Uniwersytet im. Adama Mickiewicza w Poznaniu

Wydział Biologii

Zakład Ekspresji Genów

Filogenetyczne i funkcjonalne analizy genów kodujących czynniki transkrypcyjne *SQUAMOSA PROMOTER BINDING-LIKE* u wątrobowca *Marchantia polymorpha*

Alisha Alisha

Rozprawa doktorska

Poznań, Polska 2023

Adam Mickiewicz University in Poznan

Faculty of Biology

Department of Gene Expression

Phylogenetic and functional studies of *SQUAMOSA PROMOTER BINDING-LIKE* transcription factor gene family members in the liverwort *Marchantia polymorpha*

Alisha Alisha

PhD thesis

Poznan, Poland 2023

FUNDING

PUBLICATIONS:

The results discussed in this thesis in Chapter 1 are presented in the following research article:

1. Comparative analysis of *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) gene family between bryophytes and seed plants in BioRxiv (DOI: 10.1101/2023.02.27.530190)
Authors: Alisha Alisha, Zofia Szweykowska-Kulińska, Izabela Sierocka

OTHER PUBLICATIONS:

1. A Review on "Plant conserved and non-conserved RNA-target modules network: Lessons for a better understanding of Marchantia development" in Plant Molecular Biology (Submitted: June, 2023)
Authors: Pietrykowska H, Alisha A, Aggarwal B, Watanabe Y, Ohtani M, Jarmolowski A, Sierocka I, Szweykowska-Kulinska Z

2. A Review on "Biogenesis, conservation, and function of miRNA in liverworts" in Journal of Experimental Botany (DOI: 10.1093/jxb/erac098)
Authors: Halina Pietrykowska, Izabela Sierocka, Andrzej Zielezinski, Alisha Alisha, Juan Carlo Carrasco-Sanchez, Artur Jarmolowski, Wojciech M Karlowski, Zofia Szweykowska-Kulinska

3. A Book chapter on "Advances and Applicability of Genotyping Technologies in Cotton Improvement" in Wiley Online Library (DOI: 10.1002/9781119745686.ch11)
Authors: Shubham Bhardwaj, Vikas Devkar, Amit Kumar, Alish Alisha, Shivani Sharma, Rupesh K. Deshmukh, Gunvant B. Patil

4. A Review on "Mutation Breeding in Tomato: Advances, Applicability and Challenges" in Plants (DOI: 10.3390/plants8050128)
Authors: Juhi Chaudhary[#], Alisha Alisha[#], Vacha Bhatt, Sonali Chandanshive, Nirbhay Kumar, Zahoor Mir, Ashwini Kumar, Satish K Yadav, S M Shivaraj, Humira Sonah, Rupesh Deshmukh
[#] Both authors contributed equally to the work

SUPERVISOR

Prof. dr. hab. Zofia Szweykowska-Kulińska

Department of Gene Expression, Faculty of Biology, Adam Mickiewicz University

Poznan, Poland

ASSISTANT SUPERVISOR

Dr. Izabela Sierocka

Department of Gene Expression, Faculty of Biology, Adam Mickiewicz University

Poznan, Poland

REVIEWERS

1.   Prof. Pawel Sowinski
2.   Prof. Agnieszka Mostowska
3.   Prof. Jakub Sawicki

Dedicated to my family

ACKNOWLEDGEMENTS

First and foremost, I would like to extend my gratitude towards my supervisor, **Prof. Zofia Szweykowska-Kulińska** for giving me this opportunity to be a part of such a great and wonderful team at "ZEG". Her constant support, encouragement, and guidance helped me during my whole PhD journey. Most importantly, her suggestions and useful comments have been invaluable in shaping my PhD project and this dissertation. Also, I would like to thank her for being available for me, especially during times of need. The trips to Krakow and India will always be extra memorable because of her 😊.

Another equally important member, without whom nothing of this would have been possible, is my assistant supervisor, **Dr Izabela Sierocka**. I would like to thank her for considering me worthy of working on this project. Her insightful feedback and constructive criticism helped me refine my research and produce a dissertation that I am proud of. I would like to thank her for being so patient, because of which I have always been able to share even 'tiniest' mistakes with her. Also, because of her, my scientific writing skills have improved drastically (I would like to believe that 😊).

Another grateful and warm thanks to the brilliant and always ready to help: Members of the lab at ZEG. Thanks to their camaraderie and support, my time as a PhD student was very enjoyable and productive. I feel privileged to have the opportunity to work alongside such dedicated and helpful labmates. I would like to especially thank a few members, including **Dr Łukasz Szewc**- For always listening to my failures and troubleshooting in experiments (which were really a lot) and never being short of ideas. **Dr Halina Pietrykowska**- For being a support in teaching Marchantia protocols and trying to teach me Polish (I know it was very difficult). **Monika Jóźwiak**- For being my friend first, with whom my lab time became enjoyable. Thank you for listening to my "not-so-work" problems and always coming up with smart life advice. **Dr Dawid Bielewicz**-. For organizing the beer meetings and always giving helpful advice about research, the future, and life in general. **Dr Aleksandra Chojnacka**- For making lab time enjoyable and for being a partner in crime in sharing PhD-related jokes. **Dr Aleksandra Świda-Barteczka**- Despite her help in the lab, took me to the hospital and made sure I got the treatment. **Dr Andrzej Pacak**- For always being so nice and exchanging talks about Indian and Polish cultures. **Dr Katarzyna Kruszka**- For being one of the nicest people in our lab who is always ready to help. **Dr Jakub Dolata**- For trying to

teach me Polish (for the longest duration, and I know it was really hard). **Dr Mateusz Michał Bajczyk**- For being helpful in almost any work-related problem. My fellow PhD colleagues, **Bharti Aggarwal** and **Marta Zimna-** For good times in the lab. Last but not least, **Prof. Artur Jarmołowski**- For his invaluable guidance during lab meetings and for always being so nice and kind. His cheerful energy and his passion for science continue to serve as great inspiration in both my personal and professional lives.

I would also like to thank the administration support staff, especially **Ms. Arleta Kucz**, for helping in organizing the documents for all the conferences and workshops with such ease, and **Dr. Karolina Cerbin** for always being available to solve literally any problem that arose during the work.

I am grateful to all my friends, **Akansha Tripathi**, **Pushpalata Kayastha, Amit Nagwani**, **Anand Maurya**, **Dheeraj Kumar Sarkar, Nishita Mandal, Jeena Varghese, Dr Ankur Gadgil,** and **Karolina Poczobutt** for being there for me during hard times, both in my professional and personal lives. Moreover, lunch time with all my friends acts as a detox every day and helps me go back to work with refreshed energy. A warm thanks to **Kamil Socha** for teaching me about Polish culture and being a constant support. You all made this journey easy.

Finally, I would like to thank my family for their unwavering support. A very big thanks to my **Parents, Rajesh Demra** and **Kanchan Demra,** for always encouraging me and tolerating me (especially when I asked them to stop asking when I was finishing my PhD). My brother, **Abhishek Kumar Demra**, and my sister, **Muskan Demra**, for their constant love and belief in me. I would have never been able to complete my PhD and my dissertation without them.

Thank you all from the bottom of my heart. <3

ABBREVIATIONS

aa	amino acids
AGL42	Agamous – like 42
amiRNA	artificial microRNA
AP1/2	Apetela 1/2
AP2/ERF	Apetala 2 / Ethylene-responsive element binding factors
ARF1	Auxin response Factor 1
AS2	Asymmetric leaves 2
ASL/LBD	Asymmetric leaves – 2 – like / LOB-domain
At	*Arabidopsis thaliana*
Atco-Cas9	Arabidopsis – codon – optimized Cas9
AUX/IAA	Auxin / Indole – 3 – acetic acid
bHLH	Basic helix – loop – helix
BOP1/2	Blade – on – petiole 1/2
bp	base pairs
BR	brassinosteroid
BZR1	Brassinazole – resistant 1
cDNA	complementary DNA
CDS	coding region of gene
CLE1	Clavata 3 / Embryo surrounding region – related
CLV1	Clavata1
CoNekT	Co-expression Network Toolkit
CRISPR/Cas9	clustered regulatory interspaced short palindromic repeats / CRISPR – associated protein 9
DBD	DNA binding domain
DEPC	diethylpyrocarbonate
DFR	dihydroflavonol – 4 – reductase
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
E(z)	Enhancer of zeste
EDTA	ethylenediaminetetraacetic acid
FeSOD	iron superoxide dismutase
FKF	Flavin – binding kelch repeat f – box 1
FUL	Fruitful
FT	Flowering locus T
GA	gibberellic acid
GAI	Gibberellic acid insensitive
GI	Gigantea
GW8	Grain – width 8
hCas9	human – codon – optimized Cas9
HR	homologous recombination
IPA1	Ideal plant architecture 1
JA	jasmonic acid
KAR	Karappo
KD	Knockdown
LB	luria-bertani
LFY	Leafy
MBEX	Marpolbase expression
Mp	*Marchantia polymorpha*
MPK1	Mitogen – activated protein kinase
NBS-LRR	Nucleotide – binding site leucine – rich repeat

NLS	nuclear localization signal
NMR	nuclear magnetic resonance
NREs	nitrate – responsive elements
Nt	nucleotide
ORF	open reading frame
Os	*Oryza sativa*
PAD4	Phytoalexin deficient 4
PAP1	Production of anthocyanin pigment 1
PCR	polymerase chain reaction
PRC2	Polycomb repressive complex 2
QTL	quantitative trait loci
RACE	rapid amplification of cDNA ends
RGA	Repressor of GAI
RNA	ribonucleic acid
ROS	reactive oxygen species
RT	room temperature
RT-qPCR	real time quantitative PCR
SA	salicylic acid
SAM	shoot apical meristem
SCR	Scarecrow
SDS	sodium dodecyl sulfate
SOC1	Suppressor of overexpression of constans 1
SPL	Squamosa promoter binding protein – like
SQN	Squint
SQUA	Squamosa
TAA	Tryptophan aminotransferase of arabidopsis
Tak	Takaragaike
TALENS	transcription activator-like effector nucleases
TEMED	N, N, N', N'-tetramethylethylenediamine
TF	transcription factor
TIR1	Transcport inhibitor response1
UTR	untranslated region
X-Gal	5-bromo-4-chloro-3-indolyl-d-galactopyranoside

Table of Contents

ABSTRACT (In English)

The *SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE* (*SPL*) family of transcription factors is functionally diverse in controlling a number of fundamental aspects of plant growth and development. Since the *SPL* genes exist amongst plants only, hence, it is imperative to understand their functions amongst basal lineages of land plants, to gain understanding of their evolution. The first part of my thesis presents the phylogenetic relationships between *SPL* family members from representatives of all lineages of bryophytes: two hornworts, *Anthoceros agrestis* and *A. punctatus,* liverwort *Marchantia polymorpha,* and moss *Physcomitrium patens,* and angiosperms representatives, *Arabidopsis thaliana*. The phylogenetic analysis classified SPL proteins in four phylogenetic groups. We found that the *SPL* family members within the same group share similar gene structures and protein domains which might hint towards the possible overlap in their putative functions. Moreover, there were no *SPL* genes identified in the hornwort lineage when we started our analysis and our results established that a minimal set of *SPL* genes is present in hornworts of Anthoceros lineage, which is similar to liverwort, *M. polymorpha*.

In the second part of presented thesis several molecular genetic tools were applied to characterize the function of Mp*SPL3* and Mp*SPL4* genes from model liverwort species, *M polymorpha*. First, combining *in planta* promoter activity using GUS reporter gene together with RT-qPCR analysis we have shown that both Mp*SPL3* and Mp*SPL4* genes are ubiquitously expressed during the vegetative as well as reproductive phases of Marchantia's life cycle. Next, to obtain knockout plants for each gene, CRISPR/Cas9 approach was used. The obtained two loss-of-function Mp*SPL3* plants displayed reduced thalli with delayed growth in comparison to wild-type plants. On the other hand, Mp*SPL4* loss-of-function plants displayed more severe phenotype resembling a prothallus-like stage with no production of gemma cups. As in the case of both genes' loss-of-function mutations caused very strong effect on Marchantia development, we applied artificial miRNA approach to knockdown the expression of Mp*SPL3* and Mp*SPL4* genes. The obtained knockdown plants for both genes displayed growth retardation during their vegetative growth. Moreover, gametangiophores production was completely abolished in Mp*spl3-kd* lines, while Mp*spl4-kd* plants exhibited delayed archegoniophores production which additionally showed morphological distortions. Therefore, proper level of Mp*SPL3* and Mp*SPL4* genes expression is also indispensable for Marchantia sexual organs development. In third

approach, we have prepared overexpression lines of both genes to study how the MpSPL3 and MpSPL4 protein excess will influence Marchantia development. The preliminary phenotypic analysis for gain-of-function Mp*SPL3* transgenic plants displayed no significant changes in phenotype during vegetative stage of growth as compared to wild-type plants. On the other hand, the plants overexpressing MpSPL4 protein displayed smaller and narrower thalli with bigger gemma cups in comparison to wild-type plants.

Taken together, the presented results provide significant insights into the basic functions of Mp*SPL3* and Mp*SPL4* genes from the *SPL* TF family from liverwort *M. polymorpha*, which are crucial players in controlling proper growth and development of both vegetative thallus and reproductive organs.

ABSTRACT (In Polish)

Czynniki transkrypcyjne z rodziny *SPL, SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE*, stanowią zróżnicowaną funkcjonalnie grupę białek, które kontrolują szereg podstawowych aspektów wzrostu i rozwoju roślin. Ponieważ geny *SPL* są specyficzne dla roślin, stąd konieczne jest zrozumienie ich funkcji wśród przedstawicieli żyjących linii roślin lądowych, aby zrozumieć ich ewolucję. W pierwszej części mojej pracy doktorskiej zajęłam się analizą filogenetyczną w celu zbadania relacji między członkami rodziny SPL przedstawicieli wszystkich linii mszaków: dwóch gatunków glewików, *Anthoceros agrestis* i *A. punctatus*, wątrobowca *Marchantia polymorpha* i mchu *Physcomitrium patens,* oraz rośliny okrytozalążkowej *Arabidopsis thaliana*. Na podstawie analizy filogenetycznej, białka SPL sklasyfikowano w czterech grupach filogenetycznych. Zaobserwowałam ponadto, że członkowie rodziny *SPL* w tej samej grupie posiadają podobne struktury genów i podobny zestaw domen białkowych, co może potencjalnie wskazywać na pełnienie podobnych funkcji. Przeprowadzone badania wykazały, że podobnie jak *M. polymorpha,* przedstawiciele glewików również posiadają tylko czterech członków w rodzinie *SPL,* co stanowi najprostszy zestaw genów *SPL* wśród roślin lądowych.

W drugiej części mojej pracy doktorskiej wykorzystałam szereg narzędzi molekularnych w celu określenia funkcji genów Mp*SPL3* i Mp*SPL4* z modelowego gatunku wątrobowca, *M. polymorpha*. Analiza aktywności transkrypcyjnej promotorów *in planta* badanych genów przy użyciu genu reporterowego β-glukuronidazy w połączeniu z analizą RT-qPCR wykazała, że oba geny ulegają ekspresji zarówno podczas wegetatywnej jak i reprodukcyjnej fazy cyklu życiowego *M. polymorpha*. Następnie, w celu uzyskania roślin z wyłączoną funkcją genu Mp*SPL3* lub Mp*SPL4* zastosowano podejście CRISPR/Cas9. Otrzymane mutanty ΔMp*spl3*^{ge} charakteryzowały się zredukowaną wielkością i kształtem plech oraz opóźnionym wzrostem w porównaniu do roślin typu dzikiego. Co ciekawe, mutanty ΔMp*spl4*^{ge} wykazały jeszcze silniejszy fenotyp niż rośliny ΔMp*spl3*^{ge}, gdyż rośliny te rosły jako masy komórkowe bez charakterystycznej dla Marchantia budowy plechy. Ponieważ w przypadku obu genów Mp*SPL* utrata ich funkcji spowodowało bardzo silne zaburzenie rozwoju wątrobowca, w kolejnym podejściu zastosowałam sztuczne mikroRNA w celu otrzymania roślin z obniżoną ekspresją badanych genów. W przypadku obu genów, mutanty otrzymane za pomocą sztucznego mikroRNA wykazały opóźnienie wzrostu podczas wegetatywnej fazy cyklu. Co więcej, wytwarzanie gametangioforów zostało całkowicie zablokowane w przypadku silnie obniżonej ekspresji genu Mp*SPL3,* podczas gdy

obniżenie ekspresji Mp*SPL4* spowodowało opóźnioną produkcję archegonioforów o zmienionej morfologii w porównaniu do roślin typu dzikiego. Dlatego właściwy poziom ekspresji genów Mp*SPL3* i Mp*SPL4* jest niezbędny dla rozwoju organów rozmnażania płciowego. W celu sprawdzenie efektu nadprodukcji białek MpSPL3 i MpSPL4 na rozwój *M. polymorpha,* przygotowałam rośliny transgeniczne z nadekspresją obu genów. Wstępna analiza roślin z nadekspresją genu Mp*SPL3* nie wykazała zmian w ich fenotypie podczas wegetatywnej fazy wzrostu w porównaniu do roślin typu dzikiego. Z kolei nadekspresja genu Mp*SPL4* spowodowała zmiany w morfologii plech, które były mniejsze i węższe w porównaniu do roślin typu dzikiego oraz wytwarzały powiększone miseczkowate zbiorniki produkujące wegetatywne rozmnóżki.

Podsumowując, przedstawione wyniki dają znaczący wgląd w podstawowe funkcje genów Mp*SPL3* i Mp*SPL4* z rodziny czynników transkrypcyjnych *SPL* u wątrobowca *M. polymorpha*, które są kluczowymi faktorami kontrolującymi prawidłowy wzrost i rozwój wegetatywnych plech, jak i struktur rozmnażania płciowego u tego wątrobowca.

1. INTRODUCTION

1.1 *Marchantia polymorpha* as a model organism

Despite the blurred phylogenetic relationships among bryophytes, they are suggested to be the first land plants inhabiting the earth according to fossil evidence (Edwards *et al.*, 1995; Wellman *et al.*, 2003; Nishiyama *et al.*, 2004; Qiu *et al.*, 2006; Wickett *et al.*, 2014; Rensing, 2018). Bryophytes comprise three lineages: mosses, liverworts, and hornworts, which together occupy the first nodes on the embryophyte tree life of land (Fig. 1).
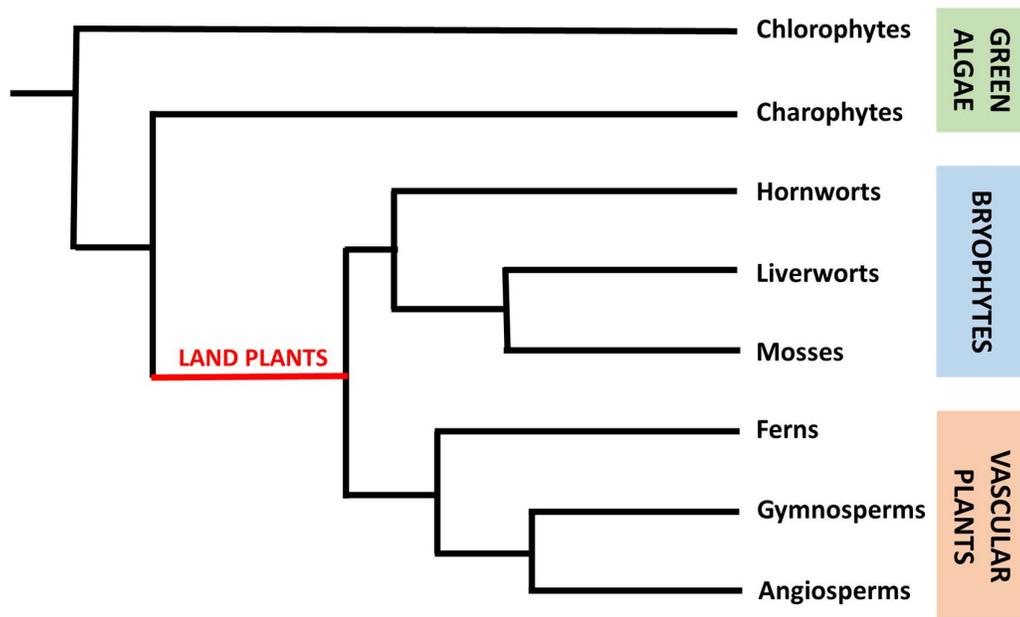
Figure 1: A simplified phylogenetic tree depicting relationships among green algae, bryophytes, and vascular plants. The divergence of land plants from algae is shown in red.

Because of bryophytes' phylogenetic position and unique features like free-living dominant gametophyte and partial or complete dependence of sporophyte on gametophyte, it is critical to study each lineage of bryophytes to understand land plant terrestrialization (Shaw and Renzaglia, 2004; Wang *et al.*, 2022; Bowman *et al.*, 2022). Since many decades, seed plants have been employed as model plants, but advancements in genomic and transcriptomic data from each lineage of bryophytes in recent years have made them model systems for studying the evolution of land plants (Cronk, 2005; Hunter, 2008; Wood *et al.*, 2013; Rensing *et al.*, 2020; Abdurakhmonov, 2022; Bowman *et al.*, 2022; Wang *et al.*, 2023*a*; Yadav *et al.*, 2023).

The bryophyte lineage, liverworts, consists of simple thalloid and complex thalloid clades. In recent years, the complex thalloid liverwort *Marchantia polymorpha* has emerged as a powerful model system in plants for the study of complex biological processes including

developmental, phylogenetic, physiological, and stress-induced cellular responses (Shimamura, 2015; Ishizaki *et al.*, 2016*a*; Bowman *et al.*, 2017*a*; Poveda, 2020, Bowman, 2022; Kohchi *et al.*, 2021; Naramoto *et al.*, 2022). Features of a model plant include ease of growth and maintenance, fast generation time, the ability to select for mutant phenotypes, knowledge of the genome via large scale sequencing, and the availability of genetic tools, including the ability to create transgenic individuals via transformation to create site-directed gene mutants. Marchantia fulfills all these criteria. Additionally, the dominance of gametophytic generation over sporophytic generation allows the recovery of developmental mutants with relative ease (Shimamura, 2015). Besides sexual reproduction, it can also be propagated asexually by gemmae, produced in gemma cups (Shimamura, 2015). The isolation of gene disrupted mutants is easier in Marchantia because of its haploid nature than in Arabidopsis*,* which has a diploid genome (Takenaka *et al.*, 2000). On the other hand, there are some limitations to working with a haploid system since the disruption of some essential genes may lead to lethality, making it difficult to obtain transgenic plants (Ishizaki *et al.*, 2013; Flores-Sandoval *et al.*, 2016). However, techniques such as artificial miRNA and conditional knockout systems have been successfully developed for Marchantia studies, expanding their potential to be used in gene-targeting strategies (Flores-Sandoval *et al.*, 2016). Hence, in order to study specific cellular and molecular processes in detail, *Marchantia* has been developed into an attractive evolutionary model plant system (Takenaka *et al.*, 2000).

1.1.1 Overview of *Marchantia polymorpha* life cycle

M. polymorpha shows an alternation of generations between haploid gametophyte and diploid sporophyte generations (Fig. 2). The haploid gametophytic stage is the dominant stage, which starts with a unicellular spore (top of the scheme). The spore germinates into a sporeling - a group of undifferentiated cells that develop from a spore and further give rise to a prothallus, which is a single-cell-layered thick structure. During the prothallus stage, the apical cell fulfilling meristematic function acquires the final three-dimensional developmental program leading to the formation of a thallus (Ishizaki *et al.*, 2016*b;* Kohchi *et al.*, 2021).
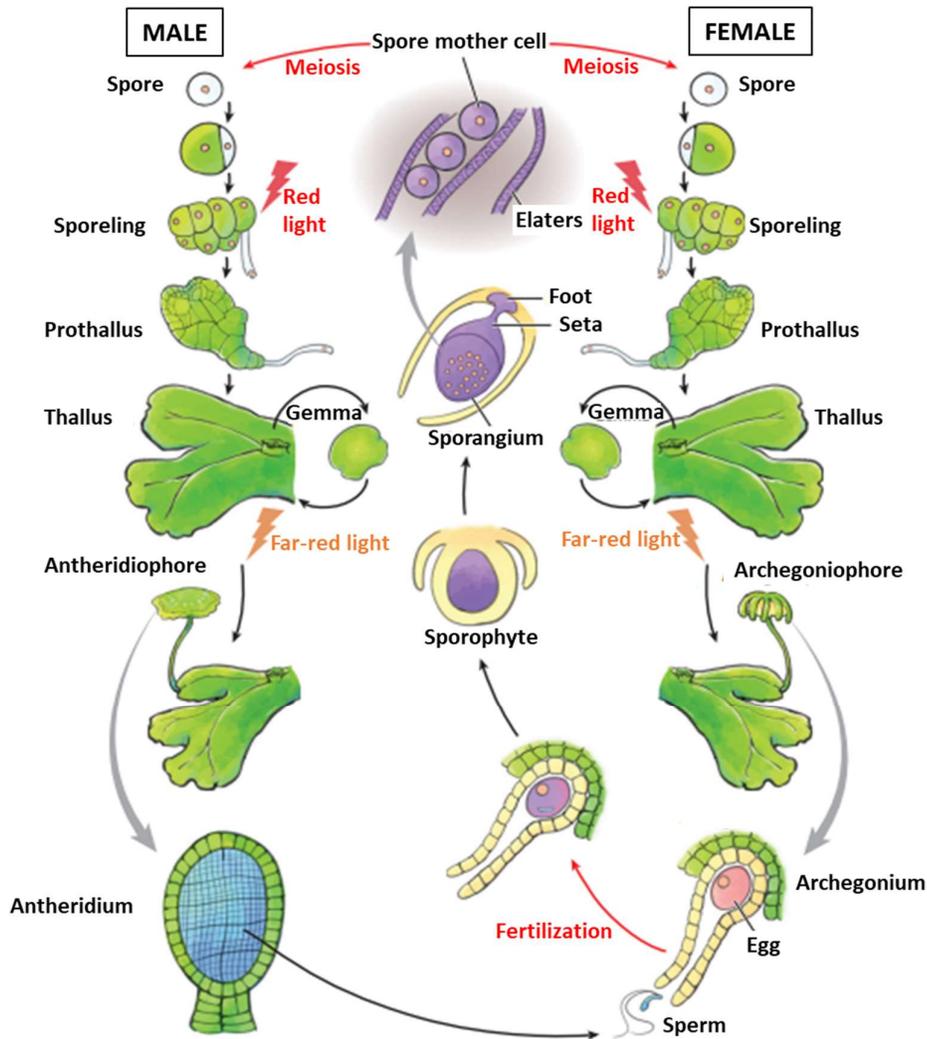
Figure 2: An outline of *M. polymorpha* life cycle (adapted from (Kohchi *et al.*, 2021)).

The thallus is a multilayered structure with a characteristic dorsal-ventral organization. The dorsal side is characterized by a presence of air chambers, while the ventral side is covered with a patterned network of scales and rhizoids (Bowman, 2022). At the dorsal side, additionally, asexual propagules, gemmae, are produced within gemma cups. A mature gemma usually remains dormant inside the gemma cup but after water intake, it disperses and establishes a new thallus. Marchantia is a dioecious species with separate male gametophyte (left) and female gametophyte (right). Under long-day conditions or by far-red light irradiation, male thallus produces antheridiophores containing antheridia and female thallus produces archegoniophores containing archegonia. Antheridia produce flagellated sperm cells which swim towards the egg cells in archegonia to fertilizes them. After fertilization, a multicellular diploid sporophyte is developed. The mature sporophyte is very small (up to 3 mm long) and consists of spore mother cells which produce spores via meiosis.

It is estimated that single capsule produces ~300,000 of haploid spores which are discharged from the sporangium by hygroscopic movements of elaters (Ishizaki *et al.*, 2016*b*; Kohchi *et al.*, 2021).

Marchantia's life cycle offers following advantages to study it as a model system in genetic, cellular, biochemical and molecular analysis:

1.   Haploid dominant life cycle: offers easier and rapid genetic analysis along with excluding the probability of heterozygosity in its gametophytic generation making it easier to study phenotypes of mutants.

2.   Asexual reproduction: offers easier propagation and rapid development of homozygous individuals since gametophytic thallus develops from a gemma or a single spore (Ishizaki *et al.*, 2016*b*).

1.1.2 Characteristic features of *Marchantia polymorpha* genome

As a representative of basal lineage of land plants, *M. polymorpha* became the first liverwort to have its genome sequenced in 2017. Interestingly, the genome of Marchantia preserves all the characteristics predicted for an ancestral land plant's genome including, presence of sex chromosomes, similar content of transcription factors (TF) and signaling pathway genes. The unique feature of Marchantia's genome is the presence of minimal set of regulatory genes due to absence of an ancient whole genome duplication. As compared to genomes of other sequenced land plants, Marchantia's genome contains single or low copy number of several regulatory genes. Moreover, the comparative genome analysis of few sequenced land plants, including Marchantia with charophytes revealed that orthologs of several signaling pathway genes were either absent in charophytes or acquired novel functions in land plants. For example, in auxin signaling pathway, *YUCCA* gene family (required for auxin biosynthesis) is absent in charophytes and found specifically within land plants whereas *TIR1* (*TRANSCPORT INHIBITOR RESPONSE1*) and *AUX/IAA* (*AUXIN/INDOLE-3-ACETIC ACID*) genes (required for auxin signaling) gained new functions in land plants which allowed them to specifically interact with auxin. Also, several genes which are specific for vascular plants were found absent in Marchantia's genome. This includes absence of gene encoding ethylene-forming enzyme (ACO) and lack of some downstream signaling genes from NBS-LRR (NUCLEOTIDE-BINDING SITE LEUCINE-RICH REPEAT) signaling pathway (Bowman *et al.*, 2017*b*).

In the genome of *M. polymorpha*, 2.1% of protein-coding genes were found to encode transcription factors (Bowman *et al.*, 2017*b*). The number of these genes is lower as compared with other land plants but higher than in algae, indicating that the number of transcription factors increases with the complexity of plants (Lang *et al.*, 2010; Catarino *et al.*, 2016; Lehti-Shiu *et al.*, 2017). However, the comparative analysis between transcription factors found in Marchantia and other land plants as compared to green algae revealed that numerous TF families diversified increasingly within land plants, implying their importance in terrestrialization on Earth. These include members of ASL (ASYMMETRIC LEAVES-2-LIKE) /LBD (LOB-domain), GRAS (named after the first 3 members: GIBBERELLIC-ACID INSENSITIVE [GAI], REPRESSOR of GAI [RGA], and SCARECROW [SCR] of GRAS protein family), NAC (named after NAM [NO APICAL MERISTEM], ATAF1/2 [*Arabidopsis thaliana* ACTIVATING FACTOR] and CUC2 [CUP-SHAPED COTYLEDON] proteins), AP2/ERF (APETALA2/ETHYLENE-RESPONSIVE ELEMENT BINDING FACTORS), WRKY (named after 'WRKY' signature amino acid sequence present in WRKY domains), bHLH (basic HELIX-LOOP-HELIX) and SPL (SQUAMOSA PROMOTER BINDING PROTEIN-LIKE). The discovery of TFs repertoire in Marchantia aids in a better understanding of land plant evolution and how novel adaptations enabled the land plant ancestors to survive under new sources of stresses encountered in terrestrial environments (Bowman *et al.*, 2017*b*).

1.1.3 Tools available for *Marchantia polymorpha* functional studies

1.1.3.1 Transformation methods

In plant developmental research, isolation of mutants using a gene disruption approach has been proven to be useful for studying the function of the respective genes in the context of plant development and survival (Takenaka *et al.*, 2000; Ishizaki *et al.*, 2016*b*). Many techniques are available for the transformation of plants,, and these techniques are also available for *Marchantia* now (Takenaka *et al.*, 2000; Ishizaki *et al.*, 2013, 2015, 2016*b*; Tsuboyama and Kodama, 2018; Tsuboyama *et al.*, 2018). In an initial transformation protocol developed for Marchantia, the suspension-cultured cells were successfully transformed by particle bombardment, but the callus-like structures obtained could not be regenerated into new plants (Irifune *et al.*, 1996). Further, Takenaka et al. successfully generated a method for direct transformation of *M. polymorpha* plants (Takenaka *et al.*, 2000). The transformed thalli obtained from this procedure generally take 8-10 weeks, which is less than the 12-36 weeks required to differentiate the same cell mass in higher plants.

Later in 2008, Chiyoda et al., developed a method for direct transformation of Marchantia where the thallus is developed from spores by particle bombardment. An advantage of this method is that it requires half the time as compared to initial methods and does not require any stepwise transfer to liquid and solid media (Chiyoda *et al.*, 2008). Additionally, Chiyoda et al., also established the same approach for plastid transformation using suspension culture cells of Marchantia (Chiyoda *et al.*, 2007, 2014). However, the mode of DNA transfer by physical delivery has some drawbacks, which include the generation of complex mutations and rearrangements that make the manipulation of the transformants obtained difficult (Chiyoda *et al.*, 2007, 2014). Meanwhile, Ishizaki et al., developed the first Agrobacterium-mediated transformation in Marchantia, as previously it has been practical only in angiosperms. This method allows transformation of Marchantia immature thalli grown from spores - sporelings. Importantly, this method has greater efficiency and lesser rearrangements than the physical DNA delivery method (Ishizaki *et al.*, 2008). Therefore, Agrobacterium-mediated transformation is the most employed transformation protocol in Marchantia studies (Kato *et al.*, 2015a; Monte *et al.*, 2018, 2019; Montgomery *et al.*, 2020; Li *et al.*, 2020a; Dierschke *et al.*, 2021).

In the previous methods requiring liquid medium, the transformants generated were aggregated when shaken in the medium, and hence, it was difficult to obtain independent transformants. This problem was resolved in the AgarTrap (Agar-utilized Transformation with Pouring Solutions) method as independent transformants can be obtained in the petri plates with higher efficiency (Sporeling - AgarTrap) (Tsuboyama and Kodama, 2014). The drawback with using sporelings for transformation is that sexual reproduction is crucial for the generation of spores, which when working on infertile mutants cannot be obtained. In order to overcome these limitations, a new modified version of this method was developed in which, intact gemmae/ gemmalings (G-AgarTrap) obtained by asexual reproduction were employed. Furthermore, this method was also developed using mature thallus pieces (T-AgarTrap) (Tsuboyama-Tanaka and Kodama, 2015; Tsuboyama and Kodama, 2018). Additionally, the development of Gateway binary vectors with different selection markers for transgenic experiments greatly enabled and eased the molecular genetic analysis (transgenic research) in Marchantia (Ishizaki *et al.*, 2015).

1.1.3.2   Genome manipulation tools

The availability of genome sequence and development of efficient *Agrobacterium*-mediated transformation protocols for *M. polymorpha*, further enabled to establish several gene targeting methods. These include gene editing by homologous recombination (HR), CRISPR/Cas9 (CLUSTERED REGULATORY INTERSPACED SHORT PALINDROMIC REPEATS/CRISPR-associated protein 9), TALENS (TRANSCRIPTION ACTIVATOR-LIKE EFFECTOR NUCLEASES), and artificial miRNA (amiRNA) (Ishizaki *et al.*, 2013; Sugano *et al.*, 2014; Kopischke *et al.*, 2017; Sugano and Nishihama, 2018). All these gene-targeting methods ease the manipulation of genome by forward and reverse genetics analysis.

HR-mediated gene targeting has been a potent tool for functional analysis of genome by reverse genetics. Gene targeting by HR has been difficult in higher plants although it is highly efficient in moss, *P. patens* (Kamisugi et al., 2006). A protocol for gene-targeting by HR has been established in Marchantia via positive/negative selection system, similar to rice. In this system, hygromycin-resistant gene, *hpt,* and diptheria toxin A fragment gene, *DT-A* were used as positive and negative markers for Marchantia, respectively (Terada *et al.*, 2002, 2007; Ishizaki *et al.*, 2013). When the whole transgene is introduced to the genome via non-homologous recombination, DT-A protein having endonuclease activity, exerts cellular toxicity. Because of introducing a double selection system within HR, higher specificity and efficiency can be obtained (Ishizaki *et al.*, 2013). This approach has successfully been employed in targeting several genes including Mp*PHOTOTROPIN* involved in chloroplast photorelocation movement (Komatsu *et al.*, 2014), Mp*FKF* (*FLAVIN-BINDING KELCH REPEAT F-BOX 1*) and Mp*GI* (*GIGANTEA*) involved in photoperiodic regulation (Kubota *et al.*, 2014), Mp*TAA* (*TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS*) involved in auxin biosynthesis (Eklund *et al.*, 2015), Mp*KAR* (*KARAPPO*) involved in gemmae initial cells specification (Hiwatashi *et al.*, 2019) and Mp*CLV1* (*CLAVATA1*) involved in CLE1 (CLAVATA3/EMBRYO SURROUNDING REGION-related) signaling pathway (Hirakawa *et al.*, 2020).

A fast and simple gene editing technology known as CRISPR-Cas has been successfully employed to modify genomes of many model plants (Sugano *et al.*, 2014; Peterson *et al.*, 2016; Noman *et al.*, 2016; Meng *et al.*, 2017; Qi, 2019; Mallett *et al.*, 2019). This targeted genome editing method has been also successfully adapted for gene editing in Marchantia in recent years (Sugano *et al.*, 2014). There are several combinations of vectors that can be used for CRISPR/Cas9-directed genome modification in Marchantia (Ishizaki *et al.*, 2015; Sugano

et al., 2018). The Cas9 protein can be driven under the expression of either Marchantia *EF1α* or *Cauliflower mosaic virus 35S* promoter. The initial CRISPR/Cas9 protocol employed human-codon-optimized Cas9 (hCas9) (Sugano *et al.*, 2014), however, because of its low gene editing efficiency, it was replaced by Arabidopsis-codon-optimized (Atco-Cas9) which greatly increased the gene-editing efficiency (Ishizaki *et al.*, 2015; Sugano *et al.*, 2018) (Fig. 3).
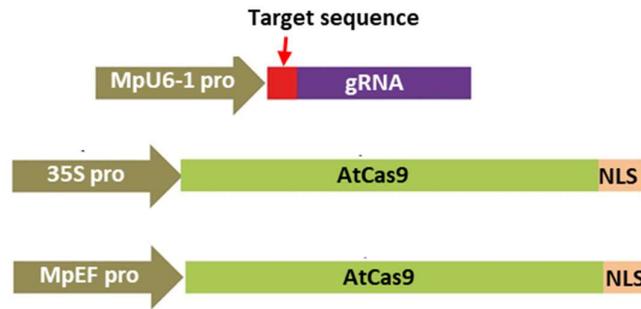
Figure 3: Constructs used for CRISPR/Cas9 studies in Marchantia in studies by Sugano et al., 2018. gRNAs are expressed under Marchantia *U6* promoter, *AtCas9* is fused with NLS (nuclear localization signal) and expressed under either 35S promoter or Marchantia *EF1α* promoter (adapted from Sugano et al., 2014; 2018).

Currently, CRISPR-Cas9 is the most frequently employed gene editing technology for obtaining stable mutants of protein coding genes but also for editing *MIR* gene loci in Marchantia (Li *et al.*, 2020*a*; Dierschke *et al.*, 2021; Ishida *et al.*, 2022). In addition to single gRNA approach, the CRISPR-Cas9 can be also applied with the usage of multiple gRNAs in single construct (Sugano *et al.*, 2014, 2018; Sugano and Nishihama, 2018).

Gene targeting by artificial miRNAs has been used in many plant species including moss, *P. patens* (Khraiwesh *et al.*, 2008), many monocots and dicots (Schwab *et al.*, 2006; Alvarez *et al.*, 2006). Artificial miRNAs are constructed by substitution of miRNA and miRNA* in the endogenous miRNA precursor which directs it to a specific target (Schwab *et al.*, 2006; Alvarez *et al.*, 2006; Ossowski *et al.*, 2008). In Marchantia, artificial miRNA constructs were generated using the backbones of either Mp*MIR160* or lycophyte *Selaginella kraussiana MIR166* and the resulting amiRNAs were expressed under MpEF1-α promoter (Fig. 4) (Flores-Sandoval *et al.*, 2016).
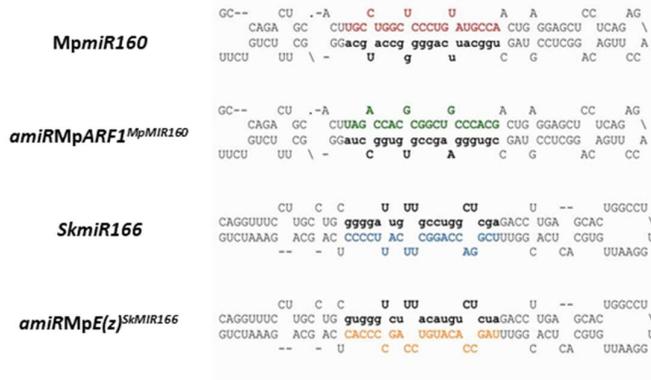
Figure 4: Artificial microRNA design used in studies by Flores-Sandoval et al., 2016. The predicted stem-loop structures formed by Mp*miR160*, *amiR*-Mp*ARF1*^{MpMIR160}, Sk*miR166* and *amiR*-Mp*E(z)*^{SkMIR166}. Coloured sequences represent miR and amiR, and miR* sequences are depicted by bold and lower case. Artificial miRNAs designed for Mp*ARF1* (*AUXIN RESPONSE FACTOR*) and Mp*E(z)* (*ENHANCER OF ZESTE*) are denoted as *amiR*-Mp*ARF1*^{MpMIR160} and *amiR*-Mp*E(z)*^{SkMIR166}, respectively.

As already mentioned, gene editing by HR or CRISPR/Cas9 have some limitations that are more prominent in a haploid organism (Ishizaki *et al.*, 2013, 2016*b*; Sugano *et al.*, 2018) . To overcome these limitations, artificial miRNAs with inducible or conditional expression system have been developed as an alternative tool for studying function of a gene of interest in Marchantia. Moreover, it has been efficiently employed in the knock-down of many genes in Marchantia. For example, CRISPR/Cas9 mutants of Mp*ARF1* (*AUXIN RESPONSE FACTOR 1*) involved in auxin signaling, resulted in strong phenotypic defects in cell patterning and differentiation (Sugano *et al.*, 2014). Meanwhile, many lines were obtained from artificial miRNA for Mp*ARF1* ($_{pro}$*EF1:amiR-MpARF1* ^{MpMIR160}) but none of them produced such strong defects, implying that the produced alleles were not null. On the other hand, artificial miRNA lines generated for Mp*E(z)* (*ENHANCER OF ZESTE*) gene encoding component of PRC2 (Polycomb repressive complex 2), under the constitutive expression of Mp*EF1* promoter ($_{pro}$*EF1:amiR-MpE(z)* ^{MpMIR160} and $_{pro}$*EF1:amiR-MpE(z)* ^{SkMIR166}) resulted in sporeling lethality. Therefore, an estrogen inducible system was developed to obtain loss-of-function allele of Mp*E(z)*. In the presence of estrogen, these mutant plants displayed arrested growth with eventual lethality under prolonged estrogen induction (Flores-Sandoval *et al.*, 2016).

1.1.4 Available database resources for *Marchantia polymorpha*

In 2017, the database for genes and genetic parts of *M. polymorpha* genome, known as MarpolBase or MarpoDB became available to the research community. The database currently contains three versions of the Marchantia reference genome: JGI 3.1 (Bowman *et al.*, 2017*b*), MpTak1_v5.1 (*M. polymorpha* subsp. *ruderalis*) and the current version,

MpTak1_v6.1 (comprises the combination of male Tak1 genome (MpTak1_v5.1) and female Tak2 U-chromosome) (Montgomery *et al.*, 2020). MpTak1_v6.1 is the newest annotation, which is based on chromosome-level genome assembly as opposed to JGI 3.1 with scaffold-based genome and hence, is a standard reference genome currently used for analytical purposes. Additionally, the database also contains genome data from *M. polymorpha* subsp. *polymorpha* BR5 (MppBR5) and subsp. *montivagans* SA2 (MpmSA2) (Linde *et al.*, 2020). In general, the database provides a platform for genome-based analysis as it contains many analytical tools. These tools can be used to perform similarity search (BLAST), to design guide RNA for CRIPR/Cas9 system (CasFinder, CRISPRdirect), or to design artificial miRNA for gene of interest (amiRNA design helper). MarpolBase also acts as a database for submitting Marchantia gene names to prevent any redundancy and maintain uniformity in literature. In general, the database serves as a platform that aids in promoting synthetic biology in *M. polymorpha* (Linde et al., 2020).

Since the RNA-seq data from *M. polymorpha* studies has been growing vastly, an expression database was created for analyzing this data in 2021. The expression database, named as 'MarpolBase expression (MBEX)' is closely linked with MarpolBase genome resources and thus continuously incorporates annotation updates. The expression database consists of expression tools, co-expression tools and analysis tools for evolutionary and functional analysis of Marchantia genes. The expression tools can be employed for generating chromatin expression images, bar plot, line plot and cluster gram after submitting an appropriate gene ID. Meanwhile, co-expression tools can be used to generate functional or co-expression network, to draw network and to obtain co-expression or rank table. Moreover, the analysis tools can be exploited to perform differential expression analysis for two selected conditions, functional enrichment analysis for a set of genes, orthogroup analysis for selected gene and set relations by integrating multiple analysis. Additionally, the database is linked to SRA (Sequence Read Archive) data source (Leinonen *et al.*, 2010). In general, the expression database serves as a platform for envisioning expression levels of genes, investigating differentially expressed genes, co-expression data and networks and analyzing functional enrichment data (Kawamura *et al.*, 2022) (Fig. 5).
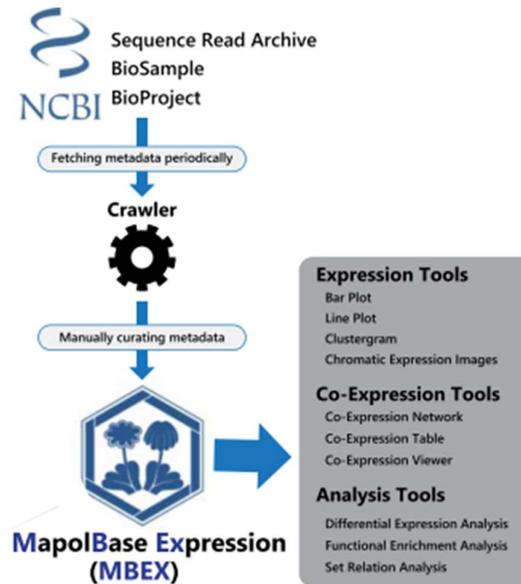
Figure 5: An outline of MBEX construction (adapted from Kawamura *et al.*, 2022)

In 2023, the gene expression atlas was created for *M. polymorpha* in different abiotic stresses. The atlas was created for seven different abiotic stresses alone as well as in different combinations. The atlas provides data on phenotype, differential gene expression, analysis of biological and cellular pathways in stress conditions (Tan *et al.*, 2023). This data is made available through eFP browser for *M. polymorpha* at Marchantia eFP Browser and CoNekT : Co-expression Network Toolkit platforms. Although both these databases are expression-based databases, eFP browser can be employed for visualization of the expression data while CoNekT database can be employed for comparative tissue-specific expression data in stress conditions under study (Tan *et al.*, 2023). In general, this data and the available databases will aid in investigating the gene expression pathways and networks involved in different abiotic stresses in Marchantia.

1.2 SPL Transcription factors functions in plants

By controlling the transcription of genes, transcription factors (TFs) are the major governors of organisms' functioning and development. The regulation of gene expression is further important for regulating various biological processes in living organisms, including plants. In plants, the activity of transcription factors is regulated by many factors including, transcription and translation control, protein-protein interactions, chromatin conformation and subcellular localization (Yanagisawa, 1998). Moreover, the comparative analysis of transcriptional regulators in Arabidopsis (plant), *Saccharomyces cerevisiae* (fungi), *Caenorhabditis elegans* and Drosophila (animals) revealed that many transcription factors

are shared between these kingdoms. However, the study also revealed the presence of lineage-specific TFs, indicating their origin after these three kingdoms' divergence. Plant-specific TF families include members from AP2/ERF, ARF, GRAS, Dof (DNA binding with one finger), WRKY and SPL (Yanagisawa, 2004; Birkenbihl *et al.*, 2005; Bowman *et al.*, 2017*b*; Li *et al.*, 2021). The majority of the plant-specific TF family members that have been identified so far are involved in gene regulation processes that are connected to the development of plant-specific organs and to response mechanisms for adapting to terrestrial conditions.

1.2.1 SPL TF family: general characteristics

In *Antirrhinum majus* (snapdragon), two genes were identified as interacting partners for *SQUAMOSA* gene (*SQUA*) promoter and named as *SBP1* and *SBP2*, derived from SQUAMOSA-PROMOTER BINDING PROTEIN. This *SQUA* gene is a floral meristem identity gene that belongs to *MADS*-box transcription factor family. The conserved domain in both *SBP* genes consists of 79 amino acids at their C-terminus and contain a bipartite nuclear localization signal (NLS) for transport to the nucleus (Davies and Schwarz-Sommer, 1994; Klein *et al.*, 1996). Moreover, to study the involvement of this conserved domain in specific binding of its interactors, truncated versions of SBP2 protein were prepared. This conserved domain was hence found to be sufficient and necessary for binding and recognition of *SQUA* promoter elements and therefore, was named as SBP-box (Davies and Schwarz-Sommer, 1994; Birkenbihl *et al.*, 2005). The characteristic feature of SBP domain is the presence of two Zn-binding domains (Zn-1 and Zn-2) and one NLS (Davies and Schwarz-Sommer, 1994; Birkenbihl *et al.*, 2005) (Fig. 6).

Figure 6: A sequence logo of SBP domain prepared from 166 SPL protein sequences from 52 species used in study conducted by (Birkenbihl *et al.*, 2005). SBP domain consists of two Zn-binding domains and one nuclear localization signal, represented by Zn-1, Zn-2 and NLS in this figure, respectively (adapted from (Birkenbihl *et al.*, 2005).

