHOT2B* are 20905, 15078, 13546, 26663 and 58579 nucleotides in length. Due to the complexity and huge size of the

soybean phototropin genes, harboring large number of duplications and mobile genetic elements, it is indeed difficult to be using the whole genomic sequences for molecular cloning and future plant transformation research purposes. Thus we decided to clone and express cDNAs for the 5 soybean phototropin genes.

Different tissue specific-expression patterns of soybean phototropin genes have been found by RNAseq experiments [58]. *GmPHOT1A* and *GmPHOT1B-1* were found to highly expressed in young leaves and in flowers, with low level of expression in seeds, roots and nodules. *GmPHOT1B-2* and *GmPHOT2A* have comparatively low expression in these tissues.

1.8 Generation of Plant transformation constructs

When plants are infected by *A. tumefaciens,* a single-stranded T-DNA region from the tumor inducing (Ti) plasmid of *A. tumefaciens* along with several virulence (vir) proteins are transferred to the host plant cells. The T-region of *A. tumefaciens* is flanked by the T-DNA borders, which are imperfect tandem repeats of 25 bp [59, 60]. Foreign DNA can be transferred to the plant cells by using the Ti plasmid of *A. tumefaciens* as the vector. The first *Agrobacterium*-mediated plant transformation was performed in 1983 [61]. In the past, it was not an easy task to clone a gene of interest into Ti plasmid because of its huge size and the lack of a unique restriction recognition sites within the T-DNA region. Different strategies were developed to deliver foreign DNA into the T-region of Ti

plasmid. One of these strategies led to the generation of T-DNA binary vectors. The binary vector method for plant transformation takes advantage of *A. tumefaciens* and its capacity to transfer T-DNA to [62].

A T-DNA binary system is composed of two different replicons, a binary plasmid that contains the T-DNA and a helper plasmid containing the virulence genes [63, 64]. The size of a binary vector can be kept small by having the vir genes on a separate plasmid. The small size of the binary plasmid enables replication in both *E. coli* and *A. tumefaciens*. The introduction of large Ti plasmids into plants was found to be possible in experiments in which Ti plasmid of 200 kbp was introduced into plants on reversing the orientation of the right border of T-DNA [65].

A. tumefaciens plasmid vectors have been modified in a variety of ways to engineer different sets of Gateway compatible binary vectors for plant transformation purpose [66]. Gateway technology is a recombination based cloning method introduced by Invitrogen, that facilitates molecular cloning via the site specific recombination system of bacteriophage lambda [67] while maintaining the orientation of the reading frame within the fragments of interest. It allows the cloning of PCR products into a plasmid (the entry vector) followed by transfer of the cloned DNA from the entry clone to other plasmids (the destination vectors) that are specific for different downstream applications resulting in the formation of expression clones.

The destination vector used in this project is pGBPGWG obtained from Nottingham *Arabidopsis* Stock Centre (NASC). It is derived from pGreenII binary vector and it is 7652 bp long. It consist of a 35S promoter (to express the transgene), attR1 site, ccdB gene (inhibits growth of most E. coli strains), attR2 site, an enhanced GFP (Green fluorescent protein) fused to the Carboxy-terminus, basta resistance marker (selective agent for plants) and a 35S terminator. It has been detected from experimental studies that the phot1-GFP fusion under the control of phot1 promoter localizes to hypocotyl, inflorescence stem and etiolated roots of *Arabidopsis*. On the contrary, under the control of 35S promoter, phot1-GFP fusion has been detected uniformly throughout the cytoplasm [51]. Thus the fluorescent tag (EGFP) will be used to enable detection of the soybean proteins in transgenic plants in future plant transformation experiments.

The cloned insert in the entry clone is flanked by *attL1* and *attL2* sites. An LR recombination reaction recombines the sequence of interest into the destination vectors flanked by *attR1* and *att R2* sites resulting in the formation of an expression clone flanked by attB1 and attB2 sites [68]. The gene of interest resides in the expression construct. This is obtained by replacing the ccdB gene from the destination vector with the gene of interest.

The LR recombination reaction:

"*attL* (entry clone) x *attR* (destination vector) → *attB* (expression clone) x *attP* (by-product)" [Gateway cloning technology, instruction manual]

PROJECT OUTLINE

Our fundamental hypothesis is that the soybean phototropins play essential roles in regulating processes that modulate photosynthetic activity, as do the *Arabidopsis* phototropins. Clones of the *GmPHOT* mRNAs and genes are not available. Thus, reverse-transcriptase-PCR was used to obtain cDNAs for all 5 of the *GmPHOT* mRNAs. Total RNA, extracted from 5 day old etiolated soybean epicotyl tissue, served as the template for RT-PCR using primers (including the stop codon and those lacking it) derived from the genomic sequences. The primers were designed to target the 5'- and 3'-ends of each *GmPHOT* protein coding region. The soybean cultivar Williams 82 was used since it was used for the soybean genome project. Three different kits with different DNA polymerase enzymes (*Taq* DNA polymerase, Easy A-high fidelity cloning enzyme, *Pfu* Ultra II DNA polymerase) have been used to obtain cDNA copies to eliminate clones containing errors.

The *GmPHOT* cDNAs were then cloned into pCR8/GW/TOPO vector or pENTR-D/TOPO vector, depending upon the type of kit used, via TOPO cloning. The identity and orientation of the clones were confirmed by diagnostic restriction analysis with specific restriction enzymes. Next, the cloned cDNAs were chosen for sequencing and compared to the published *GmPHOT* sequences from the phytozome database using Vector NTI software in order to determine the accuracy of the gene models.

Clones from at least 3 rounds of RT-PCR were generated in order to validate the accuracy of the cloned sequences. Clones containing stop codons were generated for downstream fusions to the amino-terminus of the *GmPHOT* protein coding region while clones lacking stop codons were generated for downstream fusions to its carboxy-terminus.

The predicted *GmPHOT* polypeptide sequences were compared to the protein sequences of the *Arabidopsis* phototropins by pairwise and multiple alignments using ClustalX 2.1 software. The percentage identity between the sequences was determined. Similarly the percentage identity between the LOV1, LOV2 and catalytic domains of soybean and *Arabidopsis* phototropins were determined.

One of the cloned cDNA sequences (entry clones) from each *GmPHOT,* that matched the predicted sequences from phytozome were then recombined into plant expression vector, pGBPGWG, using Gateway cloning strategies. The plant expression constructs (35S::*GmPHOT:EGFP*) obtained by LR recombination reaction were sequenced across the recombination (*att*) sites to ensure that the open reading frame is in-frame with the tag protein (EGFP). The expression clones were then introduced into *A. tumefaciens* strain GV3101 (pSoup) via electroporation.

CHAPTER 2 MATERIALS AND METHODS

2.1 Reverse transcriptase Polymerase chain reaction

Total RNA from soybean (*Glycine max*) cultivar Williams 82 was isolated by Dr. Watson from 5 day old dark-grown epicotyls, using the Qiagen RNeasy Plant Mini kit. Primers that either include the stop codon or lacked it, were designed by Dr. Watson to target the 5'- and 3'-ends of each *GmPHOT* protein coding region. The primers were obtained from Invitrogen. The reverse transcriptase-polymerase chain reaction (RT-PCR) was used to obtain cDNAs for the *GmPHOT* mRNAs using a total RNA template with gene-specific primers (Tables 3, 4). A 50 µl PCR amplification was performed using 5 µl of PCR reaction buffer, 1 µl of 10 mM dNTP, 0.5 µl of RT enzyme, 1 µl of RNA (80 ng), 1 µl of *Taq* DNA polymerase (Superscript III One-step RT-PCR with Platinum *Taq*, Invitrogen), 1 µl of 5 µM stocks of forward and reverse primers. The volume was adjusted to 50 µl by adding sterile Milli-Q water. The reaction was incubated under the following conditions: 50°C for 20 mins (RT step), 94 °C for 5 mins followed by 30 cycles of 94°C for 15 sec, 51°C for 30 sec, 68 °C for 17 sec with a final extension at 72 °C for 5 mins. A Perkin Elmer Gene Amp PCR system 2400 thermocycler was used. A 10 µl aliquot of the PCR products was electrophoresed on a 1.2% (w/v) agarose gel, using 1X TAE (Tris-acetate-EDTA) as the running buffer. The agarose gel was then subjected to staining in 0.5 µg/ml of ethidium bromide (EtBr) solution for 30 minutes, followed by de-staining

for another 30 minutes in Milli-Q water and then visualized with Quantity One 1-D Analysis software. Accuscript *PfuUltra* II RT-PCR kit (Stratagene) was used to amplify the GmPHOT mRNAs for those containing presumptive sequence errors as discussed in detail in the Results and Discussion. *PfuUltra* II Fusion HS DNA Polymerase has enhanced proof-reading activity with a hotstart DNA antibody (which reduces the possibility of non-specific amplification and primer-dimer formation) and ArchaeMaxx PCR enhancing factor (acts as a dUTPase converting harmful dUTP to dUMP preventing the accumulation of dUTP and PCR inhibition, thus increasing overall PCR performance). These features help produce PCR products with high throughput and efficiency. The AccuScript RT (included in the kit) is derived from a Moloney murine leukemia virus reverse transcriptase that has 3'-5' exonuclease activity. This is very efficient in reducing reverse transcriptase-induced errors and increasing the accuracy of cDNA synthesis. The RT-PCR reaction using this enzyme is different compared to the one-step RT-PCR reaction discussed above. It is a two-step reaction with the cDNA being synthesized from RNA in the first step. The RT reaction consists of 1 µl of 10X AccuScript RT reaction buffer, 0.6 µl of oligo dt primer, 1 µl of 10 mM dNTP mix and 460 ng of RNA. The reaction is incubated at 65 °C for 5 min followed by the addition of 1 µl of 100 mM DTT and 0.5 ul of AccuScript high-fidelity RT. The reaction is then incubated at 42 °C for 1.5 hr. Five µl of the cDNA synthesized is amplified in a 50 µl PCR reaction consisting of 5 µl of 10X PCR reaction buffer, 1 µl of dNTP mix, 2 µl of 5 µM gene specific forward and reverse primers (Four bases CACC were added to the 5' end of forward primers for allowing complementation with the GTGG overhang in the pENTR-D/TOPO vector downstream), 5 µl of synthesized cDNA, and 1 µl of *PfuUltra* II

HS DNA polymerase. The volume was adjusted to 50 µl by adding sterile Milli-Q water. The cycling parameters were 95°C for 1 min followed by 40 cycles at 95 °C for 30 sec, 51 °C for 30 sec, 68°C for 3 min 10 sec with a final extension at 68 °C for 5 min. Amplified PCR products of *GmPHOT1A* and *GmPHOT1B-1* were fractionated on a 1.2 % agarose (w/v) gel, stained with EtBr and visualized. *GmPHOT1B-2* and *GmPHOT2A*/*GmPHOT2B* were purified using QIA Quick PCR purification kit (Invitrogen) before resolving on 1.2 % agarose (w/v) gel because of the problems that were generated in the subsequent TOPO cloning reaction. PCR products were purified using QIA Quick PCR purification kit (Invitrogen). The concentrations (ng/ µl) of the products were estimated using NIH Image 1.62 software in order to calculate the amount of the product that would be used to set up TOPO cloning reaction.

2.2 Polymerase Chain Reaction (PCR)

Clones containing the endogenous stop codons were used to generate clones lacking stop codons by PCR using GoTaq Green Master Mix (Promega, Madison) and gene-specific primers lacking stop codon. 2X GoTaq Green Master mix is a mixture of GoTaq DNA polymerase, 2X Green GoTaq Reaction buffer, dNTPs and 3 mM $MgCl_2$. A 25 µl PCR reaction mixture containing 12.5 µl of 2X GoTaq Green Master mix, 1.25 µl of 5 µM stocks of forward and reverse primers and 10 ng of cDNA (containing stop codons) was set up. The volume was made up to 25 µl by adding sterile Milli-Q water. The cycling parameters were 95°C for 2 min followed by 25 cycles at 95°C for 20 sec, 51 °C for 20 sec, 72 °C for 3 min 10 sec with a final extension at 72 °C for 5 min. The PCR reaction

was followed by *DpnI* digestion to remove template DNA if any. One µl of *DpnI* enzyme was added to the PCR product and digested for 1 hour at 37 °C and then the enzyme was inactivated at 65 °C for 15 min. The *DpnI*-digested PCR products were electrophoresed on 1.2% (w/v) agarose gel followed by staining and de-staining.

It is found in early experiments that clones obtained with Easy-A High-fidelity PCR cloning enzyme are much less likely to containing PCR-induced errors than those produced with *Taq* DNA polymerase. Easy-A High-fidelity PCR cloning enzyme (Stratagene, California) has 3'-5' exonuclease activity in addition to terminal transferase activity. It adds an overhang of a single 3' A that helps in efficient TOPO cloning reactions. Some of the clones obtained from *Taq* DNA polymerase, both having and lacking stop codons, were having PCR induced errors, as evident from the sequencing results that will be discussed later. Thus Easy-A High-fidelity PCR cloning enzyme kit was used. A 25 µl PCR amplification was performed using 2.5 µl 10X Easy A reaction buffer, 0.5 µl of 10 mM dNTPs, 1.25 µl of 5 µM stocks of forward and reverse primers, 0.25 µl of Easy A Hifi PCR cloning enzyme. The volume was made up to 25 µl by adding sterile Milli-Q water. The cycling parameters were 95 °C for 2 min followed by 25 cycles at 95 °C for 40 sec, 51 °C for 30 sec, 72 °C for 3 min 10 sec with a final extension at 72 °C for 7 min. A 10 µl aliquot of the PCR product was electrophoresed on a 1.2% (w/v) agarose gel, using 1X TAE as the running buffer, followed by EtBr staining as discussed above.

2.3 TOPO cloning

PCR products obtained with *Taq* DNA polymerase and Easy-A High-fidelity PCR cloning enzyme were cloned into the pCR8/TOPO vector (pCR8/GW/TOPO TA cloning kit, Invitrogen). The vector is Gateway-adapted to facilitate a recombination reaction which incorporates the gene of interest into a Gateway destination vector. This vector contains a spectinomycin resistance marker. The deoxyadenosine (A) residues produced at the 3'-end of the PCR product by the *Taq* polymerase and Easy A Hifi cloning enzyme complements efficiently with the single overhanging 3' deoxythymidine (T) residues of the TOPO vector. *PfuUltra* II HS DNA polymerase-amplified PCR products were directionally cloned into pENTR/D-TOPO vector (pENTR Directional TOPO Cloning Kit, Invitrogen). This Gateway-vectors harbor a kanamycin resistance marker. The overhang (GTGG) in the vector compliments efficiently with the complimentary bases (CACC) in the PCR product. The efficiency of directional TOPO cloning is greater than or equal to 90%. In the TOPO cloning reaction, 1 µl (20 ng) of PCR product (as determined by the NIH Image 1.62 software) was added to l µl of appropriate TOPO vector and 1 µl of salt solution. Sterile Milli-Q water was added to a final reaction volume of 6 µl and incubated at room temperature for 2 min.

Following the TOPO cloning reaction, 2 µl of the reaction was added to a tube containing 50 µL of chemically competent TOP 10 *E. coli* cells (provided with the TOPO cloning kits) and mixed gently. The reaction was incubated on ice for 5 minutes. Then the cells were given heat shock treatment for 30 seconds at 42 °C. After that, 250 µl of S.O.C

medium (included in the TOPO cloning kit) was added to the tubes and then the tubes were shaken at 37°C for an hour. One hundred µl of the bacterial culture was then spread on LB agar plates containing spectinomycin (100 µg/ml) or kanamycin (50 µg/ml) depending on the vector used, followed by overnight incubation at 37°C.

About 20 to 50 resistant colonies were obtained. Eight to twelve colonies were selected for analysis by plasmid DNA isolation from spectinomycin or kanamycin resistant transformants. The colonies were inoculated in 2.5 ml of LB containing the respective antibiotics for 14-16 hrs. A portion of the culture was used to prepare glycerol freezer stocks and the rest was used for plasmid DNA isolation. Standard plasmid minipreps were performed using QIAGEN Plasmid Miniprep kit on the broth cultures to extract plasmid DNA.

