Functional characterization of *Arabidopsis* orthologs of *PAT1* gene in *Physcomitrella patens (P. patens)*

A thesis submitted towards partial fulfilment of

BS-MS Dual Degree Programme

Thesis submitted by-

Ms. Sukanya Vasant Jogdand

20121038

Project Supervisor

Dr. Anjan K. Banerjee

Associate professor,

Department of Biology,

IISER Pune

Project Advisor

Dr. Aurnab Ghose

Associate professor,

Department of Biology,

IISER Pune

IISER PUNE

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान पुणे

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

(An Autonomous Institution of Ministry of Human Resource Development, Govt. of India)

Dr. Homi Bhabha Road, Pune - 411 008.

Certificate

This is to certify that this dissertation entitled "**Functional characterization of *Arabidopsis* orthologs of *PAT1* gene in *Physcomitrella patens* (*P. patens*)**" towards the partial fulfilment of the BS-MS degree programme at the **Indian Institute of Science Education and Research, Pune** represents work carried out by **Ms. Sukanya Vasant Jogdand** at IISER, Pune under the supervision of **Dr. Anjan K. Banerjee**, Associate professor, Department of Biology, IISER Pune during the academic year 2016-2017.

Date: 30/03/2017

30/03/2017

Sukanya Vasant Jogdand

20121038

5th yr BS-MS Student,

IISER Pune

30/3/17

Dr. Anjan K. Banerjee

Associate professor,

Department of Biology,

IISER Pune

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान पुणे
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE
(An Autonomous Institution of Ministry of Human Resource Development, Govt. of India)
Dr. Homi Bhabha Road, Pune - 411 008.

Declaration

I hereby declare that the matter embodied in the report entitled "**Functional characterization of *Arabidopsis* orthologs of *PAT1* gene in *Physcomitrella patens* (*P. patens*)**" are the results of the work carried out by me at Department of Biology, Indian Institute of Science Education and Research, Pune under the supervision of Dr. Anjan K. Banerjee and the same has not been submitted elsewhere for any other degree.

Date: 30/03/2017

30/03/2017

Sukanya Vasant Jogdand
20121038
5th yr BS-MS Student,
IISER Pune

30/3/17

Dr. Anjan K. Banerjee
Associate professor,
Department of Biology,
IISER Pune

TABLE OF CONTENTS

TITLE **PAGE NO.**

LIST OF FIGURES

FIGURE **PAGE NO.**

Chapter 1 Introduction

Chapter 3 Materials and methods

Chapter 4 Results and discussions

LIST OF TABLES

TABLE **PAGE NO.**

Chapter 3 Materials and methods

Chapter 4 Results and discussions

Acknowledgements

I would like to offer my sincere thanks to Dr. Anjan K. Banerjee, Associate professor at the IISER, Pune for providing me with the opportunity to work on this project. His invaluable guidance at each and every step of the project helped me develop a scientific outlook and his high energy levels kept me enthused throughout the term of the project and thereafter. Thanks for being a great mentor.

I would also like to express my gratitude to Dr. Aurnab Ghose, Associate professor at IISER Pune for his valuable, inspiring and timely advice he provided during the course of the project.

I would thank Dr. Tomoaki Nishiyama and Dr. Mitsuyasu Hasebe for providing us with the vectors pTN182 and pTN186 used for knockout cloning purpose.

Also, I thank my Moss team members – Amey, Vyankatesh and M. Boominathan for their continuous help in teaching, assisting and encouraging me all throughout the project.

I am grateful to Ravi, Harpreet, Amit, Kirtikumar, Bhavani, Nilam and all the present and ex-lab members of PMB lab for their constant support and friendly atmosphere in lab. Thanks for the countless memories we made together.

I would like to thank all my friends at IISER Pune, who have provided me with fun and joy throughout my time in IISER.

Special thanks are owed to my parents, Pradnesh, Ashlesha, and the rest of my family, who have been my true support throughout my years of education, both morally and financially.

I thank IISER Pune for providing me this wonderful opportunity to grow and develop both scientifically and personally.

Abbreviations

ABA	Abscisic acid
ACT	actin 5
At	*Arabidopsis thaliana*
cDNA	complementary DNA
CDS	coding sequence
CIGR	chitin-inducible gibberellin-responsive
Cms	*Citrus medica var. sarcodactylis*
DELLA	N-terminal Della domain
E2	ubiquitin-conjugating enzyme *E2*
FR	Flanking region
GA	Gibberellic acid
gDNA	Genomic DNA
GRAS	GAI (gibberellic acid insensitive), RGA (repressor of GAI), and SCR (scarecrow)
HAM	Hairy meristem
IDR	Intrinsically Disordered Region
LAS/Ls	Lateral suppressor
ORF	Open Reading Frame
Os	*Oryza sativa*
PAT1	Phytochrome A Signal Transduction 1
PAL1	Phytochrome A Signal Transduction1-like
PEG	Polyethylene Glycol

PHYA	Phytochrome A
Pp	*Physcomitrella patens*
SCR	Scarecrow
SCL	Scarecrow-like
SHR	Short-root
TF	Transcription factor
Va	*Vitis amurensis*
Vv	*Vitis vinifera*
WT	wildtype

Abstract

Plant-specific GRAS proteins are the putative transcription factors that play critical and diverse roles in plant growth and development. GRAS family comprises of seven clades in moss with GRAS domain containing proteins- SCR, SHR, DELLA, LAS/Ls, HAM, SCL and PAT1. Previous studies have showed that, *AtPAT1* plays a role in phytochrome signal transduction in *A. thaliana*, while *VaPAT1* is involved in abiotic stress responses in *V. amurensis*. However, the role of PAT1 have been unexplored in the lower group of plants *viz* bryophytes. We are interested in understanding the functional role of PAT1 proteins in *P. patens*. A comprehensive analysis of *PpPAL1A* and *PpPAL1B* genes, including gene structure, motif search, sequence alignment, phylogeny, tissue specific expression profile and targeted mutagenesis were performed. The sequence alignment, motif search and phylogenetic tree analyses together confirmed that PpPAL1A and PpPAL1B are the members of the PAT1-branch of the GRAS domain protein family and they showed 82.95% identity at the level of amino acids. Based on the available *in silico* microarray data, we studied the expression levels of both *PpPAL1A* and *PpPAL1B* genes in moss. Interestingly, we observed that *PpPAL1A* showed constitutive expression in all tissue while *PpPAL1B* expression was dominated in the sporophytic phase. Our tissue specific gene expression analysis done by semi-quantitative RT-PCR supported the available *in silico* data, where both the genes showed relatively higher expression level in gametophore compared to protonema. Statistical significance was seen in relative expression levels, in protonema and gametophore tissues for both *PpPAL1A* and *PpPAL1B*. The full length coding sequences for *PpPAL1A* and *PpPAL1B* were successfully cloned and transformed in moss by *Agrobacterium-mediated* transformation. Putative transgenic lines are growing in BCDAT media and will be further characterized. Knockout construction of *PpΔpal1a* and *PpΔpal1b* vectors and generation of mutant lines by PEG-mediated transformation is currently under progress. These approaches would help us to understand the functional roles of PAT1-like proteins in the transcriptional regulation and modulatory functions of GRAS domain proteins in moss.

Chapter 1

Introduction

1.1 *Physcomitrella patens* – the model system

The genus *Physcomitrella* belongs to the family *Funariaceae* in the major group *Bryophytes* (mosses and liverworts). *P. patens* have relatively simple morphology, is rapidly propagated and easily transformed compared with most vascular plants. The important features that make it a desirable model system are its small size, non-flowering plant that reproduce using spores, short life cycle and well-annotated genome sequence (~500 MB). The main developmental feature of this plant is its biphasic alternation between a dominant haploid gametophytic and a minor diploid sporophytic phase. The dominant haploid gametophytic phase facilitates generation of gene knock out lines along with rapid identification of dominant and recessive mutations that affect moss development.

1.2 Eight types of stem cells in moss

The eight different types of stem cells are reported in *P. patens* that determine moss development (Kofuji and Hasebe, 2014). Moss germinates from a haploid spore to produce a two-dimensional filamentous network called as protonema. Further, each filament grows by polarized growth where stem cells are present at the apex of each cell. Phytohormones play an important role in subsequent tissue differentiation and development of moss (Russo et al., 1992). The transition from chloronema (that contains ~100 fully developed chloroplast) to caulonema (that contains fewer less-developed plastids) is based in an auxin-dependent manner. Cell plates help in differentiation of chloronema and caulonema, where cell division is observed to be transverse in chloronema and oblique angle in caulonema (Harrison et al., 2009). The side branch initials decide the fate of bud formation that produces dimorphic gametophore, shoot of moss. The sexual organs male antheridia and female archegonia are formed at the top of a single gametophore that fertilises in an appropriate moist conditions.

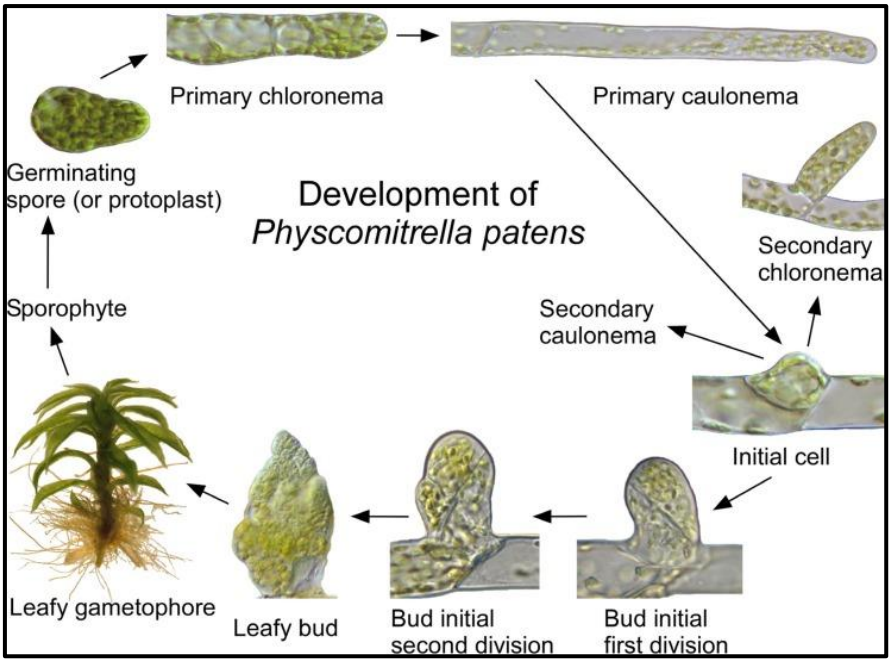


Figure 1.1 Development and life cycle of *P. patens* (moss) (adapted from (Roberts et al., 2012)).

1.3 GRAS family proteins and their domain structure

The transcriptional regulation of cell division and tissue differentiation is very important in the growth and development of multicellular organisms. One of the important plant-specific protein families are GRAS proteins. The acronym GRAS is based on the locus designations of the three members: GIBBERELLIN-ACID INSENSITIVE (GAI), REPRESSOR of GA1 (RGA) and SCARECROW (SCR) that are usually composed of 400-770 amino acid residues and they share sequence homology at their C-terminal parts (Bolle, 2004).

Sequence analysis of the products of the GRAS (*GAI, RGA* and *SCR*) gene family indicates that they possess a variable N-terminal domain (N-domain) and a widely and highly conserved C-terminal domain (GRAS domain) that contains five different recognizable sequence motifs in the following order: leucine heptad repeat I (LHR I, LRI), VHIID motif, leucine heptad repeat II (LHR II, LRII), PFYRE motif and SAW motif (shown in Figure 1.2). The VHIID motif is identified with residues like valine, leucine, isoleucine, proline, asparagine, histidine, aspartic acid and glutamine. The LRI motif contains putative nuclear localization signals (NLSs) while LRII motif contains an LXXLL (Leu-Xaa-Xaa-Leu-Leu; Xaa denotes any amino acid) and they are flanked to the VHIID motif on either side. This LRI-VHIID-LRII pattern

has been reported to involve in the interaction of GRAS proteins and their nucleic acid or protein partners. The PFYRE motif consists proline (P), aromatic phenylalanine and tyrosine (FY), and arginine/ glutamic acid (RE) residues that are conserved in all GRAS proteins. The SAW motif contains three sequential units: WX7G (X for any seven amino acids), L-W and SAW. Currently, the roles of PFYRE and SAW motifs are not known, but the residues of these motifs are highly conserved and might be essential for structural integrity and function of GRAS proteins (Sun et al., 2012).

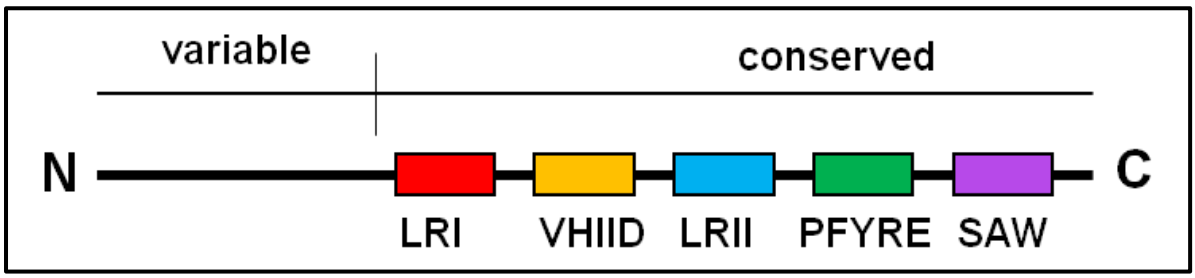


Figure 1.2 Schematic representation of domain structure of GRAS domain proteins. (Boxes indicates conserved domains and amino acid residues are shown below)

1.4 GRAS subfamilies

The GRAS family members are the plant putative transcription factors that play multiple roles in plant development. Depending on the common features that the GRAS family shares, they can be categorized into the following clades: 1. DELLA (N-terminal Della domain) proteins, 2. Scarecrow (SCR) branch, 3. Lateral Suppressor (LAS/Ls) branch, 4. Hairy Meristem (HAM) branch, 5. Phytochrome A Signal Transduction 1 (PAT1) branch, 6. Short root (SHR) branch and 7. SCR-like (SCL) branch (Bolle, 2004).