In *A. thaliana*, 16 members belonging to SBP-box gene family were identified and named as *SPL* genes (*SQUA promoter-binding protein-like*) (Cardon et al., 1999). NMR structures of Arabidopsis SPL4, SPL7 and SPL12 have proven that SBP-DBD (DNA binding domain)

contains two zinc-binding pockets (Fig. 7). The coordinating pattern of dual zinc-binding motifs is Zn-1: Cys3His – Zn-2: Cys2HisCys or Zn-1: Cys4 – Zn-2: Cys2HisCys, consisting together eight Cysteine and Histidine residues. Zinc ions are crucial for stabilisation of both motifs and for their interaction with each other through hydrophobic residues. Due to this tight packing, SBP-DBD behaves as a single domain (Yamasaki *et al.*, 2004).
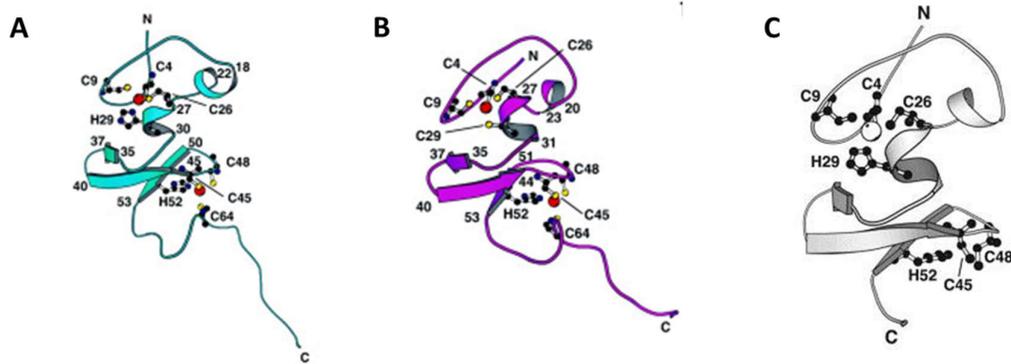
Figure 7: Ribbon diagrams of DNA binding domains of AtSPLs, prepared from NMR spectroscopy values. A) AtSPL4-DBD, B) AtSPL7-DBD and C) AtSPL12-DBD. In A) and B), zinc ions are represented as red spheres and in C) by spheres. Zn-coordinating residues are depicted as ball-and-stick presentations. DBD corresponds to residues 1-7 and the residue numbers are marked for the SBP domain (Yamasaki *et al.*, 2004).

The SPL proteins bind DNA in a sequence-specific manner. The promoter of the target genes contains a conserved sequence TNCGTACAA, where N represents any base (Cardon *et al.*, 1997; Birkenbihl *et al.*, 2005). Experimental studies revealed that the SBP domain binds to DNA with a stoichiometry of 1:1. The region of DBD was found to be positively charged for specific binding to negatively charged DNA. This region was mostly comprised of conserved basic residues, Arg and Lys (Yamasaki *et al.*, 2004). These findings suggest that SBP-DBD has a novel type of structure and binding as compared to previous-known Zn-binding domains. The sequence specificity of SPL family of transcription factor family is related to zinc-finger like structures in the DBD. Functional studies showed that the palindromic GTAC core motif is required for proper DNA-binding by these TFs (Birkenbihl *et al.*, 2005; Yamasaki *et al.*, 2009; Zhang and Li, 2013; Lei *et al.*, 2016; Wang *et al.*, 2016; Cao *et al.*, 2021; Zhang *et al.*, 2022) (Fig. 8).
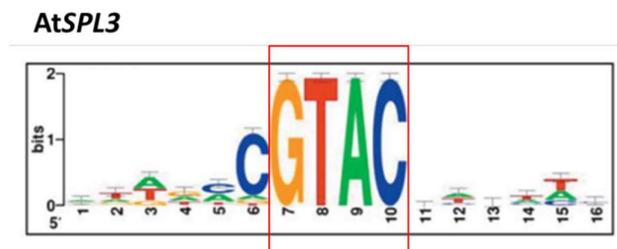
Figure 8: A sequence logo of GTAC core sequence detected in At*SPL3* sequence. The experiment first revealed that SBP-domain binds to GTAC core sequence (adapted from (Birkenbihl *et al.*, 2005).

1.2.2 Functions of *SPL* gene family in angiosperms

Although there is growing number of publications describing SPL transcription factor family in different plants, the most comprehensive studies are available for angiosperms, with *A. thaliana* and *O. sativa* being the best functionally characterized. The expression levels of many *SPL* genes are controlled by microRNA, in dicots by miR156, and in monocots by miR156/529. For instance, in *A. thaliana*, 10 out of 16 *SPL* members are being regulated by miR156/157 (Fig. 9a). On the other hand, in *O. sativa*, 11 out of 19 *SPL* genes are targets for miR529/miR156/miR535 (Xie *et al.*, 2006; Dai and Zhao, 2011; Yue *et al.*, 2017) (Fig. 9b).
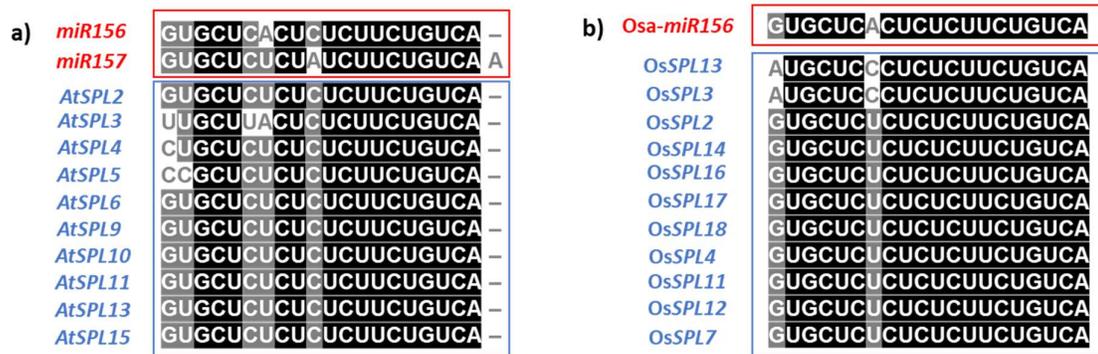
Figure 9: Sequence alignment between miRNA mature sequences and its target *SPL* genes in *A. thaliana* and *O. sativa*, respectively. a) alignment between miR156/156 mature sequences and its target 10 *SPL* genes in *A. thaliana* and b) alignment between miR156 mature sequence and its target 11 *SPL* genes in *O. sativa*. Identical sites are shaded (adapted from (Xie *et al.*, 2006; Chen *et al.*, 2010))

In Arabidopsis, miR156 is highly expressed in tissues like young seedlings but it declines as the shoot development progresses. Meanwhile in rice, studies have shown that both the miRNAs, miR156 and miR529 have overlapping expression patterns (Morea *et al.*, 2016). Interestingly, Os*SPL14* is targeted by both miRNAs (miR156a-j and miR529a-5p) during seedling stage while it is mostly targeted by miR529 (miR529a-5p) in panicle (Jeong *et al.*, 2011). Changes in the expression levels of each miRNA have their implication in Arabidopsis and rice development. The overexpression of two Os*miR156* members (*pri-miR156d* and *pri-miR156h*) resulted in severe plant morphological changes including dwarfism, increase in number of tillers, delayed flowering time, reduction in number of grains per panicle, secondary branches of panicle, and spikelets. Despite the changes observed in panicle size in these plants, their fertility was not affected as compared to wild-type rice. Meanwhile, the overexpression studies of miR529a in rice affected the panicle architecture and expression of three out of five *SPL* genes (Os*SPL2*, Os*SPL14* and Os*SPL17*) (Yue *et al.*, 2017). Therefore, the Os*SPL*-OsmiR156/529 module was found to influence various developmental processes, specifically the flower development of rice (Xie *et al.*,

2006). Moreover, similar morphological changes were also observed in Arabidopsis plants overexpressing miR156, which includes increase in number of leaves, delayed time of flowering under long day conditions and decrease in number of flowers from side shoots and decrease in apical dominance (Schwab *et al.*, 2005). Hence, miR156-*SPL* module is known to control various regulatory functions in the development of higher plants.

1.2.2.1 Functions of *SPL* gene family in developmental regulation in angiosperms

SPL family of transcription factors is known to regulate juvenile-to-adult and vegetative-to-reproductive transition in Arabidopsis. miR156 is known to control many aspects of plant development and physiology by regulating *SPL* gene expression at different times of their development in Arabidopsis (Xu *et al.*, 2016). *SPL* genes regulated by miRNA in Arabidopsis have been divided functionally into three groups based on their involvement in phase transitions:

1.   *SPL* genes which contribute to both phase transitions: *SPL2, SPL9, SPL10, SPL11, SPL13 and SPL15*

2.   *SPL* genes which contribute to floral meristem identity transition: *SPL3, SPL4 and SPL5*

3.   *SPL* genes which do not contribute to shoot morphogenesis but play a role in other physiological processes: *SPL6 (Xu et al., 2016)*.

The phase transitions in Arabidopsis are regulated by miR156 and miR172 in a coordinated manner. A subsequent decrease in the levels of *MIR156* and an increase in its target *SPL* transcription factor mRNAs causes simultaneous increase in *MIR172* levels. This leads to activation of flowering pathway by targeting *APATELA2 (AP2)*-like TF gene and an induction in *FLOWERING LOCUS T (FT)* gene. This signaling cascade of miR172-*FT* with miR156 activates *SPL* genes which further induces floral meristem identity genes in Arabidopsis (Jung *et al.*, 2016).

Vegetative phase change refers to transition from juvenile-to-adult phase. In *Arabidopsis*, it is accompanied by an increase in the expression of *SPL* genes which are regulated by decrease in levels of miRNAs, miR156 and miR157 (Hu *et al.*, 2023*a*). miR156 has been shown to regulate shoot morphology and *SPL* gene expression to higher extent than miR157. For example, At*SPL9, 10, 13* and *15* genes are essential for the development of adult vegetative traits (He *et al.*, 2018). In Arabidopsis, brassinosteroid (BR- one of the main growth-promoting steroid hormones in plants) has been shown to positively regulate leaf

morphology. Additionally, treatment by BR hormone induced the expression of At*SPL9*, *10* and *15*. Moreover, *SPL9* gain-of-function lines exhibited BR hypersensitivity. Furthermore, *SPL9* was shown to interact with BZR1 (BRASSINAZOLE-RESISTANT 1), a master TF of BR signaling pathway. Therefore, *SPL9*-BR pathway is involved in vegetative phase change in Arabidopsis (Wang *et al.*, 2021).

Moreover, At*SPL9* and At*SPL13* have been shown to promote adult leaf morphology via binding to the promoters of *blade-on-petiole1* and *2* (*BOP1/2*) and suppresses their expression. In juvenile leaf, miR156 levels are higher and as a result its targeted *SPL* levels are lower. This results in activation of *BOP1/BOP2* which causes earlier blade-petiole boundaries. In contrast, in an adult leaf, At*SPL9/13* levels are higher which results in down-regulation of *BOP1/BOP2* causing late development of blade-petiole boundaries and hence, elongated blades (Hu *et al.*, 2023*a*) (Fig. 10).
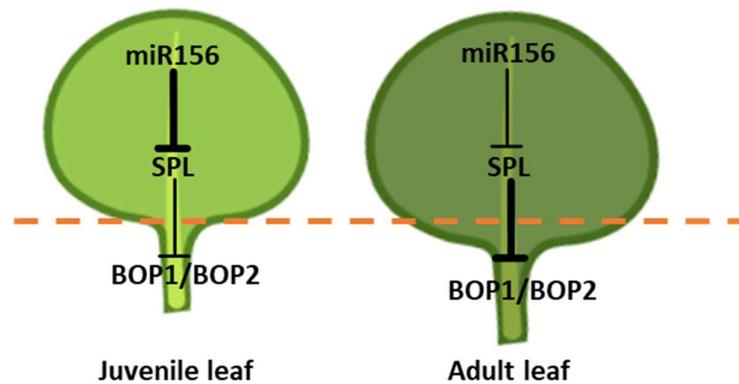
Figure 10. A schematic representation of involvement of miR156-*SPL* module in leaf development (adapted from (Hu *et al.*, 2023*a*). The bold arrows depict that miR156-*SPL* inhibition is active in juvenile leaf while *SPL-BOP1/2* inhibition is active in adult leaf, respectively. The images of juvenile and adult leaves are created with BioRender.com.

Reproductive phase change includes transition from vegetative shoot apical meristem (SAM) to inflorescence meristem. Several SPL proteins including SPL2/10/11, SPL3 and SPL9, bind to promoters of several flowering regulatory genes including *APETELA1* (*AP1*), *LEAFY* (*LFY*), *FRUITFUL* (*FUL*), *ASYMMETRIC LEAVES 2 (AS2), SOC1* (*SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1*) and *AGL42* (*AGAMOUS-LIKE 42*) and positively regulate flowering (Chen *et al.*, 2010). For instance, AtSPL2 regulates fertility rate, pollen production and elongation of various floral organs by binding to *AS2* promoter (Wang *et al.*, 2016). Another miR156 targeted *SPL*s, At*SPL3* and At*SPL5* regulates flowering via their interactions with nitrate-regulated genes. *SPL3* and *SPL5* contain nitrate-responsive elements (NREs) in their promoters and transcription factors of nitrate signaling pathway bind to these NREs (Olas *et al.*, 2019). In the case of *SPL9,* it was shown to interact with DELLA proteins

involved in Gibberellic acid (GA) signaling pathway to promote floral transition in Arabidopsis (Yu *et al.*, 2012; Zhang *et al.*, 2019). Hence, miR156-*SPL9* module play dual role in both vegetative and reproductive stages of development in Arabidopsis. Therefore, miR156-*SPL* regulatory module acts throughout the course of plant development in Arabidopsis (Fig. 11).
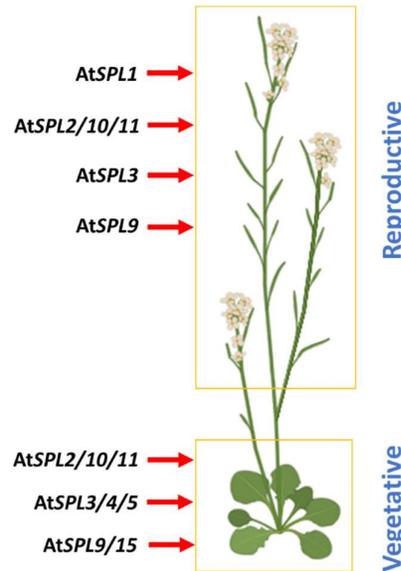
Figure 11: A summary of At*SPL* genes in development of vegetative and reproductive traits in Arabidopsis. The figure is created with BioRender.com.

In rice, a network of miRNAs and TF regulates vegetative (tiller) and reproductive (panicle) architecture/ branching (Wang *et al.*, 2015*a*). This branching pattern further determines the yield and quality of grain. The miRNA156/529-*SPL* modules are one of important players in controlling this process. For instance, over-expression of Os*SPL14* positively regulates panicle branches and grain yield (Jiao *et al.*, 2010; Miura *et al.*, 2010). Another example is *OsSPL13* which was first identified as a major quantitative trait locus (QTL), *GLW7*, affecting grain yield in rice. This *GWL7* locus promotes grain length and panicles, thereby, increases the yield (Si *et al.*, 2016). Another QTL locus *GW8* (*Grain-width 8*) was identified as synonymous with *OsSPL16.* The overexpression of *OsSPL16* causes reduction in plant height, tiller numbers, panicle branches and increase in grain width and hence, yield *(Wang et al., 2012)*. Also, upregulation of *OsSPL7* caused a decrease in number of tillers and lateral roots (Dai *et al.*, 2018*a*). As mentioned earlier, miR529a directly modulates the expression of several *OsSPL* genes (*OsSPL2*, *OsSPL14* and *OsSPL17*) and affects the panicle architecture (Yue *et al.*, 2017). Hence, different rice *SPL* genes function co-ordinately in regulating the quality and yield of rice grains (Fig. 12). Therefore, *SPL* functions can be

exploited by the breeders to improve the grain yield in rice and other agricultural crops.
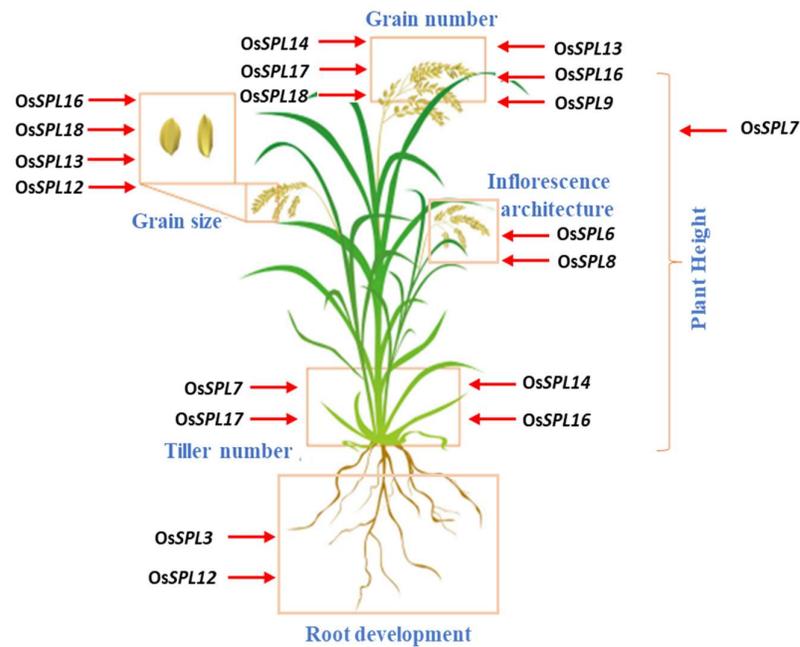
Figure 12: A summary of roles of *OsSPL* genes in growth and development of rice (modified from (Liu *et al.*, 2016).

1.2.2.2 Functions of *SPL* gene family in stress responses in angiosperms

SPL genes expression regulation is implicated in controlling various stress responses (Ling and Zhang, 2012; Zhang *et al.*, 2015; Morea *et al.*, 2016; Li *et al.*, 2022). In Arabidopsis, during stress conditions like cold, drought and salinity, miR156 upholds the plant in its juvenile stage for longer durations, by suppressing the expression of *SPL*s. After attaining favourable conditions, *SPL* inhibition is released as miR156 is suppressed hence, accelerating the developmental transition by activating genes critical for flowering (Wang *et al.*, 2009; Yu *et al.*, 2010; Sunkar, 2012; Wang, 2014). Under salt and drought treatments in Arabidopsis, studies have shown that transgenic plants with functional inhibition of miR156 become more susceptible to stress treatment. In contrast, transgenic plants overexpressing miR156 showed increased tolerance to these two stresses. Interestingly, salt and drought stress cause elevated expression levels of anthocyanin-pathway genes, *DFR* (*DIHYDROFLAVONOL-4-REDUCTASE*) and *PAP1* (*PRODUCTION OF ANTHOCYANIN PIGMENT 1*) in wild-type Arabidopsis plants that helped them to survive unfavourable growth conditions. On the other hand, *SPL9* has been shown to interact with both these genes directly and negatively influence anthocyanin biosynthesis. Therefore, under stress conditions, miR156 functions to inhibit action of several *SPLs* which in turn enables higher

anthocyanin biosynthesis to increase the plants' tolerance to the stresses (Cui *et al.*, 2014) (Fig. 13).
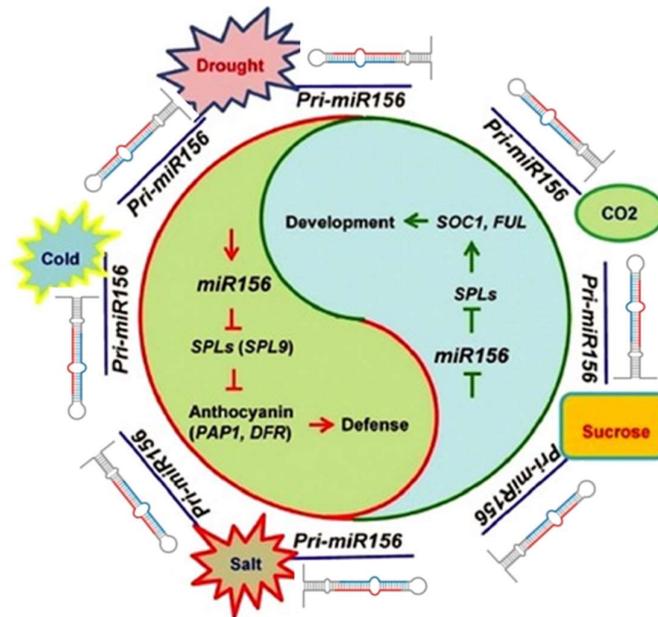
Figure 13: A working model of miR156-*SPL* module in coordinating the plant growth and development and plant tolerance to abiotic stresses. Under stress conditions, miR156 is induced, causing plant tolerance to stresses. In contrast, under appropriate conditions, miR156 is supressed, causing normal plant growth and development (adapted from (Cui *et al.*, 2014). The images of pri-miRNA are created with BioRender.com.

Apart from its involvement in vegetative development, miR156-*SPL9* module in Arabidopsis is also implicated in cold stress tolerance. AtSPL9 has been shown to activate the expression of *CBF2* gene from C-REPEAT BINDING FACTOR/DRE BINDING FACTOR1 (CBF/DREB1) family upon low temperature conditions, which led to enhanced cold tolerance of Arabidopsis plants (Zhao *et al.*, 2022) (Fig. 14).
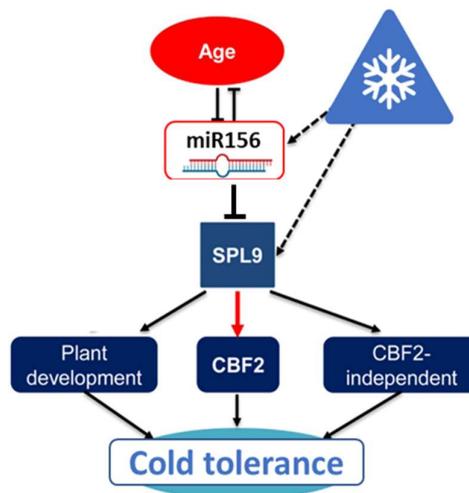
Figure 14: A working model of involvement of miR156-*SPL* modules in providing cold tolerance in Arabidopsis (adapted from (Zhao *et al.*, 2022). The image of miRNA is created with BioRender.com

Also, in rice, the regulation of several *OsSPL* genes expression by miR156 is involved in various stress responses including salinity, stress and cold stress (Wang and Wang, 2015; Cui *et al.*, 2015; Lan *et al.*, 2019). Cold stress affects plant growth and development and causes loss in crop yields (Sun *et al.*, 2020). It is especially threatening for early stages of plant development as it inhibits seedling growth. It also has adverse effects on reproductive development of rice. For example, it negatively affects plant height, panicle architecture, grain length, and anther's length and volume in rice (Cui *et al.*, 2015). In rice, the overexpression of miR156k caused reduced tolerance to cold stress in the seedling stage of rice. This overexpression led to changes in the expression of some cold responsive genes and a decrease in the expression of their target *SPL* genes (*OsSPL3, OsSPL14* and *OsSPL17*) (Cui *et al.*, 2015).

Salinity is one of the major stresses affecting the plant growth and development hence, resulting in loss of yield in crop species. Salt tolerance in rice is a complex network involving many proteins, transcription and epigenetic factors. *OsSPL10* was identified as a candidate for *SST* (*SEEDLING SALT TOLERANT*) gene in playing a role in salt tolerance and trichome development in rice. The knockout mutant of Os*SPL10* displayed more tolerance to salt stress with glabrous traits, as opposed to plants overexpressing Os*SPL10*. Glabrous rice varieties are desirable, and they are characterized by phenotypes like glabrous leaves and glumes without trichomes. Since trichomes causes dust formation during harvesting and grain manipulation, hence, glabrousness is considered an important agronomic trait in rice (Lan *et al.*, 2019). Therefore, *OsSPL10* was shown to negatively influence salt tolerance in rice (Lan *et al.*, 2019).

Apart from its involvement in salt stress, Os*SPL10* is also involved in regulating drought stress. Low level of Os*SPL10* expression in rice plants prevents reactive oxygen species (ROS) accumulation and programmed cell death processes by positively regulating the expression of its downstream gene Os*NAC2*. Additionally, knock-out of Os*SPL10* caused earlier closure of stomata and hence, prevented water loss (Li *et al.*, 2023*a*). Hence, Os*SPL10* has been shown to be a valuable target for crop improvement by imparting salt and drought tolerance in rice (Lan *et al.*, 2019; Li *et al.*, 2023*a*).

Heat stress is known to be a severe hazard to crop production worldwide. The heat damage to crop plants in the flowering season is quite prevalent as plants show higher sensitivities to increasing temperature during their reproductive development. Various mechanisms in plants have been evolved to prevent them from heat damage including heat-shock proteins (HSPs), heat-shock transcription factors (HSFs), reactive oxygen species (ROS), unfolded

protein structures, secondary messengers of calcium and phytohormone signaling (Hirt and Shinozaki, 2003; Zhao *et al.*, 2020; Suzuki *et al.*, 2022; Wang *et al.*, 2023*b*). The SPL family of transcription factors is involved in conferring heat stress tolerance (thermotolerance) in plants. For example, *SPL1* and *SPL12* genes from Arabidopsis have been shown to act redundantly to impart thermotolerance at their reproductive stage of growth during heat stress. The double knockout of both genes caused extreme plant sensitivity to heat stress, as opposed to *SPL1* and *SPL12* overexpressing mutants. Additionally, heat stress induced a large number of abscisic acid (ABA) signaling pathway genes in the double knockout plants. Therefore, both *SPL1* and *SPL12* genes confer thermotolerance via ABA signaling pathway in Arabidopsis (Chao *et al.*, 2017).

Many essential nutrients including inorganic phosphate (Pi) and copper (Cu) are required for proper plant development and hence, act as limited factors in plant growth. Under Pi and Cu deficit conditions, plants display deviations in their morphology and signaling processes. Several transcription factor genes, including *SPLs,* modulate these stress responses by monitoring the expression of Pi- and Cu-deficit responsive genes. In Arabidopsis, *SPL3, SPL7* and *SPL9* are implicated in this process (Yamasaki *et al.*, 2009; Lei *et al.*, 2016, 2022). Under Pi deficit conditions, the wild-type plants displayed higher levels of anthocyanins. Under these conditions in wild-type plants, the expression of miR156 was elevated while it's targeted several *SPL* transcripts were repressed, which includes At*SPL3* and At*SPL9*. Moreover, over-expressing plants of both these genes displayed higher Pi and lower anthocyanin levels as compared to wild-type plants. Hence, both *SPL3* and *SPL9* genes are involved in low Pi stress conditions in Arabidopsis (Lei *et al.*, 2016, 2022) (Fig. 15).
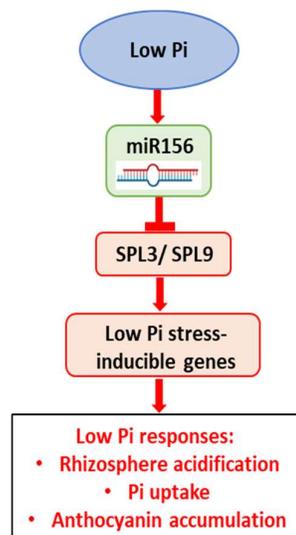
Figure 15: A working model of involvement of miR156-*SPL* modules in low Pi stress conditions in Arabidopsis. The image of miRNA is created with BioRender.com.

In the case of Cu deficient conditions, At*SPL7* caused the induction of several copper-responsive-miRNAs: miR397, miR398, miR408 and miR857 (Perea-García *et al.*, 2021; Lei *et al.*, 2022). These Cu-miRNAs under Cu-deficient conditions degrade transcripts of excessive cuproproteins to balance the levels of Cu under scarcity. (Ana Perea-Garcia et al., 2021; Lei KJ et al., 2022).

Hence, apart from regulating plant architecture and reproductive development in plants, *SPL* genes are also important factors in plant's response to different stresses. These studies can thus be exploited in agriculture to prepare varieties with enhanced agronomic traits including increased crop yield and stress tolerance.

1.2.2.3 Functions of *SPL* gene family in imparting plant immunity in angiosperms

The age-dependent plant immunity is related with vegetative phase change, during which the plant gains resistance to diseases against many pathogens. The timing of activation of this age-dependent plant immunity is related to decrease in miR156 levels. In Arabidopsis, *SPL10* along with *SPL2* and *SPL11* genes have been shown to activate the age-dependent immunity in adult stage of their life. Furthermore, *SPL10* gene is shown to provide this immunity against bacterial pathogen *Pseudomonas syringae pv. tomato* DC3000 via direct binding of SPL10 to promoter of *PAD4* (*phytoalexin deficient 4*), a component of salicylic acid (SA) signaling pathway (Hu *et al.*, 2023*b*).

The miR156-*SPL9* module was shown to impart plant immunity against this bacterial pathogen also via activation of SA signaling pathway but additionally, through accumulation of reactive oxygen species (ROS). In this study, miR156 overexpressing plants displayed elevated levels of ROS and lower transcript levels of SA signaling pathway genes, as compared with miR156 loss-of-function and *SPL9* gain-of-function transgenic plants. As a result, miR156 loss-of-function and *SPL9* gain-of-function transgenic plants displayed increased resistance to this bacterial pathogen as compared to miR156 overexpressing plants (Yin *et al.*, 2019).

In another study, miR156 loss-of-function and *SPL9* gain-of-function plants showed increased sensitivity to a fungus *Botrytis cinerea* as compared to the wild-type plants. This miR156-*SPL9* module imparts fungal resistance via SQUINT (SQN), a component of jasmonic acid (JA) pathway in Arabidopsis. The *sqn* mutants were insensitive to any JA responses or attack against this fungus, but these phenotypes were restored after constitutive

expression of miR156, which further reduced the levels of *SPL9* (Sun *et al.*, 2022) . Thus, *SPL* genes are involved in imparting plant resistance against various pathogens via different regulatory pathways in Arabidopsis.

Rice production and yield is highly affected by blast disease, which is caused by fungus, *Magnaporthe oryzae*. Many rice miRNAs have been shown to be responsive to infection during blast disease. Osa-miR535, which targets Os*SPL4* and Os*SPL14,* belongs to this group. Rice transgenic plants over-expressing miR535 showed reduced blast resistance. Additionally, these transgenic lines had increased tillers, reduced size of panicles and less filled grains. These phenotypes were mainly due to downregulation of both Os*SPL14* and Os*SPL4* genes (Zhang et al., 2022), especially that Os*SPL14* has been shown to enhance grain yield in rice and provide resistance against blast disease by activation of *WRKY45* expression (Wang *et al.*, 2018). Hence, Os*SPL14* gene regulates growth, production, and immunity in rice (Shimono *et al.*, 2007; Miura *et al.*, 2010; Zhang *et al.*, 2022). Furthermore, Os*SPL4* promotes the expression of *GH3.2* gene encoding, IAA-amido synthetase, which is involved in promoting immunity against this fungus. Additionally, the expression levels of *GH3.2* transcripts were elevated in Os*SPL4* over-expressing plants, showing that *GH3.2* acts downstream of Osa-miR535-targeted Os*SPL4*. Therefore, Osa-miR535-Os*SPL4*-GH3.2 module acts in parallel to Osa-miR535-Os*SPL14*-WRKY45 module to orchestrate together the immunity against blast disease (Zhang *et al.*, 2022) (Fig. 16).
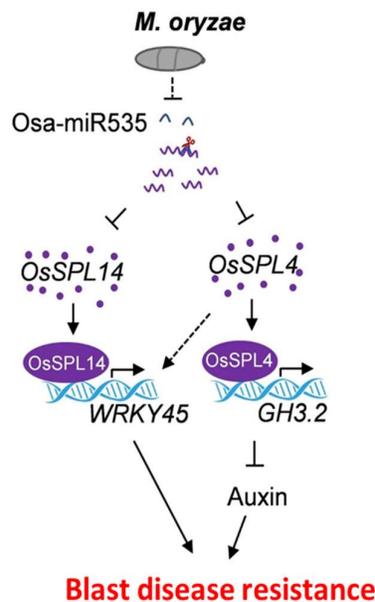
Figure 16: A schematic representation of Osa-miR535-*SPL* module in imparting immunity against *M. oryzae* in rice (adapted from (Zhang *et al.*, 2022).

Hence, SPL transcription factors impart immunity against diseases in angiosperms by activating similar signaling pathways, SA and JA signaling pathways. Moreover, SPL transcription factors act as potential targets for improving disease resistance in crops.

1.2.3 Functions of *SPL* gene family in bryophytes

So far, *SPL* transcription factor gene family has been described in only few bryophyte species, including moss *P. patens* ('Comparative analysis of the SBP-box gene families in P. patens and seed plants', 2007), liverwort *M. polymorpha* (Streubel *et al.*, 2023; Alisha *et al.*, 2023) and hornworts *Anthoceros agrestis* and *Anthoceros angustus* (Streubel *et al.*, 2023; Alisha *et al.*, 2023). However, their functions have been characterized only for three out of 13 moss *SPLs* and two out of four liverwort *SPLs,* respectively. Also, in this clade of land plants, some members of *SPL* family are post-transcriptionally regulated by miR156/529.

In moss *P. patens,* both miRNAs are present, but expressed in different developmental stages, with miR156 primarily expressed in protonema and miR529 mainly expressed in gametophores with mature sporophytes (Xie et al., 2021). Moreover, in moss, miR156 promotes transition from young protonema to leafy gametophores, as opposed to flowering plants, where it acts as an inhibitor of phase transition (Cho *et al.*, 2012). Three Physcomitrium *SPL* family members have been identified to be targets for miR156 (Pp*SBP3*, Pp*SBP6* and Pp*SBP13*) (Arazi *et al.*, 2005; Riese *et al.*, 2007; Alisha *et al.*, 2023). From these, Pp*SBP3* have been functionally characterized. Pp*SBP3* has been shown to regulate gametophore production in moss. The loss-of-function *ppsbp3* mutant plants displayed an increased number of leafy buds, a phenotype opposite to plants with functional inhibition of miR156. Hence, miR156-Pp*SBP3* module controls the developmental timing for buds and leafy gametophores production in moss (Cho *et al.*, 2012).

In the case of miRNA non-targeted moss *SPL* genes, Pp*SBP1* and Pp*SBP4* have been shown to be involved in blue light signaling pathway. In *ppsbp1* and *ppsbp4* knockout mutants, enhanced branching phenotype was observed during protonema growth (Riese *et al.*, 2008). Interestingly, an opposite phenotype was observed in a double mutant cryptochrome blue-light receptor genes, *ppcry1a/1b* (Imaizumi et al., 2002). Moreover, the expression of both Pp*SBP1* and Pp*SBP4* genes was upregulated in *ppcry1a/1b* mutant. Therefore, these two *SPL* members are negatively regulated by blue-light cryptochrome receptor (Pp*CRY)* and are involved in phototransduction in *P. patens* (Riese *et al.*, 2008).

Similarly, as in angiosperms, one of the *SPL* family member is involved in copper homeostasis in *P. patens*, namely Pp*SBP2* (Nagae *et al.*, 2008). PpSBP2 regulates copper

homeostasis by negatively regulating the expression of iron superoxide dismutase (FeSOD) by binding to GTACT-motif in its promoter. Furthermore, moss *FeSOD* gene was shown to be repressed by Cu in *Nicotiana tobacum* transgenic plants. Additionally, *FeSOD* gene promoter was also shown to be repressed by Cu in Arabidopsis. Therefore, molecular mechanisms of GTACT-motif containing genes involved in transcriptional regulation of copper were found to be evolutionary conserved in land plants (Nagae *et al.*, 2008).

In liverwort *M. polymorpha, SPL* gene family consists of four members (Tsuzuki *et al.*, 2016, 2019; Streubel *et al.*, 2023; Alisha *et al.*, 2023). A similar set of *SPL* gene family has also been identified in two hornworts, *A. agrestis* and *A. punctatus* (Streubel *et al.*, 2023; Alisha *et al.*, 2023). Therefore, the *SPL* families from liverwort and hornworts represent the simplest and the smallest set of *SPL* genes identified in land plants thus far. In Marchantia, Mp*SPL1* and Mp*SPL2* belong to the genes targeted by miRNA. Unlike other land plants, where *SPL* family members are majorly targeted by miR156, Mp*SPL1* and Mp*SPL2* are targeted by two distinct miRNAs. Mp*SPL1* is targeted by a liverwort-specific miRNA, Mpo-MR-13 (also named as MpmiR11671) while Mp*SPL2* is targeted by miR529c (Tsuzuki *et al.*, 2016, 2019; Streubel *et al.*, 2023). Both these miRNA-*SPL* modules have been functionally characterized (Tsuzuki *et al.*, 2019; Streubel *et al.*, 2023). The null mutants of miR529c showed similar phenotype as transgenic plants in which miR529-resistant Mp*SPL2* copy was introduced to the genome. Both transgenic plants developed reproductive structures even in the absence of far-red light which is required for gametangiophores induction. Hence, lack of miR529 released the repression of Mp*SPL2*, promoting plants to undergo transition to reproductive phase. Therefore, MpSPL2 is an important regulator of vegetative-to reproductive transition in *M. polymorpha* in response to inductive light conditions. Interestingly, Mp*SPL2* knock-out plants developed reproductive structures which produced fertile gametes. Hence, Mp*SPL2* was shown to be only involved but not essential for sexual reproduction in Marchantia. Therefore, miR529-*SPL2* module in *M. polymorpha* seems to play a similar role to miR156-*SPL* module in higher land plants (Tsuzuki *et al.*, 2019).
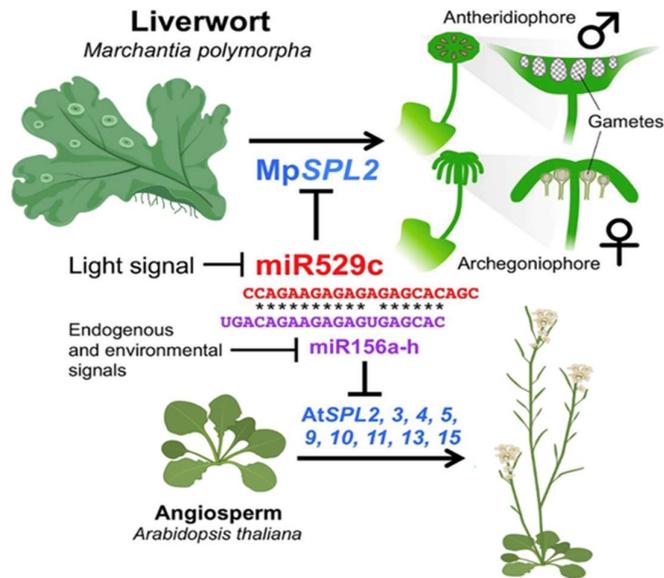
Figure 17: The miR156/529-*SPL* module is involved in reproductive transition in both *M. polymorpha* and *A. thaliana*. Schematic of roles of miR529c-*SPL* module in Marchantia and miR156-*SPL* module in Arabidopsis (adapted from (Tsuzuki *et al.*, 2019). Pictures of vegetative stages of Arabidopsis and Marchantia and reproductive stage of Arabidopsis plant are created with BioRender.com.

In the case of Mp*SPL1* gene which is regulated by Mpo-MR-13 it has been shown that it regulates branching architecture of Marchantia thallus. The branching architecture is determined by meristems developing at apices of the thallus. The active meristems will develop into branches as opposed to the dormant meristems, hence defining the shape of the thallus. Mp*spl1* loss-of-function (lof) mutants had no dormant meristems as compared to wild-type plants. On contrary, Mp*spl1* gain-of function and Mpo-*mr-13*[lof] mutant plants developed more dormant meristems (early dormancy), a phenotype opposite to Mp*spl1* mutants. Additionally, this study demonstrated that Mpo-MR-13-Mp*SPL1* module is controlled by PIF (PHYTOCHROME-INTERACTING FACTOR) - mediated signaling. Therefore, Mp*SPL1*-Mpo-MR-13 module acts in regulating meristem dormancy in *M. polymorpha* which is dependent on the light conditions (Streubel *et al.*, 2023) (Fig. 18).
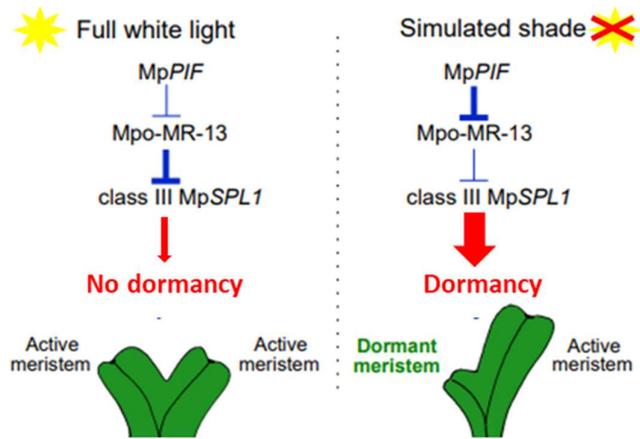
Figure 18: A schematic representation of role of Mpo-MR-13-Mp*SPL1* module in apical dominance in *M. polymorpha*. Under full white light, Mp*PIF* is inactive and hence, Mpo-MR-13 represses Mp*SPL1* while under simulated shade conditions, Mp*PIF* is active and it represses Mpo-MR-13 hence, Mp*SPL1* becomes active which induces dormancy in Marchantia thallus (adapted from (Streubel *et al.*, 2023).

2. MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Plant material and growth conditions

All the experiments were performed using Takaragaike-1 (Tak-1) as male and Takaragaike-2 (Tak-2) as female accessions. The Tak-1 and Tak-2 samples were brought from Prof. Takayuki Kohchi laboratory in Japan (Graduate School of Biostudies, Kyoto University, Japan).

For soil culture: Gemmae were first cultured at 22°C on Jiffy-7® (Jiffy International AS) 42mm peat pellets for 14 days. The plants were kept under continuous light of 50-60 µmol m^{-2}s^{-1}, by LED Neonica Growy (Neonica Polska). After 14 days of continuous light, the plants were transferred to a 16h-light/8h-dark conditions for the induction of gametangiophore development. The light during this induction phase was provided by diodes (LED Engin; OSRAM GmbH) with ~30 µmol m^{-2}s^{-1} deep red (~660nm), ~30 µmol m^{-2}s^{-1} high power blue (1-5W; ~470nm) and 20-30 µmol m^{-2}s^{-1} FR light (735-740nm).

For *in-vitro* culture: Gemmae were cultivated on solid half-strength Gamborg's B5 salts (Sigma Aldrich) supplemented with 1% (w/v) sucrose (Sigma Aldrich), and with 0.8% (w/v) agar medium (Carl Roth Laboratories). The plants were cultured in petri dishes (Sarstedt) and duchefa boxes (Unimarket) added in MLR-350H Sanyo versatile environmental test growth chamber (Panasonic cooperation) at 23°C under continuous light of 50-60 µmol m^{-2}s^{-1}.

2.1.2 Bacterial strains and growth conditions

Bacterial strains used in the experiments:

- *Escherichia coli* **DH5α**: for plasmid amplification and cloning protocols
- *E. coli* **DB3.1**: for plasmid amplification and cloning protocols
- *Agrobacterium tumefaciens* **GV3101**: for transformation of *M. polymorpha* sporelings

E. coli DH5α and DB3.1 were cultured on LB (Luria-Bertani) medium supplemented with 1.5% agar at 37°C. *A. tumefaciens* was cultured for 2 days at 28°C. Additionally, shaking was used for liquid cultures.

2.1.3 Solutions and buffers

2.1.3.1   Medium for plant culture

0M51C medium

For preparing 0M51C medium, following buffers were prepared:

a)   1000X Gamborg's B5 micro-elements stock solution:

Component	Final concentration	Amount
$Na_2MoO_4x2H_2O$ (Sigma Aldrich)	1 mM	12.5 mg
$CuSO_4x5H_2O$ (Sigma Aldrich)	0.01 mM	1.25 mg
$CoCl_2x6H_2O$ (Sigma Aldrich)	0.01 mM	1.25 mg
$ZnSO_4x7H_2O$ (Sigma Aldrich)	6 mM	100 mg
$MnSO_4xH_2O$ (Sigma Aldrich)	50 mM	500 mg
H_3BO_3 (Sigma Aldrich)	40 mM	150 mg
MilliQ H_2O	-	Up to 50 ml

All the components were dissolved completely one by one in the given order. The solution was then autoclaved at 121ºC for 20 min.

b)   Gamborg's B5 vitamin mix solution:

Component	Final concentration	Amount
Nicotinic acid (Sigma Aldrich)	8.12 μM	0.5 mg
myo-Inositol (Sigma Aldrich)	554.94 μM	50 mg
Thiamine HCl (Sigma Aldrich)	29.65 μM	50 mg
Pyridoxine HCl (Sigma Aldrich)	4.86 μM	0.5 mg
MilliQ H_2O	-	Up to 500 ml

All the components were mixed, and the solution was autoclaved at 121ºC for 20 min.

c)   10X 0M51C medium:

Component	Final concentration	Amount
KNO_3 (Sigma Aldrich)	200 mM	10 g
EDTA (Thermo Fisher)	1 mM	0.2 g
$CaCl_2x2H_2O$ (Sigma Aldrich)	20 mM	1.5 g
KH_2PO_4 (Sigma Aldrich)	20 mM	1.375 g
NH4NO3 (Sigma Aldrich)	40 mM	2 g
$MgSO_4x7H_2O$ (Sigma Aldrich)	15 mM	1.85 g
0.075% KI solution (Sigma Aldrich)	0.00075%	5 ml
1000X Gamborg's B5 micro-elements	-	5 ml
Gamborg's B5 vitamin mix	-	5 ml
MilliQ H_2O	-	Up to 500 ml

All the components were mixed properly. The aliquots of 25-50 ml were stored in -20ºC until further use.

For preparation of 0M51C medium, following components were added:

Component	Final concentration	Amount
10X 0M51C medium	1X	100 ml
L-glutamine (Sigma Aldrich)	4 mM	0.6 g
Sucrose (Sigma Aldrich)	110 mM	40 g
Casein hydrolysate (Sigma Aldrich)	0.2%	2 g
MilliQ H_2O	-	Up to 1000 ml

The pH was adjusted to 5.5 with 1M or 0.1M KOH. The medium was autoclaved at 121°C for 20 min.

2.1.3.2 Medium for bacterial culture

LB broth medium

Components	Final concentration	Amount
Bactotryptone (BioShop)	1%	1 g
Yeast extract (BioShop)	0.5%	0.5 g
NaCl (Sigma)	1%	1 g
Distilled H_2O	-	Up to 100 ml

The medium was autoclaved for 20 min at 121°C.

LB agar medium

For the preparation of LB agar medium, LB broth medium was supplemented with 1.5% agar. The medium was then autoclaved for 20 min at 121°C.

The medium was cooled to 50°C approximately before adding the appropriate antibiotics. LB agar medium after cooling was poured to sterile petri dishes and stored at 4°C until further use.

2.1.3.3 Antibiotic solutions

Antibiotic	Manufacturer	Stock concentration	Final concentration
Kanamycin	BioShop Canada	50 mg/ml	50 µg/ml
Ampicillin	BioShop Canada	50 mg/ml	50 µg/ml
Gentamycin	Sigma Aldrich	25 mg/ml	25 µg/ml
Rifampicin	BioShop Canada	50 mg/ml	100 µg/ml

2.1.3.4 Electrophoresis solutions

10X TAE buffer

Components	Final concentration	Amount
Tris base (Sigma)	400 mM	48.5 g
Glacial acetic acid (Sigma)	200 mM	11.4 ml
EDTA (Thermo Scientific)	10 mM	20 ml
Distilled H_2O	-	Up to 1000 ml

Agarose gel

Components	Final concentration	Amount
Agarose (Prona Agarose)	1.5%	1.5 g
10X TAE	1X	10 ml
Distilled H_2O	-	Up to 100 ml

The components after adding to the 250ml flask were boiled in a microwave until completely dissolved. The solution was cooled down to 50°C before adding Ethidium Bromide (EtBr) (Sigma) to a final concentration of 0.05mg/100ml.

2.1.3.5 DNA isolation solutions

1M Tris buffer

Component	Final concentration	Amount
Tris base (Sigma Aldrich)	1 M	12.10 g
MilliQ H_2O	-	Up to 100 ml

All the components were mixed and pH of solution was adjusted to 9.5 with 5M HCl.

CTAB extraction buffer

Component	Final concentration
Tris-HCl; pH 8.0	100 mM
EDTA; pH 8.0	20 mM
NaCl	1.4 M
β-mercaptoethanol*	2% (v/v)
PVP (polyvinylpyrrolidone)*	2% (w/v)
CTAB (hexadecyltrimethylammonium bromide)*	2% (w/v)

* β-mercaptoethanol, PVP, and CTAB should be added just before use.

DNA extraction buffer for genotyping

Component	Final concentration	Amount
1M Tris base; pH 9.5 (Sigma Aldrich)	0.1 M	5 ml
0.5M EDTA; pH 8.0 (Thermo Fisher)	10 mM	1 ml
3M KCl (Sigma Aldrich)	1 M	16.6 ml
MilliQ H$_2$O	-	Up to 50 ml

All the components were mixed and filtered with MILLEX-HP filters (0.45 μm) (Millipore).

DNA loading buffer: 2X HSE buffer was prepared as following:

Components	Final concentration
Urea (Sigma)	4M
EDTA (Thermo Scientific)	0.05M
Sucrose (Sigma)	50%
Xylencyanol (Sigma)	0.1%
Bromophenol blue (Sigma)	0.1%

The components were mixed in sterile milliQ water, aliquoted in 2ml microcentrifuge tubes and stored at -20ºC until further use.

2.1.3.6 RNA isolation solutions

DEPC-treated RNAse free water: To 1l of milliQ water, 1ml of Diethyl pyrocarbonate (DEPC) (Sigma Aldrich) was added and incubated at RT under fume hood for overnight. Next day, the water was autoclaved twice at 121ºC at 20 min.

Trizol-like reagent: The solution was prepared by mixing the following components:

Component	Final concentration	Amount
Ammonium thiocyanate (Sigma Aldrich)	0.4 M	19.03 g
Guanidine thiocyanate (Sigma Aldrich)	0.8 M	29.54 g
3M sodium acetate (Thermo Fisher)	0.1 M	8.35 ml
Sodium acetate saturated phenol for RNA extractions	38%	95 ml
Glycerol (Sigma Aldrich)	5%	12.5 ml
DEPC-treated H$_2$O	-	Up to 250 ml

To prepare sodium acetate saturated phenol solution for RNA extraction, following components were added:

a.   1% Roti®-Aqua-Phenol (Carl Roth Laboratories)

b.   0.05M sodium acetate (Thermo Fisher)

c.   0.01% 2-mercaptoethanol (Sigma Aldrich)

Preparation: The solution containing 1% Roti®-Aqua-Phenol and 0.05M sodium acetate was made in DEPC-treated water in a dark bottle and kept in 4°C for 4 hours, then heated to RT. Next, most of the water phase above the phenol was removed leaving approximately only 1cm. To this, 0.01% 2-mercaptoethanol was added.

