Functional characterisation of cis-acting regulatory elements in the promoters of MSL family genes from the moss *Physcomitrium patens*

A Thesis submitted to Indian Institute of Science Education and Research Pune
in partial fulfilment of the requirements for the BS-MS Dual Degree Programme
by

Sanskar Nitinkumar Agrawal

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April 2023

Supervisor: Prof. Anjan K. Banerjee
Professor, Department of Biology, IISER Pune
From May 2022 to Mar 2023

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान पुणे
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE
डॉ. होमी भाभा मार्ग, पुणे 411008, महाराष्ट्र, भारत I Dr. Homi Bhabha Road, Pune 411008, Maharashtra, India
T +91 20 2590 8001 W www.iiserpune.ac.in

IISER PUNE

Certificate

This is to certify that this dissertation entitled "**Functional characterisation of cis-acting regulatory elements in the promoters of MSL family genes from the moss *Physcomitrium patens***" towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study carried out by "**Sanskar Nitinkumar Agrawal at Indian Institute of Science Education and Research, Pune**" under the supervision of "**Prof. Anjan K. Banerjee, Professor, Department of Biology**" during the academic year 2022-2023. The results obtained from this study should not be used for any reference or publication purposes without the written consent of the supervisor.

4/5/2023

Signature of the Supervisor
(Prof. Anjan K. Banerjee)

प्रो. अंजन के. बैनर्जी/Prof. Anjan K. Banerjee
प्राध्यापक, जीवशास्त्र कार्यक्रम
Professor, Biology Program
भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान
Indian Institute of Science Education & Research
पुणे, भारत /Pune- 411 008, India

Date: 4/05/2023

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान पुणे
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE
डॉ. होमी भाभा मार्ग, पुणे 411008, महाराष्ट्र, भारत | Dr. Homi Bhabha Road, Pune 411008, Maharashtra, India
T +91 20 2590 8001 W www.iiserpune.ac.in

IISER PUNE

Declaration

I hereby declare that the matter embodied in the report entitled "**Functional characterisation of cis-acting regulatory elements in the promoters of MSL family genes from the moss *Physcomitrium patens***" should appear here are the results of the work carried out by me at the "**Department of Biology, Indian Institute of Science Education and Research, Pune**", under the supervision of "**Prof. Anjan K. Banerjee**" and the same has not been submitted elsewhere for any other degree.

Signature of the student
(Sanskar Nitinkumar Agrawal)

Date: 4/05/2023

I would like to extend my heartfelt appreciation and gratitude to my family, whose unwavering love and support have been my driving force throughout my academic journey.

To my parents, who have instilled in me the values of perseverance, hard work, and determination, and who have always believed in my abilities.

To my friends, who have provided me with a much-needed respite from my studies and have always been there to lift my spirits.

I would also like to express my sincere thanks to my thesis advisor, who has provided me with invaluable guidance, support, and patience throughout my research.

To the people who generously contributed their time and insights to my study, I am truly grateful.

Finally, I would like to extend my gratitude to all those who have touched my life in one way or another.

This thesis is dedicated to each and every one of you.

Last but not least, I want to thank me, for believing in me, for doing all this hard work, for having no days off, for never quitting, for always being a giver, for doing more right than wrong, and for just always being Me all the time. And I would like to thank the man up top: GOD, cause I am nothing without him.

Table of contents

List of Figures

List of Tables

Abstract

All organisms use mechanosensation to sense and respond to their immediate environment by translating physical stimulus to biochemical information. Plants being sessile in nature, rely heavily on mechanosensation to enable them to adapt and survive in a variety of environmental conditions. Mechanosensation is mostly mediated by a set of membrane channel proteins called Mechanosensitive (MS) ion channels. One of the most well-characterised MS ion channels is the Mechano-sensitive channel small conductance (MscS), originally reported in *E. coli*. The homologs of MscS-like channels (MSLs) have been found in all plant lineages and are extensively studied in angiosperms like *Arabidopsis thaliana*, *Oryza sativa* etc. Recently, the study of such developmentally crucial genes in extant land plants has gained popularity due to their significance in the course of land plant evolution. *Physcomitrium patens* is one such land plant and an excellent model organism that has a simple life cycle and a highly efficient homologous recombination system. Comparative studies of the molecular processes in *P. patens* can help establish an evolutionary relationship between MSLs of vascular and non-vascular plants. Our *in silico* studies reveal that *P. patens* has 16 MSL genes, of which 4 MSLs were differentially regulated under dehydration stress. The differential gene expression of these MSLs was validated by RT-qPCR analysis, suggesting a possible regulatory switch particularly activated during the dehydration stress. *In-silico* analyses of the four *PpMSL* promoters revealed that they house key dehydration-responsive regulatory elements. Based on these preliminary analyses, two *PpMSLs* were selected for further studies. To analyse the physiological function of the dehydration-responsive regulatory motifs through reporter fusion assays, we generated independent vector constructs having full-length and truncated promoters of *PpMSL3* and *PpMSL15* upstream to the GUS reporter gene. Stable transgenic lines are being generated in moss using these vectors to explore the gene expression dynamics under dehydration stress. Quantification of relative GUS activity in these transgenic lines will reveal the minimal functional promoters for the PpMSL genes and the regulatory mechanism of gene expression by the various motifs under dehydration-rehydration stress conditions.

Acknowledgements

I would like to extend my sincere thanks to my thesis advisor, Prof. Anjan K. Banerjee, for providing me with invaluable guidance, support, and encouragement throughout my research journey. Without his expert insights and thoughtful feedback, my thesis would not have been possible.

I am also deeply grateful to my committee member, Dr Kalika Prasad, for providing his expertise and constructive feedback. His insightful comments and suggestions have helped me refine my research and shape this thesis into its final form.

I am extremely thankful to Dr Amey Bhide and Ms Shirsa Palit for constantly supervising me and helping me at every step of the thesis. I am grateful to Ms Nikita Patil, Ms Nilam Malankar and Mr Gourav Arora for their support and encouragement throughout my research. Their feedback and suggestions have helped me to refine my methodology and analyze my results. I would also like to thank Mr Kishan Saha, Ms Shikha Singh, Mr Mohit Mantri, Ms Kanishka Rajoria, Ms Debasmita Pal and all other AKB lab members for their constant support and love.

I would like to appreciate the support and encouragement extended by the faculty and staff of the Department of Biology at IISER Pune. Their dedication to teaching and research has been an inspiration to me.

Lastly, I am immensely grateful to my family and friends for their unwavering support and encouragement throughout my graduate studies. Their love, patience, and understanding have been a constant source of motivation and strength.

I extend my heartfelt appreciation to everyone who has contributed to this work.

Contributions

Contributor Name	**Contributor role**
AKB, AB	Conceptualization Ideas
SNA, AB	Methodology
SNA	Software
SNA	Validation
SNA	Formal analysis
SNA, AB	Investigation
AKB	Resources
SNA, AB	Data Curation
SNA	Writing - original draft preparation
SNA, AB, AKB	Writing - review and editing
SNA	Visualization
AKB, AB	Supervision
AKB, AB, SNA	Project administration
AKB	Funding acquisition

AKB: Prof. Anjan K. Banerjee

AB: Dr Amey J. Bhide

SNA: Sanskar Nitinkumar Agrawal

Chapter 1: Introduction

1.1. *Physcomitrium patens* ecotype Gransden: an excellent model plant to study plant evolution, development, and physiology

Physcomitrium patens, also known as the spreading earth moss, belongs to the Funariaceae family of the division Bryophyta. Funariales (**Figure 1**) is a diversified lineage of small, subapically branching plants with terminal sex organs, a sporophyte with a capsule dehiscing by lid loss and two opposing pairs of peristome teeth along the capsule mouth. The peristome and a distinct line of sporangial dehiscence are absent from *Physcomitrium*. (Beike *et al.*, 2014; Medina *et al.*, 2019). It colonises muddy areas near the shores of waterbodies and is mostly found in temperate regions worldwide.

P. patens Gransden was collected by H. Whitehouse from Gransden Wood and has been used as a potential model organism since then. It has been extensively used to study plant evolution, development, and physiology (Lang *et al.*, 2008). It has been used as a model organism for more than 80 years. Studying molecular processes can help establish relationships between vascular and non-vascular plants (Nishiyama *et al.*, 2003).

P. patens life cycle shows alternation of generation like all other mosses. Both the diploid sporophyte and haploid gametophyte are multicellular, with the haploid gametophyte being the dominant generation. The sporophytic generation is monosporangiate (Cove, 2005).

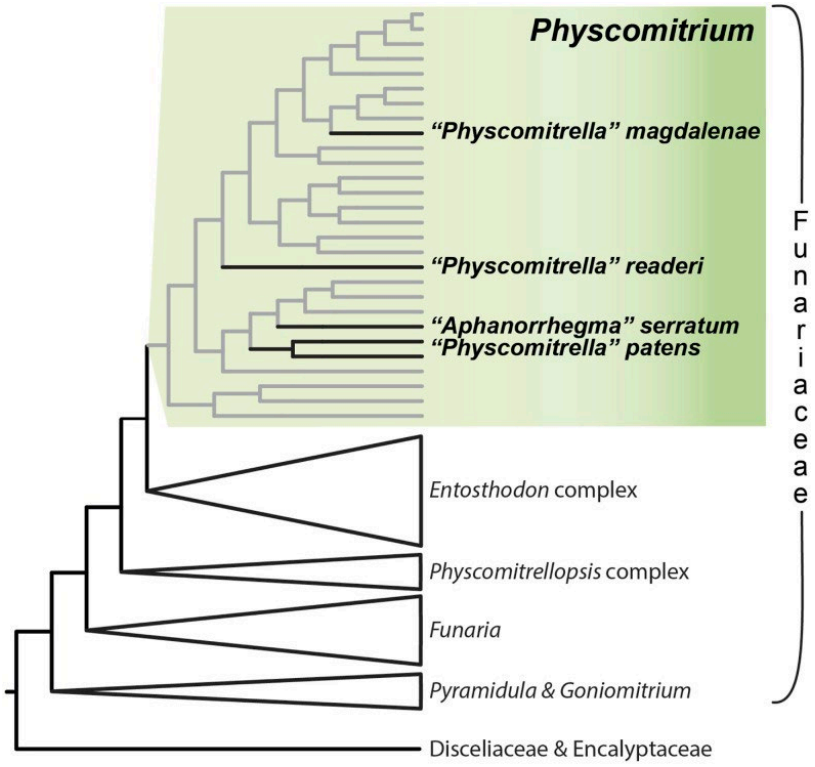


Figure 1: Representative phylogenetic tree of moss family Funariaceae*. The phylogenetic tree was constructed to illustrate the polyphyletic distribution from diversity in 648 target capture nuclear exons within Physcomitrium using the maximum likelihood method. (Figure adapted from Rensing et al., 2020).*

1.2. *Characteristics of Physcomitrium patens* as a model organism

Physcomitrium patens is the most studied and widely used moss model system (Cove' and Knight, 1993). It was the first non-seed plant whose genome had been sequenced (Rensing *et al.*, 2020). Some features that make it an excellent model organism include: its small size, relatively short life span, and easily accessible and fully annotated genome. *P. patens* has a dominant haploid gametophyte phase, making it advantageous for creating gene knock-out lines and other genetic studies compared to diploid models. It is the only plant model system known for its highly efficient gene targeting through homologous recombination (Cove *et al.*, 2009), allowing for a unique approach to studying plant gene functions. It is easy to culture in Petri plates due to its small size. Its relatively simple development and few tissues with a limited number of cells help to track single-cell development easily. Approximately one-quarter of its genome is uncategorised,

raising the likelihood of successful discovery efforts (Rensing *et al.*, 2020). Although it lacks vascular tissue, it has many signalling pathways similar to that in angiosperms. However, there is still a significant challenge in fully understanding the roles of genes in this plant species.

1.3. Morphology and Anatomy

Mosses undergo a haplodiplontic life cycle, which is the same as that of all terrestrial plants (**Figure 2**) that encompasses the multicellular haploid gametophyte and diploid sporophyte generations. However, unlike current polysporangiophytes, mosses have a dominant haploid phase, while the sporophytic phase is characterised by a single terminal capsule. The juvenile stage of the gametophyte produces protonema, which are tip-growing filaments that branch out and spread by apical extension once the spores have germinated. Chloronema and caulonema are the two main cell types that make up these filaments. The first cell type to form inside a germinating spore is called a chloronema, and it has transverse cell plates and a huge number of chloroplasts. The apical cell of the chloronema divides continuously and transforms into a caulonemal cell after a few days, with thinner walls, fewer chloroplasts, cell plates positioned obliquely, and faster growth compared to chloronema.

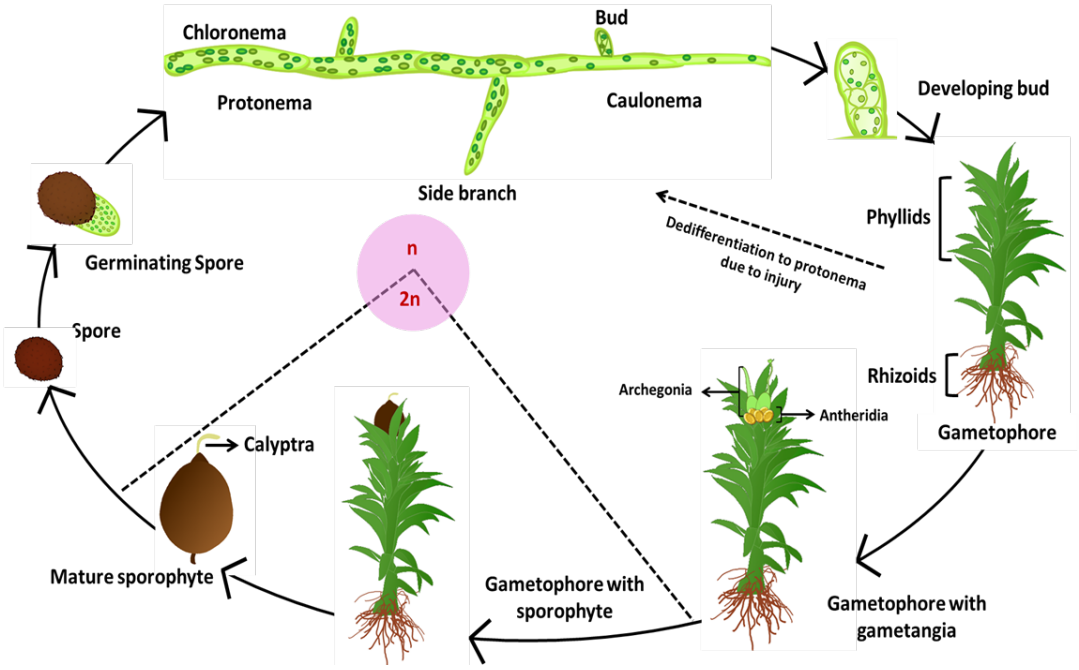


Figure 2: Life cycle of Physcomitium patens. *The figure displays the life cycle of Physcomitrium patens, a moss that undergoes an alternation of generations. In this life cycle, the haploid gametophyte (n) is the predominant generation and produces both male and female gametes. Male gametes are produced by antheridia, whereas female gametes are produced by archegonia. Fertilisation takes place when sperm from an antheridium fertilises an egg in an archegonium, resulting in a diploid zygote. The zygote matures into a sporophyte (2n), which relies on the gametophyte for sustenance. The sporophyte is made up of a foot, a seta, and a capsule, the latter of which contains spores that, when the capsule breaks, release their contents and produce new gametophytes. This life cycle provides genetic variation and adaptation as the haploid gametophyte phase allows for genetic diversity through sexual reproduction, while the diploid sporophyte stage enables the creation of genetic combinations that can be beneficial for survival. n and 2n represent the haploid and diploid generations which have single and paired chromosomes, respectively. (Figure credit: Shirsa Palit).*

The buds of the plant progress into upright, leafy gametophores, which are made up of a stem and phyllids. Similar to the development of vascular plants, the formation of specialised cells for water and nutrient transportation in these structures is controlled by NAC transcription factors (Xu *et al.*, 2014). The inner cells develop into a core strand of

elongated, narrow, and thin-walled cells, whereas the stem is normally short and does not branch (W *et al.*, 2019). The gametophores have only a few leaves, which are arranged in a spiral pattern and spread out when moist.

The gametophores grow gametangia, which generate both male (antheridia) and female (archegonia) gametes in a single cluster at the apex of mature gametophores(Landberg *et al.*; Hiss *et al.*, 2017) when exposed to short days (8h light: 16h dark) (Hohe *et al.*, 2002). The male and female gametangia are divided into apical clusters, indicating that these organisms are bisexual. Under ideal development conditions, both male and female sex organs are terminal; nevertheless, when vegetative growth starts after male gametangiogenesis, female sex organs may appear to be present in a separate cluster.

Moss male gametes swim toward the female archegonia with the help of two flagella in the presence of liquid water (Renzaglia and Garbary, 2001). The sperm cells fertilise the egg cell once they have passed through the archegonial neck canal cells and have reached the archegonial venter, resulting in the formation of a diploid zygote. This zygote subsequently transforms into an embryo, which then matures into a sporophyte with a single apical capsule, where meiosis occurs and haploid spores are produced (Landberg *et al.*; Hiss *et al.*, 2017). On a single gametophore, just one sporophyte typically grows. The spores are partly covered by the substance sporopollenin, which is also found in the male gametophytic pollen of seed plants (Daku *et al.*, 2016).

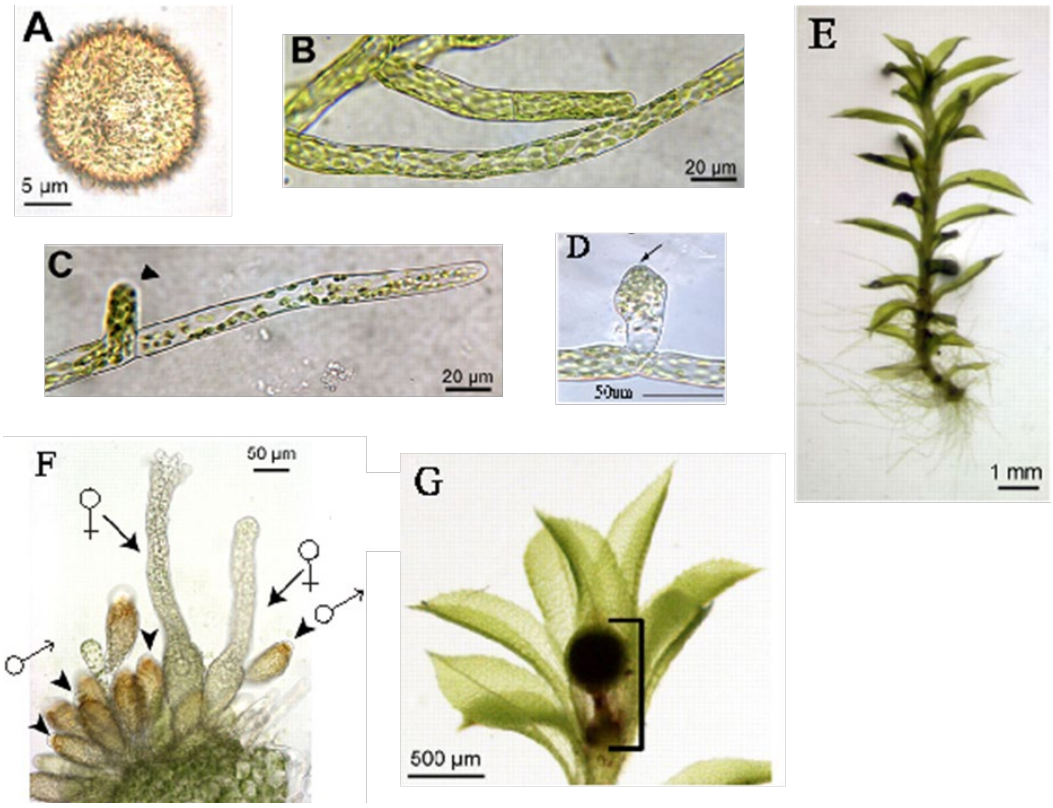


Figure 3: Different stages of moss development. *(A) Haploid Spore, (B) Chloronema filaments in protonema, (C) Caulonema filaments in protonema, (D) Developing bud, (E) Gametophore, (F) Male and Female sex organs, (G) Sporophyte. (Figure adapted from Prigge and Bezanilla, 2010).*

1.4. Bacterial *MscS* and *MscS*-like proteins in plants

Mechanosensation is the phenomenon by which animals and plants are able to detect mechanical forces like touch, sound, temperature, body movements, changes in cell shape and size etc., by converting these into electrochemical signals (García-Añoveros and Corey, 1997). This signalling is mediated by a set of membrane channels called the Mechanosensitive (MS) ion channels. These are intriguing proteins functioning as both sensors and effectors. They are embedded in membranes and transform mechanical stimuli into electrical or biochemical signals, resulting in the regulation of a diverse repertory of cellular processes that allow adaptive response (Peyronnet *et al.*, 2014). When an MS ion channel is activated, it causes a substantial but temporary flow of ions across a membrane, resulting in fast changes in cell volume, increased intracellular levels

of secondary messengers like Ca2+, or the generation of an electrical current (Kung and Blount, 2004).