2.4 Diagnosis of the *GmPHOT* entry clones

Once the DNA was extracted, the identity of the inserts was diagnosed by restriction digestion with *PstI* or *HindIII*. The correct orientation of the inserts was diagnosed by restriction enzymes *EcoRV* and *BamHI* or *HindIII* and *HaeII* (Tables 5, 6).

After identifying a positive isolate (with correct identity and orientation), its freezer stock was used to inoculate LB broth containing 100 µg/ml spectinomycin or 50 µg/ml of

kanamycin for plasmid DNA purification by QIAGEN Plasmid Midiprep kit prior to being sequenced by Eurofins MWG Operon (Huntsville, Alabama) using the M13 forward and M13 reverse primers (provided by the Eurofins MWG Operon) as well as internal gene-specific sequencing primers, which were designed by Dr. Watson (Tables 7, 8, 9, 10, 11). The sequencing results were compared to the *GmPHOT* sequences obtained from the Phytozome database, which contains the published soybean genome sequence, using the Vector NTI Advance Software. Clones from at least three separate rounds of RT-PCR were obtained for each *GmPHOT* to validate the accuracy of the cloned sequences.

2.5 LR cloning using Gateway Technology

LR cloning using Gateway Technology is a cloning method that helps one to insert the gene of interest into a Gateway destination vector of choice for downstream analysis. When using Gateway cloning, the entry vector (pENTR-D/TOPO) and Gateway destination vector harbor the same Kanamycin resistance marker. pGBPGWG, obtained from Nottingham *Arabidopsis* Stock Centre (NASC), was used as the destination vector [69].

In order to avoid recovering entry clones after the LR recombination reaction, the entry clones (only those in pENTR-D/TOPO vector) were linearized by *NruI* restriction digest. This was done because both the pENTR-D/TOPO vector and the destination vector

harbor the same kanamycin antibiotic resistance marker. *NruI* has one restriction site in the pENTR-D/TOPO vector but none in the Gateway vector nor in the gene of interest. This digestion was not required for the *GmPHOT* clones in pCR8/TOPO vector because of different antibiotic resistance markers of pCR8/TOPO vector and the destination vector. Following the digestion, equal amounts of the entry clones and the destination vector DNA were mixed (1 µl at 150 ng/1µl of each plasmid). The volume of the reaction was adjusted to 8 µl by adding 1X TE buffer (pH 8.0), then 2 µl of LR clonase II mix (Gateway® LR Clonase™ II enzyme mix, Invitrogen) was added to a final volume of 10 µl. The reactions were incubated at 25 °C for 16 hours. After that, each reaction was terminated with 1 µl of proteinase K solution (included in the kit). Following the termination, the reactions were incubated at 37 °C for 10 minutes. One microlitre of each reaction was then transformed with 50 µl of chemically competent TOP 10 *E. coli* cells via the heat shock method from the transformation reactions mentioned before. Two hundred µl of the bacterial culture was spread on LB plates containing 50 µg/ml of kanamycin and incubated overnight at 37 °C. Kanamycin resistant colonies were selected for plasmid DNA isolation using QIAGEN Plasmid Miniprep kit.

The expression clones were diagnosed for identity by restriction digestion with *EcoRI* enzyme (Table 12) and then electrophoresed on 1.2% agarose (w/v) gel. The expression clones were then sequenced at the Eurofins MWG Operon across the recombination sites (*attB1* and *attB2* sites flanking the transgene) in order to make sure that the open reading frame is in-frame with the tag/reporter protein.

2.6 Introduction of pSoup into *A. tumefaciens*

The *A. tumefaciens* strain GV3101 is resistant to gentamycin and rifampicin and it is not spontaneously resistant to tetracycline. It is a good strain to work with as it typically does not clump in culture, thus facilitating in the preparation of competent cells and transformation purposes. The disarmed Ti plasmid present in the GV3101 strain provides the virulence (vir) gene functions. Due to this plasmid, the size of the pGreen-based vector (pGBPGWG) can be kept small so that it can replicate both in *E. coli* and in *Agrobacterium*. The T-DNA, within which genes to be transferred reside, is provided on the vector. The vector uses pSa replication locus, which has been subdivided into pSa ori and pSa replicase gene (*repA*). The replication functions of the vector are provided by repA gene present on a compatible plasmid, pSoup.

pSoup DNA was a generous gift from Dr. Angus Murphy, University of Maryland. The identity of pSoup was confirmed by restriction analysis with *NdeI* and *SalI* enzymes. The GV3101 strain was streaked on LB plates containing 50 μg/ml gentamycin and 10 μg/ml rifampicin prior to transformation and incubated at 30 °C for 48 hours. After 48 hours, a single resistant colony was picked to inoculate 3 ml of LB (containing gentamycin and rifampicin antibiotics). The culture was grown at 30 °C overnight on a roller drum. Then 0.5 mL of the overnight culture was used to inoculate 100 ml of LB and grown at 30 °C for about 6 hours until OD_{600} of the culture was between 0.5-1.0. The culture was uniformly distributed into six autoclaved Corex tubes and was centrifuged at 4000 rpm for 5 minutes in a JA-20 rotor at 4 °C. After discarding the supernatants, the pelleted cells

were resuspended in 15 ml of ice-cold 10% glycerol. The process was repeated several times. Aliquots of 50 µl of the electrocompetent GV3101 cells were placed in 1.5 ml microfuge tubes and stored at -80 °C.

A tube containing the electrocompetent GV3101 cells was removed from the -80 °C freezer and placed on ice for transformation purpose. Next 50 ng of pSoup DNA was added to the tube and were mixed gently. GV3101 cells and the pSoup DNA were then transferred to electroporation cuvettes (which were pre-chilled on ice) followed by electroporation at 2.5 V and 400 Ω using Bio-Rad Gene pulsar. The cuvette was then transferred to the ice bucket followed by the addition of 1 ml of S.O.C medium. The bacterial culture was shaken for 2 hours at 30 °C in a culture tube. After 2 hours, 100 µl of the culture was plated on LB plates containing kanamycin (50 µg/ml) and tetracycline (5 µg/ml) to select for both the bacterial host strain and the plasmid). The plates were incubated for 2 days at 30 °C, at which time colonies were visible. Eight colonies were picked and were grown in 3 ml of LB (containing kanamycin and tetracycline). After 2 days, dense growth was observed. Plasmid DNA was isolated and the presence of pSoup was confirmed by restriction digestion with *NdeI* and *SalI* enzymes for 2 hours at 37 °C. The restriction-digested DNA fragments were resolved on 1.2% (w/v) agarose gel followed by staining and de-staining process as mentioned before. Next, electrocompetent cells of GV3101 (pSoup) was prepared by the method mentioned above and were stored as aliquots at -80 °C.

2.7 Introduction of *GmPHOT* expression clones into GV3101 (pSoup)

As the expression clones of *GmPHOT1A*, *GmPHOT1B-1*, *GmPHOT2A* and *GmPHOT2B* were assembled and verified in *E. coli*, they were transformed into GV3101 (pSoup) by electroporation using 50 ng of DNA per transformation. The transformants were selected on LB agar plates containing tetracycline (5 μg/ml) and kanamycin (50 μg/ml). Kanamycin and tetracycline resistant colonies were individually picked with sterile toothpicks, spotted into fresh selective LB plates containing tetracycline (5 μg/ml) and kanamycin (50 μg/ml) for subsequent steps. The plates were incubated at 30 °C for 2 days. Then the same toothpick was dipped into a prepared reaction mixture for colony PCR. A 25 μl PCR reaction mixture containing 12.5 μl of 2X *GoTaq* Green Master mix (Promega, Madison) and 2 μl from 5 μM stocks of gene specific primers. The cycling parameters were 95 °C for 2 min followed by 30 cycles at 95 °C for 20 sec, 51 °C for 20 sec, 72°C for 1 min 20 sec with a final extension at 72 °C for 3 min.

The PCR products were electrophoresed on 1.2% agarose (w/v) gel using 1X TAE as the running buffer, followed by EtBr staining and de-staining. Two colonies from the LB plates were used to inoculate LB containing tetracycline (5 μg/ml) and kanamycin (50 μg/ml) and shaken in culture tubes at 30°C. Glycerol stocks of the overnight culture of two of the putative transformants from each of the four *GmPHOT* constructs, were prepared and stored at -80 °C until needed for the transformation of the *Arabidopsis*.

CHAPTER 3 RESULTS AND DISCUSSIONS

3.1 Experimental plan

The Goal of this project was to clone and sequence the protein coding region of all the 5 G*mPHOT*s. In order to accomplish this, reverse transcriptase-PCR was used to obtain cDNAs of the *GmPHOT* mRNAs using total soybean RNA as the template and gene-specific primers. The *GmPHOT* cDNAs obtained with either Platinum *Taq* DNA polymerase or Easy-A high fidelity cloning enzyme were then cloned into pCR8/TOPO vector while those obtained with *PfuUltra* II fusion HS DNA Polymerase were cloned into pENTR-D/TOPO vector. The cDNA clones were subjected to analysis with restriction enzymes in order to confirm the identity and correct orientation of the inserts.

The cDNA clones from each *GmPHOT* were then sequenced to ensure that the sequences in the Phytozome database are correct. The sequences of the *GmPHOT* cDNA clones were compared to the soybean sequences from the Phytozome database. Clones from at least 3 rounds of RT-PCR were generated to ensure that no PCR-induced mutations were present in the clones. Clones containing and lacking stop codons were generated for downstream fusions to the amino- and carboxy- termini respectively.

The next goal was to generate plant transformation constructs to express *GmPHOT*:*EGFP* fusions under the control of the 35S promoter using Gateway Technology. Gateway cloning is an efficient cloning method for transferring gene of interests into large-plasmid vectors. This facilitates the production of transgenic plants by *Agrobacterium*-mediated T-DNA transformation. The verified *GmPHOT* cDNA clones (entry clones) were recombined into a plant expression vector (pGBPGWG) by LR cloning reaction. The pGBPGWG is derived from pGreenII vector which obtains its replication function from pSoup plasmid that will be subsequently introduced into the *A. tumefaciens* strain. The entry clones (flanked by *attL1* and *attL2* sites) replace the *ccdB* gene present in the destination vector (flanked by attR1 and attR2 sites) and results in the production of an expression clone (flanked by *attB* sites) after transformation into *E.coli*. The EGFP fused to the carboxy-terminus of the entry clones would help detect the tissue and subcellular localization of the *GmPHOT*s in plants in future experiments. The identity of the gene of interest within the expression clone was confirmed by digestion with restriction enzymes. Next, the expression clones were sequenced across the recombination sites to confirm that the *GmPHOT* ORF is in frame with the open reading frame of EGFP.

The assembled expression clones were then transferred into *A. tumefaciens* strain GV3101 (pSoup) and verified by colony PCR. Plant expression constructs (35S::*GmPHOT*:*EGFP*) in *A. tumefaciens* will be used for transformation of the *Arabidopsis* in future experiments.

3.2 Obtaining cDNA copies of *GmPHOT* mRNAs

The cDNAs of each *GmPHOT* mRNA were obtained by RT-PCR using gene-specific primers (Tables 3, 4). The RT-PCR products were analyzed by agarose gel electrophoresis to assess quality and quantity. The predicted lengths were 2934 bp, 2949 bp, 2889 bp, 2973 bp and 2991 bp for *GmPHOT1A*, *GmPHOT1B-1*, *GmPHOT1B-2*, *GmPHOT2A* and *GmPHOT2B,* respectively. The RT-PCR products (containing stop codons) obtained using Superscript III One-step RT-PCR with Platinum *Taq*kit were analyzed by agarose gel electrophoresis. When I started working in this project, there was only one PHOT2 gene (*GmPHOT2A*) in the database. *GmPHOT2B* was identified during restriction analysis of *GmPHOT2A* clones that will be discussed below. Both *GmPHOT2A* and *GmPHOT2B* are obtained with the same primer pair for RT-PCR amplification. So the RT-PCR products of *GmPHOT2A* and *GmPHOT2B* co-migrate in agarose gel electrophoresis. The lengths of the cDNAs closely matched the predicted lengths from the phytozome database. One band for each of the *GmPHOT*s was obtained in each lane (Figure 1). Clones containing stop codons [(+) stop clones] will be used for fusions to the amino-terminus of the *GmPHOT* protein coding region. Thus (+) stop primers (primers containing stop codons) were used for the initial RT-PCR amplifications in most, but not all, cases.

Clones lacking stop codons [(-) stop clones] are also required for downstream analysis. This is required for creating in-frame fusions to carboxy-terminus of the *GmPHOT* coding regions in plant expression constructs. Thus clones containing the endogenous

stop codons were used to generate (-) stop clones by PCR using GoTaq Green Master Mix and gene-specific (-) stop primers (primers lacking stop codons). In order to reduce the rate of PCR-induced errors, Easy-A High-fidelity PCR cloning enzyme was used. The PCR reaction was followed by restriction digestion with *DpnI*. This restriction enzyme digests dam methylated template DNA minimizing the chance of the template DNA transforming into *E. coli*. The *DpnI*-digested PCR product was electrophoresed on an agarose gel. The PCR products were then cloned into appropriate TOPO vectors.

The Accuscript *PfuUltra* II RT-PCR kit was used to amplify *GmPHOT* mRNAs with (-) stop primers for those containing presumptive sequence errors. For each of the PCR products mentioned above, the cloning and sequencing would be subsequently discussed in detail.

3.3 Generation of entry clones by TOPO cloning

The RT-PCR products were cloned into pCR8/TOPO or pENTR-D/TOPO (depending upon the kit used for RT-PCR) in TOPO cloning reactions. The PCR products obtained with *Taq* DNA polymerase or Easy-A High-fidelity PCR cloning enzyme were cloned into pCR8/TOPO vector. PCR products from *PfuUltra* II fusion HS DNA Polymerase were cloned into pENTR-D/TOPO vector. Since we chose the Gateway technology for creating the final expression constructs for plant transformation, we selected pCR8/TOPO and pENTR-D/TOPO vectors as they are Gateway compatible. These

vectors have *attL1* and *attL2* sites flanking the insert. This facilitates site specific recombination reaction between the entry clone and the Gateway destination vector to produce expression clones. In addition to this, pENTR-D/TOPO vector has a directional TOPO cloning site for rapid and efficient cloning of the blunt ended PCR products. For this reason, the PCR products were cloned by adding four nucleotides, CACC, to the 5'-end of the forward primer. The pENTR-D/TOPO vector has a GTGG overhang which invades the PCR product and anneals to the added complimentary bases and stabilizes the directionality.

The entry clones (in either of the vectors) were transformed into *E. coli* cells as described in the materials and methods. The transformed cells were selected on LB agar plates containing spectinomycin (for pCR8/TOPO vector) or kanamycin (for pENTR-D/TOPO vector). The number of resistant colonies varied between 20 and 50. Eight to twelve transformants were picked and subjected to plasmid DNA isolation for subsequent analysis.

3.4 Analysis of the *GmPHOT* entry clones

Restriction analyses were performed to ensure the correct identity and orientation of the *GmPHOT* cDNA clones. Each *GmPHOT* clone was analyzed by digestion with restriction enzymes (Table 5). The table shows the predicted DNA fragment sizes for each digestion. The digestion patterns with *HindIII* or *PstI* restriction enzymes

distinguish between the different *GmPHOT* cDNA clones.The pCR8/TOPO *GmPHOT1A* and the pCR8/TOPO *GmPHOT1B-1* clones were first digested with *HindIII*, which should have 2 restriction sites in *GmPHOT1A* but no site in *GmPHOT1B-1* based on the predicted sequence. Next they were digested with *PstI*, which should have 2 restriction sites in *GmPHOT1B-1* but should not digest *GmPHOT1A*. The predicted fragment lengths produced by *HindIII* in *GmPHOT1A* are 5216 bp and 532 bp and that of PstI in *GmPHOT1B-1* are 3528 bp, 1166 bp, 835 bp. The DNA fragments resulting from restriction digestion were resolved by agarose gel electrophoresis. The *HindIII* fragment lengths of *GmPHOT1A* matched the predicted lengths and there was no digestion with *PstI* (Figure 2). Similarly, the *PstI* fragment lengths of *GmPHOT1B-1* matched the predicted lengths and there was no digestion with *HindIII* (Figures 3, 4). Thus the restriction patterns revealed the correct identity of the clones.