RNA loading buffer: 2X RNA loading buffer was prepared as following:

Component	Final concentration
Tris-HCl; pH 7.5 (Sigma Aldrich)	0.01M
EDTA (Thermo Scientific)	2.5mM
Formamide (Sigma Aldrich)	95%
Xylencyanol (Sigma Aldrich)	0.01%
Bromophenol blue (Sigma Aldrich)	0.01%

The solution was prepared in DEPC-treated water and aliquoted in 2ml microcentrifuge tubes before storing them at -20°C until further use.

2.1.3.7 Protein isolation solutions

1M Na-phosphate buffer

Component	Final concentration	Amount
1M Na_2HPO_4 (Sigma Aldrich)	0.39 M	19.5 ml
1M NaH_2PO_4 (Sigma Aldrich)	0.61 M	30.5 ml
MilliQ H_2O	-	Up to 50 ml

The pH was adjusted to 7.0 and all the components were mixed properly and filtered with MILLEX-HP filters (0.45 μm).

Grinding buffer

Component	Final concentration	Amount
1M Na-phosphate buffer, pH=7.0	100 mM	10 ml
1M DTT (Carl Roth Laboratories)	10 mM	100 μl[#]
EDTA-free protease inhibitor cocktail (Roche)		8 tablets[##]
100% glycerol	20%	20 ml
MilliQ H_2O	-	Up to 100 ml

All the components were mixed properly and filtered with MILLEX-HP filters (0.45 μm).

[#] Added just before protein isolation procedure.
[##] Tablets were first dissolved in ~2ml of milliQ H_2O before adding it to buffer. It was added just before isolation procedure.

6X SDS Sample buffer

Component	Final concentration	Amount
1M Tris-Cl, pH=6.8	0.375 M	7.5 ml
DTT (Carl Roth Laboratories)	0.6 M	1.86 g
SDS (BioShop)	12%	2.4 g
100% glycerol	60%	12 ml
Bromophenol Blue (POCH S.A.)	0.06%	12 mg
MilliQ H_2O	-	Up to 20 ml

All the components were mixed properly and filtered with MILLEX-HP filters (0.45 μm) and stored in 0.5ml aliquots in -20°C.

30% acrylamide solution

Component	Final concentration	Amount
1M Tris-Cl, pH=6.8	4 M	14.61 g
DTT (Carl Roth Laboratories)	50 mM	0.39 g
MilliQ H_2O	-	Up to 50 ml

All the components were mixed properly and filtered with MILLEX-HP filters (0.45 μm).

1M Tris-Cl buffer

Component	Final concentration	Amount
Tris Base (Sigma 7-9®)	1 M	12.114 g
MilliQ H_2O	-	Up to 100 ml

The pH was adjusted to 6.8 and 8.6, using 5M or 1M HCl. All the components were mixed properly and filtered with MILLEX-HP filters (0.45 μm).

5% stacking gel

Component	Final concentration	Amount
1M Tris-Cl, pH=6.8	132 mM	1.32 ml
10% SDS	0.01%	105 μl
30% acrylamide	5%	1.8 ml
10% APS	0.06%	60 μl
TEMED (Sigma)	0.12%	12 μl
MilliQ H_2O	-	Up to 100 ml

All the components were mixed properly and filtered with MILLEX-HP filters (0.45 μm).

10% separating gel

Component	Final concentration	Amount
1M Tris-Cl, pH=8.6	365 mM	7.3 ml
10% SDS	0.01%	200 µl
30% acrylamide	10%	6.6 ml
10% APS	0.16%	160 µl
TEMED (Sigma)	0.08%	16 µl
MilliQ H_2O	-	Up to 20 ml

All the components were mixed properly and filtered with MILLEX-HP filters (0.45 µm).

10X Laemmli buffer

Component	Final concentration	Amount
Tris Base (Sigma 7-9®)	250 mM	3 g
SDS	35 mM	5 g
Glycine	1.9 M	72 g
MilliQ H_2O	-	Up to 100 ml

All the components were mixed properly and filtered with MILLEX-HP filters (0.45 µm).

Semi-wet transfer buffer

Component	Final concentration	Amount
Tris Base (Sigma 7-9®)	25 mM	0.3 g
10% SDS	1%	1 ml
Glycine	200 mM	1.4 g
Methanol	10%	10 ml
MilliQ H_2O	-	Up to 100 ml

All the components were mixed properly and filtered with MILLEX-HP filters (0.45 µm).

10X TBS buffer

Component	Final concentration	Amount
Tris Base (Sigma 7-9®)	200 mM	12 g
NaCl	1.5 M	44 g
MilliQ H_2O	-	Up to 500 ml

All the components were mixed properly and filtered with MILLEX-HP filters (0.45 µm).

TBS-T buffer

Component	Final concentration	Amount
10X TBS	1X	50 ml
Tween 20	0.1%	0.5 ml
MilliQ H_2O	-	Up to 500 ml

All the components were mixed properly and filtered with MILLEX-HP filters (0.45 µm).

2.1.3.8 GUS staining solutions

Phosphate buffer

Component	Final concentration	Amount
1M Na_2HPO_4 (Sigma Aldrich)	0.72 M	36 ml
1M NaH_2PO_4 (Sigma Aldrich)	0.29 M	14.5 ml
MilliQ H_2O	-	Up to 50 ml

All the components were mixed properly and filtered with MILLEX-HP filters (0.45 µm).

GUS premix solution

Component	Final concentration	Amount
10% Triton X-100 (Sigma Aldrich)	0.1%	0.1 ml
0.5 M $NaPO_4$; pH 7.2	50 mM	10 ml
0.5 M EDTA; pH 8.0 (Thermo Fisher)	100 mM	2 ml
MilliQ H_2O	-	Up to 100 ml

All the components were mixed properly and filtered with MILLEX-HP filters (0.45 µm) (Millipore).

GUS staining solution

Component	Final concentration	Amount
100mM $K_4[Fe(CN)_6]$ (Sigma Aldrich)	0.5 mM	250 µl
100mM X-Gluc (Thermo Fisher)	1 mM	500 µl
100mM $K_3[Fe(CN)_6]$ (Sigma Aldrich)	0.5 mM	250 µl
GUS premix solution	-	49 ml
MilliQ H_2O	-	Up to 50 ml

GUS staining solution must be prepared just before use. Potassium-ferrocynide and ferricyanide stocks need to be kept in 4°C. Dissolve X-Gluc in DMF (#227056; Sigma Aldrich) and store at -20°C until further use.

2.1.4 Kits

Kit	Manufacturer	Objective
Direct-zol RNA Mini Prep kit	Zymo research	RNA isolation
GenElute Plasmid Mini Prep kit	Sigma	Plasmid isolation
GenElute Gel extraction kit	Sigma	DNA extraction from agarose gel
GenElute PCR Clean-up kit	Sigma	Cleanup of PCR products
Gateway LR Clonase II enzyme mix	Thermo Fisher	Gateway cloning
TURBO DNAse free kit	Thermo Fisher	DNAse treatment of RNA

2X Power SYBR Green PCR Master Mix	Applied Biosystems	Quantitative Real time PCR
SMARTer® RACE 5'/3'	Takara Bio	RACE amplification reactions

2.1.5 Enzymes

Enzyme	Manufacturer	Objective
DNAse TURBO (2U/µl)	Ambion	DNAse treatment
CloneAmp HiFi Polymerase	CloneTech	Gene amplification
Advantage® 2 Polymerase mix	CloneTech	RACE reactions
DreamTaq DNA polymerase	Thermo Fisher	Gene amplification and colony PCR
SuperScript III Reverse transcriptase	Thermo Fisher	cDNA preparation
Gateway LR Clonase II enzyme mix	Thermo Fisher	Gateway cloning reactions
T4 DNA Ligase	Thermo Fisher	Ligation of insert and vector
FastAP phosphatase	Thermo Fisher	Dephosphorylate DNA ends
Restriction enzymes (Fast digest and 1U/µl)	Thermo Fisher	Digestion of vector and insert ends
RNAsein ®	Promega	Ribonuclease inhibitor

2.1.6 Vectors

Vector	Manufacturer	Objective
pENTR™/D-TOPO®	Thermo Fisher	Entry vector for gateway cloning
pGEM-T Easy	Promega	For cloning of PCR and RACE products
pMpGE_En03	Addgene (#71535)	Entry vectors for CRISPR/Cas9 reactions
pMpGE_En04	Prof. Kohchi's lab	
pBC-GE14	Prof. Kohchi's lab	
pMpGE010	Addgene (#71536)	Destination vectors for CRISPR/Cas9 reactions
pMpGE017	Prof. Kohchi's lab	
pMpGWB103	Addgene (#68557)	Destination vector for artificial miRNA reactions
pMpGWB104	Addgene (#68558)	Destination vector for promoter GUS reactions
pMpGWB110	Addgene (#68564)	Destination vectors for over-expression of *SPL3* transcript
pMpGWB111	Addgene (#68565)	
pMpGWB310	Addgene (#68638)	Destination vectors for over-expression of *SPL4* transcript
pMpGWB311	Addgene (#68639)	
pUC57-art-SPL3	GenScript Biotech	Vectors for artificial miRNA reactions
pUC57-art-SPL4		

2.1.7 List of oligonucleotides used

Name	Sequence (5' to 3')	Objective
3R-Mp030	AGGCAGAGTTAATGAGGGGACGCAGA	RACE analysis
5R-Mp030	GAGACTCGGGAGAGGGAACCAAACAA	
3nR-Mp030	CATGATGATGGCAGTGATGCAGTGG	
5nR-Mp030	AGTCTCGCGACCCCGCTTTCTACTC	
3R-Mp031	TTGGACTGTCACACACGTCAAACGAG	
5R-Mp031	CTCTCCAATCCGTGTTCACTCCCATC	
3nR-Mp031	TACTTTTGGACTGGGCTGGCCTTGTT	
5nR-Mp031	TTATCGTGCTCGTGGTTCTGATCTGC	
3R-Mp0135	TCGCCACTCGTACTTCTACCGCTACG	
5R-Mp0135	CTGTCTGTCTGTCTCGTCGTCGTCGT	
3nR-Mp0135	CCCGGCGAAAGTTATCTCCAGCAAT	
5nR-Mp0135	AGACCAGCAGCAGCAAGACAGGAATC	
3R-Mp0134	AGACAGACAGAGGAGCACGAGGAGGA	
5R-Mp0134	TCCACCGTTCACAGTCTCACGAACTC	
3nR-Mp0134	AGACGAGAAGGACGAGGACGAGGAGT	
5nR-Mp0134	TAGGTGTCCAGGGCAACTTGTGATCC	
SPL4_g1_F	CTCGATGGGACAGTGTATTGCTCC	Single gRNA for CRISPR/Cas9
SPL4_g1_R	AAACGGAGCAATACACTGTCCCAT	
SPL4_g2_F	CTCGATCATTGTCCGTGCACGCAC	
SPL4_g2_R	AAACGTGCGTGCACGGACAATGAT	
SPL4_g3_F	CTCGCTGATTTCGCAGTACTACAC	
SPL4_g3_R	AAACGTGTAGTACTGCGAAATCAG	
SPL4_g4_F	CTCGAGAGGCACAACAACCGGCGC	
SPL4_g4_R	AAACGCGCCGGTTGTTGTGCCTCT	
SPL4_g5_F	CTCGGTTCTATGTCCCCGCTGTCC	
SPL4_g5_R	AAACGGACAGCGGGGACATAGAAC	
SPL3_dn_gRNA1_F	CTCGGCCTTCATCAAACGCCGTAA	Double gRNA for CRISPR/Cas9
SPL3_dn_gRNA1_R	AAACTTACGGCGTTTGATGAAGGC	
SPL3_dn_gRNA2_F	CTCGCCGGTCATAATAGTCGTAGA	
SPL3_dn_gRNA2_R	AAACTCTACGACTATTATGACCGG	
SPL4_dn_gRNA1_F	CTCGCATTGCTGACAGTATCGGTG	
SPL4_dn_gRNA1_R	AAACCACCGATACTGTCAGCAATG	
SPL4_dn_gRNA2_F	CTCGCCCATCTGTGACTTCGATGA	
SPL4_dn_gRNA2_R	AAACTCATCGAAGTCACAGATGGG	
Pro_SPL3_TF	CACCATGTGCTTTTGAGAATTTAAAACA	
Pro_SPL3_TR	GATGTGCCGTCAAGAAAGCC	

pro_SPL4_TF	CACCGAGTGACTTCGATCCGAAGT	pENTR cloning
pro_SPL4_TR	CACAGGCCCATTCTGAAGG	
SPL3_CDS_Ox_For	CACCATGGACAGCGAGGGTGGAT	
SPL3_CDS_Ox_Rev	TTGAAACCCAAATTTCACAGAGTT	
SPL4_CDS_Ox_For	CACCATGGCACACGGGCATGAGACAG	
SPL4_CDS_Ox_Rev	CACAGGCCCATTCTGAAGGCCC	
proSPL3_1	TTGTTTCATCCACCAAAGATT	Sequencing of promoter
proSPL3_2	TTCTGTCTGATTTTAAATACGATTACA	
proSPL3_3	GCTCGCCACTCGTACTTCTA	
proSPL3_4	CATGCCAGCTCAGGTGTAGA	
proSPL3_5	GCCTCAATGGAATTGTACGG	
proSPL4_1F	TCCTCGTTGTGACATGTGGT	
proSPL4_1R	CCCGGCGTATACACAGTTCT	
proSPL4_2R	CAGAAGCATCCTTCGCCATA	
M13F_seq	GTAAAACGACGGCCAGT	Sequencing of inserts from vectors
M13R_seq	CAGGAAACAGCTATGAC	
m13F_L	GTAAAACGACGGCCAGTCTTAAG	
m13r_L	CTGCCAGGAAACAGCTATGACC	
proSPL3_cF	CGGATTCTTCCTCTCTTGGA	Colony PCR
proSPL3_cR	GATGTGCCGTCAAGAAAGCC	
proSPL4_cF	ATTCGGAGGTGAGATTT	
proSPL4_cR	CATTGTGCATGACAAAACCT	
SPL3_KO_F1	ATGCGTTTGTTGGTCAGATAGC	Genotyping
SPL3_KO_R1	GCTGCCGTTATCGTCAGTTT	
SPL3_KO_F2	GCTCCAGTGATCATGACGATCC	
SPL3_KO_R2	CTTCATCAAACGCCGTAAGGG	
SPL3_KO_F3	CTTATGCAGCGGTTCTGTCAG	
SPL3_KO_R3	CTCACTAAAGCCTCACGCTC	
SPL3_KO_INTRON	TGCCTGCAAAATCTACTGTGTG	
SPL4_KO_F1	GGATTCGAGATTCGGAGGTGAG	
SPL4_KO_R1	GCCACATTGCTGACAGTATCG	
SPL4_KO_F2	CTGGATTGCCCGAATTTTTTGG	
SPL4_KO_R2	CGAAGTCACAGATGGGATGAAAC	
SPL4_KO_F3	CGATACTGTCAGCAATGTGGC	
SPL4_KO_R3	GAAGCCTTGTCTTCTCCTTCAAG	
SPL4_KO_F4	CATCCCATCTGTGACTTCGATG	
SPL4_KO_INT_F	CAGTGAGCTGGCAAGTGTTAC	
SPL4_KO_INT_R	AAGGAAAGAAGTCGCTACACG	
new_rbm27-F	ACTTTTGCAACAGCGACTTC	

new_rbm27-R	GCCTGCAATATAGCCTTCAA	Sex determination
new_rhf73-F	GAACCCGAAACTCAGGTTTT	
new_rhf73-R	ATAACAGCCAAACGGATCAA	
SPL3_cds1_f	CACGCACAATGGATCTGC	RT-qPCR
SPL3_cds1_r	CAAAATACCGCTGGCATGA	
SPL3_cds2_f	CGATGGTGACTAGGCCACTT	
SPL3_cds2_r	AAAGTCTCACCATCGGCTGT	
SPL3_f	ATCCAGGAGAACTTCCGCAGTC	
SPL3_r	AACAACACAGCCAGGACGAATG	

2.2 METHODS

2.2.1 *M. polymorpha* transformation using *Agrobacterium*

For all the transformation experiments, protocol by Ishizaki *et al*., 2008 was followed with few modifications.

a) 1-2 sporangium were used for one transformation. The dried sporangia were collected in a 1.5 ml microcentrifuge tube.

b) To tube with sporangia, 500 µl of sterile H_2O_{milliQ} was added and the spores were crushed out of the sporangia with a sterile tip.

c) The tube was filled with 1 ml of 1% NaDCC (Sigma Aldrich) to surface-sterilize the spores.

d) The tube was vortexed for ~1 minute and centrifuged at 18,620 g for 1 minute.

e) The supernatant was removed, and the spores were washed with 1 ml of H_2O_{milliQ}. This step was repeated 3 times.

f) The spores were finally suspended in 200-500 µl of H_2O_{milliQ}. 200 µl of spore suspension was added into the 250 ml flask containing 50 ml of liquid 0M51C medium (Table 1) for 7 days under continuous white light at 22°C with 120 rpm shaking.

g) On the 5th day, a single colony of *A. tumefaciens* containing the vector of interest was inoculated in 5ml of liquid LB medium substituted with antibiotics (100 µg/ml rifampicin, 50 µg/ml spectinomycin and 25 µg/ml gentamycin). The liquid culture was incubated at 28°C with 120 rpm shaking for 2 days.

h) On the 7th day, the bacterial culture was centrifuged at 2000g for 15 minutes. The supernatant was discarded, and pellet was suspended in fresh 10 ml of liquid LB medium containing 100 µM acetosyringone (from 100mM stock solution) (Sigma Aldrich)

i) The resuspended culture was incubated at 28°C for around 5-6 hours with shaking at 120 rpm.

j) 1 ml of this *Agrobacterium*-induced medium was transferred into 50 ml of 7-day old sporelings culture. To this culture, 100 µM of acetosyringone was also added.

k) The flask was then incubated at 28°C for an additional 2 days with 120 rpm shaking.

l) The transformed sporelings (on 9th day) from each transformation were collected onto 40µm BD sterile cell strainer (Becton Dickinson). These were then rinsed with ~250ml of autoclaved milliQ H_2O.

m) The collected and washed sporelings were transferred onto petri plates containing solid half-strength Gamborg's B5 medium, 100 µg/ml cefotaxime (for inhibiting

Agrobacterium growth) and appropriate antibiotic depending upon the vector used (for selection of transformed plants).

n) Transformed plants (T1 generation) were visible after 2-3 weeks. All the genotypic and phenotypic analysis was conducted on G2 generation.

2.2.2 Bacterial transformation

A. *tumefaciens* transformation

a) The electrocompetent cells were thawed on ice for around 15-20 minutes.

b) 1µl of plasmid solution containing the desired cassette was added to these thawed cells by gentle pipetting.

c) The mixture was then transferred to the pre-chilled electroporation cuvette (Bio-Rad, USA)

d) The cuvette containing the bacterial mixture was transferred into Gene Pulser X cell system's ShockPod cuvette chamber (Bio-Rad, USA). A pulse of 2.5kV was applied and quickly 1ml of liquid LB media was added and mixed by pipetting.

e) The bacterial mixture was transferred to 1.5ml microcentrifuge tube and put into incubator for 1 hour at 28°C with shaking at 120 rpm.

f) After incubation, 150-200µl of bacterial solution was spread onto the petri plate containing solid LB media with appropriate antibiotics and put to 28°C degree chamber for 1 hour.

E. *coli* transformation

a) The chemical electrocompetent cells were thawed on ice for around 15-20 mins.

b) The vector containing the desired product/cassette was added to the thawed cells by gentle pipetting.

c) The mixture was then incubated on ice for half an hour with gentle tapping at an interval of 5 minutes each.

d) After incubation, the bacteria cells were subjected to heat shock by incubation at 42°C for 1 min.

e) The cells were next quickly placed on ice for around 2 mins.

f) To these cells, 1ml of pre-warmed (37°C) liquid LB medium was added and mixed by gently pipetting.

g) The tube was then transferred to incubator (Thermomixer Comfort, Eppendorf) set at 37°C, for 1hour with shaking at 350rpm.

h) After 1 hour, 150-200µl of bacterial solution was spread onto petri plate containing solid LB media with appropriate antibiotics to select transformed colonies.

2.2.3 DNA isolation

Genomic DNA isolation using CTAB protocol

The following steps were used for the isolation of genomic DNA from Marchantia tissue. All the centrifugation steps were performed at 19720g.

a) To each microcentrifuge tube containing 100mg of finely grounded plant material, 700µl of CTAB extraction buffer was added and vortexed to dissolve.

b) The samples were incubated at 65°C for 30-60min with occasional mixing.

c) After incubation, the samples were centrifuged for 5 mins, supernatant was transferred to a fresh sterile tube.

d) 700µl of chloroform:isoamyl alcohol (24:1) was added and mixed by inverting tube. The samples were centrifuged for 5 mins. The lower layer was discarded, and the upper phase was transferred to a fresh sterile tube.

e) 400µl of pre-chilled isopropanol (4°C) was added and the tube was inverted gently to mix all the components followed by 30 mins incubation in -20°C.

f) Next, the sample was centrifuged for 10 mins. The supernatant obtained was discarded carefully to leave only the pellet at the bottom of the tube.

g) The pellet was washed twice with 200µl of 70% ethanol and left to air-dry for 10-15 mins.

h) The DNA pellet was finally dissolved in 50-200µl of DNAse-free water (pre-warmed at 37°C).

The quality of DNA was examined by DS-11 Denovix spectrophotometer measurement (Denovix, Wilmington, Delaware, USA) and by 0.8% agarose gel electrophoresis.

DNA isolation for genotyping

a) To each collection microtube (Qiagen), a part of thallus (5mm-1cm in diameter) was collected.

b) To each microtube, 100µl of DNA extraction buffer was added together with two 3mm or one 5mm glass beads (Sigma).

c) The samples were crushed for 3 mins in TissueLyserII (Qiagen).

d) After crushing, 400µl of autoclaved milliQ water was added to each microtube and mixed properly.

2.2.4 5' and 3' RACE analysis

For both 5' and 3' RACE reactions, Advantage® 2 Polymerase Mix (Clonetech) was used. For RACE reactions, gene specific primers were designed (both 'standard' and 'nested' for first and second rounds of PCR, respectively) and the reactions were prepared according to the manufacturer's instructions. The products obtained in the PCR reaction were separated on a 1-1.5% agarose gel. Products of the appropriate length were cut from the gel and purified using GeneJET Gel Extraction and DNA Cleanup Micro Kit (Thermo Fisher). The purified products were cloned into vector pGEM T-easy (Promega). The chosen plasmids, after colony PCR, were sent for sequencing. PCR reaction. The following reaction conditions were used for first and second rounds of RACE PCR reactions:

Program 1- 1st round of RACE PCR:

Temperature	Duration	
94°C	5 min	
94°C	30 sec	
72°C	3 min	5 cycles
94°C	30 sec	
70°C	30 sec	5 cycles
72°C	3 min	
94°C	30 sec	
68°C	30 sec	25 cycles
72°C	3 min	
72°C	10 min	

Program 2- 2nd round of RACE PCR:

Temperature	Duration	
94°C	5 min	
94°C	30 sec	
68°C	30 sec	25 cycles
72°C	3 min	
72°C	10 in	

2.2.5 RNA isolation

RNA isolation was performed using Direct-zol^{TM} RNA kit (Zymo-Research) with 200mg of finely ground plant material per one isolation. To this plant material, 1ml of Trizol-like reagent was added. The samples were vortexed to dissolve all the components, incubated for 5 mins at RT and centrifuged three times for 10 mins, 4°C and at max speed each time. The

supernatant was transferred to RNase-free 2ml Eppendorf tube. For further steps of RNA isolation, manufacturer's protocol was followed. RNA hence obtained was quantified using DS-11 Denovix spectrophotometer (Denovix, Wilmington, Delaware, USA) the quality of RNA was checked using agarose gel electrophoresis. The DNAse treatment was performed using TURBO DNAse according to manufacturer's instruction (TURBO DNA-free kit, Thermo Fisher). The quality of RNA was examined again using above two methods.

2.2.6 cDNA preparation

For routine gene expression analysis, cDNA was prepared from 1µg of DNase-treated RNA as starting material and the first strand synthesis was performed using Oligo $(dT)_{18}$ primers (Thermo Fisher) and SuperScript III Reverese transcriptase (Thermo Fisher) according to manufacturer's protocol.

For 5' and 3' RACE reactions, cDNA was prepared from 1µg of DNase-treated RNA using SMARTer® RACE 5'/3' kit according to manufacturer's protocol.

2.2.7 Polymerase chain reaction (PCR)

For amplification of PCR products used for cloning and sequencing, the reactions were performed using CloneAmp HiFi PCR Premix (CloneTech) according to manufacturer's protocol.

For colony PCR and RT-qPCRs (that require no cloning and sequencing), amplifications were performed using DreamTaq polymerase (Thermo Fisher) according to manufacturer's protocol.

For genotyping PCR reactions, amplifications were performed using KAPA3G plant PCR kit (Sigma Aldrich) accordingly manufacturer's protocol.

2.2.7.1   Quantitative Real time PCR (RT-qPCR)

For reactions with quantitative real time PCR, Power YBR^{TM} Green PCR MasterMix (Applied Biosystems) or SsoAdvanced^{TM} Universal SYBR^{TM} Green Supermix (Biorad) was used. A 10µl reaction was set up according to manufacturer's protocol. The reaction was set up in a minimum of two biological and 2-3 technical replicates. All reactions were performed on Quant Studio 7 & Flex Real-Time PCR system (Thermo Fisher).

The following reaction conditions were used for RT-qPCR with SYBR Green PCR MasterMix (Applied Biosystems):

Temperature	Duration	
95ºC	10 min	
95ºC	15 sec	40 cycles
60ºC	1 min	
95ºC	15 sec	
60ºC	15 sec	Melting curve
95ºC	15 sec	

The following reaction conditions were used for RT-qPCR with SsoAdvanced Universal SYBR Green Supermix (Biorad):

Temperature	Duration	
95ºC	30 min	
95ºC	15 sec	40 cycles
60ºC	1 min	
95ºC	15 sec	
60ºC	15 sec	Melting curve
95ºC	15 sec	

The obtained results were analysed using SDS 2.4 software (Thermo Fisher). The error bars were calculated using the SD function in Microsoft Excel (year) software The expression levels of pri-miRNAs and mRNAs were calculated using the relative quantification (2-ΔCt) and the fold change values using the 2-ΔΔCt method. The statistical significance of the obtained values was determined using a Student's t-test at three significance levels: p-values with $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$.

2.2.8 Agarose gel electrophoresis

For PCR products: In case of DreamTaq PCR, 5-10µl of product was loaded directly onto the 0.8-2% of agarose gel (depending upon the length of the PCR products) containing EtBr. In case of CloneAmp HiFi PCR reactions, 5µl of product was first mixed with DNA loading buffer before loading onto the agarose gel. The electrophoresis reaction was run in 1X TBE buffer at 60-70mA in a Hoefer HE33 Mini Submarine System (Hoefer INC, Holliston, USA).

For RNA: 1-2µl of RNA was mixed with 2X RNA loading buffer, denatured at 95ºC for 2 minutes and then transferred quickly to the ice. The reaction was then run on a 2% agarose

gel in 1X TBE buffer at 60mA in a Hoefer HE33 Mini Submarine System (Hoefer INC, Holliston, USA).

After completion of the electrophoresis run, the gel was visualised with a Gene Snap software in Gene Box (Syngene).

2.2.9 Cloning reactions

To prepare the entry vectors for *in planta* over-expression experiments, the coding sequences of respective genes were amplified on cDNA template prepared from RNA isolated from 3-week-old male thallus. To prepare the entry vectors for *in planta* promoter analysis, the promoter region 4-5 kb in length upstream the start codon of respective gene was amplified on genomic DNA isolated from 3-week-old male thallus.

PCR products after amplification were run on agarose gel for electrophoresis, the products corresponding to specific required lengths were eluted and sent for sequencing. After confirming the sequences of respective products by sequencing, pENTR-dTOPO cloning was used. TOPO® cloning reaction was performed with PCR product: TOPO® vector ratio of 2:1 according to manufacturer's protocol.

For cloning to destination vectors, LR cloning reactions were performed using Gateway LR Clonase II enzyme mix according to manufacturer's protocol. For each reaction, 30 ng of an entry vector and 50 ng of destination vector was used.

2.2.10 Restriction digestion

The purified PCR products: insert as well as vector were digested using appropriate restriction enzymes. The reaction conditions were followed according to the manufacturer's instructions for each restriction enzyme. After digestion, the reactions were run on agarose gel. The obtained digested products (insert or vector or both) were eluted and purified using GeneJET Gel Extraction and DNA Cleanup Micro Kit according to the manufacturer's instructions. The concentration of samples was measured using DS-11 Denovix spectrophotometer (Denovix, Wilmington, Delaware, USA).

2.2.11 Vector dephosphorylation

In order to prevent self-ligation of vectors after restriction digestion reactions, they were treated with FastAP thermosensitive alkaline phosphatase, and the reaction was performed according to the manufacturer's instructions. Later, the samples were eluted and purified using GeneJET Gel Extraction and DNA Cleanup Micro Kit according to the manufacturer's

instructions. The concentration of samples was measured using DS-11 Denovix spectrophotometer (Denovix, Wilmington, Delaware, USA).

2.2.12 DNA ligation

The purified inserts were ligated into appropriate vectors using T4 DNA ligase (Thermo Fisher). In general, insert to vector ratio of 5:1 was used per each reaction. The reaction conditions were followed according to manufacturer's instructions. The reaction was incubated for either 1 hour at RT or overnight at 4°C and later used for the transformation reactions.

2.2.13 Plasmid DNA isolation

For isolation of plasmid DNA GeneJET Plasmid Miniprep Kit was used according to manufacturer's protocol. The concentration of the obtained DNA was measured using DS-11 Denovix spectrophotometer.

2.2.14 DNA sequencing

All routine DNA sequencing reactions were performed using Sanger method. The sequencing was performed in the Laboratory of Molecular Biology Techniques at Faculty of Biology of Adam Mickiewicz University in Poznan, Poland.

2.2.15 Protein isolation and precipitation

a) To 200 mg of grinded plant material, 800 µl of grinding buffer was added.

b) The mixture was mixed properly by vortexing and incubated at 4°C for 1 hour with shaking at 300 rpm.

c) The samples were centrifuged at 18620 g for 10 mins at 4°C.

d) The supernatant was transferred to a sterile eppendorf tube.

e) To this ~400 µl of supernatant, 100 µl of chloroform and 400 µl of methanol was added.

f) The samples were vortexed and centrifuged at 18620 g for 10 mins. at RT.

g) The supernatant was discarded and 400 µl of methanol was added. The solution was inverted few times.

h) The samples were centrifuged again at 18620 g for 10 mins. at RT.

i) The protein pellet obtained after discarding the supernatant was air-dried at RT.

j) To this dried-pellet, 100 µl of 1% SDS (containing EDTA-free protease inhibitor cocktail) was added and incubated at 99°C for 10 mins.

k) To this, 12 µl of 6X SDS sample buffer was added, and the samples were stored at -20°C.

2.2.16 Western blotting

a) 5% stacking gel and 1% separating gel was prepared for SDS-PAGE run.

b) Before adding samples to the gel, it was pre-run for 5 mins at 30 mA.

c) Each sample with 50 µl of protein extract was pre-heated at 95°C and loaded into gel, along with ladder (PageRuler™ Prestained Protein Ladder; Thermo Fisher).

d) For each gel, electrophoresis was run at 30 mA for 2.5 hours in 1X Laemmli buffer.

e) After electrophoresis, transfer was performed using semi-wet transfer buffer and membrane (Immobilon-P PVDF; Merck).

f) The transfer was performed for 1 hour at a constant 15V.

g) After completing of transfer, the membrane was incubated in 5% non-fat milk (dissolved in TBS-T buffer) overnight at 4°C on rocker-shaker.

2.2.17 Antibody staining and detection

a. 10 µl of FLAG antibody was added to 10 ml of 5% milk (freshly prepared).

b. After overnight shaking, the milk was discarded and milk containing antibody was added.

c. The membrane was incubated at RT for 1 hour and kept on rocker-shaker.

d. After incubation, the membrane was washed with TBT-T buffer for 10 mins. This washing step was repeated 3 times.

e. The membrane was incubated again for 1 hour with milk containing secondary antibody.

f. The washing step was again repeated for 3 times in TBS-T buffer.

g. After washing, 1ml of substrate (Pierce™ ECL Western Blotting Substrate) was added.

h. The membrane was incubated for 3 mins on imaging tray.

i. The images were acquired using G:Box Chemi XR5 system (Syngene; Synoptics Ltd; Scientific Digital Imaging plc.) and GeneSys software (ver.1.5.4.0).

2.2.18 Amido Black staining

After imaging, the membrane was washed in TBS-T buffer, once and incubated with 0.1% Amido Black 10B at RT for 5 mins on rocker-shaker. This staining shows RuBisCO protein bands on gel which are used as gel loading control.

2.2.19 GUS staining

For GUS staining transgenic plants containing proSPL3/SPL4:*GUS* gene were used with WT as control. For staining, protocol by Ishizaki et al., 2012 was followed with slight modifications.

a) The plant material was placed in a 12- or 24-well plate (Corning, Sigma Aldrich). To each well, 1ml of GUS staining solution was added.

b) The plants were then vacuum-infiltrated for 3 times (15 min each) inside Vacutherm™ Oven (Thermo Fisher).

c) The plants were then incubated at 37°C in incubator (New Brunswick; Eppendorf) overnight.

d) Next day, the staining solution was removed and the plants were washed with 70% ethanol for 3 times (the plants were kept for 30 min at R.T. in 70% ethanol during each time). The plants were visualised under Leica M60 microscope.

2.2.20 Microscope imaging

All the phenotypic analysis for vegetative and reproductive structures of *M. polymorpha* were performed under Leica M60 microscope (Leica Microsystems CMG GmbH) or VHX-7000 Keyence Digital microscope (Keyence). All the images were taken in Leica Application Suite (LAS) v4.5. software or VHX-7000 software, respectively.

2.2.21 CRISPR/Cas9 genome editing

For CRISPR/Cas9 by single and double gRNA approach, protocol by Sugano et al., 2018 and protocol provided by Prof. Takayuki Kohchi was followed, respectively. For single gRNA approach, five gRNAs and for double gRNA approach, two gRNAs were designed for each gene using CRISPRdirect tool (Naito *et al.*, 2015) and CRISPOR tool (Concordet and Haeussler, 2018). The off-targets were screened using BLAST in MarpolBase (which version) (Kawamura *et al.*, 2022). The gRNA spacer sequences with the highest specificity to the target sequence and with minimum off-targets were finally chosen and the 20-nt oligonucleotides were ordered for gRNA cloning.

For single gRNA approach: gRNAs were first ligated into an entry vector, pMpGE_En03 using BsaI restriction sites. The gRNA cassette was next transferred to destination vector pMpGE010 containing Cas9 cassette using LR cloning. The resultant destination vectors were further used for *Agrobacterium* mediated transformation of *M. polymorpha* sporelings.

For double gRNA approach: One gRNA was ligated into entry vector pMpGE_En04 and the second gRNA was ligated into entry vector pBC-GE14, using BsaI restriction sites. Next, both vectors were digested using BglI restriction digestion enzyme. The digestion gave two products on gel, one product at a height of ~3200 bp corresponding to gRNA1 cassette and vector backbone from pMpGE_En04 and second product at a height of ~700 bp corresponding to gRNA2 cassette from pBC-GEs. DNA was eluted from excised gel fragments and was further assembled as one vector via ligation through BglI restriction sites. The presence of both gRNAs cassettes was examined by EcoRI restriction digestion. which should give two products: 3.2 kb vector backbone containing first gRNA cassette and 0.7 kb fragment resembling second gRNA cassette. The resulting entry clone was used in LR reaction to transfer the two gRNA cassettes to destination vector pMpGE017 containing Cas9 "nickase" (Cas9^{D10A}) (Ran *et al.*, 2013; Shen *et al.*, 2014). The resultant destination vectors were used for *Agrobacterium* mediated transformation of *M. polymorpha* sporelings.

2.2.22 Artificial miRNA approach

For the artificial miRNA (amiR) approach, protocol by (Flores-Sandoval *et al.*, 2016) was followed. amiR for the gene of interest were designed using WMD3 (Web MicroRNA Designer) online tool (Mickiewicz *et al.*, 2016). The off-targets were screened using BLAST in MarpolBase version (Kawamura *et al.*, 2022). The artificial miRNA with the highest specificity to the target sequence and with no off-targets were finally chosen. In the backbone of pre-miR160, miR160 sequence was substituted by the designed 21 nucleotide sequence of artificial miRNA. amiRs were synthesized using by GenScript Biotech (Netherlands) and designed to have EcoRI/HindII sites at the 5' and 3' end of art-pre-miRNA, respectively, within the puC57 vector. All amiRs were PCR-amplified using primers ME537+ME538. PCR products were cloned into pENTR/D-TOPO and then cloned to the destination vector, pMpGWB103 using LR cloning. The resultant destination vectors used for *Agrobacterium* mediated transformation of *M. polymorpha* sporelings.

2.2.23 Phylogenetic analysis

For the phylogenetic analysis, SPL protein sequences from dicot *A. thaliana*, moss *P. patens,* liverwort *M. polymorpha*, and two hornwort species, *Anthoceros agrestis* [Bonn] and *A. punctatus* were used. *A. thaliana*, *P. patens* and *M. polymorpha* SPL sequences were retrieved from TAIR (Poole, 2007), Phytozome (Goodstein *et al.*, 2012), and MarpolBase databases (Kawamura *et al.*, 2022), respectively.

For the identification of *SPL* genes and their respective protein sequences from two Anthoceros species, the available genome sequences were used. SPL protein sequences from *A. thaliana*, *P. patens* and *M. polymorpha* were used as query against the two Anthoceros genomes. Using local BLASTp, many hits were obtained and after applying a selective criterion (bit-value score >100 and an e-value cut-off of <10^{-5}), putative SPL sequences were filtered out. In order to confirm the identity of the putative SPL protein sequences, they were analysed by Simple Modular Architecture Research tool (SMART) (Letunic *et al.*, 2004) and InterPro tool (Blum *et al.*, 2021) for the presence of their characteristic SBP-domain.

The phylogenetic tree was prepared in MEGA 11 (Tamura *et al.*, 2021). The full length SPL protein sequences were first aligned using CLUSTAL W tool available in MEGA software. After the multiple sequence alignment, the phylogenetic tree was constructed using maximum-likelihood method with a bootstrap value of 1000. CRR1 protein was used as a SPL representative from green algae, *Chlamydomonas reinhardtii* and set as an outgroup (Kropat *et al.*, 2005; Strenkert *et al.*, 2011).

2.2.24 Bioinformatic analysis of gene structure, protein motif and domain composition

The genomic and coding sequences of analysed genes were used in Gene Structure Display Server (GSDS) software (Hu *et al.*, 2015) to predict the exon-intron structures of the *SPLs*. The motif search was performed using MEME software with number of predicted motifs set to 20 (Bailey *et al.*, 2006). The co-ordinates of SBP domains within each SPL protein were obtained by Pfam database (Bateman *et al.*, 2004). The SBP domain alignment was performed using CLUSTAL W tool in Jalview software (Clamp *et al.*, 2004). For the construction of SBP-domain logo, WebLogo was used (Crooks *et al.*, 2004). For generation of *C. reinhardtii* SBP-domain logo, Chlamydomonas SPL protein sequences were retrieved from Phytozome database (Goodstein *et al.*, 2012) and confirmed using SMART tool scan (Letunic and Bork, 2018; Letunic *et al.*, 2021). After applying this criterion of selection, the candidates not containing the conserved Zn-binding motifs were removed from the analysis, and ten sequences containing both zinc-finger sites (C3H and C2HC) were selected for further analysis.

The miRNA-targeting sites in the Anthoceros *SPL* transcripts were identified by psRNATarget server (Dai and Zhao, 2011). The molecular features of Anthoceros SPL proteins including their molecular weights (Mw) and theoretical isoelectric points (p*I*) were

measured by Compute pI/Mw tool (Walker, 2007) and their subcellular localizations by WoLFSPORT (Horton *et al.*, 2007).

2.2.25 Promoter *cis*-elements analysis within the *SPL* gene sequences

For the promoter *cis*-elements analysis of *SPL* genes, a 1500bp sequence upstream of start codon of each gene were retrieved from the respective genomic resources for *A. thaliana*, *P. patens* and *M. polymorpha*. To retrieve promoter sequence from Anthoceros genomes, bedtools utilities were used (Quinlan and Hall, 2010). The *cis*-elements were predicted using PlantCARE online software (Lescot *et al.*, 2002).

2.2.26 Expression analysis of *SPL* genes in bryophytes and angiosperms

The expression data of *SPL* genes for *A. thaliana* and *P. patens* was retrieved from expression atlas databases, EMBL-EBI and PEATmoss, respectively (Lescot *et al.*, 2002; Liu *et al.*, 2012; Ortiz-Ramírez *et al.*, 2016; Fernandez-Pozo *et al.*, 2020). The expression data of *SPL* genes for *A. agrestis* and *M. polymorpha* was retrieved from published studies, (Li *et al.*, 2020*b*; Kawamura *et al.*, 2022), respectively and available at Sequence Read Archive (Leinonen *et al.*, 2010). The heat map to depict the expression profiles was created using RStudio (Verzani, 2011).

3. RESULTS

3.1 Chapter 1 – Phylogenetic, structural and functional relationships between SPL transcription factors from bryophytes and angiosperms

Plant-specific transcription factors encoded by *SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE* (*SPL*) genes are essential regulators of numerous plant developmental processes. In the first chapter of the thesis, we investigated the connections of *SPL* genes between bryophytes and *Arabidopsis thaliana* using their available genome sequences. We found four *SPL* genes in both *Anthoceros agrestis* and *Anthoceros punctatus*, which is similar to four *SPL* genes found in the liverwort *Marchantia polymorpha*. Thus, in comparison to other land plants, the examined hornwort and liverwort genomes encode a minimal set of *SPL* genes, which may reflect an archetype of *SPL*s present in the progenitor of existing embryophytes. The presence of four *SPL* phylogenetic groups with comparable exon-intron organization was found by phylogenetic and comparative gene structure analysis, with few deviations in hornworts. While we found shared protein motifs between bryophytes and Arabidopsis in three of the four evolutionary groups (Groups 2–4), the motif content differed clearly in the fourth (Group 1). Because current knowledge of *SPL* genes is primarily derived from seed plants, the comparative and phylogenetic analyses presented in this chapter provide a deeper understanding of the *SPL* gene family from some of the oldest extant land plants.

3.1.1 Identification of *SPL* genes in two Anthoceros species

The SPL protein sequences from *M. polymorpha, P. patens* and *A. thaliana* were downloaded from publicly available genomic databases according to *SPL* genes annotation. In order to identify *SPL* genes from hornwort genomes, BLASTp analysis was performed, and the resulting SPL sequences obtained were confirmed by SMART (Letunic and Bork, 2018; Letunic *et al.*, 2021) and ScanProsite (de Castro *et al.*, 2006) tools. Additionally, after some filtering criterion (i.e. removing sequences with incomplete SBP domain and redundant sequences) we obtained four *SPL* genes each in both Anthoceros genomes. Nomenclature of each *SPL* from Anthoceros was based on their identity with four *SPL* from Marchantia and named as Aa*SPL1-4* (*Anthoceros angustus*) and Ap*SPL1-4* (*Anthoceros punctatus*). While all four *SPL*s genes in *A. agrestis* generate only one transcript each, *SPL1* and *SPL2* genes in *A. punctatus* generate two transcript isoforms (Fig. 3.1). Two Ap*SPL1* and Ap*SPL2* gene transcripts, variants Ap*SPL1.2* and Ap*SPL2.2* encoding longer proteins showed higher

sequence similarity to Mp*SPL1* and Mp*SPL2* as compared to Ap*SPL1.1* and Ap*SPL2.1* variants. Therefore, Ap*SPL1.2* and Ap*SPL2.2* were selected as ApSPL1 and ApSPL2 for our further analysis.
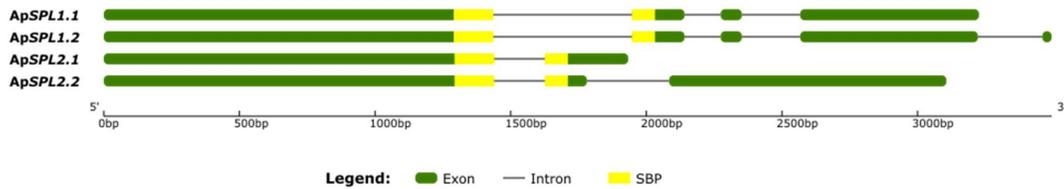
Figure 3.1: Exon-intron organization of *SPL1* and *SPL2* gene transcripts from *A. punctatus.* The gene structures were created using GSDS2.0 online server (Hu *et al.*, 2015). The exons are depicted by green boxes, conserved SBP domain by yellow boxes and introns as dark black lines in each gene model. The scale below depicts the length of genes in base pairs (bp).

Furthermore, the characteristics of each *SPL* gene identified in two Anthoceros species were determined (Table 3.1). The number of introns amongst Anthoceros *SPL*s ranged from 1 to 5 and the subcellular localization of all of them was predicted to be in the nucleus. Comparing the lengths of coding and protein sequences showed that it ranged from 774 - 2895 bp and 257 – 964 amino acids, respectively. Moreover, their molecular weights and isoelectric points varied from 27.7 – 103.7 kDa and 5.86 – 10.20, respectively. The obtained results revealed diversity within structural, physical and chemical properties in SPLs of both Anthoceros species.

Many *SPL* transcripts are known to contain target sites for conserved miRNAs (miR156/157, miR535, miR529, Mpo-mr-13). Group 1 and 2 consists of *SPL* transcripts targeted by miRNAs. Since Anthoceros SPL1 and SPL2 belong to Groups 1 and 2 respectively, we expected the presence of miRNA targeting sites within their transcripts. Meanwhile there is no micro-transcriptome data available for Anthoceros genomes. Also, we could not identify the presence of any *SPL*-targeting miRNA (miR156, miR529c or Mpo-mr-13) within Anthoceros genome. Hence, we used homology-based search tool, psRNATarget (Dai *et al.*, 2018*b*) to identify miRNA targeting sites within Anthoceros *SPL* (Aa*SPL1*, Aa*SPL2*, Ap*SPL1* and Ap*SPL2*) transcripts. For the search, we used miR156 from *Pellia endiviifolia*, miR529c and Mpo-mr-13 from *M. polymorpha* as queries (Table 3.2).

Table 3.1: Characteristics of *SPL* genes identified in two *Anthoceros* species: Aa – *Anthoceros agrestis*, Ap – *Anthoceros punctatus*.

Gene name [a]	Gene ID [b]	Exon No [c]	Transcript [d]	miR156/529c target site [e]	CDS [f] (bp)	Protein [g] (aa)	Mw [h] (kDa)	pI [i]	Subcellular Localization [j]
Aa*SPL1*	AagrBONN_evm.model.Sc2ySwM_344.856.1	6	Aa*SPL1*	No	774	257	27.67	10.2	Nucleus
Aa*SPL2*	AagrBONN_evm.model.Sc2ySwM_344.857.1	3	Aa*SPL2*	Yes	1611	536	57.08	9.11	Nucleus
Aa*SPL3*	AagrBONN_evm.model.Sc2ySwM_344.2221.1	2	Aa*SPL3*	No	1395	464	49.92	7.33	Nucleus
Aa*SPL4*	AagrBONN_evm.model.Sc2ySwM_369.244.1	3	Aa*SPL4*	No	2787	928	101.12	6.02	Nucleus
Ap*SPL1*	Apun_evm.model.utg000107l.74.1	4	Ap*SPL1.1*	No	2367	788	83.62	8.87	Nucleus
	Apun_evm.model.utg000107l.74.2	5	Ap*SPL1.2*	No	2397	798	84.65	8.87	Nucleus
Ap*SPL2*	Apun_evm.model.utg000107l.75.1	2	Ap*SPL2.1*	Yes	1746	581	61.2	8.98	Nucleus
	Apun_evm.model.utg000107l.75.2	3	Ap*SPL2.2*	Yes	2616	871	91.71	8.83	Nucleus
Ap*SPL3*	Apun_evm.model.utg000185l.396.1	2	Ap*SPL3*	No	1383	460	49.54	7.33	Nucleus
Ap*SPL4*	Apun_evm.model.utg000116l.202.1	2	Ap*SPL4*	No	2895	964	103.7	5.86	Nucleus

[a]Name referred to Anthoceros *SPL*s in this work. [b]Gene accession number in database. [c]Exon number in Anthoceros *SPL* genes. [d]Transcript name referred to Anthoceros *SPL* Gene ID. [e]Presence of the recognition site for miR156 in *SPL* transcript. [f]Length of coding DNA sequence. [g]Length of deduced SPL protein. [h]Molecular weight. [i]Theoretical isoelectric point. [j]predicted subcellular localization by WoLFPSORT tool (Horton *et al.*, 2007)

Table 3.2: Identification of miRNA binding sites in *SPL* gene transcripts of *A. agrestis* [bonn] and *A. punctatus*.

miRNA Acc.	Target Acc.	Expect	UPE	Alignment	Inhibition	Multiplicity
Pellia miR156[1]	Mp*SPL2*	1	23.287	miRNA 20 CACGAGUGAGAGAAGACAGU 1 :::::: :::::::::::: Target 1730 GUGCUCUCUCUCUUCUGUCA 1749	Cleavage	1
	Ap*SPL2*	1	20.826	miRNA 20 CACGAGUGAGAGAAGACAGU 1 :::::: :::::::::::: Target 2069 GUGCUCUCUCUCUUCUGUCA 2088	Cleavage	1
	Aa*SPL2*	1	20.826	miRNA 20 CACGAGUGAGAGAAGACAGU 1 :::::: :::::::::::: Target 1064 GUGCUCUCUCUCUUCUGUCA 1083	Cleavage	1
Mpo-miR529c[2]	Mp*SPL2*	2	17.661	miRNA 21 CGACACGAGAGAGAGAAGACC 1 ::::::::::::::::: Target 1727 GGCGUGCUCUCUCUCUUCUGU 1747	Cleavage	1
	Ap*SPL2*	2	18.611	miRNA 21 CGACACGAGAGAGAGAAGACC 1 ::::::::::::::::: Target 2066 GGCGUGCUCUCUCUCUUCUGU 2086	Cleavage	1
	Aa*SPL2*	2	18.597	miRNA 21 CGACACGAGAGAGAGAAGACC 1 ::::::::::::::::: Target 1061 GGCGUGCUCUCUCUCUUCUGU 1081	Cleavage	1
Mpo-mr-13[2]	Mp*SPL1*	2	16.787	miRNA 21 CCUAGAGGUAGGGUGAAAAGU 1 :::::.::::::: :::::::: Target 262 GGAUUUCCAUCCAACUUUUCA 282	Cleavage	1

This homology-based search revealed only *SPL2* transcripts (Aa*SPL2* and Ap*SPL2*) to be targets for either miR529c or miR156 (Table 3.2). Moreover, we cannot rule out the possibility of presence of a species-specific *SPL*-targeting miRNA for Anthoceros *SPL1* transcripts like Mpo-mr-13 for Mp*SPL1*. The detailed micro-transcriptomic studies are needed to verify the presence of these miRNAs in the genome of Anthoceros.