The GRAS proteins have been shown to function in a diverse set of plant physiological and developmental processes like the initiation of shoot meristem, radial patterning, maintenance of organ indeterminacy, gibberellin signalling and nodulation, abiotic stress responses and phytochrome signal transduction (Engstrom, 2011). This has raised several questions in the field of molecular, developmental and evolutionary aspects regarding the specific classes of GRAS

proteins as to how and when did they emerged, expanded and acquired novel functions.

A generalized model for the domain structure of GRAS proteins is shown in Figure 1.3. It broadly classifies the GRAS proteins into a variable N-terminal region consisting of the intrinsically disordered region (IDRs) and a highly conserved and structurally folded C-terminal region.

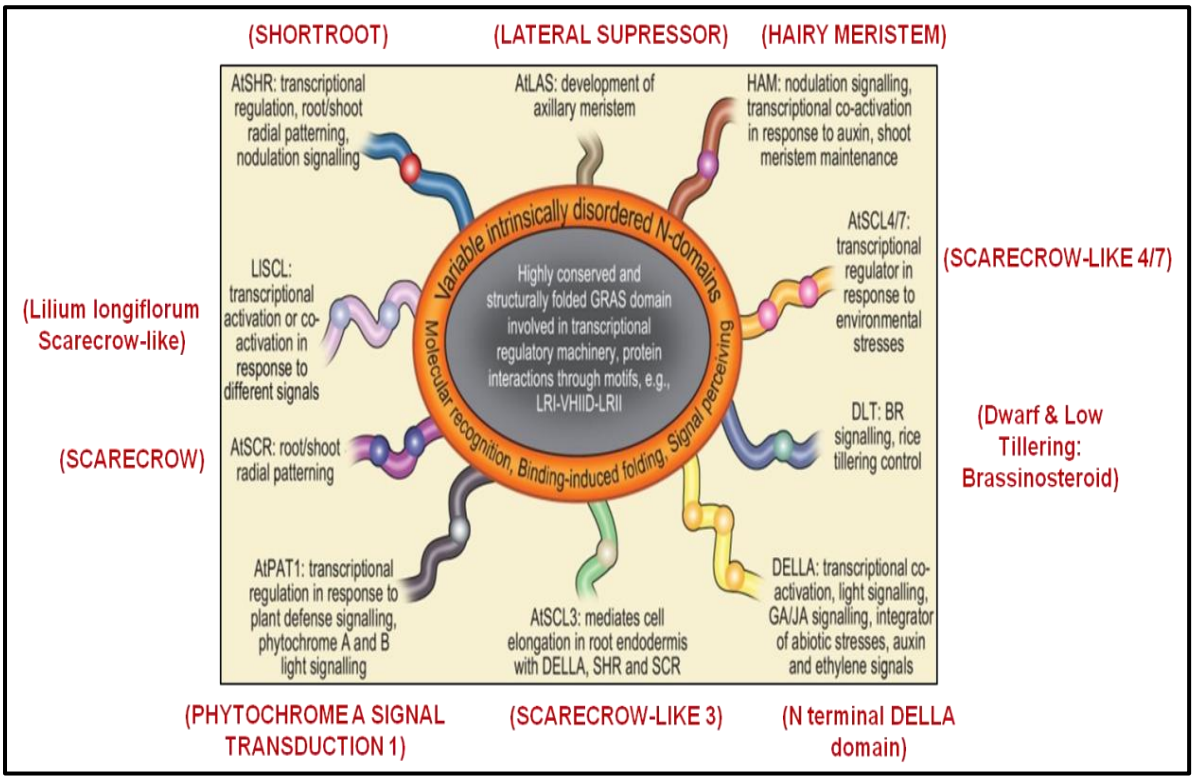


Figure 1.3 Representative model of a functional polymorphism of GRAS domain proteins determined by their IDR (adapted from (Sun et al., 2012)). The coiled trails denote the intrinsically disordered nature of the N-domains of each GRAS subfamily in molecular recognition, binding-induced folding and signal perceiving. The grey part denotes the conserved GRAS domains found in GRAS family, involved in transcriptional regulatory machinery and protein interaction through motifs. The beads on coils denote the potential binding sites present in the disordered N-domains that might be involved in specific signally pathways.

1.5 PAT1 branch proteins:

In higher plants, members of PAT1 and DELLA protein (GAI, RGA and RGL1-3) clades act as negative regulators of the gibberellins signal transduction (Peng et al., 1997). It has been shown that PAT1 is involved in light signalling mechanism that controls basic plant developmental processes, including de-etiolation and hypocotyl elongation, via phytochrome A (phyA) photoreceptor in *A. thaliana* (Bolle et al., 2000). The *pat1-1* mutant was shown to be deficient in most phyA-regulated processes and suggested to act in an early step of phyA signal transduction. The PAT1 protein was found to be localized to the cytoplasm in *Arabidopsis*. As a member of the PAT1 branch, the *Arabidopsis* Scarecrow-like 13 (*At*SCL13) is shown to positively regulate red light signalling downstream of phytochrome B (phyB) (Torres-Galea et al., 2006).

Along with role in signal transduction in *A. thaliana*, recently, it was shown that *VaPAT1* act as a positive regulator involved in grapevine abiotic stress responses (Yuan et al., 2016). Overexpression of *VaPAT1* was shown to enhance cold, drought and high salinity tolerance in transgenic *Arabidopsis* via modulation of the expression of a series of stress-related genes. Also, the transgenic lines were observed with higher levels of the stress-related genes under normal growth conditions.

The GRAS family transcription factors were first found to be emerged in bacteria and included in the Rossmann-fold methyltransferase superfamily. It was deduced that the GRAS genes might have derived from the bacterial genome in plants, but their phylogenetic data revealed that the typical plant GRAS genes first appeared in moss, with 40 members seen in *P. patens* (Wang et al., 2016). According to Plant Transcription Factor Database (v4.0), 129 GRAS transcription factors are known to be present in *P. patens.*

Chapter 2

Objectives

GRAS being an important transcription factor in plant growth and development, most of the GRAS domain proteins are not well studied in *P. patens*. This project emphasises to understand the role of PAT1-like proteins in *P. patens* and functionally characterise this protein. To elucidate the role of PAT1-like GRAS domain containing proteins that are involved in light signalling or plant defence response, we have chosen a reverse genetics approach. This project aims at investigating the role of *A. thaliana* orthologs of *PAT1* gene in moss-

1. Identification of *AtPAT1* gene orthologs in *P. patens* using bioinformatics tools
2. Tissue-specific gene expression profile of *AtPAT1* gene orthologs in *P. patens*
3. Construction of vectors for overexpression of *PpPAL1A* and *PpPAL1B* genes and generation of transgenic mutant lines by targeted mutagenesis
4. Construction of vectors to generate knockout lines of *PpPAL1A* and *PpPAL1B* genes by targeted mutagenesis

Approach

1. Identification of *AtPAT1* gene orthologs in *P. patens* using bioinformatics tools

    a. Identification of the GRAS domain containing protein sequences and isolation of *AtPAT1*-like gene orthologs in *P. patens*
    b. Sequence alignment and phylogenetic analysis of *PpPAL1A* and *PpPAL1B* with other GRAS domain containing proteins

2. Tissue-specific gene expression profile of *AtPAT1* gene orthologs in *P. patens*

    a. *In silico* analysis of tissue specific gene expression of *PpPAL1A* and *PpPAL1B* in *P. patens*
    b. Detection and amplification of *PpPAL1A* and *PpPAL1B* in *P. patens* using both gDNA and cDNA

c. Semi-quantitative RT-PCR validation of tissue-specific gene expression profile of *PpPAL1A* and *PpPAL1B* in protonema and gametophore tissue types of *P. patens*

3. Construction of vectors for overexpression of *PpPAL1A* and *PpPAL1B* genes and generation of transgenic mutant lines by targeted mutagenesis

    a. Cloning of full-length *PpPAL1A* and *PpPAL1B* genes in pCAMBIA1300 and pBI121 binary vectors respectively under 35S promoter
    b. To generate *Agrobacterium-mediated* transgenic lines by over expressing *PpPAL1A* and *PpPAL1B* genes in moss

4. Construction of vectors to generate knockout lines of *PpPAL1A* and *PpPAL1B* genes by targeted mutagenesis

    a. Construction of *PpΔpal1a* and *PpΔpal1b* knockout vectors
    b. Generation of knockout lines of pTN182::*PpΔpal1a* and pTN186:: *PpΔpal1b* vectors through PEG-mediated transformation in moss

Chapter 3

Materials and methods

3.1 Plant material and growth conditions

Physcomitrella patens (Hedw.) subsp. *patens, Gransden* wild type, were cultured in BCDAT agar (0.8%) medium in a 9-cm Petri dish sealed with surgical tape at 25ºC under a 16/8-hr (light/dark) photoperiod with continuous white light for 7days. For homogeneous protonemal culture, 7-days old culture is homogenised with Polytron homogenizer in sterile 4-5ml liquid BCDAT medium in a scintillation vial, and ~2ml of this suspension was spread on solid BCDAT medium laid with sterile cellophane disc.

Figure 3.1 Routine subculture and maintenance of moss tissue by homogenization and protonemal inoculation. A) Polytron Homogenizer; B) Blade and shaft; C) Homogenization of tissue in scintillation vial with Polytron homogenizer; D) Spreading homogenized tissue on BCDAT plate; E) Homogenized tissue grown for 5 days; F) Maintenance of moss cultures in a 9 cm Petri plate; G) 2-weeks old moss colony (Image courtesy A-F: Vyankatesh).

3.2 BCDAT stock solution

Using the standardised and shared protocols from PHYSCObase (www.moss.nibb.ac.jp), moss cultures were grown in regular BCDAT media stock composition mentioned below (Table 3.1). After autoclaving, all the stock media were stored at 4ºC.

Stock B (x100)	$MgSO_4$ $7H_2O$	100 mM
Stock C (x100)	KH_2PO_4 Adjust to pH6.5 with 4M KOH	184 mM
Stock D (x100)	KNO_3 $FeSO_4.7H_2O$	1 M 4.5 mM
Stock AT (x100)	Ammonium Tartrate	500 mM
TES (x1000)	$CuSO_4$ $5H_2O$ H_3BO_3 $CoCl_2$ $6H_2O$ Na_2MoO_4 $2H_2O$ $ZnSO_4$ $7H_2O$ $MnCl_2$ $4H_2O$ KI	0.22 mM 10 mM 0.23 mM 0.1 mM 0.19 mM 02 Mm 0.17 mM
$CaCl_2$ (x50)	$CaCl_2$.2H2O	50 mM

Table 3.1 Stock media preparation.

3.3 RNA isolation and gene expression analysis

a. RNA extraction and cDNA synthesis

Moss tissue was harvested in liquid nitrogen and stored in -80ºC until used for further analysis. Total RNA was extracted using TRIzol™ reagent (Invitrogen,

Catalogue # 15596026) according to the manufacturer's instructions and further subjected to DNase treatment. Two micrograms (2 µg) of DNase-treated RNA from different tissue types were transcribed into cDNA using Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT) enzyme (Invitrogen, Catalogue #28025013) by following the user's protocol. The cDNA synthesised was used to detect *PpPAL1A* and *PpPAL1B* gene transcripts to study their expression profile and to amplify full-length genes to be cloned in overexpression constructs.

b. Semi-quantitative RT-PCR analysis

The semiquantitative RT-PCR was carried out using cDNA synthesised from protonema and gametophore tissue. PCR reaction was set as follows using Taq DNA polymerase (Himedia, Catalogue #MBT060B).

PCR mix content	**Volume (µl)**
10X HiBuffer A	2.5
50mM MgCl2	0.75
Forward primer (10 µM)	0.5
Reverse primer (10 µM)	0.5
cDNA template	0.2
Taq DNA Polymerase	0.5
MilliQ water	18.7
Final volume	25

PCR condition	**Time**	**#cycle**
95 ºC	3 min	1
95 ºC	30 sec	36
*56ºC	30 sec	
74 ºC	45 sec	
74ºC	2 min	1
4 ºC	Hold	-

Table 3.2 PCR reaction mix and PCR thermal conditions. *Annealing temperature was calculated according to the specific primer set while extension time was set based on the amplicon size (1kb/min).

Reactions were carried out in BIO-RAD S1000^{TM} Thermal Cycler PCR machine. Each set of reactions were always included with necessary controls and determination of specific parameters along with a no template control. For gene

expression profile study, it is important to choose the appropriate number of cycles so that the amplification product is clearly seen on an agarose gel and can be quantified. The conditions were chosen such that none of the transcripts analysed should reach a plateau at the end of the amplification protocol. Similarly, the optimal number of cycles was chosen in the same range for a specific transcript of interest (i.e. *PpPAL1A* and *PpPAL1B*) and the reference genes (*ACT* and *E2*) such that both can be quantified on the same gel. Thus, we have chosen cycles varying in number as 28,30,32,34 and 36 (cycles), and accordingly, the reaction tubes were removed from the PCR block and later loaded on agarose gel.

3.4 Gel electrophoresis and semi-quantitative analysis

The PCR products were loaded onto Ethidium Bromide containing, 1 to 2% (depending on the size of the amplification products) agarose gels in 1x TAE buffer. A 50bp (NEX-GEN DNA Ladder, Cat. No. PG210-500DI) or 1kb (NEX-GEN DNA Ladder, Cat. No. PG010-500DI) DNA ladder was run on every gel to compare expected the molecular weight of the amplification product. Images were acquired with Syngene gel documentation system equipped with Gene Snap software for image analysis.

ImageJ software was used for gel quantification where band intensities were expressed as relative absorbance units. The ratio between the amplified samples (*PpPAL1A* and *PpPAL1B*) and the reference genes (*ACT* or *E2)* was calculated to normalise the initial variations in sample concentration and to increase the reaction efficiency. In all the experiments, mean and standard deviation were calculated after normalisation of values with *ACT* and *E2* reference genes. Semi-quantitative RT-PCR was performed with three biological replicates to ensure the accuracy of the results and a student's t-test was performed to calculate the level of significance.

3.5 *In silico* tissue-specific gene expression analysis

Based on gene expression data from Joerg Becker's laboratory, Botany Array Resource (BAR) Physcomitrella eFP Browser (http://bar.utoronto.ca/efp_physcomitrella/) was used to analyse the *PpPAL1A* and *PpPAL1B* gene expression in different moss tissues. Pp1s346_13V6.1 was used as the probe set identifier for the primary query, Phypa_98188 (*scarecrow-like 5*).