MscS (Mechano-sensitive channel, Small conductance) from *Escherichia coli* is one of the most well-studied MS ion channels in any system. It is a non-selective ion channel that is activated immediately by membrane tension (Martinac *et al.*, 1987). In the event of severe hypo-osmotic down shock, the cell is kept from rupturing by the YggB, an EcMscS translation product (Levina *et al.*, 1999).

MscS-like channels, also known as MSLs, are found in bacteria, archaea, certain fungi, algae, and plants: *Arabidopsis*, papaya, rice, and common bean (**Figure 4**). MSL gene families have been identified and characterised to varying degrees (Basu and Haswell 2017). Plant MSLs have been shown to localise to the membranes of various cellular compartments. In the model plant *Arabidopsis thaliana*, a total of 10 MSLs have been fully characterised, and they are engaged in a number of physiological and developmental processes (Hamilton *et al.*, 2015).

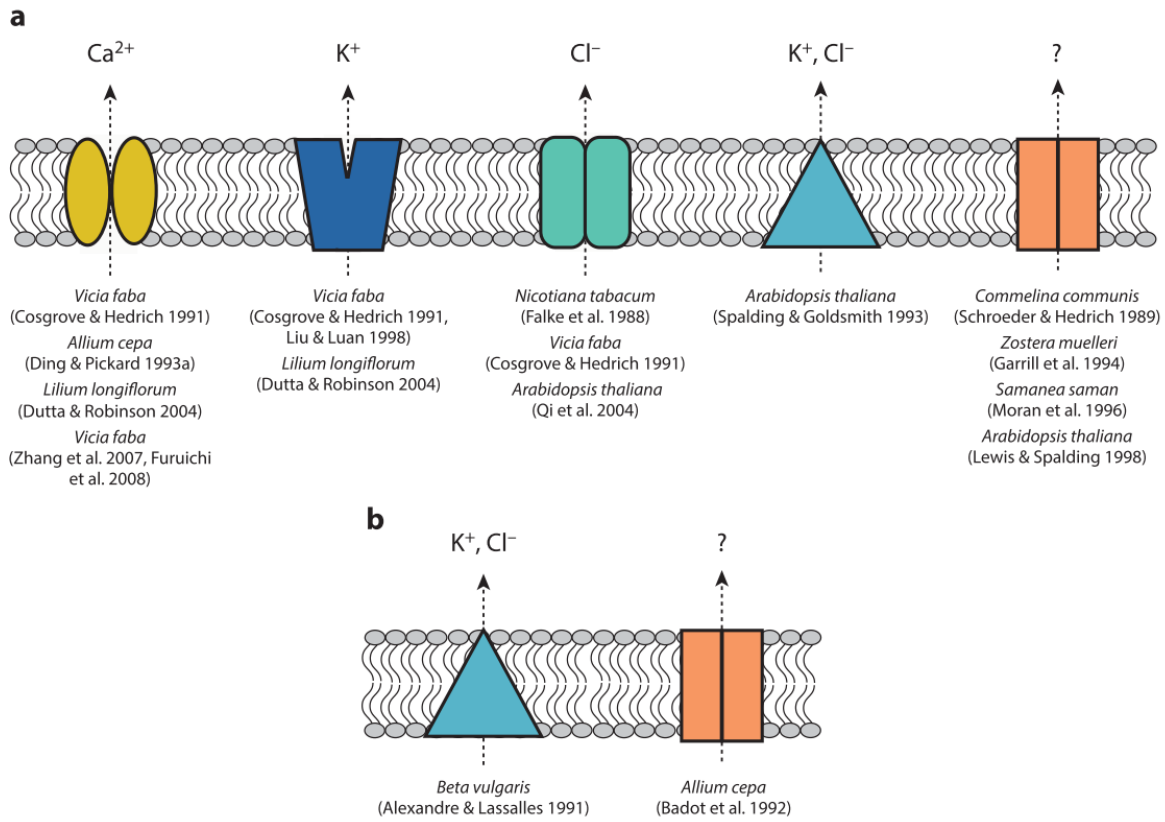


Figure 4: Uncharacterised MS ion channels in various plant membranes. *The figure shows various MS ion channels in different plant species localised to (a) Plasma*

membrane and (b) Vacuolar membrane. Underneath each channel, pertinent references are noted. Arrows show the direction of ion flow into or out of the cell or vacuole but not the ion permeability. (Figure adapted from Hamilton, Schlegel and Haswell, 2015).

Chapter 2: Motivation and Objectives

2.1. Motivation

One essential process that plants employ to detect diverse mechanical stimuli in their surroundings is mechanosensation. These include stimuli like touch, temperature, gravity, and turgor pressure, all of which have an impact on how plants grow and develop. Various adaptations would have been required for plants to survive the extremely harsh environment prevalent on the land as they evolved and moved from the water to the land. Mechanosensitive ion channel proteins could have helped plants to be detected and respond to such harsh environmental conditions. These fascinating proteins serve as both effectors and sensors. They are integrated into membranes and convert mechanical inputs into electrical or biochemical signals, which control a variety of cellular processes and enable adaptive response (Peyronnet *et al.*, 2014).

MSLs have been studied in various plants like *Arabidopsis*, Maize, Papaya etc., but very few MSLs have been molecularly characterised. There are even fewer studies on the promoters of MSLs, which raises the question of how the expression of these genes is regulated on the chromatin level.

Although the structure and function of MSL proteins have been well characterised in *Arabidopsis*, the function of MSL homologs in bryophytes is not yet known. It is interesting to study the function of these homologs in moss from a land-plant evolutionary viewpoint. *Physcomitrium patens* is a member of the Funariaceae family of the division Bryophyta and an excellent model organism. It has been extensively used in our lab to conduct developmental and evolutionary studies.

We have predicted 16 MSL homologs in *P. patens* using the available whole genome sequencing database. Interestingly, we have noted that four *P. patens* MSL genes (*PpMSL*3, 7, 13, and 15) have distinctive expression patterns specific to the dehydration and rehydration condition. Here, using domain deletions and reporter fusion assays, we proposed to study promoters of *PpMSL15* and *PpMSL3* to delineate the dehydration and rehydration stress-responsive regulatory mechanism in the moss *P.patens*.

We will be using various molecular biology approaches like bioinformatic studies, expression analysis, cloning of promotor-reporter vector constructs, stable transgenic line

generation in moss and detection of reporter gene expression to validate and decipher the molecular mechanism leading to the expressional switch.

2.2. Objectives

We proposed the following objectives to characterise the function of cis-acting regulatory elements in selected *PpMSL* (MSL3 and MSL15) promoters responsible for conditional gene expression under dehydration-rehydration stress:

1. Cloning of full-length and sequentially deleted *PpMSL3* and *PpMSL15* promoters upstream to the reporter gene GUS.
2. Transgenic line generation in the moss with sequence confirmed promoter-reporter constructs.
3. Scoring of reporter gene (GUS) expression and validation of conditional gene regulation under dehydration-rehydration stress in stable moss lines.

Chapter 3: Methodology

3.1. Data Collection and Analysis

All target gene sequences were retrieved from the **oneKP Project** (Matasci *et al.*, 2014). Gene ID for all ten *Arabidopsis thaliana* MSL genes was collected (Hamilton *et al.*, 2015). Amino acid sequences of all *AtMSLs* were collected and blasted against the genome of *P. patens* using **Phytozome v13** (https://phytozome-next.jgi.doe.gov/). Genomic, mRNA and protein sequences for all PpMSLs were collected from Phytozome v13. RNAseq data was downloaded from the **oneKP Project**. To find potential cis-regulatory elements on the putative promoters of MSLs in *Physcomitrium*, **New Place** (https://www.dna.affrc.go.jp/PLACE/) was used. Protein domain analysis was carried out using **InterPro** (https://www.ebi.ac.uk/interpro/). **Snapgene viewer** was used for gene and protein domain annotations.

3.2. Plant culture and maintenance

Physcomitrium patens ecotype "Gransden" was used in this study. It was raised in conformity with growth standards (Cove *et al.*, 2009). Tissues were grown under a 16:8 hr (light: dark) cycle at 24℃ in a plant growth incubator. One-month-old moss tissues were used for all experiments.

3.3. Dehydration-rehydration experiments

To deduce whether the observed pattern of differential expression is due to dehydration-rehydration stress, we designed the following experiment: One-month-old gametophores grown on BCDAT media (**Appendix 1**) (overlayed with a cellophane disc at 24℃ with 16 h day-8 h light conditions were used for dehydration. The cellophane disc was then placed on two layers of sterile Whatman paper in a Petri plate for 2 and 12 h at 24 °C so that they were properly dehydrated. Some colonies were harvested by causing minimal damage to the plant at set time points along with a control population. The remaining colonies were transferred to Petri plates containing MilliQ water for them to rehydrate for 12 h. Subsequently, rehydrated tissue was harvested at set time points along with a

control population. RNA was isolated from these tissues, and cDNA was synthesised. This cDNA was then used for RT-qPCR analysis to check the relative expression.

3.4. RT-qPCR analysis

Using RNAiso Plus (Takara Bio USA Inc.), total RNA was extracted from gametophore tissue. Two micrograms (2μg) of RNA samples were reverse-transcribed to create cDNA using oligo dT primers and SS-IV reverse transcriptase (Invitrogen, MA, USA) (Mohanasundaram *et al.*, 2021). SYBR Premix Ex Taq II (Tli RNaseH Plus) was used for the relative quantification of target transcripts in a Bio-Rad CFX96 Touch Real-Time PCR Detection System. *P. patens* β-Actin was used as a reference gene for relative expression analysis of target genes. The 2-ΔΔCt method was used to calculate fold change (Schmittgen and Livak, 2008). Primers for respective genes were designed such that the amplicon size ranges from 150-200bp, with annealing temperature ranging from 60-65℃, GC content above 50%, and primer size around 20bp (**Table 1**).

Table 1: Details of primers used for RT-qPCR analysis.

Sr. No.	**Primer Name**	**Primer Sequence**	**Amplicon Size**	**%GC**
1.	Pp3c3_12760_qF	CCCCTAGCTTTCCTTCCGCT	195 bp	60%
2.	Pp3c3_12760_qR	CGCCTGCTCCTTACCCTGAA		60%
3.	Pp3c7_19980_qF	GGCTTCCGTACTCAGGCGAA	165 bp	60%
4.	Pp3c7_19980_qR	CATCCCGAGGCGAATTGTGC		60%
5.	Pp3c13_20730_qF	AAAAGAGCGGATGCCCAGGT	182 bp	55%
6.	Pp3c13_20730_qR	CAGCGGCAGAAAACCAAGGG		60%
7.	Pp3c15_19020_qF	ATCCGGAGTGGTGTGACGAC	180bp	60%
8.	Pp3c15_19020_qR	TTCAGCCAACGCCATGCTTC		55%

3.5. Genomic DNA isolation from moss

The whole genomic DNA of *P.patens* tissue was isolated using the CTAB method. Approximately 100mg of moss tissue was ground using mortar and pestle. The plant cells were lysed using 400μL of 2x CTAB buffer and 5μL of 2-mercaptoethanol. The sample

was then incubated in a 60˚C water bath for one hour and centrifuged. The supernatant was treated with RNAseA in a 37˚C water bath for 30 minutes to get rid of complete RNA. The sample was washed with a mixture of Chloroform and Isoamylalcohol in a 1:1 ratio until the solution became clear. DNA was precipitated using 100% isopropyl alcohol. The pellet was washed with 70% ethanol, dried and eluted in 35µL MilliQ water.

3.6. Cloning of promoter sequences

To develop the GUS control construct in the pTFH15.3 vector backbone, the β-glucuronidase (GUS) gene was amplified from the binary vector pBI121, using GUS_XmaI_F and GUS_ApaI_R primer pair. Using the restriction enzymes XmaI and ApaI, the GUS fragment was cloned into the pTFH15.3 binary vector downstream to the rice *Actin* promoter. This construct was used for PEG-mediated protoplast transformation of the moss to generate lines expressing GUS under the activity of the rice *Actin* promoter. This pTFH15.3:GUS control vector was further used for cloning the full-length and sequentially deleted *PpMSL3* (Pp3c3_12760, 2.5 Kb) and *PpMSL15* (Pp3c15_19020, 2.5 Kb) promoters. Initially, amplified sequences from the moss genomic DNA were cloned into the pGEM-T Easy subcloning vector (Promega Corporation) and confirmed through sequencing. Later, destination vector-specific primers were used to amplify promoter sequences. A detailed cloning strategy is mentioned in the results **section 5.6**.

3.7. Transformation of plasmids in *Ultra Competent Cells* and verification of constructs

All plasmid constructs were transformed into Ultra Competent Cells. For transformation, competent cells stored at -80˚C were thawed on ice for 10-15 minutes. 50µL of cells were mixed with 5µL of the ligation reaction, incubated on ice for 20 minutes and heat shocked in a 42˚C water bath for exactly 90 seconds. 800µL of Lysogeny Broth (LB) was added, and the sample was incubated in a 37˚C incubator for one hour at 180 rpm. In 200µL of LB, the cells were resuspended after being pelleted. Cells were spread on LA plates containing suitable antibiotics to a final concentration of 50µg/mL with the help of sterile

glass beads. The bacterial colonies were allowed to grow overnight on plates while being incubated at 37 °C.

To screen for putative positive clones colony PCR was performed using suitable primer pairs. Taq polymerase from (Takara Bio USA Inc.) was used for the screening of putative-positive clones. PCR cycling was performed as follows: 94˚C for 10 minutes, 35 cycles of 94˚C for 30 seconds, the appropriate annealing temperature for 30 seconds, 72˚C for an appropriate extension time of 1Kb/minute and 72˚C final extension for 5 minutes.

PCR-positive clones were grown overnight in 10mL LB containing the suitable antibiotic with a final concentration of 50μg/mL. Plasmid was isolated using Alkaline Lysis Method. PCR and restriction digestion confirmation were carried out using suitable primer pairs and restriction enzymes, respectively. The putative positive plasmids were then sent for sequencing using suitable primer pairs.

After sequence confirmation, plasmids were linearised using the restriction enzyme NotI. To get a transformation-ready plasmid, digested DNA was precipitated using sodium acetate: ethanol and resubstituted in 30μL of MilliQ water.

3.8. Preparation of moss protonema

One-month-old moss colonies were harvested and placed in scintillation vials containing liquid BCDAT media. The tissue was then crushed with the help of a homogeniser. With the help of cut tips, the media containing the homogenised tissue was spread on Petri plates containing solid BCDAT media overlayed with a cellophane disc. This tissue was then allowed to grow. The above process was repeated until all of the tissue was completely homogenised, i.e. no gametophores were seen, and the tissue contained only protonema,

3.9. PEG-mediated protoplast transformation

The moss *P. patens* was transformed with the plasmid constructs using the PEG-mediated protoplast transformation method (**Figure 5**) (Cove *et al.*, 2009). 7-day-old wild-type moss protonema was incubated in a 4% Driselase solution (prepared in 8% mannitol). The protoplast solution obtained was then filtered using a 100μm filter. The protoplasts were then collected using centrifugation. The protoplasts were then washed

three times using 8% mannitol solution by resuspending and centrifugation. After that, the pellet was resuspended in 900µL of MaMg (**Appendix 1**) solution. In another round bottom tube containing 30µL of plasmid (containing 30-90µg of DNA), 300µL of the protoplast solution was added along with 300µL of polyethylene glycol. This was then heated in a water bath to 45 ℃ for 5 minutes. 10mL of PRML (**Appendix 1**) was then slowly added to the sample after it had been cooled to room temperature. Protoplasts were pelleted by centrifugation, resuspended in 2mL of PRML and kept in the dark overnight.

The next day the protoplasts were kept under light for some time, centrifuged and resuspended in 2mL of PRMT (**Appendix 1**) solution maintained at 45 ˚C. This solution was spread on PRMB (**Appendix 1**), plated overlayed with a cellophane disc and incubated at 24 ˚C for six days. Long-day conditions (16 h light- 8 h dark) were used to grow all plant tissues. 100µg/mL of cefotaxime and 20µg/mL of amphotericin-B were used to prevent bacterial and fungal contamination, respectively.

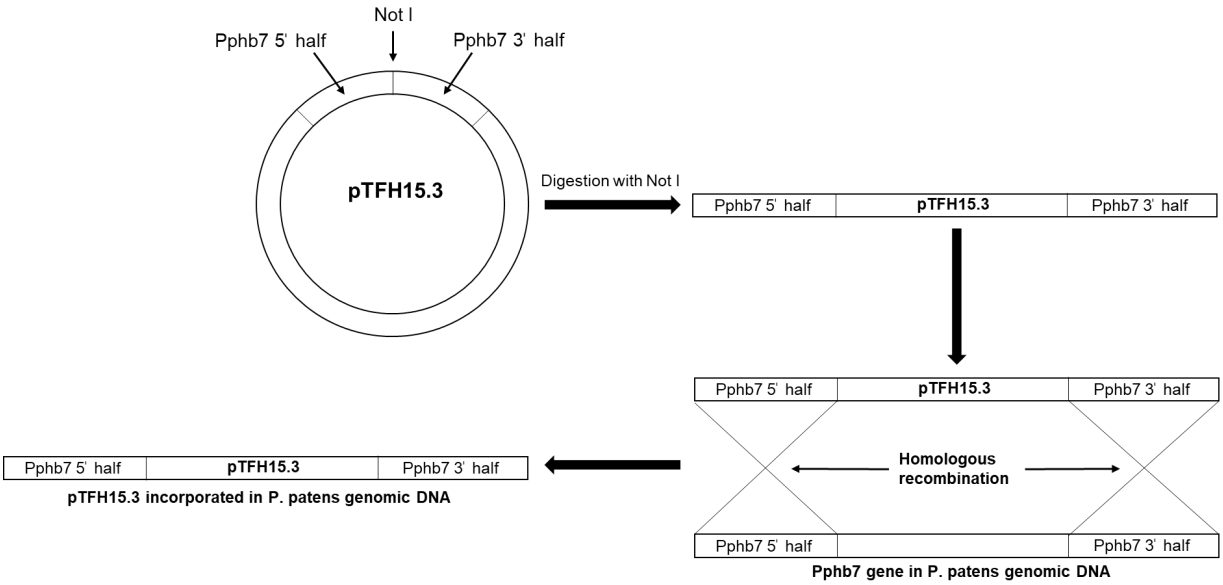


Figure 5: Strategy for DNA insertion by homologous recombination in Physcomitrium patens. *The binary vector pTFH15.3 contains the 5' and 3' halves of the Pphb7 gene from the P. patens genome. Digestion with the restriction enzyme NotI leads to the linearisation of the vector. When the linearised vector is transformed into the moss protoplasts, the Pphb7 sites recombine due to its intrinsic homologous recombination*

machinery. This leads to the insertion of the DNA sequence of the vector into the moss genome and the disruption of the Pphb7 gene.

3.10. Selection and characterisation of putative moss lines

3.10.1. Selection of putative moss lines

The moss protoplasts were allowed to regenerate for six days on PRMB plates overlayed by a cellophane disc. After six days, the cellophane discs were transferred to BCDAT media containing kanamycin with a final concentration of 20μg/mL for selection until surviving protoplasts were visible. These were then transferred to BCDAT plates for relaxation. After they grew, a few gametophores from each colony were transferred back to BCDAT plates containing kanamycin (20μg/mL) for the second selection. The surviving colonies were marked as stable transformants.