3.1.2 Evolutionary relationships of *SPL* gene family in seed plants and bryophytes

To evaluate evolutionary relationships of SPL protein sequences between bryophytes and seed plants, we conducted phylogenetic analysis. For this phylogenetic analysis we used full length protein sequences of all species under study. Based on phylogenetic tree, SPL proteins are divided into four groups (Group 1- Group 4), with each group consisting at least of one SPL from each species used for the analysis (Fig. 3.2). Both liverwort (*M. polymorpha*) and hornwort (two Anthoceros species) genome encodes only four SPL proteins as compared to moss (*P. patens*) and angiosperm (*A. thaliana*) which encode 13 and 16 SPL proteins, respectively (Table 3.3). In general, as expected, members from bryophyte clades are more phylogenetically related to each other than to angiosperms in all four groups. Additionally, SPL from liverwort and hornwort representatives are clustered closer to each other, what may indicate their common evolutionary origin. Furthermore, the paralogous protein pairs AtSPL3/6 and AtSPL4/5 are grouped separately and hence, not have any related SPL proteins from bryophytes representatives. Therefore, they were excluded from all four groups.

Moreover, Group 4 in the phylogenetic tree consist of the least number of SPL proteins with only one member from each species. The distinguishable feature of members belonging to this group is the presence of C4 motif in their first Zn-finger domain (Zn-1) as compared to the canonical C3H motif which is characteristic for SPL proteins belonging to other three groups. The presence of the least number of SPLs within Group 4 indicates their highly conserved nature along with their resistance to expansion during the course of evolution (Fig. 3.2).
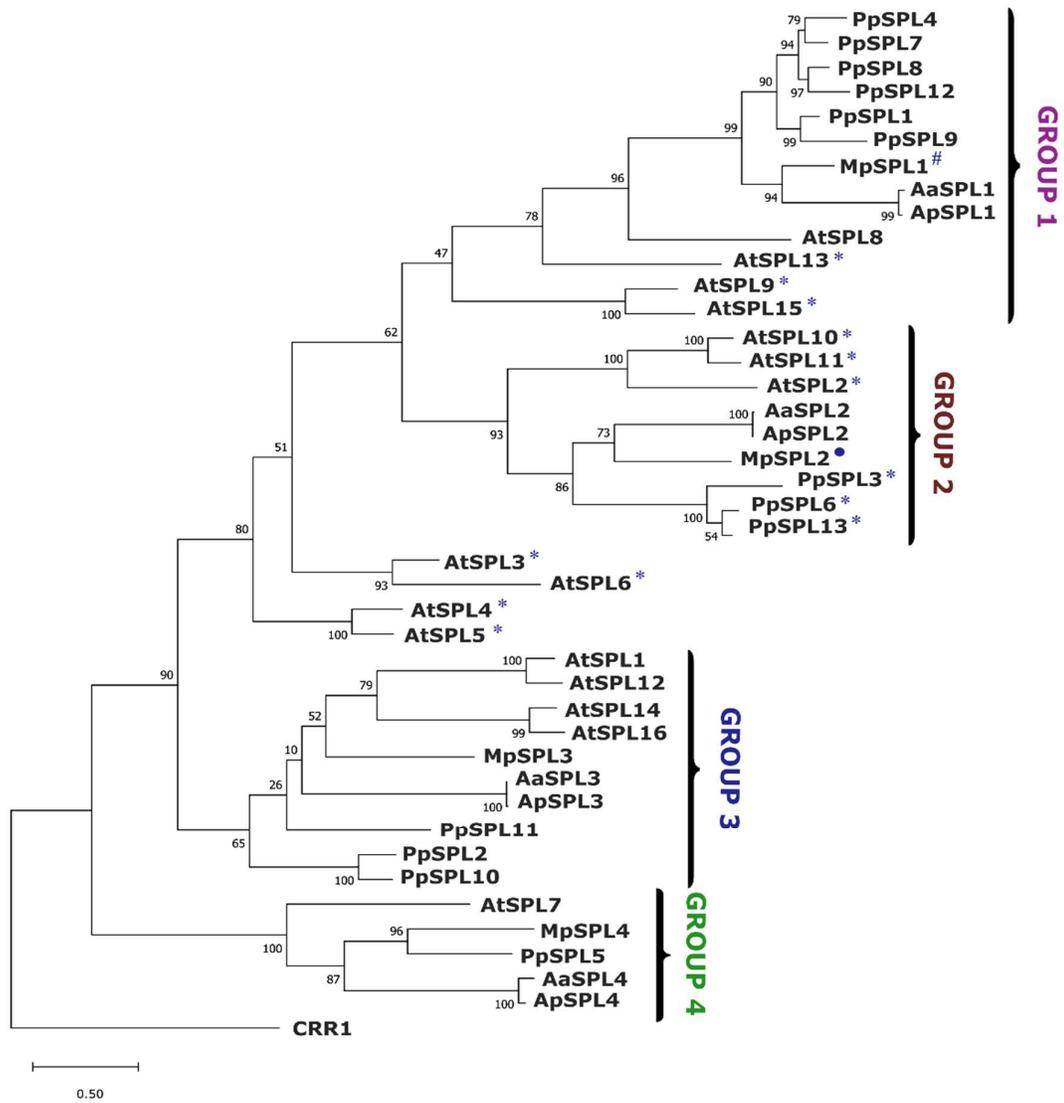
Figure 3.2: Phylogenetic tree of SPL protein sequences from bryophytes (*M. polymorpha*, *P. patens*, *A. agrestis* and *A. punctatus*) and angiosperm (*A. thaliana*) representative species. The tree was constructed in MEGA 11 software (Tamura *et al.*, 2021) using maximum-likelihood method, 1000 bootstrap replicates and *C. reinhardtii* CRR1 as an outgroup. SPL proteins marked by *,º and # are regulated at their post-transcription levels by miR156/ miR529c.

Table 3.3: Gene names, accession numbers and database links of *SPL*s from *A. thaliana*, *P. patens*, *M. polymorpha*, *A. agrestis* [bonn], *A. punctatus* and *C. reinhardtii*.

Plant species	Gene name	Accession number	Database
Arabidopsis thaliana	At*SPL1*	AT2G47070	https://www.arabidopsis.org/
	At*SPL2*	AT5G43270	
	At*SPL3*	AT2G33810	
	At*SPL4*	AT1G53160	
	At*SPL5*	AT3G15270	
	At*SPL6*	AT1G69170	
	At*SPL7*	AT5G18830	
	At*SPL8*	AT1G02065	
	At*SPL9*	AT2G42200	
	At*SPL10*	AT1G27370	
	At*SPL11*	AT1G27360	
	At*SPL12*	AT3G60030	
	At*SPL13*	AT5G50570	
	At*SPL14*	AT1G20980	
	At*SPL15*	AT3G57920	
	At*SPL16*	AT1G76580	
Physcomitrium patens	Pp*SPL1*	Pp3c12_24350	https://phytozome-next.jgi.doe.gov/
	Pp*SPL2*	Pp3c17_12760	
	Pp*SPL3*	Pp3c16_7480	
	Pp*SPL4*	Pp3c5_22750	
	Pp*SPL5*	Pp3c3_31330	
	Pp*SPL6*	Pp3c25_8630	
	Pp*SPL7*	Pp3c6_6900	
	Pp*SPL8*	Pp3c16_7540	
	Pp*SPL9*	Pp3c4_6000	

	PpSPL10	Pp3c14_18960	
	PpSPL11	Pp3c3_23940	
	PpSPL12	Pp3c25_8610	
	PpSPL13	Pp3c16_7490	
Marchantia polymorpha	MpSPL1	Mp1g10020	https://marchantia.info/
	MpSPL2	Mp1g10030	
	MpSPL3	Mp1g13640	
	MpSPL4	Mp8g11850	
Anthoceros agrestis [Bonn]	AaSPL1	AagrBONN_evm.model.Sc2ySwM_344.856.1	https://www.hornworts.uzh.ch/en.html
	AaSPL2	AagrBONN_evm.model.Sc2ySwM_344.857.1	
	AaSPL3	AagrBONN_evm.model.Sc2ySwM_344.2221.1	
	AaSPL4	AagrBONN_evm.model.Sc2ySwM_369.244.1	
Anthoceros punctatus	ApSPL1.1	Apun_evm.model.utg0001071.74.1	https://www.hornworts.uzh.ch/en.html
	ApSPL1.2	Apun_evm.model.utg0001071.74.2	
	ApSPL2.1	Apun_evm.model.utg0001071.75.1	
	ApSPL2.2	Apun_evm.model.utg0001071.75.2	
	ApSPL3	Apun_evm.model.utg0001851.396.1	
	ApSPL4	Apun_evm.model.utg0001161.202.1	
Chlamydomonas reinhardtii	CrSBP1	CHLRE_07g345050v5	https://phytozome-next.jgi.doe.gov/
	CrSBP2	CHLRE_01g012200v5	
		CHLRE_02g104700v5	
		CHLRE_05g233551v5	
		CHLRE_06g278229v5	
		CHLRE_06g300600v5	
		CHLRE_07g325738v5	
	CrCRR1	CHLRE_09g390023v5	
		CHLRE_09g399289v5	
	CrSBP3	CHLRE_17g698233v5	

3.1.3 Gene and protein structure analysis of *SPL* gene family from bryophytes and angiosperms

The gene structural diversity within *SPL* gene families of Arabidopsis and four bryophyte species was explored to check whether there is any conservation of the exon-intron structure between the analysed species. The analysis revealed large variations in the number and length of introns within each class of bryophyte and angiosperm (Fig 3.3). Members belonging to Group 1 showed the highest diversity within their gene structures as *SPL* genes from Arabidopsis and Marchantia consist of 2 introns as compared to Anthoceros *SPL* genes having 4-5 introns and *P. patens* with 6-7 introns within their *SPL* genes. Moreover, the average intron number within *A. thaliana*, *P. patens, M. polymorpha* and Anthoceros *SPL* genes was found to be 4.2, 6.8, 5.4 and 2.2, respectively. The highest number of introns (12) was found in Pp*SPL11* and the lowest number of introns (1) was found within Anthoceros *SPL3* (Aa*SPL3* and Ap*SPL3*) and *SPL4* (Ap*SPL4*). In general, members belonging to Groups 2 consist of *SPL* genes with the lowest number of introns (up to 3 introns) as compared to the other three groups (up to 12 introns). Additionally, the average intron lengths within *A. thaliana*, *P. patens, M. polymorpha* and Anthoceros *SPL* genes was calculated to be 51, 156, 275 and 104 bp, respectively. The average intron length calculations revealed that the shortest intron length was found to be in Arabidopsis and hornworts while the longest within Marchantia, respectively. This data correlates with studies calculating average intron length within the genomes of Arabidopsis, *P. patens*, Marchantia and hornworts to be 164, 278, 329 and 104 bp, respectively (Swarbreck *et al.*, 2008; Lang *et al.*, 2008; Bowman *et al.*, 2017*b*; Li *et al.*, 2020*b*).

The highest similarity of gene structure composition is observed within each group, with some exceptions. For example, in Groups 1 and 2, all genes share SBP domain (separated by an intron) while in Groups 3 and 4 both Anthoceros *SPL*s (*SPL3* and *SPL4*) and moss Pp*SPL2* and Pp*SPL10* contain SBP domain within one exon (not separated by an intron). Hence, it seems that in the first land plants, *SPL* genes had two types of structures in which SBP domain encoding fragment was either encoded by two exons or by one, suggesting that what we observe now in Anthoceros, and *P. patens* might be a relict of that state. Another possibility could be that Anthoceros, and *P. patens* genes might have lost intron within the SBP encoding fragment during the course of evolution and this phenomenon is widely observed in different genes.
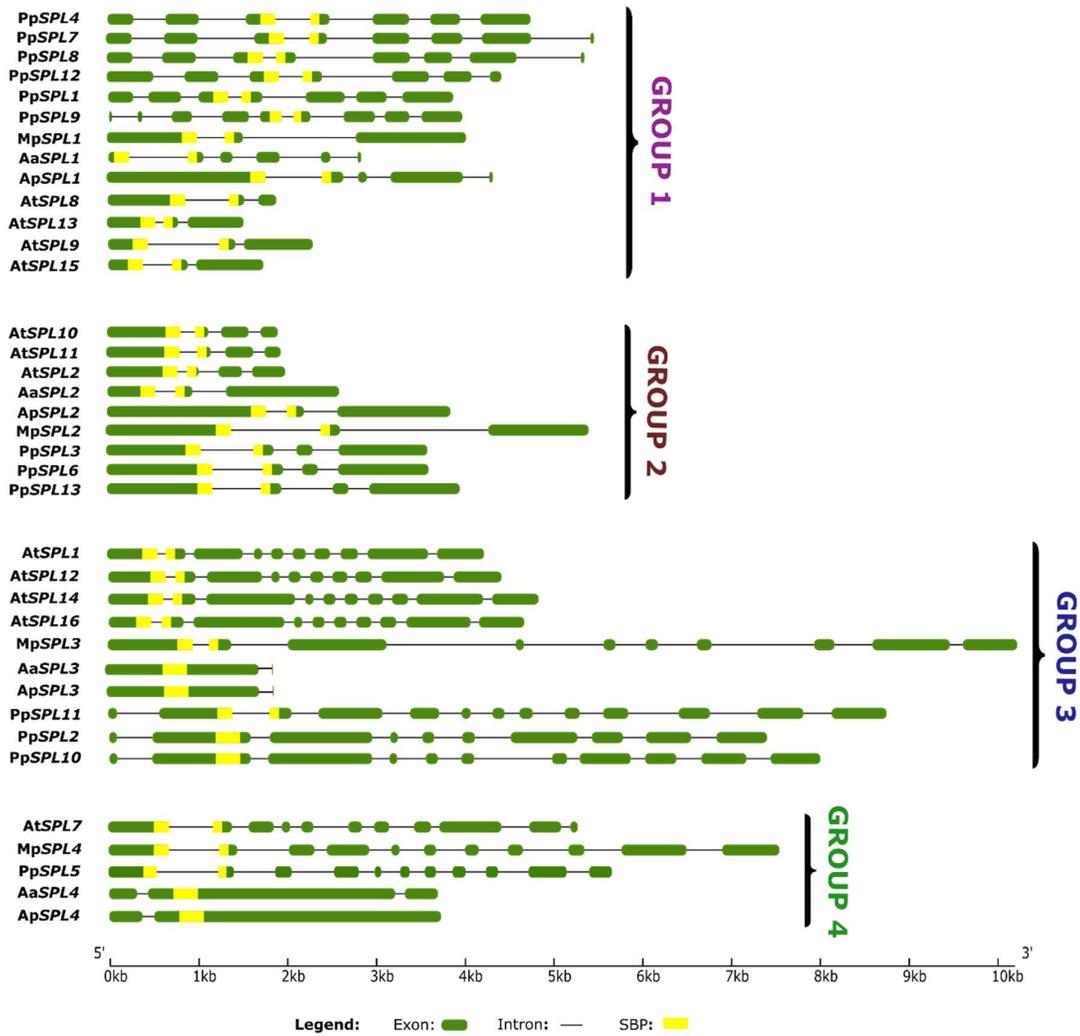
Figure 3.3: Exon-intron organization of *SPL* genes from *A. thaliana*, *P. patens*, *A. agrestis*, *A. punctatus* and *M. polymorpha* according to divisions in four groups. All the gene structures were created using GSDS2.0 online server (Hu *et al.*, 2015). The exons are depicted by green boxes, conserved SBP domain are depicted by yellow boxes and introns as dark black lines in each gene model. The scale below depicts the length of genes in base pairs (bp).

It has been reported in a previous study that members belonging to Groups 3 and 4 usually consist of genes with higher number of exons and encode longer proteins than the members of other two group. But with the addition of Anthoceros *SPL* in determining the phylogenetic relationships, there seems exceptions to this study. Usually, it is observed that Anthoceros *SPL*s do not follow this pattern since they comprise of 2-3 exon numbers only and therefore, encode shorter proteins. Hence, it seems likely that either the ancestor of Anthoceros *SPL* has lost introns during the course of evolution or this might be due to the poor annotation of currently available Anthoceros genome.

To have an improved understanding of SPL protein characteristics, their motif composition was examined using MEME software (Bailey *et al.*, 2006) (Table 3.4). Additionally, the coordinates of SBP domain for each SPL protein were attained from Pfam database (Bateman *et al.*, 2004). Only SBP domain (represented on each protein structure by motifs 2, 1 and 4) was found to be conserved between all SPL protein analysed (Fig. 3.4). Additionally, members belonging to Group 3 (except for both Anthoceros SPL3) contain ANK domain (motif 6) which is shown to be involved in protein-protein interactions (Michaely and Bennett, 1992). Mostly, similar motif composition was observed within each group. Members belonging to Group 2 exhibited simpler motif composition with the lowest number of motifs found. There is high conservation of motif composition found between members belonging to Groups 3 and 4 (apart from both Anthoceros SPL3). Additionally, members targeted by miRNA from Groups 1 and 2 share similar motif composition with the smallest number of motifs as compared to members not targeted by any miRNA. The similarity between shared motifs within SPL proteins from different plant species might reflect at similarity in their roles and biochemical properties between different plant species.

The highest variations were found within members belonging to Group 1 as *P. patens* SPLs share the highest number of motifs but the SPL proteins from the remaining plant species possess lower number of motifs. Moreover, on the basis of their protein lengths, members from Group 1 can be sub-grouped into two divisions. The first division consisting of all Group 1 *P. patens* SPLs, MpSPL1 and ApSPL1 as longer proteins and second division consisting of all Group 1 Arabidopsis SPLs and AaSPL1 as shorter proteins. Despite the similarity in the protein lengths between members from first division of Group 1, there are still differences in the arrangement of motifs within these members. For example, apart from the conserved SBP domain, ApSPL1 and MpSPL1 only share motif 9 (localized upstream to SBP domain) with PpSPLs from Group 1. Additionally, ApSPL1 also shares motif 16 (localized downstream to SBP domain) with PpSPLs from Group 1. Since MpSPL1 and ApSPL1 share similar motifs and they both originate from the same phylogenetic branch, which may indicate that their biological functions are similar.
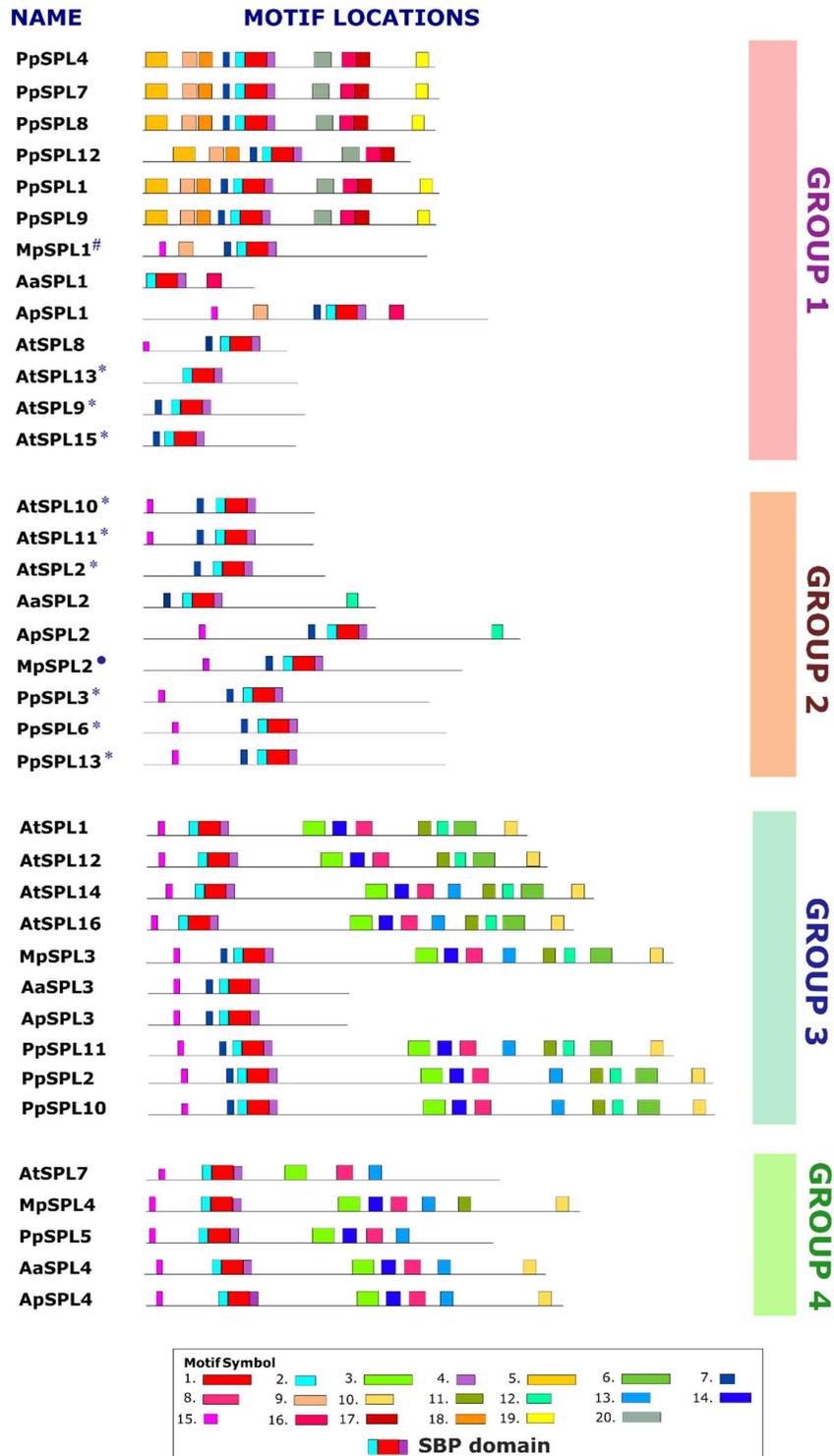
Figure 3.4: Motif analysis in SPL proteins from *A. thaliana*, *P. patens*, *A. agrestis*, *A. punctatus* and *M. polymorpha* according to divisions in four groups. The motifs were identified using MEME online tool (Bailey *et al.*, 2006). Each motif is represented in different colours with SBP domain shared between motifs 1, 2 and 4 represented by red, blue and violet boxes respectively. SPL proteins marked by *,° and # are regulated at their post-transcription levels by miR156 /miR529c and Mpo-miR13, respectively.

Table 3.4: Consensus sequences of motifs identified and shown in Fig. 3.4

Motif	Length	Consensus sequence
Motif 1	50	YHRRHKVCELHSKAPKVIVAGQEQRFCQQCSRFHLLSEFDEGKR SCRRRL
Motif 2	50	RISFKLFDKBPGEFPRRLRQQILEWLAHMPSDMEGYIRPGCTILTJ FLSM
Motif 3	15	MCQVEGCKADLSNAK
Motif 4	20	AGHNRRRRKPQPDASAAGTT
Motif 5	15	JGLKLGVRTYFETED
Motif 6	21	DWBRSEWDWDSVIFVAQPASG
Motif 7	33	RAYNSNMLSAGFPSNFTQNPMGIFSSAGIRSFG
Motif 8	41	LQSVRPLAVEAGQSTNLTVKGKNLRRPGTRLLCSFGGKYLA
Motif 9	50	PGGLTPLHVAASMEGAEDIIDALTNDPQQIGLHAWKHKRDSTGE TPFDYA
Motif 10	29	MNPSANNNEQQGDPSWSTENWDNPGAAGL
Motif 11	15	PGKRHRSSSPGSQVP
Motif 12	41	KNSANAGLVPEHKANGSLDSSEQRQEZQQQQQQQQQQQQQQQ
Motif 13	29	SSGIPGLAATSGNSQENDIRAFDPSSQEL
Motif 14	29	RGRLYRPFMVSMVAVAAVCVCVCLLFRGP
Motif 15	50	YNGVQPGVPWLRPIGARATETMSGQSIPRPAMSLPGSGGSVNVN VTVNDG
Motif 16	29	FWCNGRFJVQVGRQLALVVNGKVVDSKSN
Motif 17	21	KSLLVFAVERGWCAVVEKLLD
Motif 18	21	PLIVADAEVCSEJRTLESELE
Motif 19	50	QASMLGVDQQRRLFLGLGGVADEGNKGSGGEEGKEAGAHSNG VTVVGESP
Motif 20	50	DFDLQQKGASASAAAAFNLPPAPPPAPQASAAQKASFSTGFSSG APSPMT

Many group-unique motifs are found within the same phylogenetic group to which no functional characterisation can be annotated. For example, in Group 1, all *P. patens* SPL proteins share motifs 16-20, in Group 3 (with an exception of both Anthoceros SPL3), all members share seven motifs (Motifs 3, 5, 6, 8, 11, 12 and 14) and in Groups 3 and 4 (with the exception of AtSPL7), all members share motifs 3, 8 and 14. The presence of similar group-unique motifs within members of the same phylogenetic group and between members of two groups might hint towards their similar and shared roles in diverse plant species. Additionally, the presence of similar and group-unique motif composition between PpSPLs from Group 1 might indicate their involvement in species-specific (moss-specific) functions. Functional studies are needed to test this hypothesis.

As shown in Fig. 3.4, only the SBP domain is the conserved protein domain shared by all SPL proteins from the studied plant species. The SBP-domain consists of two Zn-finger

binding motifs (Zn-1 and Zn-2) and one nuclear localization (NLS) motif. The Zn-1 and Zn-2 motifs are characterized by the presence of conserved amino acid structures, C2HC or C4 and C3H, respectively. The members belonging to Group 4 are usually characterized by the presence of C4 motif in the Zn-1 of SBP-domain while members belonging to other three groups (Group 1, 2 and 3) are characterized by the presence of C2HC motif in the Zn-1 of SBP-domain. To examine the conservation of SBP-domain structure between Arabidopsis and bryophytes, multiple sequence alignment was performed using CLUSTAL W in Jalview program (Clamp *et al.*, 2004) (Fig. 3.5). The alignment showed the highly conserved amino acid compositions in Zn-1, Zn-2 and NLS motifs. In general, Zn-2 motif showed higher conservation than Zn-1 motif.

To gain the understanding of structural differences between SBP domains among species analysed, we used Weblogo to depict the alignment results for each species separately (Crooks *et al.*, 2004) (Fig. 3.6). The sequence logos revealed a high conservation at amino acid level in two Zn-finger and NLS motifs in land plants and green algae SPL proteins. When comparing SBP-domain structure of land plants with green algae, *C. reinhardtii*, several differences can be noticed. For example, in Zn-1 motif, only C3H composition is observed in algae in contrast to both C3H and C4 compositions observed in land plants. Also, the conserved basic amino acid residues from positions 17-22 which are highly conserved in all land plants under study are missing from green algae, except arginine at position 19. Additionally, the signature of NLS motif, KRRRK, showed higher conservation in land plants than in algae. This comparative SBP domain analysis showed higher similarities of bryophytes SBP domain with that of Arabidopsis than green algae. Therefore, this may suggest that SPL proteins pre-exist before the origin of land plants and the conserved SBP domain from land plants and green algae arose from a common ancestor.
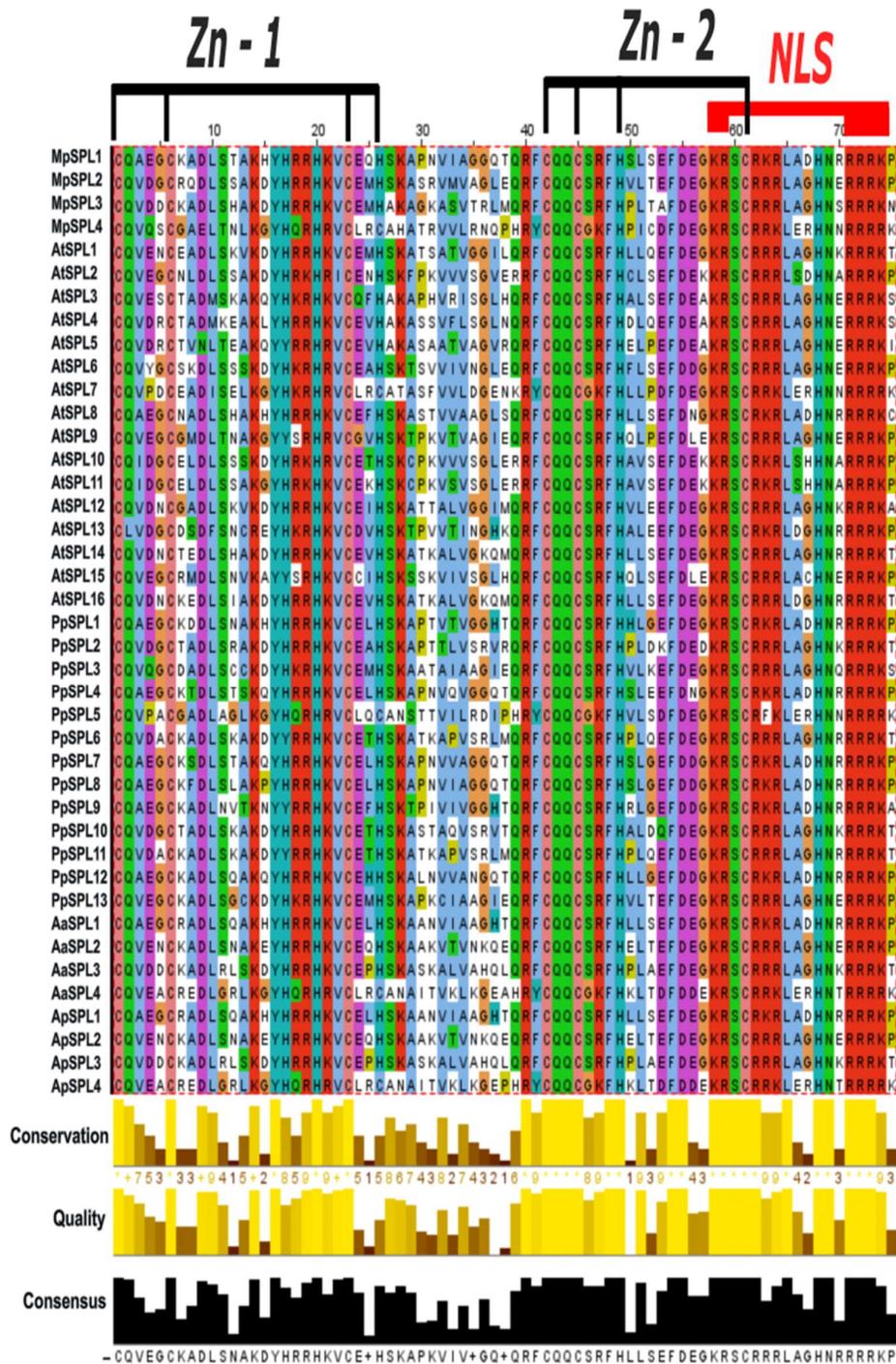
Figure 3.5: Multiple sequence alignment between SBP-box domain sequences of *A. thaliana*, *P. patens*, *A. agrestis*, *A. punctatus* and *M. polymorpha*. The alignment was performed in Jalview software using ClustalW (Clamp *et al.*, 2004). The level of conservation and consensus SBP-domain sequence is shown below the alignment.
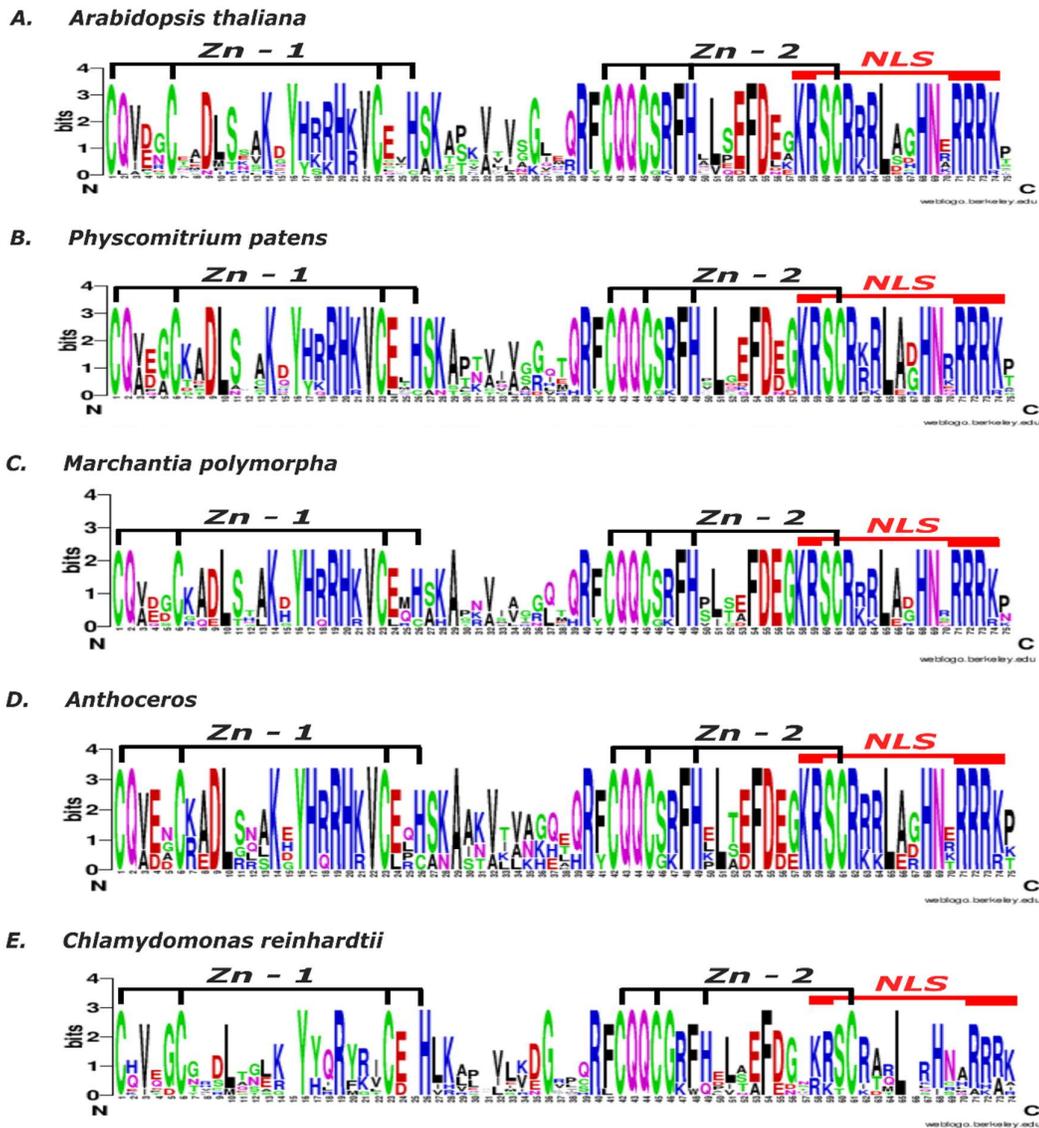
Figure 3.6: Sequence logos of SBP domain in A. *A. thaliana*, B. *P. patens*, C. *M. polymorpha*, D. *Anthoceros* and E. *C. reinhardtii*. The sequence logos were designed using Weblogo online software (Crooks *et al.*, 2004). The overall height of the stack depicts the extent of sequence conservation at that position. The height of letters within each stack depicts the relative frequency of each amino acid at that position.

3.1.4 *Cis*-element analysis in the promoter regions of *SPL* genes

The knowledge about motif composition in SPL protein sequences and *cis*-acting elements in the promoter regions of *SPL* genes can collectively help to gain insights into potential functions of *SPL* genes in various plant species. *Cis*-elements present in the promoter regions of genes have been known to regulate gene transcription and serve in adaptive mechanisms to respond to changing environmental conditions (Walther *et al.*, 2007*a*; 2007*b*). The PlantCARE database search was used to identify *cis*-elements in promoter regions of each

SPL gene (Lescot *et al.*, 2002). Numerous *cis*-elements were detected, and they were categorised into four divisions (full list in Table 3.5):

a) Growth and Development: includes elements like A-box, CAAT-box, TATA-box, AT-rich elements, circadian clock elements

b) Phytohormone: includes elements like HD-ZIP, TGA-box, TGA-elements, P-box

c) Light: includes elements like ACE motif, Box II-like elements, GA-motif, MRE elements

d) Stress: includes elements like GC-motif, TC-rich repeats, ARE, LTR and MBS motifs.

Among the selected categories, *cis*-elements involved in growth and development were most numerous in all species, followed by phytohormone or light responsive category (Fig. 3.7A). In general, number of growth and development elements expanded with increasing plant diversity. The number of *cis*-elements involved in phytohormone was the highest in moss and the lowest in hornworts. In contrast, the number of *cis*-elements involved in light response was the highest in angiosperm. The stress-responsive elements were the highest in angiosperm and moss. Additionally, *cis*-elements within each *SPL* gene from all four phylogenetic groups were calculated and their numbers were visualized on the chart (Fig. 3.7B). In general, all four categories of *cis*-elements were detected in promoter regions of almost all *SPL* genes with the exception of Ap*SPL2*, At*SPL10* and At*SPL16*. Promoter sequences of Ap*SPL2* and At*SPL10* from Group 2 do not contain phytohormone response elements while in the At*SPL16* gene promoter from Group 3, no stress responsive elements were detected.
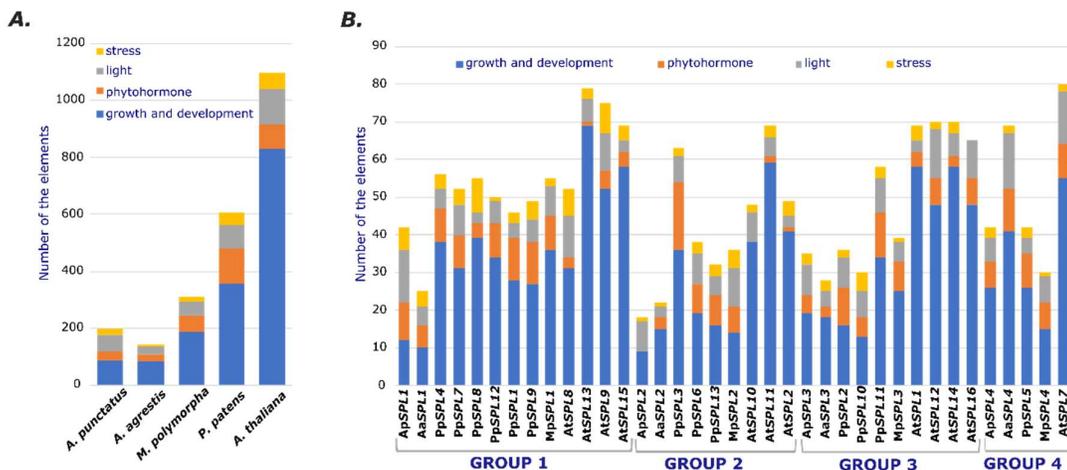
Figure 3.7: Cis-element analysis in promoter regions A) of 4 *SPL*s each in *A. punctatus*, *A. agrestis* and *M. polymorpha*, 13 *SPL*s in *P. patens* and 16 *SPL*s in *A. thaliana* B) within each *SPL* gene from Groups 1-4.

Table 3.5: Cis-acting analysis of *SPL* gene promoter regions from *A. thaliana*, *P. patens*, *A. agrestis*, *A. punctatus* and *M. polymorpha*. The regulatory *cis*-elements in promoter of each *SPL* gene were identified using PlantCARE database (Lescot *et al.*, 2002). The elements were divided into four categories of growth and development, phytohormone, light and stress and shown in table below.

Function	Site Name	*A.punctatus*	*A.agrestis*	*P.patens*	*M.polymorpha*	*A.thaliana*
Growth and Development	A-box	4	2	6	4	1
	CAAT-box	73	58	246	139	396
	CAT-box	2	1	13	6	4
	CCAAT-box	2	0	3	6	2
	GCN4_motif	0	0	2	3	3
	NON-box	0	0	0	0	1
	O2-site	1	2	3	5	3
	RY-element	0	0	1	0	2
	TATA-box	5	21	80	24	409
	AT-rich element	0	0	1	0	4
	circadian	0	0	2	0	3
	total	87	84	357	187	828
Phytohormone	ABRE	9	8	29	14	31
	AuxRR-core	2	0	1	1	5
	CGTCA-motif	10	6	33	16	13
	GARE-motif	3	1	0	0	4
	TGA-element	0	0	9	3	2
	P-box	0	0	8	1	8
	HD-Zip 3	0	0	1	0	0
	TATC-box	0	0	2	0	3
	TCA-element	3	2	7	3	8
	TGA-box	0	0	0	3	0
	TGACG-motif	5	6	33	16	13
	total	32	23	123	57	87
Light	ACE	2	0	1	0	2
	AE-box	0	0	4	3	9
	ATCT-motif	0	0	1	3	2
	Box 4	1	4	8	0	16
	Box II	0	0	1	0	0
	Box II -like sequence	0	0	1	0	0
	CAG-motif	0	0	0	1	0
	chs-CMA1a	1	0	1	0	2
	GA-motif	0	0	0	0	1
	GATA-motif	5	1	6	0	3
	GATT-motif	0	0	1	0	0
	3-AF1 binding site	2	1	2	3	1
	GT1-motif	0	1	3	2	18

	ATC-motif	0	0	2	1	0
	MRE	0	0	0	1	7
	G-box	10	6	31	15	34
	GTGGC-motif	1	1	0	0	0
	I-box	1	2	2	0	5
	LAMP-element	0	0	0	0	2
	sbp-CMA1c	0	0	1	0	1
	Sp1	23	6	4	12	1
	TCCC-motif	8	1	3	8	1
	TCT-motif	4	4	8	2	19
	total	58	27	80	51	124
Stress	ARE	3	2	14	3	29
	TC-rich repeats	0	0	1	1	9
	GC-motif	10	5	5	6	2
	LTR	3	0	9	1	10
	MBS	4	3	18	3	7
	Total	20	10	47	14	57

3.1.5 Expression analysis of *SPL* genes in different developmental tissues

In order to gain a general understanding about the expression profiles of investigated *SPL* gene family across different developmental tissues in bryophytes and Arabidopsis, the publicly available RNA-seq data were collected and analysed. Amongst hornworts, only expression data for *A. agrestis* was available, hence, only *A. agrestis* data are shown below. The detected expression levels were plotted as heat-map for each plant species separately (Fig 3.8).

From the selected RNA-seq data for *A. thaliana*, the expression of 14 out of 16 *SPL* genes was identified. The RNA-seq values for two members of Group 1, At*SPL13* and At*SPL15* were not found. The expression studies involving transgenic lines with β-glucuronidase tagged to AtSPL13 and AtSPL15 showed that both these SPLs are expressed at extremely low levels for a brief duration during leaf development and early inflorescence developmental stages (Xu *et al.*, 2016). Such distinct expression profiles are most likely the reason that both genes were missing in the analysis.

In general, the expression analysis of *SPL* genes in Arabidopsis revealed that members belonging to the same phylogenetic group have rather similar expression pattern, implying that *SPL* paralogs are involved in the regulation of similar processes. Group 1 had the most

distinct expression pattern, and it was largely associated with flower development (Fig. 3.8A, Table 3.6). Although Arabidopsis Group 2 *SPL* genes were expressed in more developmental stages than Group 1 *SPL* genes, they also showed enhanced expression during floral organ development. In contrast, the expression levels of *SPL* genes from Groups 3 and 4 were higher and relatively similar in all organs and developmental stages under analysis. Most commonly, Group 1 and 2 *SPL* genes which are targeted by miRNA (except At*SPL8*) used in this study showed specific expression in flower development as compared with ubiquitous expression patterns of Group 3 and 4 *SPL* gene members.

In general, Arabidopsis *SPL* genes can be categorised into two groups based on their expression patterns:

1. *SPL* genes with relatively constitutive and stable expression levels across all developmental stages of Arabidopsis, and
2. *SPL* genes with elevated expression levels during specific growth and reproductive developmental stages of Arabidopsis.

Table 3.6: Expression data values for *SPL* genes from *Arabidopsis thaliana.* The values were downloaded from Moreno et al. 2022.

	Cotyledon	Vascular leaf	Flower	Stamen	Carpel	Pollen	Silique	Seed
At*SPL8*			21		50		2	
At*SPL4*	10	4	24	1	20		15	
At*SPL5*	1	1	5	1	4			3
At*SPL9*		1	57	2	52		3	
At*SPL11*	4	6	77	14	132		32	
At*SPL2*	6	26	41	15	74		21	
At*SPL10*	5	6	57	6	76		8	0.5
At*SPL16*	40	43	25	16	31		7	11
At*SPL6*	5	9	19	8	21	1	10	0.9
At*SPL1*	66	64	119	33	154		50	12
At*SPL14*	118	123	143	43	128		28	19
At*SPL12*	41	80	55	20	39	1	22	19
At*SPL7*	14	21	39	21	40	21	25	14

The expression analysis of *SPL* genes in *P. patens* revealed a similar division as Arabidopsis. It can be seen that Pp*SPL* genes from Group 1 are either not or weakly expressed in the spores and protonema but are prominently expressed in sporophytes and gametophores (Fig. 3.8B, Table 3.7). Pp*SPL7* from Group 1 is highly expressed in archegonia and at various developmental stages of sporophyte, implying its role in sexual reproduction of moss and its sporophyte maturation. The Group 2 members, most of which are targeted by miRNA (Pp*SPL3*, Pp*SPL6* and Pp*SPL13*) showed variable expression in different tissues. For example, Pp*SPL* genes from Group 2 were found to be most active throughout the sporophytic premeiotic-to-meiotic stage (sporophyte PM-M), with the exception of Pp*SPL3* as it also showed higher expression levels in moss chloronema and rhizoids. In contrast, the other two groups (Groups 3 and 4) Pp*SPL* genes were found to be constitutively expressed in all of the moss developmental stages and tissues under analysis.
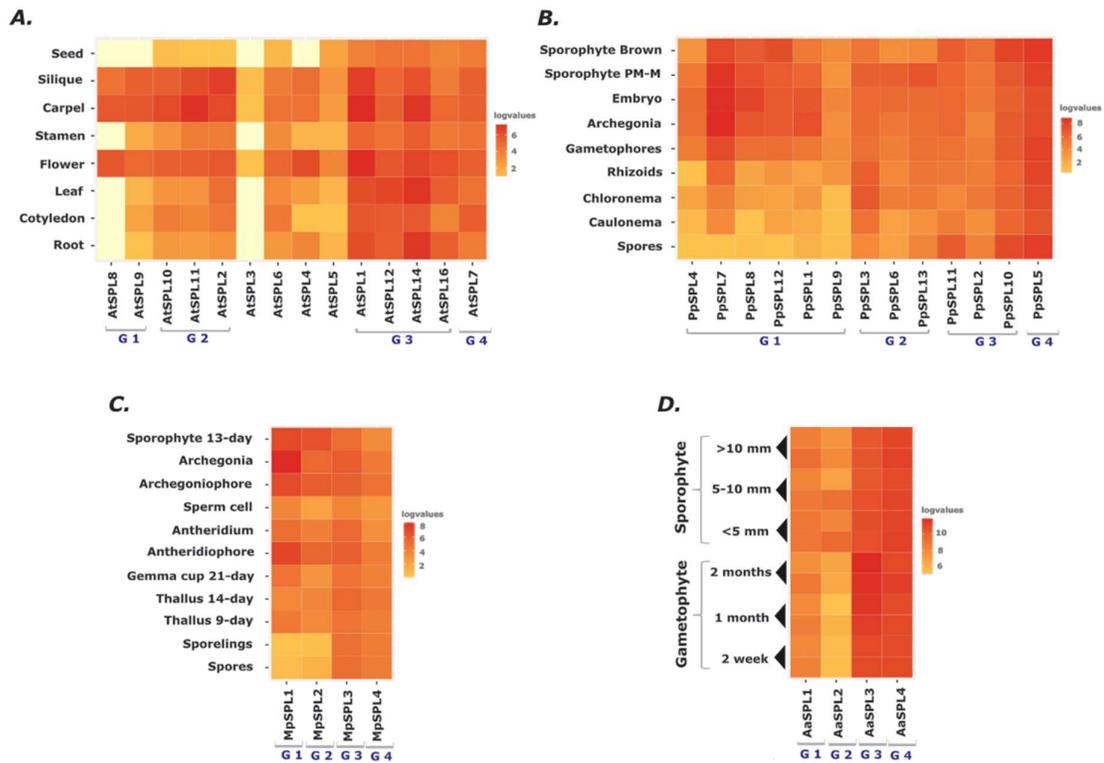
Figure 3.8: The expression profiles of *SPL* genes from different developmental stages and organs from *A. thaliana*, *P. patens*, *A. agrestis*, *A. punctatus* and *M. polymorpha* according to divisions in four groups. The expression data values are used as in tables 3.6-3.9. The heatmap is generated in RStudio (Verzani, 2011).