Pp1s456_3V6.1 was used as the probe set identifier for the primary query, Phypa_200712 (*scarecrow-like 5*).

3.6 Identification of putative *At*PAT1-like proteins in *P. patens*

Phytozome search was carried out with *Arabidopsis* PAT1 containing GRAS domain proteins as a query to identify potential PAT1-like proteins in moss with a maximum E-value of $1e^{-5}$. Functional domains were analysed using SMART (http://smart.embl-heidelberg.de/) and PROSITE (http://prosite.expasy.org). UniGene search was performed to identify transcripts from the same locus for *Arabidopsis* PAT1-like proteins in *P. patens*. OrthoMCL (http://orthomcl.org/,v2.0.3) (Li et al., 2003) was used to search for *AtPAT1* orthologous genes in *P. patens* and other species using the entire GRAS protein sequences with a maximum E-value of $1e^{-5}$.

3.7 Multiple sequence alignment and GRAS domain analysis

The annotated genome sequences were retrieved from NCBI Nucleotide database (http://www.ncbi. nlm.nih.gov/) using BLAST searches based on GRAS domain containing proteins for PAT1 branch representing five different species – *Physcomitrella patens (*PpPAL1A, XP_001782936 and PpPAL1B, XP_001785259*)*, *Vitis vinifera (*VvPAT1, XP_002282942.1*), Arabidopsis thaliana* (AtPAT1, At5g48150; AtSCL1, AT1G21450; AtSCL5, At1g50600; AtSCL13, At4g17230; AtSCL21, At2g04890), *Oryza sativa* (OsCIGR1, AAL61820.1 and OsCIGR2, Q8GVE1.1) and *Citrus medica var. sarcodactylis* (CmsGRAS, JF440647.1). Multiple sequence alignment of all PAT1-like amino acid sequences containing GRAS domain were performed using DNAMAN 9.0 software based on conservation and sequence homology with default parameters.

3.8 Phylogenetic analyses

The amino acid sequences encoded by the complete PAT1-like gene families from other species including *Arabidopsis thaliana*, *Vitis vinifera, Oryza sativa* and *Citrus medica var. sarcodactylis* were retrieved from NCBI or Phytozome using BLAST searches. PpTIR1-like (XP_001760786.1) was used as an outlier to construct rooted phylogenetic tree and which does not belongs to the PAT1 branch using MEGA 6.0 software (http://www.megasoftware.net/history.php) and applying the Neighbor-Joining (NJ) method. The bootstrap method was used for test of

phylogeny with 1000 number of replicates. The Poisson model was used as a substitution type for amino acids.

3.9 Primer designing

Based on the region of interest for real-time PCR, full-length cloning and knockout cloning purpose of both PAL1A and PAL1B genes, primers were designed using NetPrimer primer analysis software. Primers were selected manually considering various parameters as length, T_m, C+G content, repetitive bases, self-dimer (ΔG) or cross dimer (ΔG) formation and specificity.

For RT-PCR, primers were designed across exonic junctions to avoid amplification from genomic regions.

No.	Purpose	Primer name	Oligonucleotides (5'-3')	# of bases	Amplicon size (bp)
P1	For qRT-PCR analysis of the reference gene (Actin) in moss	Act_qF	ACCGAGTCCAACATTCTACC	20	112 bp (cDNA)
P2		Act_qR	GTCCACATTAGATTCTCGCA	20	
P3	For qRT-PCR analysis of the reference gene (E2) in moss	E2_qF	TACGGACCCTAATCCAGATGAC	22	116 bp (cDNA)
P4		E2_qR	CAACCCATTGCATACTTCTGAG	22	
P5	Gene detection & tissue specific analysis of PpPAL1A by qRT-PCR	Pp_PAL1-A_q5'F	CCTGAATATGGGAAGCAGTGTCG	23	365 bp (gDNA) 124 bp (cDNA)
P6		Pp_PAL1-A_q3'R	GCTGTTCTGATCTACTCCTTGTTCC	25	
P7	Gene detection & tissue specific analysis of PpPAL1B by qRT-PCR	Pp_PAL1-B_q5'F	TATGGGAAGCAGCAATGCAACG	22	374bp (gDNA) 138bp (cDNA)
P8		Pp_PAL1-B_q3'R	TTCTGCACAAGCTACAAGCAGC	22	
P9[a]	Full length gene amplification to study 35S::*PpPAL1A*	SalI_PAL1-A_FL_5'F	GTCGACGCGGGTGAGTGAGTATTGAAG	27	4481 bp (gDNA) 2745 bp (cDNA)
P10		KpnI_PAL1-A_FL_3'R	GGTACCGAGACAGGTCCTCAGATAACTCG	29	
P11	Full length gene amplification to study 35S::*PpPAL1B*	BamHI_PAL1-B_FL_5'F	GGATCCGCGAAGGTGAATACGAACGTG	27	4468 bp (gDNA) 2292 bp (cDNA)
P12		SacI_PAL1-B_FL_3'R	GAGCTCGGCAACTATTTACCCAGGACT	27	
P13	5'FR amplification to study knockout of PpΔPAL1A	KO_5'FR_PAL1A_KpnIF	AAAAGGTACCCTAGTGTAGCGTGAGGGCTAGTAAG	35	1085 bp (gDNA)
P14		fuseGUS_5'PAL1A_R	CAGGACGTAACAATGTTAGAACCATAGTACCCAGCTCC	38	
P15	GUS gene amplification to study gene promoter activity in PpΔPAL1A	Fuse5'PAL1A_GUS_F	TATGGTTCTAACATTGTTACGTCCTGTAGAAACCCCAA	38	2148 bp (pBI121)
P16		KO_5'FR_GUS_EcoRIR	AAAAGAATTCGATCTAGTAACATAGATGACACCGC	35	
P17	3'FR amplification to study PpΔPAL1A	KO_3'FR_PAL1A_BamHIF	AAAAGGATCCATCGGTCCCTCATTGTTTCCTCTG	34	1185 bp (gDNA)
P18		KO_3'FR_PAL1A_SacIIR	AAAACCGCGGACTAGATCAAATGCCGGAGGTCAG	34	
P19	5'FR amplification to study PpΔPAL1B	KO_5'FR_PAL1B_KpnIF	AAAAGGTACCGAGTAGGCGTTTGTGGTACTTGAC	34	1233 bp (gDNA)
P20		fuseGUS_5'PAL1B_R	CAGGACGTAACAAAGTAAGGACCATAGAACCCAGCTCT	38	
P21	GUS gene amplification to study gene promoter activity in PpΔPAL1B	Fuse5'PAL1B_GUS_F	TATGGTCCTTACTTTGTTACGTCCTGTAGAAACCCCAA	38	2148 bp (pBI121)
P22		KO_5'FR_GUS_HindIIIR	AAAAAAGCTTGATCTAGTAACATAGATGACACCGC	35	
P23	3'FR amplification to study PpΔPAL1B	KO_3'FR_PAL1B_BamHIF[a]	AAATGGATCCGGCCCTAGCCTGTTCTATTTGTAA	39	1328 bp (gDNA)
P24		KO_3'FR_PAL1B_NotIR	AAAGCGGCCGCGTTGGAAGCCAAGTGACATTAGTGC	36	

Table 3.3 List of primers used in present study. ([a] Red marked are the restriction sites included in the primer; Grey colored section represents the primers used as reference gene for semi-quantitative RT-PCR; Pink colored section represents the primers used for gene expression analysis of both *PpPAL1A* and *PpPAL1B* genes; Yellow colored section represents the primers used for the cloning of overexpression

construct *viz* 35S::*PpPAL1A* and *35S:: PpPAL1B*; Blue colored section represents the primers used for the cloning of *viz PpΔpal1a* and *PpΔpal1b*).

3.10 Plasmids

The plasmid vectors used in the present study were listed as follows:

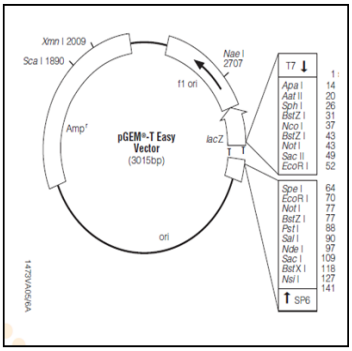
a. pGEM-T Easy Vector (3015 bp) as subcloning vector with ampicillin resistance in *E. coli*

b. pCAMBIA1300 binary vector (10,067 bp) for cloning 35S:: *PpPAL1A* overexpression construct with kanamycin resistance in *E. coli* and hygromycin resistance in *P. patens*

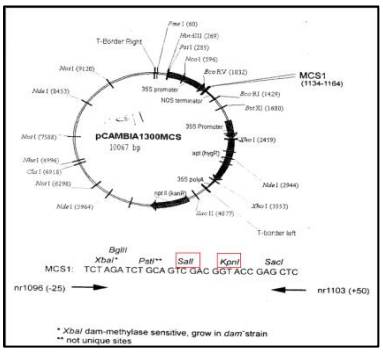
c. pBI121 binary vector for cloning 35S:: *PpPAL1B* overexpression construct with kanamycin resistance in *E. coli* and Geneticin (G418) resistance in *P. patens*

d. pTN182 (*Floxed modified nptII cassette*) vector for cloning *PpΔpal1a* knockout construct with kanamycin resistance in *E. coli* and Geneticin (G418) resistance in *P. patens*

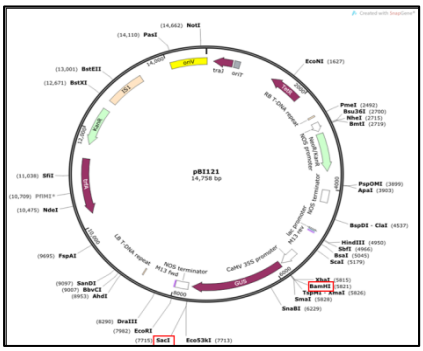
e. pTN186 (*Floxed modified aphIV cassette*) vector for cloning *PpΔpal1b* knockout construct with ampicillin resistance in *E. coli* and Geneticin (G418) resistance in *P. patens.*



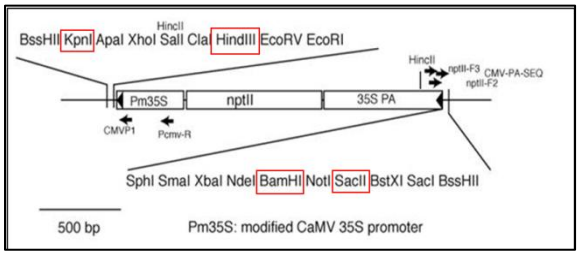
(a) pGEM-T Easy Vector



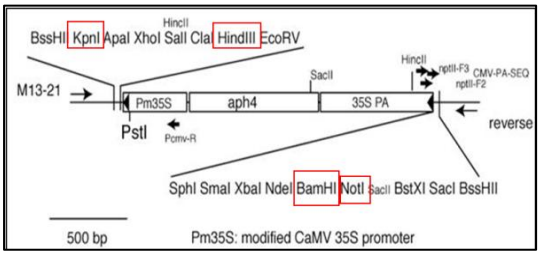
(b) pCAMBIA1300



(c) pBI121



(d) pTN182



(e) pTN186

Figure 3.2 Map of plasmid vectors used in present study. (a) pGEM-T Easy Vector; (b) pCAMBIA1300; (c) pBI121; (d) pTN182; and (e). pTN186. (A red box marked indicates the restriction sites chosen for cloning purpose).

3.11 Cloning of overexpression constructs of 35S:: *PpPAL1A* and 35S:: *PpPAL1B*

To produce 35S::*PpPAL1A* & *35S::PpPAL1B* overexpression constructs, full-length coding sequences along with 5' and 3'UTR regions, 2745 bp and 2292 bp respectively were amplified from cDNA using Phusion High-Fidelity DNA Polymerase (NEB) with primers containing restriction sites (P9-P10 for 35S:: *PpPAL1A* and P11-P12 for *35S:: PpPAL1B*, Table 3.3). The PCR amplified fragments were digested with *SalI-HF – KpnI-HF* (NEB) for 35S:: *PpPAL1A* and with *BamHI – SacI* (Promega) for *35S:: PpPAL1B* and subcloned into a pGEM-T Easy vector (Promega, USA).

The pGEM-T clones A7 for *PAL1A* and SB6 for *PAL1B* were confirmed by restriction digestion and also by sequencing from both T7 and SP6 promoter ends. Later the digested products were gel-excised and purified using Wizard® SV Gel and PCR Clean-Up System (Promega), following the manufacturer's instructions. The qualitative and quantitative analysis was done using Nanodrop spectrophotometer.

Further, the purified products were inserted into the binary vectors pCAMBIA1300 and pBI121 by ligation using T4 DNA ligase (NEB). Overexpression clones A1 for *PAL1A* and B2 for *PAL1B* were confirmed by PCR amplification, restriction digestion and sequencing from gene-specific primers used for cloning. The recombinant clone of 35S:: *PpPAL1A* and 35S:: *PpPAL1B* was mobilized to *Agrobacterium tumefaciens* strain C58 GV2260 for plant transformation purpose.