2.10.2. Molecular characterisation of stable transgenic moss lines

3.10.2.1. Genomic DNA PCR

To confirm transgenic lines, the genomic DNA of colonies that survived the second selection was carried out using the Shorty Method (Bezanilla, 2008). For this, 3-4 moss gametophores were harvested and put into Shorty Extraction Buffer (composition in **Table 2**). The tissue was then crushed using a micro pestle and centrifuged at maximum speed for five minutes. 300μL of supernatant was transferred to a fresh Eppendorf tube. An equal volume of Isopraponal was added and centrifuged at maximum speed for ten minutes. The supernatant was discarded, and the pellet was washed with 70% ethanol. The pellet was dried and eluted in 35μL of fresh MilliQ. 2μL of this genomic DNA was used for PCR confirmation. A PCR was set up using Taq polymerase with gene-specific as well as kanamycin-specific primers.

Table 2: Shorty Extraction Buffer composition

Stock Solutions	**Working Concentration**	**Volume in 50ml buffer**
1M tris-HCl	02.M Tris-HCl, pH 9.0	10mL
2M LiCl	0.4M LiCl	10mL
0.5M EDTA	25mM EDTA	2.5mL

10% SDS	1% SDS	5mL
-	MilliQ	22.5mL

3.10.2.2. GUS staining of PCR confirmed lines

One-month-old transgenic moss colonies were subjected to β-glucuronidase (GUS) staining assay (protocol adapted from JEFFERSON *et al.*, 1987 with minor modifications). Colonies were kept in the GUS solution (see composition in **Table 3**) overnight at 37℃ in the dark. The solution was then removed, and 50% ethanol was added and kept on a tube roller for 1-2 h. Next, 70% ethanol was added and kept for 1-2 h. Finally, 100% ethanol was added, and the colony was kept overnight on the tube roller. Complete colonies, as well as individual gametophores, were then imaged under the microscope.

Table 3: Composition of GUS staining buffer.

Stock solution	Working Concentration	Volume in 10mL buffer
Potassium ferricyanide (5mM)	0.5mM	1mL
Potassium ferrocyanide (5mM)	0.5mM	1mL
EDTA (500mM)	10mM	200µL
Triton (10%)	0.1%	100µL
X-glucuronide	-	4.8mg in 500µL DMSO
Sodium hydrogen phosphate (1M)	61mM	610µL
Monosodium phosphate (1M)	39mM	390µL
MilliQ	-	Upto 10mL

Chapter 4: Results

4.1. The moss *Physcomitrium patens* has 16 MSL homologs

The Plant genome database Phytozome v13 was used to collect information about all the predicted MSL genes in *P. patens*, given in **Table 4**. A total of 16 genes were identified using 10 *Arabidopsis thaliana AtMSL* gene sequences (Haswell, 2007) as query sequences. Protein domain analysis of these 16 predicted MSLs revealed that each of them has either the complete protein predicted as an MSL domain or a part of it predicted to be an MscS domain (**Figure 6A**). These proteins have multiple transmembrane domains, which is a characteristic of mechanosensitive ion-channel genes. **Figure 6B** shows the genomic sequence of each gene with the 5' and the 3' UTRs along with the intronic and exonic regions.

Table 4: Details of the 16 predicted MSL genes in P.patens.

Gene ID	**Predicted sub-cellular localisation**	**Basal expression level in gametophores (RPKM)**
Pp3c1_10620	Unknown	2.098
Pp3c1_35260	Chloroplast	17.7
Pp3c3_38000	Chloroplast	9.425
Ppc3c3_12760	Mitochondria	47.04
Pp3c7_19980	Unknown	0.135
Pp3c9_20120	YNAI-Related	0.026
Pp3c9_20150	Mitochondria	0.745
Pp3c10_8380	Chloroplast	4.8
Pp3c10_30950	Mitochondria	14.874
Pp3c11_3390	Unknown	2.156
Pp3c13_20730	Mitochondria	7.274
Pp3c14_810	Mitochondria	0
Pp3c14_8610	Chloroplast	8.001
Pp3c15_19020	Mitochondria	0.114
Pp3c26_12710	Unknown	1.029
Pp3c27_4760	Protein 6-Related	0.124

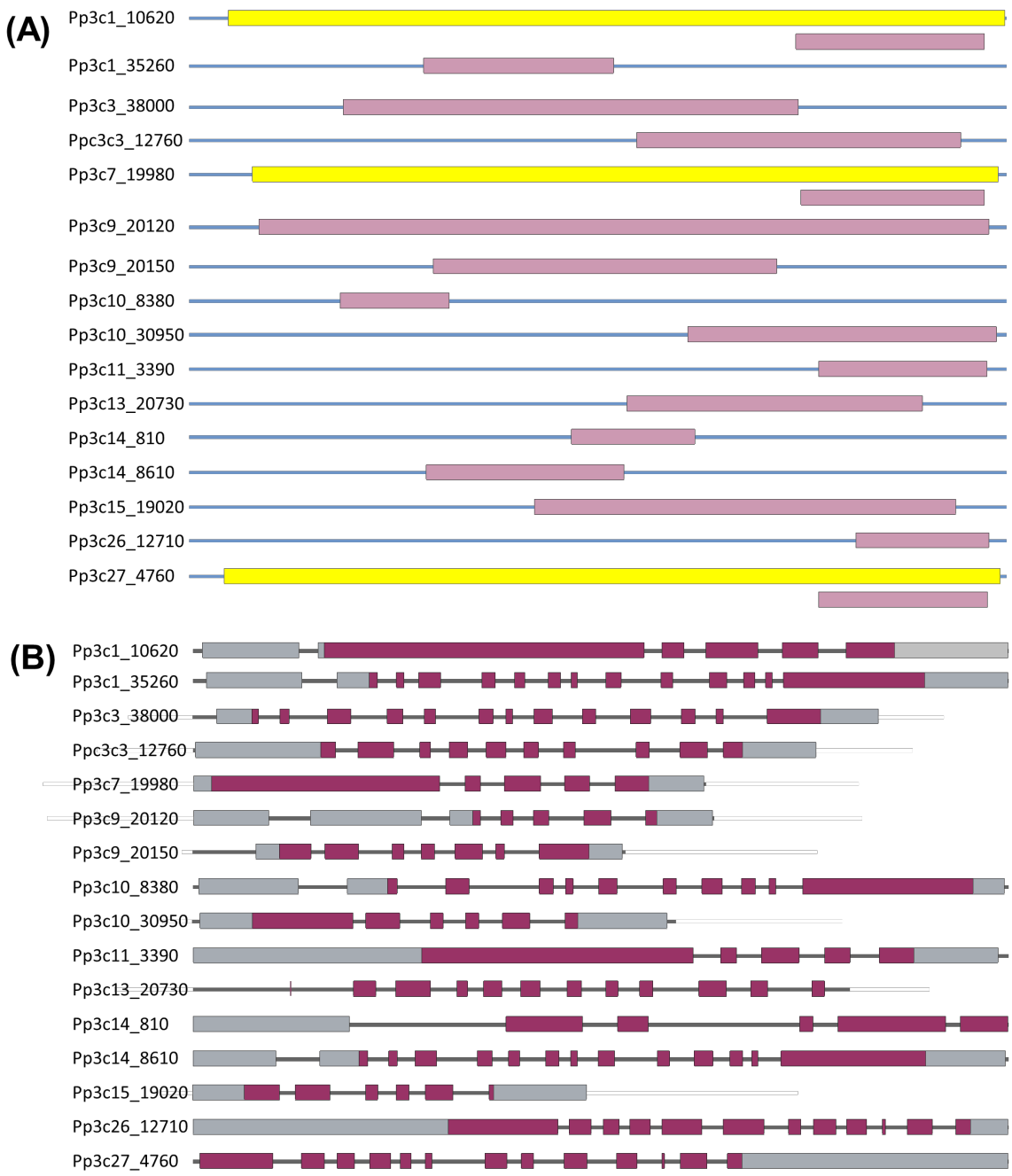


Figure 6: All 16 predicted PpMSL proteins contain the MscS domain. *(A) Domain prediction in the protein sequence of PpMSLs. Pink regions represent the MscS channels, whereas the yellow region represents MSL channels. (B) Comparison of the genomic locus of PpMSLs. Grey regions represent 5′ and 3′ untranslated regions (UTRs), and purple regions represent the coding DNA sequence (CDS).*

4.2. MSL gene homologs from *P. patens* and *Arabidopsis* show distribution into three groups through phylogenetic comparison

A multiple protein sequence alignment of the MscS domain of all *AtMSL* and *PpMSL* genes was performed. The alignment revealed the positional conservation of amino acids across the MscS domain. These results show that the MscS domain is conserved across bryophytes and angiosperms. The Neighbour-joining method was used to create an unrooted phylogenetic tree. (**Figure 7**). The tree is divided into three groups. No clade-specific distribution was observed for *PpMSL* and *AtMSL* MscS domains. Three groups were observed, which are consistent with the three groups observed in the previous unrooted phylogenetic tree of *AtMSL*s (Haswell, 2007). Six MSL homologs were grouped together with *AtMSL2*, while the other four *PpMSL*s were grouped together with *AtMSL1* and *AtMSL3*. The last group contains five *PpMSL*s grouped together with other 7 Arabidopsis MSLs.

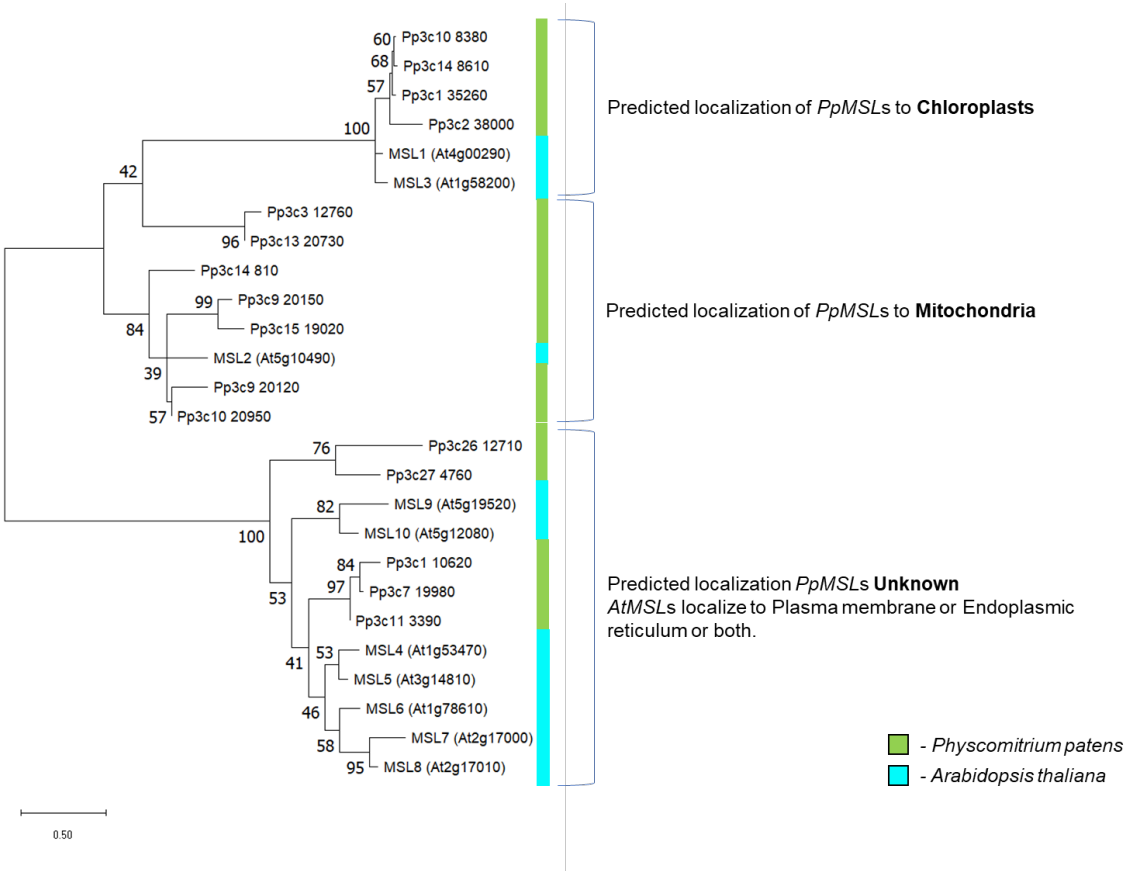


Figure 7: Comparative phylogenetic analysis of MSL homologs from Arabidopsis and Physcomitrium patens: *Alignment of MscS domain of 10 AtMSL and 16 PpMSL*

protein sequences using ClustalW. The maximum likelihood method was used for the construction of the unrooted phylogenetic tree with a 1000 bootstrap value. Amino acid sequences were retrieved from the 1KP project. Bootstrap values above 40 are mentioned at each node. The number near branches represents bootstrap values.

4.3 *PpMSL3* and *PpMSL15* gene expression alter during dehydration and rehydration stress conditions

Further, the expression of these genes in moss gametophores subjected to the dehydration and rehydration conditions were analyzed using the available moss transcriptome database. Transcriptomic data collected for these genes is appended in **Appendix 2**. Expression patterns of these genes in gametophore and protonema tissues grown under BCD liquid media were compared (**Figure 8A**). When the expression of these genes in the native state was compared to those subjected to dehydration stress (**Figure 8B**), it was observed that genes are differentially regulated in dehydration stress conditions. Notably, the expression of Pp3c15_19020 (*PpMSL15*) and Pp3c7_19980 (*PpMSL7*) genes was highly upregulated by 125 folds during the dehydration condition, whereas the expression of Pp3c13_20730 (*PpMSL13*) and Pp3c3_12760 (*PpMSL3*) genes was strongly downregulated. Other genes show no change in expression levels when subjected to the dehydration condition. Orchestration of the expression pattern of these genes during dehydration conditions suggests a possible regulatory switch particularly activated during dehydration conditions in moss gametophores.

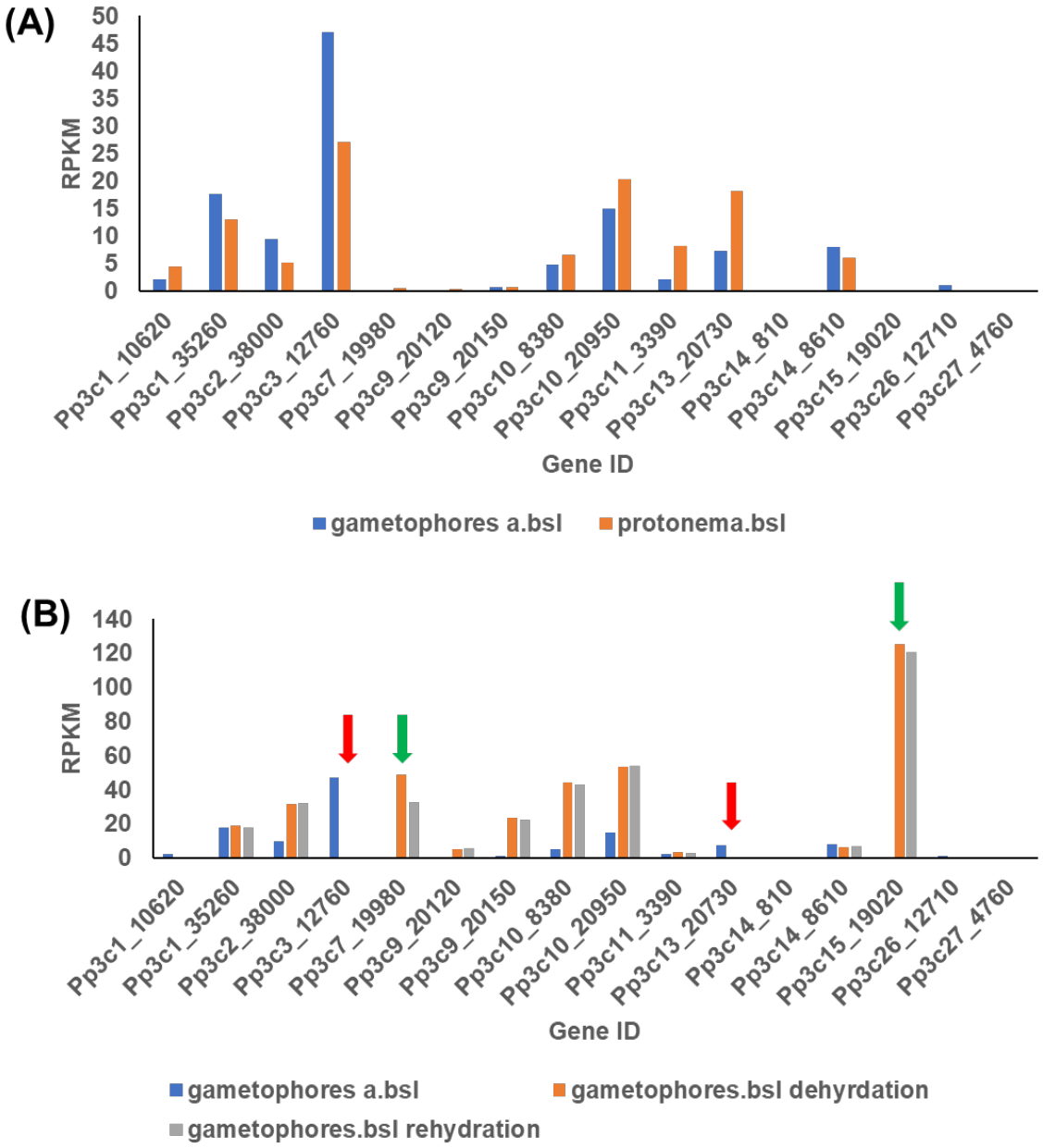


Figure 8: Gene expression analysis of 16 MSL genes in P.patens. *(A) RPKM values for all predicted PpMSL genes in both gametophores and protonema grown in BCD media. (B) Comparison of RPKM values of PpMSL genes in gametophores grown under normal, dehydration and rehydration conditions, respectively. Some of the genes are differentially expressed (downregulated genes marked in red and upregulated genes marked in green) under normal and stressed conditions. Data were adapted from* Perroud *et al.*, 2018.

4.4. Promoters of four PpMSL genes are enriched with binding sites for dehydration-rehydration specific Cis-regulatory elements

A 2.5 Kb upstream sequence, along with the 5'UTR of four *PpMSL* genes (MSL 3, 7, 13, and 15), was considered for promoter analysis. Regulatory element binding sites were identified and compared between the promoters of these genes (**Figure 9**). The details of each motif are mentioned in **Appendix 3**. These motifs are predicted to be dehydration responsive in plants like *Arabidopsis*, Rice, and *Brassica*. Some motifs were also associated with cold stress along with dehydration stress. We also identified a motif (LTRECOREATCOR15) only present in the promoters of MSL13 and MSL3, which are genes getting downregulated under dehydration stress. Since the gametophores used for RNAseq analysis were incubated at a lower temperature (15˚C) than the normal (25˚C) during the dehydration and rehydration stress conditions, we speculate that the cold stress condition could also have contributed to the regulation of gene expression. To deduce this, we have conducted dehydration-rehydration experiments on moss *P. patens,* explained in **section 4.2**.