Table 3.7: Expression data values for *SPL* genes from *Physcomitrium patens.* The values were downloaded from (Fernandez-Pozo *et al.*, 2020)

	Spores	Caulonema	Chloronema	Rhizoids	Gametophores	Archegonia	Sporophyte-PM-M	Sporophyte-Brown
Pp*SPL1*	1.8	2.86	9.48	5.14	34.54	153.27	78.19	33.9
Pp*SPL2*	23.99	13.19	36.88	30.79	31.39	19.35	38.08	51.73
Pp*SPL3*	17.5	44.85	119.57	97.29	48.65	63.32	90.37	42.04
Pp*SPL4*	0.38	1.08	4.54	0.46	27.79	58.22	32.31	11.04
Pp*SPL5*	313.84	152.18	191.66	235.07	272.44	200.95	263.78	323.69
Pp*SPL6*	5.04	5.29	27.91	18.82	25.54	36.76	96.49	13.83
Pp*SPL7*	0.42	16.46	26.49	72.98	186.97	488.61	377.98	212.5
Pp*SPL8*	0.6	0.48	6.24	4.05	56.39	136.89	163.35	110.26
Pp*SPL9*	0.78	0.67	0.81	10.73	14.83	12.83	10.71	14.32
Pp*SPL10*	193.55	75.54	85.23	73.8	114.71	117.93	122.39	259.86
Pp*SPL11*	115.51	31.26	17.93	35.01	45.98	66.52	74.31	103.48
Pp*SPL12*	0.47	4.95	4.65	10.38	58.56	110.78	106.25	173.49
Pp*SPL13*	19.63	12.17	33.16	40.88	51.19	56	146.59	18.47

Like Arabidopsis and *P. patens*, Marchantia and *A. agrestis SPL* genes from Groups 3 and 4 showed constitutive expression profiles in all developmental stages and organs under analysis. *SPL* members from Groups 1 and 2 in Marchantia displayed tissue specific expression pattern, with the highest levels detected during development of reproductive organs as well as young sporophytes (Fig. 3.8C, Table 3.8). This suggests that Mp*SPL* genes are engaged during entire growth and developmental stages of liverwort life cycle. Additionally, Mp*SPL1* and Mp*SPL2* genes which are targeted by miRNAs are upregulated in sexual organs, implying their additional roles during sexual reproduction in Marchantia.

Table 3.8: Expression data values for *SPL* genes from *Marchantia polymorpha.* The values were downloaded from (Kawamura *et al.*, 2022)

	Mp*SPL1*	Mp*SPL2*	Mp*SPL3*	Mp*SPL4*
Spores 0 h (Spores)	0.244328	1.1361	36.82445	19.82057
Spores 96 h (Sporelings)	0	0.189475	30.7971	17.81448
Tak-1 Thallus 9 day (Thallus 9 day)	21.11429	11.20411	25.17865	16.65979
Tak-1 Thallus 14 day (Thallus 14 day)	11.91286	12.47148	44.34023	22.24595
Tak-1 Gemma cup 21 day (Gemma cup 21 day)	25.01036	6.822147	29.25353	14.94946
Antheridiophore	135.8983	49.97364	59.99973	18.27822
Antheridium	33.94589	16.28736	40.47235	8.322781
Sperm cell	12.25275	3.553252	13.57561	4.980176
Archegoniophore	116.5309	61.88261	55.20973	27.39149
Archegonia	290.2252	44.04714	66.87528	21.62104
Young sporophyte 13 day (sporophyte 13 day)	124.1062	97.53377	36.25326	9.618658

Meanwhile in *A. agrestis*, Aa*SPL2* showed specific expression pattern as compared to other three *SPL*s as it is predominantly expressed in sporophytic generation of this hornwort species. In contrast, Aa*SPL1* from Group 1 showed equally higher expression during its both gametophytic and sporophytic stages (Fig. 3.8D, Table 3.9).

Table 3.9: Expression data values for *SPL* genes from *Anthoceros punctatus.* The values were downloaded from (Li *et al.*, 2020*b*)

	Aa*SPL1*	Aa*SPL2*	Aa*SPL3*	Aa*SPL4*
2weeks Gam 1 (2 week)	350.7077	42.65364	1954.052	1956.143
2weeks Gam 2 (2 week)	282.0749	39.96061	2038.77	1805.867
4weeks Gam 1 (1 month)	390.4545	61.41981	2734.223	1941.305
4weeks Gam 2 (1 month)	264.0865	32.46965	3255.087	1822.63
2month Gam 1 (2 months)	459.0806	86.79136	3451.609	2542.073
2month Gam 2 (2 months)	229.0108	91.15963	3919.629	2054.427
upto5mm Sporo 1 (<5 mm)	509.4232	726.2486	1643.666	2435.206

upto5mm Sporo 2 (<5 mm)	464.275	291.5447	1744.84	2345.837
5mm to 1cm Sporo 1 (5-10 mm)	464.5007	614.9586	1659.482	2340.68
5mm to 1cm Sporo 2 (5-10 mm)	310.6818	114.4617	1320.546	1889.845
over1cm Sporo 1 (>10 mm)	699.5649	269.5671	1462.744	2381.358
over1cm Sporo 2 (>10 mm)	354.8668	192.655	1350.954	2096.032

In conclusion, the analysis of expression data in all analysed plant species revealed that they could be divided into two categories. First, *SPL* gene from Groups 3 and 4, which are highly expressed in practically all tissues and hence, may operate in a similar manner as housekeeping genes maintaining basal cellular functions. Furthermore, the genes in this category are not regulated by any miRNAs. Second, *SPL* genes from Groups 1 and 2, which show specific expression pattern, implying their roles during specific developmental stages of growth and reproduction. Additionally, many genes in this category are post-transcriptionally regulated by miRNA. *SPL* genes from Groups 1 and 2, whose expression is strongly associated with sexual reproduction were detected in three of the four plant species studied, which includes the dicot, Arabidopsis, the moss *P. patens* and the liverwort, Marchantia. Given that there is no available expression data for hornwort *A. agrestis* of their reproductive organs, accordingly the conserved evolutionary mode of action within SPL family representatives, it could be hypothesized that at least one of the *SPL* family members might be involved in regulation of the reproductive pathway in Anthoceros. Moreover, despite the hints from expression profiles about the putative functions of SPL in different developmental tissues, functional studies need to be conducted to have a clear understanding of involvement of SPL transcription factors in various stages of plant growth and development.

3.2 Chapter 2 - Functional characterization of Mp*SPL3* and Mp*SPL4* transcription factors from *Marchantia polymorpha*

3.2.1 Characterization of 5' and 3' cDNA ends of overlapping neighbouring genes of Mp*SPL3* and Mp*SPL4*

Based on the genomic and transcriptomic data available for *M. polymorpha* (MarpolBase, annotation version 3.1, 2019), the analysis of Mp*SPL3* and Mp*SPL4* (Accession nos. *Mapoly0019s0134.1* and *Mapoly0008s0031.1,* respectively) genomic organisation revealed that both these genes consist of 11 exons that when transcribed gave rise to single mRNAs. Mp*SPL3* and Mp*SPL4* genes encode proteins of 1158 aa and 1004 aa in length, respectively. During this analysis, it was observed that short neighbouring genes, *Mapoly0019s0135.1* and *Mapoly0008s0030.1* (Fig. 3.9 and 3.10) are present in the promoter region of Mp*SPL3* and Mp*SPL4* genes, respectively. Each of the neighbouring gene was annotated at the same strand as Mp*SPL3* and Mp*SPL4* genes and covered a portion of 5'UTRs of their loci.

Figure 3.9: Genomic organisation of Mp*SPL3* (*Mapoly0019s0134.1*) locus along with its neighbouring gene *Mapoly0019s0135.1*, based on data from MarpolBase version JGIv3.1. Introns are represented by black lines, exons by boxes, white boxes indicate 5' and 3' UTRs while coloured boxes indicate CDS.
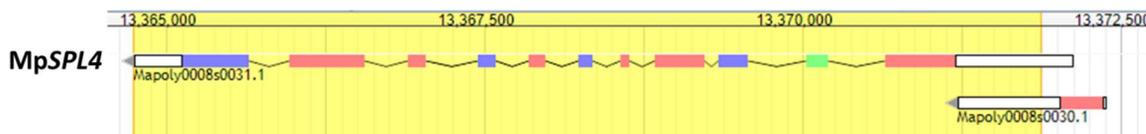
Figure 3.10: Genomic organisation of Mp*SPL4* (*Mapoly0008s0031.1*) locus along with its neighbouring gene *Mapoly0008s0030.1*, based on data from MarpolBase version JGIv3.1. Introns are represented by black lines, exons by boxes, white boxes indicate 5' and 3' UTRs while coloured boxes indicate CDS.

In order to obtain a clear picture of the 5' and 3' ends of these neighbouring genes and to confirm the lengths of 5' UTRs of Mp*SPL3* and Mp*SPL4* genes, we performed 5' RACE for Mp*SPL3* and Mp*SPL4* genes and 5' and 3' RACE for their neighbouring genes, respectively. The RACE-cDNA templates were prepared from 3-week-old male and female tissues.

3.2.1.1 RACE analysis for Mp*SPL3* and its neighbouring gene, *Mapoly0019s0135.1*

According to MarpolBase version JGIv3.1 annotation, *Mapoly0019s0135.1* shares 267 bp with the 5'UTR of Mp*SPL3* gene. The primers for RACE analysis of *Mapoly0019s0135.1* were designed outside of the shared region, to be specific for this gene (Fig. 3.11).
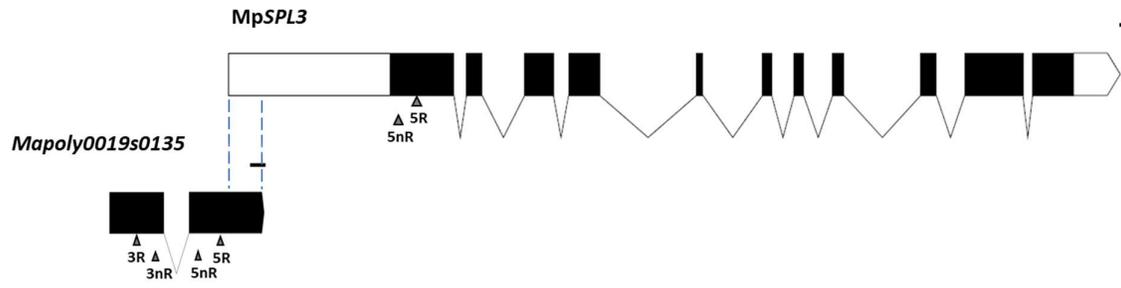
Figure 3.11: A schematic structure of primers designed for RACE experiments for Mp*SPL3* along with its neighbouring gene, *Mapoly0019s0135.1*. The bar above each gene corresponds to 100 bp. Primers are abbreviated as 5R: 5' RACE, 5nR: 5' nested RACE, 3R: 5' RACE and 3nR: 3' nested RACE, for each respective gene. The dashed blue lines depict shared region between two genes.

5' RACE reaction for Mp*SPL3* yielded a PCR product corresponding to ~2.5 kb length (Fig. 3.12A). Since the expected product size was 2240 bp, hence the product highlighted in red circle was eluted, cloned, and used for sequencing. Two out of six clones were specific and mapped precisely to the 5'UTR of Mp*SPL3* while the other four products were unspecific (Fig. 3.12B). The results indicated 5' UTR of Mp*SPL3* to be 232 and 248 bp shorter than annotated (1966 bp) in the database in v3.1.
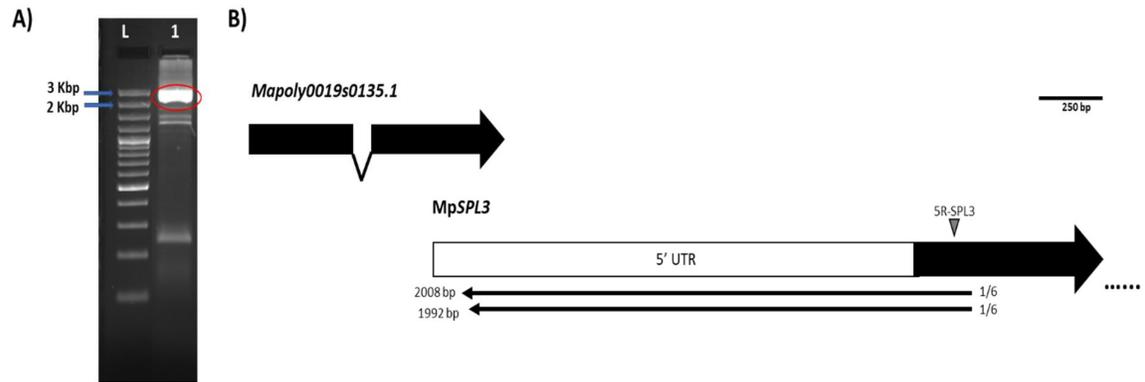
Figure 3.12: 5' RACE analysis for Mp*SPL3* gene transcript. A) Gel electrophoresis of 5' RACE product for Mp*SPL3*. L – 100bp plus ladder; 1 denotes the Mp*SPL3* 5'RACE products with primer 5R from Fig. 3.11. The product highlighted in red circle was gel eluted, cloned, and sequenced, B) Schematic representation of the length of obtained 5' RACE products for Mp*SPL3* gene transcript in the context of gene annotation. The scheme shows a part of Mp*SPL3* gene (5' UTR and first exon) along with its neighbouring gene *Mapoly0019s0135.1*. The arrows below Mp*SPL3* show the length and position of the ends of 5'RACE products obtained after sequencing. The scale above represents 250bp.

During several attempts of amplifying and cloning the 5' and 3' RACE products for *Mapoly0019s0135.1* gene, no specific cDNA products were obtained. Summarizing these results, there could be two possibilities:

1.   The neighbouring gene is rather wrongly annotated in the genome in the version 3.1.

2.   The tissues used for the RACE experiment (3 week male and female vegetative tissues) might either have very low expression of this gene or this gene is not expressed in this developmental stage.

However, in the new revision of *M. polymorpha* standard reference genome, MpTak_v6.1r2, released in 2023, the neighbouring gene has been removed from annotation (Fig. 3.13) which supported the hypothesis no. 1.
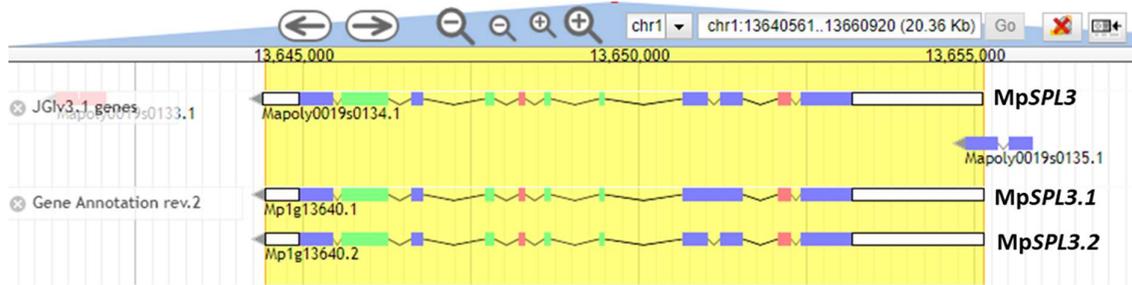
Figure 3.13: Genomic organisation of Mp*SPL3* (*Mapoly0019s0134.1/ Mp1g13640*) locus along with its neighbouring gene *Mapoly0019s0135.1*, overlapping with 5'UTR of Mp*SPL3*. The schemes are based on data from MarpolBase version JGIv3.1 and version Gene annotation rev.2 (v6.1r2). Introns are represented by black lines, exons by boxes: white boxes indicate 5' and 3' UTRs while coloured boxes indicate CDS.

Moreover, according to the newest revision of *M. polymorpha* standard reference genome, MpTak_v6.1r2, Mp*SPL3* gene encodes two isoforms of mRNA, Mp*SPL3.1* and Mp*SPL3.2* (Fig. 3.13). The only difference between the two gene isoforms is the presence of intron in 3rd exon. Hence, Mp*SPL3.2* is shorter than Mp*SPL3.1*.

3.2.1.2   RACE analysis For Mp*SPL4* and its neighbouring gene, *Mapoly0008s0030.1*

According to MarpolBase version JGIv3.1 annotation, *Mapoly0008s0030.1* shares 913 bp with the 5'UTR of Mp*SPL4* gene. The primers for RACE analysis of *Mapoly0008s0030.1* were designed outside of the shared region, to be specific for this gene (Fig. 3.14).
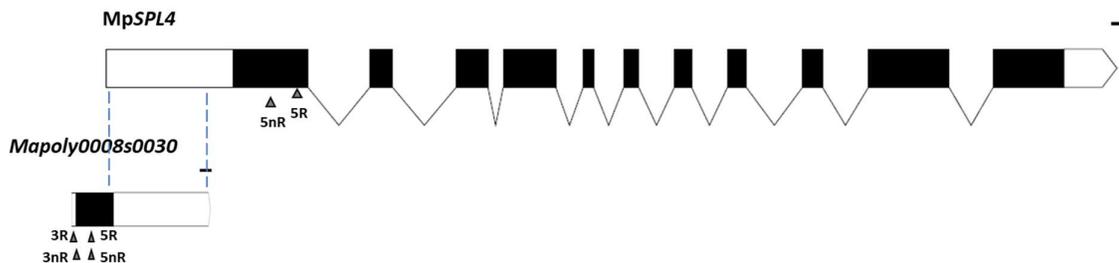
Figure 3.14: A schematic structure of primers designed for RACE experiments for Mp*SPL4* locus along with its neighbouring gene, *Mapoly0008s0030.1*. The bar above each gene corresponds to 100 bp. Primers are abbreviated as 5R: 5' RACE, 5nR: 5' nested RACE, 3R: 5' RACE and 3nR: 3' nested RACE, for each respective gene. The dashed blue lines depict shared region between two genes.

The experiment with 5'RACE for Mp*SPL4* yielded multiple PCR products and the expected size of PCR product from 5'RACE reaction was 1223 bp. The two products (highlighted by red circles in Fig. 3.15a) corresponding to 1kb and ~1.4kb were cloned and used for sequencing. 10 out of 12 products obtained from these products mapped to 5'UTR of Mp*SPL4* gene (Fig. 3.17). All the products were of heterogenous lengths. 9 out of these 10 clones showed 5'UTR to be shorter (467 bp – 683 bp) than annotated (931 bp) while one clone revealed 5'UTR to be longer (1048 bp) than annotated (931 bp) in the database in v3.1.

Simultaneously, nested PCR was performed for the 5'RACE products and the expected size of PCR product from 5' nested RACE reaction was 1089 bp. Out of multiple PCR products obtained, the four products with highest intensity (highlighted by red circles in Fig. 3.15b) were cloned and used for sequencing. 9 out of 10 clones obtained from these products also mapped to 5'UTR of Mp*SPL4* gene. All the products obtained were of heterogenous lengths. 8 out of 9 clones indicated 5'UTR to be shorter (297 bp – 656 bp) than annotated (931 bp) while one clone revealed 5'UTR to be longer (1006 bp) than annotated (931 bp) in the database in v3.1 (Fig. 3.17).
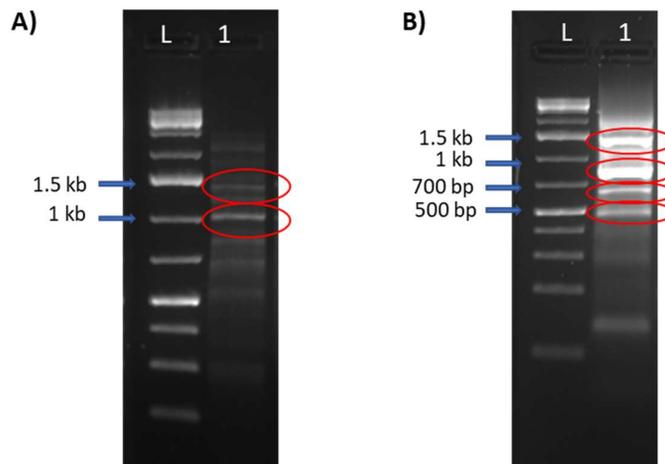
Figure 3.15: 5' RACE analysis for Mp*SPL4* gene transcript. A) Gel electrophoresis of 5'RACE products for *MpSPL4* transcript using 5R_SPL4 primer. B) Gel electrophoresis of nested 5'RACE products for *MpSPL4* transcript using 5nR_SPL4 primer. L denotes ladder and 1 denotes in A) 5'RACE and in B) 5' nested RACE products obtained with primers 5R and 5nR from Fig. 3.14*,* respectively. The product highlighted in red circles were eluted, cloned, and sent for sequencing.

According to the genome annotation JGIv3.1, the expected sizes of PCR products from 5' RACE, 5' nested RACE, 3' RACE and 3' nested RACE for *Mapoly0008s0030.1* should be 181bp, 123bp, 1168bp and 1089bp, respectively (Fig. 3.14). The first round of 5' and 3' RACE experiment for *Mapoly0008s0030.1* transcript yielded multiple PCR products at varying lengths (Fig. 3.16A). Therefore, a second round of RACE PCR reaction was performed with

nested primers to specify PCR products. After nested PCR of 5'RACE, three populations of products corresponding to ~600bp, ~800bp and ~5kb were obtained (lane 1 in Fig. 3.16B) which were cloned and sequenced. Five out of 10 clones from 'product b' were specific and mapped precisely to the 3' UTR of *Mapoly0008s0030.1* gene, all these clones contained a poly(A) sequence (Fig. 3.17). Therefore, 3' end of *Mapoly0008s0030.1* gene is shorter (~450 bp) than that annotated in the database (~820 bp).
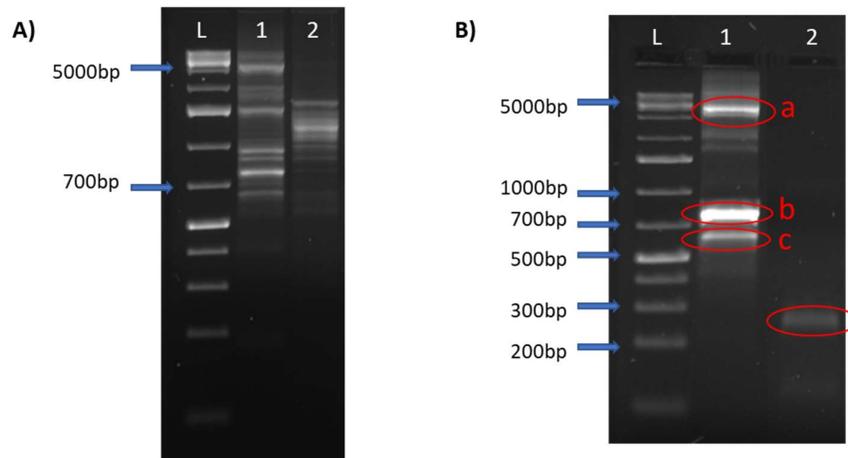
Figure 3.16: 5' and 3' RACE analysis for *Mapoly0008s0030.1* gene transcript. A) Gel electrophoresis of 5' and 3' RACE products for *Mapoly0008s0030.1* transcript using 5R and 3R primers in Lanes 1 and 2, respectively and B) Gel electrophoresis of nested 5' and 3' RACE products for *Mapoly0008s0030.1* transcript using 5nR and 3nR primers in Lanes 1 and 2, respectively. L denotes ladder and the product highlighted in red circles were eluted, cloned, and sent for sequencing.

In the case of 5'RACE for *Mapoly0008s0030.1* gene transcript, nested PCR of 5'RACE revealed a homogenous product corresponding to ~250 bp length (lane 2 in Fig. 3.16B). Two out of six clones obtained from this product were specific and mapped to the 5' UTR of this gene (Fig. 3.17). Moreover, the sequencing revealed 5'UTR to be longer (~128 bp) instead of 36 bp, as annotated in the database v3.1 (annotated 5'UTR is depicted as 'b' while the 'a and b' together denotes the new length of 5'UTR in Fig. 3.17).
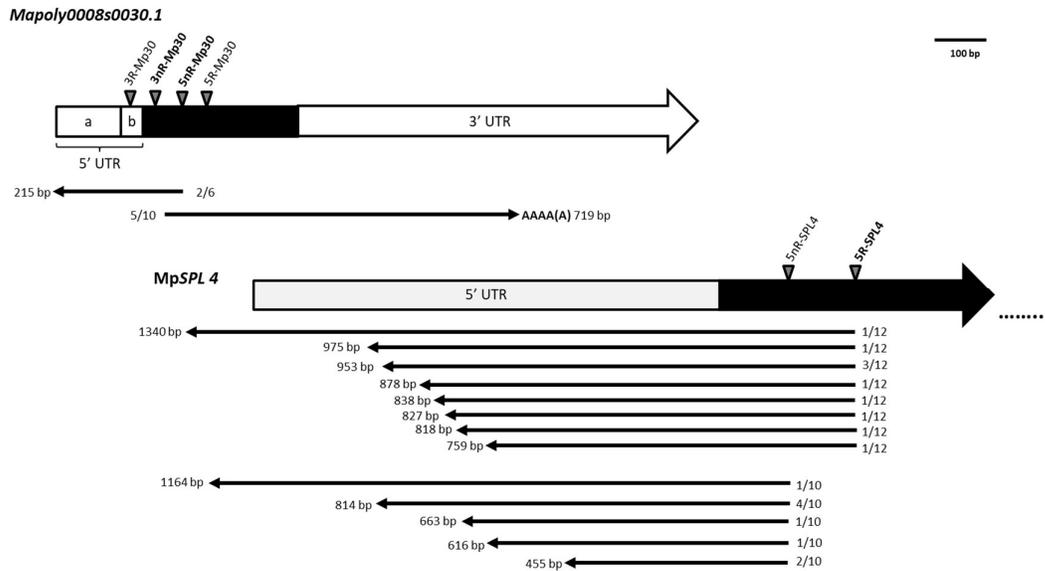
Figure 3.17: A schematic representation of RACE products for Mp*SPL4* and its neighbouring gene, *Mapoly0008s0030.1*. The scheme shows a part of Mp*SPL4* gene (5' UTR and first exon) along with its neighbouring gene *Mapoly0008s0030.1*. The arrows below each gene show the length and position of 5' and 3' RACE products obtained after sequencing. The annotated 5'UTR is depicted as 'a' while the 'a and b' denotes the new length of 5'UTR obtained after sequencing. The bar above represents 100 bp.

However, in the newest revision of *M. polymorpha* standard reference genome, MpTak_v6.1r2, released in 2023, the Mp*SPL4* neighbouring gene, *Mapoly0008s0030.1* has been removed from annotation (Fig. 3.18). However, our results support the hypothesis that *Mapoly0008s0030.1* might be an independent transcriptional unit, denoted from having an independent 5' UTR and 3' UTR.
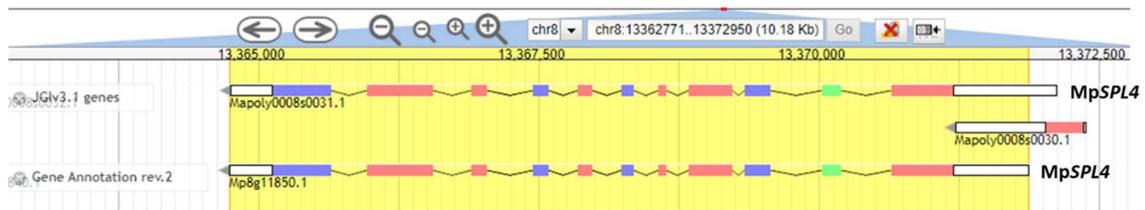
Figure 3.18: Genomic organisation of Mp*SPL4* (*Mapoly0008s0031.1/ Mp8g11850.1*) locus along with its neighbouring gene *Mapoly0019s0135.1/ Mp8g1160.1*, overlapping with 5'UTR of Mp*SPL4*. The schemes are based on data from MarpolBase version JGIv3.1 and version Gene annotation rev.2 (v6.1r2). Introns are represented by black lines, exons by boxes: white boxes indicate 5' and 3' UTRs while coloured boxes indicate CDS.

3.2.2 Profiling of Mp*SPL3* and Mp*SPL4* genes expression pattern during different developmental stages of Marchantia life cycle

In order to investigate the expression pattern of Mp*SPL3* gene transcripts in different developmental tissues of Marchantia, RT-qPCR analysis was performed. The primers were designed to be specific for each transcript isoforms. Therefore, for Mp*SPL3.1* gene, forward primer was designed in the region of intron 3 and reverse primer in exon 4 of Mp*SPL3.2*. While

for Mp*SPL3.2* gene, forward primer was designed in exon 3 while reverse primer was designed at the exon 3 – exon 4 junction (Fig. 3.19).
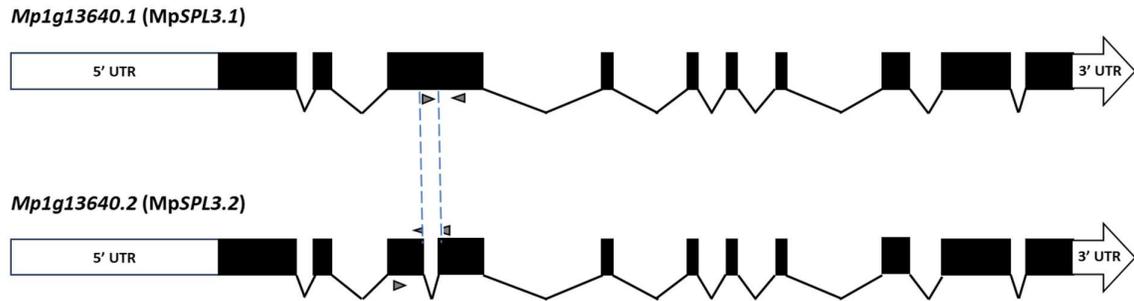
Figure 3.19: A schematic representation of primers position designed for analysis of Mp*SPL3* (*Mp1g13640*) gene transcripts level. Upper scheme represents isoform 1, Mp*SPL3.1* (*Mp1g13640.1*) lower scheme represents isoform 2, Mp*SPL3.2* (*Mp1g13640.2*). The schemes are based on data from MpTak_v6.1r2 annotation. The dashed blue lines marks the region covered by Intron 3 and triangles marks the position of designed primers.

In general, both transcripts are expressed in all the tissues under analysis (Fig. 3.20). Moreover, Mp*SPL3* mRNA level is the highest in archegoniophore and the lowest in 1-week-old male and female tissues. In general, both Mp*SPL3* transcripts were expressed at comparable levels in most of the tissues except antheridiophores, where Mp*SPL3.2* transcript level is more than double/ almost triple the level of Mp*SPL3.1*.
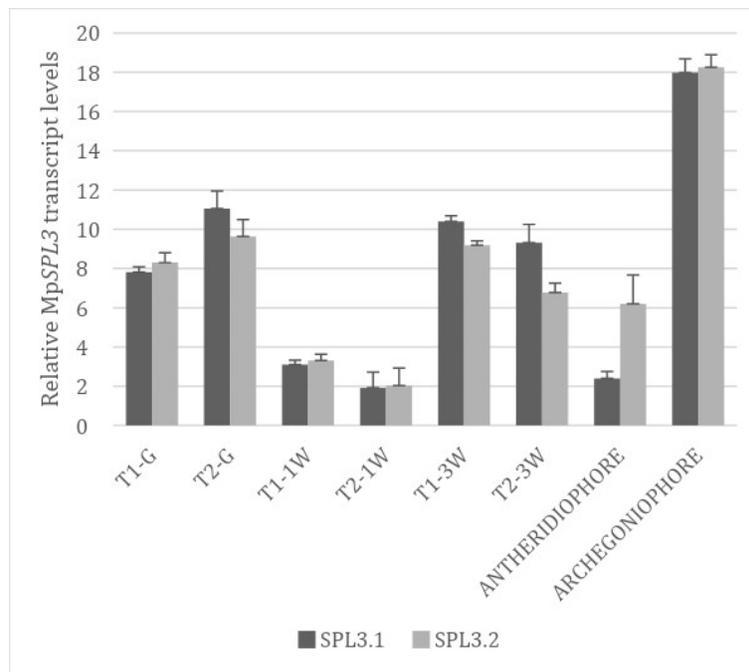
Figure 3.20: RT-qPCR analysis of expression level of Mp*SPL3* gene transcripts: Mp*SPL3.1* and Mp*SPL3.2*. The analysis was performed with two biological replicates in two technical replicates each. X-axis represents different developmental tissues of Marchantia and Y-axis represents the fold change. T1-G: Tak1-gemmae, T2-G: Tak2-gemmae, T1-1w: Tak1-1 week old, T2-1w: Tak2-1 week old, T1-3w: Tak1-3 week old, T2-3w: Tak2-3 week old.

Mp*SPL4* gene encodes a single mRNA isoform, *Mp8g11850.1* (Fig. 3.18 and 3.21). In order to investigate the expression pattern of Mp*SPL4* gene transcript in different developmental tissues of Marchantia, RT-qPCR analysis was performed. The location of primers designed is shown in Fig. 3.21.

Mp8g11850.1 (Mp*SPL4*)

Figure 3.21: A schematic representation of primers position designed for analysis of Mp*SPL4* (*Mp8g11850.1*) gene transcripts level. The scheme is based on data from MpTak_v6.1r2 annotation. The triangles marks the position of designed primers.

In general, Mp*SPL4* transcript is expressed in all the tissues under analysis (Fig. 3.22). The highest expression was observed in female reproductive organs, archegoniophores while the lowest was in female 1-week-old tissue (T2-1W in Fig. 3.22).
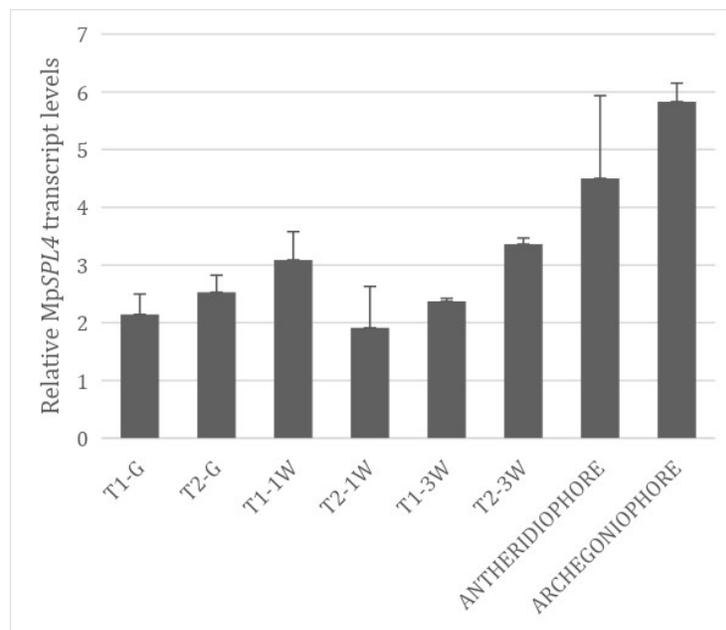
Figure 3.22: RT-qPCR analysis of expression level of Mp*SPL4* gene transcript. The analysis was performed with two biological replicates in two technical replicates each. X-axis represents different developmental tissues of Marchantia and Y-axis represents the fold change. T1-G: Tak1-gemmae, T2-G: Tak2-gemmae, T1-1w: Tak1-1 week old, T2-1w: Tak2-1 week old, T1-3w: Tak1-3 week old, T2-3w: Tak2-3 week old.

3.2.3 Characterisation of Mp*SPL3* and Mp*SPL4* gene promoters *in planta* activity

To investigate *in vivo* expression levels of Mp*SPL3* and Mp*SPL4* genes, their promoters' sequences corresponding to 5134 bp and 4107 bp upstream of the start codon, respectively were fused to the GUS reporter gene. The obtained recombinant transgenes were introduced into the Marchantia genome via Agrobacterium-mediated transformation. The histochemical GUS

staining for both gene promoters was observed in vegetative parts of male and female thalli (Fig. 3.23) as well as in the antheridial and archegonial receptacles of gametangiophores (Fig. 3.24). Importantly, the GUS expression pattern driven by $_{pro}$Mp*SPL3* and $_{pro}$Mp*SL4* are in coherent with the expression data available on Marchantia expression database (top right images in Fig. 3.23 and 3.24) (Kawamura *et al.*, 2022).
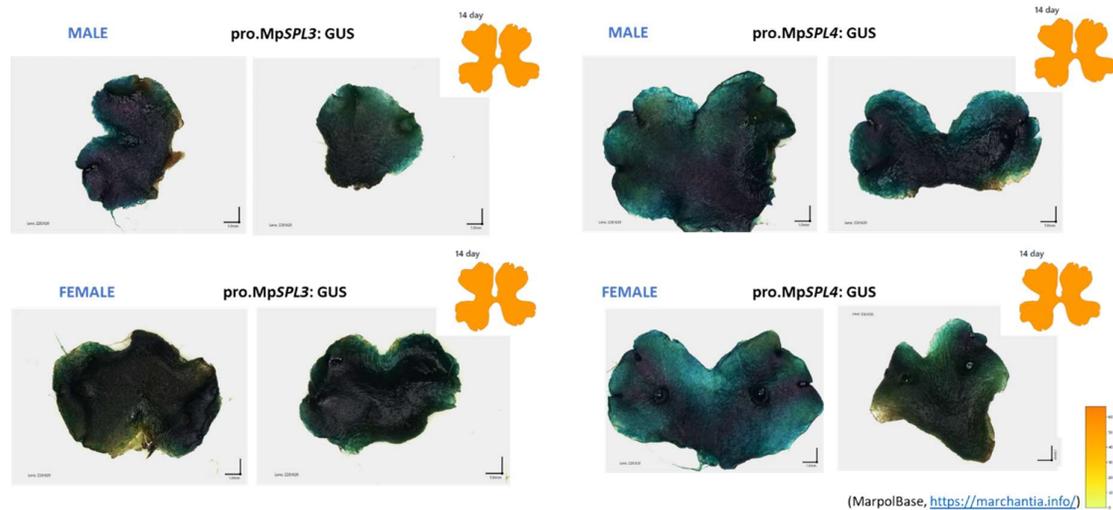
Figure 3.23: Promoter activity of Mp*SPL3* and Mp*SPL4* genes in Marchantia's vegetative tissue (male and female). The images on the top right of each GUS-expression images are a screenshot of expression pattern of Mp*SPL3* and Mp*SPL4* genes in Marchantia expression database. The scale below each image represents 1mm.
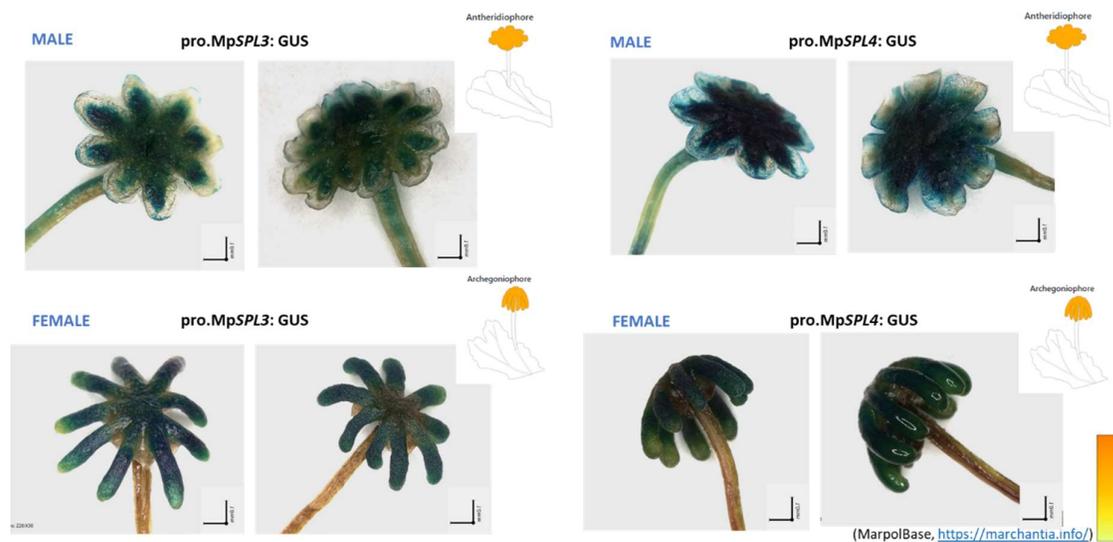
Figure 3.24: Promoter activity of Mp*SPL3* and Mp*SPL4* genes in Marchantia's reproductive organs (male and female). The images on the top right of each GUS-expression images are a screenshot of expression pattern of Mp*SPL3* and Mp*SPL4* genes in Marchantia expression database. The scale below each image represents 1mm.

3.2.4 Generation of knockout mutants of Mp*SPL3* using CRISPR/Cas9 approach

To learn about the function of Mp*SPL3* in Marchantia's life cycle, CRISPR/Cas9 system was used. In order to generate CRISPR/Cas9 knockout lines for Mp*SPL3* gene, five guide RNAs (gRNAs) were designed (Fig. 3.25): two gRNAs (gRNA 1 and 2) were designed upstream of the SBP domain coding region (blue), two gRNAs (gRNA 3 and 4) were designed within the gene region encoding SBP domain, and one gRNA (gRNA 5) was designed downstream of the SBP domain coding region.
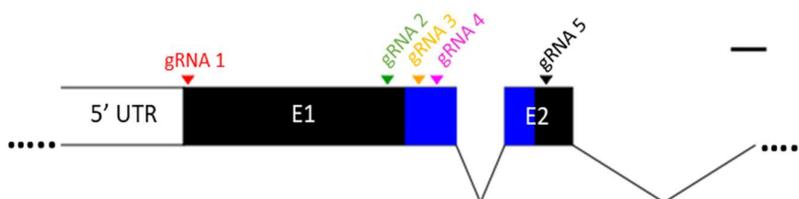
Figure 3.25: Schematic representation of five gRNA positions to target the Mp*SPL3* gene region. 5' UTR is marked as white box, E1 and E2 represent exons 1 and 2, marked as black boxes, SBP domain-coding region is marked as blue box, introns are marked as black lines. Each gRNA position is shown as red, green, yellow, pink and black arrows. The scale above represents 100bp.

After genotyping of transgenic lines at G1 generation, two knock-out lines were obtained with edited Mp*SPL3* locus (Mp*spl3-1.1* and Mp*spl3-1.3*). These two mutant lines were generated from gRNA 1. Both these transgenic lines were male. While there were two mutant lines obtained from gRNA 1, no plants with mutated MpSPL3 locus were obtained from other four gRNAs (Table 3.10).

Table 3.10: A summary of the number of plants genotyped and the number of plants obtained with edited locus at G1 generation for each gRNA for Mp*SPL3* gene.

	No. of plants genotyped (G1)	No. of plants with edited Mp*SPL3* locus (G1)
gRNA 1	47	2
gRNA 2	91	0
gRNA 3	96	0
gRNA 4	80	0
gRNA 5	76	0

Mp*spl3-1.1* has one nt substitution and 20 nt insertions and Mp*spl3-1.3* has 7 nt deletions (Fig. 3.26). Both types of mutations lead to shorter and truncated version of protein because of a premature stop codon.
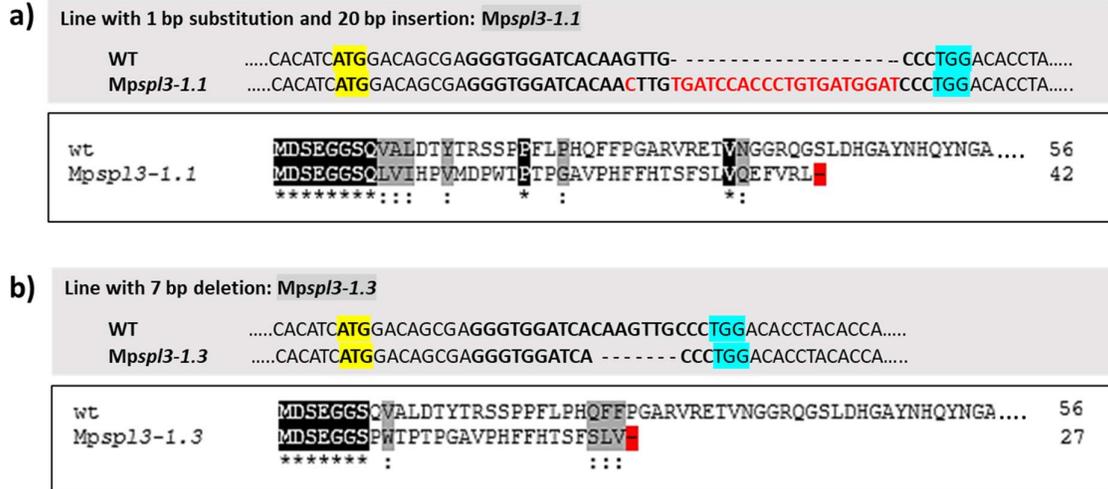
Figure 3.26: Schematic representation of the resulting mutations in the CRISPR/Cas9-generated Mp*spl3* alleles, Mp*spl3-1.1* (a) and Mp*spl3-1.3* (b). Alignment of genomic DNA and amino acid sequences between Mp*spl3*^ko allele and wild-type allele are shown in shaded and framed boxes, respectively. gRNA is shown in bold, PAM sequence is highlighted in blue. Start codon is highlighted in yellow and stop codon is highlighted in red.

In order to investigate mRNA level of Mp*SPL3* transcript in Mp*spl3*^ko lines, initially RT-PCR was performed with two pairs of primers (cds F + cds R1 and cds F + cds R2): primer cds F was designed 141 bp upstream the start codon and within 5' UTR and primers cds R1 and cds R2 were designed 66 bp and 219 bp downstream the start codon of Mp*SPL3* transcript. Additionally, primer cds R2 was designed downstream the stop codons of Mp*spl3-1.1* and Mp*spl3-1.3* transcripts. (Fig. 3.27A). No transcript was detected in both transgenic lines with both primer pairs, indicating the knockout of Mp*SPL3* gene (Fig. 3.27C).

Additionally, the transcript levels of Mp*SPL3* were analysed by RT-qPCR (with primers designed for RT-qPCR analysis: SPL3_f and SPL3_r, Fig. 3.27B) revealed that in both Mp*spl3*^ko lines, the Mp*SPL3* mRNA levels were almost undetectable in comparison to the WT plant (Fig. 3.27D).
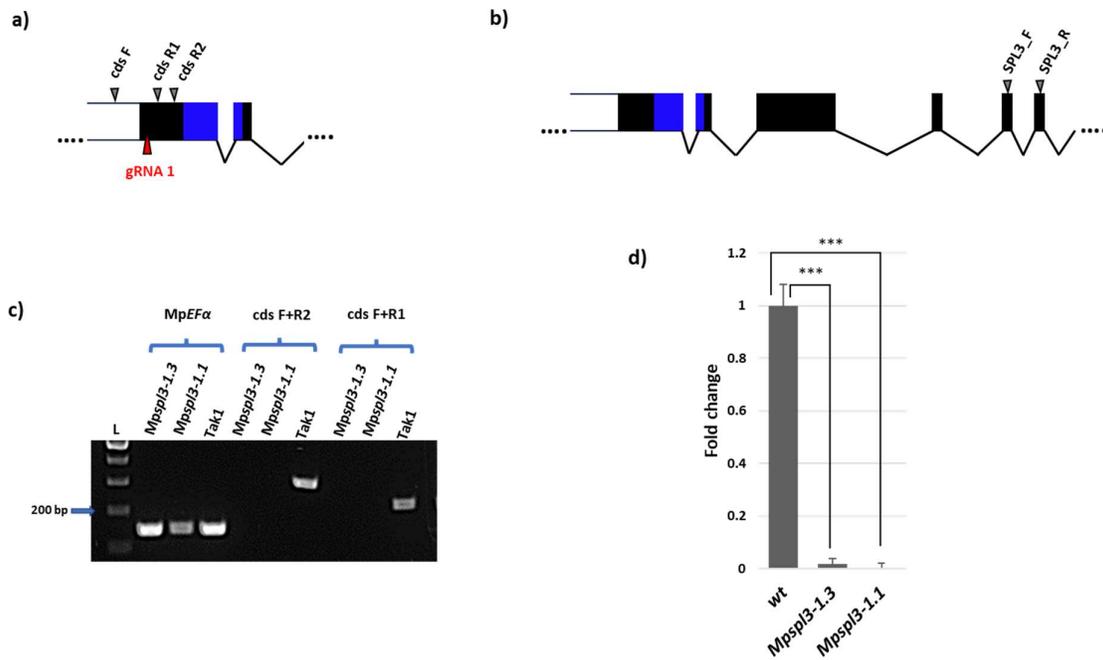
Figure 3.27: Analysis of Mp*SPL3* gene transcript levels in CRISPR/Cas9 mutant plants. A schematic representation of primers position designed for RT-PCR (a) and RT-qPCR (b). (c) Gel electrophoresis of RT-PCR products for detection of Mp*SPL3* mRNA levels in Mp*SPL3*^ko^ lines as compared to wild-type plants (Tak-1). (d) RT-qPCR for detection of Mp*SPL3* mRNA levels in Mp*SPL3*^ko^ lines as compared to wild-type plants. The analysis was performed with two biological replicates in two technical replicates each. p-values to determine the statistical significance were calculated using paired t-test ($p \leq 0.05$ = *, $p \leq 0.01$ = ** and $p \leq 0.001$ = ***.

Both transgenic lines exhibited strong alteration in phenotypes when compared to the wild-type plants (Fig. 3.28). Plants from both transgenic lines were observed under microscope at different time intervals, from 10th day to 5th week of growth. Plants were cultured on Gamborg's medium in phytotron under standard growth conditions.