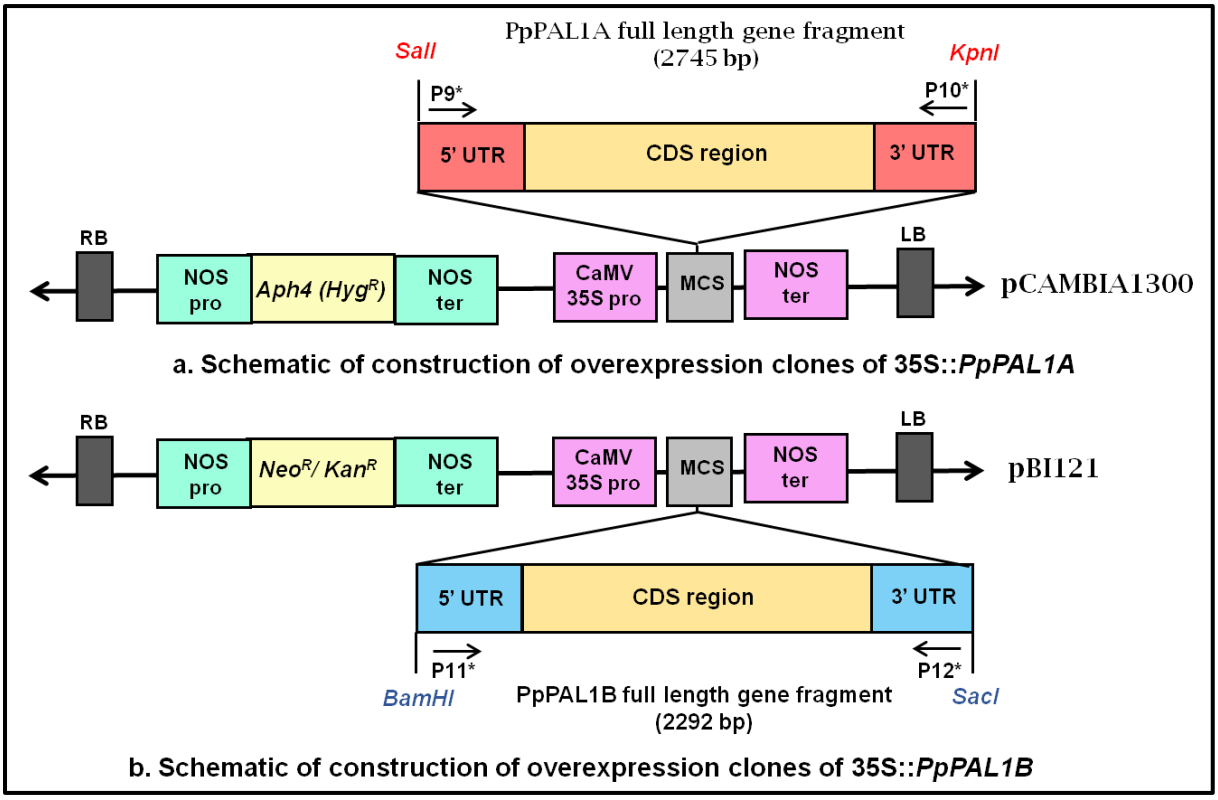


Figure 3.3 Schematics of cloning of full-length genes *PpPAL1A* and *PpPAL1B* in pCAMBIA1300 and pBI121 binary vector. (* represents the restriction sites present in the primers)

3.12 *Agrobacterium*-mediated transformation of protonemata

Over expression lines were generated using *Agrobacterium*-mediated transformation technique (www.moss.nibb.ac.jp/) as mentioned below-

a. ***Agrobacterium* Culture and Co-Culture**

i. Overexpression construct was transformed into *Agrobacterium tumefaciens* strain C58 GV2260 with ~5-10 µg plasmids.

ii. Transformed *Agrobacterium* colony was inoculated in 5ml LB broth with the addition of appropriate antibiotics (50 mg/ml Kanamycin and 50 mg/ml Rifampicin) and was grown at 180rpm for 24 hrs at 28ºC.

iii. The culture was centrifuged at 3,000 rpm for 7 min at room temperature, and the supernatant was discarded.

iv. The resulting pellet was washed by adding 5 ml BCDAT +5% Glucose and re-suspend slowly by careful pipetting. Resuspended culture was centrifuged again at 3,000 rpm for 7 min at room temperature, and the supernatant was discarded.

v. The previous step (iv) was repeated once.

vi. Filter sterilised Acetosyringone was added to the co-culture media prepared in the BCDAT +5% Glucose solution with 200 mM final concentration.

vii. The pellet was re-suspended in 2 ml co-culture media and grown at 28ºC for 2 hrs at 180 rpm.

viii. The absorbance of bacterial culture was checked at 600 nm on spectrophotometer in UV-VIS mode. Path correction was performed by using the formula given below:

Volume needed of bacterial culture = 0.1*10/Nanodrop reading path length correction

ix. 10 ml of co-culture media was pipette out into a 9 cm Petri plates and swirled gently to cover the bottom.

x. 4-5 days old homogenised protonemata tissue were scraped into the Petri plate containing co-culture media.

xi. Petri plates were sealed with two strips of Parafilm to avoid contamination and leakage.

xii. Petri plates were incubated at 25ºC under continuous light for two days.

b. Washing and plating

This step requires to be performed carefully as left over Agrobacterium is capable of killing the generated transformants.

i. Two-day-old tissue grown in co-cultured plates was taken in a fresh Petri plate using forceps, and excess liquid was removed by cut-tip.

ii. 10 ml of washing media was added to the plate and tissue was re-suspended in media and washing media was removed using cut-tip.

iii. Step ii was repeated thrice.

iv. After multiple washes, the tissue was overlaid on cellophane layer placed on BCDAT agar media containing appropriate antibiotics for selection (Hygromycin B, final 20mg/L; Geneticin, final 20 mg/L) + Claforan (final 100 µg/ml) + Augmentin (final 50 µg/ml). Plates were sealed with surgical tape and kept at 24ºC for two weeks.

v. After two weeks on selection-I, cellophane layer was transferred to relaxation media containing only anti-bacterial agents like Claforan and Augmentin (no specific antibiotic that was used for transformant selection) and kept at 24ºC for about 1-2 weeks**.**

vi. A small piece of each line was transferred from the relaxation plate to BCDAT plates with antibiotic selection-II.

Moss lines grown on selection-II for two weeks were considered as positive transformants.

3.13 Construction of knockout clones of *PpΔpal1a* and *PpΔpal1b* with *GUS* fusion

Along with homologous recombination strategy, we have used PCR gene fusion method for construction of knockout vectors for *PpPAL1* genes. Oligo primers were designed and used for the knockout construction of *PpΔpal1a* (P13-P18, Table 3.3), and similarly for the knockout construction of *PpΔpal1b* (P19-P24, Table 3.3). Fragments like 5' flanking region (FR) from *PpPAL1* gene and *GUS* gene from the pBI121 binary vector, which need to be fused were amplified in separate PCR reactions. The primers were designed so that the ends of the products should contain complementary sequences (as shown in Figure 3.5 for 2R and 3F). When these PCR products were mixed, denatured, and annealed, the strands having the matching sequences at their respective ends overlap and act as primers for each

other. Extension of this overlap by DNA polymerase produces a molecule in which the original sequences are fused together as a single amplicon.

Approximately around 1kb of 5' flanking region consisting promoter of *PpPAL1* gene was amplified from gDNA and fused with *GUS* ORF amplified from pBI121 binary vector plasmid DNA. This fusion fragment will be digested and cloned into a vector containing antibiotic cassette. Then around 1Kb of 3' flanking region of *PpPAL1* gene amplified from gDNA will be cloned into the vector containing cloned 5'flanking region fused with *GUS* ORF. The final knockout construct will contain 5' flanking region of *PAL1- GUS*- Antibiotic cassette- 3' flanking region of *PAL1* gene sequence.

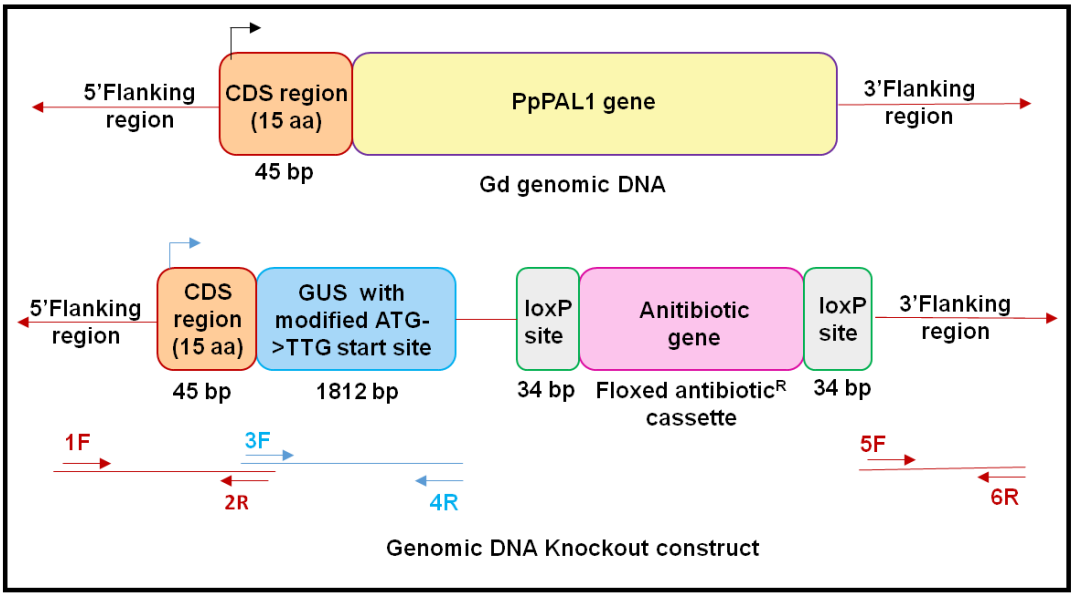


Figure 3.4 Schematic of knockout cloning constructs of *PpΔpal1a* and *PpΔpal1b*.

3.14 Microscopy imaging

Leica microscope was used for imaging of moss cultures and phenotypic characterization of the overexpression mutant lines. The *Agrobacterium-mediated* transformants were routinely observed for the growth of mutant lines.

Chapter 4

Results and discussion

4.1 Identification of *AtPAT1* gene orthologs in *P. patens* using bioinformatics tools

a. Identification of the GRAS domain containing protein sequences and isolation of *AtPAT1*-like gene orthologs in *P. patens*

To identify putative *P. patens* PAT1-like protein, we first searched Phytozome database using a published *Arabidopsis* PAT1 protein with conserved GRAS domain protein sequences as a query. The deduced AtPAT1-like protein sequences were searched on UniGene as well as OrthoMCL, and the results exhibited the presence of two orthologs of *AtPAT1* genes in *P. patens*. The OrthoMCL blast sequence results showed E-value of $1e^{-171}$ with score value 496 for PpPAL1A (#Accession- ppat|fgenesh1_pm.scaffold_346000001) and E-value of $3e^{-166}$ with score value 482 for PpPAL1B (#Accession- ppat|estExt_gwp_gw1.C_4560001). The Pfam search showed that there is only one GRAS domain (~355 aa) present in both PpPAL1A and PpPAL1B. These two genes from *P.patens* were named as *PpPAL1A* and *PpPAL1B,* were further classified and analysed in the present study. The detailed information on the *PpPAL1A* and *PpPAL1B* genes in P. patens, including their gene IDs and the classification of both the genes, are mentioned in Table 4.1.

Gene	*PpPAL1A*	*PpPAL1B*
Accession no. (NCBI)	XP_001782936	XP_001785259
Protein ID (JGI)	61285	200712
Protein length (aa)	657	649
Gene ID (NCBI)	5946141	5948465
Location (Cosmoss)	Chr01:1,864,870..1,869,443 (- strand)	Chr02:20533174..20537852 (+ strand)
Gene length (bp)	4574	4679
Transcript ID	Pp1s346_13V6.5	Pp1s456_3V6.2
Landmark (Cosmoss)	scaffold_346:146,641..151,214	scaffold_456:8,847..13,525
mRNA length (bp)	2838	2503

Table 4.1 Classification of *PpPAL1A* and *PpPAL1B* genes.

The GRAS protein family seems unique to plants and currently consists 40 members with 129 GRAS transcription factors in *P. patens*. In 2000, Bolle *et al.* showed that AtPAT1 showed the highest homology (45-70% identity) with AtSCL1/ 5 and 13 subgroups in *A. thaliana*. A Recent finding from Yuan *et al.,* in 2015, showed that in PAT1-branch, VaPAT1 showed 74% identity with CmsGRAS and 52% identity with OsCIGR1. Thus, to isolate *AtPAT*-like orthologous genes - *PpPAL1A* and *PpPAL1B*, it was important to find out the identities and similarities between and among the PAT1-like GRAS domain containing protein sequences. In the present study, we have looked across five different plant species like *Physcomitrella patens*, *Arabidopsis thaliana*, *Vitis vinifera, Oryza sativa* and *Citrus medica var. sarcodactylis*.

Our findings suggest that, both PpPAL1A and PpPAL1B show 82.95% identity at the level of amino acids. When compared to other PAT1-like GRAS domain containing proteins, both PpPAL1A and PpPAL1B showed 24 to 40% identity.

Identity Similarity	PpPAL1A	PpPAL1B	CmsGRAS	VvPAT1	OsCIGR1	OsCIGR2	AtPAT1	AtSCL1	AtSCL5	AtSCL13	AtSCL21
PpPAL1A	-	-	-	-	-	-	-	-	-	-	-
PpPAL1B	0.8295 0.8873	-	-	-	-	-	-	-	-	-	-
CmsGRAS	0.3495 0.4589	0.3547 0.4556	-	-	-	-	-	-	-	-	-
VvPAT1	0.3938 0.5072	0.3953 0.5014	0.4846 0.5767	-	-	-	-	-	-	-	-
OsCIGR1	0.3206 0.4446	0.3279 0.4475	0.3773 0.4734	0.4796 0.6003	-	-	-	-	-	-	-
OsCIGR2	0.3649 0.4948	0.3831 0.5088	0.4517 0.5464	0.4313 0.5383	0.3312 0.4595	-	-		-	-	-
AtPAT1	0.3766 0.4843	0.3757 0.4878	0.4846 0.5767	0.4363 0.5504	0.4796 0.6003	0.4955 0.6373	-	-	-	-	-
AtSCL1	0.3854 0.5238	0.3727 0.5089	0.3174 0.4457	0.3328 0.4711	0.2728 0.4039	0.3333 0.4746	0.3536 0.5032	-	-	-	-
AtSCL5	0.3916 0.5231	0.4040 0.5329	0.4108 0.5057	0.4071 0.5402	0.3107 0.4374	0.4786 0.6213	0.4754 0.5898	0.3213 0.4693	-	-	-
AtSCL13	0.2435 0.3165	0.2438 0.3148	0.4355 0.5158	0.2933 0.3636	0.2226 0.2956	0.3308 0.4044	0.3795 0.4632	0.2495 0.3338	0.3031 0.3768	-	-
AtSCL21	0.3237 0.4372	0.3328 0.4417	0.5625 0.6898	0.4000 0.5000	0.3117 0.4139	0.4332 0.5415	0.5694 0.6780	0.3127 0.4413	0.4032 0.5081	0.4109 0.5011	-

Table 4.2 Sequence identities and similarities among and between PAT1 branch containing GRAS domain proteins from different plant species. Name of the protein was coloured black; Values coloured in blue indicates the identity value

between two proteins; Values coloured in red indicates the similarity value between two proteins.

b. Sequence alignment and phylogenetic analysis of PpPAL1A and PpPAL1B with other GRAS domain containing proteins

The amino acid sequences for alignment were selected based on the conservation and sequence homology for PpPAL1A and PpPAL1B with other PAT1-like GRAS domain containing proteins and were aligned using DNAMAN 9.0 software. Alignment of PpPAL1A and PpPAL1B with their homologues derived from other five plant species showed a variable N-terminal and a conserved C-terminal domain similar to other members of the GRAS family. The five distinct typical motifs at the C-terminal part: LRI, VHIID, LRII, PFYRE, and SAW defined for GRAS proteins were also conserved in both PpPAL1A and PpPAL1B proteins. The leucine-rich domains LRI and LRII flanked a conserved V/I HIID domain (Figure 4.1).