Determining the correct orientation of the *GmPHOT* cDNA clones was crucial for the downstream fusion reactions. For identifying the correct orientation, the clones were digested with restriction enzymes that have sites in the insert as well as in the vector (Table 6). The *GmPHOT1A* and *GmPHOT1B-1* clones were double digested with *EcoRV* and *BamHI*. *EcoRV* should have 1 restriction site in the vector while *BamHI* should have 1 site in the insert. It is necessary to have restriction sites both in the insert and in the vector for diagnosing the orientation of the insert. The predicted fragment lengths for the forward orientation of *GmPHOT1A* are 3357 bp and 2391 bp while that for reverse orientation are 4960 bp and 800 bp. Similarly, the predicted fragment lengths for the

forward orientation of *GmPHOT1B-1* are 3327 bp and 2436 bp while that for reverse orientation are 5200 bp and 750 bp. When the digestion products were resolved on agarose gel, two bands were observed for each clone. The fragment lengths matched the predicted lengths. An equal proportion of forward and reverse clones were obtained in case of *GmPHOT1A* (Figure 5) whereas 7 out of 10 forward clones were produced in case of *GmPHOT1B-1*(Figure 6).

Similarly, pENTR-D/TOPO *GmPHOT1B-2* was digested with *HindIII* or *PstI*, which should have 3 and 2 restriction sites respectively in the insert based on the predicted sequence. The predicted fragment lengths produced by *HindIII* are 3350 bp, 1749 bp and 370 bp while that produced by *PstI* are 4906 bp and 563 bp. The DNA fragments resulting from restriction digestion were resolved by agarose gel electrophoresis. The fragment lengths matched the predicted lengths, confirming the identity of *GmPHOT1B-2* (Figure 7). Next the clones were double digested with *EcoRV* and *BamHI* for orientation analysis. When the digestion products were resolved by agarose gel electrophoresis, the fragment lengths matched the predicted lengths of 3081 bp and 2388 bp. Therefore, all of the clones diagnosed were in the forward orientation (Figure 8).

The pENTR-D/TOPO *GmPHOT2A* clones were analyzed in a similar fashion with *HindIII* and *PstI*, each of which should have 2 sites in the insert based on the predicted sequence. The restriction fragment lengths matched the predicted lengths of 5382 bp and

171 bp for *HindIII* or 4397 bp and 1156 bp for *PstI* (Figures 9, 10). Interestingly, 5 clones of *GmPHOT2A* when digested with *PstI* had a restriction digestion pattern that was markedly different from the other clones of *GmPHOT2A* (Figure 11). Upon additional analysis of the Phytozome database, a paralog of *GmPHOT2A*, named as *GmPHOT2B*, was identified that would produce a PCR product using the same primer pair as that of *GmPHOT2A* because of sequence identity across the 5' and 3' ends of the protein coding regions. The pENTR-D/TOPO *GmPHOT2A* clones, that have *GmPHOT2B* insert, when digested with *PstI* should produce fragments of size 5043 bp and 528 bp based on the predicted sequence. The fragment lengths of these 5 clones indeed matched the predicted lengths (Figure 11). Digestion of the *GmPHOT2A* and *GmPHOT2B* clones with *HindIII* were similar as predicted, and produced the same 2 fragments when resolved on agarose gel (Figure 12).

Next, the clones were double digested with *HindIII* and *HaeII*, for orientation analysis. *HindIII* is predicted to have 2 restriction sites in the insert while *HaeII* should have 1 site in the insert and one in the vector. When the restriction digests were electrophoresed on an agarose gel, 4 fragments of the expected size were obtained for each clone. For *GmPHOT2A* clones, the predicted fragment lengths are 3468 bp, 1544 bp, 470 bp and170 bp for forward clones and 2880 bp, 2540 bp, 480 bp, 170bp for reverse clones. For *GmPHOT2B* clones, the predicted fragment lengths are 3471 bp, 1559 bp, 370 bp and 170 bp for forward clones and 2915 bp, 2885 bp, 370 bp and 170 bp for reverse clones. The fragment lengths of 11 out of 12 *GmPHOT2A* clones diagnosed matched the

predicted lengths for forward orientation. The fragment lengths of 3 out of 5 *GmPHOT2B* clones matched the predicted lengths for forward orientation (Figure 12, 13). The directional cloning by pENTR-D/TOPO vector is quite efficient as most of the clones were produced in the correct forward orientation, compared to the non-directional cloning by the pCR8/TOPO vector which produced an equal proportion of forward and reverse clones.

Clones from at least three separate rounds of RT-PCR were obtained for each *GmPHOT* to confirm the accuracy and integrity of the sequences of the gene models in the database and also to validate the accuracy of the cloned cDNA sequence. The cDNA inserts, of two clones for each *GmPHOT* that had the correct identity and orientation, were sequenced. DNA sequencing was necessary to ensure that the clones were free of misincorporations induced by reverse transcriptase enzyme or DNA polymerase enzyme during RT-PCR reaction or thermal cycler induced mutations. The entry clones were sequenced with M13 forward and M13 reverse primers (that flank the insert in the entry clone) along with other gene specific primers listed in tables 7, 8, 9, 10, 11. Each sequencing run spanned approximately 600 bp such that the entire clone was covered. A consensus nucleotide sequence was generated for each clone by majority rule combining the sequencing runs against the reference file of the respective *GmPHOT* obtained from the Phytozome database. The consensus sequence obtained was then translated into amino acid sequences and aligned with the predicted amino acid sequences from the Phytozome.

GmPHOT cDNA clones obtained with Platinum *Taq* DNA polymerase frequently contained PCR-induced mutations. Only the (+) stop entry clones of *GmPHOT1A* and *GmPHOT1B-1* in pCR8/TOPO perfectly matched the sequence from the Phytozome database. The (+) stop entry clones of *GmPHOT1B-2*, *GmPHOT2A* and *GmPHOT2B* in pCR8/TOPO had few PCR-induced errors and interestingly the *GmPHOT1B-2* clones possessed a 60 nucleotide sequence insertion that was absent in the reference file. The error free (+) stop clones of *GmPHOT1A* and *GmPHOT1B-1* were subsequently used to obtain (-) stop clones by PCR using *GoTaq* green master mix initially, which failed to produce any error-free clones. In order to avoid obtaining clones with PCR generated errors, Easy A high fidelity cloning enzyme was used as discussed in the materials and methods. Even when using this enzyme, PCR-induced mutations were obtained though at a much lower rate. The (-) stop entry clones for *GmPHOT1A* and *GmPHOT1B-1* in pCR8/TOPO that were obtained with this enzyme perfectly matched the sequences in the Phytozome database.

Due to scarcity of time, we proceeded with obtaining (-) stop clones of the remaining *GmPHOT*s that contained presumptive sequence errors, using *PfuUltra* II Fusion HS DNA polymerase (Accuscript *PfuUltra* II RT-PCR kit, Stratagene). *PfuUltra* II Fusion HS DNA polymerase was used to amplify *GmPHOT* mRNAs using gene-specific (-) stop primers. This enzyme, having an enhanced proof reading activity lowers the possibility of non-specific amplification and primer-dimer formation thus helping in producing PCR products with high throughput and efficiency. The (-) stop entry clones of *GmPHOT2A*

and *GmPHOT2B* in pENTR-D/TOPO produced using this enzyme perfectly matched the sequences in the Phytozome database. These (-) stop clones of *GmPHOT2A* and *GmPHOT2B* will be used to generate (+) stop clones when required.

The (-) stop entry clones from *GmPHOT1B-2* in pENTR-D/TOPO contain a 60 nucleotide insertion that is absent in the reference sequence from Phytozome. A consensus sequence for *GmPHOT1B-2* was generated by comparing the clones from all 7 rounds of RT-PCR (six rounds were performed using the Platinum Taq DNA polymerase to obtain (+) stop clones while 1 round was performed with the *PfuUltra*II Fusion HS DNA polymerase to obtain (-) stop clones. A pairwise alignment was performed between the consensus sequence and the database sequence. The consensus sequence matched the reference sequence except for the 60 nucleotide sequence insertion in the consensus sequence relative to the reference. The nucleotide sequences were then translated into amino acid sequences and then pairwise alignment was performed with the predicted amino acid sequences from the database. There was an additional 20 amino acids encoded by the entry clones thus producing a total of 982 amino acids for *GmPHOT1B-2* whereas the amino acid length of the database sequence is 962 (Figure 14). It seems likely that the database contains an inaccurate gene model for *GmPHOT1B*-2. Thus this work reveals a *GmPHOT1B-2* polypeptide length (as derived from the cDNA sequence) of 20 amino acids greater than that predicted in the database. This is probably due to a putative splice site at that region in the genomic sequence. But our work showed that the mRNA is not getting spliced at that site. This result is intriguing and might have

important physiological implications, which would be explored in later studies. The additional inserted sequence is present in the linker region in between the LOV2 domain and the catalytic domain. The current status of all the 5 *GmPHOT*s is summarized in Table 13.

The amino acid sequences encoded by the protein coding region of the 5 *GmPHOT* entry clones and polypeptide sequences from common bean, broad bean, garden pea, *Arabidopsis* were imported into the ClustalX 2.1 software. A multiple alignment of phototropin polypeptide sequences from all these above mentioned plant species was done. The phylogenetic analyses revealed that the soybean phototropins are highly conserved proteins with *GmPHOT1A*, *GmPHOT1B-1* and *GmPHOT1B-2* belonging to the phot1 clade while *GmPHOT2A* and *GmPHOT2B* belonging to the phot2 clade (Figure 15).

The *GmPHOT* polypeptide sequences were compared to the polypeptide sequences of the *Arabidopsis* phototropins in order to identify the structural similarity between soybean and *Arabidopsis* phototropins. Pairwise and multiple alignments were done using ClustalX 2.1 and then the aligned sequence was imported into BoxShade server (Figure 16). The aligned sequences were next imported into the sequence similarity and identity software for predicting the percentage of identity (Table 14). *GmPHOT1A, GmPHOT1B-1* and *GmPHOT1B-2* are 64.99%, 67.2% and 66.29% identical to *AtPHOT1* respectively

whereas *GmPHOT2A* and *GmPHOT2B* are 69.28% and 69.18% identical to *AtPHOT2*. *GmPHOT2A* and *GmPHOT2B* share 87.17% identity within their protein coding region. The percentage identity between the protein sequences of *GmPHOT* and *AtPHOT* depict their structural likeliness and also putatively indicate their functional resemblance.

Next the percentage identity between the LOV1 (Table 15), LOV2 (Table 16) and catalytic domains (Table 17) of *GmPHOT*s and *AtPHOT*s were predicted. These domains were found to be highly conserved between the two gene families. The LOV1 and LOV2 domains of *GmPHOT*s were more identical amongst each other than they were with the *Arabidopsis* sequence. Similar scenario was obtained in case of the catalytic domains. The LOV2 domains of *GmPHOT2A* and *GmPHOT2B,* which belong to the same clade, were 100% identical to each other. This is because *GmPHOT*s have arose by recent duplication. The polypeptide sequence encoded by the *GmPHOT1*A, *GmPHOT1B-1* and *GmPHOT1B-2* cDNA clones are 977, 982 and 982 amino acids in length and is co-linear throughout its entire length with *AtPHOT1*. Similarly the polypeptide sequence encoded by the *GmPHOT2A* and *GmPHOT2B* cDNA clones are 990 and 996 amino acids in length and is co-linear throughout its entire length with *AtPHOT2*. The *GmPHOT1* cloned protein sequences were identified to be smaller in length compared to that of *AtPHOT1*. In contrast to that, *GmPHOT2* cloned protein sequences were identified to be longer in length compared to that of *AtPHOT2*. Despite the differences in length, the soybean protein sequences are quite identical to that of *Arabidopsis* with the presence of

the conserved LOV1, LOV2 and catalytic domains, suggesting that *GmPHOT*s might be functional phototropins and thus encode an active photoreceptor.

3.5 Generation of plant expression constructs by LR cloning

The gene of interest from the each of the 4 entry clones except from *GmPHOT1B-2* for which we do not have any error free clone, was transferred into the destination vector by a recombination reaction using Gateway LR Clonase II enzyme mix as discussed in the materials and methods, resulting in the formation of an expression clone. The Gateway destination vector has *attR1* and *attR2* sites. The LR clonase enzyme mix catalyzes an in vitro recombination reaction between the *attL*-flanked entry clone and an *attR*-containing destination vector to generate *attB* sites flanking the transgene in the expression clone. The resulting expression clone (35 S::*GmPHOT:EGFP*) was then transformed into *E. coli* cells. A mixture of colonies containing expression clones and entry clones will be obtained because the entry clones obtained from pENTR-D/TOPO vector and the destination vectors have the same kanamycin antibiotic resistance marker. The entry clones were thus digested with *NruI* restriction enzyme, which has one recognition site in the pENTR-D/TOPO vector but not in the Gateway vector or the insert, prior to LR cloning. This reduced the chance of getting back colonies containing entry clones. The entry clones in pCR8/TOPO do not require digestion with *NruI* because of the presence of different antibiotic resistance marker.

For each resulting expression clone, ten kanamycin resistant colonies were subjected to plasmid DNA isolation and then analyzed by digestion with restriction enzymes listed in Table 12. The restriction analysis was performed to verify the identity of the transgenes. *EcoRI* was used to diagnose the *GmPHOT1*s because it has multiple restriction sites within the insert while *PstI* was used for the *GmPHOT2*s. These enzymes are predicted to have 4 restriction sites in the insert. The restriction fragments were resolved on an agarose gel. The fragment lengths of expression clones of *GmPHOT1A* (4442 bp, 2590 bp, 984 bp, 580 bp and 359 bp), *GmPHOT1B-1* (4442 bp, 2964 bp, 984 bp and 580 bp), *GmPHOT2A* (4655 bp, 2489 bp, 1156 bp, 706 bp) and *GmPHOT2B* (4655 bp, 1961 bp, 1945 bp, 528 bp) matched the predicted lengths from the phytozome database (Figures 17, 18, 19, 20). This confirmed the identity of *GmPHOT* expression clones after the LR recombination reaction. Two expression clones from each *GmPHOT* were then sequenced from within the inserts across the recombination sites (*attB1* and *attB2* sites flanking the transgene) in order to make sure that the *GmPHOT* ORFs were in-frame with the coding region of EGFP. The sequencing runs covered the 5’ end of the insert and the *attB1* site and also 3’ end of the insert and the *attB2* site. The nucleotide sequences obtained were translated into amino acid sequences and aligned with the predicted amino acid sequences from the database. The expression clones from each of the 4 *GmPHOT*s were found to have the *GmPHOT* protein coding region in-frame with the EGFP tag in the transgene. A schematic representation of the plant expression construct has been demonstrated (Figure 21).

3.6 Transformation of *A. tumefaciens*

As our Gateway vector is derived from pGreen, its replication functions are provided by the pSoup plasmid. Thus pSoup needs to be introduced into the *A. tumefaciens* strain GV3101 prior to its transformation with the expression clones. This strain is resistant to gentamycin and rifampicin and it gains tetracycline resistance when it harbors pSoup. This is a suitable strain to work with binary vectors that confer kanamycin resistance in bacteria, as this strain is sensitive to kanamycin. The identity of pSoup DNA was confirmed by digestion with *NdeI* and *SalI* restriction enzymes (each of which should have 1 site in the pSoup). Electrocompetent cells of GV3101 were prepared and then transformed with pSoup DNA by the process of electroporation as discussed in the materials and methods. Subsequently electrocompetent cells of GV3101 strain (pSoup) was prepared by the above discussed method.