The Mp*spl3-1.3* transgenic plants throughout their vegetative growth showed very strong growth retardation together with abolished thallus bifurcation, which is characteristic for wild-type thallus growth. Additionally, Mp*spl3-1.3* mutant plants showed significant decrease in surface area with a distinguishable callus-like phenotype. Meanwhile, after two weeks, few plants from Mp*spl3-1.1* transgenic line showed slightly less severe abnormalities as compared to Mp*spl3-1.3* mutant plants (Fig. 3.28). Moreover, Mp*spl3-1.1* transgenic line produced gemma cups but with reduced number as compared to wild-type plants. In conclusion, both Mp*spl3*^ko^ lines showed growth retardation as compared to wild-type plants in *in vitro* growth conditions.
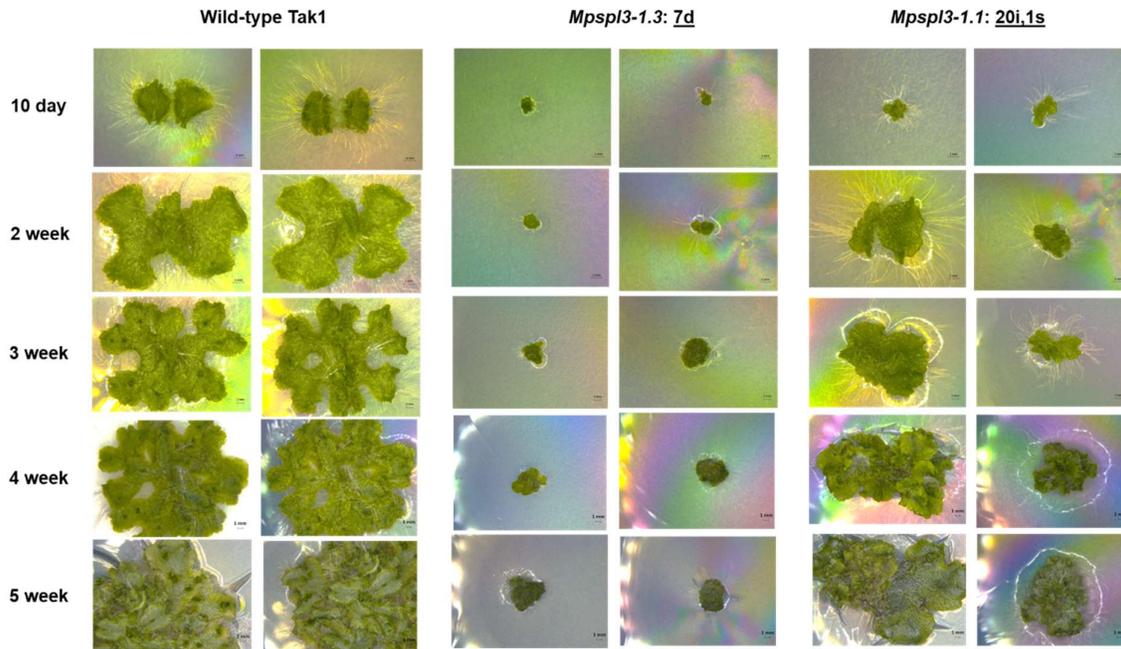
Figure 3.28: Phenotypic analysis of Mp*spl3*[ko] lines: Mp*spl3-1.1* and Mp*spl3-1.3* mutant plants were grown together with wild-type, Tak1 plants in *in vitro* culture. The phenotypic analysis is shown from 10-day old plants to 5-week-old plants. All the images were taken at 0.63X magnification in Leica M60 microscope. The scale below depicts 5mm for each image.

At the same time, the Mp*spl3*[ko] transgenic plants were transferred to soil to observe whether growth conditions will have impact on the plant phenotype. Their phenotypes in the presence of infra-red light along with wild-type Tak1 plants were examined. After 3 weeks of infra-red irradiation, Mp*spl3-1.3* transgenic lines showed severe growth abnormalities as compared to wild-type plants (Fig. 3.29a). These abnormalities included decrease in surface area, not characteristic wild-type-like thallus bifurcations and already visible necrotic tissue. Mp*spl3-1.1* transgenic lines also showed retarded growth as compared to wild-type plants. Although Mp*spl3-1.1* mutant plants showed wild-type-like thallus bifurcations at 3-week after far-red irradiation, they have reduced surface area and delayed growth as compared to wild-type plants at same stage of growth. At 5 and 7 weeks after far-red irradiation, when most of the wild-type plants started to produce gametangiophores (antheridiophores in this case), Mp*spl3-1.3* transgenic lines still showed severe growth retardation but with increase in necrotic tissue as compared to wild-type plants while Mp*spl3-1.1* transgenic lines showed wild-type-like thallus bifurcation (Fig. 3.29b and c).

a)

Wild-type Tak1   *Mpspl3-1.3*: <u>7d</u>   *Mpspl3-1.1*: <u>20i,1s</u>

4 cm   4 cm   4 cm

3 weeks after far-red irradiation

b)

Wild-type Tak1   *Mpspl3-1.3*: <u>7d</u>   *Mpspl3-1.1*: <u>20i,1s</u>

4 cm   4 cm   4 cm

5 weeks after far-red irradiation

c)

Wild-type Tak1   *Mpspl3-1.3*: <u>7d</u>   *Mpspl3-1.1*: <u>20i,1s</u>

4 cm   4 cm   4 cm

7 weeks after far-red irradiation
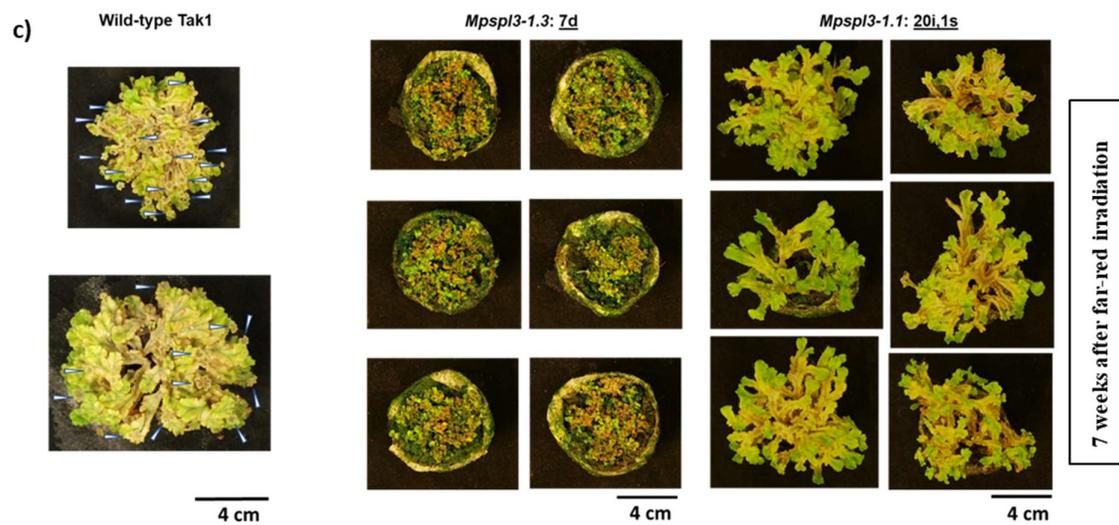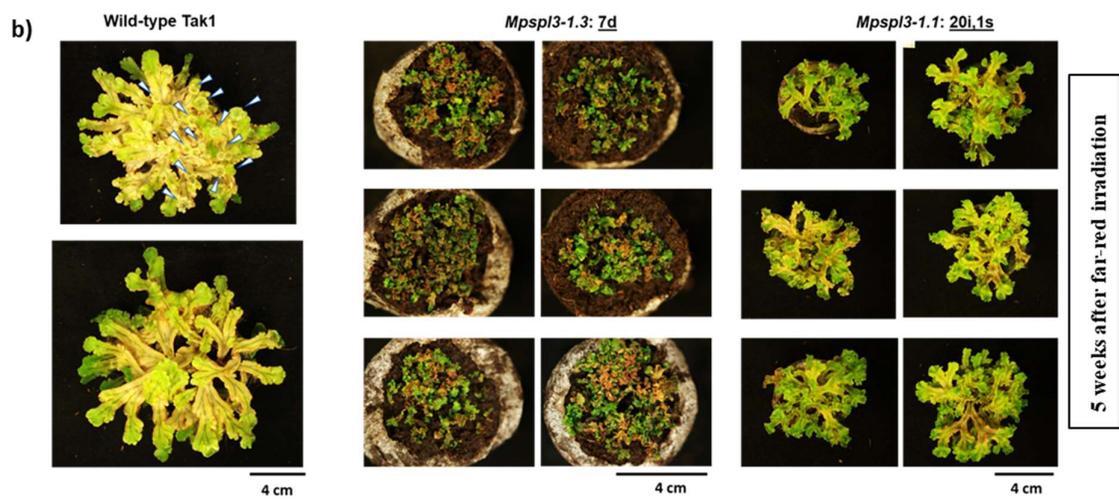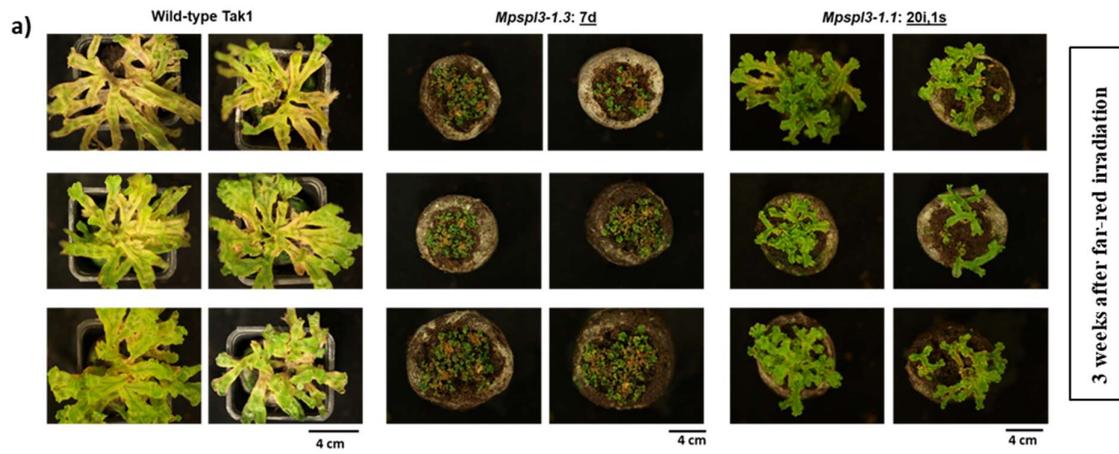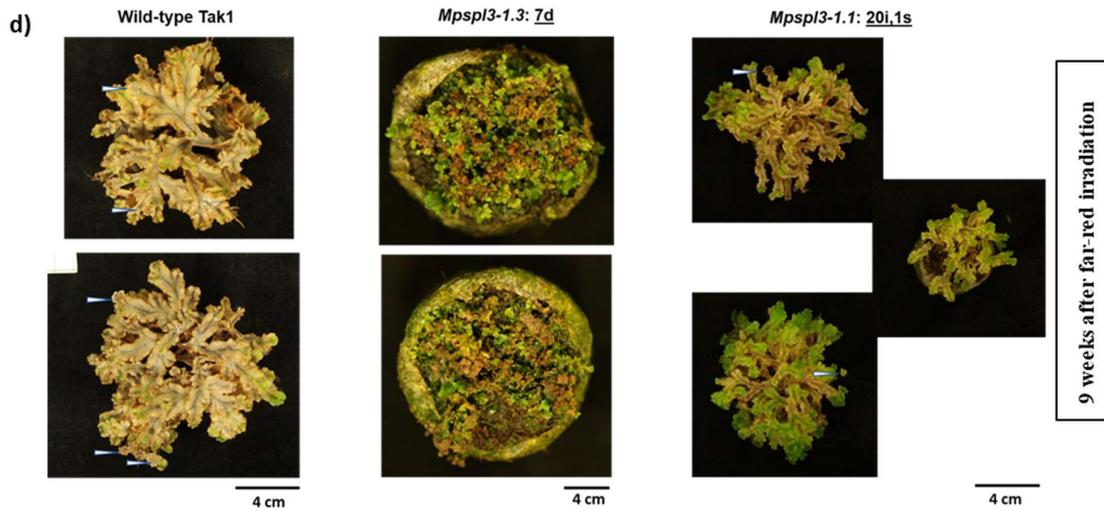
Figure 3.29: Phenotypic analysis of Mp*spl3*[ko] lines. Mp*spl3-1.1* and Mp*spl3-1.3* mutant plants were grown together with wild-type, Tak1 plants on soil. The images were taken at (a) 3 weeks (b) 5 weeks (c) 7 weeks and (d) 9 weeks after far-red induction. The arrows show the antheridiophores.

Interestingly, after 9 weeks of infra-red irradiation, when wild-type plants already started to show necrosis, Mp*spl3-1.3* transgenic plants also showed necrosis with growth retardation while Mp*spl3-1.1* transgenic lines started to produce antheridiophores, depicting delayed growth as compared with wild-type plants (Fig. 3.30d). The antheridiophores hence produced by Mp*spl3-1.1* plants on soil showed similar morphology as wild-type Tak-1 plants (Fig. 3.30).
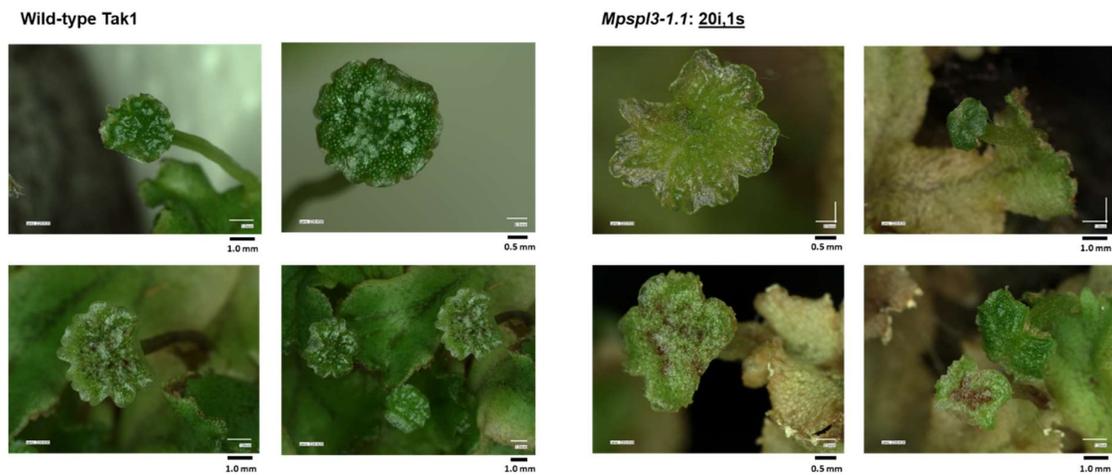
Figure 3.30: Phenotypic analysis of antheridiophores in Mp*spl3*[ko] line, Mp*spl3-1.1* as compared to wild-type, Tak1 plants on soil. All the images were taken in VHX-700 Keyence Digital microscope.

Furthermore, in order to perform in-depth analysis between antheridiophores of Mp*spl3-1.1,* the plants were cultured on Gamborg's media to have sterile conditions and were exposed to infra-red irradiation. To our surprise, Mp*spl3-1.1* transgenic lines did not produce any antheridiophores on Gamborg's media even after longer exposure to infra-red irradiation (upto

3-months) (Fig. 3.31). We propose this might be due to the presence of different nutrients in the soil as compared to Gamborg's media, which is used for culturing Marchantia.
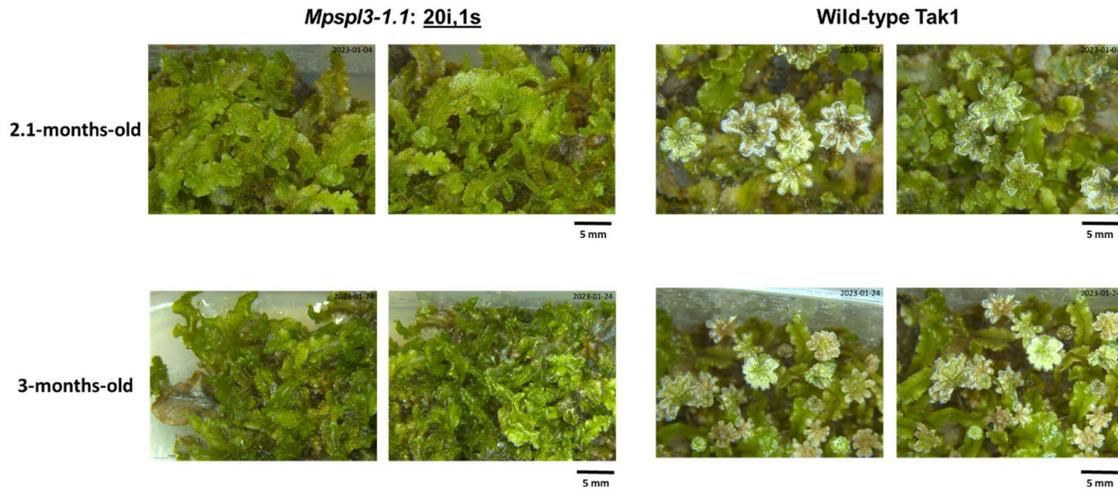
Figure 3.31: Phenotypic analysis of Mp*spl3*ko line, Mp*spl3-1.1* as compared to wild-type, Tak1 plants on Gamborg's media. The plants were imaged at 2.1 months and 3 months after far-red irradiation. The scale for Mpspl3-1.1 at 2.1-months-old stage is not correct as plants were photographed after tilting.

3.2.5 Generation of knockout mutants of Mp*SPL4* using CRISPR/Cas9 approach

To learn about the function of Mp*SPL4* in Marchantia's life cycle, CRISPR/Cas9 system was used. In order to generate CRISPR/Cas9 knock-out lines for Mp*SPL4* gene, five guide RNA (gRNA) were designed (Fig. 3.32) two gRNAs (gRNA 1 and 2) were designed upstream of the SBP domain coding region (blue), two gRNAs (gRNA 3 and 4) were designed within the gene region encoding SBP domain, and one gRNA (gRNA 5) was designed downstream of SBP domain coding region.
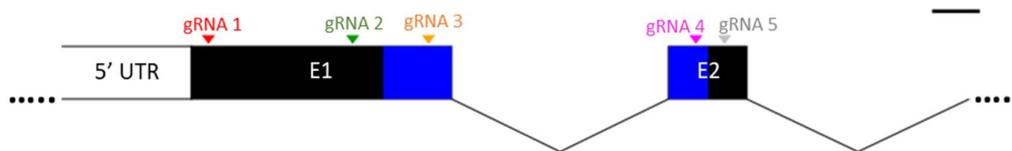
Figure 3.32: Schematic representation of five gRNA positions to target the Mp*SPL4* gene region. 5' UTR is marked as white box, E1 and E2 represent exons 1 and 2, marked as black boxes, SBP domain-coding region is marked as blue box, introns are marked as black lines. Each gRNA position is shown as red, green, yellow, pink and black arrows. The scale above represents 100bp.

After genotyping of transgenic lines at G1 generation, seven and five mutant plants with edited Mp*SPL4* locus were obtained for gRNA 3 and 5, respectively (Table 3.11).

Table 3.11: A summary of the number of plants genotyped and the number of plants with edited Mp*SPL4* locus at G1 generation for each gRNA.

SPL4	No of plants genotyped (G1)	No. of plants with edited Mp*SPL4* locus (G1)
gRNA 1	84	0
gRNA 2	119	0
gRNA 3	192	7
gRNA 4	74	0
gRNA 5	127	6

Surprisingly, all the knock-out lines obtained from gRNA 3 and five out of six knockout lines obtained from gRNA 5 converted to wild-type, when genotyped at their G2 generations (Table 3.12). To obtain mutants with same mutations in G2 generation, we went back to their respective G1 generation transgenic plants to culture and genotype more gemmae. These plants were overgrown similar to wild-type plants and, they lost their phenotype obtained from mutation, after longer culture on solid-culture. More surprisingly, these G1 plants when genotyped again also converted to wild-type. Hence, these G1 plants might be chimeric.

Table 3.12: A summary of sex and mutations of each mutant line with edited Mp*SPL4* locus at G1 and G2 generations. The edited mutant lines were obtained from gRNA 3 and 5.

Line No.	Sex	Mutation in G1 generation	Mutation in G2 generation
gRNA 3			
4	M	1 bp insertion	
17	M	1 bp insertion	
60	M	1 bp insertion	
75	M	1 bp insertion	Converted to wild-type
52	M	1 bp deletion	
47	M	7 bp deletion	
43	F	21 bp deletion	
gRNA 5			
17	M	1 bp insertion	Did not obtain G2 generation
16	M	1 bp insertion	
41	F	1 bp insertion	
58	M	1 bp insertion	Converted to wild-type
15	F	1 bp insertion	
9	M	1 bp deletion	

Meanwhile, one knock-out line obtained (Mp*spl4_17*) from gRNA5 obtained at G1 generation did not produce any gemma or gemma cups to propagate G2 generation. Hence, all analysis

was performed on G1 generation only. Mp*spl4* transgenic line has 1 bp insertion and this mutation caused premature stop codon and hence, a truncated version of protein (Fig. 3.33).

Figure 3.33: Schematic representation of the resulting mutations in the CRISPR/Cas9-generated Mp*spl4* locus. Alignment of genomic DNA and amino acid sequences between Mp*spl4*^ko allele and wild-type allele are shown in shaded and framed boxes, respectively. gRNA is shown in bold, PAM sequence is highlighted in blue. Start codon is highlighted in yellow and stop codon is highlighted in red. SBP domain is marked as navy blue residues.

Mp*spl4*^ko transgenic plants showed very strong phenotypic alterations in comparison to wild-type plants. These knockout plants were very small and highly retarded in growth as compared to wild-type plants. In general, these plants displayed prothallus-like phenotype (Fig. 3.34). Single wild-type Marchantia plant when cultured *in-vitro*, overgrows a petri plate within ~1.5 months. Meanwhile, Mp*spl4*^ko transgenic plant was strongly reduced in size and shape, remained green up to 4 months of culture when eventually, started to show necrosis in some parts of the plant (Fig. 3.34 – right panel).
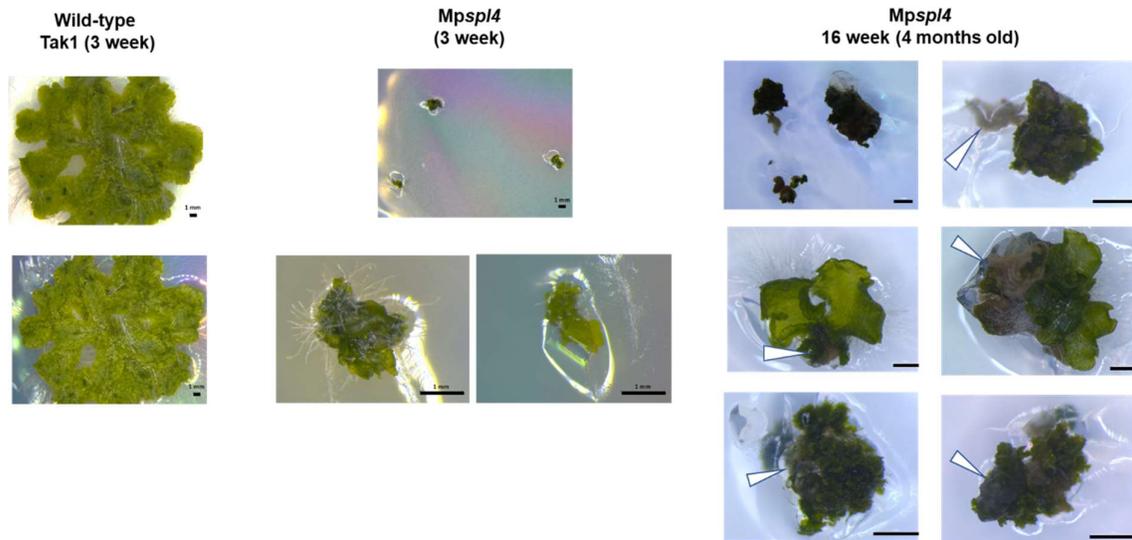
Figure 3.34: Phenotypic analysis of Mp*spl4*^ko line. Mp*spl4* was grown together with wild-type, Tak1 plants in *in-vitro* culture. The phenotypic analysis is shown at 3-weeks and 4-months after transfer of gemmae to *in-vitro* culture. Plants were photographed at G1 generation. All images were taken in Leica M60 stereo microscope. The scale below each image depicts 1mm.

The necrotic tissue was discarded, and the young green tissue was transferred to fresh medium and cultured further. After two months in *in-vitro* culture (in-total six months of culture), the Mp*spl4*^ko transgenic plants produced gemma cups (Fig. 3.35).
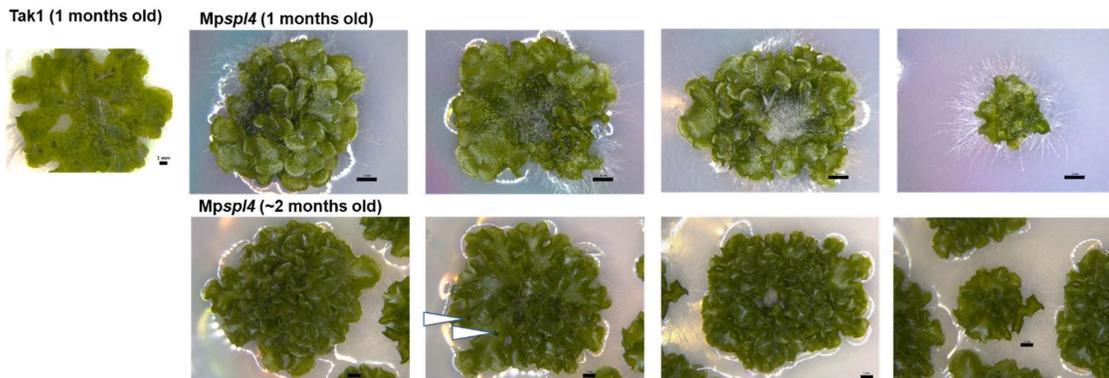
Figure 3.35: Phenotypic analysis of Mp*spl4*^ko line. Mp*spl4* was grown together with wild-type, Tak1 plants in *in-vitro* culture. The phenotypic analysis is shown at 1-month and 2-months after transfer of green tissues from Fig. 3.34 to fresh medium. All images were taken using Leica M60 stereo microscope. The arrow shows the position of gemma cups. The scale below each image depicts 1mm.

To our surprise, the G2 generation plants showed no phenotypic differences as compared to wild-type up to two weeks old (Fig. 3.36). Therefore, G2 plants were genotyped to confirm the presence of mutation.
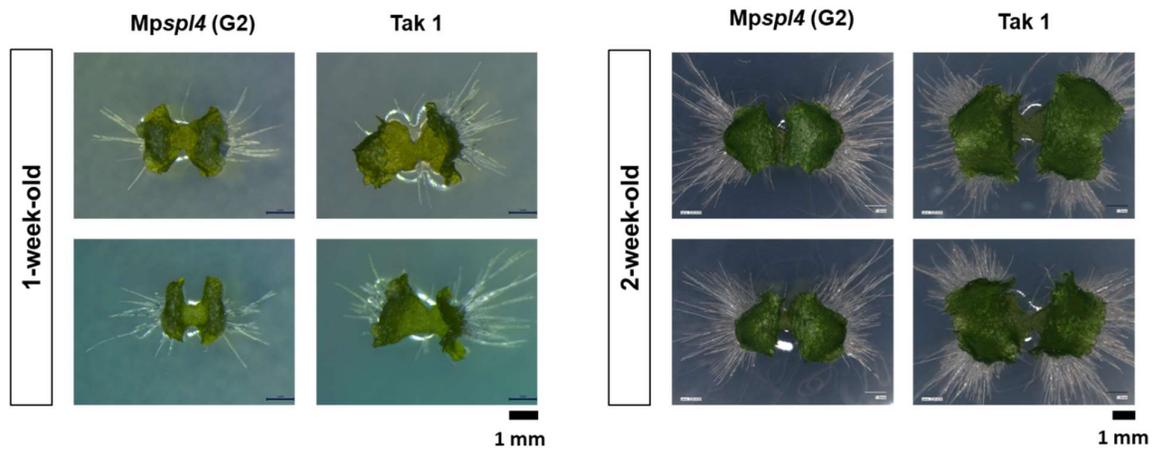
Figure 3.36: Phenotypic analysis of Mp*spl4*[ko] line in G2 generation. Mp*spl4* was grown together with wild-type, Tak1 plants in *in vitro* culture. The phenotypic analysis was shown at 1-week and 2-weeks after transfer of gemma in Fig. 3.36 to *in vitro* culture. The plants were photographed at G2 generation. All the images were taken in Leica M60 microscope.

Surprisingly, G2 generation showed no mutation in its genome. Therefore, Mp*spl4*[ko] G1 plants that produced gemma used for G2 generation growth (Fig. 3.35) were again genotyped. To our surprise, all G1 transgenic plants in addition to the primary 1 bp insertion had an additional 1-bp deletion (depicted as Mp*spl4_b* in Fig. 3.37) which in final resulted in one amino acid substitution (D→E, position 222) in comparison to wild-type allele. In this way the premature stop codon was removed and a functional copy of Mp*SPL4* locus was generated (Fig. 3.37).

Figure 3.37: Schematic representation of the resulting mutations in the CRISPR/Cas9-generated Mp*spl4* locus. Mp*spl4* young tissue had original Mpspl4 mutation while old tissue showed new mutation, depicted as Mp*spl4_b*. Alignment of genomic DNA and amino acid sequences between Mp*spl4*[ko] allele and wild-type allele are shown in shaded and framed boxes, respectively. gRNA is shown in bold; PAM sequence is highlighted in blue, the conversion from D→E at position 222 is shaded in grey. Start codon is highlighted in yellow and stop codon is highlighted in red. SBP domain is marked as navy-blue residues.

The fresh young tissue obtained after 4 months of growth in *in vitro* culture for Mp*spl4*[ko] transgenic line were also transferred to soil. After 9 weeks on soil, in the case of single plant an outgrowth was obtained (depicted as '1' in Fig. 3.38a) that showed better growth in comparison to three other plants. After 19 weeks, this outgrowth (depicted as '1' in Fig. 3.3b) over-grew other plants and started to grow like wild-type plants. Since we already experienced the genotype conversion of Mp*spl4*[ko] to wild-type in *in-vitro* culture, therefore this outgrowth and other tissues of Mp*spl4*[ko] (depicted as '1' and '2' in Fig. 3.38b) of transgenic line were genotyped.
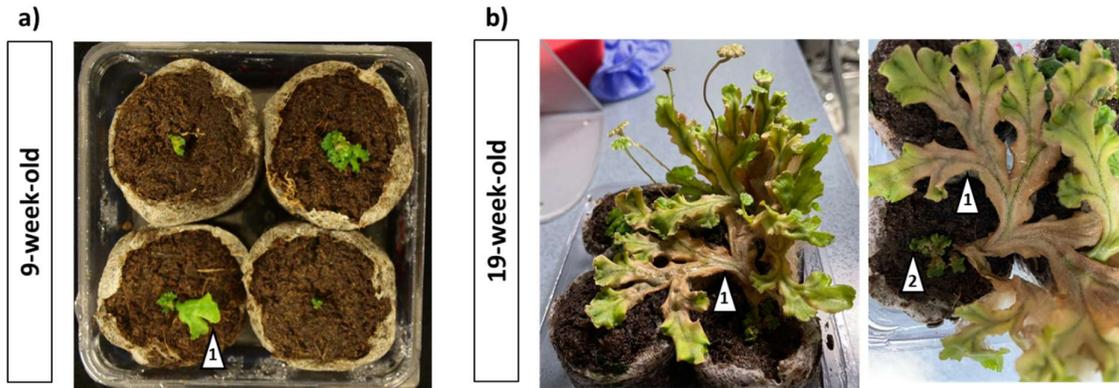
Figure 3.38: Phenotypic analysis of Mp*spl4*[ko] line, Mp*spl4* on soil. The images were taken at a) and 9 weeks b) 19 weeks after far-red induction.

Again, to our surprise, the outgrowth also showed 1 bp deletion, the same as in the case which caused only one amino-acid difference (D→E, position 222) between wild-type and mutant copy of Mp*SPL4* (Fig. 3.39 and Fig. 3.38) while the other plants still showed the same mutation of 1 nt insertion (Fig. 3.39).

Figure 3.39: Schematic representation of the resulting mutations in the CRISPR/Cas9-generated Mp*spl4* locus. Mp*spl4_2* had original Mp*spl4* mutation while Mp*spl4_1* showed similar mutation to Mp*spl4_b*. Alignment of genomic DNA and amino acid sequences between Mp*spl4*ᵏᵒ allele and wild-type allele are shown in shaded and framed boxes, respectively. gRNA is shown in bold; PAM sequence is highlighted in blue, the conversion from D ⯈ E at position 222 is shaded in grey. Start codon is highlighted in yellow and stop codon is highlighted in red. SBP domain is marked as navy-blue residues.

The plant tissues which still retained the mutation (depicted as '2' in Fig. 3.3b) in Mp*spl4*ᵏᵒ transgenic line, were transferred to fresh soil again to be separated from the plant with reverted genotype and used for phenotype analysis. Wild-type plants started to show visible necrosis at 7 weeks after infra-red, hence these plants are shown as control in fig. 3.40. Similarly, as in *in vitro* culture, the Mp*spl4*ᵏᵒ transgenic plants showed very severe retardation of growth with decreased thallus surface area, no formation of gemma cups and gametangiophores, even after 7 months under infra-red irradiation (in-total 19 weeks before and 7 months later, so almost a year from starting the G1 generation culture) (Fig. 3.40).
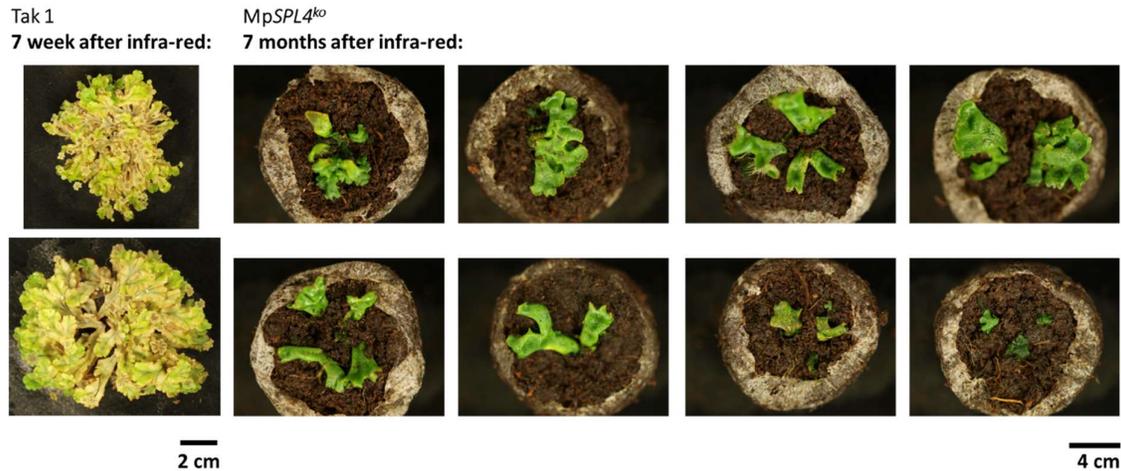
Figure 3.41: Phenotypic analysis of Mp*spl4*^ko line grown on soil. Mp*spl4* was grown together with wild-type, Tak1 plants. Images of wild-type and Mp*spl4*^ko mutant plants were taken 7 weeks and 7 months after far-red induction, respectively.

Since, all transgenic lines obtained for Mp*SPL4* gene were shown to possess unstable mutations, in the next attempt, CRISPR/Cas9 experiment was repeated but with double gRNA approach. Since with the use of single guide RNA approach, there is a risk of off-target editing, using double guide RNA approach can overcome this obstacle. In double guide RNA approach, Cas9 nickase (Cas9^{D10A}) is used (mutant Cas9 nickase generate single stranded breaks as opposed to wild-type Cas9 which produces double stranded breaks) which minimizes off-target double strand breaks, hence there is no need to scan off-target sites. Therefore, this approach is more efficient than conventional CRISPR/Cas9 approach. The two gRNAs were designed within the region encoding SBP domain of Mp*SPL4* gene (Fig. 3.41).
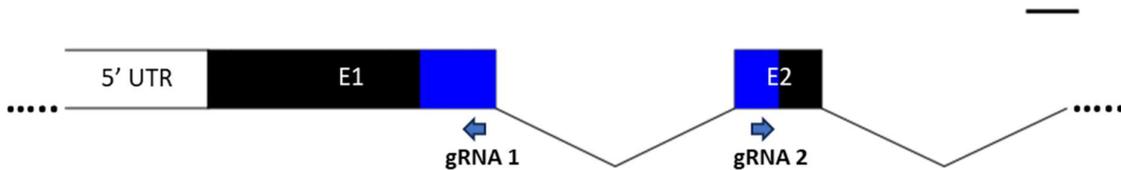
Figure 3.41: Schematic representation of two gRNA positions to target the Mp*SPL4* gene region by double gRNA approach. 5' UTR is marked as white box, E1 and E2 represent exons 1 and 2, marked as black boxes, SBP domain-coding region is marked as blue box, introns are marked as black lines. Both gRNA positions are marked as blue arrows. The scale above represents 100bp.

From this approach, 12 mutant lines with edited Mp*SPL4* locus were obtained in G1 generation (Table 3.13). All these 12 mutant lines produced gemma cups at the G1 generations and hence, could be propagated to the next G2 generation, respectively. To our surprise again, 10 out of 12 transgenic lines lost the mutation in their G2 generation and had reverted to wild-type genotype. However, two transgenic lines, #3 and #54.2, retained the same mutation in G2

generation as in G1 generation (Table 3.13). Importantly, G1 generation of #3 and #54.2 lines had reduced number of gemmae and gemma cups.

Table 3.13: A summary of mutations in each mutant line with edited Mp*SPL4* locus at G1 and G2 generations. The edited mutant lines were obtained from double gRNA approach.

Line No.	Mutation at G1 generation	Mutation at G2 generation
80.2	27 bp deletion and 7 bp substitution	Converted to wild-type
51.3	47 bp deletion	
2.2	29 bp deletion and 6 bp substitution	
4.4	34 bp deletion and 5 bp substitution	
91.1	13 bp deletion and 3 bp substitution	
66.3	16 bp deletion and 3 bp substitution	
55.4	24 bp deletion	
52.2	7 bp substitution and 20 bp insertion	
79.1	3 bp deletion, 2 bp insertion and 13 bp substitution	
91.2	65 bp deletion and 1 bp substitution	
3	513 bp deletion	513 bp deletion
54.2	11 bp deletion and 2 bp substitution	11 bp deletion and 2 bp substitution

Mp*spl4*^ko transgenic line #3 (Mp*spl4_3*) revealed 29 bp deletion in exons 1 and 2 and 484 bp deletion in intron sequence (whole intron 1 deletion) while line #54.2 (Mp*spl4_54.2*) revealed 11 bp deletion and 2 bp substitution in exon 1. In both transgenic lines, the Mp*SPL4* coding sequence was changed to encode only one zinc-finger binding motif of the conserved SBP-domain (Fig. 3.42).

b)

```
WT          .....TCAGCCCCACCGATACTGTCAGCAATGTGGCAA.....

Mpspl4_54.2 .....TCAGCCCCACCGATATA - -CAG - - - - - - - - -AA.....
```

```
wt            MAHGHETGLAWEWDSVLLLANPSLTSHSLEGDGSTSCGLPGSSAADQNHEHDNHCHEHNH    60
Mpspl4_54.2   MAHGHETGLAWEWDSVLLLANPSLTSHSLEGDGSTSCGLPGSSAADQNHEHDNHCHEHNH    60
              ************************************************************

wt            EHNHDKFHAHETSSADHVQTDSISSGSQQMGVNTDWRDPRLDCPNFLAGRVPCACTDNDD    120
Mpspl4_54.2   EHNHDKFHAHETSSADHVQTDSISSGSQQMGVNTDWRDPRLDCPNFLAGRVPCACTDNDD    120
              ************************************************************

wt            DDSGVSRKRSKPVPRCQVQSCGAELTNLKGYHQRHRVCLRCAHATRVVLRNQPHRYCQQC    180
Mpspl4_54.2   DDSGVSRKRSKPVPRCQVQSCGAELTNLKGYHQRHRVCLRCAHATRVVLRNQPHRYTEVS    180
              *******************************************************

wt            GKFHPICDFDEGKRSCRRKLERHNNRRRRKALESEDTLPPLDSGDIEPRSDSSPAD.....   236
Mpspl4_54.2   SHL                                                             183
```

Figure 3.42: Schematic representation of the resulting mutations in the CRISPR/Cas9-generated Mp*spl4* locus by double gRNA approach. a) Mp*spl4_3* and b) Mp*spl4_54.2* mutant lines. Alignment of genomic DNA and amino acid sequences between Mp*spl4*[ko] allele and wild-type allele are shown in shaded and framed boxes, respectively. gRNA is shown in bold; PAM sequence is highlighted in blue; stop codon is highlighted in red. SBP domain is marked as navy-blue residues.

The phenotypic analysis of both transgenic lines, Mp*spl4_3* and Mp*spl4_54.2* at their G2 generation, revealed strong growth retardation, as 3-month-old plants resembled few days old wild-type gemmae. Additionally, the mutant plants showed severe decrease in size with no signs of gemma cup formation (Fig. 3.43). After 5 months, all the transgenic lines started to die due to necrosis (Fig. 3.44).
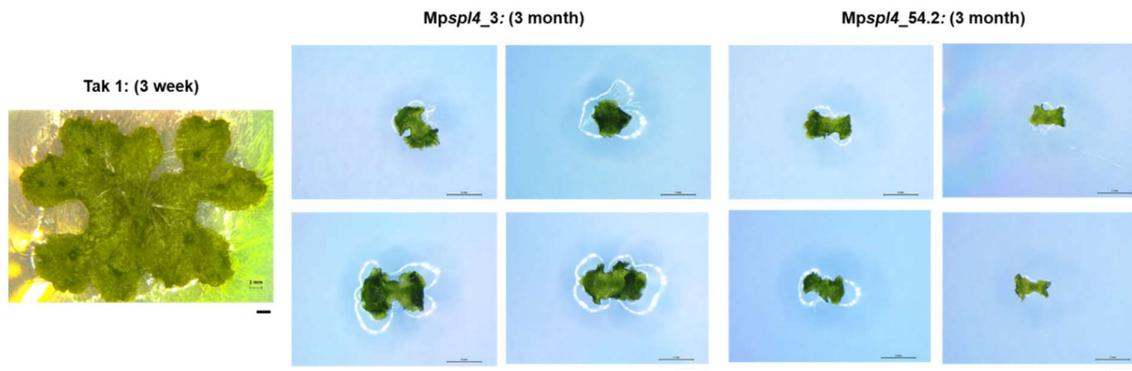
Figure 3.43: Phenotypic analysis of Mp*spl4*[ko] lines, Mp*spl4_3* and Mp*spl4_54.2*. The mutant lines were grown together with wild-type, Tak1 plants in *in vitro* culture. The images of wild-type plants are taken at 3-week-old while the images of Mp*spl4*[ko] lines is taken at 3 months. All images were taken in Leica M60 microscope. The scale below wild-type and Mp*spl4*[ko] lines depicts 1mm and 5mm, respectively.
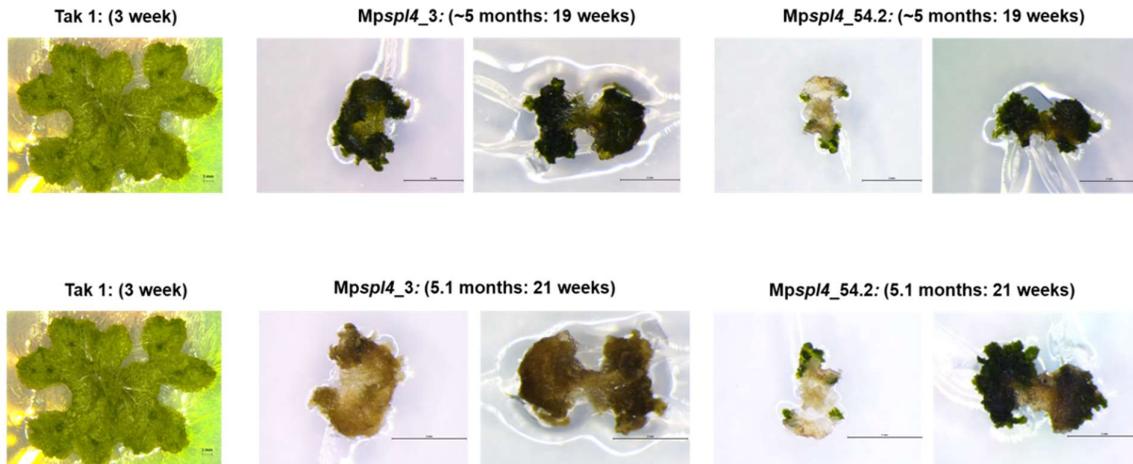
Figure 3.44: Phenotypic analysis of Mp*spl4*ᵏᵒ lines, Mp*spl4_3* and Mp*spl4_54.2*. The mutant lines were grown together with wild-type, Tak1 plants in *in vitro* culture. The images of wild-type plants are taken at 3-week-old while the images of Mp*spl4*ᵏᵒ lines is taken at 19 weeks and 21 weeks. All the images were taken in Leica M60 microscope. The scale below wild-type and Mp*spl4*ᵏᵒ lines depicts 1mm and 5mm, respectively.

3.2.6 Generation of knockdown mutants by artificial miRNA approach

Since knock-out lines obtained for Mp*SPL3* and Mp*SPL4* genes by using CRISPR/Cas9 approach showed very serious abnormalities and retardation of growth, hence alternative approach using artificial miRNA was used to investigate the impact of Mp*SPL3* and Mp*SPL4* genes depletion on Marchantia's development. For both genes, artificial miRNAs were designed according to protocol described by Sandoval et al., 2016.

3.2.6.1 Generation of Mp*SPL3* knock-down lines using artificial miRNA

Based on the protocol published by Sandoval et al., 2016, artificial miRNA for Mp*SPL3* was designed to target Mp*SPL3* gene transcript within the coding sequence in exon 1 (Fig. 3.45).
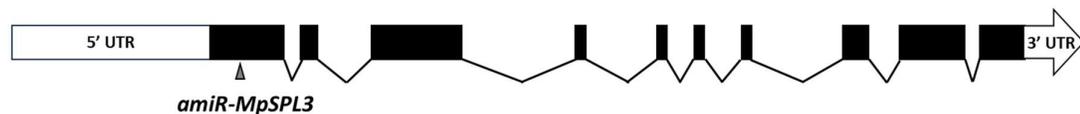
Figure 3.45: Schematic representation of artificial miRNA position to target the Mp*SPL3* gene transcript. The scheme is based on data from MpTak_v6.1r2 annotation.

After transformation of Marchantia wild-type plants harbouring the *amiR*-Mp*SPL3* cassette, seven mutant lines were obtained in G2 generation. All these mutant lines showed changes of plant phenotype in comparison to wild-type plants. Hence, all these seven transgenic lines were analysed for the expression levels of of pri-amiR and Mp*SPL3* mRNA. Out of these seven mutant lines, five: *amiR*-Mp*SPL3-1a, 1b, 2b, 2c, 4b*, and *6*, had significantly higher pri-amiR levels (Fig. 3.46a). From these lines, four (*amiR*-Mp*SPL3-2b, 2c*, *4b* and *6*) showed

significantly reduced levels of Mp*SPL3* gene transcripts (Fig. 3.46b), what is consistent with amiR-mediated knockdown of Mp*SPL3* expression. Hence, these four mutant lines were chosen for further phenotypic analysis.
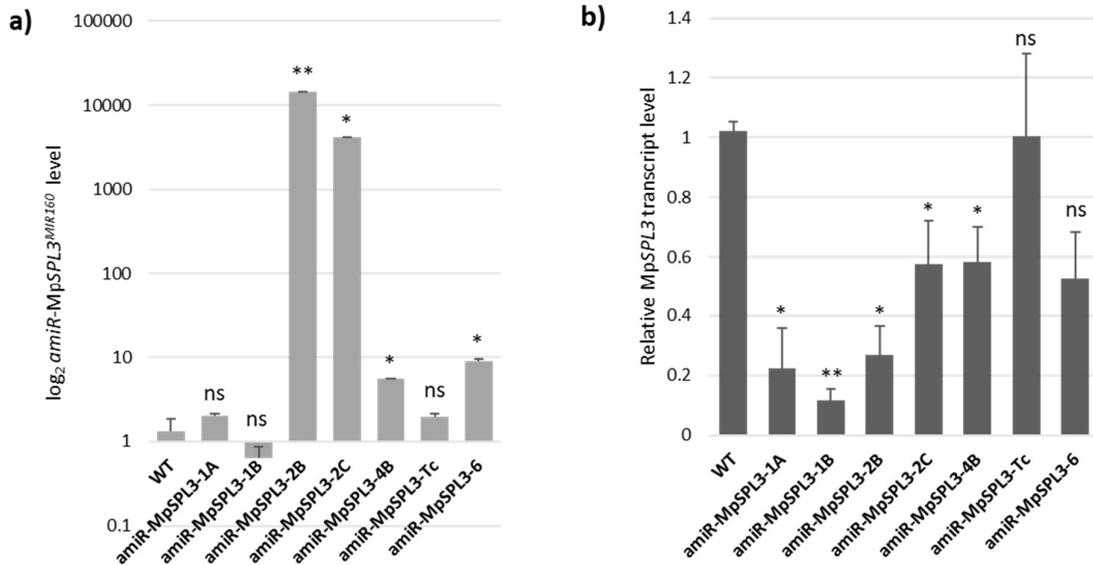
Figure 3.46: Analysis of Mp*SPL3* and pri-amiR-Mp*SPL3* expression level in Mp*spl3-amiR^{MIR160}* plants. (a) *amiR-*Mp*SPL3^{MIR160}* expression relative to wild-type mRNA, and (b) Mp*SPL3* expression relative to wild-type mRNA measured by RT-qPCR. The analysis was performed with two biological replicates in two technical replicates each. p-values to determine the statistical significance were calculated using paired t-test ($p \leq 0.05$ = *, $p \leq 0.01$ = ** and $p \leq 0.001$ = ***).