The phylogenetic tree was constructed using a Neighbour-joining (NJ) method by aligning the full-length PAT1-like protein sequences from five distinct species. The bootstrap value shown at each branching point reflects the percentage of 1,000 iterations (Figure 4.2). The sequence alignment and phylogenetic tree together confirmed that PpPAL1A and PpPAL1B are the members of the PAT1-branch of the GRAS domain protein family.

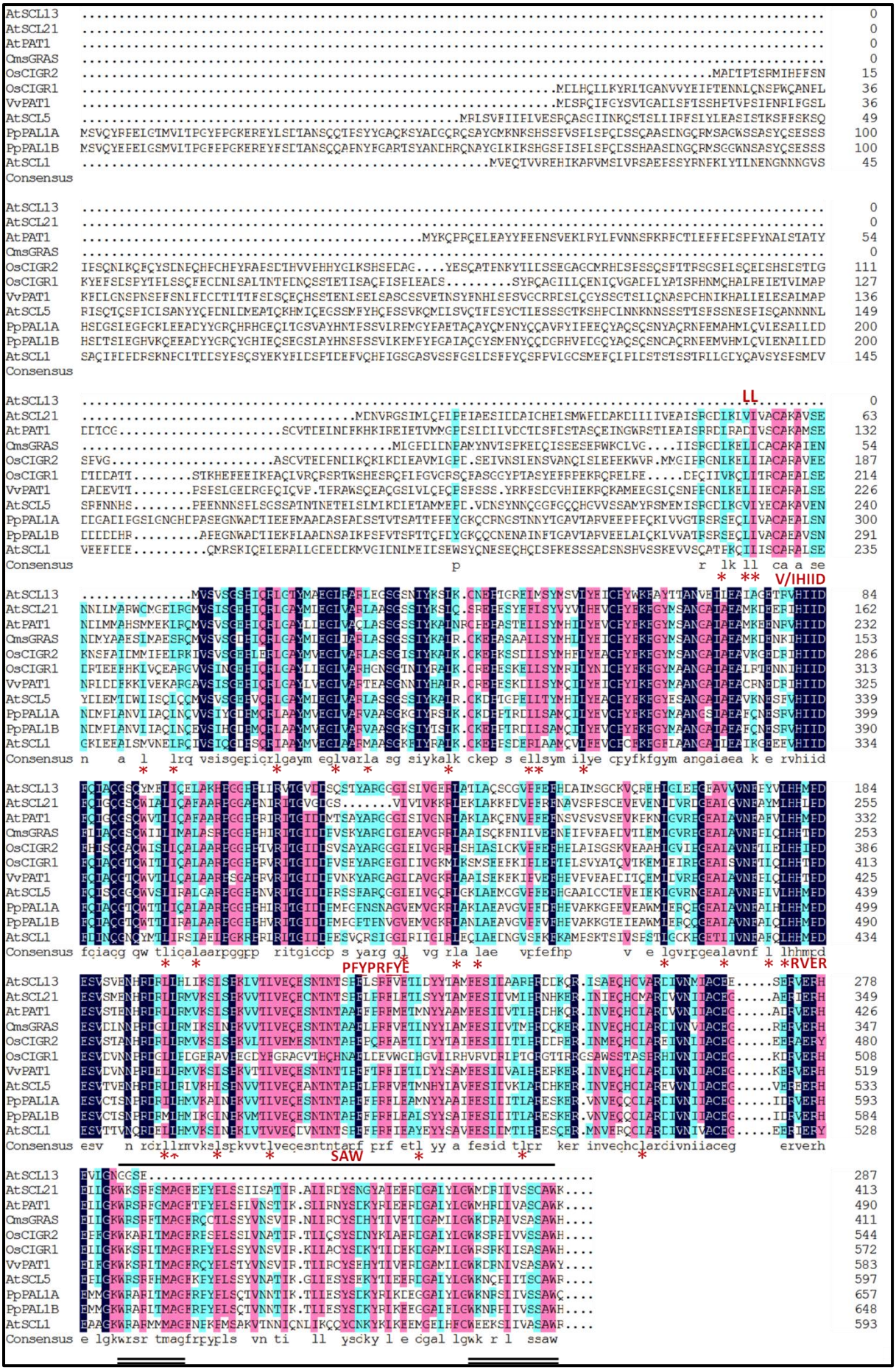

LL
V/IHIID
PFYPRFYE
RVER
SAW

Figure 4.1 Sequence alignment of PpPAL1A and PpPAL1B with other GRAS domain proteins. (Following sequences were used for alignment: *Physcomitrella patens (*PpPAL1A, XP_001782936 and PpPAL1B, XP_001785259*)*, *Vitis vinifera (*VvPAT1, XP_002282942.1*)*, *Arabidopsis thaliana* (AtPAT1, At5g48150; AtSCL1, AT1G21450; AtSCL5, At1g50600; AtSCL13, At4g17230; AtSCL21, At2g04890), *Oryza sativa* (OsCIGR1, AAL61820.1 and OsCIGR2, Q8GVE1.1) and *Citrus medica var. sarcodactylis* (CmsGRAS, JF440647.1). Conserved sequences are coloured. The dashes (-) represent gaps introduced to facilitate alignment. *Asterisks* (*) mark denotes the conserved leucine residues in the leucine-rich domains flanking the V/I HIID motif. $W(X)_7G$, $W(X)_{10}W$, the two pairs of conserved residues in SAW domain are *double-underlined*. The five main conserved domains in the C-terminal are indicated with uppercase letters above the alignment.

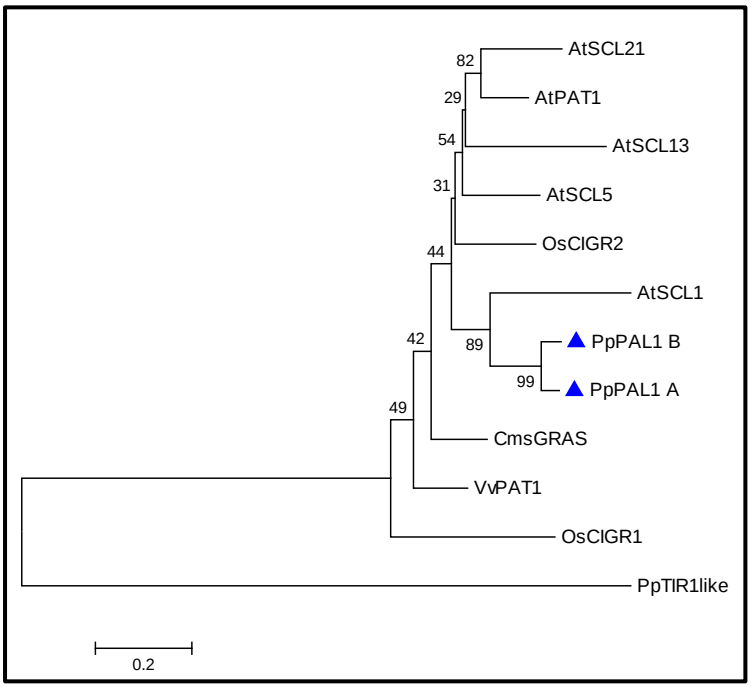


Figure 4.2 Phylogenetic analyses of PpPAL1A and PpPAL1B with other GRAS domain proteins. All *Arabidopsis* PAT1 branch proteins (AtPAT1; AtSCL1; AtSCL5; AtSCL13; AtSCL21) as well as several stress-related proteins from other species *viz Vitis vinifera* (VvPAT1); *Oryza sativa* (OsCIGR1; OsCIGR2); *Citrus medica var. sarcodactylis* (CmsGRAS) were also included. PpTIR1-like protein was used as an outlier for the construction of the rooted tree. The *blue coloured triangle* indicates PpPAL1A and PpPAL1B proteins. The bootstrap values are indicated at the branch

points. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances that were used to infer the phylogenetic tree (Scale bar, 0.2 indicates the amino acid substitutions per site). All positions that contain alignment gaps and missing data were eliminated only in pairwise sequence comparisons.

4.2 Tissue-specific gene expression profile of *AtPAT1* gene orthologs in *P. patens*

a. *In silico* analysis of tissue specific gene expression of *PpPAL1A* and *PpPAL1B* in *P. patens*

Using *Physcomitrella* eFP Browser, 'electronic fluorescent pictographic' representations of *PpPAL1A* and *PpPAL1B* gene expression patterns were showed based on *Physcomitrella* microarray data set (Winter et al., 2007). For *PpPAL1A*, we observed that it is constitutively expressed in all tissue types (as shown in Figure 4.3 (i)). Interestingly, compared to our primary query identifier (Phypa_98188, *scarecrow-like 5*), rhizoid and caulonemal filaments showed moderate expression. With respect to the gene expression profile of *PpPAL1B*, we see an interesting pattern where expression of the gene is concentrated only in the sporophytic phase. S2 stage shows the highest expression of *PpPAL1B,* whereas the lowest expression is seen in protonemal stage (Figure 4.3 (ii)).

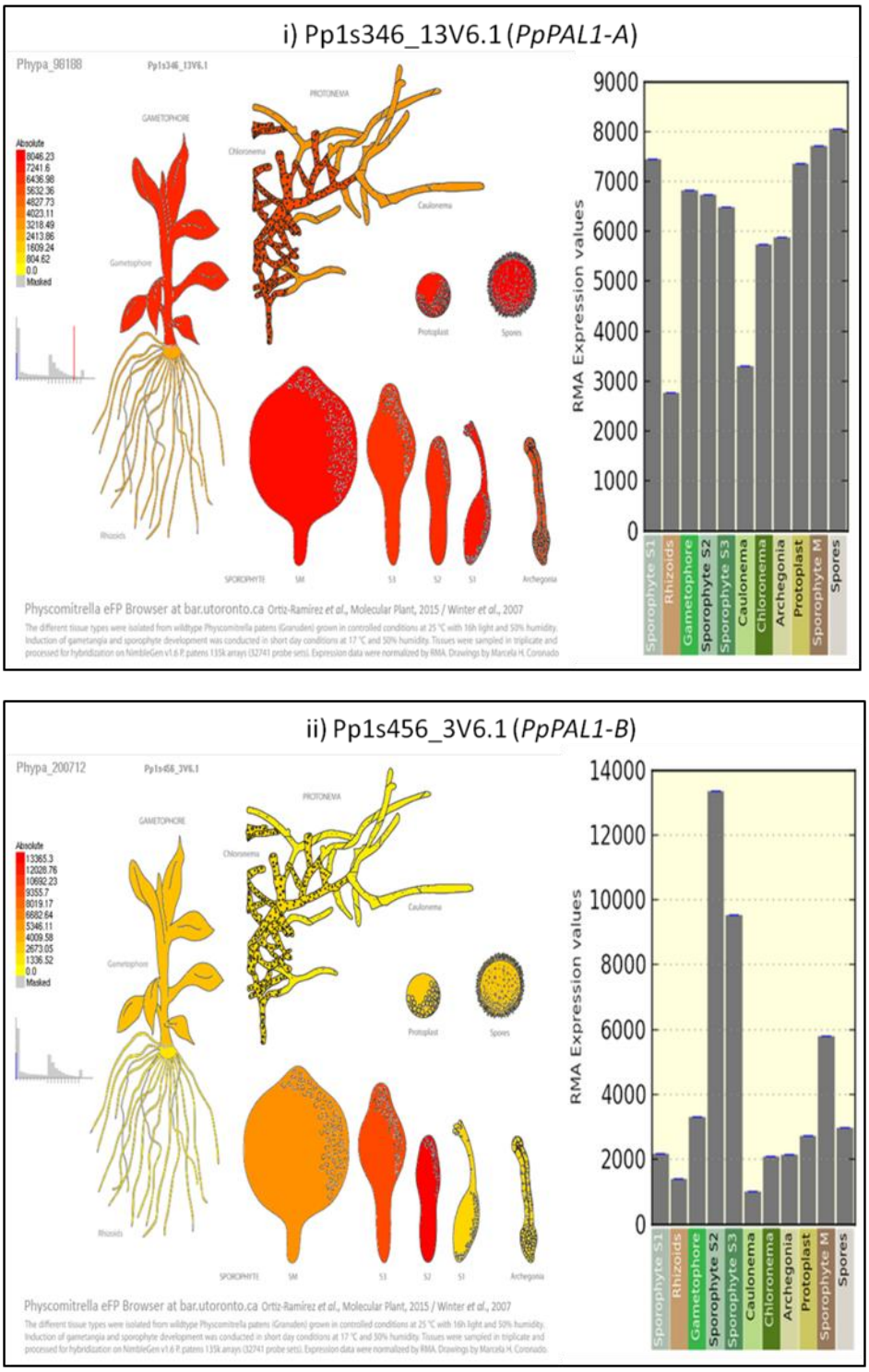


Figure 4.3 *In silico* analysis of tissue specific gene expression of *PpPAL1A* and *PpPAL1B* in *P. patens.* (i) Tissue-specific gene expression of *PpPAL1A.* For *PpPAL1A* gene expression data, this probe set reaches its maximum expression level (expression potential) of 8046.23 in the *Physcomitrella* data source. *Pp1s346_13V6.1* was used as the probe set identifier for the primary query, Phypa_98188 *(scarecrow-like 5);* (ii) Tissue-specific gene expression of *PpPAL1B.*

For *PpPAL1B* gene expression data, this probe set reaches its maximum expression level (expression potential) of 13365.31 in the *Physcomitrella* data source. Pp1s456_3V6.1 was used as the probe set identifier for the primary query, Phypa_200712 (scarecrow-like 5). (Using *Physcomitrella* eFP Browser *online tool).*

b. Detection and amplification of *PpPAL1A* and *PpPAL1B* genes in wild-type P. patens using both gDNA and cDNA

gDNA was isolated from 2weeks-old whole moss tissue. RNA was isolated from 7days-old protonema and 14days-old gametophore tissue of *P. patens,* and ~2µg of isolated RNA was converted to cDNA. The *PpPAL1A* and *PpPAL1B* short gene fragments were amplified and detected by semi-quantitative RT-PCR using specific primers. P5- P6 primers were used to amplify specific 124 bp fragment of *PpPAL1A* gene from cDNA and 365 bp fragment from gDNA. Similarly, P7- P8 primers were used to detect *PpPAL1B* with amplification of 138 bp from cDNA and 374 bp from gDNA.