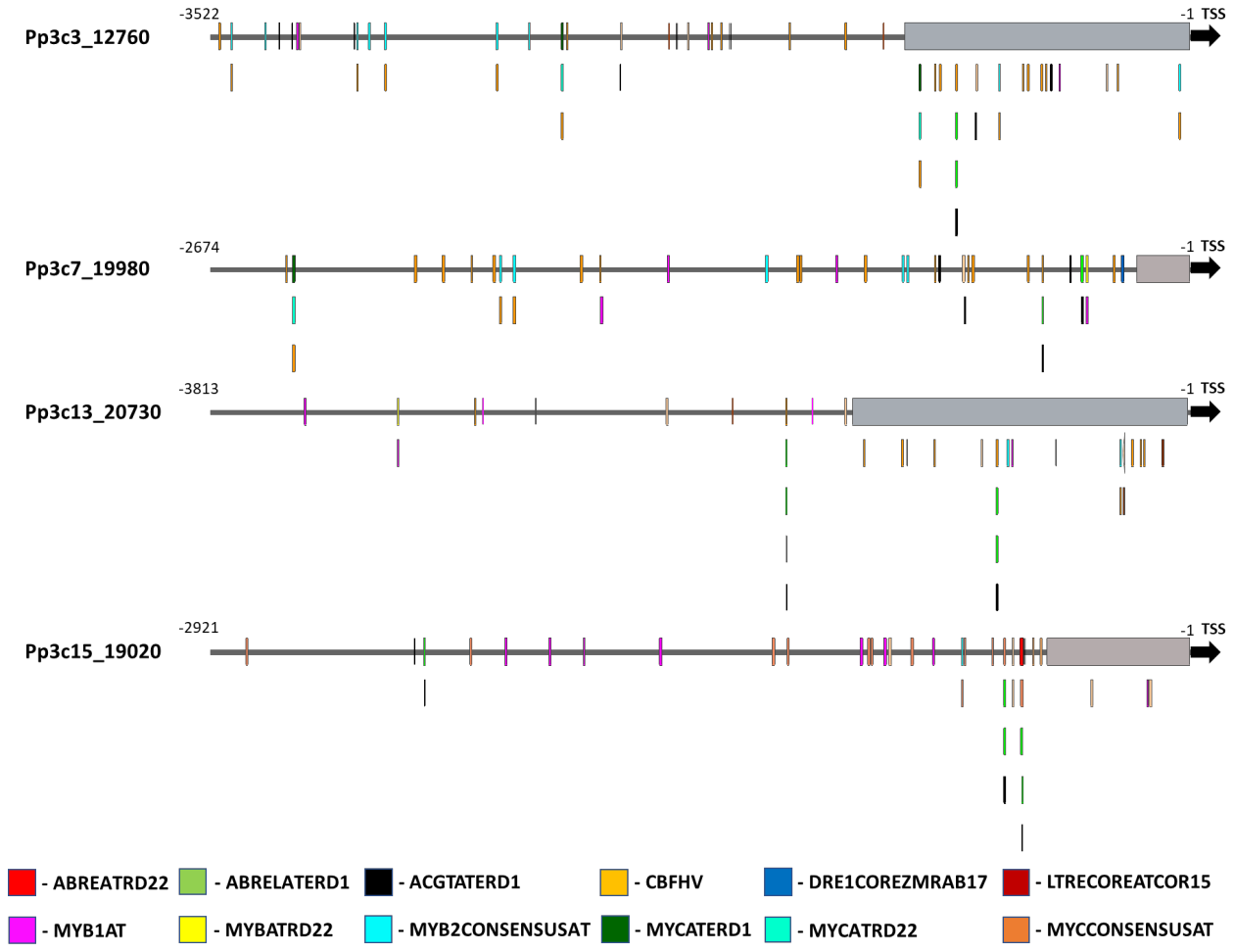


Figure 9: Distribution of cis-regulatory elements over the four PpMSL promoters. *Analysis of the putative promoters of four selected MSL genes: PpMSL15, PpMSL13, PpMSL7, and PpMSL3, showing key dehydration-related motifs. Each colour represents a particular motif, as shown in the colour key panel. These motifs could act as switches, allowing various cellular machinery to bind and regulate the transcription of the respective downstream genes.*

4.5. RT-qPCR analysis reconfirms the predicted expressional switch depicted through available RNAseq data

Tissues subjected to dehydration lost up to 85% of their wet weight after two hours of dehydration. Among all PpMSL homologs, *PpMSL15* showed the highest gene upregulation during dehydration stress, whereas *MSL3* showed the highest gene

downregulation (**Figure 8B**). We have validated this transcriptome data analysis through RTq-PCR. *PpMSL3* showed a 0.21-fold and 0.78-fold decrease in expression after 2 and 12 h dehydration and an 8-fold increase in expression after 2 h rehydration (**Figure 10A**). On the contrary, *PpMSL15* showed an approximately 21-fold and 146-fold increase in the relative expression level after 2 and 12 h dehydration conditions, whereas after 2 h rehydration, its expression level reduced to 28-fold (**Figure 10B**). From these results, we were able to validate the RNAseq data from Phytozome. With this, we proceeded with the cloning and line generation of the *PpMSL15* and *PpMSL3 promoters*.

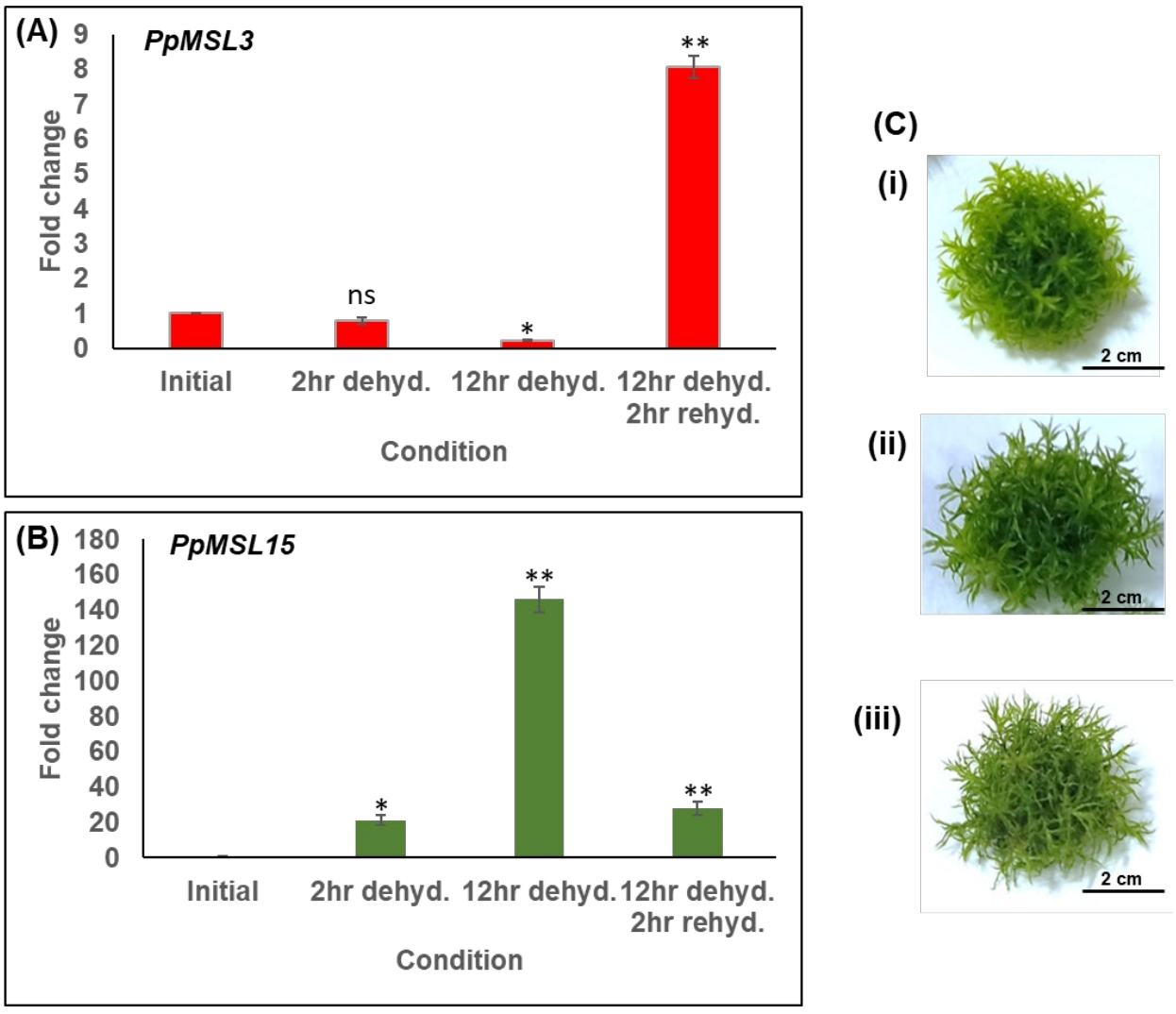


Figure 10: PpMS15 and PpMSL3 expression alters under various dehydration and rehydration conditions. *The graph represents the fold change in (A) PpMSL3 and (B) PpMSL15 expression compared to β-Actin under dehydration for 2 h, 12 h and*

*rehydration for 2 h, respectively. (C) One-month-old moss colonies (i), subjected to 2 h dehydration (ii) and 12 h dehydration (iii). Statistical analysis was performed using the Two-tailed Paired t-test, and P-values <0.05 and <0.01 were marked as * and **, respectively. Gene expression was normalised with the respective time point control samples, and then fold-change was calculated compared to the initial sample.*

4.6. Reporter gene (GUS) was cloned downstream of the rice Actin-1 promoter and used as a control vector

The β-glucuronidase (GUS) gene was amplified from the binary vector pBI121 and digested using XmaI and ApaI restriction enzymes (**Figure 11A**). Similarly, the pTFH15.3 vector was digested using XmaI and ApaI, and ligation was set up (refer to **Appendix 4** for vector map). Colony PCR was performed to screen for putative-positive clones, and plasmids were isolated. Plasmid PCR was done to confirm the clones (**Figure 11B**). Digestion of PCR-positive plasmids was performed using the EcoRV restriction enzyme (**Figure 11C**), and positive clones were confirmed through sequencing (**Appendix 5**).

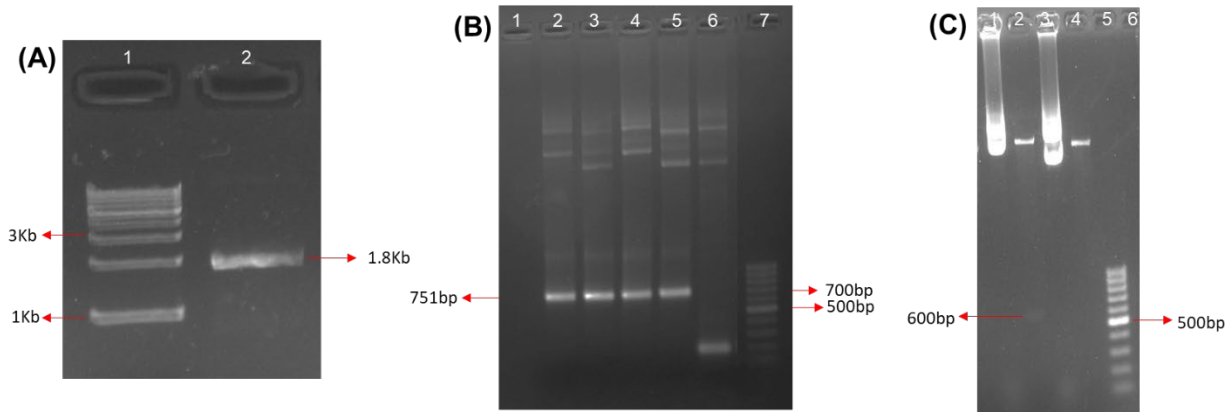


Figure 11: Generation of pTFH15.3:GUS control construct. (A) Amplification of GUS from pBI121 plasmid (Expected Size: 1.8Kb). (B) Plasmid PCR to confirm putative positive plasmids with an amplicon size of 751bp (**1.** Negative control, **2-6.** Putative positive plasmids. (C) Digestion of plasmids using EcoRV (Expected release: 600bp).

4.7. *PpMSL3* and *PpMSL15* promoters were cloned upstream of the reporter gene GUS

For amplification of chosen putative MSL3 and MSL15 promoters, *P. patens* genomic DNA was isolated, as mentioned in section 4.4. MSL3 promoter was amplified from the genomic DNA of the wild-type moss using the MSL3 Prom_KpnI_F and MSL3 Prom_AscI_R. Similarly, the MSL15 promoter was amplified using the MSL15 Prom_KpnI_F and MSL15 Prom_AscI_R primer pair. Phusion High-fidelity DNA polymerase was used for the amplification. PCR reactions were set up as follows:

	For MSL15		**For MSL3**	
	Temperature	**Time**	**Temperature**	**Time**
Initial denaturation	98˚C	2 minutes	98˚C	2 minutes
Denaturation	98˚C	30 seconds	98˚C	30 seconds
Annealing	55˚C	45 seconds	58˚C	45 seconds
Extension	65˚C	2 minutes	72˚C	1:30 minutes
Final Extension	65˚C	5 minutes	65˚C	5 minutes
Hold	4˚C	Forever	4˚C	Forever

Note that the extension temperature for the amplification of MSL15 promoter was lowered to 65℃ as compared to the standard 72℃ used for Phusion polymerase. Also, the $MgCl_2$ concentration was increased to 3mM to get the band at the expected size.

The amplification bands (**Figure 12A**, **Figure 13A**) were then gel cut, eluted (**Figure 12B**, **Figure 13B**), and ligated into the sub-cloning vector pGEM-T Easy Easy (MSL3) and pJET1.2 (MSL15). Colony PCR was performed to get putative-positive clones. Next, plasmid isolation of putative positive clones using the Alkaline lysis method was done, followed by digestion conformation. pGEM-T Easy:MSL3_Promoter clones were confirmed using the restriction enzyme EcoRI (**Figure 12C**), whereas pJET:MSL15_Promoter clones were confirmed by digestion using XmnI (**Figure 13C**).

Clones that were found positive in the digestion were sent for sequencing, and sequence-confirmed clones were used further for cloning.

Both pGEM-T Easy:MSL15_Promoter and pGEM-T Easy:MSL3_Promoter clones were digested using KpnI and AscI restriction enzymes (**Figure 12D**, **Figure 13D**). The obtained fallout was gel cut, purified (**Figure 12E**, **Figure 13E**), and ligated to the pTFH15.3:GUS vector, which was also digested with KpnI and AscI. Putative positive clones were identified using colony PCR. Plasmid isolation was carried out for putative positive clones, which were then confirmed by plasmid PCR (**Figure 12F**, **Figure 13F**) and restriction digestion. For restriction digestion confirmation of pTFH15.3:MSL3_Promoter:GUS clone XmnI enzyme was used (**Figure 12G**), whereas for pTFH15.3:MSL15_Promoter:GUS clone SwaI was used (**Figure 13G**). Positive clones were further sent for sequence confirmation (refer to **Appendix 5** for sequencing results). Positive clones were digested with NotI and Sodium acetate purified for PEG-mediated moss transformation. Vector maps of all constructs are given in **Appendix 4**.

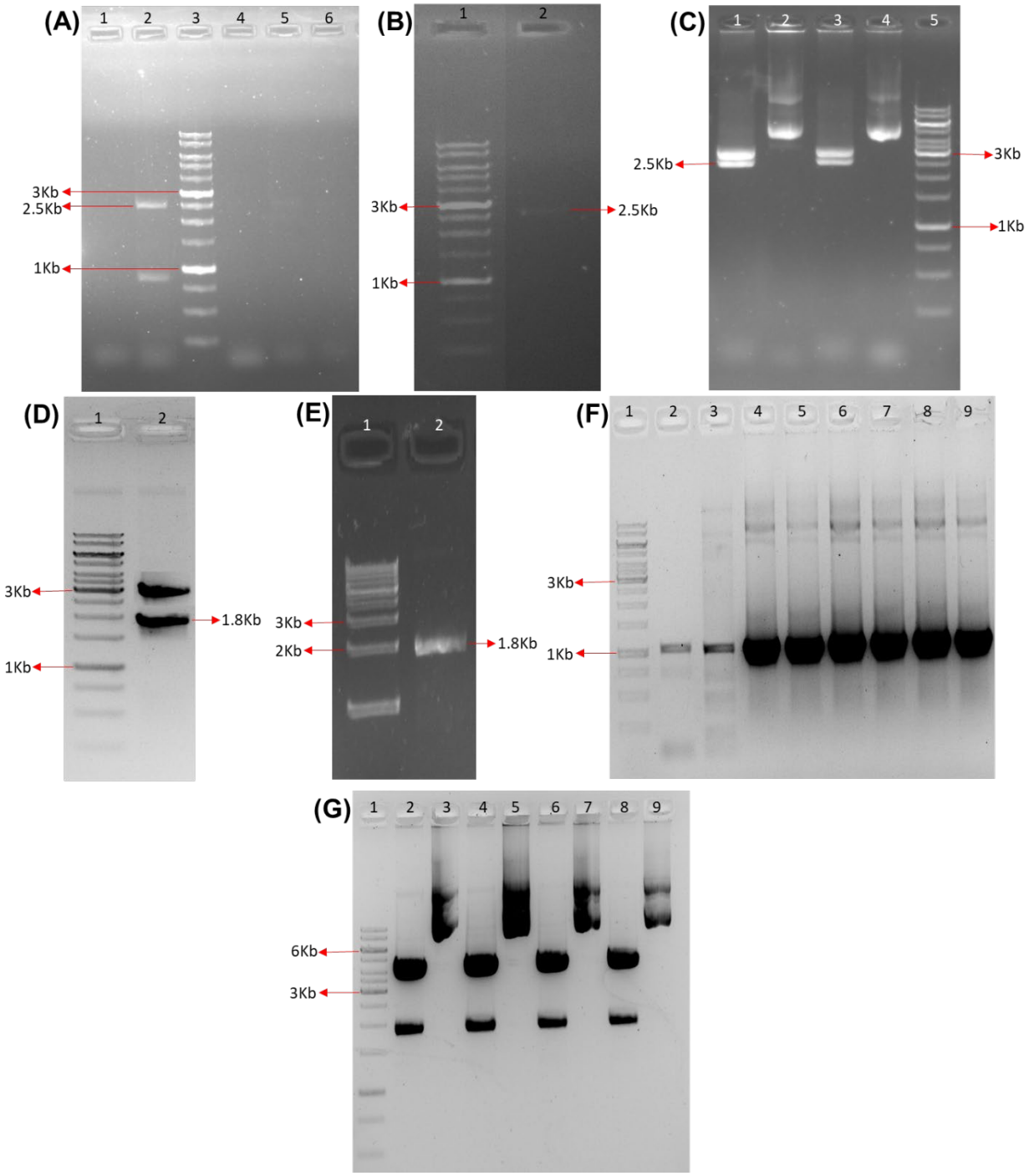


Figure 12: Cloning of PpMSL3 promoter in pTFH15.3:GUS vector. *(A) Gradient PCR to amplify MSL3 promoter (**1.** 58˚C, **2.** 60˚C, **3.** 1Kb ladder, **4.** 62˚C, **5.** 64˚C, **6.** 66˚C annealing temperature). (B) Verification of gel-purified PpMSL3 promoter amplicon. (C) Digestion confirmation of pGEM-T Easy:MSL3_promoter using EcoRI (expected bands: 3Kb, 2.5Kb). (D) Digestion of pGEM-T Easy:MSL3_promoter with KpnI and AscI and gel*

*elution of lower 1.8Kb promoter band. (E) Verification of 1.8Kb band. (F) Plasmid PCR of pTFH15.3:MSL3_Promoter:GUS using MSL3_SCE_F and GUS_SCE_R screening primer pair with an expected amplicon size of 1.1Kb (**1.** 1Kb ladder, **2-3.** Positive control, **4-9.** Putative positive clones). (G) Digestion of PCR-positive clones with XmnI (Expected bands: 4.2Kb, 4.0Kb, 1.9Kb).bands: 4.2Kb, 4.0Kb, 1.9Kb).*