For phenotypic analysis, plants were observed first during their vegetative stage of growth (at 3-weeks) and later in their reproductive stage of growth (at >2 months; after induction of gametangiophores). These two timelines were chosen according to wild-type life cycle. Two mutant lines, *amiR-*Mp*SPL3-4b* and *6* (both male), revealed similar phenotype i.e. no proper thallus bifurcation as compared with characteristic thallus bifurcations of wild-type plants and absence of gametangiophores production even after ~2 months of infra-red irradiation. Additionally, both these mutant lines after 3 months of growth, had most of their thalli sunken inside the solid medium with only young apical thallus part and gemma cups protruded out of medium (hence, brown colour of agar in the Fig. 3.47) Another selected mutant lines, *amiR-*Mp*SPL3-2b* (Female) and *2c* (Male), also showed similar phenotype i.e. delayed growth and no production of gametangiophores after ~1 month of infra-red irradiation (Fig. 3.47). Hence, it can be clearly seen that with elevated levels of *amiR-*Mp*SPL3^{MpMIR160}*, gametangiophores production (as observed in all mutant lines), thallus bifurcation (as observed in *amiR-*Mp*SPL3-4b* and *6* lines) and thallus growth (as observed in *amiR-*Mp*SPL3-2b* and *2c* lines) were affected. Therefore, these four mutant lines currently are under in-depth phenotypic analysis.
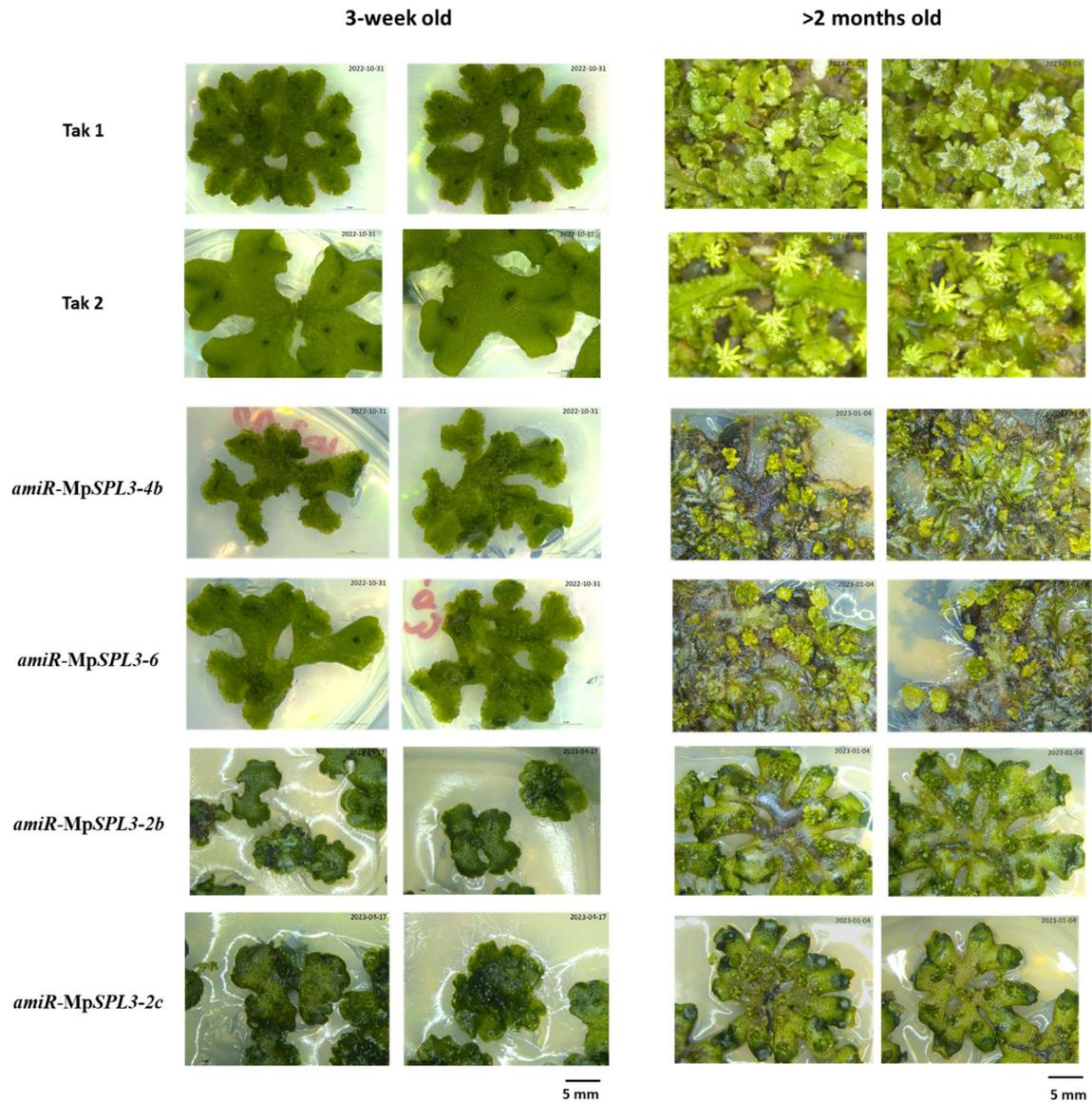
Figure 3.47: Phenotypic analysis of Mp*SPL3* knockdown lines generated using artificial miRNA. The mutant lines were grown together with wild-type, Tak1 and Tak2 plants in *in-vitro* culture. The images were taken at 3-weeks old and 3 months old (except *amiR*-Mp*SPL3-2b* and *2c*, at 2 months old). All the images were taken in Leica M60 microscope.

3.2.6.2 Generation of Mp*SPL4* knock-down lines using artificial miRNA

Based on the protocol published by Sandoval et al., 2016, artificial miRNA for Mp*SPL4* was designed to target Mp*SPL4* gene transcript within the coding sequence in exon 10 (Fig. 3.48).

Mp*SPL4*
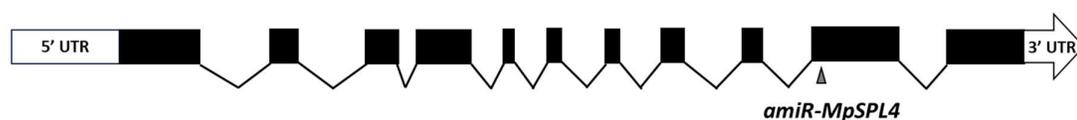
Figure 3.48: Schematic representation of artificial miRNA position to target the Mp*SPL4* gene transcript. The scheme is based on data from MpTak_v6.1r2 annotation.

By using artificial miRNA to knockdown Mp*SPL4* gene expression, more than 90% of regenerating plants in T1 generation grew in a clump-like structure with no wild-type thallus-like bifurcations and gemma cups production (Fig. 3.49).
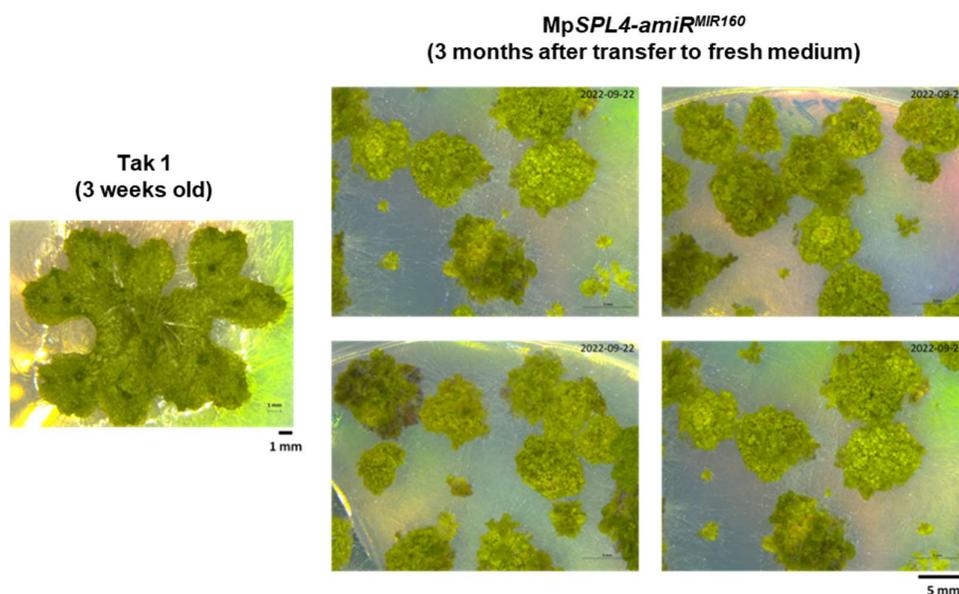
Figure 3.49: Phenotypic analysis of Mp*SPL4* knock-down lines generated using artificial miRNA, Mp*spl4-amiR*^{MIR160}. The mutant lines were grown together with wild-type, Tak1 in *in-vitro* culture. The images were taken for 3-weeks old wild-type Tak-1 plants and 3 months old Mp*spl4-amiR*^{MIR160} plants at T1 generation. All the images were taken using Leica M60 stereo microscope.

Moreover, these lines did not produce any gemmae, hence, the molecular analyses were performed using plants from T1 generation. Three independent transgenic T1 mutant lines were tested for the presence of the *amiR*-Mp*SPL4*^{MpMIR160} transcript and for the expression level of Mp*SPL4* gene. All three mutant lines revealed overexpression of *amiR*-Mp*SPL4*^{MpMIR160} transcript (Fig. 3.50a) with simultaneous significant reduction of the Mp*SPL4* gene transcript (Fig. 3.50b) as compared to wild-type plants. Therefore, high levels of *amiR*-Mp*SPL4*^{MpMIR160} caused amiR-mediated knockdown of Mp*SPL4* what resulted in very strong defects of plant growth and morphology resembling Mp*SPL4*^{ko} CRISPR/Cas9 obtained plants phenotype (Fig. 3.34 and 3.43).
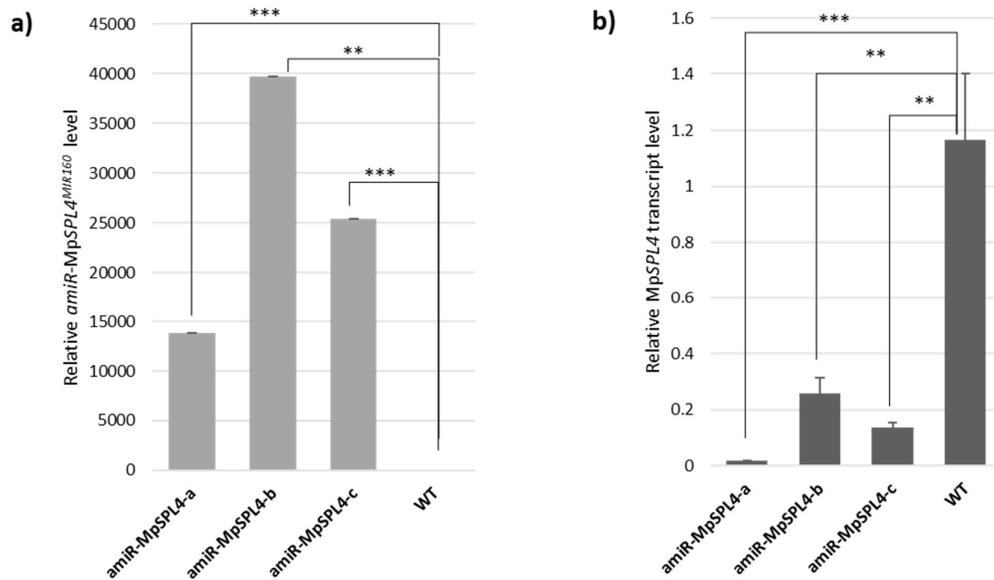
Figure 3.50: Analysis of Mp*SPL4* and pri-amiR-Mp*SPL4* expression level in Mp*spl4-amiR^{MIR160}* plants in T1 generation. (a) *amiR*-Mp*SPL4^{MIR160}* expression relative to wild-type mRNA, and (b) Mp*SPL4* expression relative to wild-type mRNA measured by RT-qPCR. The analysis was performed with one biological replicates in three technical replicates each (since each transgenic T1 line was treated as one biological replicate). p-values to determine the statistical significance were calculated using paired t-test ($p \leq 0.05 = *$, $p \leq 0.01 = **$ and $p \leq 0.001 = ***$.

While most of the transgenic *amiR*-Mp*SPL4*^{Mp*MIR160*} plants grew only in T1 generation, there were four transgenic lines that produced gemmae cups and hence, their G2 generations were propagated. All these transgenic lines were tested for the presence of *amiR*-Mp*SPL4*^{Mp*MIR160*} transcript and for the expression levels of Mp*SPL4* gene transcript (Fig. 3.52a and b). Although the *amiR*-Mp*SPL4*^{Mp*MIR160*} transcript was overexpressed in all four transgenic lines, in none of them the Mp*SPL4* expression was significantly reduced. However, we did experiments only for one biological replication and the experiments have to be continued later.
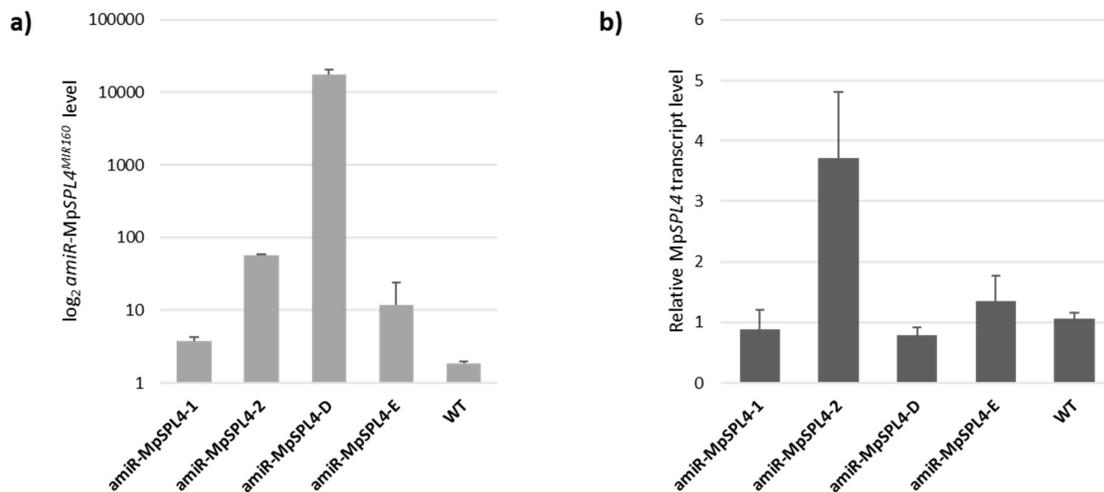
Figure 3.51: Analysis of Mp*SPL4* and pri-amiR-Mp*SPL4* expression level in Mp*spl4-amiR^{MIR160}* plants in G2 generation. (a) *amiR*-Mp*SPL4^{MIR160}* expression relative to wild-type mRNA, and (b) Mp*SPL4* expression relative to wild-type mRNA measured by RT-qPCR. The analysis was performed with one biological replicate in three technical replicates each.

The mutant line *amiR*-Mp*SPL4-D* had highest levels of *amiR*-Mp*SPL4^{MpMIR160}* but with no downregulation of Mp*SPL4* transcript. This might be because of either using only one biological replicate for our analysis or because of miRNA off-target effects. However, this mutant line displayed significant differences in plant phenotype as compared to wild-type plants i.e. layered thallus (hyperbranching) and absence of any gametangiophores production after ~2 months of far-red irradiation (Fig. 3.52).
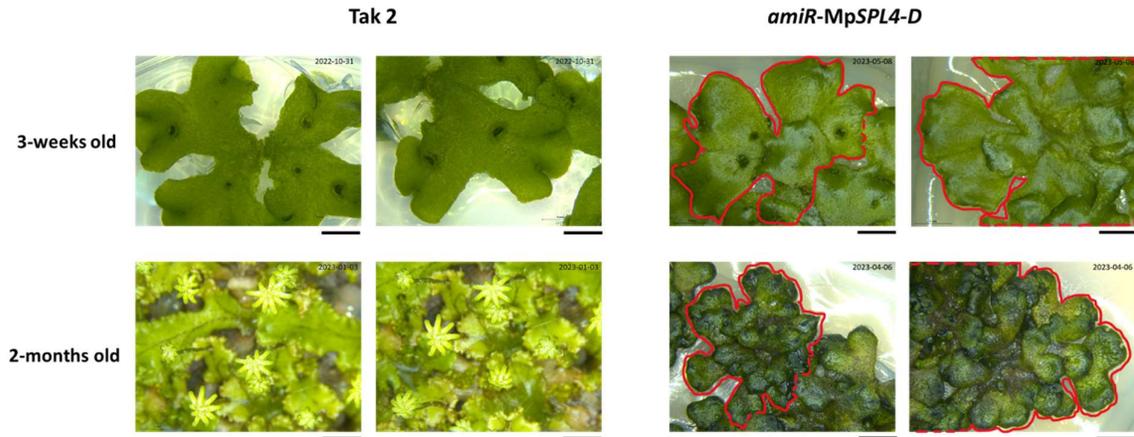
Figure 3.52: Phenotypic analysis of Mp*SPL4* knockdown line (*amiR*-Mp*SPL4-D*) generated using artificial miRNA. The mutant lines were grown together with wild-type, Tak2 plants in *in vitro* culture. The images were taken at 3-weeks-old and 2-months-old. All the images were taken in Leica M60 microscope. The scale below each image depicts 5 mm. The outline of *amiR*-Mp*SPL4-D* plant generated from a single gemma is outlined in red.

Only after 4 months of far-red irradiation, *amiR*-Mp*SPL4-D* mutant plants started to develop archegoniophores. The *amiR*-Mp*SPL4-D* archegoniophores were very difficult to observe (marked by arrows in Fig. 3.53) at smaller magnification despite the visible archegoniophores in wild-type Tak-2 plants, at same magnifications. At higher magnifications, the archegoniophores of *amiR*-Mp*SPL4-D* looked distorted with smaller stalks and immature receptacles (Fig. 3.53). Although for now the *amiR*-Mp*SPL4-D* plants were not proved to be knock down line for Mp*SPL4* gene, however based on the phenotype observation this line is a good candidate for further molecular and phenotypic analysis which will be performed in nearest future.
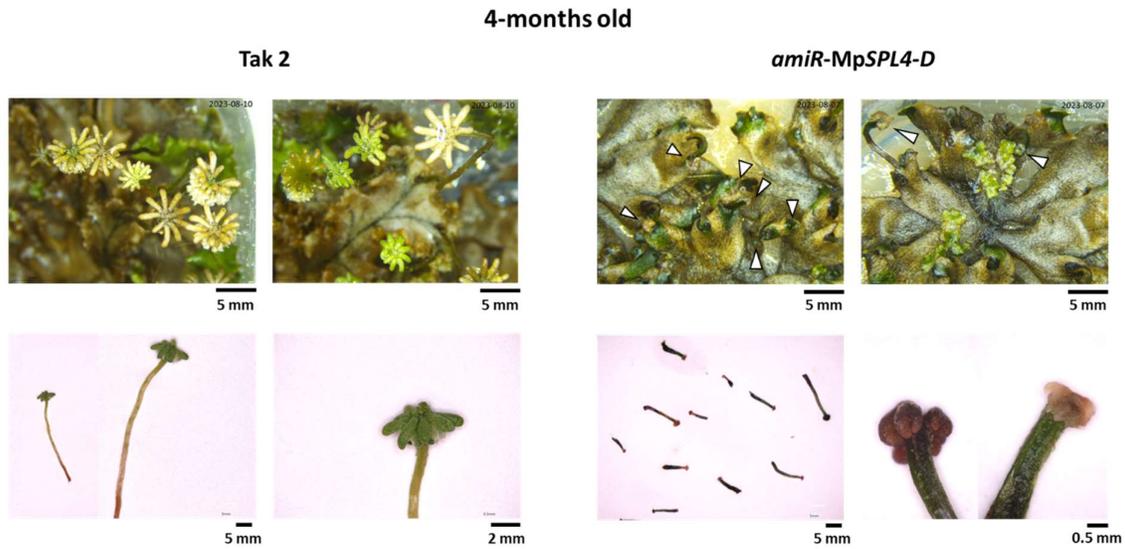
Figure 3.53: Phenotypic analysis of Mp*SPL4* knockdown line (*amiR*-Mp*SPL4-D*) generated using artificial miRNA. The mutant lines were grown together with wild-type, Tak2 plants in *in vitro* culture. The images were taken at 4-months-old. The position of archegoniophores in *amiR*-Mp*SPL4-D* are marked by arrows. The images on top were taken in Leica M60 microscope and the magnified images below were taken in VHX-700 Keyence Digital microscope.

3.2.7 Generation of mutant plants overexpressing MpSPL3.1 and MpSPL3.2 proteins

To examine the effect of MpSPL3 protein overexpression on the Marchantia development, the coding sequences of both Mp*SPL3* mRNA isoforms were constitutively expressed under CaMV35S or Mp*EF1* gene promoter. Moreover, the transcript sequences were cloned in frame with 3x FLAG tag to enable further protein detection. Many transgenic lines were obtained after transformation with Mp*SPL3*_cds1 (Mp*SPL3.1*) and Mp*SPL3*_cds2 (Mp*SPL3.2*). Six transgenic lines for Mp*SPL3*_cds1 overexpression, three under CaMV35S and three under Mp*EF1* promoter were chosen as representatives for measuring the expression levels of Mp*SPL3.1* transcript. Additionally, Mp*SPL3.2* transcript level was also investigated. The RT-qPCR analysis revealed significantly higher level of Mp*SPL3.1* transcripts in all these transgenic lines in comparison to wild-type plants. Moreover, Mp*SPL3.2* transcript levels were also elevated in the transgenic lines under investigation (Fig. 3.54).
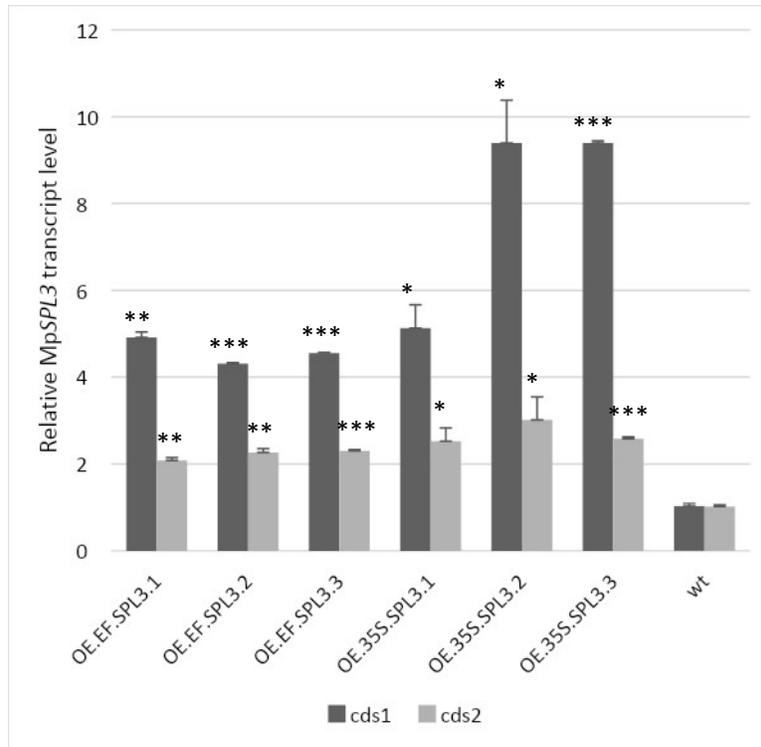
Figure 3.54: RT-qPCR for detection of Mp*SPL3.1* and Mp*SPL3.2* transcript levels in Mp*SPL3.1* overexpression lines as compared to wild-type mRNA. The analysis was performed with two biological replicates in two technical replicates each. p-values to determine the statistical significance were calculated using paired t-test ($p \leq 0.05$ = *, $p \leq 0.01$ = ** and $p \leq 0.001$ = ***.

For phenotypic analysis, three lines were chosen as representatives, one male (*OE.35S.MpSPL3.1_1*) and two female (*OE.35S.MpSPL3.1_2* and *OE.EF.MpSPL3.1_2*). The phenotypic analysis of these three lines revealed no significant differences of growth as compared to wild-type plants at 21-days and 30-days-old stage (Fig. 3.55).
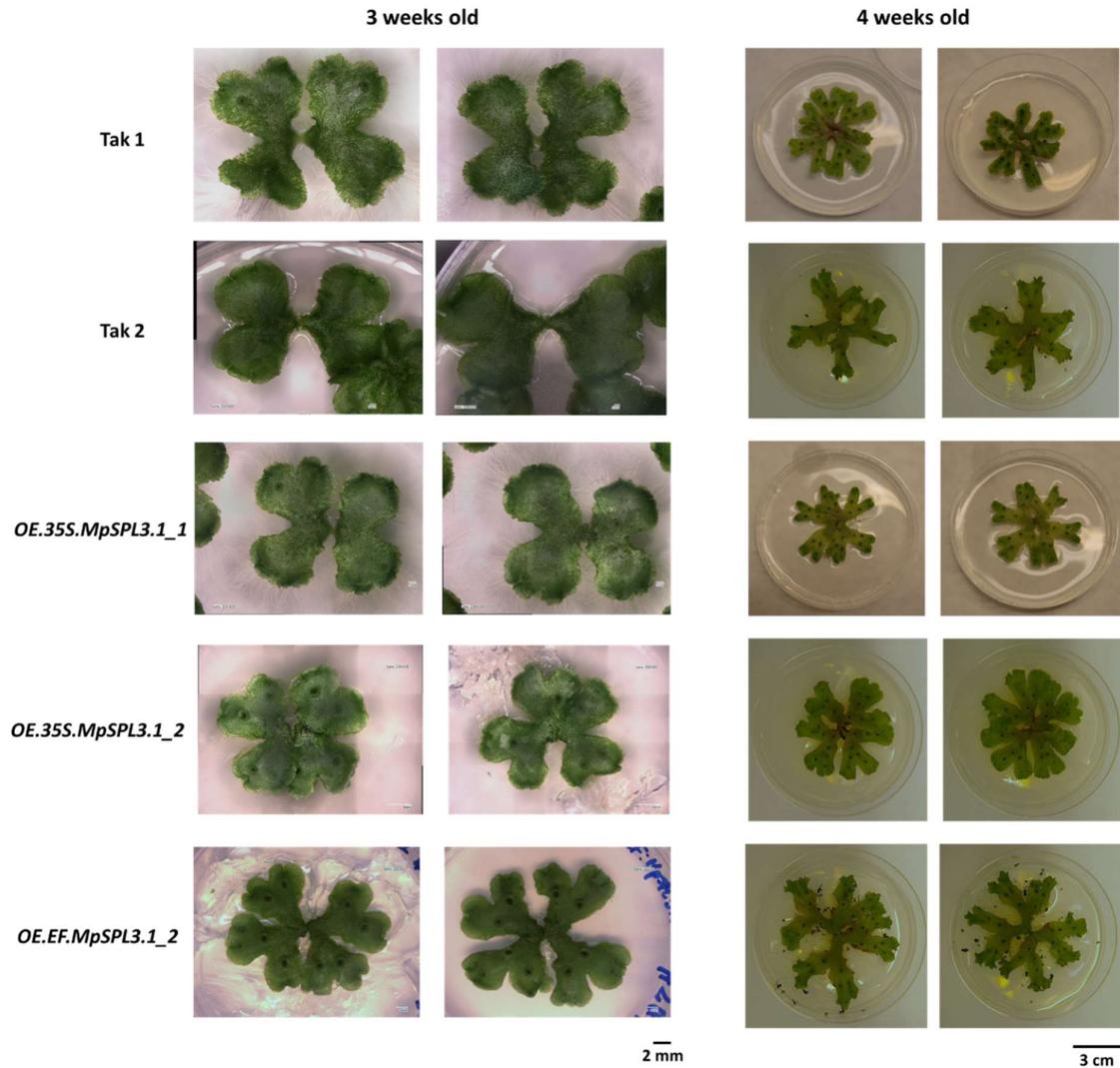
Figure 3.55: Phenotypic analysis of plants overexpressing Mp*SPL3.1* coding sequence. The transgenic plants were grown together with wild-type, Tak1 and Tak2 plants in *in-vitro* culture. The images of plants at 21-days-old were taken by VHX-700 Keyence Digital microscope and 30-days-old were taken by camera.

Moreover, to examine the level of MpSPL3.1 protein isoform, vegetative thalli of all six transgenic lines of Mp*SPL3.1* were tested using western blot approach. The calculated molecular weight of MpSPL3.1 protein isoform is 132 kDa. Hence, membrane above 70 kDa was only used for incubation with FLAG-antibody. As a negative control of FLAG antibody, *A. thaliana* Col-0 plant extract (Fig. 3.56, lane 8) and as a positive control, *A. thaliana* U1-70K-GFP-FLAG (transgenic line prepared by Dr Łukasz Szewc; unpublished data; Fig. 3.56, lane 9), with an expected size of ~100 kDa were used. In this experiment, a signal at ~130 kDa was observed from the protein extract isolated from all six overexpression mutant plant (Fig. 3.56b, lanes 2 – 7) but at the same time, it was also observed from the protein extract isolated from wild-type plants (Fig. 3.56, lane 1). Because of unspecific binding of FLAG antibody in wild-

type plants protein extract, the western experiment needs to be further tested and optimized to check whether FLAG antibodies can be successfully applied for Marchantia studies.
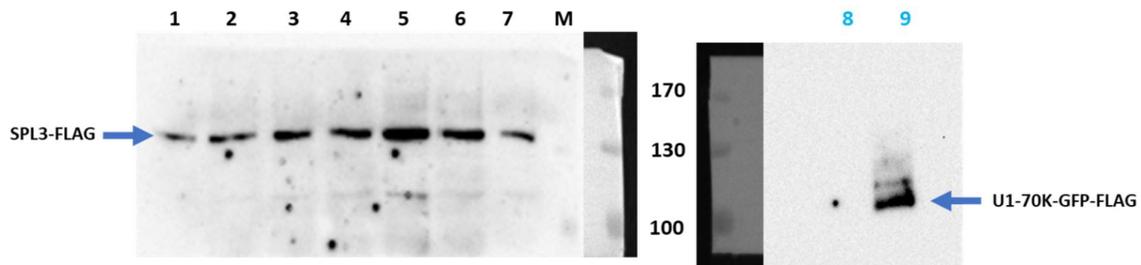
Figure 3.56: Western blot analysis of MpSPL3.1 protein isoform (molecular weights of 132 kDa). Protein extracts were extracted from 3-week old vegetative thalli from 1: Wild-type, 2-4: OE.EF.Mp*SPL3.1_1-3*, 5-7: OE.35S.Mp*SPL3.1_1-3*, 8: *A. thaliana* Col-0 (negative control for antibody) and 9: *A. thaliana* U1-70K-GFP-FLAG (~100kDa; positive control for antibody; unpublished data). Membrane above 70 kDa was incubated with FLAG-antibody.

Simultaneously, four transgenic lines, three male (#2, #4 and #5) and one female (#1) have been chosen as representatives for phenotypic analysis of overexpression of MpSPL3.2 protein isoform. The phenotypic analysis did not reveal any significant changes as compared to wild-type at 21-days and 30-days-old stages (Fig. 3.57). However, in-detail analysis will be conducted in near future at both vegetative and reproductive stages for transgenic lines overexpressing Mp*SPL3* coding sequences. Furthermore, the transgenic lines overexpressing Mp*SPL3.2* coding sequence also require expression analysis to measure the transcript and protein levels by RT-qPCR and western blotting, respectively.
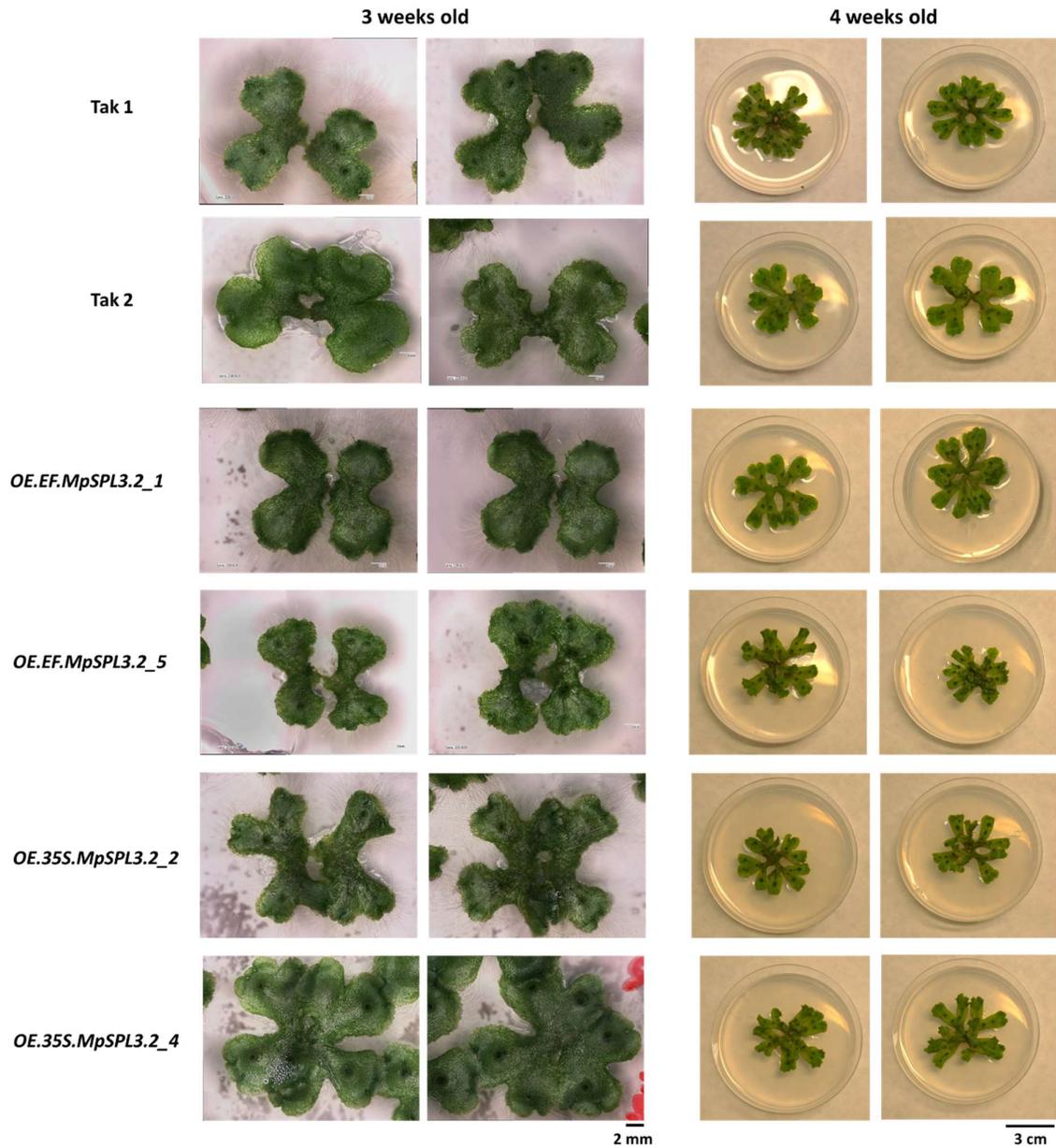
Figure 3.57: Phenotypic analysis of plants overexpressing Mp*SPL3.2* coding sequence. The transgenic plants were growth together with wild-type, Tak1 and Tak2 plants in *in-vitro* culture. The images of plants at 21-days-old were taken by VHX-700 Keyence Digital microscope and 30-days-old were taken by camera.

3.2.8 Generation of lines over-expressing Mp*SPL4* gene

To examine the effect of overexpression of MpSPL4 protein on the Marchantia development, the coding sequence of Mp*SPL4* gene was constitutively expressed under CaMV35S promoter. Out of these lines, five lines were chosen to measure the expression levels of Mp*SPL4* transcript. In all these transgenic lines, Mp*SPL4* transcript levels were significantly upregulated (Fig. 3.58).
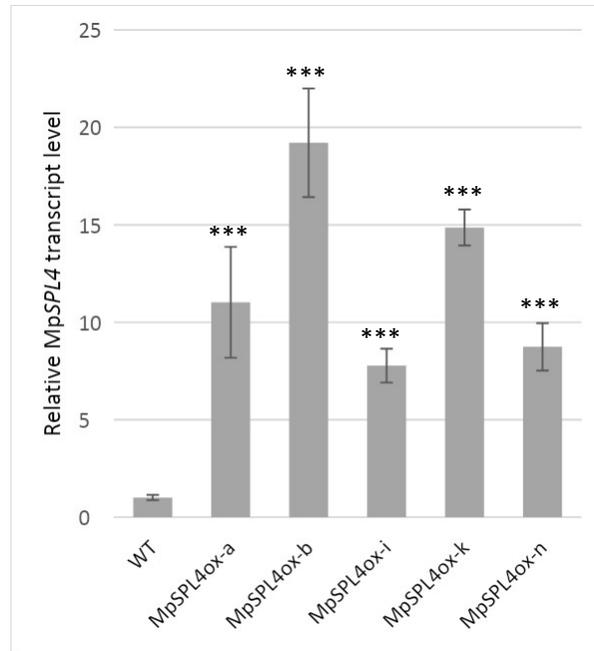
Figure 3.58: RT-qPCR for detection of Mp*SPL4* transcript levels in MpSPL4 overexpression lines as compared to wild-type mRNA. The analysis was performed with two biological replicate in two technical replicates each. p-values to determine the statistical significance were calculated using paired t-test ($p \leq 0.05$ = *, $p \leq 0.01$ = ** and $p \leq 0.001$ = ***).

Henceforth, all five transgenic lines were chosen for phenotypic analysis at 3-week-old stages. The five transgenic lines with significant higher levels of Mp*SPL4* transcripts showed similar phenotype. Plants overexpressing MpSPL4 protein have reduced thallus size with fewer and enlarged gemma cups as compared to wild-type plants (Fig. 3.59).
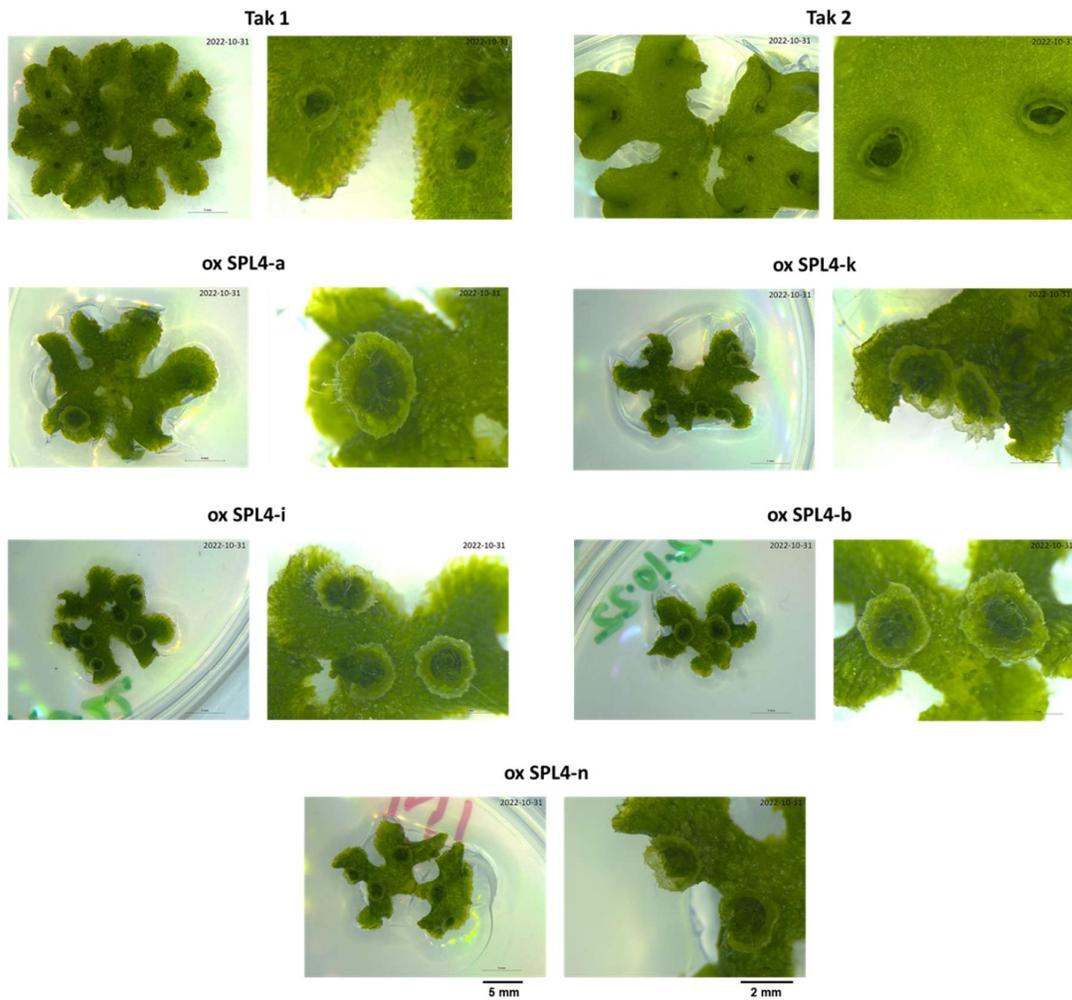
Figure 3.59: Phenotypic analysis of plants overexpressing Mp*SPL4* coding sequence. The transgenic plants were grown together with wild-type, Tak1 and Tak2 plants in *in-vitro* culture. The images of plants at 3-weeks-old were taken by VHX-700 Keyence Digital microscope.

4. DISCUSSION

4.1 Hornworts and liverworts possess a minimal set of *SPL* genes in comparison to other land plants

The first goal of the presented PhD dissertation was to characterize the phylogenetic relationships of *SPL* TF genes from representatives of all bryophyte lineages with angiosperm, *A. thaliana*. For liverwort *M. polymorpha*, moss *P. patens*, and *A. thaliana* we collected the annotated set of *SPL* protein encoding genes for each species. In the case of hornworts, there were no genes annotated as *SPL* at the time we started our analysis. Therefore, using available genomic resources for two hornwort species, *A. agrestis* and *A. punctatus,* we have identified four *SPL* genes for each hornwort what is similar to the number of *SPL* genes described in the liverwort *M. polymorpha*. Our results are in agreement with currently published data by Streubel et al. who also identified four *SPL* genes in *A. agrestis* and several liverwort species (Streubel *et al.*, 2023). Interestingly, the number of *SPL* gene family members in liverworts and hornworts stands in contrast to the data published for other land plants. This is because the previous studies have demonstrated that moss *P. patens* possesses 13 members in the *SPL* family, and further expansion of the *SPL* family has been observed in many angiosperm genomes, with 23 members in *Medicago truncatula*, 31 in *Zea mays*, and a staggering number 56 in *Triticum aestivum* (Zhang *et al.*, 2016; Wang *et al.*, 2019; Li *et al.*, 2022). These variations demonstrate substantial differences in the number of *SPL* genes among land plant genomes. Moreover, the identification of only four SPL members in liverworts and hornworts, in comparison to other land plants, highlights that the evolutionary process of all land plant *SPL* genes involved several rounds of gene duplication and subsequent speciation events. These events gave rise to paralog genes originating from only four *SPL* representatives that are present in the basal lineages of bryophytes. This hypothesis is supported by the phylogenetic tree of SPL proteins from bryophytes and Arabidopsis, where each of the four phylogenetic groups (Groups 1-4) recognizes only a single representative from Anthoceros and Marchantia (Fig. 3.2). Furthermore, the SPL proteins from bryophytes within all identified phylogenetic groups are grouped on nearby branches, indicating their close evolutionary relationships.

The analysis of *SPL* genes structure from selected land plant species revealed a similar exon-intron organization between *SPL* genes from bryophytes and *A. thaliana* within the same phylogenetic group, with the exception of both Anthoceros species *SPL* genes from Groups 3 and 4. Notably, these hornworts *SPL* genes exhibited one or two very short introns, in contrast to the more complex structures observed in *SPL* genes from representatives of liverwort, moss

and dicot (Fig. 3.3). Furthermore, the average intron length of hornwort *SPL*s is the shortest when compared to *SPL* genes from the other land plants employed in our analysis. The specific intron length and number within *SPL* genes of both Anthoceros species might be correlated with the specificity of these hornworts' genome organisation. In both genomes, high gene density is observed, which is achieved through the existence of numerous intron-less genes. Moreover, hornworts *SPL* gene structures reflect another characteristic feature of hornwort genomes, the occurrence of three to four exons per gene on average (Szövényi, 2016; Li *et al.*, 2020*b*). Conversely, the liverwort *SPL* genes exhibit the highest average intron length, which aligns with the average intron length calculated from genome analysis in *M. polymorpha* (Bowman et al., 2017b). To summarize, the available genomic evidence reveals the conservation of exon-intron structures within SPL clades, with only minor variations in the number of exons and introns, primarily reported in hornworts. This conservation is even evident between distantly related species, such as Marchantia and Arabidopsis. Nonetheless, exceptions to this rule of *SPL* gene structure conservation, as seen in Anthoceros, can be attributed to differences between genome structure and composition.

The four phylogenetic SPL protein families identified in our investigation were next examined in terms of protein motifs conservation. Apart from *A. agrestis* and *A. punctatus* SPL3 proteins, the SPL proteins in Groups 2, 3, and 4, revealed a similar pattern of conserved motifs between bryophytes and Arabidopsis (Fig. 3.4). However, in Group 1, there were explicit differences in protein motif set between SPLs amongst the analysed plant species. In general, only the SBP domain was identified as a common motif in all SPL proteins, irrespective of land plant lineage. Furthermore, we found a remarkably high degree of amino acid conservation within the SBP domain, particularly in two zinc-finger type structures and the NLS signal (Fig. 3.5 and 3.6A-D). Structural studies involving SPL proteins have demonstrated that all conserved basic amino acids from Zn-1, Zn-2, and NLS signal form a positively charged surface which is involved in binding the negatively charged DNA (Yamasaki *et al.*, 2004). Although SPL proteins have been also identified in green algae, however, the SBP domain from *C. reinhardtii* proteins showed a lower degree of conservation in the number of basic amino acids, especially within the first zinc-finger like structure (Fig. 3.6E). In fact, studies have shown that the Chlamydomonas CRR1 protein exhibited significantly lower affinity to the Arabidopsis-derived 15 bp *AP1* gene promoter fragment and the Chlamydomonas-derived copper response element (CuRE) as compared to Arabidopsis AtSPL1, AtSPL3, AtSPL8, and moss PpSPL1 proteins (Birkenbihl *et al.*, 2005). The poor efficiency of SBP domain of CRR1 protein in

binding to DNA might be a result of the number of reduced basic amino acids in its SBP domain as compared to land plant representatives. Among the conserved Arg/Lys residues, those located in the SBP domain's N-terminus (Lys14, Arg/Lys18, Arg19, Lys/Arg21) are considered to be the candidate residues that dictate sequence specificity through directly recognizing DNA bases (Yamasaki *et al.*, 2004). Interestingly, all these conserved amino acid residues are present in the SBP domains of all three bryophyte species, showing that these sites were established relatively early in the evolution of land plants.

In addition to the SBP domain, we discovered additional motifs in the analysed SPL proteins, which exhibited high conservation across evolutionary distant plant species, particularly in Groups 3 and 4 (Fig. 3.4). Although the precise function of these motifs remains unknown, their significant evolutionary conservation suggests that they may serve as structural units crucial for the proper functioning of encoded SPL proteins. Based on the analysed expression profiles, all *SPL* genes in Groups 3 and 4 displayed constitutive expression across various organs and developmental stages in both bryophytes and Arabidopsis (Fig. 3.8). Consequently, these *SPL* genes may play vital roles in regulating fundamental cellular processes in all land plants. Since the SBP domain is a common feature of all SPL proteins and is essential for precise recognition and binding to *cis*-elements in gene promoters to regulate their expression, there must be some specific characteristics of each SPL protein to direct each of them to their specific target gene promoters. Therefore, the additional conservation observed in the C-terminus region of these proteins may indicate that these conserved motifs might be crucial for orchestrating the proper expression profile by Group 3 and 4 SPL proteins in diverse tissues and organs throughout the life cycle of a plant. This orchestration could potentially involve the interaction of these SPL proteins with other proteins via conserved motifs localized in their C-terminals, such as ankyrin repeats (members belonging to Group 3, with exception of both Anthoceros SPL3 in Fig. 3.4), which are known to be important for protein-protein interactions. However, the precise significance of these conserved motifs remains unknown and needs to be further investigated, particularly through cross-species studies. Exploring the functions and interaction of these motifs can provide valuable insights into the regulatory mechanisms governed by SPL proteins across different plant species.

The phylogenetic Group 1 SPL proteins displayed most significant variability in motif composition amongst the investigated SPL proteins, accompanied by observable variations in protein lengths (Fig. 3.4 and Table 3.4). It appears that during the course of evolution, the *P. patens* Group 1 SPLs underwent independent diversification from other SPL family members

in the moss lineage, resulting in their functional specialization. Research on *P. patens* has shown that the Pp*SPL1* and Pp*SPL4* genes influence gametophore development, protonema branching, and spore germination (Riese *et al.*, 2008). Interestingly, although the PpSPL7 protein shares the same set of conserved protein motifs as PpSPL1 and PpSPL4 (Fig. 3.4), its expression profile suggests its involvement in archegonia and sporophyte development (Fig. 3.8B). Thus, the presence of similar motif compositions within proteins from the same family does not necessarily imply to their involvement in the same developmental processes. Whether the unique motifs identified in the *P. patens* Group 1 SPL proteins are required for fulfilling their specific roles still remains a topic for future investigations. Further studies are needed to determine the functional significance of these unique motifs in the context of their specialized developmental processes.

The composition of the promoter regions plays a pivotal role in the regulatory control of gene expression in different tissues or in response to different stimuli. In the promoter regions of *SPL* genes from studied bryophytes and Arabidopsis, numerous *cis*-elements were identified, primarily associated with growth and development, light, hormone, and stress responses (Fig. 3.7). This evidence suggests that in each of the investigated plant species, the *SPL* family is intricately regulated at the transcriptional level, enabling them to be expressed and respond to diverse developmental and environmental signals. Interestingly, no similar distribution of *cis*-elements was detected in the promoter regions of *SPL* genes within the same phylogenetic group. This observation implies that alterations in the *cis*-regulatory elements took place during the evolution of land plant *SPL* genes.