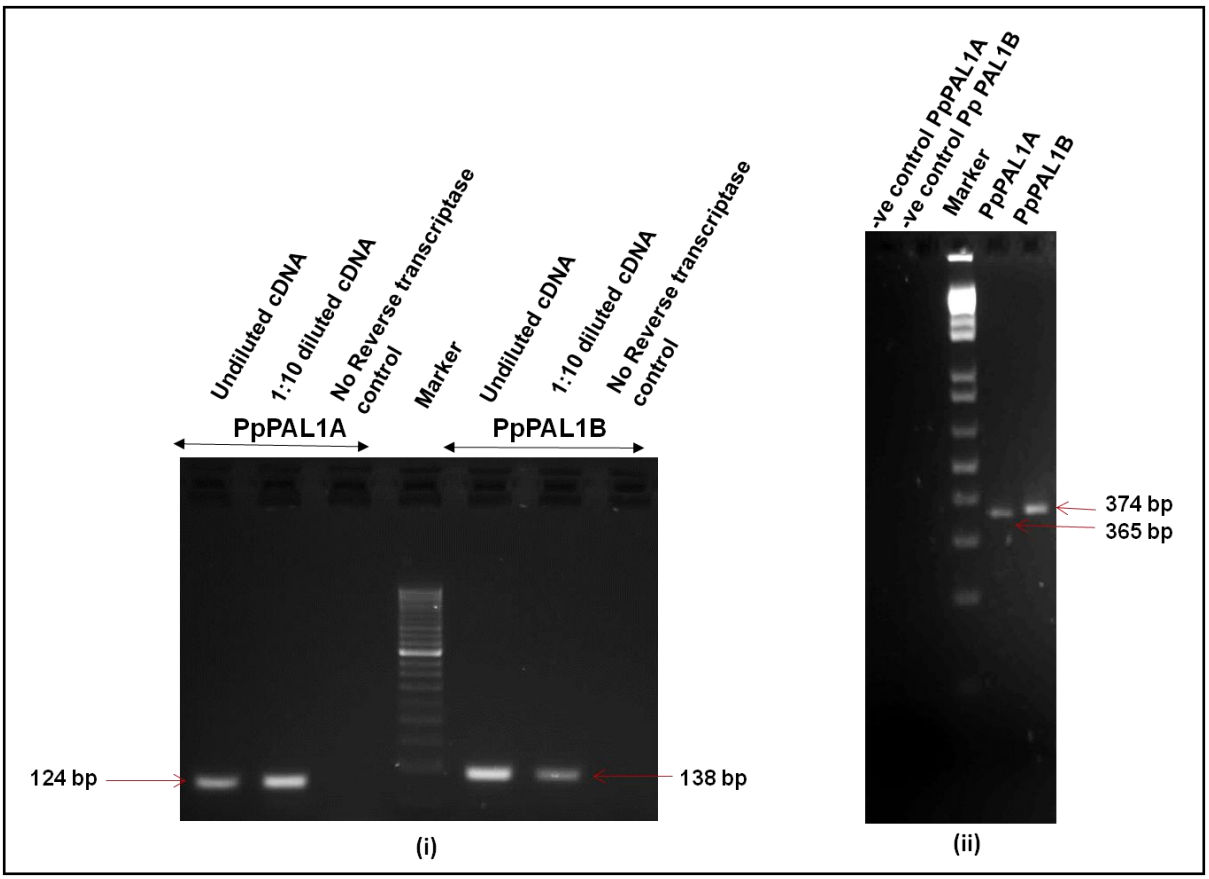


Figure 4.4 Detection of *PpPAL1A* and *PpPAL1B* in wild-type *P. patens*. (i) Detection of *PpPAL1A* (~124 bp) and *PpPAL1B* (~138 bp) from cDNA. cDNA prepared using no reverse transcriptase was used as negative control; (ii) Detection of *PpPAL1A* (~365 bp) and *PpPAL1B* (~374 bp) from gDNA. PCR reactions run without gDNA template were considered as negative control.

c. Semi-quantitative RT-PCR validation of tissue-specific gene expression profile of *PpPAL1A* and *PpPAL1B* in protonema and gametophore tissue types of *P. patens*

For tissue specific gene expression of *PpPAL1A* and *PpPAL1B*, two tissue types were chosen: 7days-old protonema and 2weeks-old gametophore tissue. Exponential amplification of both *PpPAL1A* and *PpPAL1B* genes occur during PCR cycles 28 to 34, compared to reference genes *ACT* and *E2* (as shown in Figure 4.5 (a) and (b)). Expression analysis profile supported the *in silico* gene expression analysis by comparing the level of expression of *PpPAL1A* and *PpPAL1B* in protonema and gametophore tissues. Our data showed a relative fold change of 1.79 for *PpPAL1A* on comparing levels in gametophore to protonema. Similarly, a relative fold change of 1.49 was observed for *PpPAL1B* on comparing levels in gametophore to protonema. Statistical t-test (paired two samples for means) showed a significant difference for both the genes, with a p- value of 0.01 for *PpPAL1A* and a p- value of 0.03 for *PpPAL1B*.

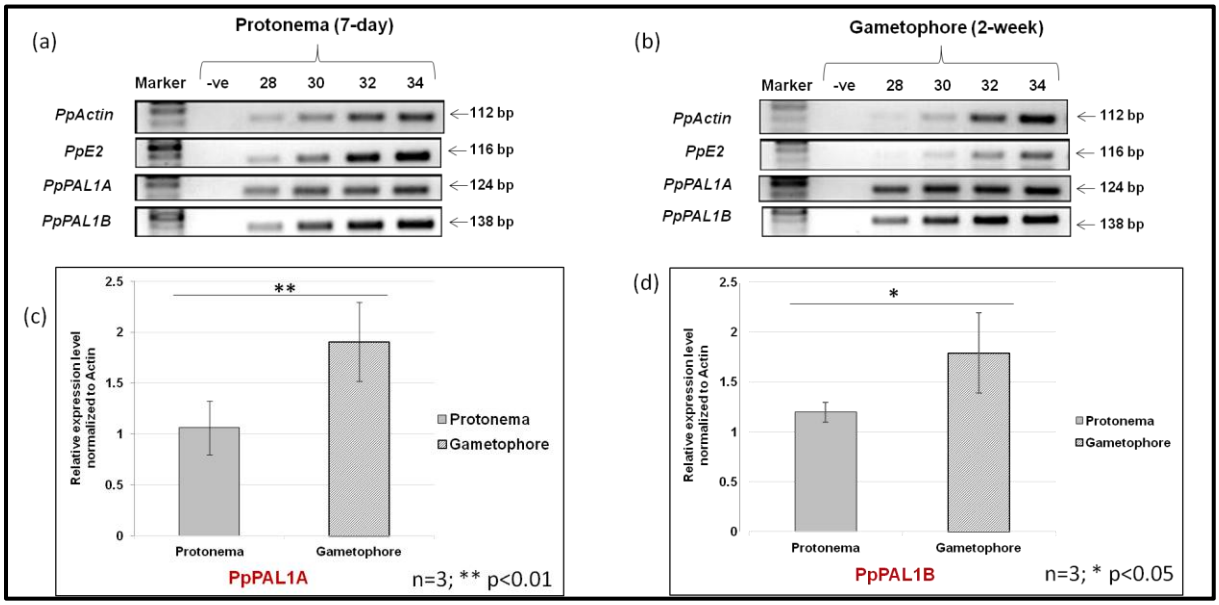


Figure 4.5 Validation of tissue-specific gene expression profiles of *PpPAL1A* and *PpPAL1B* by semiquantitative RT-PCR. (a) Gene expression profile in 7days-old protonema tissue; (b) Gene expression profile in 2weeks-old gametophore tissue (Negative control contains no cDNA as a template) ; (c) Relative expression levels for *PpPAL1A* were analysed after normalisation with *ACT* gene; (d) Relative expression levels for *PpPAL1B* were analysed after normalisation with *ACT* gene. (Two reference genes *ACT* and *E2* were chosen as an internal control. PCR cycles from 28 to 34 were chosen depending on transcripts abundance. A t-test was carried out with three biological replicates where '**' represents $p < 0.01$; '*' represents $p < 0.05$)

4.3 Generation of overexpression constructs and transgenic mutant lines by targeted mutagenesis

a. Cloning of full-length *PpPAL1A* and *PpPAL1B* genes in pCAMBIA1300 and pBI121 binary vectors respectively under 35S promoter

The full-length gene sequences from gene transcript IDs Pp1s346_13V6.5 and Pp1s456_3V6.2 were retrieved for *PpPAL1A* and *PpPAL1B* respectively. Around 2745 bp, the full-length gene was amplified for *PpPAL1A* from cDNA using P9-P10 primers. Similarly, ~2292 bp full-length gene was amplified for *PpPALB* from cDNA using P11-P12 primers (Figure 4.6).

The PCR-amplified full-length gene fragments were subcloned into a pGEM-T Easy vector (~3 kb). Plasmids were screened for clone confirmation. Restriction digestion of pGEM-T clones A1 and A7 with EcoRI enzyme generated desired fragments as for *PpPAL1A* showing 2997 bp, ~1500 bp and 759 bp bands (Figure 4.7 (c)). Similarly, pGEM-T clones for *PpPAL1B viz* SB5 and SB6 generated desired 2997 bp and 2323 bp fragments (Figure 4.7 (d)).

The pGEM-T clones A7 and SB6 were confirmed by sequencing from T7, and SP6 promoter ends. The sequence of Clone A7 was additionally confirmed from qRT-PCR primers (P5-P6) which cover the specific gene sequence (shown in Table 4.3). Sequencing results revealed that 413 bp sequences were missing out of 551 bp from the 5' UTR.1 resulting in a *PpPAL1A* splice variant (Figure 4.7 (a) and (b)).

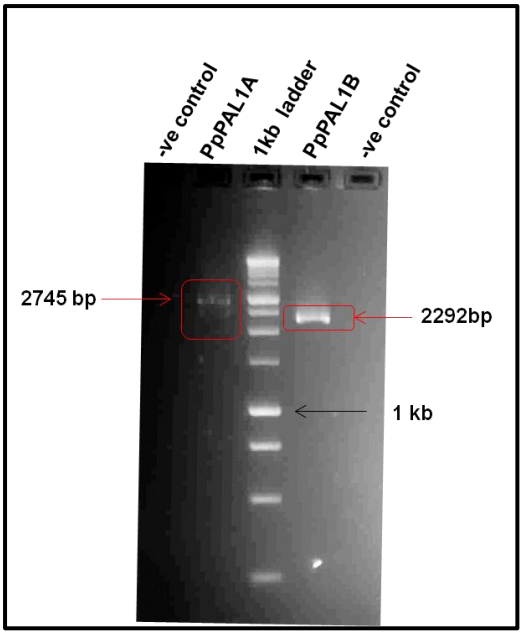


Figure 4.6 Full-length gene amplification of *PpPAL1A* and *PpPAL1B* using WT cDNA. (Negative control contains no template cDNA)

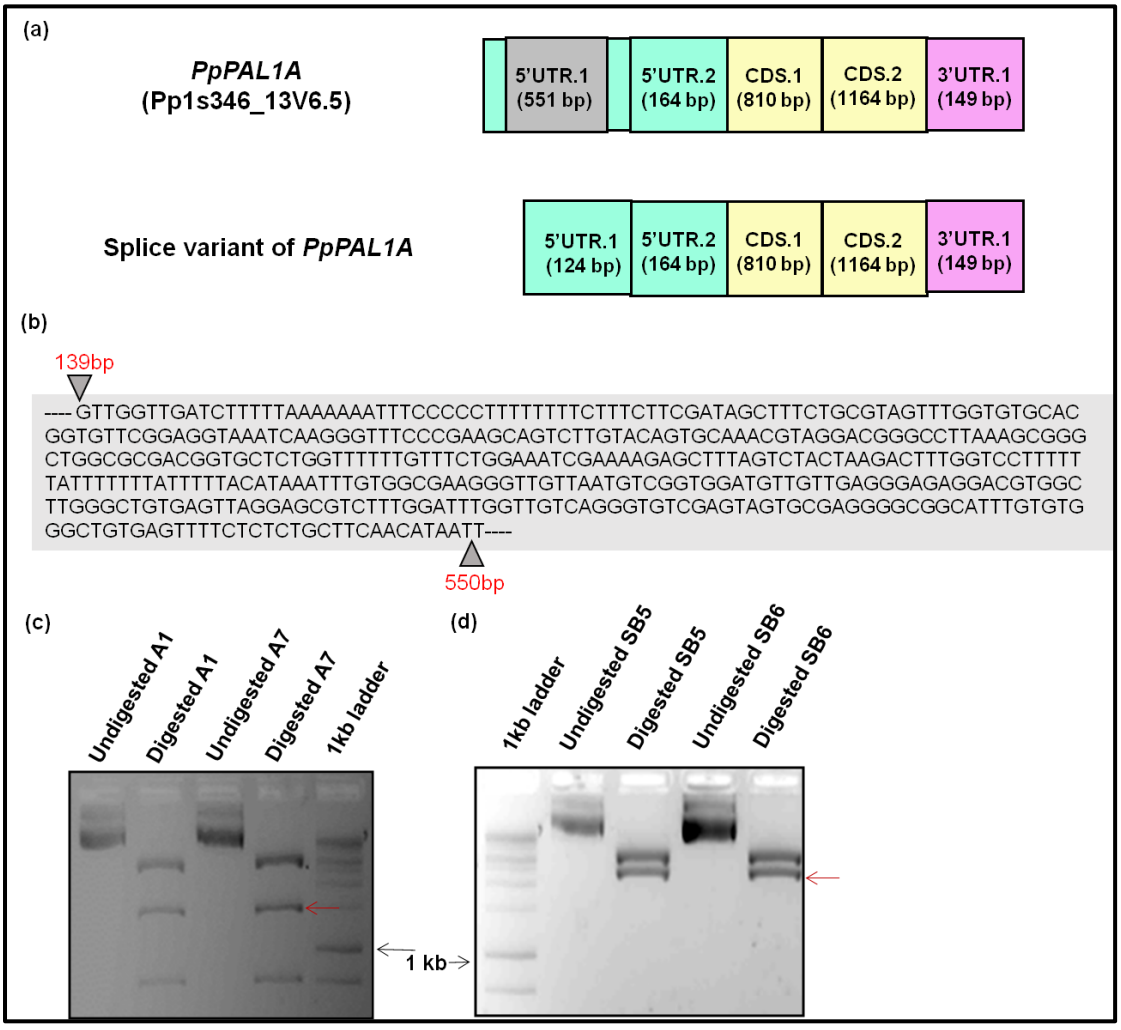


Figure 4.7 Confirmation of subcloning of full-length *PpPAL1A* and *PpPAL1B* gene into a pGEM-T easy vector. (a) Representative schematic of splice variant of *PpPAL1A* (Grey region indicates the missing region from 5'UTR.1); (b) Nucleotide sequence of 413 bp missing out of 551 bp of 5'UTR.1 resulting in *PpPAL1A* splice variant (triangle indicates the location of start and end nucleotide of missing region); (c) Confirmation of subcloning of full length *PpPAL1A* gene into pGEM-T Easy vector; (d) Confirmation of subcloning of full length *PpPAL1B* gene into pGEM-T easy vector.