Next, the assembled expression clones of *GmPHOT1A*, *GmPHOT1B-1*, *GmPHOT2A* and *GmPHOT2B* were transformed with GV3101 (pSoup) by electroporation. Following transformation, the identity of the *GmPHOT* inserts was analyzed by colony PCR using gene-specific primers. The PCR products were then analyzed by agarose gel electrophoresis. The fragment lengths matched the predicted lengths that were 2934 bp, 2949 bp, 2973 bp and 2991 bp for *GmPHOT1A*, *GmPHOT1B-1*, *GmPHOT2A* and *GmPHOT2B* respectively (Figures 22, 23, 24, 25).

CHAPTER 4 FUTURE PROSPECTS

PsPHOT1A has been found to be a functional phototropin as evident from the phototropism assays [52]. The functions of the soybean phototropins are not yet known. Soybean, being an important crop plant, is on the top list of crops whose productivity increase is of prime concern. Determination of the functions of its phototropins, which have a role in optimizing photosynthetic efficiency, is crucial. Our hypothesis is that soybean phototropins encode an active photoreceptor for mediating phototropic responses. This can be resolved only if the *Arabidopsis* phototropin null mutants can be complemented by soybean phototropins. The plant transformation constructs have been created for the four soybean phototropins via Gateway technology and *A. tumefaciens* transformation. *GmPHOT1* constructs will be introduced into *Arabidopsis phot1* mutant, while *GmPHOT2* constructs will be introduced into *Arabidopsis phot2* mutant by modified floral dip method [71]. Apart from this, the *GmPHOT1* and *GmPHOT2* constructs will be introduced separately into *Arabidopsis phot1 phot2* double mutant plants.

After creating several independent transgenic lines that express *GmPHOT* transgenes, the transgenes can be detected by amplification of genomic DNA (extracted from leaf tissues) by PCR with gene specific primers. Once the presence of transgene is confirmed,

several analyses will be performed to check if the transgenes are able to complement the expected phototropin defects. One of the analyses includes a phototropism assay, in which the young seedlings from the transgenic lines will be grown on vertical agar plates and subjected to unilateral blue light at a fluence rate of 0.8 µmole $m^{-1}s^{-1}$.

Although there has been success in complementing the Arabidopsis *phot1* mutants using the 35S promoter to drive PsPHOT1A expression, it might be that normal expression levels with proper cell type specificity will be needed for the GmPHOTs. Thus, a second set of plant transformation constructs will be made with authentic Arabidopsis PHOT promoters (PHOT1 promoter to drive the GmPHOT1 series and the PHOT2 promoter for GmPHOT2 series) to drive expression of fusions of the GmPHOTs to Green Fluorescent Protein (GFP) coding region to determine the tissue and subcellular localization of the GmPHOTs, which is currently unknown but has important implications about the function of these photoreceptors.

Light absorption characteristics of the leaves will be measured using a LICOR 6400. Other phototropin-dependent processes including light-dependent chloroplast movements, cotyledon growth, stomatal opening, leaf expansion, leaf positioning, regulation of hypocotyl elongation will also be investigated. Light dependent orientation of primary leaves, stomatal opening and leaf solar tracking in older soybean plants will also be investigated.

These experiments will provide a solid understanding of blue light responses in soybean that are essential for us to understand the effects of manipulating GmPHOT expression levels on processes that will impact photosynthesis in transgenic soybean. These studies are important because much of what we know about specific phototropin actions is based only on Arabidopsis. Since soybean and other legumes have more than one form of phot1 and phot2, unlike Arabidopsis, they may possess a specialization of function not present in Arabidopsis.

It has been shown experimentally that OsPHOT1 mRNA transcript levels remain mostly unaffected by irradiation with white light showing basically basal level expression in the leaves of dark grown seedlings. While OsPHOT2 mRNA transcript levels in leaves significantly increased (5 fold increase) relative to the control plants when irradiated with 40 µmole m^{-1} s^{-1} of white light for 4 hours. When 6 day old etiolated seedlings of Arabidopsis are irradiated with white, red, blue/UV-A lights, they showed an elevation in the PHOT2 gene expression levels. mRNA transcript levels were maximum in leaves [24, 25].

When dark grown pea seedlings were illuminated with continuous white light, a significant decrease (10 fold) in the mRNA transcript levels of PsPK5 was observed while there was only about 4 fold decreases in the mRNA transcript levels of PsPK3. The transcript levels of PsPK5 and PsPK3 were unaffected when subjected to single pulses of red, far-red and blue light. Thus PsPK5 and PsPK3 genes are responsive to continuous light but not to single pulses of light [55].

Different tissue specific expression of soybean phototropin genes have been interpreted by RNA sequencing [58]. *GmPHOT1A* and *GmPHOT1B-1* have been identified to be highly expressed in young leaves and in flowers, with low level of expression in seeds, roots and nodules. *GmPHOT1B-2* and *GmPHOT2A* have comparatively low expression in these tissues. Thus, the expression of all the 5 *GmPHOT* genes at the mRNA level will be analyzed using RT-PCR. Gene-specific primers will be designed and their authenticity will be ensured by subjecting the generated *GmPHOT* clones to PCR amplification using these primers. Once the primers are ensured to be full-proof, template RNAs from a variety of organs and tissues from seedlings and mature plants will be analyzed by RT-PCR. With seedlings, it would be interesting to find out if the two PHOT1 clades exhibit differential expression when etiolated seedlings are exposed to light. PsPHOT1B transcripts from pea seedlings were observed to be rapidly down regulated by light whereas PsPHOT1A transcripts remain relatively unchanged [56, 72]. Basic knowledge of such developmental differences will allow us to determine which phototropin genes are important for controlling blue light photosensitivity during different developmental phases.

Transgenic soybean lines will be created with altered expression levels of the endogenous phototropins. Lines that either under-express the phototropin gene products or that ectopically express these genes will be generated and will be tested for alterations in phototropin dependent pathways. The knowledge gained will provide new approaches for modification of useful traits to include in commercial soybean production. By examining a broad range of conditions to assess the outcome of altering phototropin expression in

soybean, it would provide a means to optimize the photosynthetic efficiency in soybean. Such modifications are expected to result in improved biomass and ultimately crop yield of a major Indiana crop species.

LIST OF REFERENCES

LIST OF REFERENCES

1. Chen, M. and J. Chory, *Phytochrome signaling mechanisms and the control of plant development.* Trends Cell Biol, 2011. **21**(11): p. 664-71.
2. Franklin, K.A. and P.H. Quail, *Phytochrome functions in Arabidopsis development.* J Exp Bot, 2010. **61**(1): p. 11-24.
3. Kendrick RE, K.G., *Photomorphogenesis in Plants*. 1994, Dordrecht, The Netherlands: Kluwer Academic Publishers.
4. Briggs, W.R., J.M. Christie, and M. Salomon, *Phototropins: a new family of flavin-binding blue light receptors in plants.* Antioxid Redox Signal, 2001. **3**(5): p. 775-88.
5. Ahmad, M., C. Lin, and A.R. Cashmore, *Mutations throughout an Arabidopsis blue-light photoreceptor impair blue-light-responsive anthocyanin accumulation and inhibition of hypocotyl elongation.* Plant J, 1995. **8**(5): p. 653-8.
6. Lin, C., et al., *Enhancement of blue-light sensitivity of Arabidopsis seedlings by a blue light receptor cryptochrome 2.* Proc Natl Acad Sci U S A, 1998. **95**(5): p. 2686-90.
7. Noh, B. and E.P. Spalding, *Anion channels and the stimulation of anthocyanin accumulation by blue light in Arabidopsis seedlings.* Plant Physiol, 1998. **116**(2): p. 503-9.
8. Zimmerman, B.K. and W.R. Briggs, *Phototropic Dosage-Response Curves for Oat Coleoptiles.* Plant Physiol, 1963. **38**(3): p. 248-53.
9. Briggs, W.R., *Light Dosage and Phototropic Responses of Corn and Oat Coleoptiles.* Plant Physiol, 1960. **35**(6): p. 951-62.
10. Darwin, C., *The power of movement in plants*. 1966, New York: Da Capo Press.
11. Christie, J.M. and A.S. Murphy, *Shoot phototropism in higher plants: new light through old concepts.* Am J Bot, 2013. **100**(1): p. 35-46.
12. Salomon, M., Zacherl, M., and Rudiger, W., *Asymmetric, blue light-dependent phosphorylation of a 116-kilodalton plasma membrane protein can be correlated with the first- and second-positive phototropic curvature of oat coleoptiles.* Plant Physiology, 1997. **115**: p. 485-491.
13. Iino, M., *Phototropism in higher plants*, in *Photomovement: ESP Comprehensive Series in Photosciences*, D.H.a.M. Lebert, Editor. 2001, Elsevier: Amsterdam.
14. Janoudi, A.K. and K.L. Poff, *Characterization of adaptation in phototropism of Arabidopsis thaliana.* Plant Physiol, 1991. **95**: p. 517-21.
15. Parker, K., T.I. Baskin, and W.R. Briggs, *Evidence for a phytochrome-mediated phototropism in etiolated pea seedlings.* Plant Physiol, 1989. **89**(2): p. 493-7.
16. Gallagher, S., et al., *Light-mediated changes in two proteins found associated with plasma membrane fractions from pea stem sections.* Proc Natl Acad Sci U S A, 1988. **85**(21): p. 8003-7.

17. Salomon, M., et al., *Exposure of oat seedlings to blue light results in amplified phosphorylation of the putative photoreceptor for phototropism and in higher sensitivity of the plants to phototropic stimulation.* Plant Physiol, 1997. **115**(2): p. 493-500.
18. Salomon, M., M. Zacherl, and W. Rudiger, *Asymmetric, blue light-dependent phosphorylation of a 116-kilodalton plasma membrane protein can be correlated with the first- and second-positive phototropic curvature of oat coleoptiles.* Plant Physiol, 1997. **115**(2): p. 485-91.
19. Short TW, B.W., *The transduction of blue light signals in higher plants. 45:.* Annual Review of Plant Physiology and Plant Molecular Biology, 1994. **45**: p. 143-171.
20. Palmer, J.M., et al., *Blue Light-Induced Phosphorylation of a Plasma Membrane-Associated Protein in Zea mays L.* Plant Physiol, 1993. **102**(4): p. 1211-1218.
21. Liscum, E. and W.R. Briggs, *Mutations in the NPH1 locus of Arabidopsis disrupt the perception of phototropic stimuli.* Plant Cell, 1995. **7**(4): p. 473-85.
22. Huala, E., et al., *Arabidopsis NPH1: a protein kinase with a putative redox-sensing domain.* Science, 1997. **278**(5346): p. 2120-3.
23. Christie, J.M., et al., *Arabidopsis NPH1: a flavoprotein with the properties of a photoreceptor for phototropism.* Science, 1998. **282**(5394): p. 1698-701.
24. Jarillo, J.A., et al., *Phototropin-related NPL1 controls chloroplast relocation induced by blue light.* Nature, 2001. **410**(6831): p. 952-4.
25. Kagawa, T., et al., *Arabidopsis NPL1: a phototropin homolog controlling the chloroplast high-light avoidance response.* Science, 2001. **291**(5511): p. 2138-41.
26. Christie, J.M., et al., *LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (nph1): binding sites for the chromophore flavin mononucleotide.* Proc Natl Acad Sci U S A, 1999. **96**(15): p. 8779-83.
27. Nakasone, Y., et al., *Kinetic measurement of transient dimerization and dissociation reactions of Arabidopsis phototropin 1 LOV2 domain.* Biophys J, 2006. **91**(2): p. 645-53.
28. Knieb, E., M. Salomon, and W. Rudiger, *Tissue-specific and subcellular localization of phototropin determined by immuno-blotting.* Planta, 2004. **218**(5): p. 843-51.
29. Kong, S.G., et al., *Blue light-induced association of phototropin 2 with the Golgi apparatus.* Plant J, 2006. **45**(6): p. 994-1005.
30. Inada, S., et al., *RPT2 is a signal transducer involved in phototropic response and stomatal opening by association with phototropin 1 in Arabidopsis thaliana.* Plant Cell, 2004. **16**(4): p. 887-96.
31. Lariguet, P., et al., *PHYTOCHROME KINASE SUBSTRATE 1 is a phototropin 1 binding protein required for phototropism.* Proc Natl Acad Sci U S A, 2006. **103**(26): p. 10134-9.
32. Baum, G., et al., *Stimulation of the blue light phototropic receptor NPH1 causes a transient increase in cytosolic Ca2+.* Proc Natl Acad Sci U S A, 1999. **96**(23): p. 13554-9.
33. Vanneste, S. and J. Friml, *Auxin: a trigger for change in plant development.* Cell, 2009. **136**(6): p. 1005-16.

34. Peer, W.A., et al., *Seven things we think we know about auxin transport.* Mol Plant, 2011. **4**(3): p. 487-504.
35. Haga, K. and T. Sakai, *PIN auxin efflux carriers are necessary for pulse-induced but not continuous light-induced phototropism in Arabidopsis.* Plant Physiol, 2012. **160**(2): p. 763-76.
36. Friml, J., et al., *Lateral relocation of auxin efflux regulator PIN3 mediates tropism in Arabidopsis.* Nature, 2002. **415**(6873): p. 806-9.
37. Briggs, W.R. and J.M. Christie, *Phototropins 1 and 2: versatile plant blue-light receptors.* Trends Plant Sci, 2002. **7**(5): p. 204-10.
38. Hanks, S.K. and T. Hunter, *Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification.* Faseb j, 1995. **9**(8): p. 576-96.
39. Taylor, B.L. and I.B. Zhulin, *PAS domains: internal sensors of oxygen, redox potential, and light.* Microbiol Mol Biol Rev, 1999. **63**(2): p. 479-506.
40. Crosson, S. and K. Moffat, *Structure of a flavin-binding plant photoreceptor domain: insights into light-mediated signal transduction.* Proc Natl Acad Sci U S A, 2001. **98**(6): p. 2995-3000.
41. DeBlasio, S.L., D.L. Luesse, and R.P. Hangarter, *A plant-specific protein essential for blue-light-induced chloroplast movements.* Plant Physiol, 2005. **139**(1): p. 101-14.
42. Sakai, T., et al., *Arabidopsis nph1 and npl1: blue light receptors that mediate both phototropism and chloroplast relocation.* Proc Natl Acad Sci U S A, 2001. **98**(12): p. 6969-74.
43. Inoue, S., et al., *Blue light-induced autophosphorylation of phototropin is a primary step for signaling.* Proc Natl Acad Sci U S A, 2008. **105**(14): p. 5626-31.
44. Matsuoka, D. and S. Tokutomi, *Blue light-regulated molecular switch of Ser/Thr kinase in phototropin.* Proc Natl Acad Sci U S A, 2005. **102**(37): p. 13337-42.
45. Kasahara, M., et al., *Photochemical properties of the flavin mononucleotide-binding domains of the phototropins from Arabidopsis, rice, and Chlamydomonas reinhardtii.* Plant Physiol, 2002. **129**(2): p. 762-73.
46. Christie, J.M., et al., *Phototropin LOV domains exhibit distinct roles in regulating photoreceptor function.* Plant J, 2002. **32**(2): p. 205-19.
47. Kagawa, T., et al., *Function analysis of phototropin2 using fern mutants deficient in blue light-induced chloroplast avoidance movement.* Plant Cell Physiol, 2004. **45**(4): p. 416-26.
48. Okajima, K., D. Matsuoka, and S. Tokutomi, *LOV2-linker-kinase phosphorylates LOV1-containing N-terminal polypeptide substrate via photoreaction of LOV2 in Arabidopsis phototropin1.* FEBS Lett, 2011. **585**(21): p. 3391-5.
49. Christie, J.M., et al., *phot1 inhibition of ABCB19 primes lateral auxin fluxes in the shoot apex required for phototropism.* PLoS Biol, 2011. **9**(6): p. e1001076.
50. Kinoshita, T., et al., *Phot1 and phot2 mediate blue light regulation of stomatal opening.* Nature, 2001. **414**(6864): p. 656-60.
51. Sakamoto, K. and W.R. Briggs, *Cellular and subcellular localization of phototropin 1.* Plant Cell, 2002. **14**(8): p. 1723-35.