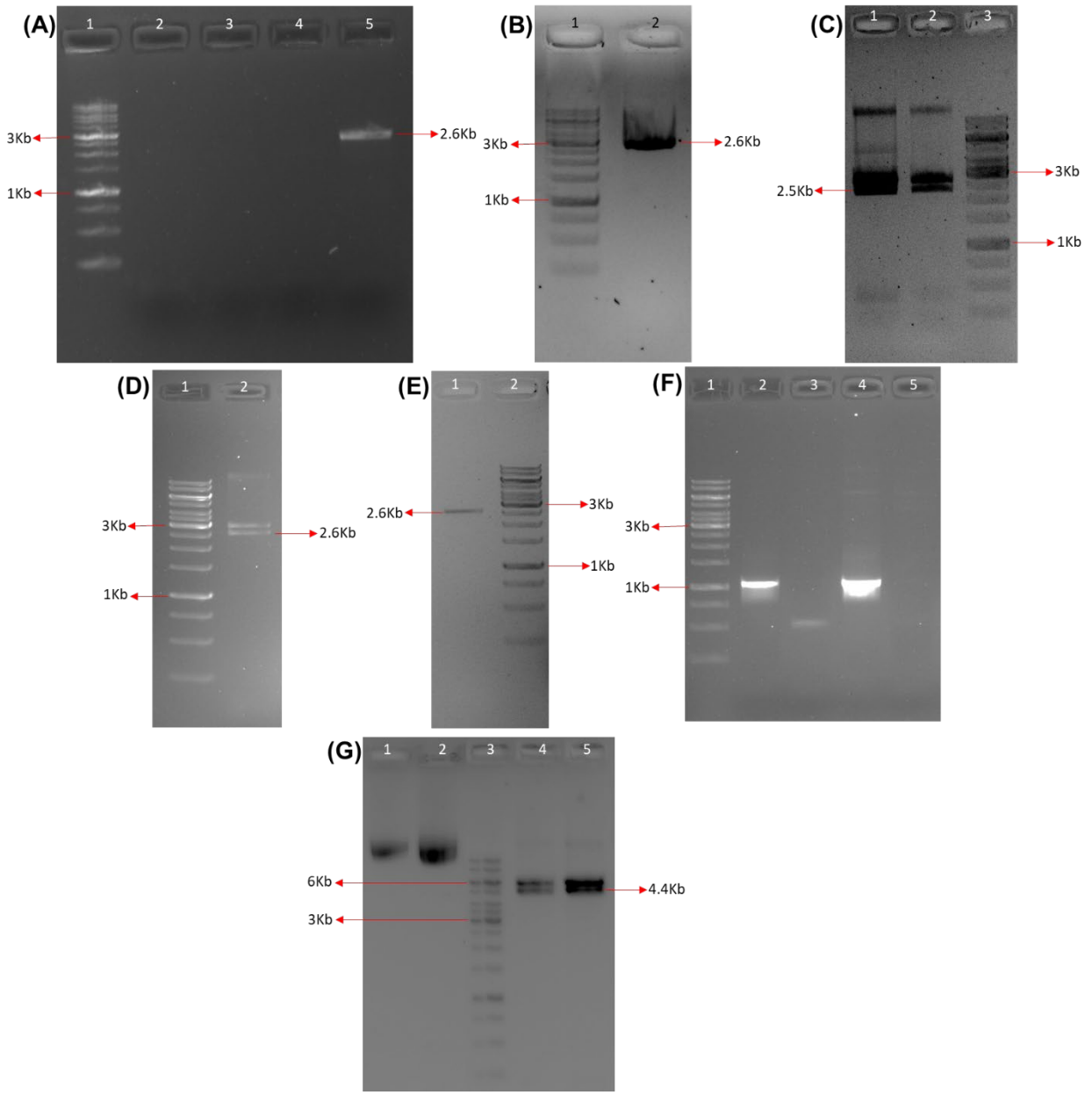


Figure 13: Cloning of PpMSL15 promoter in the pTFH15.3:GUS vector. *(A) PCR with $MgCl_2$ gradient to amplify MSL15 promoter (**1.** 1Kb ladder, **2.** 1.5mM, **3.** 2mM, **4.** 2.5mM,*

5. *3mM $MgCl_2$ gradient). (B) Verification of gel-purified PpMSL15 promoter amplicon. (C) Digestion confirmation of pGEM-T Easy:MSL15_promoter using PvuII (expected bands: 3Kb, 2.5Kb). (D) Digestion of pGEM-T Easy: MSL15_promoter with KpnI and AscI and gel elution of lower 2.6Kb band. (E) Verification of 2.6Kb promoter band. (F) Plasmid PCR of pTFH15.3:MSL15_Promoter:GUS using MSL15_SCE_F and GUS_SCE_R screening primer pair with an expected amplicon size of 1.1Kb (**1.** 1Kb ladder, **2-4.** Putative positive clones, **5.** Negative control). (G) Digestion of PCR-positive plasmids with SwaI (Expected bands: 6Kb, 4.4Kb).*

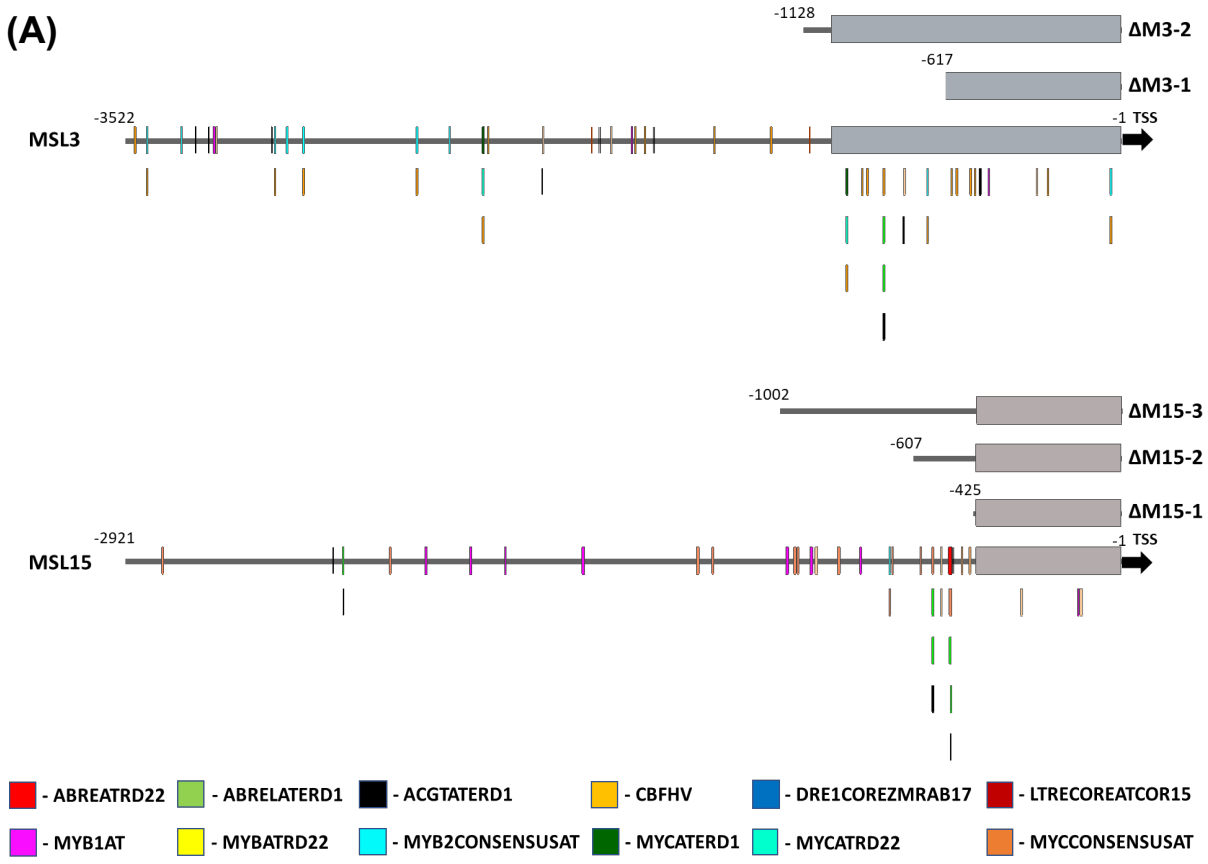
Table 5: Details of primers used for the cloning and confirmation of full-length and promoter deletion constructs.

	Full-length promoter cloning	
Sr. No.	**Primer Name**	**Primer Sequence**
1.	GUS_XmaI_F	AAAACCCGGGATGTTACGTCCTGTAGAAACCCCAACC
2.	GUS_ApaI_R	AAAAGGGCCCTCATTGTTTGCCTCCCTGCTG
3.	MSL3_Prom-KpnI-F	TGATTACGAATTCGAGCTCGGTACCCTTGTTCAGGGA TATGGGGTTGAC
4.	MSL3_Prom_AscI_R	AAAGGCGCGCCCGCCAATGTTCACTTTACCCAATCA
5.	MSL15_Prom_KpnI_F	AAAAGGTACCGGACATTTTGTTAAAAAGAAATAAG
6.	MSL15_Prom_AscI_R	AAAAGGCGCGCCCTACTCTTCAAATACTGTAC
	Deletion constructs cloning	
7.	MSL3 Prom_1Kb_F	AAAGGTACCGAACGAGAGCGAATGAGAGATCTTG
8.	MSL3 Prom_617bp_F	AAAGGTACCCGTAGGGCAGCCCTGCG
9.	MSL15 Prom_1Kb_F	AAAGGTACCGGTAATCATGATGAATGTG
10.	MSL15 Prom_607bp_F	AAAGGTACCCGTTTCCAGTTTCCTAC
11.	MSL15 Prom_5'UTR_F	AAAGGTACCGCACTTAGTTAGTGCC

	Screening and Sequencing	
12.	pTFH15.3_VSF	CAAGATCAGGAAGAGGGGAAAAGGG
13.	pTFH15.3 MCS_R	CAAACTTAGTAGGATTCTGGTGTGTGGGCA
14.	MSL3_Prom_SCE_R	TAACGTAATGGATGCTTCGTAGTTC
15.	MSL15_Prom_SCE_F	CCCTGCCACGTGTTTACGTC
16.	GUS_SCE_R	GCGATGGATTCCGGCATAGT

4.8. Motifs in the promoters of *PpMSL* genes are distributed into distinct groups

To determine the minimal functional promoter and motifs responsible for the regulatory switch, we designed deletion constructs based on the distribution of motifs. As shown in **Figure 14**, motifs on the putative promoters of MSL15 and MSL3 are distributed as distinct patches. For *PpMSL15*, three constructs corresponding to 1002bp, 607bp, and 425bp (5' UTR) were designed. On the other hand, for the *PpMSL3* promoter, two constructs of 1128bp and 617bp were designed. All the deletion constructs were also cloned in pTFH15.3:GUS control vector using KpnI and AscI restriction enzymes (refer to **Appendix 4** for vector maps). PCR for amplifying the deletion constructs was carried out using the forward primers, as shown in **Table 5**, and the same reverse primer used for the cloning of full-length promoters (**Figure 15A, 15B**). Amplification bands were gel eluted (**Figure 15C, 15D**). All constructs were subcloned in subcloning vector pGEM-T Easy. All subcloned constructs were confirmed by plasmid PCR using M13_F and M13_R primer pair. Suncloned vectors were then digested with KpnI and AscI restriction enzymes (**Figure 15E**). The promoter fallouts were then gel eluted (**Figure 15F, 15G**) and ligated to KpnI, and AscI digested pTFH15.3:GUS binary vector. The binary vector constructs were PCR-confirmed using M13_F and GUS_SCE_R primer pair. Digestion confirmation of MSL15 deletion constructs was done using XbaI (**Figure 15G, 15H**) and MSL3 deletion constructs with SwaI (**Figure 15H**). Digestion-confirmed vectors were further confirmed by sequencing (refer to **Appendix 5** for sequencing results).



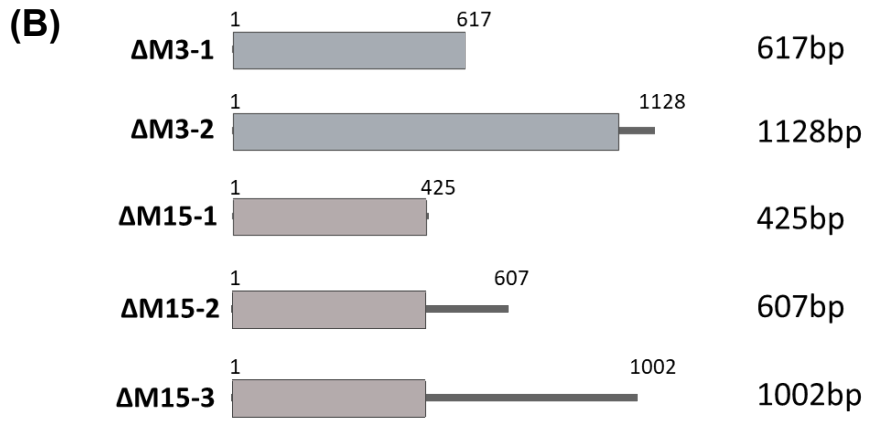


Figure 14: Schematic representation of the design of promoter deletion constructs according to the motif distribution. *(A) Two deletion constructs of 1128bp and 607bp (ΔM3-1 and ΔM3-2) lengths for the PpMSL3 promoter and three for the PpMSL15 promoter (425bp, 607bp, and 1002bp, ΔM15-1, ΔM15-2 and ΔM15-3 respectively) were designed considering the motif distribution and the maximum number of motifs covered.*

These will be cloned upstream of the GUS sequence in the pTFH15.3:GUS control vector. (B) Schematics of all five PpMSL3 and PpMSL15 promoter deletion constructs.

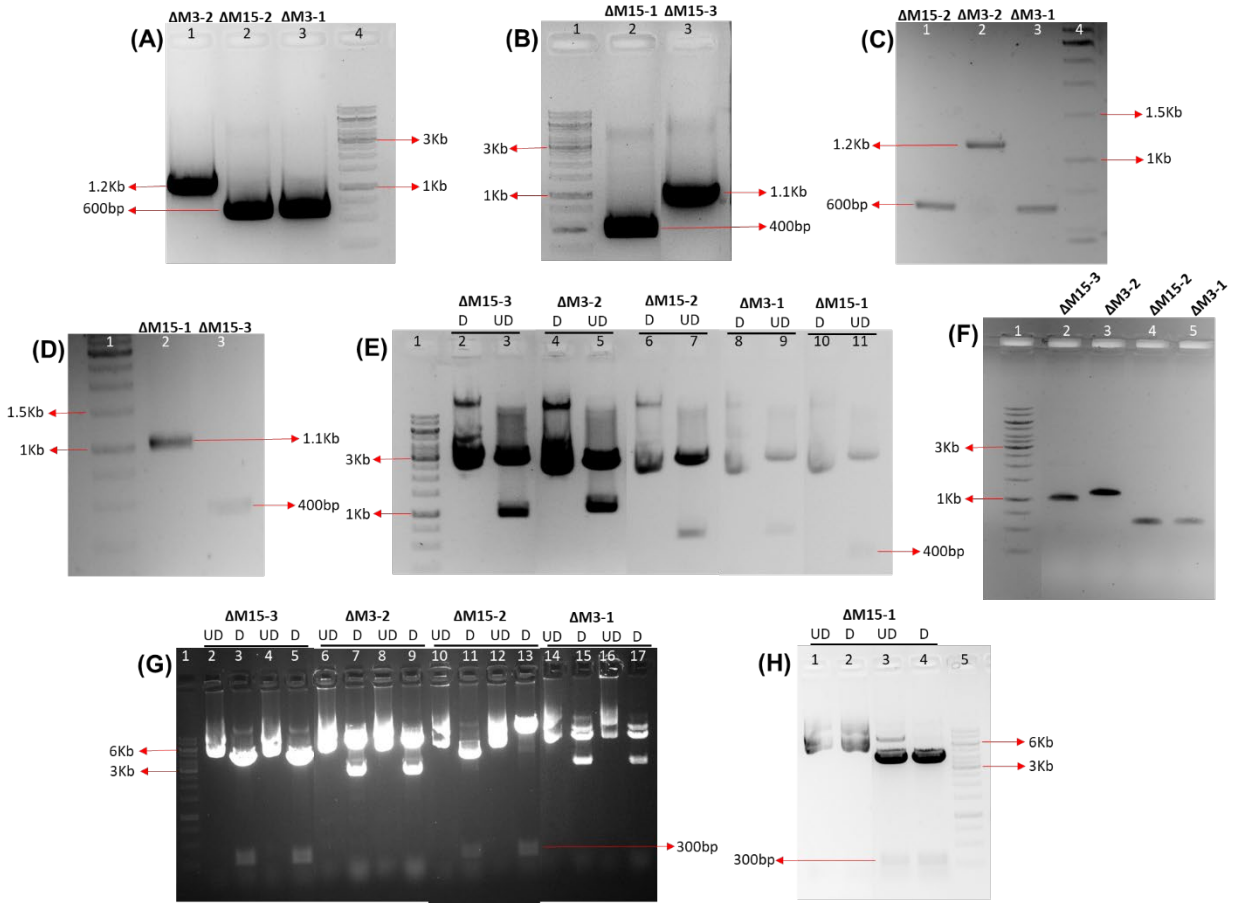


Figure 15: Generation of PpMSL15 and PpMSL3 promoter deletion constructs in pTFH15.3:GUS vector. *(A) Amplification images of constructs (**1.** ΔM3-2, **2.** ΔM15-2, **3.** ΔM3-1). (B) Amplification of ΔM15-1 (2) and ΔM15-3 (3). (C) Verification of amplified and gel-purified promoters (**1.** ΔM15-2, **2.** ΔM3-2, **3.** ΔM3-1). (D) Verification of gel-purified ΔM15-3 (2) and ΔM3-1 amplicons (3). (E) Digestion of pGEM-T Easy clones of all deletion constructs using KpnI and AscI. (F) Verification of deletion constructs digested from the respective pGEM-T Easy clones (**2.** ΔM15-3, **3.** ΔM3-2, **4.** ΔM15-2, **5.** ΔM3-2). (G) Digestion of MSL3 deletion constructs clones using SwaI and MSL15 clones using XbaI (Expected bands: ΔM15-3: 4.4Kb, 4.1Kb and 300bp, ΔM15-2: 4Kb, 4.1Kb and 300bp, ΔM3-2: 6.4Kb and 2.6Kb, ΔM3-1: 5.9Kb and 2.6Kb). (H) Digestion of pTFH15.3:ΔM15-1:GUS plasmids using XbaI (Expected bands: 3.9Kb, 4.1Kb and 300bp).*

4.9. pTFH15.3:GUS and *PpMSL*3 promoter lines were successfully confirmed through genomic DNA PCR

Transgenic lines were generated for the pTFH15.3:GUS control vector and pTFH15.3:MSL3_promoter:GUS and pTFH15.3:MSL15_promoter:GUS constructs. Although we did not quantify any phenotypic traits, gametophores from all the transgenic lines did not show any visible phenotypic change (**Figure 16**). Genomic DNA was isolated from all lines that survived in the second selection using the Shorty method. PCR was performed using both insert-specific as well as kanamycin-specific primers. Wild-type genomic DNA was used as a negative control, and the plasmids of the respective constructs were used as positive controls in the PCR. A reaction without any template was labelled as the water control. The negative control and water control did not show any bands, which confirms that there was no contamination while setting up the PCR. Bands of expected sizes were obtained in the positive control reaction as well as the genomic DNA of lines (**Figure 17**).