Using appended restriction sites from the primers, the full-length gene fragments were further digested and cloned into the binary vectors (Figure 3.2). *SalI* and *KpnI* were used for cloning of 35S:: *PpPAL1A* (~2.3 kb) insert into the pCAMBIA1300 binary vector (~10 kb). Similarly, *BamHI* and *SacI* were used for cloning of 35S:: *PpPAL1B* (~2.3 kb) into the pBI121 binary vector (~13kb) (Figure 4.8).

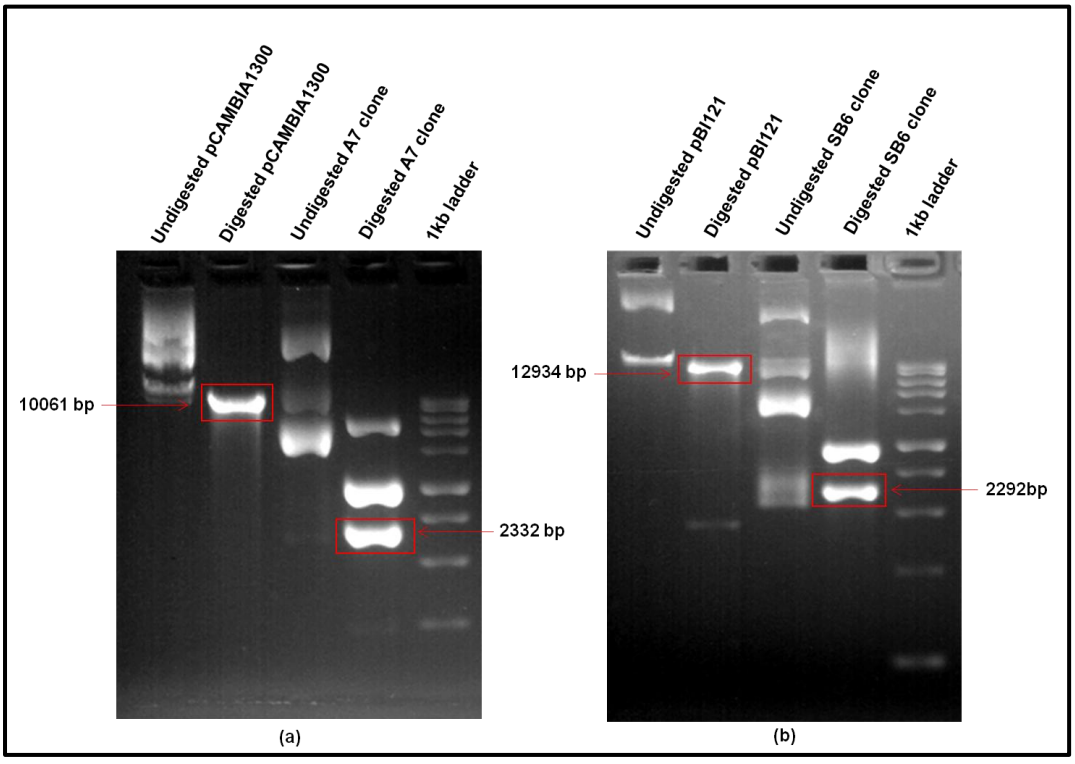


Figure 4.8 Cloning of *PpPAL1A* and *PpPAL1B* genes in binary vectors *viz* pCAMBIA1300 and pBI121 respectively. (a) Cloning of 35S::*PpPAL1A* in binary vector pCAMBIA1300; (b) Cloning of 35S::*PpPAL1B* in binary vector pBI121.

For overexpression construct, 35S:: *PpPAL1A,* clone C1 was confirmed by PCR amplification while for overexpression construct, of 35S:: *PpPAL1B,* clone B2 was confirmed by both PCR amplification and restriction digestion with *BamHI- SacI* restriction enzymes (Figure 4.9).

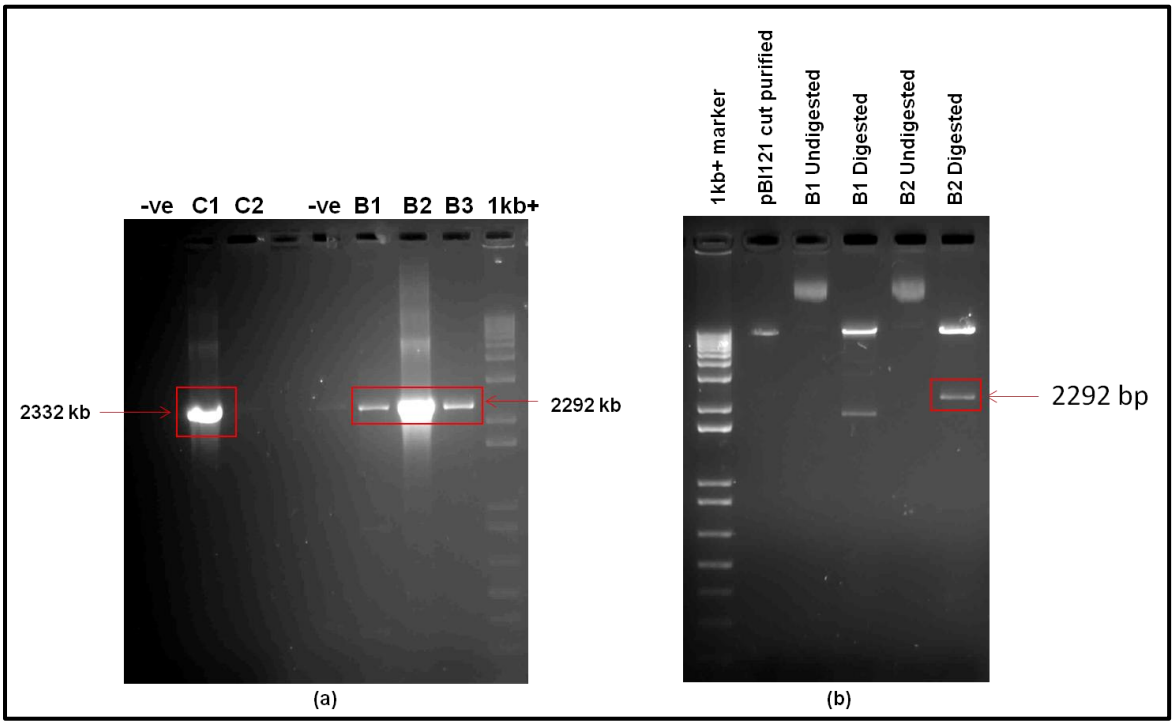


Figure 4.9 Confirmation of cloning of *PpPAL1A* and *PpPAL1B* genes into the binary vectors (pCAMBIA1300 and pBI121). (a) PCR amplification using gene-specific primers to confirm clones. C1 and C2 represent clones for 35S:: *PpPAL1A* while B1, B2, B3 represents clones for 35S:: *PpPAL1B.* The negative control contains no template DNA as template. (b) Restriction digestion of 35S:: *PpPAL1B* clones (B1 and B2) using *BamHI* and *SacI* restriction enzymes to confirm cloning event.

Clones C1 and B2 were confirmed by sequencing using gene specific primers (shown in Table 4.3). Further C1 and B2 were transformed and cloned in *Agrobacterium tumefaciens* strain C58 GV2260. PCR confirmation of these *Agrobacterium* clones is shown in Figure 4.10.

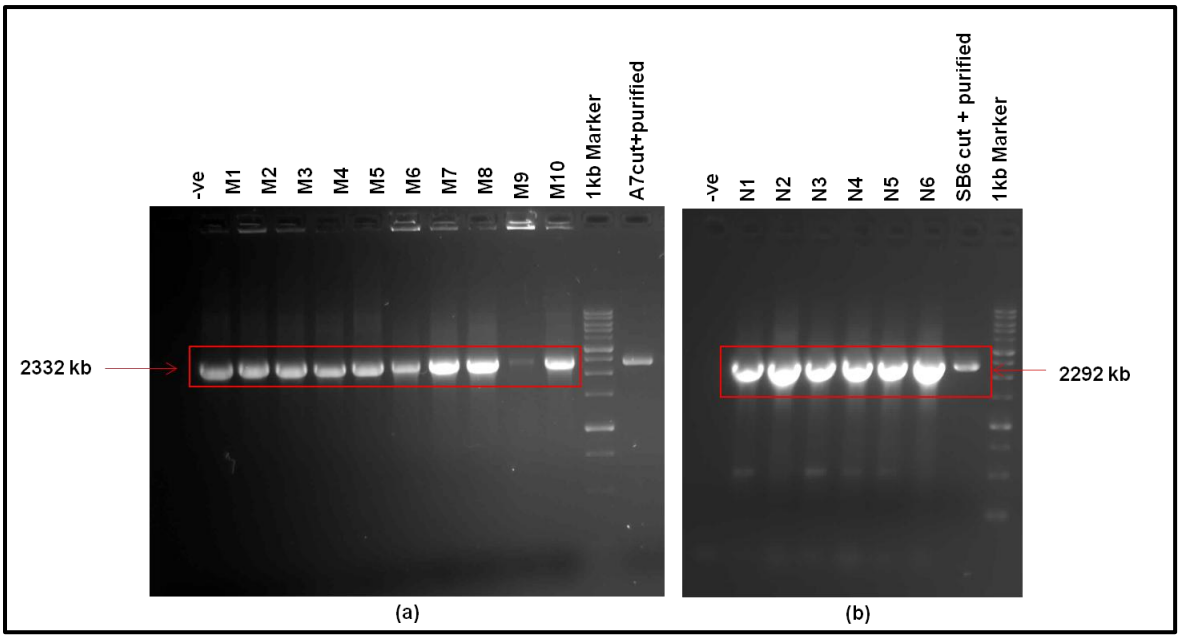


Figure 4.10 Confirmation of clones of full-length *PpPAL1A* and *PpPAL1B* gene transformed into *Agrobacterium tumefaciens* strain C58 GV2260. (a) Clone confirmation by RT-PCR of 35S:: *PpPAL1A* clones (M1 to M10); (b) Clone confirmation by RT-PCR of 35S:: *PpPAL1B* clones (N1 to N6).

Clone	Vector	Sequencing from primer-	#Chromosome	Score (bits)	E-value	Identity	Frame
A7	pGEM-T	T7 promoter	1	588	0	99%	+1/-1
A7	pGEM-T	SP6 promoter	1	807	0	99%	+1/+1
A7	pGEM-T	Pp_PAL1-A_q5'F	1	928	0	99%	+1/-1
A7	pGEM-T	Pp_PAL1-A_q3'R	1	954	0	99%	+1/+1
SB6	pGEM-T	T7 promoter	2	852	0	97%	+1/+1
SB6	pGEM-T	SP6 promoter	2	1047	0	98%	+1/-1
C1	pCAMBIA1300	SalI_PAL1-A_FL_5'F	1	570	0	97%	+1/-1
C1	pCAMBIA1300	KpnI_PAL1-A_FL_3'R	1	650	0	97%	+1/+1
C1	pCAMBIA1300	Pp_PAL1-A_q5'F	1	1096	0	99%	+1/-1
C1	pCAMBIA1300	Pp_PAL1-A_q3'R	1	948	0	99%	+1/+1
B2	pBI121	BamHI_PAL1-B_FL_5'F	2	819	0	98%	+1/-1
B2	pBI121	SacI_PAL1-B_FL_3'R	2	413	0	99%	+1/+1

Table 4.3 List of blast results and sequence analysis for the confirmation of the constructs used for overexpression of *PAL1* genes in the present study.

b. To generate overexpression transgenic lines of 35S:: *PpPAL1A* and 35S:: *PpPAL1B* in *P. patens* by *Agrobacterium-mediated* transformation

The confirmed overexpression constructs 35S:: *PpPAL1A* (clone M1) and *35S:: PpPAL1B* (clone N2) transformed into *Agrobacterium tumefaciens* strain C58 GV2260 using rifampicin and kanamycin as selection markers were inoculated, and the *Agrobacterium-mediated* transformation was performed in moss. The inoculum was co-cultured with moss five-days-old homogenised protonemata tissue. Further, plating and relaxation (for 2-weeks) were carried out between two selection events (for 2-weeks). Six independent events were performed to generate desired mutant lines.

Mutant lines surviving after selection-II were observed under Leica microscope for their phenotypic characterization at regular intervals (Figure 4.11).