52. Oakley, R.V., *The PsPK4 gene from the garden pea encodes phototropin1,* in *Biology*. 2002, Purdue University: Indianapolis.
53. Jain M, S.P., Tyagi SB, Tyagi AK, Khurana JP. , *Light regulation and differential tissue-specific expression of phototropin homologues from rice (Oryza sativa ssp. indica).* . Plant Science, 2007. **172**: p. 164-171.
54. Suzuki, H., et al., *Blue-light regulation of ZmPHOT1 and ZmPHOT2 gene expression and the possible involvement of Zmphot1 in phototropism in maize coleoptiles.* Planta, 2014.
55. Khanna, R., X. Lin, and J.C. Watson, *Photoregulated expression of the PsPK3 and PsPK5 genes in pea seedlings.* Plant Mol Biol, 1999. **39**(2): p. 231-42.
56. Elliott, R.C., et al., *Phytochrome regulation of pea phototropin.* J Plant Physiol, 2004. **161**(3): p. 265-70.
57. Schmutz, J., et al., *Genome sequence of the palaeopolyploid soybean.* Nature, 2010. **463**(7278): p. 178-83.
58. Severin, A.J., et al., *RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome.* BMC Plant Biol, 2010. **10**: p. 160.
59. Yadav, N.S., et al., *Short direct repeats flank the T-DNA on a nopaline Ti plasmid.* Proc Natl Acad Sci U S A, 1982. **79**(20): p. 6322-6.
60. Jouanin, L., et al., *Analysis of TR-DNA/plant junctions in the genome of a Convolvulus arvensis clone transformed by Agrobacterium rhizogenes strain A4.* Plant Mol Biol, 1989. **12**(1): p. 75-85.
61. Ruvkun, G.B. and F.M. Ausubel, *A general method for site-directed mutagenesis in prokaryotes.* Nature, 1981. **289**(5793): p. 85-8.
62. Gelvin, S.B., *Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool.* Microbiol Mol Biol Rev, 2003. **67**(1): p. 16-37, table of contents.
63. A. hoekema, p.r.h., P. j. j. hooykaas & R. a. schilperoort, *A binary plant vector strategy based on separation of vir- and T-region of the Agrobacterium tumefaciens Ti-plasmid.* Nature, 1983. **303**: p. 179-180.
64. de Frammond, A.J., K. A. Barton, and M.-D. Chilton. , *Mini-Ti-: a new vector strategy for plant genetic engineering.* Bio/Technology, 1983. **5**: p. 262-269.
65. Miranda, A., et al., *Agrobacterium tumefaciens transfers extremely long T-DNAs by a unidirectional mechanism.* J Bacteriol, 1992. **174**(7): p. 2288-97.
66. Karimi, M., D. Inze, and A. Depicker, *GATEWAY vectors for Agrobacterium-mediated plant transformation.* Trends Plant Sci, 2002. **7**(5): p. 193-5.
67. Hartley, J.L., G.F. Temple, and M.A. Brasch, *DNA cloning using in vitro site-specific recombination.* Genome Res, 2000. **10**(11): p. 1788-95.
68. Katzen, F., *Gateway((R)) recombinational cloning: a biological operating system.* Expert Opin Drug Discov, 2007. **2**(4): p. 571-89.
69. Zhong, S., et al., *Improved plant transformation vectors for fluorescent protein tagging.* Transgenic Res, 2008. **17**(5): p. 985-9.
70. Christie, J.M., *Phototropin blue-light receptors.* Annu Rev Plant Biol, 2007. **58**: p. 21-45.

71. Miguel Martinez-Trujillo, V.L.-B., José Luis Cabrera-Ponce, Luis Herrera-Estrella, *Improving transformation efficiency ofArabidopsis thaliana by modifying the floral dip method.* Plant Molecular Biology Reporter, 2004. **22**(1): p. 63-70
72. Lin, C., et al., *Association of flavin adenine dinucleotide with the Arabidopsis blue light receptor CRY1.* Science, 1995. **269**(5226): p. 968-70.

TABLES

Table 1. **The range of phototropin-induced responses in plants**. Both phototropin1 and phototropin2 have some redundant functions as well as some specific functions. Listed below are the responses that are mediated either by phot1 or phot2 and that require both the photoreceptors [70]

phot1 alone	phot2 alone	Both phot1 and phot2
Phototropism (low fluence)	Phototropism (high fluence)	Leaf expansion
Chloroplast accumulation	Chloroplast accumulation	
Hypocotyl inhibition	Chloroplast avoidance (high fluence)	
Stomatal opening	Stomatal opening	
Leaf movement	Leaf movement	
Leaf positioning		
mRNA stability		

Table 2. The 5 phototropin genes of the soybean genome: The soybean genome encodes 5 phototropin genes, 3 of which belong to the phot 1 class and 2 belong to the phot2 class.

Proposed Name	**Gene ID (Phytozome)**
GmPHOT1A	Glyma12g07890
GmPHOT1B-1	Glyma13g40550
GmPHOT1B-2	Glyma15g04850
GmPHOT2A	Glyma16g19560
GmPHOT2B	Glyma08g33545

Table 3. PCR primers used to amplify soybean phototropin1 mRNAs. These primers were used to amplify GmPHOT1A (Gm12), GmPHOT1B-1 (Gm13) and GmPHOT1B-2 (Gm15) genes. The sense primers for GmPHOT1B-2 included with CACC that are necessary for the directional TOPO cloning are also listed. (F=forward or sense primer, FC=forward construct with CACC, R=reverse or antisense primer)

Primer name	Primer sequence	Primer orientation
Gm 12 F	ATGGAGGCGTTTCCAAGAGACCAAC	Sense
Gm 12 FC	CACCATGGAGGCGTTTCCAAGAGACCAAC	Sense
Gm12 R +stop	TTAGAAGACGTTCATGTCCTCTTGCACTTG	Antisense
Gm 12 R -stop	GAAGACGTTCATGTCCTCTTGCACTTGAT	Antisense
Gm 13 F	ATGGAGCAGTCAGAAAAATCGCCCACG	Sense
Gm 13 FC	CACCATGGAGCAGTCAGAAAAATCGCCCACG	Sense
Gm 13 R +stop	TTAGAAAATATTTGTCTGCAGATCCTCTAGCCCAGG	Antisense
Gm 13 R -stop	GAAAATATTTGTCTGCAGATCCTCTAGCCCAGG	Antisense
Gm 15 F	ATGGAGCAGTCAGAAAAATCGCCTAAA	Sense
Gm 15 FC	CACCATGGAGCAGTCAGAAAAATCGCCTAAA	Sense
Gm 15 R +stop	TCAAAAAACATTTGCCTGTAGATCCTCTAGC	Antisense
Gm 15 R -stop	AAAAACATTTGCCTGTAGATCCTCTAGCC	Antisense

Table 4. PCR primers used to amplify soybean phototropin2 mRNAs. These primers were used to amplify GmPHOT2A (Gm16) and GmPHOT2B (Gm8) genes. Both GmPHOT2A and GmPHOT2B have the same primer pair. The sense primers included with CACC that are necessary for the directional TOPO cloning are also listed.

Primer name	Primer sequence	Primer orientation
Gm16 F	ATGGAGAAACTGAAAGTGTCTGCTAAGAATGATCC	Sense
Gm16 FC	CACCATGGAGAAACTGAAAGTGTCTGCTAAGAATGATCC	Sense
Gm16 R +stop	TCAAAAAATATCCATATCAATAGAGCTAACTAGCACTCC	Antisense
Gm16 R -stop	AAAAATATCCATATCAATAGAGCTAACTAGCACTCCATC AT	Antisense

Table 5. Predicted restriction fragment sizes of the entry clones diagnosed for identity. The pCR8/TOPOor pENTR-D/TOPO *GmPHOT* entry clones diagnosed by specific restriction endonucleases. Eight to twelve plasmid DNA isolates for each clone were analyzed by restriction digestion.

Entry clone	Predicted fragment sizes (bp) when digested by *HindIII*	Figure no.	Predicted fragment sizes (bp) when digested by *PstI*	Figure no.
pCR8/ TOPO GmPHOT1A	5216 bp, 532 bp	5	No digestion	5
pCR8/ TOPO GmPHOT1B-1	No digestion	6	3528 bp, 1166	7
pENTR-D/ TOPO GmPHOT1B-2	3350,1749,370	10	4906,563	10
pENTR-D/ TOPO GmPHOT2A	5382,171	12	4397,1156	13
pENTR-D/ TOPO GmPHOT2B	5400,171	15	5043,528	14

Table 6. Predicted restriction fragment sizes of the entry clones diagnosed for orientation. The pCR8/TOPO or pENTR-D/TOPO *GmPHOT* entry clones diagnosed by specific restriction endonucleases. Eight to twelve plasmid DNA isolates for each clone were analyzed by restriction digestion

Entry clone	Predicted fragment sizes (bp) when digested by *EcoRV* and *BamHI*	Figure no.	Predicted fragment sizes (bp) when digested by *HindIII*/*HaeII*	Figure no.
pCR8/TOPO GmPHOT1A	a) 3357,2391 (forward orientation) b) 4960,800 (reverse orientation)	8		
pCR8/TOPO GmPHOT1B-1	a) 3327,2436 (forward orientation) b) 5200,750 (reverse orientation)	9		
pENTR-D/TOPO GmPHOT1B-2	a) 3081,2388 (forward orientation) b) 5088,780 (reverse orientation)	11		
pENTR-D/TOPO GmPHOT12A			a) 3468,1544,470, 170 (forward orientation) b) 2880,2540,480, 170 (reverse orientation)	15, 16
pENTR-D/TOPO GmPHOT12B			a) 3471,1559,370, 170 (forward orientation) b) 2915,2885,370, 170 (reverse orientation)	15

Table 7. GmPHOT1A (Gm12) sequencing primers. The primer sequences start at the 5' end and the position is relative to the transcription start codon (ATG).

Primer name	Primer sequence	Primer orientation	Primer position from ATG
12-1	GATGACGGGCTATACATCAAAG	Sense	600-1200
12-1R	CTTTGATGTATAGCCCGTCATC	Antisense	600-1
12-2	CATCAGGAAAAGCCAGTCTAACTT	Sense	1200-1800
12-2R	AAGTTAGACTGGCTTTTCTGATG	Antisense	1200-600
12-3	TGCTAATATGAAACCAGAAGATTTATG	Sense	1800-2400
12-3R	CATAAATCTTCTGGTTTCATATTAGCA	Antisense	1800-1200
12-4	ACAGCTTCTAGTTCCAGTTATAAATGA	Sense	2400-2934
12-4R	TCATTTATAACTGGAACTAGAAGCTGT	Antisense	2400-1800

Table 8. GmPHOT1B-1 (Gm13) sequencing primers. The primer sequences start at the 5' end and the position is relative to the transcription start codon (ATG).

Primer name	Primer sequence	Primer orientation	Primer position from ATG
13-1	TGGGAGAAATTGTCGTTTTTTAC	Sense	600-1200
13-1R	GTAAAAAACGACAATTTCTCCCA	Antisense	600-1
13-2	GTTCAGGAGGAAAAGTCAGTCC	Sense	1200-1800
13-2R	GGACTGACTTTTCCTCCTGAAC	Antisense	1200-600
13-3	GGATGCTAATAAGAAACCAGATGATT	Sense	1800-2400
13-3R	AATCATCTGGTTTCTTATTAGCATCC	Antisense	1800-1200
13-4	GCCACAGCTTATAATTCCAGCT	Sense	2400-2929
13-4R	AGCTGGAATTATAAGCTGTGGC	Antisense	2400-1800

Table 9. GmPHOT1B-2 (Gm 15) sequencing primers. The primer sequences start at the 5' end and the position is relative to the transcription start codon (ATG).

Primer name	Primer sequence	Primer orientation	Primer position from ATG
15-1	AAATTGCCGGTTTTTACAGG	Sense	600-1200
15-1R	CCTGTAAAAACCGGCAATTT	Antisense	600-1
15-2	ATTCAGGAGGAAAAGTCAGTCC	Sense	1200-1800
15-2R	GGACTGACTTTTCCTCCTGAAT	Antisense	1200-600
15-3	GGATGCTAACAAGATGATTTATGACA	Sense	1800-2400
15-3R	TGTCATAAATCATCTTGTTAGCATCC	Antisense	1800-1200
15-4	GAAGCCAGAAAATGTGTTACTCAA	Sense	2400-3000
15-4R	TTGAGTAACACATTTTCTGGCTTC	Antisense	2400-1800
15-5	GCTAGAGGATCTACAGGCAAATGT	Sense	3000-3030
15-5R	ACATTTGCCTGTAGATCCTCTAGC	Antisense	3000-2400

Table 10. GmPHOT2A (Gm16) sequencing primers. The primer sequences start at the 5' end and the position is relative to the transcription start codon (ATG).

Primer name	Primer sequence	Primer orientation	Primer position from ATG
16-1	GAATCTTCTCACTATCACCCCC	Sense	600-1200
16-1R	GGGGGTGATAGTGAGAAGATTC	Antisense	600-1
16-2	GTTTCTTCAAGGACCAGAAACAG	Sense	1200-1800
16-2R	CTGTTTCTGGTCCTTGAAGAAAC	Antisense	1200-600
16-3	GATTATATCATTATTGGATCATTCTTT	Sense	1800-2400
16-3R	AAAGGATGATCCAATAATGATATAATC	Antisense	1800-1200
16-4	GGCTAGTCTTGCAGGCAGG	Sense	2400-2929
16-4R	CCTGGCTGCAAGACTAGCC	Antisense	2400-1800

Table 11. GmPHOT2B (Gm8) sequencing primers

Primer name	Primer sequence	Primer orientation	Primer position from ATG
8-1	TGTTGTCTCAGATGCAACAAAAC	Sense	600-1200
8-1R	GTTTTGTTGCATCTGAGACAACA	Antisense	600-1
16-2	GTTTCTTCAAGGACCAGAAACAG	Sense	1200-1800
16-2R	CTGTTTCTGGTCCTTGAAGAAAC	Antisense	1200-600
16-3	GATTATATCATTATTGGATCATTCTTT	Sense	1800-2400
16-3R	AAAGGATGATCCAATAATGATATAATC	Antisense	1800-1200
8-4	TGTGTTAGCTGATTTTGATCTATCATT	Sense	2400-2995
8-4R	AATGATAGATCAAAATCAGCTAACACA	Antisense	2400-1800

Table 12. **Predicted restriction fragment sizes of the expression clones**. The expression clones were analyzed with specific restriction endonucleases. Eight plasmid DNA isolates for each clone were analyzed by restriction digestion

Expression clone	Restriction enzyme	Number of DNA fragments	Predicted fragment sizes (bp)	Figure No.
GmPHOT1A	*EcoRI*	4	4442,2590,984,580,359	19
GmPHOT1B-1	*EcoRI*	4	4442,2964,984,580	20
GmPHOT1B-2	*EcoRI*	4	4655,2179,1525,563	-
GmPHOT2A	*PstI*	4	4655,2489,1156,706	21
GmPHOT2B	*PstI*	4	4655,1961,1945,528	22

Table 13. Current status of GmPHOTs: Soybean phototropins are designated as 1A (for GmPHOT1A), 1B-1 (for GmPHOT1B-1), 1B-2 (for GmPHOT1B-2), 2A (for GmPHOT2A) and 2B (for GmPHOT2B) in this table.Superscript III One-step RT-PCR with Platinum *Taq* kit has been denoted as Platinum *Taq* andAccuscript *PfuUltra* II RT-PCR kit as *PfuUltra.*