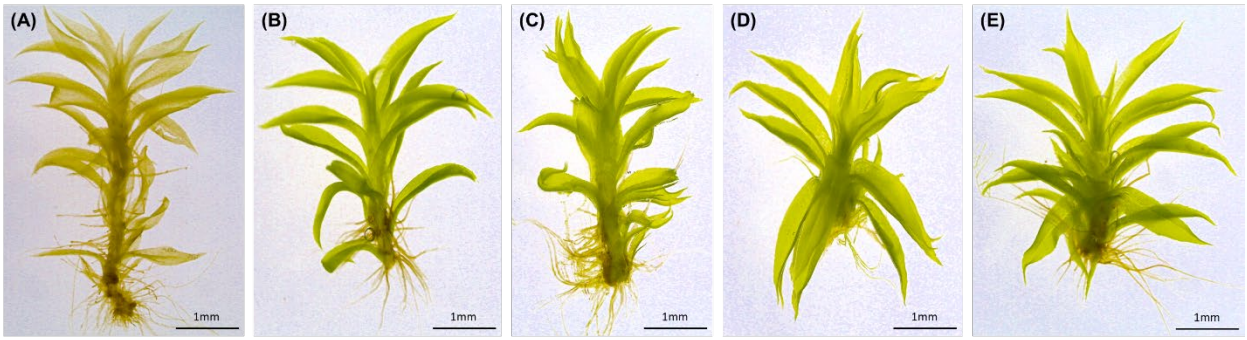


Figure 16: Gametophores from all transgenic lines do not show any visual phenotypic changes. *The Control vector (B), PpMSL3 line #2 (C), PpMSL3 line #9 (D) and PpMSL15 line #28 (E) gametophores show no change in the phenotype as compared to the wild-type moss gametophore (A).*

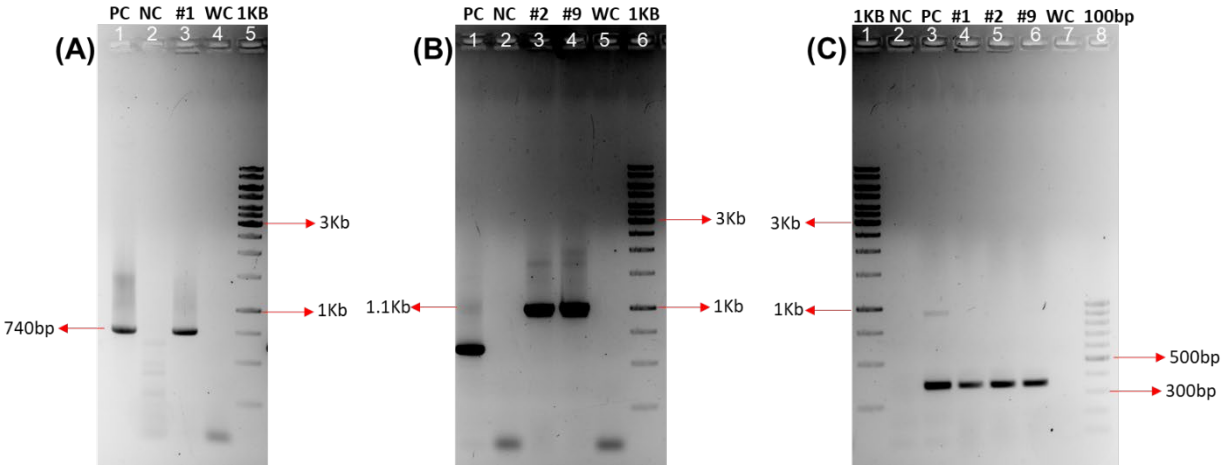


Figure 17: PCR confirmation of control and PpMSL3 promoter transgenic lines. *(A) PCR Confirmation of pTFH15.3:GUS control line with an expected band size of 740bp (**1.** Positive control, **2.** Negative control, **3.** Transgenic line #1, **4.** Water control, **5.** 1Kb ladder). (B) PCR confirmation of PpMSL3 promoter line (**1.** Positive control, **2.** Negative control, **3.** Transgenic line #2, **4.** Transgenic line #9, **5.** Water control, **6.** 1Kb ladder). The expected size of the band was 1.1Kb. (C) PCR confirmation of all lines using Kanamycin-specific primers with an expected band size of 350bp (**1.** 1Kb ladder, **2.** Negative control, **3.** Positive control, **4.** pTFH15.3:GUS transgenic line #1, **5.** PpMSL3 promoter transgenic line #2, **6.** PpMSL3* promoter *transgenic line #9, **7.** Water control, **8.** 100bp ladder).*

4.10. pTFH15.3:GUS vector control line showed ubiquitous GUS activity

When subjected to GUS staining, pTFH15.3:GUS vector transgenic line #1 showed intense blue staining in the stem region of the gametophores (**Figure 18**). Some gametophores also showed staining in the phylids. This indicated that the GUS sequence in the vector is able to form a functional protein detected through staining.

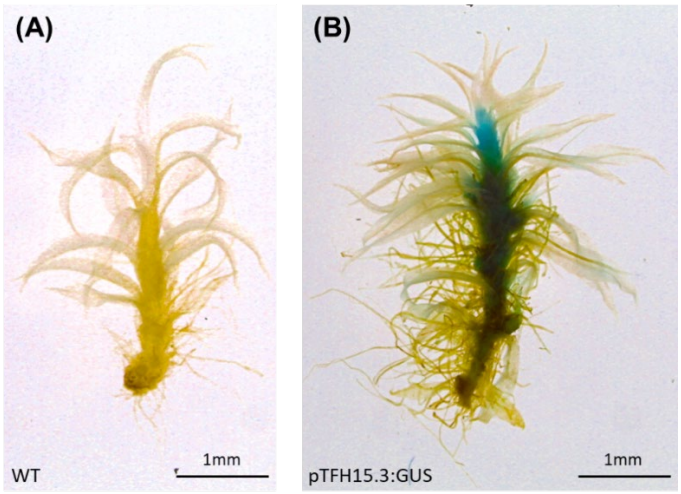


Figure 18: GUS activity was evident in the pTFH15.3:GUS vector control line #1 colony*. Compared to wild-type moss gametophore (A), the transgenic moss colony infused with GUS staining solution showed vibrant blue colour throughout the stem in all gametophores and in the phylids of some gametophores (B).*

Chapter 5: Discussion

5.1. Mechanosensation: a key biological process for plants

Mechanosensation is a crucial mechanism that plants use to sense various mechanical stimuli present in their environment. These include touch, temperature, gravity and turgor pressure which affect plant growth and development in various ways. During the course of evolution, while plants were transitioning from water to land, various adaptations would have been necessary for plants to cope with the extremely harsh conditions present on the land. Mechanosensitive ion channels could have been one possible way for plants to sense and respond to such changes in the environment. These channels could either trigger momentary changes in the state of the cell through the flux of ions or lead to the activation of downstream factors, which further leads to changes in the expression of key development regulatory genes. Previous studies have shown MSL activities in various plants like *Arabidopsis*, Maize, Rice etc. MSLs are shown to localise to the membranes of mitochondria, plastids, endoplasmic reticulum, as well as plasma membrane of plant cells (Haswell, 2007). There are 10 MSL genes reported in Arabidopsis, and their localization and gene function is actively investigated (Hamilton *et al.*, 2015). In the current study, we aimed to understand the regulatory mechanism of MSL gene activation under dehydration-rehydration stress conditions in moss *P. patens*.

5.2. MSLs show differential expression in *Physcomitrium patens*

Detailed BLAST and phylogenetic analyses indicated that there are 16 MSL genes present in the moss *Physcomitrium patens* (**Table 4**). Expression analysis revealed that four of these MSL genes (*PpMSL3*, *PpMSL7*, *PpMSL13* and *PpMSL15*) were differentially regulated under dehydration and rehydration stress conditions (**Figure 8**). RT-qPCR analysis showed that PpMSL15 is upregulated when the moss is subjected to dehydration stress, and its levels normalize to the native state after rehydration. On the other hand, PpMSL3 expression was downregulated upon dehydration, and its levels increased when subjected to rehydration (**Figure 10**). We hypothesized that there is a possible regulatory switch, particularly activated during dehydration conditions in moss gametophores. This

switch could help the plant cope with the dehydration stress. Moss is also known to be a very versatile plant which could explain such changes to its transcriptome under stress.

5.3. *PpMSL* promoter house key dehydration responsive motifs

Further, we found that promoters of PpMSL genes harbour various binding sites for regulatory elements predicted to be functional during dehydration stress conditions (**Figure 9**). These motifs were distributed on these promoters in the form of distinct groups. With the sequential deletions studies coupled with the reporter fusion assay system, we aimed to investigate the core promoter sequence and the function of associated regulatory motifs/elements to understand the regulation of MSL gene expression during dehydration stress. Most of these motifs were clustered near or on the 5' UTR region of these genes (**Figure 14**). We have successfully cloned and coupled full-length promoters (**Figure 12**, **Figure 13**) and promoter deletion sequences (**Figure 15**) to the reporter (GUS) system and transformed them into WT *P. patens* protoplasts. GUS expressed under rice actin promoter was used as a control (**Figure 11**). We were able to generate and confirmed transgenic moss lines expressing GUS under actin (control) and PpMSL3 full-length promoters (**Figure 17**). Control lines showed intense GUS staining throughout the stem (**Figure 18**). GUS staining and analysis are under progress for full-length promoters and promoter deletion MSL lines.

Salient findings of the study

- A total of 16 Mechanosensitive-ion channel-like (MSL) genes were identified through *in-silico* analysis in the moss *Physcomitrium patens*.
- Out of the 16 MSLs, four MSL genes: *PpMSL3*, *PpMSL7*, *PpMSL13*, and *PpMSL15,* showed significant change in expression under dehydration and rehydration stress conditions available through the oneKP project transcriptomic data.
- The differential expression pattern for *PpMSL*3 and *PpMSL*15 was successfully validated through RT-qPCR analysis under dehydration-rehydration stress conditions.
- After 12 h of dehydration, *PpMSL3* showed a 0.78-fold decrease, while *PpMSL15* showed a 146-fold increase in their gene expression. Subsequently, after 2 h rehydration, *PpMSL3* expression increased by 8-fold and *PpMSL15* expression reduced to 28-fold, compared to the initial value.
- Full-length promoters of *PpMSL3* and *PpMSL15* were successfully cloned upstream of the GUS reporter gene to understand the native spatiotemporal expression pattern.
- Promoter analysis of *PpMSL3* and *PpMSL15* genes using New Place revealed that they harbour key dehydration-responsive elements such as MYB2CONSENSUSAT, LTRECOREATCOR15, CBFHV and several others.
- The motifs were present in the form of distinct groups. Sequential deletion of these groups was aimed at through the deletion constructs.
- To understand the effect of these motifs on promoter activity, we generated two promoter deletion constructs for *PpMSL3* and three for *PpMSL15* promoters.
- Using the control vector, two full-length promoters and five promoter deletion constructs, several events of PEG-mediated moss transformation were conducted in an attempt to generate stable transgenic lines.
- One control vector and two *PpMSL*3 full-length promoter lines were confirmed using genomic DNA PCR with insert-specific as well as kanamycin-specific primers.
- The control vector and the full-length PpMSL promoter transgenic lines did not show any visible phenotypic changes relative to the wild-type moss gametophores.
- The pTFH15.3:GUS control vector transgenic line showed ubiquitous GUS expression, indicating that the GUS protein is functional and is not leading to any

visible phenotypic changes in the moss. This strategy will be applied to understand the activity of the full-length promoters and their deletion constructs under dehydration-rehydration stress conditions.

Future Perspectives

In the current study, we aimed to explore the molecular regulatory switch functional in moss under drought stress conditions. Two MSL genes (*PpMSL3* and *PpMSL15*) showed dynamic expression patterns under dehydration-rehydration conditions. Promoter analysis of these genes revealed the presence of various cis-regulatory elements shown to act under dehydration conditions. To decipher the role of these elements, we designed deletion constructs for these promoters.

The comparative analysis of the GUS activity in the deletion constructs' transgenic lines will help in determining the minimal promoter sequence required for the expression of *PpMSL3* and *PpMSL15* genes in *P.patens*. This analysis will reveal the significance of these elements in switching gene expression. We could also identify whether any enhancer elements are present upstream to the minimal promoter sequence.

Line generation for all five deletion constructs and full-length *PpMSL*3 and *PpMSL*15 promoters are in progress. Once lines are confirmed, GUS staining and microscopic imaging for the positive lines will be carried out. All these results will be used for the validation of minimal promoter activity with reporter fusion assays. Dehydration-rehydration experiments will be carried out to explore the role of the regulatory motifs in causing the switch in gene expression. Further, this study can be extended to characterise the *PpMSL* genes and decipher the chain of molecular processes they activate, helping the moss *P. patens* cope with the dehydration stress. The overall outcomes of this study could help in unravelling several evolutionary questions and have biotechnological applications in developing improved drought-tolerant plants.

References

1. Beike, AK, Von Stackelberg, M, Schallenberg-Rüdinger, M, Hanke, ST, Follo, M, Quandt, D, McDaniel, SF, Reski, R, Tan, BC, and Rensing, SA (2014). Molecular evidence for convergent evolution and allopolyploid speciation within the Physcomitrium-Physcomitrella species complex. BMC Evol Biol 14.
2. Bezanilla, M (2008). Bezanilla Lab Moss Methods Manual.
3. Cove', DJ, and Knight, CD (1993). The Moss Physcomitrella patens, a Model System with Potential for the Study of Plant Reproduction. Plant Cell 5, 1483.
4. Cove, D (2005). The Moss Physcomitrella patens. Https://DoiOrg/101146/AnnurevGenet39073003110214 39, 339–358.
5. Cove, DJ, Perroud, PF, Charron, AJ, McDaniel, SF, Khandelwal, A, and Quatrano, RS (2009). The moss Physcomitrella patens: a novel model system for plant development and genomic studies. Cold Spring Harb Protoc 4.
6. Daku, RM, Rabbi, F, Buttigieg, J, Coulson, IM, Horne, D, Martens, G, Ashton, NW, and Suh, DY (2016). PpASCL, the Physcomitrella patens anther-specific chalcone synthase-like enzyme implicated in sporopollenin biosynthesis, is needed for integrity of the moss spore wall and spore viability. PLoS One 11.
7. García-Añoveros, J, and Corey, DP (1997). The molecules of mechanosensation. Annu Rev Neurosci 20, 567–594.
8. Hamilton, ES, Schlegel, AM, and Haswell, ES (2015). United in diversity: mechanosensitive ion channels in plants. Annu Rev Plant Biol 66, 113–137.
9. Haswell, ES (2007). MscS-Like Proteins in Plants. Curr Top Membr 58, 329–359.
10. Hiss, M, Meyberg, R, Westermann, J, Haas, FB, Schneider, L, Schallenberg-R€ Udinger, M, Ullrich, KK, and Rensing, SA (2017). Sexual reproduction, sporophyte development and molecular variation in the model moss Physcomitrella patens: introducing the ecotype Reute. Wiley Online Libr 90, 606–620.
11. Hohe, A, Rensing, SA, Mildner, M, Lang, D, and Reski, R (2002). Day Length and Temperature Strongly Influence Sexual Reproduction and Expression of a Novel MADS-Box Gene in the Moss Physcomitrella patens. Plant Biol 4, 595–602.
12. JEFFERSON, R, BEVAN, M, and KAVANAGH, T (1987). The use of the Escherichia coli β-glucuronidase gene as a gene fusion marker for studies of gene

expression in higher plants.
13. Kung, C, and Blount, P (2004). Channels in microbes: so many holes to fill. Mol Microbiol 53, 373–380.
14. Landberg, K, Pederson, E, TV-P, and 2013, undefined The moss Physcomitrella patens reproductive organ development is highly organized, affected by the two SHI/STY genes and by the level of active auxin in the SHI. AcademicOupCom.
15. Lang, D, Zimmer, AD, Rensing, SA, and Reski, R (2008). Exploring plant biodiversity: the Physcomitrella genome and beyond. Trends Plant Sci 13, 542–549.
16. Levina, N, Tötemeyer, S, Stokes, NR, Louis, P, Jones, MA, and Booth, IR (1999). Protection of Escherichia coli cells against extreme turgor by activation of MscS and MscL mechanosensitive channels: identification of genes required for MscS activity. EMBO J 18, 1730–1737.
17. Martinac, B, Buechner, M, Delcour, AH, Adler, J, and Kung, C (1987). Pressure-sensitive ion channel in Escherichia coli. Proc Natl Acad Sci U S A 84, 2297–2301.
18. Matasci, N et al. (2014). Data access for the 1,000 Plants (1KP) project. Gigascience 3.
19. Medina, R, Johnson, MG, Liu, Y, Wickett, NJ, Shaw, AJ, and Goffinet, B (2019). Phylogenomic delineation of Physcomitrium (Bryophyta: Funariaceae) based on targeted sequencing of nuclear exons and their flanking regions rejects the retention. Wiley Online Libr 57, 404–417.
20. Mohanasundaram, B, Bhide, AJ, Palit, S, Chaturvedi, G, Lingwan, M, Masakapalli, SK, and Banerjee, AK (2021). The unique bryophyte-specific repeat-containing protein SHORT-LEAF regulates gametophore development in moss. Plant Physiol 187, 203–217.
21. Nishiyama, T et al. (2003). Comparative genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: Implication for land plant evolution. Proc Natl Acad Sci U S A 100, 8007.
22. Perroud, PF et al. (2018). The Physcomitrella patens gene atlas project: large-scale RNA-seq based expression data. Plant J 95, 168–182.
23. Peyronnet, R, Tran, D, Girault, T, and Frachisse, JM (2014). Mechanosensitive

channels: Feeling tension in a world under pressure. Front Plant Sci 5, 558.

24. Prigge, MJ, and Bezanilla, M (2010). Evolutionary crossroads in developmental biology: Physcomitrella patens. Development 137, 3535–3543.
25. Rensing, SA, Goffinet, B, Meyberg, R, Wu, SZ, and Bezanilla, M (2020). The Moss Physcomitrium (Physcomitrella) patens: A Model Organism for Non-Seed Plants. Plant Cell 32, 1361–1376.
26. Renzaglia, KS, and Garbary, DJ (2001). Motile gametes of land plants: Diversity, development, and evolution. CRC Crit Rev Plant Sci 20, 107–213.
27. Schmittgen, TD, and Livak, KJ (2008). Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc 3, 1101–1108.
28. W, D, Y, W, and H, T (2019). CLE-CLAVATA1 Signaling Pathway Modulates Lateral Root Development under Sulfur Deficiency. Plants (Basel, Switzerland) 8.
29. Xu, B et al. (2014). Contribution of NAC transcription factors to plant adaptation to land. Science (80-) 343, 1505–1508.

Appendix

Appendix 1: Different media compositions used to culture and transform *P.patens*

Stock	**Composition**	**Final Concentration in media**	**Volume in 100mL media**
BCDAT media			
Stock B	$MgSO_4.7H_2O$	1mM	1mL
Stock C	KH_2PO_4 (adjust pH to 6.5 using 4M KOH)	1.84mM	1mL
Stock D1	KNO_3	10mM	1mL
Stock D2	$FeSO_4.7H_2O$	45µM	1mL
Stock AT	$C_4H_{12}N_2O_6$	5mM	1mL
Stock TES	$Al_2K(SO_4)_3.K_2SO_4.24H_2O$ $CoCl_2.6H_2O$ $CuSO_4.5H_2O$ H_3BO_3 KBR KI LiCl $MnCl_2.4H_2O$ $SnCl_2.2H_2O$ $ZnSO_4.7H_2O$	0.006% (w/v) 0.006% (w/v) 0.006% (w/v) 0.061% (w/v) 0.003% (w/v) 0.003% (w/v) 0.003% (w/v) 0.003% (w/v) 0.003% (w/v) 0.006% (w/v)	100µL
Agar	-	0.8%	0.8g
Distilled H_2O	-	-	To 100mL
PRML			
Stock B	-	-	1mL
Stock C	-	-	1mL
Stock D1	-	-	1mL
Stock D2	-	-	1mL

Stock AT	-	-	1mL
Stock TES	-	-	100µL
D mannitol	-	6%	6g
Agar	-	0.8%	0.8g
Distilled H_2O	-	-	To 100mL
PRMT			
Stock B	-	-	1mL
Stock C	-	-	1mL
Stock D1	-	-	1mL
Stock D2	-	-	1mL
Stock AT	-	-	1mL
Stock TES	-	-	100µL
D-mannitol	-	6% (w/v)	6g
Agar	-	0.4% (w/v)	0.4g
Distilled H_2O	-	-	To 100mL
PRMB			
Stock B	-	-	1mL
Stock C	-	-	1mL
Stock D1	-	-	1mL
Stock D2	-	-	1mL
Stock AT	-	-	1mL
Stock TES	-	-	100µL
D-mannitol	-	6% (w/v)	6g
Agar	-	0.56% (w/v)	0.56g
Distilled H_2O	-	-	To 100mL
MMg solution			
D-mannitol	-	9.1%	405mg
MES	-	10%	500µL
$MgCl_2$	-	15mM	75µL
Distilled H_2O	-	-	To 5mL

PEG-Ca			
$Ca(NO_3)_2$	1M	0.1M	500µL
Tris (pH 8)	1M	0.01M	50µL
PEG 6000	-	-	2g
Mannitol	8%	-	4.5mL

Appendix 2: RPKM values for 16 PpMSLs under normal, dehydration, and rehydration conditions.