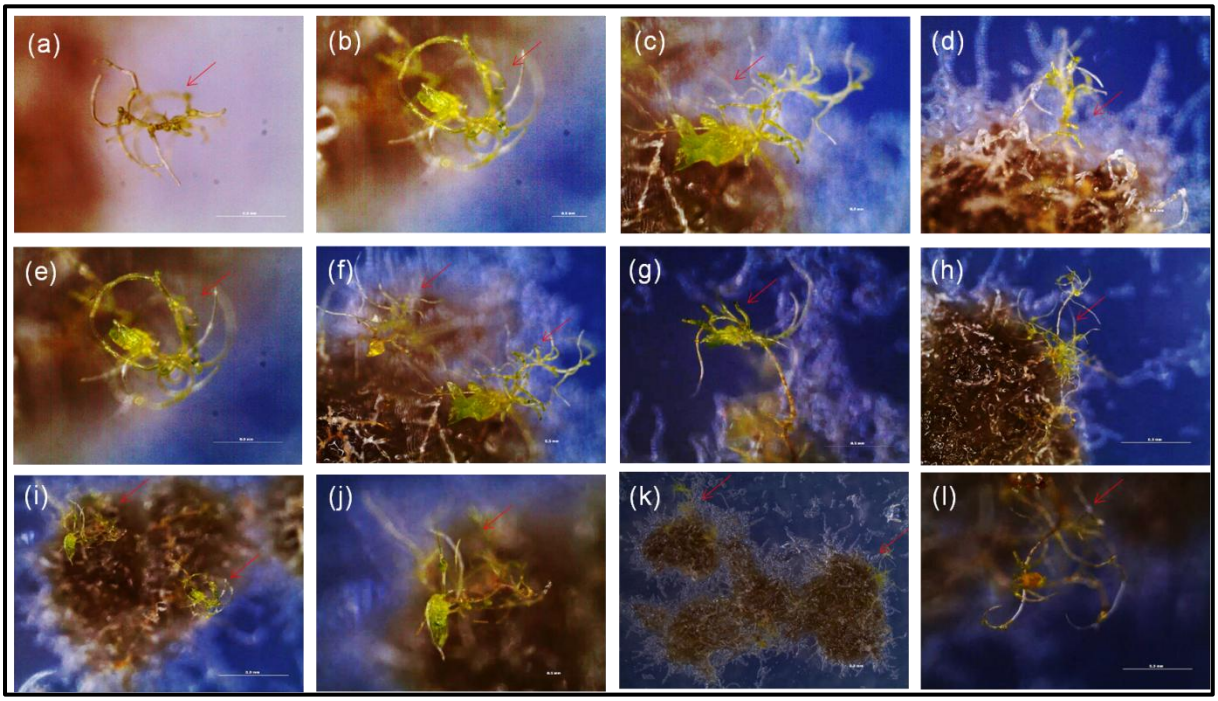


Figure 4.11 Representative micrographs of transgenic 35S:: *PpPAL1A* and 35S:: *PpPAL1B* overexpression lines survived after three antibiotic selection events. Red marked arrow in (a to l) represents the putative mutant lines emerging from the transformation events.

4.4 Generation of knockout constructs by targeted mutagenesis

a. Construction of knockout clones of *PpΔpal1a* and *PpΔpal1b* by homologous recombination along with *GUS* promoter characterization in *P. patens*

Along with homologous recombination, we have used PCR fusion method that allowed us to characterise GUS promoter along with the knockout study of *PpΔpal1a* and *PpΔpal1b.* For *PpΔpal1a,* using P13 & P14 primers, 1072 bp of 5' flanking region was amplified from gDNA further; P15 & P16 primers were used to amplify 2135 bp fragment of *GUS* ORF from pBI121 plasmid vector. P13 & P16 primers were used to amplify 3207 bp fragment, a fusion of 5' flanking region and *GUS* ORF from previously amplified 5' flanking region PCR product and *GUS* gene amplified PCR product. 3' flanking region of 1161 bp was amplified from gDNA using P17 & P18 primers (Figure 4.12 (a)).

Similarly, for *PpΔpal1b*, 1233 bp of 5' flanking region was amplified from gDNA using P19 &P20 primers; P21 & P22 were used to amplify 2148 bp of *GUS* ORF from pBI121 plasmid vector. P19 & P22 primers were used to amplify 3381 bp fragment of fusion of 5' flanking region and *GUS*. 987 bp of 3' flanking region was amplified from gDNA using P23 & P24 primers (Figure 4.12 (b)).

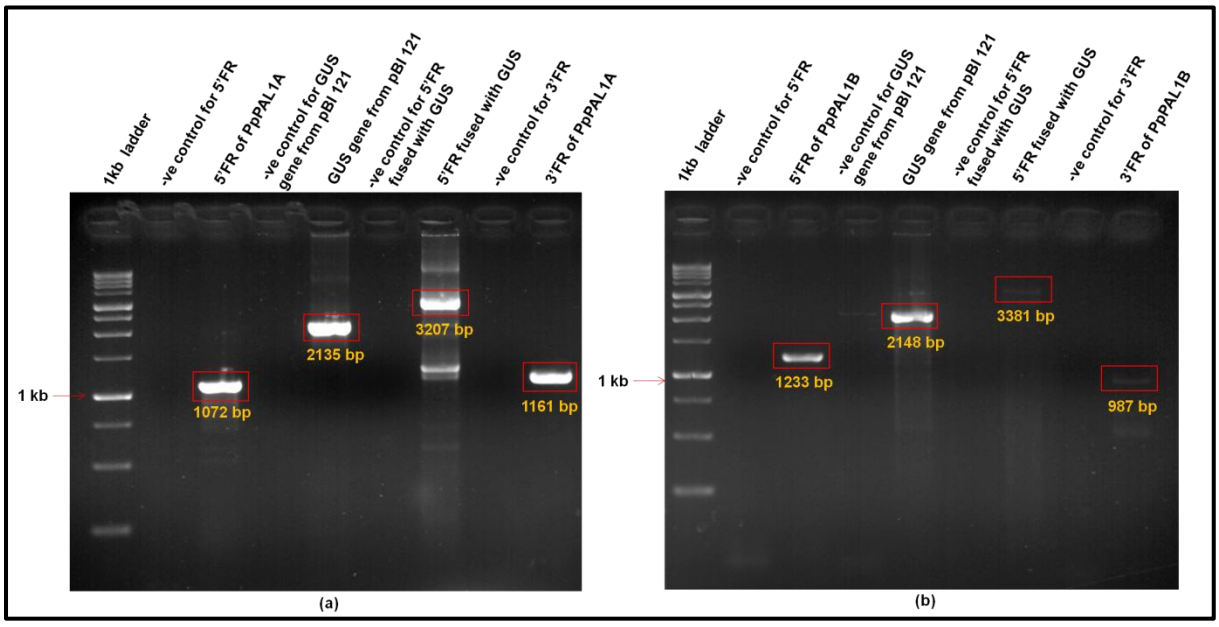


Figure 4.12 Sequential PCR amplification of 5' flanking region, *GUS* ORF, 5'+*GUS* ORF fusion product and 3' flanking region for cloning of *PpΔpal1a* and

***PpΔpal1b* to generate knockout constructs**. (a) PCR amplification of 5' flanking region, *GUS* gene, 5' flanking region +*GUS* gene fusion and 3' flanking regions for knockout cloning of *PpΔpal1a*; (b) PCR amplification of 5' flanking region, *GUS* gene, 5' flanking region +*GUS* gene fusion and 3' flanking regions for knockout cloning of *PpΔpal1b.*

For *PpΔpal1b* cloning, the amplified 5' flanking region fused with *GUS* was subcloned into a pGEM-T vector and confirmed by colony PCR. Clones X1 and X8 were confirmed by sequencing from T7, and SP6 promoter ends. Similarly, the 3' flanking region amplified was subcloned into a pGEM-T vector and confirmed by colony PCR and restriction digestion. Clones Y6 and Y8 confirmed by sequencing from T7 and SP6 ends. Further cloning into pTN186 is under progress.

Similarly, for *PpΔpal1a* cloning, the amplified 5' flanking region fused with *GUS* was subcloned into the pGEM-T vector. Confirmation of insertion into the generated clones will be done by PCR and restriction digestion. Also, amplified 3' flanking region was subcloned into pGEM-T vector and confirmation by colony PCR, and restriction digestion is in progress.

Chapter 5

Conclusion

In this study, we have identified two *PAT1*-like genes from the genome sequences of *P. patens* and named them as – *PpPAL1A* and *PpPAL1B*. A comprehensive analysis of both the genes, including gene structure, motif search, phylogeny, tissue-specific expression profile, targeted mutagenesis, were performed. Our sequence alignment showed that both the PAT1-like proteins PpPAL1A and PpPAL1B belong to GRAS proteins. The high degree of amino acid sequence homology was identified among all five motifs (LRI-VHIID-LRII-PFYRE-SAW). These amino acid residues were conserved in the C-terminal part of GRAS proteins across five different plant species like *Physcomitrella patens*, *Arabidopsis thaliana*, *Vitis vinifera, Oryza sativa* and *Citrus medica var. sarcodactylis*. Around 24-40% identity was seen between PpPAT1-like proteins and other GRAS proteins. However, within same species, PpPAL1A and PpPAL1B showed 82.95% identity in *P. patens* at the level of amino acids.

Using the available microarray data, *in silico* expression analysis was carried out for both *PpPAL1A* and *PpPAL1B* genes using Phypa_98188 as a primary query. Interestingly, *PpPAL1A* showed constitutive expression in all moss tissue types while *PpPAL1B* was observed to be highly concentrated in the sporophytic stage and least expressed in the initial protonemal stage. Further, semi-quantitative RT-PCR was performed with three independent biological replicates, to study the tissue-specific expression analysis of both *PpPAL1A* and *PpPAL1B* genes in protonema and gametophore tissue using *ACT* and *E2* as reference genes. Our tissue-specific gene expression analysis supported the *in silico* data, where both genes showed relatively higher expression levels in gametophore compared to protonema. Also, the relative expression levels of both *PpPAL1A* and *PpPAL1B* showed statistical significance when normalised to *ACT*.

The full-length coding sequences for both the genes were amplified and subcloned into the pGEM-T vector. The clones for *PpPAL1A* and *PpPAL1B* were sequence confirmed from both T7 and SP6 ends. Further, *PpPAL1A* was successfully cloned into pCAMBIA1300, and *PpPAL1B* was cloned into pBI121

binary vectors. These confirmed clones were transformed in moss by *Agrobacterium-mediated* transformation method to generate overexpression lines. The putative mutant lines are coming up and will be further characterised.

For the construction of knockout vectors- *PpΔpal1a* and *PpΔpal1b,* sequential 5' flanking region, *GUS* ORF, 5'+*GUS* ORF fusion product and 3' flanking region were PCR amplified. The amplified 5' flanking region fused with *GUS* ORF and 3' flanking region were subcloned into pGEM-T vector independently. However, the construction of *PpΔpal1a* and *PpΔpal1b* vectors and generation of its knockout lines is currently in progress.

Salient features of this study:

i. Using the sequence alignment and phylogenetic analysis, we identified two *PAT1*-like genes in moss -*PpPAL1A* and *PpPAL1B*.
ii. In silico expression analysis showed that PpPAL1A is constitutively expressed in all tissue while PpPAL1B is highly expressed in the sporophytic phase.
iii. Validation of in silico expression data by semi-quantitative RT-PCR showed that PpPAL1A is highly expressed in gametophore compared to protonema (1.79 relative fold change). Similarly, PpPAL1B expression level is higher in gametophore compared to protonema (1.49 relative fold change). Also, both genes showed significant statistical change.
iv. The full-length coding sequences of *PpPAL1A* and *PpPAL1B* were successfully cloned in pCAMBIA1300 and pBI121 binary vectors respectively. Also, generation of overexpression lines was carried by *Agrobacterium-mediated* transformation, and putative lines are coming up that will be further characterised.
v. Construction of knockout vectors of *PpPAL1A* and *PpPAL1B* is currently in progress.
vi. Generation of knockout lines and characterization of all the transgenic lines would be our future interests to understand the function of these two genes in moss.

References

Bolle, C. (2004). The role of GRAS proteins in plant signal transduction and development. Planta *218*, 683–692.

Bolle, C., Koncz, C., and Chua, N.H. (2000). PAT1, a new member of the GRAS family, is involved in phytochrome A signal transduction. Genes Dev. *14*, 1269–1278.

Engstrom, E.M. (2011). Phylogenetic analysis of GRAS proteins from moss, lycophyte and vascular plant lineages reveals that GRAS genes arose and underwent substantial diversification in the ancestral lineage common to bryophytes and vascular plants. Plant Signal Behav *6*, 850–854.

Harrison, C.J., Roeder, A.H.K., Meyerowitz, E.M., and Langdale, J.A. (2009). Local Cues and Asymmetric Cell Divisions Underpin Body Plan Transitions in the Moss Physcomitrella patens. Curr. Biol. *19*, 461–471.

Kofuji, R., and Hasebe, M. (2014). Eight types of stem cells in the life cycle of the moss Physcomitrella patens. Curr. Opin. Plant Biol. *17*, 13–21.

Li, L., Stoeckert, C.J.J., and Roos, D.S. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes -- Li et al. 13 (9): 2178 -- Genome Research. Genome Res. *13*, 2178–2189.

Peng, C., P., R., D. E., K., and K. E., C. (1997). The *Arabidopsis GAI* gene defines a signalling pathway that negatively regulates gibberellin responses. Genes Dev *11*, 3194–3205.

Roberts, A.W., Roberts, E.M., and Haigler, C.H. (2012). Moss cell walls: structure and biosynthesis. Front. Plant Sci. *3*, 1–7.

Russo, V.E.A., Brody, S., Cove, D., Ottolenghi, S., and Cove, D.J. (1992). Chapter 12: Regulation of Development in the Moss, Physcomitrella Patens. 179–193.

Sun, X., Jones, W.T., and Rikkerink, E.H. a (2012). GRAS proteins: the versatile roles of intrinsically disordered proteins in plant signalling. Biochem. J. *442*, 1–12.

Torres-Galea, P., Huang, L.F., Chua, N.H., and Bolle, C. (2006). The GRAS protein SCL13 is a positive regulator of phytochrome-dependent red light signaling, but can

also modulate phytochrome A responses. Mol. Genet. Genomics *276*, 13–30.

Wang, Y., Shi, S., Zhou, Y., Zhou, Y., Yang, J., and Tang, X. (2016). Genome-wide identification and characterization of GRAS transcription factors in sacred lotus (*Nelumbo nucifera*). PeerJ *4*, e2388.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., and Provart, N.J. (2007). An “electronic fluorescent pictograph” Browser for exploring and analyzing large-scale biological data sets. PLoS One *2*, 1–12.

Yuan, Y., Fang, L., Karungo, S.K., Zhang, L., Gao, Y., Li, S., and Xin, H. (2016). Overexpression of VaPAT1, a GRAS transcription factor from Vitis amurensis, confers abiotic stress tolerance in Arabidopsis. Plant Cell Rep. *35*, 655–666.