GmPHOT	RT PCR (+) or (-) stop clone	RT-PCR kit	Round of RT-PCR	TOPO vector	Consensus matches Phytozome	Error free clone	PCR (+) or (-) stop clone obtained
1A	(+) stop	Platinum *Taq*	3	pCR8	Yes	Yes	(-) stop
1B-1	(+) stop	Platinum *Taq*	3	pCR8	Yes	Yes	(-) stop
1B-2	(+) stop (-) stop	Platinum *Taq* *PfuUltra*	6 1	pCR8 pENTR-D	No* No*	No No	
2A	(+) stop (-) stop	Platinum *Taq* *PfuUltra*	3 1	pCR8 pENTR-D	Yes Yes	No Yes	
2B	(+) stop (-) stop	Platinum *Taq* *PfuUltra*	2 1	pCR8 pENTR-D	Yes Yes	No Yes	

*****matches phytozome except for the 60 nucleotide insertion

Table 14: Percentage identity between the protein coding regions of Arabidopsis and soybean phototropins. *Arabidopsis* phototropins are designated as AtPHOT1 and AtPHOT2 while soybean phototropins are designated as 1A (for GmPHOT1A), 1B-1 (for GmPHOT1B-1), 1B-2 (for GmPHOT1B-2), 2A (for GmPHOT2A) and 2B (for GmPHOT2B). All the numbers are in percentage**.**

	AtPHOT1	**1A**	**1B-1**	**1B-2**	**AtPHOT2**	**2A**	**2B**
AtPHOT1	100						
1A	64.99	100					
1B-1	67.2	74	100				
1B-2	66.29	73.79	93.99	100			
AtPHOT2	55.84	55.08	57.26	56.93	100		
2A	52.82	52.81	53.97	53.66	69.28	100	
2B	52.4	51.89	53.25	52.95	6918	87.17	100

Table 15: Percentage identity between the LOV1 domains of Arabidopsis and soybean phototropins. *Arabidopsis* phototropins are designated as AtPHOT1 and AtPHOT2 while soybean phototropins are designated as 1A (for GmPHOT1A), 1B-1 (for GmPHOT1B-1), 1B-2 (for GmPHOT1B-2), 2A (for GmPHOT2A) and 2B (for GmPHOT2B). All the numbers are in percentage**.**

	AtPHOT1	**1A**	**1B-1**	**1B-2**	**AtPHOT2**	**2A**	**2B**
AtPHOT1	100						
1A	83.17	100					
1B-1	82.24	89.71	100				
1B-2	83.17	92.52	97.19	100			
AtPHOT2	71.96	73.83	73.83	74.76	100		
2A	74.76	78.5	76.63	77.57	89.71	100	
2B	73.83	77.57	75.7	76.63	90.65	99.06	100

Table 16: Percentage identity between the LOV2 domains of *Arabidopsis* and soybean phototropins. *Arabidopsis* phototropins are designated as AtPHOT1 and AtPHOT2 while soybean phototropins are designated as 1A (for GmPHOT1A), 1B-1 (for GmPHOT1B-1), 1B-2 (for GmPHOT1B-2), 2A (for GmPHOT2A) and 2B (for GmPHOT2B). All the numbers are in percentage**.**

	AtPHOT1	**1A**	**1B-1**	**1B-2**	**AtPHOT2**	**2A**	**2B**
AtPHOT1	100						
1A	93.45	100					
1B-1	94.39	98.13	100				
1B-2	94.39	99.06	99.06	100			
AtPHOT2	88.78	89.71	90.65	90.65	100		
2A	86.91	87.85	88.78	88.78	96.26	100	
2B	86.91	87.85	88.78	88.78	96.26	100	100

Table 17: Percentage identity between the catalytic domains of *Arabidopsis* and soybean phototropins. *Arabidopsis* phototropins are designated as AtPHOT1 and AtPHOT2 while soybean phototropins are designated as 1A (for GmPHOT1A), 1B-1 (for GmPHOT1B-1), 1B-2 (for GmPHOT1B-2), 2A (for GmPHOT2A) and 2B (for GmPHOT2B). All the numbers are in percentage**.**

	AtPHOT1	**1A**	**1B-1**	**1B-2**	**AtPHOT2**	**2A**	**2B**
AtPHOT1	100						
1A	80.55	100					
1B-1	83.44	86.45	100				
1B-2	82.41	85.41	98.28	100			
AtPHOT2	71.87	71.52	76.04	76.04	100		
2A	69.68	69.68	72.47	72.12	85.01	100	
2B	69.33	69.33	72.47	72.12	84.66	98.6	100

FIGURES

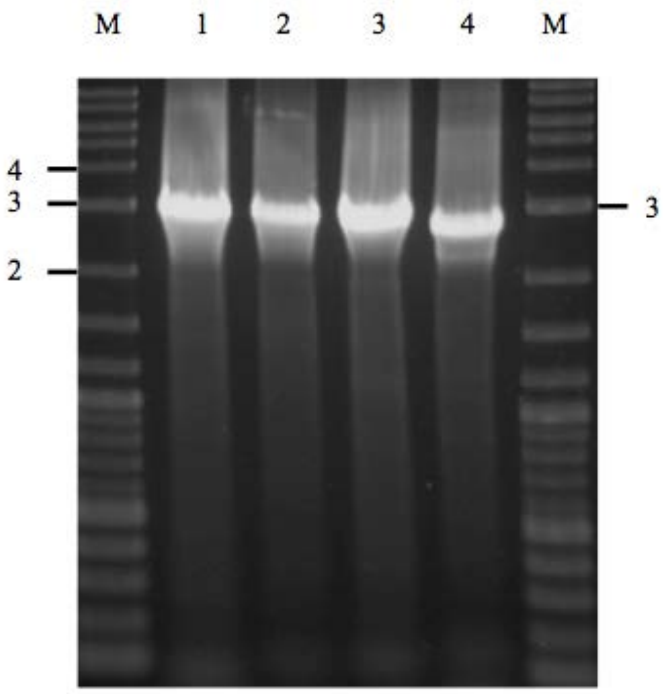

Figure 1. **RT-PCR of *GmPHOT* mRNAs**: cDNAs of GmPHOT1A (lane 1), GmPHOT1B-1 (lane 2), GmPHOT2A (lane 3), GmPHOT1B-2 (lane 4) obtained by RT-PCR and then visualized by agarose gel electrophoresis. Gene Ruler 1 kb plus DNA markers were used in this gel and all the following gels.

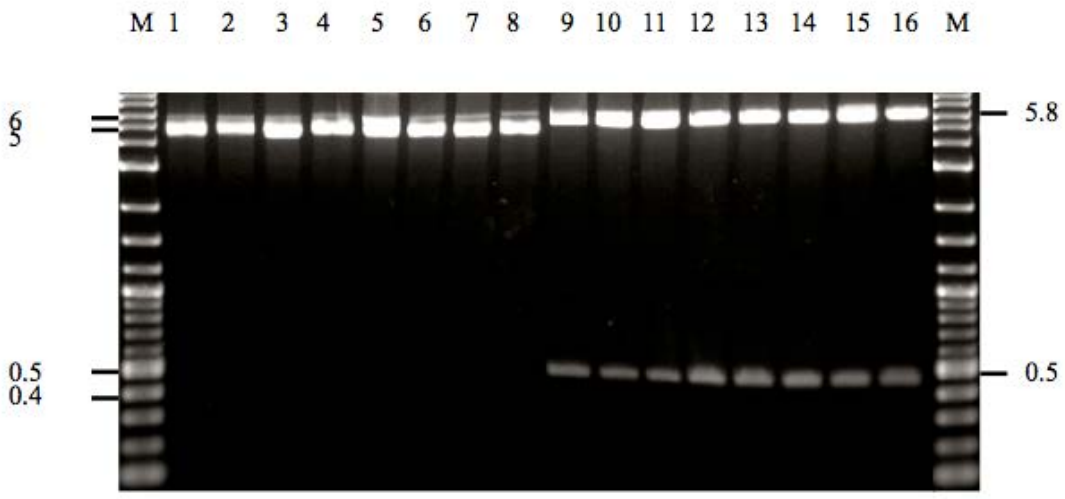

Figure 2. **Restriction analyses of pCR8/TOPO *GmPHOT1A* entry clones for identity.** The clones were digested with *PstI* (lane 1-8) and *HindIII* (lane 9 to 16) restriction enzymes. The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb. Lane 3(=lane 11) and lane 5(=lane 13) clones were chosen for sequencing.

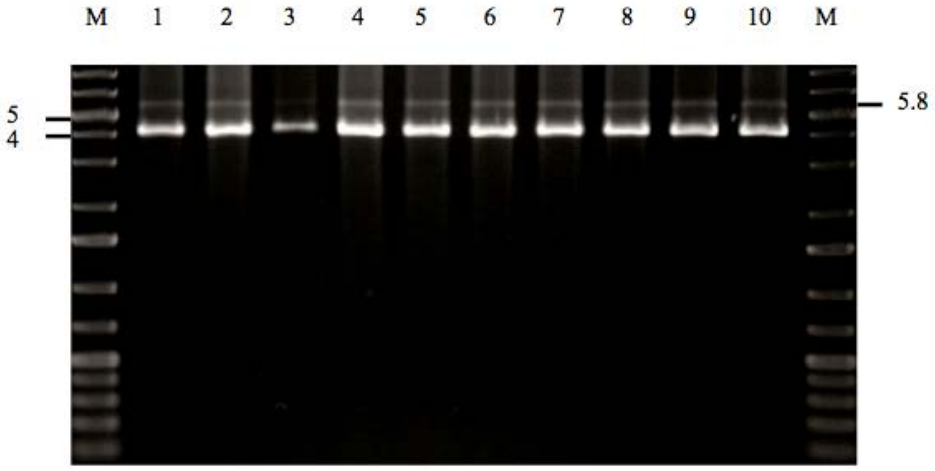

Figure 3. **Restriction analyses of pCR8/TOPO *GmPHOT1B-1* entry clones for identity:** The clones were digested with *HindIII* restriction enzyme. The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb. Lane 2 and lane 5 clones were chosen for sequencing.

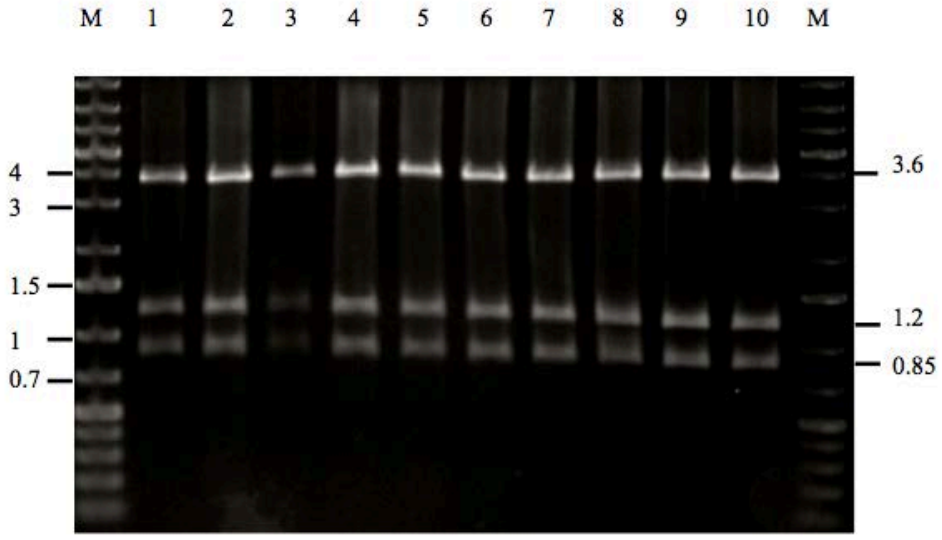

Figure 4. Restriction analyses of pCR8/TOPO *GmPHOT1B-1* entry clones for identity: The clones were digested with *PstI* restriction enzyme. The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb. Lane 2 and lane 5 clones were chosen for sequencing.

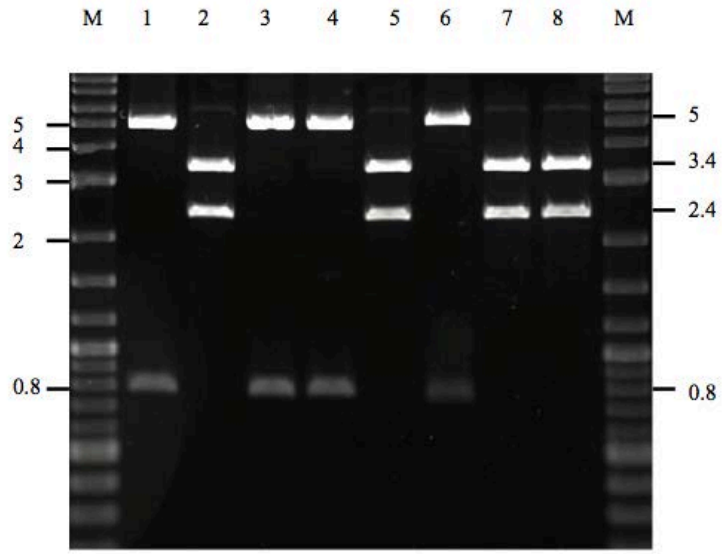

Figure 5. **Restriction analyses of pCR8/TOPO *GmPHOT1A* entry clones for orientation:** The clones were double digested with *EcoRV* and *BamHI* restriction enzymes. The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb. Lane 2 and lane 5 clones were chosen for sequencing.

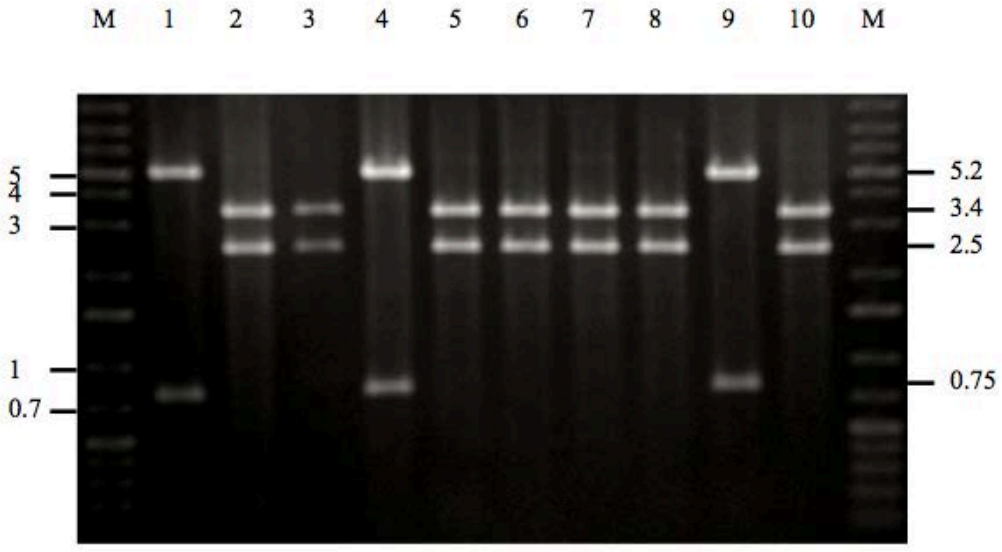

Figure 6. **Restriction analyses of pCR8/TOPO *GmPHOT1B-1* entry clones for orientation:** The clones were double digested with *EcoRV* and *BamHI* restriction enzymes. The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb. Lane 2 and lane 5 clones were chosen for sequencing.

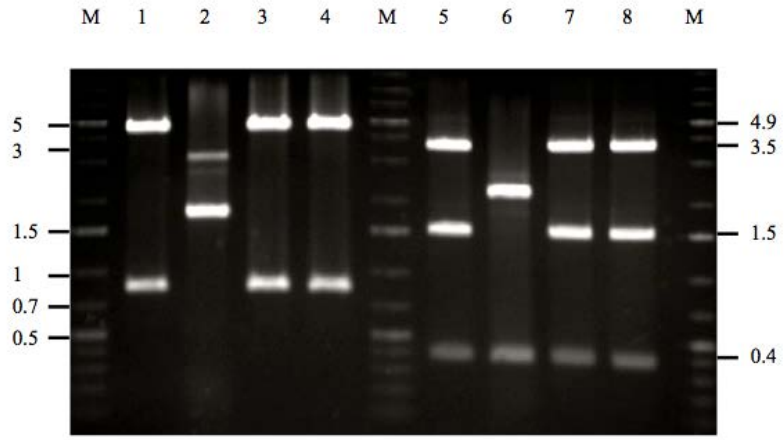

Figure 7. **Restriction analyses of pENTR-D/TOPO *GmPHOT1B-2* entry clones for identity:** The clones were digested with *PstI* (lane 1-4) and *HindIII* (lane 5-8) restriction enzyme. The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb. Lane 1(=lane 5) and lane 3(= lane 7) clones were chosen for sequencing. Lane2 DNA was an empty vector DNA that was selected by mistake.