Gene ID	RPKM (normal conditions)	RPKM (dehydrated conditions)	RPKM (rehydrated conditions)
Pp3c1_10620	14.874	53.375	54.091
Pp3c1_35260	17.7	18.771	17.918
Pp3c3_38000	8.001	5.957	6.609
Ppc3c3_12760	2.098	0.072	0.028
Pp3c7_19980	0	0	0
Pp3c9_20120	9.425	31.584	32.121
Pp3c9_20150	7.274	0.199	0.184
Pp3c10_8380	47.04	0.033	0.09
Pp3c10_30950	0.026	5.001	5.678
Pp3c11_3390	0.745	23.603	22.054
Pp3c13_20730	0.124	0.016	0.022
Pp3c14_810	1.029	0.023	0.021
Pp3c14_8610	4.8	44.326	42.855
Pp3c15_19020	0.114	125.421	120.671
Pp3c26_12710	0.135	48.732	32.915
Pp3c27_4760	2.156	3.134	2.641

Appendix 3: Description of all dehydration-responsive motifs present on the *PpMSL* promoters*

Motif Name	Motif Sequence	Description
ABREATRD22	RYACGTGGYR	ABRE (ABA-responsive element) in Arabidopsis dehydration-responsive gene rd22; R=A/G; Y=C/T.
ABRELATERD1	ACGTG	ABRE-like sequence (from -199 to -195) required for etiolation-induced expression of erd1 (early responsive to dehydration) in Arabidopsis.
ACGTATERD1	ACGT	ACGT sequence (from -155 to -152) required for etiolation-induced expression of erd1 (early responsive to dehydration) in Arabidopsis.
CBFHV	RYCGAC	The binding site of barley (H.v.) CBF1, and also of barley CBF2; CBF = C-repeat (CRT) binding factors; CBFs are also known as dehydration-responsive element (DRE) binding proteins (DREBs); (SQ=GTCGAC); R=A/G; Y=C/T.
LTRECOREATCOR15	CCGAC	The core of low-temperature responsive element (LTRE) of cor15a gene in Arabidopsis (A.t.); A portion of repeat-C (C-repeat), TGGCCGAC, which is repeated twice in cor15a promoter; ABA responsiveness; Involved in cold induction of BN115 gene from winter Brassica napus; LTRE; Light signalling mediated by phytochrome is necessary for cold- or

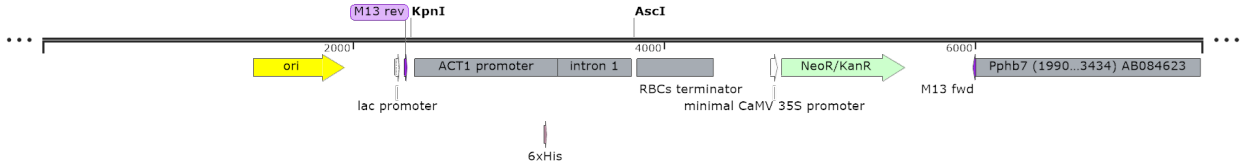
		drought-induced gene expression through the C/DRE in Arabidopsis.
MYB1AT	WAACCA	MYB recognition site found in the promoters of the dehydration-responsive gene rd22 and many other genes in Arabidopsis; W=A/T.
MYB2CONSENSUSAT	YAACKG	MYB recognition site found in the promoters of the dehydration-responsive gene rd22 and many other genes in Arabidopsis; Y=C/T; K=G/T.
MYBATRD22	CTAACCA	The binding site for MYB (ATMYB2) in the dehydration-responsive gene, rd22; MYB binding site in rd22 gene of Arabidopsis thaliana; ABA-induction; Located at ca. -141 of rd22 gene; Also MYC at ca. -200 of rd22 gene.
MYCATERD1	CATGTG	MYC recognition sequence (from -466 to -461) necessary for expression of erd1 (early responsive to dehydration) in dehydrated Arabidopsis; NAC protein bound specifically to the CATGTG motif; NAC protein bound specifically to the CATGTG motif.
MYCATRD22	CACATG	The binding site for MYC (rd22BP1) in Arabidopsis dehydration-responsive gene, rd22; MYC binding site in rd22 gene of Arabidopsis thaliana; ABA-induction; Located at ca. -200 of rd22 gene; Also MYB at ca. -141 of rd22 gene.

MYCCONSENSUSAT	CANNTG	MYC recognition site found in the promoters of the dehydration-responsive gene rd22 and many other genes in Arabidopsis; Binding site of ATMYC2 (previously known as rd22BP1); see S000144 (E-box; CANNTG); N=A/T/G/C; MYC recognition sequence in CBF3 promoter; Binding site of ICE1 (inducer of CBF expression 1) that regulates the transcription of CBF/DREB1 genes in the cold in Arabidopsis; ICE1; This sequence is also known as RRE (R response element).

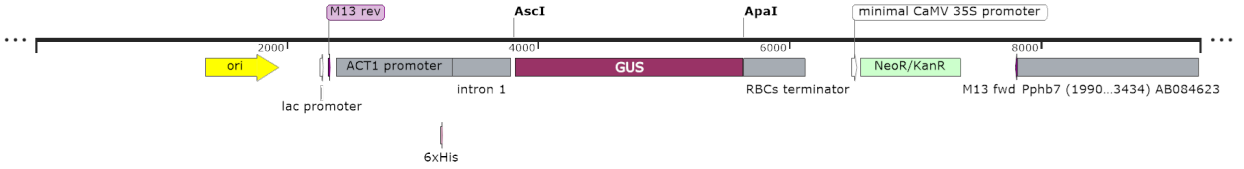
* source - New Place (https://www.dna.affrc.go.jp/PLACE/).

Appendix 4: Vector maps of all constructs

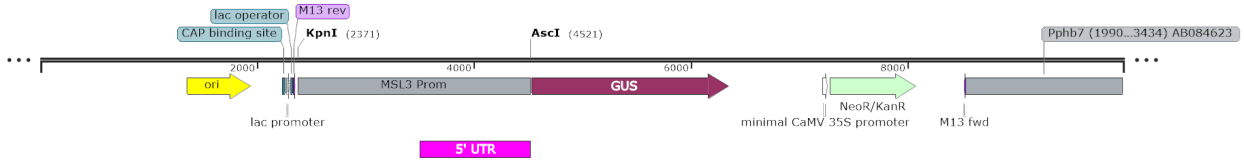
1. pTFH15.3



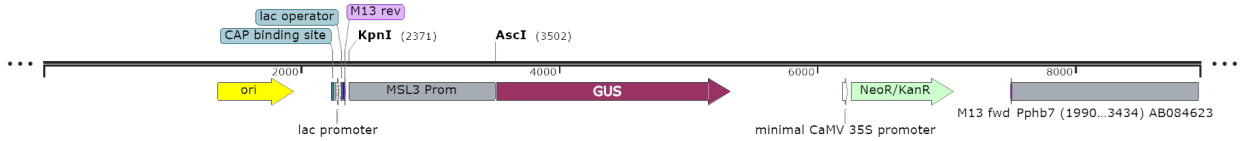
2. pTFH15.3:GUS



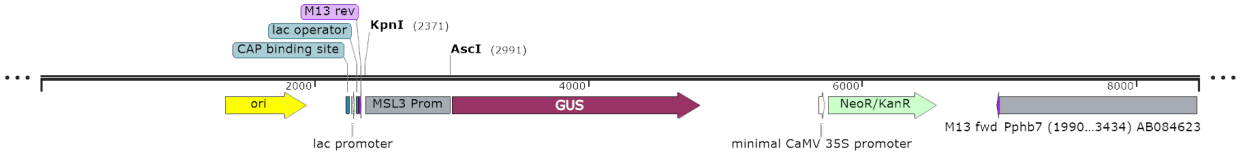
3. pTFH15.3:MSL3_Promoter:GUS



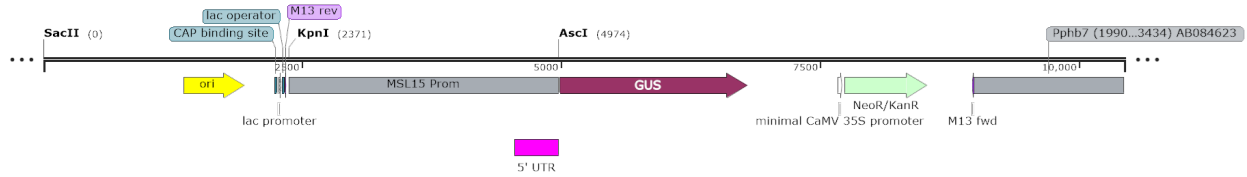
4. pTFH15.3:MSL3_Promoter_1Kb:GUS



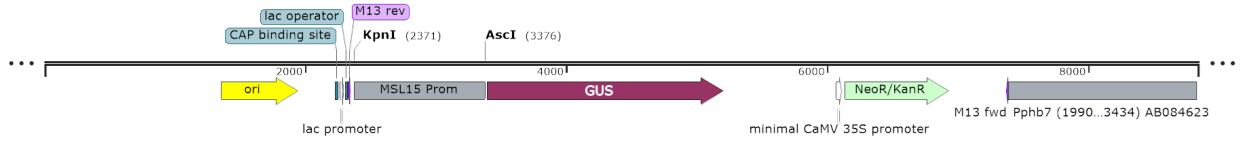
5. pTFH15.3:MSL3_Promoter_600bp:GUS



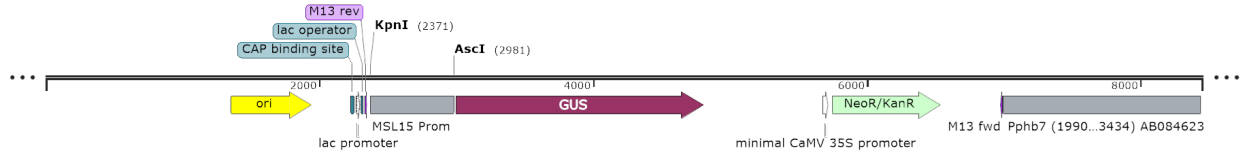
6. pTFH15.3:MSL15_Promoter:GUS



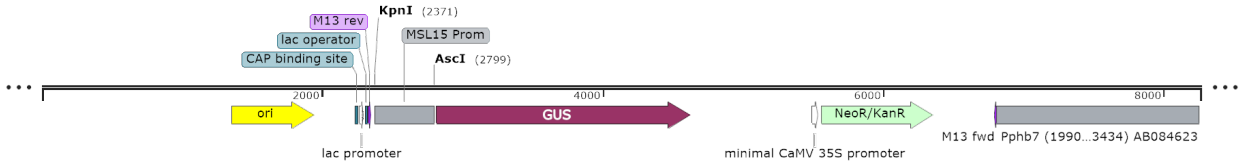
7. pTFH15.3:MSL15_Promoter_1KB:GUS



8. pTFH15.3:MSL15_Promoter_600bp:GUS



9. pTFH15.3:MSL15_Promoter_5' UTR:GUS



Appendix 5: Sequencing results for all eight *PpMSL* full-length and deletion construct clones

1. pTFH15.3:GUS

```
Query  1     ATGTTACGTCCTGTAGAAACCCCAACCCGTGAAATCAAAAAACTCGACGGCCTGTGGGCA  60
Sbjct  1     ............................................................  60

Query  61    TTCAGTCTGGATCGCGAAAACTGTGGAATTGATCAGCGTTGGTGGGAAAGCGCGTTACAA  120
Sbjct  61    ............................................................  120

Query  121   GAAAGCCGGGCAATTGCTGTGCCAGGCAGTTTTAACGATCAGTTCGCCGATGCAGATATT  180
Sbjct  121   ............................................................  180

Query  181   CGTAATTATGCGGGCAACGTCTGGTATCAGCGCGAAGTCTTTATACCGAAAGGTTGGGCA  240
Sbjct  181   ............................................................  240

Query  241   GGCCAGCGTATCGTGCTGCGTTTCGATGCGGTCACTCATTACGGCAAAGTGTGGGTCAAT  300
Sbjct  241   ............................................................  300

Query  301   AATCAGGAAGTGATGGAGCATCAGGGCGGCTATACGCCATTTGAAGCCGATGTCACGCCG  360
Sbjct  301   ............................................................  360

Query  361   TATGTTATTGCCGGGAAAAGTGTACGTATCACCGTTTGTGTGAACAACGAACTGAACTGG  420
Sbjct  361   ............................................................  420

Query  421   CAGACTATCCCGCCGGGAATGGTGATTACCGACGAAAACGGCAAGAAAAAGCAGTCTTAC  480
Sbjct  421   ............................................................  480

Query  481   TTCCATGATTTCTTTAACTATGCCGGAATCCATCGCAGCGTAATGCTCTACACCACGCCG  540
Sbjct  481   ............................................................  540

Query  541   AACACCTGGGTGGACGATATCACCGTGGTGACGCATGTCGCGCAAGACTGTAACCACGCG  600
Sbjct  541   ............................................................  600

Query  601   TCTGTTGACTGGCAGGTGGTGGCCAATGGTGATGTCAGCGTTGAACTGCGTGATGCGGAT  660
Sbjct  601   ............................................................  660

Query  661   CAACAGGTGGTTGCAACTGGACAAGGCACTAGCGGGACTTTGCAAGTGGTGAATCCGCAC  720
Sbjct  661   ............................................................  720

Query  721   CTCTGGCAACCGGGTGAAGGTTATCTCTATGAACTGTGCGTCACAGCCAAAAGCCAGACA  780
Sbjct  721   ............................................................  780

Query  781   GAGTGTGATATCTACCCGCTTCGCGTCGGCATCCGGTCAGTGGCAGTGAAGGGCGAACAG  840
Sbjct  781   ............................................................  840

Query  841   TTCCTGATTAACCACAAACCGTTCTACTTTACTGGCTTTGGTCGTCATGAAGATGCGGAC  900
Sbjct  841   ............................................................  900

Query  901   TTGCGTGGCAAAGGATTCGATAACGTGCTGATGGTGCACGACCACGCATTAATGGACTGG  960
Sbjct  901   ............................................................  960

Query  961   ATTGGGGCCAACTCCTACCGTACCTCGCATTACCCTTACGCTGAAGAGATGCTCGACTGG  1020
Sbjct  961   ............................................................  1020

Query  1021  GCAGATGAACATGGCATCGTGGTGATTGATGAAACTGCTGCTGTCGGCTTTAACCTCTCT  1080
Sbjct  1021  ............................................................  1080

Query  1081  TTAGGCATTGGTTTCGAAGCGGGCAACAAGCCGAAAGAACTGTACAGCGAAGAGGCAGTC  1140
Sbjct  1081  ............................................................  1140

Query  1141  AACGGGGAAACTCAGCAAGCGCACTTACAGGCGATTAAAGAGCTGATAGCGCGTGACAAA  1200
Sbjct  1141  ............................................................  1200

Query  1201  AACCACCCAAGCGTGGTGATGTGGAGTATTGCCAACGAACCGGATACCCGTCCGCAAGGT  1260
Sbjct  1201  ............................................................  1260

Query  1261  GCACGGGAATATTTCGCGCCACTGGCGGAAGCAACGCGTAAACTCGACCCGACGCGTCCG  1320
Sbjct  1261  ............................................................  1320

Query  1321  ATCACCTGCGTCAATGTAATGTTCTGCGACGCTCACACCGATACCATCAGCGATCTCTTT  1380
Sbjct  1321  ............................................................  1380

Query  1381  GATGTGCTGTGCCTGAACCGTTATTACGGATGGTATGTCCAAAGCGGCGATTTGGAAACG  1440
Sbjct  1381  ............................................................  1440

Query  1441  GCAGAGAAGGTACTGGAAAAAGAACTTCTGGCCTGGCAGGAGAAACTGCATCAGCCGATT  1500
Sbjct  1441  ............................................................  1500

Query  1501  ATCATCACCGAATACGGCGTGGATACGTTAGCCGGGCTGCACTCAATGTACACCGACATG  1560
Sbjct  1501  ............................................................  1560

Query  1561  TGGAGTGAAGAGTATCAGTGTGCATGGCTGGATATGTATCACCGCGTCTTTGATCGCGTC  1620
Sbjct  1561  ............................................................  1620

Query  1621  AGCGCCGTCGTCGGTGAACAGGTATGGAATTTCGCCGATTTTGCGACCTCGCAAGGCATA  1680
Sbjct  1621  ............................................................  1680

Query  1681  TTGCGCGTTGGCGGTAACAAGAAAGGGATCTTCACTCGCGACCGCAAACCGAAGTCGGCG  1740
Sbjct  1681  ............................................................  1740

Query  1741  GCTTTTCTGCTGCAAAAACGCTGGACTGGCATGAACTTCGGTGAAAAACCGCAGCAGGGA  1800
Sbjct  1741  ............................................................  1800

Query  1801  GGCAAACAATGA  1812
Sbjct  1801  ............  1812
```

2. *PpMSL3* promoter

(i) Forward reaction

```
Query  42    CTTGTTCAGGGATATGGGGTTGACAACTGCGCGCCACCGCCTTGCCCTGTATGTACCTTT  101
Sbjct  1     ............................................................  60

Query  102   AGGAAGTCTTGCTGTTCTAATTGCAGATTTGTCTGCTAGAAAATTACCCTTGATAGGTCC  161
Sbjct  61    ............................................................  120

Query  162   CATGTATCGTGCTGGAGGACCGTTGATGCCAGGTTTGCTTCTTGGCTCTGTGGTTGGAGT  221
Sbjct  121   ............................................................  180

Query  222   TGTCGTCTCCACACAATTCAACTTGAGCTGAAGCACAGAATGTTTTATACTTGTAACATA  281
Sbjct  181   ............................................................  240

Query  282   TCATCTCTGGACATAGGACACATGGGGCTTCTCTTTGGCATATGTTTGTAAGTGCCAAGC  341
Sbjct  241   ............................................................  300

Query  342   AATTCTTATGTCCTGTATATTCATTGCGTTAGACAACGATTTTGCCTTGCCAACTTCAAG  401
Sbjct  301   ............................................................  360

Query  402   TTGAGGTACCTAGGTGACCTTAAAAGGTGCACTCTGAACCACAAGATTGTGGTGCTAATT  461
Sbjct  361   ............................................................  420

Query  462   TGCCCTTCTTGCATTAAAAGATTTCATTGTGGAGAAGATCTTTTGAGAACGTCGACTGCA  521
Sbjct  421   ............................................................  480

Query  522   GCCCTTCGTATGTATTTTCGATACGAAATGTTCAATAAGATTGAATGCTTCGAAGGGGGA  581
Sbjct  481   ............................................................  540

Query  582   CCTCATGGTATATTTCTGTATAGCACAACTTGCATGGTTGTTACTGGGAGTCCACCAACA  641
Sbjct  541   ............................................................  600

Query  642   TCATTCGTCAATCTGTGATTGATTTTTCAAAGGTAGTTTGCTTTCCGACTGTTTACTCTG  701
Sbjct  601   ............................................................  660

Query  702   GCTGCCGTGCATACGTAGGTCAAGTTTAGAGATTCTGCAGGTCTCAGAATACTGTCGACT  761
Sbjct  661   ............................................................  720

Query  762   ACAGAATCAAGCATCGAAACGGAACGATCTGATCTAAGGCTGTGCCACATTATTGTTCAG  821
Sbjct  721   ............................................................  780

Query  822   GGAAAAACCAGCTTATCAAGTGTTAGCATATCCTAATGCAATTCGATTTTCCATTTGTGT  881
Sbjct  781   ............................................................  840

Query  882   GGAAAATATCGCGTAATATAGCAACGTTACCTGCTTTCTTTGCTTGCGATGTGCTTGCTA  941
Sbjct  841   ............................................................  900

Query  942   GGCTGAAGAAAGATATTTGCGCTCGAGTCaaaaaaaaCATATCGTCTAGGGCAACCTTAG  1001
Sbjct  901   ............................................................  960

Query  1002  ATTTTCATACTTCCACCCAGGATATTATGTCCAGCCTTCAAGCTCAGCCACACCCTTCAA  1061
Sbjct  961   ............................................................  1020

Query  1062  GTCAATGATGATGCAAACTTCGGCAATATGTGGTAATACAATCTATCCGCATGGAGCACC  1121
Sbjct  1021  ............................................................  1080

Query  1122  TGATCCAGCAATAGGAATATGGATGGGCGTTGTCTAAGATGCCGTGCGACTCTACCAGTT  1181
Sbjct  1081  ............................................................  1140
```