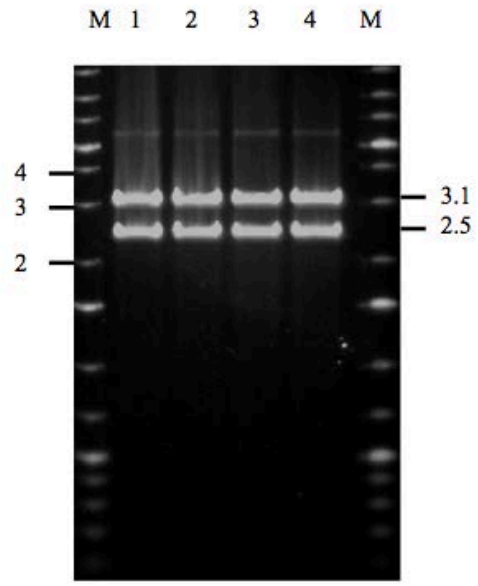

Figure 8. **Restriction analyses of pENTR-D/TOPO *GmPHOT1B-2* entry clones for orientation:** The clones were double digested with *EcoRV* and *BamHI* restriction enzymes. The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb. Lane 1 and lane 3 clones were chosen for sequencing.

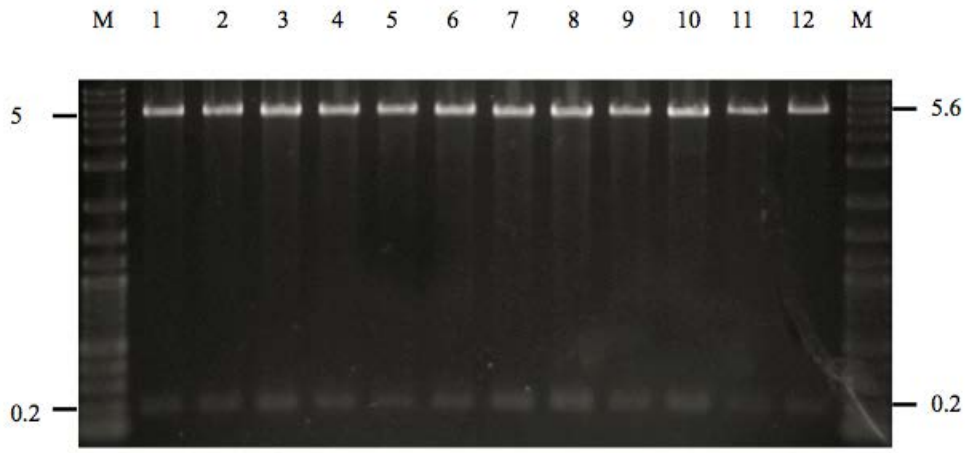

Figure 9. Restriction analyses of pENTR-D/TOPO *GmPHOT2A* entry clones for identity: The clones were digested with *HindIII* restriction enzyme. The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb. Lane 1 and lane 2 clones were chosen for sequencing.

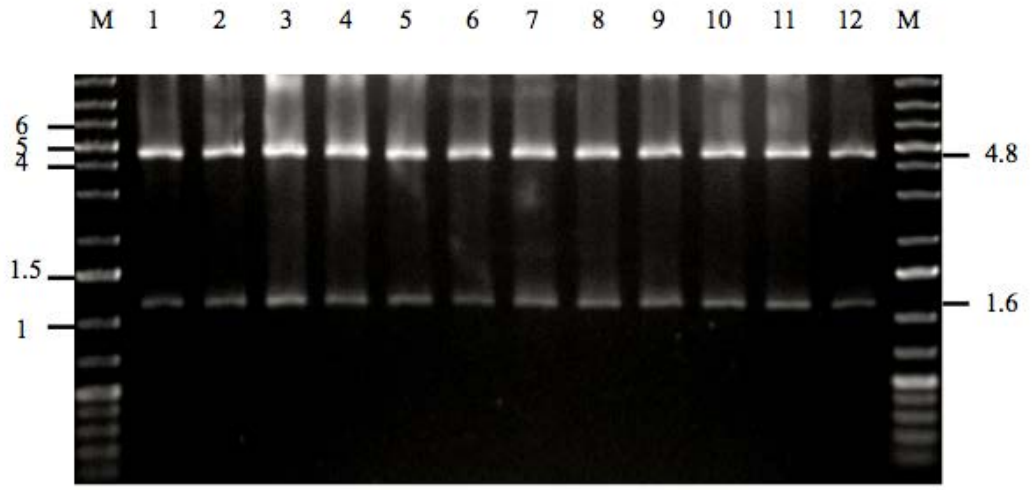

Figure 10. **Restriction analyses of pENTR-D/TOPO *GmPHOT2A* entry clones for identity:** The clones were digested with *PstI* restriction enzyme. The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb. Lane 1 and lane 2 clones were chosen for sequencing.

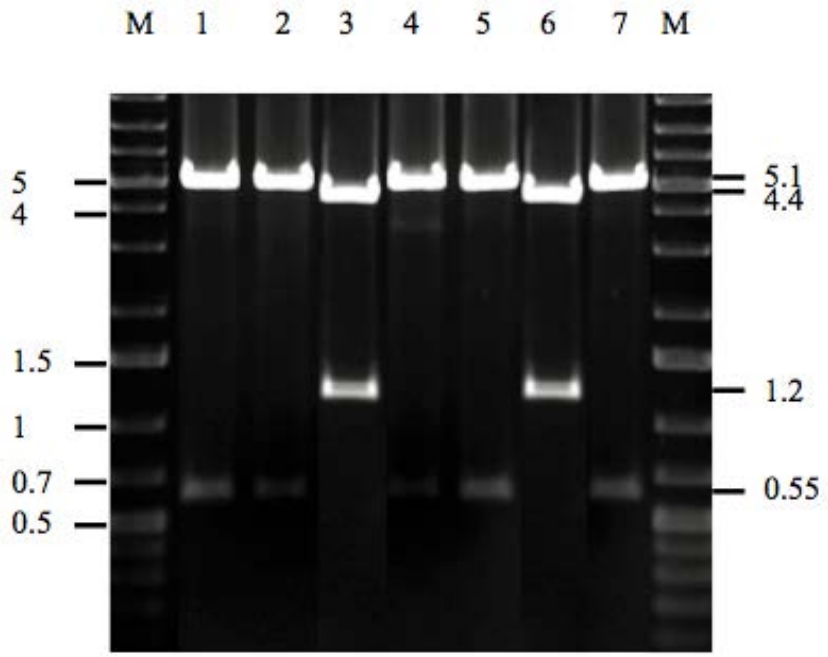

Figure 11. **Restriction analyses of pENTR-D/TOPO *GmPHOT2A* and *GmPHOT2B* entry clones for identity:** The clones were digested with *PstI* restriction enzyme. The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb. Lane 1, 2, 4, 5, 7 were GmPHOT2B clones while only lane 3 and 6 were GmPHOT2A clones. lane1 and lane 4 GmPHOT2B clones were chosen for sequencing.

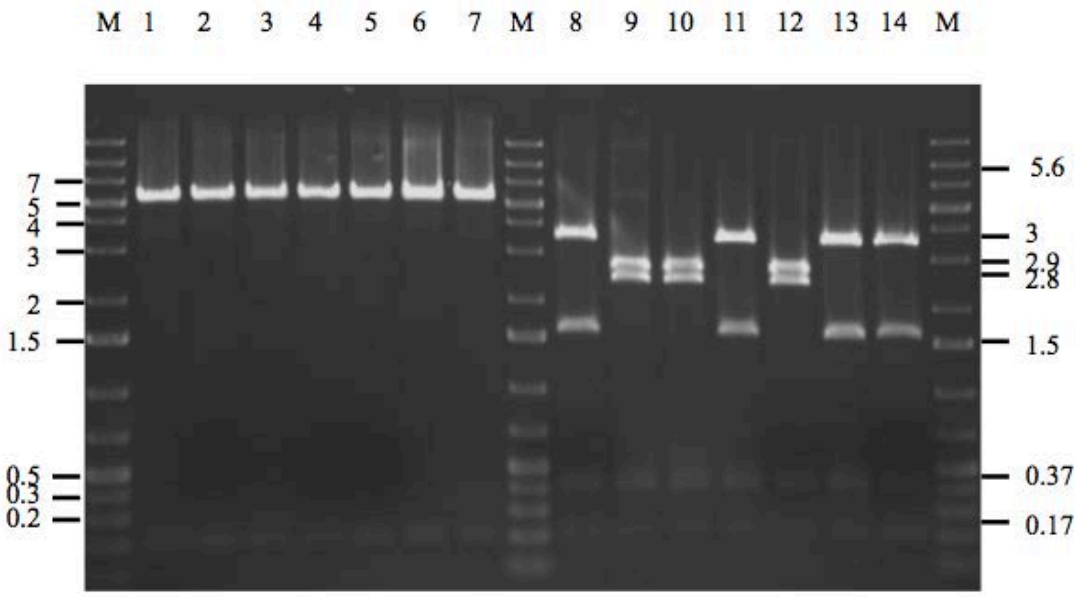

Figure 12.Restriction analyses of pENTR-D/OPO *GmPHOT2A* and *GmPHOT2B* entry clones for identity and orientation: The clones were digested with *HindIII* restriction enzyme for identity (as depicted by lane1 to lane7) and the same clones in the same order were digested with *HindIII* and *HaeII* for orientation analysis (as depicted by lane8 to lane 14). The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb. Lane 1, 2, 4, 5, 7 were GmPHOT2B clones while lane 3 and 6 were GmPHOT2A clones. Lane1 (=lane9) and lane 2(=lane12) clones were chosen for sequencing.

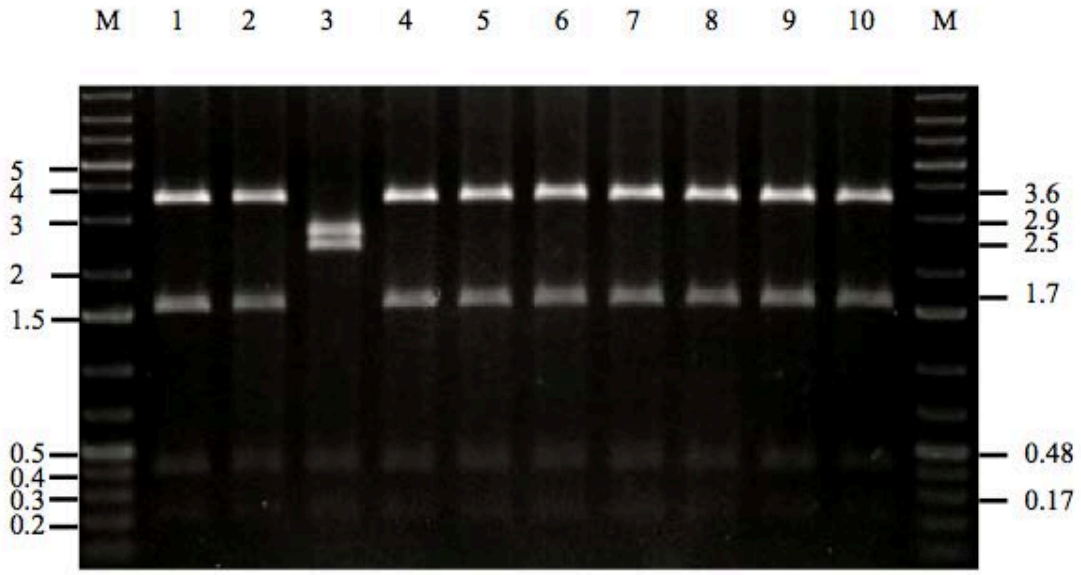

Figure 13.**Restriction analyses of pENTR-D/TOPO *GmPHOT2A*entry clones for orientation:** The clones were digested with *HindIII* and *HaeII* restriction enzymes. The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb. Lane 1 and lane 2 clones were chosen for sequencing.

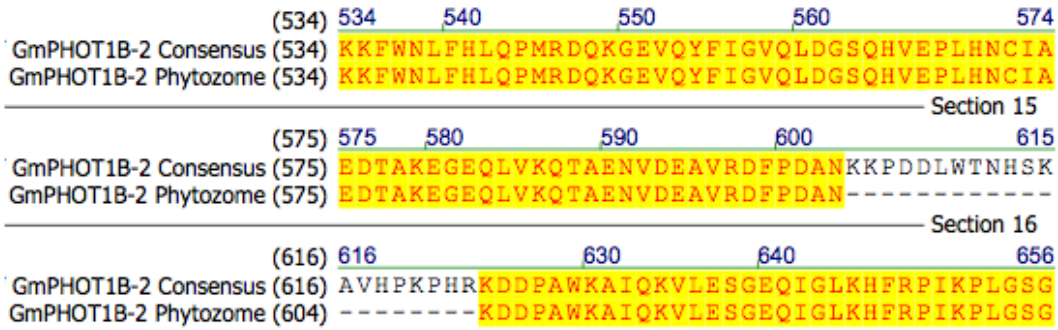

Figure 14. **Comparison of GmPHOT1B-2 consensus and GmPHOT1B-2 Phytozome polypeptide**: Amino acid sequence alignment between the GmPHOT1B-2 sequence obtained from the Phytozome and the consensus sequence of GmPHOT1B-1 clones obtained with the help of Vector NTI software.

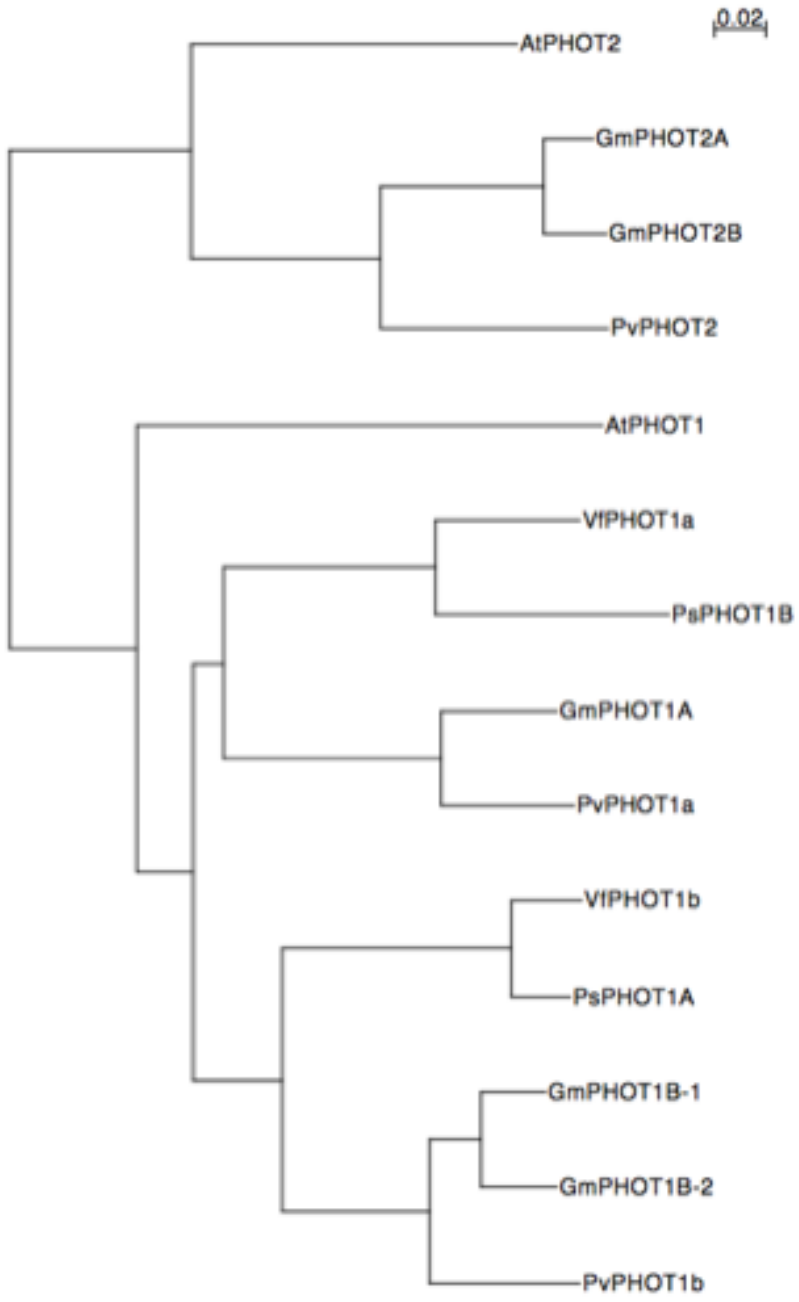

Figure 15. Phylogeny of legume and *Arabidopsis* phototropins. A multiple alignment of full-length phototropin protein sequences from the garden pea (PsPHOT), common bean (PvPHOT), broad bean (VfPHOT), *Arabidopsis* (AtPHOT and the protein coding regions of all the 5 *GmPHOT* entry clones was created with ClustalX 1.82. The Tree was generated with 10,000 bootstrap trials and visualized with Njplot


AtPHOT1
GmPHOT1A
GmPHOT1B-1
GmPHOT1B-2
AtPHOT2
GmPHOT2A
GmPHOT2B
consensus

Figure 16. **Alignment between the polypeptide sequences of *Arabidopsis* and soybean phototropins**. Pairwise and multiple alignments were done using ClustalX 2.1 software and then the aligned sequences were imported into the BoxShade server software. The black boxes indicate the region of similarity. LOV1 and LOV2 domain locations are indicated by blue lines and that of the catalytic domains by a red line.

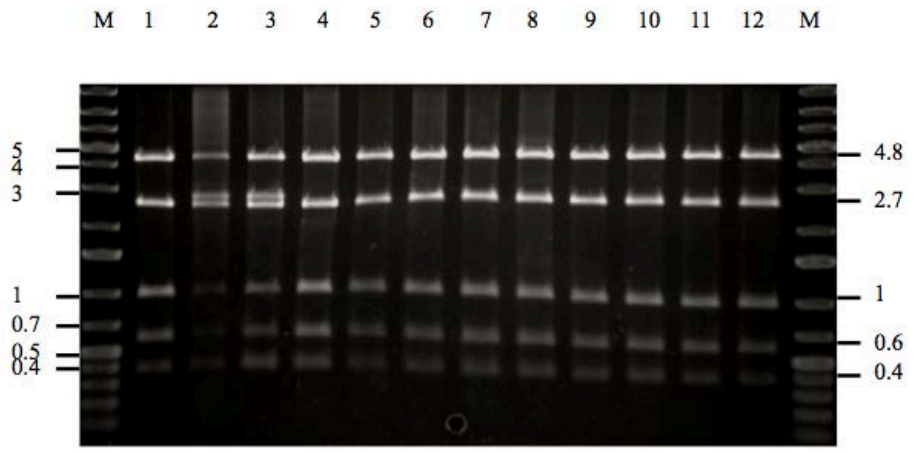

Figure 17. **Restriction analyses of the *GmPHOT1A* expression clones for identity**: The expression clones obtained by LR cloning were digested with restriction enzyme *EcoRI*. The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb. Lane 6 and lane 7 were chosen for sequencing.

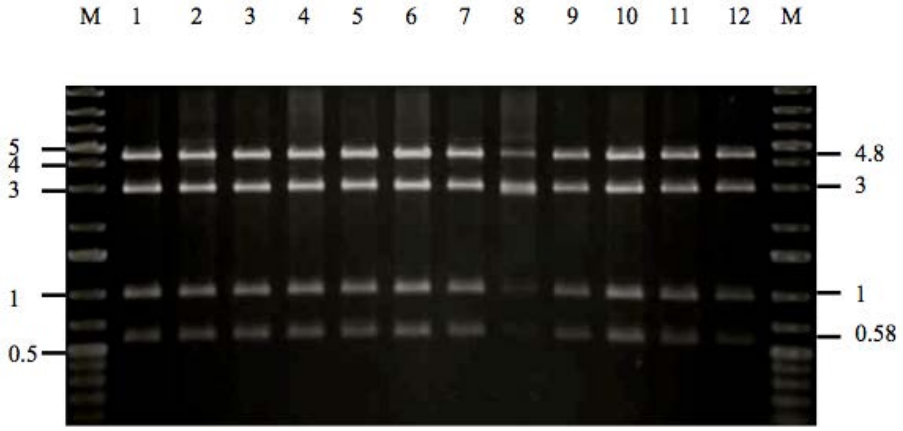

Figure 18. **Restriction analyses of *GmPHOT1B-1* expression clones for identity:** The clones were digested with *EcoRI* restriction enzyme. The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb. Lane 1 and lane 2 clones were chosen for sequencing.

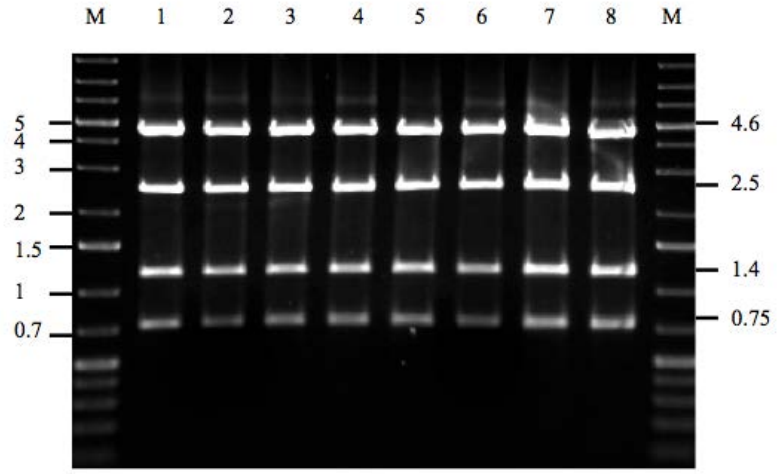

Figure 19. **Restriction analyses of *GmPHOT2A* expression clones for identity:** The clones were digested with *EcoRI* restriction enzyme. The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb. Lane 1 and lane 2 clones were chosen for sequencing.

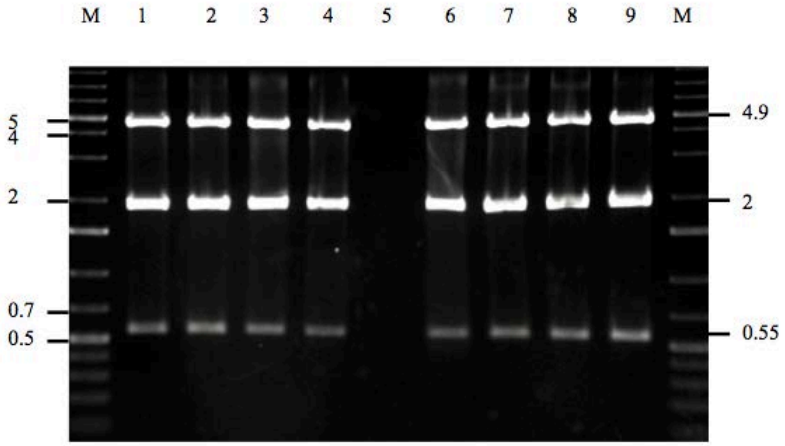

Figure 20. Restriction analyses of *GmPHOT2B* expression clones for identity: The clones were digested with *EcoRI* restriction enzyme. The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb. Lane 1 and lane 2 clones were chosen for sequencing.

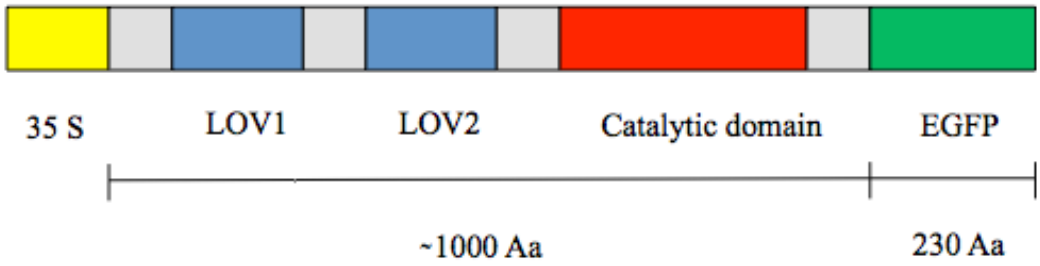

Figure 21. **Plant expression construct (35S::*GmPHOT:EGFP*):** The *GmPHOT* protein coding region including the LOV domains and the catalytic domains are shown. The EGFP is fused to its carboxy-terminus. *GmPHOT*s encode protein of length~1000 Amino acids while EGFP protein is 230 amino acids long. 35S promoter drives the expression of *GmPHOT*.

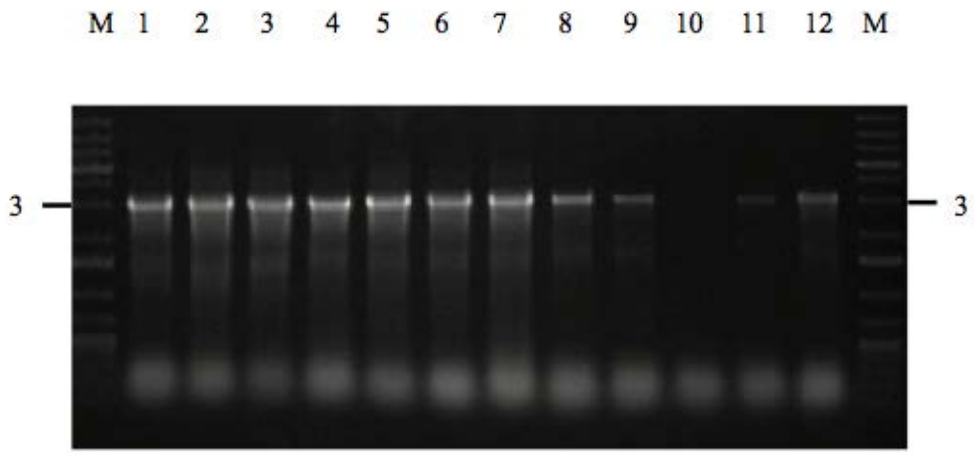

Figure 22. **Colony PCR of *A. tumefaciens* colonies harboring expression clones of *GmPHOT1A***: The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb.

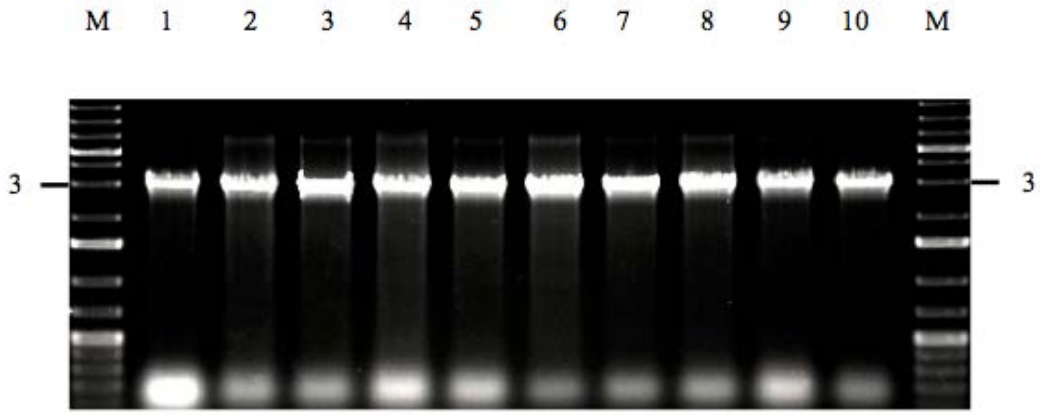

Figure 23. **Colony PCR of *A. tumefaciens* colonies harboring expression clones of *GmPHOT1B-1***: The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb.

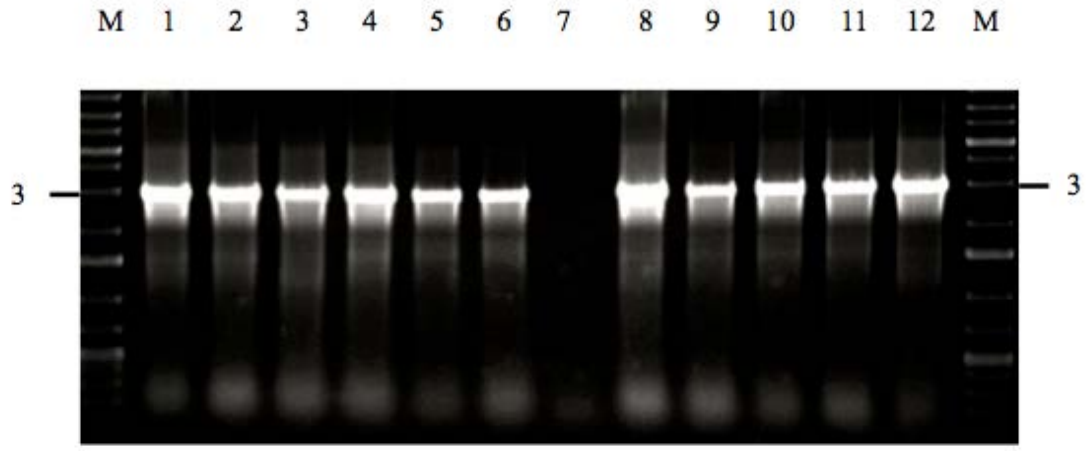

Figure 24. **Colony PCR of *A. tumefaciens* colonies harboring expression clones of GmPHOT2A**: The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb.

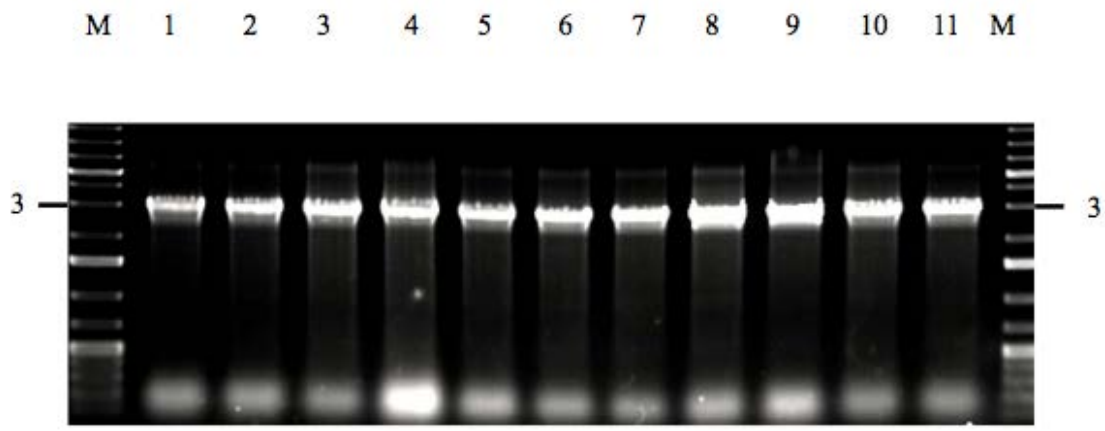

Figure 25. **Colony PCR of *A. tumefaciens* colonies harboring expression clones of *GmPHOT2B***: The DNA fragments were resolved on 1.2% agarose gel. All the DNA fragments are in Kb.