(ii) Reverse reaction

```
Query  32    CGCCAATGTTCACTTTACCCAATCAAGCAGCAACTGGGTTCGACCTTGTCATTACTACGC  91
Sbjct  2517  ............................................................  2458

Query  92    TACAGTCGCAACTAAGGGATCGCTCATGGCAGTATCGAAGTTCCAAATAGAAAAAGTATA  151
Sbjct  2457  ............................................................  2398

Query  152   CATTCTGAACAACTTGCTCAAATCTCATTTACTCTGAGTACGAGTTGGGGCTCAATCTAT  211
Sbjct  2397  ............................................................  2338

Query  212   GCTACATAATTTTGTAAATCGTTGGATAAAACAGCGGTGGCTCCTTGTTTCTGCGGGTGG  271
Sbjct  2337  ............................................................  2278

Query  272   CTTGATTTAAATTCAAATGTAAACAGTAGGTGCAGCGAGGACAAGGTCACTCGGTCGACT  331
Sbjct  2277  ............................................................  2218

Query  332   TGTAGATGCAAAGGCTTAATCAGTCGCTGCGCGCGACGACTGGCTGGGCAACCTCGGCTC  391
Sbjct  2217  ............................................................  2158

Query  392   CGCTAAGCACCCTGTTTCAAATCAAACAATAGCTCACTCATTCTCTAAACAAGCCCTGCA  451
Sbjct  2157  ............................................................  2098

Query  452   CTGGTTCCGAGAATCAATCTTCCACTACATACATAAGTACAATAAACCACACACATTGAACT  511
Sbjct  2097  ............................................................  2038

Query  512   ACGAAGCATCCATTACGTTATGTTCAACTCCCCAAATGTTACGACGAACAATTGAAACAA  571
Sbjct  2037  ............................................................  1978

Query  572   GAGTCTCATAAACACACAACAGCTTAAGCAAGCTGCACTTGGCTAAACAACCGTCAGGTG  631
Sbjct  1977  ............................................................  1918

Query  632   CGCAGGGCTGCCCTACGATCAGAAGGGGAGCATTGTTCTCCACTGTTCGTACAGCGATGA  691
Sbjct  1917  ............................................................  1858

Query  692   TTTGTGTTCAAGAAGGGACAACTGCATGCCAGAGCATCTCAGTGCCTTCGACGCCACCAA  751
Sbjct  1857  ............................................................  1798

Query  752   ATCGTGCTCGAGGTGAAGCCTTGAAGCCATATCATTTCGTAATCGACGTCGTATTTCGCA  811
Sbjct  1797  ............................................................  1738

Query  812   TGCGAAGTACTTTCATGGAGGACGCAACTATGCTAATTACGCACCCCCAAGAATCACGTG  871
Sbjct  1737  ............................................................  1678

Query  872   GACGGGTGGGGGTGGGCGGGGCAATGTTGCTCAGAGAGACGAAAAGGCTGAAACATCTGT  931
Sbjct  1677  ............................................................  1618

Query  932   GGAGTTTGACAATTGAAAGCTGGCAACAGAGTACCGATAAACCGGAGGTACAACATCAGA  991
Sbjct  1617  ............................................................  1558

Query  992   TCATCACATGATGGGCTCAATAAATAACTCTACACTACCTTTTCTCGAGTATTTTCGACA  1051
Sbjct  1557  ............................................................  1498

Query  1052  AGTTAAACAAAACCTGCAAGATTCTTATCATGTAACATCTTCCATTTATCTTCCAAA-CT  1110
Sbjct  1497  ............................................................A..  1438

Query  1111  ATCTTCCAGCATCATGATCCGACCAAGATCTCTCATTCGCTCTCGTT  1157
Sbjct  1437  ............................................  1391
```

3. *PpMSL15* promoter

(i) Forward reaction

```
Query  38    TTACGAATTCGAGCTCGGTACCGGACATTTTGTTAAAAAGAAATAAGTATAAGATGAAAT  97
Sbjct  4     ............................................................  63

Query  98    TTGATAAGAATAATAATTATTAGAGttattaaatattaatgaatttaaatttttttctac  157
Sbjct  64    ............................................................  123

Query  158   aatatatattttattaaataatttattagttatatataatatttggtcaatatttattat  217
Sbjct  124   ............................................................  183

Query  218   tttagtataatattattaataaaattcaagattttaaaagtaattctttaataCTTGATT  277
Sbjct  184   ............................................................  243

Query  278   ATAGCTTAATATAATTTACATTACACAAGTAATGtatatatatatatatatatatatatG  337
Sbjct  244   ............................................................  303

Query  338   GAAATTGATACGTTGATTGGAATTTTTACTAAATTGACCACGTTTCAGTCAATTAAAGCA  397
Sbjct  304   ............................................................  363

Query  398   TATACACTCATTACAGCATGCATGGCTAAATCCATGCTACACGCCATAGTCCTAAAAGCA  457
Sbjct  364   ............................................................  423

Query  458   CAAAAATATTAAGAACGGTGATCTAATTTTACAAATCCTTTGAGTCATGGACAAGGCAAA  517
Sbjct  424   ............................................................  483

Query  518   TGCtataaaaataattatattatataaaatataaaaaatatttttttttGACGAATAATA  577
Sbjct  484   ............................................................  543

Query  578   CAACAACAATTTGTTTCTAAAAGaaaaaaaaGATATAAAGTGGTTTTACAAAAACTTAAA  637
Sbjct  544   ............................................................  603

Query  638   ATATTTTTTGAAATTTTGTAATTATCATAAAATCAACTACCATTAGATAATTAAGTACCT  697
Sbjct  604   ............................................................  663

Query  698   AATTAAAATTGATTTGATGTGATAAATTAATTTAAAGAACATAAATTACTTTTGGTTTTA  757
Sbjct  664   ............................................................  723

Query  758   TAATTGATGGTGTATAGCTTAGACTGTTTCTTTTATACACAATATGTAGTTTTTATGAAT  817
Sbjct  724   ............................................................  783

Query  818   TTTAATATACTTACTTGCAATCATAGTTTGTAGTTGGTTTACATTACTATCAATTTATTT  877
Sbjct  784   ............................................................  843

Query  878   ATGAACCTTTTGTAAATTCTATAGTATTTATGTATACTAAGTCaattatattttaaaaaa  937
Sbjct  844   ............................................................  903

Query  938   tttaactatattaattcgtgatattattttgatttagtaaatttataatatatgcttaaa  997
Sbjct  904   ............................................................  963
```

(ii) Reverse reaction

```
Query  32    GACGTAACATCTACTCTTCAAATACTGTACTCCCTCCCTCTTGGCACTGGGTCTCTGATA  91
Sbjct  2635  ............................................................  2576

Query  92    CCCTCCCAGTCAATCGGATGGCTCGGTTCAGACTCTCTGCAACACCCTTGAACAAATTGT  151
Sbjct  2575  ............................................................  2516

Query  152   GTGTCGACTCAAACCAAAGTCACTGTCTAGAGATGAGAGACGGATCCAGGGACAAGTAAG  211
Sbjct  2515  ............................................................  2456

Query  212   GTGACGATGTCGAAGCTCTGATACGGACCTGAAAAAGGATTCCCATGCGCTTCAGGATGG  271
Sbjct  2455  ............................................................  2396

Query  272   CGCAGACACTAGATTTTCGTACCTGATCGAATCACAAAACTGCAGcgtcgtcgtcgtcgt  331
Sbjct  2395  ............................................................  2336

Query  332   cgATATTACTAGCGAAAAATGAACCCCAACTTCGATCGATCGATCAATCTATGAGTTCTA  391
Sbjct  2335  ............................................................  2276

Query  392   ACTCCATCCATGGTATGGGTCCATGCTCCTGGATTTCGCAGGCCACTCTACATCAGGCGG  451
Sbjct  2275  ............................................................  2216

Query  452   CACTAACTAAGTGCGTGTCATCGCGTGAGAAATAACACACCAACATAACaaaaaataaaa  511
Sbjct  2215  ............................................................  2156

Query  512   aaaacaaagacaaaaaTAGACGTAAACACGTGGCAGGGATCCATGCGTCGTTGTCGACTT  571
Sbjct  2155  ............................................................  2096

Query  572   GGGATGACTAGCGTGAGACACGTGTACGGTAGGCAGTGTGTGACTTGGCCCACAAGTGGT  631
Sbjct  2095  ............................................................  2036

Query  632   AGGAAACTGGAAACGGGAAACTGTGGAACCCAGGTCATCAGATTTGTGAAGTACTCGTCC  691
Sbjct  2035  ............................................................  1976

Query  692   CGGATGCTAAAGCAGACATTTGACAGTTGGATCCTTGAGTTAGCTTGGGTGCAGCTTTCC  751
Sbjct  1975  ............................................................  1916

Query  752   GATGGCCAGGTTTCGTGCTGGAGCTTTCCTCGGGACCAATGCTGCTGCTAAAACCATGGA  811
Sbjct  1915  ............................................................  1856

Query  812   TGGATTAATGGATATATGGATGAATGGATTGGTTGGGAGGGGATAGAGTGGTGACAATTG  871
Sbjct  1855  ............................................................  1796
```

4. *PpMSL3* deletion constructs

(i) ΔM3-1

```
Query  19   CGTAGGGCAGCCCTGCGCACCTGACGGTTGTTTAGCCAAGTGCAGCTTGCTTAAGCTGTT  78
Sbjct  1    ............................................................  60

Query  79   GTGTGTTTATGAGACTCTTGTTTCAATTGTTCGTCGTAACATTTGGGGAGTTGAACATAA  138
Sbjct  61   ............................................................  120

Query  139  CGTAATGGATGCTTCGTAGTTCAATGTGTGGTTTATTGTACTTATGTATGTAGTGGAAGA  198
Sbjct  121  ............................................................  180

Query  199  TTGATTCTCGGAACCAGTGCAGGGCTTGTTTAGAGAATGAGTGAGCTATTGTTTGATTTG  258
Sbjct  181  ............................................................  240

Query  259  AAACAGGGTGCTTAGCGGAGCCGAGGTTGCCCAGCCAGTCGTCGCGCGCAGCGACTGATT  318
Sbjct  241  ............................................................  300

Query  319  AAGCCTTTGCATCTACAAGTCGACCGAGTGACCTTGTCCTCGCTGCACCTACTGTTTACA  378
Sbjct  301  ............................................................  360

Query  379  TTTGAATTTAAATCAAGCCACCCGCAGAAACAAGGAGCCACCGCTGTTTTATCCAACGAT  438
Sbjct  361  ............................................................  420

Query  439  TTACAAAATTATGTAGCATAGATTGAGCCCCAACTCGTACTCAGAGTAAATGAGATTTGA  498
Sbjct  421  ............................................................  480

Query  499  GCAAGTTGTTCAGAATGTATACTTTTTCTATTTGGAACTTCGATACTGCCATGAGCGATC  558
Sbjct  481  ............................................................  540

Query  559  CCTTAGTTGCGACTGTAGCGTAGTAATGACAAGGTCGAACCCAGTTGCTGCTTGATTGGG  618
Sbjct  541  ............................................................  600

Query  619  TAAAGTGAACATTGGCG  635
Sbjct  601  .................  617
```

(ii) ΔM3-2

```
Query  19    GAACGAGAGCGAATGAGAGATCTTGGTCGGATCATGATGCTGGAAGATAGTTTTGGAAGA  78
Sbjct  1     ............................................................  60

Query  79    TAAATGGAAGATGTTACATGATAAGAATCTTGCAGGTTTTGTTTAACTTGTCGAAAATAC  138
Sbjct  61    ............................................................  120

Query  139   TCGAGAAAAGGTAGTGTAGAGTTATTTATTGAGCCCATCATGTGATGATCTGATGTTGTA  198
Sbjct  121   ............................................................  180

Query  199   CCTCCGGTTTATCGGTACTCTGTTGCCAGCTTTCAATTGTCAAACTCCACAGATGTTTCA  258
Sbjct  181   ............................................................  240

Query  259   GCCTTTTCGTCTCTCTGAGCAACATTGCCCCGCCCACCCCCACCCGTCCACGTGATTCTT  318
Sbjct  241   ............................................................  300

Query  319   GGGGGTGCGTAATTAGCATAGTTGCGTCCTCCATGAAAGTACTTCGCATGCGAAATACGA  378
Sbjct  301   ............................................................  360

Query  379   CGTCGATTACGAAATGATATGGCTTCAAGGCTTCACCTCGAGCACGATTTGGTGGCGTCG  438
Sbjct  361   ............................................................  420

Query  439   AAGGCACTGAGATGCTCTGGCATGCAGTTGTCCCTTCTTGAACACAAATCATCGCTGTAC  498
Sbjct  421   ............................................................  480

Query  499   GAACAGTGGAGAACAATGCTCCCCTTCTGATCGTAGGGCAGCCCTGCGCACCTGACGGTT  558
Sbjct  481   ............................................................  540

Query  559   GTTTAGCCAAGTGCAGCTTGCTTAAGCTGTTGTGTGTTTATGAGACTCTTGTTTCAATTG  618
Sbjct  541   ............................................................  600

Query  619   TTCGTCGTAACATTTGGGGAGTTGAACATAACGTAATGGATGCTTCGTAGTTCAATGTGT  678
Sbjct  601   ............................................................  660

Query  679   GGTTTATTGTACTTATGTATGTAGTGGAAGATTGATTCTCGGAACCAGTGCAGGGCTTGT  738
Sbjct  661   ............................................................  720

Query  739   TTAGAGAATGAGTGAGCTATTGTTTGATTTGAAACAGGGTGCTTAGCGGAGCCGAGGTTG  798
Sbjct  721   ............................................................  780

Query  799   CCCAGCCAGTCGTCGCGCGCAGCGACTGATTAAGCCTTTGCATCTACAAGTCGACCGAGT  858
Sbjct  781   ............................................................  840

Query  859   GACCTTGTCCTCGCTGCACCTACTGTTTACATTTGAATTTAAATCAAGCCACCCGCAGAA  918
Sbjct  841   ............................................................  900

Query  919   ACAAGGAGCCACCGCTGTTTTATCCAACGATTTACAAAATTATGTAGCATAGATTGAGCC  978
Sbjct  901   ............................................................  960

Query  979   CCAACTCGTACTCAGAGTAAATGAGATTTGAGCAAGTTGTTCAGAATGTATACTTTTTCT  1038
Sbjct  961   ............................................................  1020

Query  1039  ATTTGGAACTTCGATACTGCCATGAGCGATCCCTTAGTTGCGACTGTAGCGTAGTAATGA  1098
Sbjct  1021  ............................................................  1080

Query  1099  CAAGGTCGAACCCAGTTGCTGCTTGATTGGGTAAAGTGAACATTGGCG  1146
Sbjct  1081  ................................................  1128
```

5. PpMSL15 deletion constructs

(i) ΔM15-1

```
Query  22   CGCACTTAGTTAGTGCCGCCTGATGTAGAGTGGCCTGCGAAATCCAGGAGCATGGACCCA  81
Sbjct  1    ............................................................  60

Query  82   TACCATGGATGGAGTTAGAACTCATAGATTGATCGATCGATCGAAGTTGGGGTTCATTTT  141
Sbjct  61   ............................................................  120

Query  142  TCGCTAGTAATATcgacgacgacgacgacgCTGCAGTTTTGTGATTCGATCAGGTACGAA  201
Sbjct  121  ............................................................  180

Query  202  AATCTAGTGTCTGCGCCATCCTGAAGCGCATGGGAATCCTTTTTCAGGTCCGTATCAGAG  261
Sbjct  181  ............................................................  240

Query  262  CTTCGACATCGTCACCTTACTTGTCCCTGGATCCGTCTCTCATCTCTAGACAGTGACTTT  321
Sbjct  241  ............................................................  300

Query  322  GGTTTGAGTCGACACACAATTTGTTCAAGGGTGTTGCAGAGAGTCTGAACCGAGCCATCC  381
Sbjct  301  ............................................................  360

Query  382  GATTGACTGGGAGGGTATCAGAGACCCAGTGCCAAGAGGGAGGGAGTACAGTATTTGAAG  441
Sbjct  361  ............................................................  420

Query  442  AGTAG  446
Sbjct  421  .....  425
```

(ii) ΔM15-2

```
Query  18   CCGTTTCCAGTTTCCTACCACTTGTGGGCCAAGTCACACACTGCCTACCGTACACGTGTC  77
Sbjct  2    ............................................................  61

Query  78   TCACGCTAGTCATCCCAAGTCGACAACGACGCATGGATCCCTGCCACGTGTTTACGTCTA  137
Sbjct  62   ............................................................  121

Query  138  ttttgtctttgtttttttattttttGTTATGTTGGTGTGTTATTTCTCACGCGATGAC  197
Sbjct  122  ............................................................  181

Query  198  ACGCACTTAGTTAGTGCCGCCTGATGTAGAGTGGCCTGCGAAATCCAGGAGCATGGACCC  257
Sbjct  182  ............................................................  241

Query  258  ATACCATGGATGGAGTTAGAACTCATAGATTGATCGATCGATCGAAGTTGGGGTTCATTT  317
Sbjct  242  ............................................................  301

Query  318  TTCGCTAGTAATATcgacgacgacgacgacgCTGCAGTTTTGTGATTCGATCAGGTACGA  377
Sbjct  302  ............................................................  361

Query  378  AAATCTAGTGTCTGCGCCATCCTGAAGCGCATGGGAATCCTTTTTCAGGTCCGTATCAGA  437
Sbjct  362  ............................................................  421

Query  438  GCTTCGACATCGTCACCTTACTTGTCCCTGGATCCGTCTCTCATCTCTAGACAGTGACTT  497
Sbjct  422  ............................................................  481

Query  498  TGGTTTGAGTCGACACACAATTTGTTCAAGGGTGTTGCAGAGAGTCTGAACCGAGCCATC  557
Sbjct  482  ............................................................  541

Query  558  CGATTGACTGGGAGGGTATCAGAGACCCAGTGCCAAGAGGGAGGGAGTACAGTATTTGAA  617
Sbjct  542  ............................................................  601

Query  618  GAGTAG  623
Sbjct  602  ......  607
```

(iii) ΔM15-3

```
Query  16    GGT-ATCATGATGAATGTGGTTGGTTTTTATTGGTTGCAAATCCAATTGTGGACATATGA  74
Sbjct  1     ...A.........................................................  60

Query  75    ATGTAGAGGTCAttttttttCCTTTTCATCATGGTTAAAATGGGTGGGTCGATTTACTAA  134
Sbjct  61    ............................................................  120

Query  135   AAACTTGGTTCTATTAAACATTCTTATTTAATAAATAAAGGCAAGAAAAGGGCAATTGTC  194
Sbjct  121   ............................................................  180

Query  195   ACCACTCTATCCCCTCCCAACCAATCCATTCATCCATATATCCATTAATCCATCCATGGT  254
Sbjct  181   ............................................................  240

Query  255   TTTAGCAGCAGCATTGGTCCCGAGGAAAGCTCCAGCACGAAACCTGGCCATCGGAAAGCT  314
Sbjct  241   ............................................................  300

Query  315   GCACCCAAGCTAACTCAAGGATCCAACTGTCAAATGTCTGCTTTAGCATCCGGGACGAGT  374
Sbjct  301   ............................................................  360

Query  375   ACTTCACAAATCTGATGACCTGGGTTCCACAGTTTCCCGTTTCCAGTTTCCTACCACTTG  434
Sbjct  361   ............................................................  420

Query  435   TGGGCCAAGTCACACACTGCCTACCGTACACGTGTCTCACGCTAGTCATCCCAAGTCGAC  494
Sbjct  421   ............................................................  480

Query  495   AACGACGCATGGATCCCTGCCACGTGTTTACGTCTAttttgtctttgttttttttattt  554
Sbjct  481   ............................................................  540

Query  555   tttGTTATGTTGGTGTGTTATTTCTCACGCGATGACACGCACTTAGTTAGTGCCGCCTGA  614
Sbjct  541   ............................................................  600

Query  615   TGTAGAGTGGCCTGCGAAATCCAGGAGCATGGACCCATACCATGGATGGAGTTAGAACTC  674
Sbjct  601   ............................................................  660

Query  675   ATAGATTGATCGATCGATCGAAGTTGGGGTTCATTTTTCGCTAGTAATATcgacgacgac  734
Sbjct  661   ............................................................  720

Query  735   gacgacgCTGCAGTTTTGTGATTCGATCAGGTACGAAAATCTAGTGTCTGCGCCATCCTG  794
Sbjct  721   ............................................................  780

Query  795   AAGCGCATGGGAATCCTTTTTCAGGTCCGTATCAGAGCTTCGACATCGTCACCTTACTTG  854
Sbjct  781   ............................................................  840

Query  855   TCCCTGGATCCGTCTCTCATCTCTAGACAGTGACTTTGGTTTGAGTCGACACACAATTTG  914
Sbjct  841   ............................................................  900

Query  915   TTCAAGGGTGTTGCAGAGAGTCTGAACCGAGCCATCCGATTGACTGGGAGGGTATCAGAG  974
Sbjct  901   ............................................................  960

Query  975   ACCCAGTGCCAAGAGGGAGGGAGTACAGTATTTGAAGAGTAG  1016
Sbjct  961   ........................................  1002
```