To further investigate the expression patterns of *SPL* genes in selected plant species, we analysed their expression profiles across various organs and developmental stages (Fig. 3.8). Heat maps of expression profiles revealed distinct patterns: *SPL* genes from Groups 3 and 4 exhibit constitutive expression in both bryophytes and Arabidopsis. In contrast, *SPL* genes from Groups 1 and 2 displayed developmentally specific expression or higher expression in specific organs/tissues. This differential expression pattern aligns with the posttranscriptional regulation exerted by miR156 or miR529 family members on all Group 2 *SPL* genes and three Arabidopsis *SPL*s in Group 1 (Fig. 3.2). miR156 is conserved across all land plant lineages, whereas miR529 is predominantly found in bryophytes and monocots. Functional and evolutionary analysis suggest that miR529 has been lost in certain taxonomic groups, including core eudicots like Arabidopsis (Xie *et al.*, 2021; Li *et al.*, 2023*b*). Therefore, in Arabidopsis miR156 acts as a master regulator of juvenile plant development in Arabidopsis and represses

the transition from vegetative to reproductive phase (Chen *et al.*, 2010; Sunkar, 2012; Zheng *et al.*, 2019). However, both miRNAs are found in rice where they fine-tune the expression of *SPL* genes at different stages of development. First, miR156 represses *SPL* genes during the vegetative stage to govern branch expansion, and then miR529 suppresses *SPL* genes during the reproductive stage to appropriately define panicle architecture and size (Wang *et al.*, 2015*b*; Li *et al.*, 2023*b*). Also in moss *P. patens,* both miRNAs are present, but expressed in different developmental stages, with miR156 primarily expressed in protonema and miR529 mainly expressed in gametophores with mature sporophytes (Xie *et al.*, 2021). Another example of bryophyte expressing both, miR156 and miR529 is liverwort *Pellia endiviifolia*, however their mode of action remains unexplored (Alaba *et al.*, 2015; Pietrykowska *et al.*, 2022). In contrast, miR156 is absent in *M. polymorpha*, and only miR529c governs the control of reproductive transition by inhibiting Mp*SPL2* gene expression during vegetative growth, thus preventing the development of reproductive branches and organs (Tsuzuki *et al.*, 2019). Therefore, miR529-*SPL* module appears to control the vegetative-to-reproductive phase shift in liverwort development, analogous to miR156/529-*SPL* module in angiosperms. While we found no direct evidence of miR156 or miR529 presence in the genomes of *A. agrestis* and *A. punctatus*, our research demonstrated the presence of conserved miR156/529-targeted elements in both Aa*SPL2* and Ap*SPL2* genes. Therefore, it is quite plausible that at least one of these miRNAs is present in the examined Anthoceros species, especially considering that miR156 has been identified in another Anthoceros species, *A. angustus* (Zhang et al., 2020).

As mentioned above, in *A. thaliana* the expression of *SPL* genes from Group 1 and Group 2 is regulated by evolutionary conserved miR156. Interestingly, in Marchantia Mp*SPL1* gene from Group 1 is regulated by liverwort-specific miRNA, Mpo-MR-13 (Tsuzuki *et al.*, 2016; Streubel *et al.*, 2023). According to transcriptomic studies, this Mpo-MR-13-Mp*SPL1* module may be important in directing the transition from the vegetative to reproductive life cycle. Mp*SPL1* exhibits a characteristic expression pattern, with a distinct expression peak in gametangiophores, accompanied by simultaneous down-regulation of Mpo-MR-13 precursors during this developmental stage (Flores-Sandoval *et al.*, 2018). However, recent functional studies, have revealed an additional role for this Mpo-MR-13-Mp*SPL1* module in the regulation of meristem dormancy, with PIF-mediated phytochrome signalling exerting superior control over this module. Under shade-imitated conditions, *Mpo-MR-13* expression is repressed by Mp*PIF* and thus, releasing Mp*SPL1* gene transcript from Mpo-MR-13 regulation, which further results in promotion of meristem dormancy by repressing meristem activity

(Streubel *et al.*, 2023). A similar dependence on PIF-mediated regulation was proposed in Arabidopsis, where PIF-mediated repression of several *MIR156* genes expression releases the miR156-taregeted *SPL*s from posttranscriptional control, allowing them to participate in the shade avoidance mechanism (Xie *et al.*, 2017). Based on these observations, Streubel et al proposed that the miRNA-*SPL* regulatory module involved in meristem dormancy and essential for the shade avoidance mechanism, evolved separately in the liverwort and angiosperm lineages. However, it is important to note that the authors focus was mainly on vegetative thallus development, and no data related to sexual reproduction were provided (Streubel *et al.*, 2023). As a result, it cannot be ruled out that the Mpo-MR-13-Mp*SPL1* module will serve a dual role during the Marchantia life cycle, contributing to both vegetative and reproductive processes.

4.2 The role of Mp*SPL3* and Mp*SPL4* transcription factors in the development of liverwort *Marchantia polymorpha*

Following the phylogenetic analysis of *SPL* genes from bryophytes and angiosperm representatives, we proceeded with the functional characterization of two *SPL* genes, Mp*SPL3* and Mp*SPL4*, from liverworts model species, *M. polymorpha*. Transcripts of these genes are not targeted by miRNAs in comparison to Mp*SPL1* and Mp*SPL2* gene transcripts (Fig. 3.2).

Our first experiment focused on clarifying whether the 5' region of Mp*SPL3* and Mp*SPL4* gene loci are correctly annotated as the 5' end of Mp*SPL3* and Mp*SPL4* genes overlapped with neighbouring genes (according to MarpolBase, v3.1) (Fig. 3.9 and 3.10). In the case of 5' RACE for Mp*SPL3* gene transcript, the obtained products were specific and aligned with Mp*SPL3* gene locus (Fig. 3.12). Conducted RACE analysis for the 5' and 3' cDNA ends of Mp*SPL3* neighbouring gene resulted in unspecific products. This outcome implies that either the neighbouring gene was incorrectly annotated in the initial versions of Marchantia genome database, or it is expressed at very low levels in the tissue under analysis. In the latest versions of Marchantia genome database (MpTak_v6.1r2), the neighbouring gene of Mp*SPL3* has been removed (Fig. 3.13), thereby supporting our initial hypothesis.

In the case of 5'RACE for Mp*SPL4* gene transcript*,* all obtained products were specific and aligned with Mp*SPL4* gene locus (Fig. 3.17). According to obtained results, different transcription start site can be selected by transcriptional machinery at Mp*SPL4* locus which results in different length of its 5'UTR. The 5'/3'RACE analysis of Mp*SPL4* neighbouring gene revealed differences in the length of the 5' UTR and 3' UTR in comparison to the MarpolBase annotation, v3.1, with 5'UTR being 92 bp longer and 3'UTR being 370 bp shorter

to the annotated gene (Fig. 3.17). In summary, based on RACE experiments we can assume that Mp*SPL4* neighbouring gene represents an independent transcription unit. Alternatively, the transcript obtained from *Mapoly0008s0030.1* locus might be prematurely polyadenylated part of the transcribed 5'UTR of Mp*SPL4* gene, especially that both genes are annotated on the same DNA strand. Interestingly, in the latest version of the Marchantia genome database (MpTak_v6.1r2), the neighbouring gene of Mp*SPL4* has been removed from annotation (Fig. 3.18). To clarify these two loci annotation, further experiments are required, e.g. RNA-sequencing.

From the transcriptomic dataset analysis (Fig. 3.8), we have learned that Mp*SPL3* and Mp*SPL4* genes are constantly expressed in different types of Marchantia tissues. By using RT-qPCR approach, we have confirmed that both investigated genes are expressed ubiquitously during the vegetative and reproductive phase of Marchantia life cycle (Fig. 3.20 and 3.22). Also, histochemical study of transgenic plants expressing GUS reporter protein either under control of Mp*SPL3* or Mp*SPL4* promoter revealed that promoters of both tested genes are responsible for omnipresent GUS protein expression (Fig. 3.23 and 3.24). Interestingly, similar spatiotemporal GUS activity pattern was observed in transgenic plants expressing *GUS* reporter gene under the regulation of Mp*IAA* promoter. Notably, $_{pro}$Mp*IAA*:*GUS* plants displayed high GUS activity in all analysed tissues, including thallus, gemmae and gemmae cups, rhizoids, as well as reproductive tissues. Specifically, in reproductive tissues, the GUS activity was predominantly observed in the receptacles of antheridia and the digitate rays of archegoniophores, rather than their stalks (Kato *et al.*, 2015*a*). Taking together these information, interesting question can be addressed, whether MpSPL3 and MpSPL4 proteins could be potentially implicated in co-operative action within auxin signaling pathway. Data showing the expression of MpSPL3, MpSPL4 and MpIAA proteins under their native promoters combined with GUS expression could provide one of evidence to support this hypothesis. Additionally, similar approach but with fluorescent reporter protein could give insight on the putative protein interaction at a cellular level. Therefore, further studies are necessary to delve into this putative interplay between MpSPL3 and MpSPL4 proteins with auxin signalling pathway regulators in Marchantia.

To gain insights into the functions of Mp*SPL3* and Mp*SPL4* genes, at first, we applied CRISPR/Cas9 approach to generate knockout mutants. In our efforts to enhance the chances of gene editing, we designed five different guide RNAs in different locations for both genes under study (Fig. 3.25 and 3.33). For Mp*SPL3* gene locus, we succeeded in obtaining two mutant

lines at G2 generations but only from gRNA 1 (Table 3.10 and Fig. 3.26)), designed 11 bp downstream the start codon. Both these mutant lines, Mp*spl3-1.1* and Mp*spl3-1.3* were male. Therefore, we speculate that perhaps for female, it is not possible to obtain complete knock-out of Mp*SPL3* gene. To clarify this issue, it would be interesting to back-cross male Mp*spl3-1.1* and Mp*spl3-1.3* mutant plants with WT plants and analyse next generation sporelings for mutant female individuals. Out of these two mutant lines obtained, Mp*spl3-1.3* mutant line showed more growth abnormalities with a significant callus-like phenotype and severely reduced thallus area than Mp*spl3-1.1*, although both mutants displayed overall growth retardation when compared to wild-type plants (Fig. 3.29). The differences in phenotypes observed might be the result of different mutations (deletion in Mp*spl3-1.3* and insertion and substitution in Mp*spl3-1.1* mutant lines, respectively). In both Mp*spl3* mutant lines mutations were introduced at the very beginning of CDS sequence changing the reading frame and introducing premature stop codons (Fig. 3.26). In general, incorrectly transcribed mRNAs with premature termination codon are eliminated by nonsense-mediated RNA decay (NMD) process, thus preventing the accumulation of potentially harmful truncated proteins (Chan *et al.*, 2007; Hug *et al.*, 2016). Although NMD is an evolutionarily conserved pathway across eukaryotes (Causier *et al.*, 2017), how it works and with what specificity in Marchantia is not known. Therefore, if not recognized and utilized by NMD, mutated transcripts from Mp*spl3* loci may result in the production of truncated version of proteins: 42 aa in length in the case of Mp*spl3-1.1* mutant line and 27 aa in the case of Mp*spl3-1.3* mutant line instead of 1158 aa native wild-type protein (Fig. 3.26). Furthermore, the comparative motif search between the wild-type MpSPL3 protein sequence (only first 42 aa) with protein sequences of both mutant lines showed that MpSPL3 shared more motifs with Mpspl3-1.1 than with Mpspl3-1.3 predicted proteins. Additionally, both mutant lines because of their respective mutations, produce an additional motif (Motif 2 in Fig. 4.1). The function of these five identified motifs is unknown. The presence of only 3 motifs in Mpspl3-1.3 predicted protein as compared to 5 motifs in Mpspl3-1.1 predicted protein might also be the cause of differences in phenotypes resulted from distinct mutations.

Figure 4.1: Analysis of protein motif composition within first 42 aa of MpSPL3 protein and within putative truncated alleles, Mpspl3-1.1 and Mpspl3-1.3. The schematic structure is obtained using MEME online tool (Bailey *et al.*, 2006). Each motif is represented in different colours.

Another explanation why these two Mp*SPL3* mutant lines differ in phenotypes might be due to off-target effects caused by gRNA1 in the genome of one of obtained lines. Therefore, we searched for potential off-target sites using online databases, CRISPRdirect (Naito *et al.*, 2015) and CRISPOR (Concordet and Haeussler, 2018). While our bioinformatic search with CRISPRdirect predicted no off-target sites up to 8 mismatches but with CRISPOR, we predicted three off-target sites. Two of these three sites were identified in intergenic region, while one was found to be in exonic region of a gene (Mp2g25480: annotated as encoding mitochondrial pyruvate carrier 2). Hence, these predicted sites will be tested in the future to observe any off-target mutations in the genome of both mutant lines of Mp*SPL3*. Additionally, whole genome sequencing of both mutant lines could give us a better picture of the putative off-target effects which might not be detected by these databases.

For Mp*SPL4* gene locus, many problems were encountered while working on obtaining CRISPR/Cas9 genome edited plants. As mentioned in results, we observed many mutant lines at G1 generation for Mp*SPL4* (Table 3.12). However, later, when the plants developed gemmae cups, we cultured gemmae within them to obtain G2 generation. The G2 generation at 14-day old stage were used for collecting plant material for genotyping again. After sequencing, we found most of them turned out to be wild-type, genotypically (Table 3.12). Hence, it seemed that there might be two populations of cells within Marchantia young thalli, one with mutation and one with wild-type genotype. As the plants grew older and developed gemmae, the population of cells with wild-type genotype outgrew the mutant population, resulting in wild-type plants in the next generation. It also appears that these wild-type cells divide and give rise to gemmae cups. Therefore, most of the mutant plants at their G1 generations might be mosaic. It has been suggested in the Arabidopsis studies that CRISPR/Cas9 system may not be active just immediately following Agrobacterium-mediated transformation and hence, giving rise to distinct cell lineages (Impens *et al.*, 2022). This might be the case in our studies for Mp*SPL4* gene knock-out by CRISPR/Cas9 in Marchantia, despite it being haploid-dominant plant.

As mentioned in the introduction and results chapters, the CRISPR/Cas9 approach by single gRNA could have disadvantage of having off-target effects which could be eliminated by multiple gRNA approach employed with Cas9-D10A nickase. Hence, we used this approach to knock-out Mp*SPL4* (Fig. 3.42). But again, we encountered similar problems in obtaining a stable mutant line at G2 generation. We observed many mosaic plants for Mp*SPL4* loci also with this approach (Table 3.13). Surprisingly, all these plants have different mutations in T1 and G1 generations (not observed with single gRNA approach), similar to what has been observed in Arabidopsis, where over half of the mutations observed in T2 generation were not present in T1 generation at all (Feng *et al.*, 2014). Additionally, (Sugano *et al.*, 2018) demonstrated that Marchantia CRISPR/Cas9 generated mutant plants, initially lacking target mutations in their T1 generations, eventually produce gemmae with *de novo* mutations, which seems to be the case for Mp*SPL4* knock-out by double gRNA CRISPR/Cas9 approach in Marchantia.

These findings clearly indicate the existence of mosaicism when generating plants using CRISPR/Cas9 technology in haploid-dominant plants such as Marchantia as is the case in diploid-dominant plants. Moreover, the studies by (Sugano *et al.*, 2018) introducing an improved CRISPR/Cas9 method by using Arabidopsis codon-optimized Cas9, also reported the presence of mosaicism in Marchantia. In this study, which utilizes the gateway-based technology, involving vector pMpGE010, which was also employed in our studies, resulted in 28% and 3% mosaics for *ARF1* (responsible for auxin biosynthesis) and *NOP1* (responsible for air-chamber formation) loci, respectively. Another gene Mp*MPK1* (mitogen-activated protein kinase genes) was also edited by CRISPR/Cas9 using the same vector, yielding a small number of mosaic mutants (~10%). Interestingly, even the transformation of Marchantia thalli with an empty vector resulted in over 20% mosaic plants. The authors in this study suggested that constant expression of CRISPR/Cas9 may also allow mutants without any alterations to potentially acquire mutations later at the target site in random cells during growth. In subsequent studies by (Sugano *et al.*, 2018) in which they developed a detailed protocol for transformation of transcription factors, which we followed in our analysis, also mentioned the presence of mosaics in the G1 generation. Despite the presence of mosaics in reported studies, no detailed investigations for the presence of mosaic individuals in haploid-dominant Marchantia have been documented. However, (Feng *et al.*, 2014) addressed the mechanism of transgenic chimera formation in diploid plants, suggesting that various cells with dissimilar genetic transformation events, discrepancies in the timing of mutation, along with transient

expression of transgenes could contribute to this phenomenon. Therefore, for essential genes involved in plant development, extra attention is required when employing the CRISPR/Cas9 approach in Marchantia studies. Consequently, functional analysis of essential genes may necessitate alternative strategies such as artificial miRNA expression system established by (Flores-Sandoval *et al.*, 2016) or conditional knockout mutants developed by (Nishihama *et al.*, 2016).

Despite many mutant lines obtained from knockout of Mp*SPL4* gene locus were mosaic, we finally obtained one mutant line at G1 generation from single gRNA CRISPR/Cas9 approach (Mp*spl4*) (Fig. 3.34) and two mutant lines at the G2 generation from double gRNA CRISPR/Cas approach (Mp*spl4_3* and Mp*spl4_54.2*) (Fig. 3.43). All three lines revealed severe growth abnormalities including reduction in thallus area, no production of gemma cups, delayed growth, and a prothallus-like phenotype (Fig. 3.35 and 3.44). In conclusion, we observed that the knockout of Mp*SPL4* gene locus caused more drastic effects on plant vegetative growth than observed after knocking out of Mp*SPL3* gene. From their phenotypic analysis, we can conclude that both *SPL* genes are crucial players for proper thallus growth and development in Marchantia.

Considering that complete knockout of Mp*SPL3* and Mp*SPL4* genes led to either severe defects in Marchantia, or the presence of unstable mutations in plants, we employed the artificial miRNA approach to aim at knocking down the expression of both studied genes. From knock-down of Mp*SPL3* gene locus by artificial miRNA, we obtained four plant lines with decrease in Mp*SPL3* transcript levels with simultaneous increase of the expression of *pri-amiRNA* (Fig. 3.48). Two transgenic knock-down lines, *amiR*-Mp*SPL3-4b* and *6*, with similar levels of *pri-amiRNA*, produced thallus with improper bifurcations as compared to wild-type plants. Another two transgenic lines, *amiR*-Mp*SPL3-2b* and *2c*, with significantly higher levels of *pri-amiRNA* than *amiR*-Mp*SPL3-4b* and *6* lines, showed delayed growth as compared to wild-type plants. In common, all these mutant lines did not produce any gametangiophores even after 2-months far-red irradiation. Our ongoing investigations are aimed at determining whether these plants eventually produce gametangiophores but with a delayed onset or if they fail to produce them altogether.

However, in the case of Mp*SPL4* amiR approach, we have obtained many lines at T1 generation which showed significantly high levels of *pri-amiRNA* with simultaneous reductions in Mp*SPL4* transcript levels, respectively (Fig. 3.50). At the same time, we have obtained several

lines growing more efficiently which gave rise G1 and G2 generations. For now, we have chosen one transgenic line, *amiR*-Mp*SPL4-D,* for further analysis which showed very high level of *pri-amiRNA,* although with no statistically significant reductions in Mp*SPL4* transcript levels (Fig. 3.52). However, currently only single biological replicate was investigated to characterize the expression patterns of *pri-amiRNA* levels and its target gene. We also can not rule out the possibility that all these changes observed might be an effect of miRNA off-target action. Further expression analysis including more biological replicates will give more comprehensive view on the Mp*SPL4* transcript level. Nonetheless, this mutant line was chosen also because of its characteristic phenotype. In comparison to female Tak-2 plants, *amiR*-Mp*SPL4-D* plants displayed delayed growth with thallus showing hyperbranching with few thallus layers, one over the another (Fig. 3.54). In addition, these mutant plants did not produce archegoniophores even after 2 months of far-red irradiation, while wild-type female plants cultured simultaneously, already developed mature archegoniophores. Interestingly, after 4 months of growth in gametangiophore inducing conditions, finally *amiR*-Mp*SPL4-D* plants produced archegoniophores, however with characteristically changed morphology, shorter stalk with immature receptacles.

Along with the approach to disable Mp*SPL3* and Mp*SPL4* gene's function, we generated plants overexpressing MpSPL3 and MpSPL4 proteins to investigate whether overaccumulation of these proteins will influence Marchantia growth and development. Although in each out of the tested overexpressing Mp*SPL3.1* mutant line, accumulation of Mp*SPL3* transcript isoform 1 was identified, preliminary observation of the plants with constitutive Mp*SPL3.1* overexpression showed no significant changes in phenotype during vegetative stage of growth as compared to wild-type plants. The overexpression of MpSPL3 isoform 1 protein not showing any noticeable phenotypic differences might be the result of degradation of excess proteins by endogenous proteases (Ito and Akiyama, 2005; Sakoh *et al.*, 2005; Zolkiewski, 2006). Since TFs are known to function interdependently instead of alone, hence, overexpressing one TF does not produce any phenotype. This has been observed in previous studies where overexpression of four TFs (PISTILLATA (PI), APETALA3 (AP3), SEPALLATA3 (SEP3) and AGAMOUS (AG)) involved in specifying the identity of floral organs cause the overexpression phenotype as opposed to overexpression of each TF alone (Pelaz *et al.*, 2000; Honma and Goto, 2001). Still, phenotypic analysis during the reproductive stage of growth needs to be performed for these plants in nearest future.

On the other hand, plants over-expressing Mp*SPL4* exhibited reduced thallus size with enlarged gemmae cups when compared to wild-type plants during vegetative growth. The production of large gemme cups in MpSPL4 overexpressing plants might show the involvement of MpSPL4 in molecular network responsible for gemma and gemma cup formation. This hypothesis might indirectly be also supported by the phenotype of Mp*SPL4* CRISPR/Cas9 mutant plants which did not produce gemma and gemma cups and that is one of the reasons we could not obtained G1/G2 isogenic lines. In Marchantia, several genes responsible for gemma and gemma cup production have been characterized functionally from which some are also implicated in auxin signaling (Kato *et al.*, 2020; Suzuki *et al.*, 2023). Mp*RRB*, a type-B response regulator and *GCAM1*, a R2R3-MYB transcription factor gene acting downstream of MpRRB-mediated cytokinin signaling have been shown to play role in gemmae cup development. Knock-out of Mp*gcam1* and Mp*rrb* resulted in the absence of gemma and gemmae cups. More interestingly, plants overexpressing Mp*GCAM1* in the background of Mp*rrb* knockout lines under inducible system produce undifferentiated cell clumps (Aki *et al.*, 2022), a phenotype similar to what we have observed in MpSPL4 artificial miRNA lines in T1 generations. Taken together, these findings suggest complex interactions and potential involvement of MpSPL4 in the pathways associated with gemmae cup formation in Marchantia like auxin and cytokinin signaling pathways. Further investigations are necessary to elucidate the precise involvement of MpSPL4 in mechanisms and regulatory networks responsible for gemma and gemma cup production.

Many of the phenotype changes observed in the obtained knockout, knockdown and over-expression plant lines of Mp*SPL3* and Mp*SPL4* genes bear some resemblance to Marchantia mutant lines associated with genes involved in auxin-biosynthesis pathway. For instance, downregulation or overexpression of genes involved in auxin biosynthesis, Mp*TAA,* Mp*IAA*, Mp*ARF1*, Mp*TIR* and Mp*YUC2* caused strong phenotypic effects on thallus phenotype (Flores-Sandoval *et al.*, 2015; Eklund *et al.*, 2015; Kato *et al.*, 2015*b*). These phenotypes include callus-like growth, hyperbranching, no gemmae cups production and not proper dorsiventral patterning (Fig. 4.2). These findings indicate a notable convergence between the phenotypes resulting from the manipulation of Mp*SPL3* and Mp*SPL4* genes and those associated with genes involved in auxin-mediated processes. This suggests a potential interplay and shared pathways between Mp*SPL3*, Mp*SPL4*, and the auxin signalling network. However, further investigations including RNA-seq experiments to investigate the changes in gene expression profiles in Mp*SPL3* and Mp*SPL4* knock-out/ knock-down and over-expressing transgenic plants are required to establish if these proposed interactions are true.
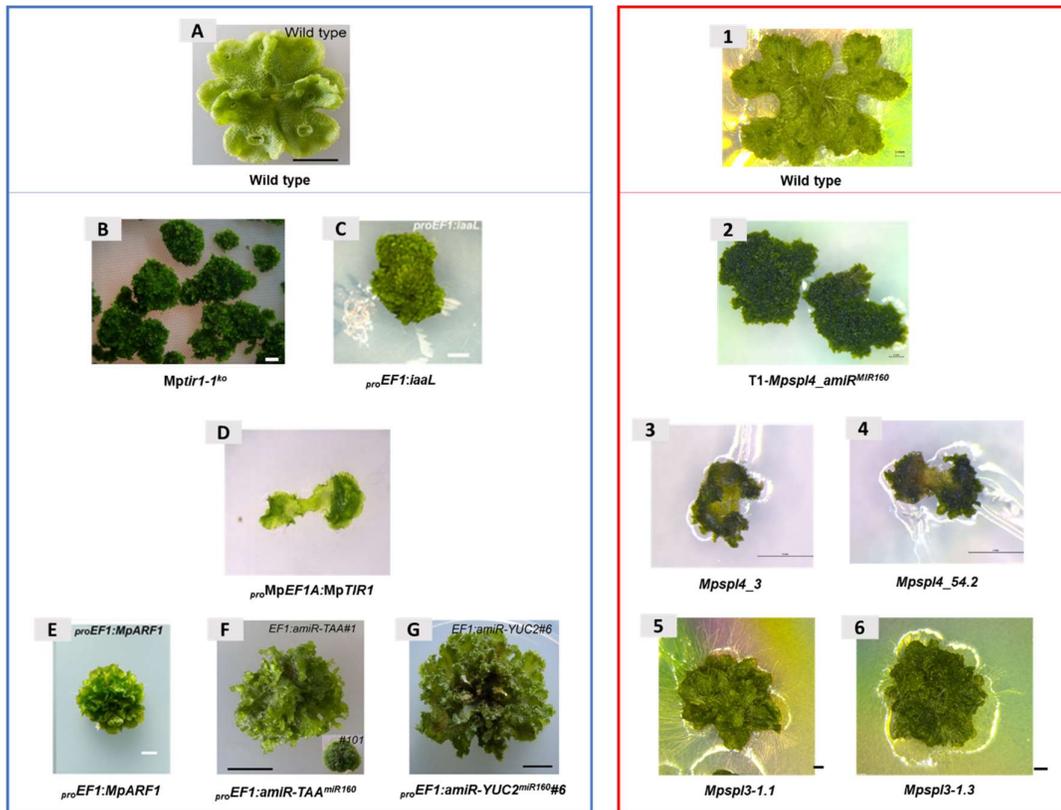
Figure 4.2: Comparison between phenotypes of mutant plants involved in auxin signaling pathway (blue frame) to Mp*SPL3* and Mp*SPL4* transgenic plants (red frame). In blue frame: (A) 3-week-old wild-type plants (Eklund *et al.*, 2015), (B) Mp*tir1-1ᵏᵒ* – Mp*TIR1* knock-out lines (90-days old) obtained by homologous recombination (Suzuki *et al.*, 2023), (C) *pro*EF1:iaaL – plants overexpressing a heterologous bacterial auxin conjugation enzyme iaaL (Flores-Sandoval *et al.*, 2015), (D) *pro*EF1:Mp*TIR1* – overexpressing Mp*TIR1* plants (in the presence of IAA, exogenous auxin) (Suzuki *et al.*, 2023), (E) *pro*EF1:Mp*ARF1* – overexpressing Mp*ARF1* plants (Flores-Sandoval *et al.*, 2015), (F) and (G) *pro*EF1:amiR-TAA*miR160* and *pro*EF1:amiR-YUC2*miR160* – plants with constitutive expression of artificial miRNA targeting Mp*TAA* and Mp*YUC2*, respectively (Eklund *et al.*, 2015). In red frame: (1) 3-week-old wild-type plant (Fig. 3.28), (2) T1-Mp*spl4_amiRᴹᴵᴿ¹⁶⁰* – Mp*SPL4* knock-down lines generated using artificial miRNA at T1 generation (Fig. 3.49), (3) and (4) 3-months-old Mp*spl4_3* and Mp*spl4_54.2* – Mp*SPL4* knock-out lines generated using double gRNA CRISPR/Cas9 approach Fig. 3.43), (5) and (6) Mp*spl3-1.1* and Mp*spl3-1.3* – 3-week-old Mp*SPL3* knock-out lines generated using single gRNA CRISPR/Cas9 approach (Fig. 3.28).

Our functional studies involve characterization of two *SPL* genes, Mp*SPL3* and Mp*SPL4* in the liverwort model species *M. polymorpha*. The presented results provide a significant insight into the basic functions of Mp*SPL3* and Mp*SPL4* which are crucial players in controlling the proper growth and development of both vegetative thallus and reproductive organs in Marchantia. Interestingly, the observed changes of phenotype in the Mp*SPL3* and Mp*SPL4* knockout or knockdown mutant plants revealed similar phenotypic effects as those observed in mutant plants of several genes involved in auxin biosynthesis pathways. These findings suggest potential interactions and shared pathways between MpSPL3, MpSPL4, and the auxin signaling network, highlighting the complexity of gene regulation and function in Marchantia.

Further investigations, including RNA-seq experiments or protein interaction studies, are needed to reveal what is the impact of the down regulation or overexpression of Mp*SPL3* and Mp*SPL4* genes on the transcriptomic and proteomic landscape of *M. polymorpha* vegetative and reproductive stage of life.

5. BIBLIOGRAPHY

Abdurakhmonov IY. 2022. *Model Organisms in Plant Genetics*. BoD – Books on Demand.

Aki SS, Morimoto T, Ohnishi T, Oda A, Kato H, Ishizaki K, Nishihama R, Kohchi T, Umeda M. 2022. R2R3-MYB transcription factor GEMMA CUP-ASSOCIATED MYB1 mediates the cytokinin signal to achieve proper organ development in Marchantia polymorpha. Scientific reports **12**, 21123.

Alaba S, Piszczalka P, Pietrykowska H, *et al.* 2015. The liverwort Pellia endiviifolia shares microtranscriptomic traits that are common to green algae and land plants. The New phytologist **206**, 352–367.

Alisha A, Szweykowska-Kulińska Z, Sierocka I. 2023. Comparative analysis of SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) gene family between bryophytes and seed plants. bioRxiv, 2023.02.27.530190.

Alvarez JP, Pekker I, Goldshmidt A, Blum E, Amsellem Z, Eshed Y. 2006. Endogenous and synthetic microRNAs stimulate simultaneous, efficient, and localized regulation of multiple targets in diverse species. The Plant cell **18**, 1134–1151.

Arazi T, Talmor-Neiman M, Stav R, Riese M, Huijser P, Baulcombe DC. 2005. Cloning and characterization of micro-RNAs from moss. The Plant journal: for cell and molecular biology **43**, 837–848.

Bailey TL, Williams N, Misleh C, Li WW. 2006. MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic acids research **34**, W369–73.

Bateman A, Coin L, Durbin R, *et al.* 2004. The Pfam protein families database. Nucleic acids research **32**, D138–D141.

Birkenbihl RP, Jach G, Saedler H, Huijser P. 2005. Functional dissection of the plant-specific SBP-domain: overlap of the DNA-binding and nuclear localization domains. Journal of molecular biology **352**, 585–596.

Blum M, Chang H-Y, Chuguransky S, *et al.* 2021. The InterPro protein families and domains database: 20 years on. Nucleic acids research **49**, D344–D354.

Bowman JL, Arteaga-Vazquez M, Berger F, *et al.* 2022. The renaissance and enlightenment of Marchantia as a model system. The Plant cell **34**, 3512–3542.

Bowman JL, Kohchi T, Yamato KT, *et al.* 2017*a*. Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. Cell **171**, 287–304.e15.

Bowman JL, Kohchi T, Yamato KT, *et al.* 2017*b*. Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. Cell **171**, 287–304.e15.

Cao J, Liu K, Song W, *et al.* 2021. Pleiotropic function of the SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE gene TaSPL14 in wheat plant architecture. Planta **253**, 44.

Cardon GH, Höhmann S, Nettesheim K, Saedler H, Huijser P. 1997. Functional analysis

of the Arabidopsis thaliana SBP-box gene SPL3: a novel gene involved in the floral transition. The Plant journal: for cell and molecular biology **12**, 367–377.

de Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N. 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic acids research **34**, W362–5.

Catarino B, Hetherington AJ, Emms DM, Kelly S, Dolan L. 2016. The Stepwise Increase in the Number of Transcription Factor Families in the Precambrian Predated the Diversification of Plants On Land. Molecular biology and evolution **33**, 2815–2819.

Causier B, Li Z, De Smet R, Lloyd JPB, Van de Peer Y, Davies B. 2017. Conservation of Nonsense-Mediated mRNA Decay Complex Components Throughout Eukaryotic Evolution. Scientific reports **7**, 16692.

Chan W-K, Huang L, Gudikote JP, Chang Y-F, Imam JS, MacLean JA 2nd, Wilkinson MF. 2007. An alternative branch of the nonsense-mediated decay pathway. The EMBO journal **26**, 1820–1830.

Chao L-M, Liu Y-Q, Chen D-Y, Xue X-Y, Mao Y-B, Chen X-Y. 2017. Arabidopsis Transcription Factors SPL1 and SPL12 Confer Plant Thermotolerance at Reproductive Stage. Molecular plant **10**, 735–748.

Chen X, Zhang Z, Liu D, Zhang K, Li A, Mao L. 2010. SQUAMOSA promoter-binding protein-like transcription factors: star players for plant growth and development. Journal of integrative plant biology **52**, 946–951.

Chiyoda S, Ishizaki K, Kataoka H, Yamato KT, Kohchi T. 2008. Direct transformation of the liverwort Marchantia polymorpha L. by particle bombardment using immature thalli developing from spores. Plant cell reports **27**, 1467–1473.

Chiyoda S, Linley PJ, Yamato KT, Fukuzawa H, Yokota A, Kohchi T. 2007. Simple and efficient plastid transformation system for the liverwort Marchantia polymorpha L. suspension-culture cells. Transgenic research **16**, 41–49.

Chiyoda S, Yamato KT, Kohchi T. 2014. Plastid transformation of sporelings and suspension-cultured cells from the liverwort Marchantia polymorpha L. Methods in molecular biology **1132**, 439–447.

Cho SH, Coruh C, Axtell MJ. 2012. miR156 and miR390 regulate tasiRNA accumulation and developmental timing in Physcomitrella patens. The Plant cell **24**, 4837–4849.

Clamp M, Cuff J, Searle SM, Barton GJ. 2004. The Jalview Java alignment editor. Bioinformatics **20**, 426–427.

Riese M, Höhmann S, Saedler H, Münster T, Huijser P. 2007. Comparative analysis of the SBP-box gene families in P. patens and seed plants. Gene **401**, 28–37.

Concordet J-P, Haeussler M. 2018. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic acids research **46**, W242–W245.

CoNekT : Co-expression Network Toolkit.

Cronk QCB. 2005. Plant eco-devo: the potential of poplar as a model organism. The New phytologist **166**, 39–48.

Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome research **14**, 1188–1190.

Cui L-G, Shan J-X, Shi M, Gao J-P, Lin H-X. 2014. The miR156-SPL9-DFR pathway coordinates the relationship between development and abiotic stress tolerance in plants. The Plant journal: for cell and molecular biology **80**, 1108–1117.

Cui N, Sun X, Sun M, Jia B, Duanmu H, Lv D, Duan X, Zhu Y. 2015. Overexpression of OsmiR156k leads to reduced tolerance to cold stress in rice (Oryza Sativa). Molecular breeding: new strategies in plant improvement **35**, 1–11.

Dai Z, Wang J, Yang X, Lu H, Miao X, Shi Z. 2018*a*. Modulation of plant architecture by the miR156f-OsSPL7-OsGH3.8 pathway in rice. Journal of experimental botany **69**, 5117–5130.

Dai X, Zhao PX. 2011. psRNATarget: a plant small RNA target analysis server. Nucleic acids research **39**, W155–9.

Dai X, Zhuang Z, Zhao PX. 2018*b*. psRNATarget: a plant small RNA target analysis server (2017 release). Nucleic acids research **46**, W49–W54.

Davies B, Schwarz-Sommer Z. 1994. Control of floral organ identity by homeotic MADS-box transcription factors. Results and problems in cell differentiation **20**, 235–258.

Dierschke T, Flores-Sandoval E, Rast-Somssich MI, Althoff F, Zachgo S, Bowman JL. 2021. Gamete expression of TALE class HD genes activates the diploid sporophyte program in. eLife **10**.

Edwards DI, Duckett JG, Richardson JB. 1995. *Hepatic characters in the earliest land plants*.

Eklund DM, Ishizaki K, Flores-Sandoval E, *et al.* 2015. Auxin Produced by the Indole-3-Pyruvic Acid Pathway Regulates Development and Gemmae Dormancy in the Liverwort Marchantia polymorpha. The Plant cell **27**, 1650–1669.

Feng Z, Mao Y, Xu N, *et al.* 2014. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **111**, 4632–4637.

Fernandez-Pozo N, Haas FB, Meyberg R, *et al.* 2020. *PEATmoss (Physcomitrella Expression Atlas Tool): a Unified Gene Expression Atlas for the Model Plant Physcomitrella Patens*.

Flores-Sandoval E, Dierschke T, Fisher TJ, Bowman JL. 2016. Efficient and Inducible Use of Artificial MicroRNAs in Marchantia polymorpha. Plant & cell physiology **57**, 281–290.

Flores-Sandoval E, Eklund DM, Bowman JL. 2015. A Simple Auxin Transcriptional Response System Regulates Multiple Morphogenetic Processes in the Liverwort Marchantia

polymorpha. PLoS genetics **11**, e1005207.

Flores-Sandoval E, Romani F, Bowman JL. 2018. Co-expression and Transcriptome Analysis of Transcription Factors Supports Class C ARFs as Independent Actors of an Ancient Auxin Regulatory Module. Frontiers in plant science **9**, 1345.

Zhang W, Bei LI, & Bin YU. 2016. Genome-wide identification, phylogeny and expression analysis of the SBP-box gene family in maize (Zea mays). Journal of integrative agriculture **15**, 29–41.

Goodstein DM, Shu S, Howson R, *et al.* 2012. Phytozome: a comparative platform for green plant genomics. Nucleic acids research **40**, D1178–86.

He J, Xu M, Willmann MR, McCormick K, Hu T, Yang L, Starker CG, Voytas DF, Meyers BC, Poethig RS. 2018. Threshold-dependent repression of SPL gene expression by miR156/miR157 controls vegetative phase change in Arabidopsis thaliana. PLoS genetics **14**, e1007337.

Hirakawa Y, Fujimoto T, Ishida S, Uchida N, Sawa S, Kiyosue T, Ishizaki K, Nishihama R, Kohchi T, Bowman JL. 2020. Induction of Multichotomous Branching by CLAVATA Peptide in Marchantia polymorpha. Current biology: CB **30**, 3833–3840.e4.

Hirt H, Shinozaki K. 2003. *Plant Responses to Abiotic Stress*. Springer Science & Business Media.

Hiwatashi T, Goh H, Yasui Y, *et al.* 2019. The RopGEF KARAPPO Is Essential for the Initiation of Vegetative Reproduction in Marchantia polymorpha. Current biology: CB **29**, 3525–3531.e7.

Honma T, Goto K. 2001. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature **409**, 525–529.

Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. 2007. WoLF PSORT: protein localization predictor. Nucleic acids research **35**, W585–7.

Hug N, Longman D, Cáceres JF. 2016. Mechanism and regulation of the nonsense-mediated decay pathway. Nucleic acids research **44**, 1483–1495.

Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. 2015. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics **31**, 1296–1297.

Hu T, Manuela D, Xu M. 2023*a*. SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 9 and 13 repress BLADE-ON-PETIOLE 1 and 2 directly to promote adult leaf morphology in Arabidopsis. Journal of experimental botany **74**, 1926–1939.

Hunter P. 2008. The paradox of model organisms. The use of model organisms in research will continue despite their shortcomings. EMBO reports **9**, 717–720.

Hu L, Qi P, Peper A, Kong F, Yao Y, Yang L. 2023*b*. Distinct function of SPL genes in age-related resistance in Arabidopsis. PLoS pathogens **19**, e1011218.

Imaizumi T, Kadota A, Hasebe M, Wada M. 2002. Cryptochrome light signals control

development to suppress auxin sensitivity in the moss Physcomitrella patens. The Plant cell **14**, 373–386.

Impens L, Jacobs TB, Nelissen H, Inzé D, Pauwels L. 2022. Mini-Review: Transgenerational CRISPR/Cas9 Gene Editing in Plants. Frontiers in genome editing **4**, 825042.

Irifune K, Ono K, Takahashi M, Murakami H, Morikawa H. 1996. Stable transformation of cultured cells of the liverwortMarchantia polymorpha by particle bombardment. Transgenic research **5**, 337–341.

Ishida S, Suzuki H, Iwaki A, *et al.* 2022. Diminished Auxin Signaling Triggers Cellular Reprogramming by Inducing a Regeneration Factor in the Liverwort Marchantia polymorpha. Plant & cell physiology **63**, 384–400.

Ishizaki K, Chiyoda S, Yamato KT, Kohchi T. 2008. Agrobacterium-Mediated Transformation of the Haploid Liverwort Marchantia polymorpha L., an Emerging Model for Plant Biology. Plant & cell physiology **49**, 1084–1091.

Ishizaki K, Johzuka-Hisatomi Y, Ishida S, Iida S, Kohchi T. 2013. Homologous recombination-mediated gene targeting in the liverwort Marchantia polymorpha L. Scientific reports **3**, 1532.

Ishizaki K, Nishihama R, Ueda M, Inoue K, Ishida S, Nishimura Y, Shikanai T, Kohchi T. 2015. Development of Gateway Binary Vector Series with Four Different Selection Markers for the Liverwort Marchantia polymorpha. PloS one **10**, e0138876.

Ishizaki K, Nishihama R, Yamato KT, Kohchi T. 2016*a*. Molecular Genetic Tools and Techniques for Marchantia polymorpha Research. Plant & cell physiology **57**, 262–270.

Ishizaki K, Nishihama R, Yamato KT, Kohchi T. 2016*b*. Molecular Genetic Tools and Techniques for Marchantia polymorpha Research. Plant & cell physiology **57**, 262–270.

Ito K, Akiyama Y. 2005. Cellular functions, mechanism of action, and regulation of FtsH protease. Annual review of microbiology **59**, 211–231.

Jeong D-H, Park S, Zhai J, Gurazada SGR, De Paoli E, Meyers BC, Green PJ. 2011. Massive analysis of rice small RNAs: mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage. The Plant cell **23**, 4185–4207.

Jiao Y, Wang Y, Xue D, *et al.* 2010. Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. Nature genetics **42**, 541–544.

Jung J-H, Lee H-J, Ryu JY, Park C-M. 2016. SPL3/4/5 Integrate Developmental Aging and Photoperiodic Signals into the FT-FD Module in Arabidopsis Flowering. Molecular plant **9**, 1647–1659.

Kamisugi Y, Schlink K, Rensing SA, Schween G, von Stackelberg M, Cuming AC, Reski R, Cove DJ. 2006. The mechanism of gene targeting in Physcomitrella patens: homologous recombination, concatenation and multiple integration. Nucleic acids research **34**, 6205–6214.

Kato H, Ishizaki K, Kouno M, Shirakawa M, Bowman JL, Nishihama R, Kohchi T.

2015*a*. Auxin-Mediated Transcriptional System with a Minimal Set of Components Is Critical for Morphogenesis through the Life Cycle in Marchantia polymorpha. PLoS genetics **11**, e1005084.

Kato H, Ishizaki K, Kouno M, Shirakawa M, Bowman JL, Nishihama R, Kohchi T. 2015*b*. Correction: Auxin-Mediated Transcriptional System with a Minimal Set of Components Is Critical for Morphogenesis through the Life Cycle in Marchantia polymorpha. PLoS genetics **11**, e1005365.

Kato H, Yasui Y, Ishizaki K. 2020. Gemma cup and gemma development in Marchantia polymorpha. The New phytologist **228**, 459–465.

Kawamura S, Romani F, Yagura M, *et al.* 2022. MarpolBase Expression: A Web-Based, Comprehensive Platform for Visualization and Analysis of Transcriptomes in the Liverwort Marchantia polymorpha. Plant & cell physiology **63**, 1745–1755.

Khraiwesh B, Ossowski S, Weigel D, Reski R, Frank W. 2008. Specific gene silencing by artificial MicroRNAs in Physcomitrella patens: an alternative to targeted gene knockouts. Plant physiology **148**, 684–693.

Klein J, Saedler H, Huijser P. 1996. A new family of DNA binding proteins includes putative transcriptional regulators of the Antirrhinum majus floral meristem identity gene SQUAMOSA. Molecular & general genetics: MGG **250**, 7–16.

Kohchi T, Yamato KT, Ishizaki K, Yamaoka S, Nishihama R. 2021. Development and Molecular Genetics of. Annual review of plant biology **72**, 677–702.

Komatsu A, Terai M, Ishizaki K, Suetsugu N, Tsuboi H, Nishihama R, Yamato KT, Wada M, Kohchi T. 2014. Phototropin encoded by a single-copy gene mediates chloroplast photorelocation movements in the liverwort Marchantia polymorpha. Plant physiology **166**, 411–427.

Kopischke S, Schüßler E, Althoff F, Zachgo S. 2017. TALEN-mediated genome-editing approaches in the liverwort yield high efficiencies for targeted mutagenesis. Plant methods **13**, 20.

Kropat J, Tottey S, Birkenbihl RP, Depège N, Huijser P, Merchant S. 2005. A regulator of nutritional copper signaling in Chlamydomonas is an SBP domain protein that recognizes the GTAC core of copper response element. Proceedings of the National Academy of Sciences of the United States of America **102**, 18730–18735.

Kubota A, Kita S, Ishizaki K, Nishihama R, Yamato KT, Kohchi T. 2014. Co-option of a photoperiodic growth-phase transition system during land plant evolution. Nature communications **5**, 3668.

Lang D, Weiche B, Timmerhaus G, Richardt S, Riaño-Pachón DM, Corrêa LGG, Reski R, Mueller-Roeber B, Rensing SA. 2010. Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. Genome biology and evolution **2**, 488–503.

Lang D, Zimmer AD, Rensing SA, Reski R. 2008. Exploring plant biodiversity: the Physcomitrella genome and beyond. Trends in plant science **13**, 542–549.

Lan T, Zheng Y, Su Z, Yu S, Song H, Zheng X, Lin G, Wu W. 2019. OsSPL10, a SBP-Box Gene, Plays a Dual Role in Salt Tolerance and Trichome Formation in Rice (Oryza sativa L.). G3 Genes|Genomes|Genetics **9**, 4107–4114.

Lehti-Shiu MD, Panchy N, Wang P, Uygun S, Shiu S-H. 2017. Diversity, expansion, and evolutionary novelty of plant DNA-binding transcription factor families. Biochimica et Biophysica Acta, Gene Regulatory Mechanisms **1860**, 3–20.

Lei K-J, Lin Y-M, Ren J, Bai L, Miao Y-C, An G-Y, Song C-P. 2016. Modulation of the Phosphate-Deficient Responses by MicroRNA156 and its Targeted SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 3 in Arabidopsis. Plant & cell physiology **57**, 192–203.

Leinonen R, Sugawara H, Shumway M. 2010. The Sequence Read Archive. Nucleic acids research **39**, D19–D21.

Lei KJ, Zhou H, Gu DL, An GY. 2022. The involvement of abscisic acid-insensitive mutants in low phosphate stress responses during rhizosphere acidification, anthocyanin accumulation and Pi homeostasis in Arabidopsis. Plant science: an international journal of experimental plant biology **322**, 111358.

Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S. 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic acids research **30**, 325–327.

Letunic I, Bork P. 2018. 20 years of the SMART protein domain annotation resource. Nucleic acids research **46**, D493–D496.

Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P. 2004. SMART 4.0: towards genomic data integration. Nucleic acids research **32**, D142–4.

Letunic I, Khedkar S, Bork P. 2021. SMART: recent updates, new developments and status in 2020. Nucleic acids research **49**, D458–D460.

Li D, Flores-Sandoval E, Ahtesham U, Coleman A, Clay JM, Bowman JL, Chang C. 2020*a*. Ethylene-independent functions of the ethylene precursor ACC in Marchantia polymorpha. Nature plants **6**, 1335–1344.

Li Y, Han S, Sun X, Khan NU, Zhong Q, Zhang Z, Zhang H, Ming F, Li Z, Li J. 2023*a*. Variations in OsSPL10 confer drought tolerance by directly regulating OsNAC2 expression and ROS production in rice. Journal of integrative plant biology **65**, 918–933.

Li Y, He Y, Qin T, Guo X, Xu K, Xu C, Yuan W. 2023*b*. Functional conservation and divergence of miR156 and miR529 during rice development. The crop journal **11**, 692–703.

Li S, Liu G, Pu L, Liu X, Wang Z, Zhao Q, Chen H, Ge F, Liu D. 2021. WRKY Transcription Factors Actively Respond to in. Phytopathology **111**, 1625–1637.

Linde A-M, Sawangproh W, Cronberg N, Szövényi P, Lagercrantz U. 2020. Evolutionary History of the. Frontiers in plant science **11**, 829.

Ling L-Z, Zhang S-D. 2012. Exploring the evolutionary differences of SBP-box genes

targeted by miR156 and miR529 in plants. Genetica **140**, 317–324.

Li F-W, Nishiyama T, Waller M, *et al. 2020*b*. Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. Nature plants **6**, 259–272.

Li L, Shi F, Wang G, *et al. 2022. Conservation and Divergence of () Gene Family between Wheat and Rice. International journal of molecular sciences **23**.

Liu Q, Harberd NP, Fu X. 2016. SQUAMOSA Promoter Binding Protein-like Transcription Factors: Targets for Improving Cereal Grain Yield. Molecular plant **9**, 765–767.

Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua N-H. 2012. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. The Plant cell **24**, 4333–4345.

Mallett DR, Chang M, Cheng X, Bezanilla M. 2019. Efficient and modular CRISPR-Cas9 vector system for. Plant direct **3**, e00168.

Marchantia eFP Browser.

Poveda J . 2020. Marchantia polymorpha as a model plant in the evolutionary study of plant-microorganism interactions. Current Plant Biology **23**, 100152.

Meng Y, Hou Y, Wang H, Ji R, Liu B, Wen J, Niu L, Lin H. 2017. Targeted mutagenesis by CRISPR/Cas9 system in the model legume Medicago truncatula. Plant cell reports **36**, 371–374.

Michaely P, Bennett V. 1992. The ANK repeat: a ubiquitous motif involved in macromolecular recognition. Trends in cell biology **2**, 127–129.

Mickiewicz A, Rybarczyk A, Sarzynska J, Figlerowicz M, Blazewicz J. 2016. AmiRNA Designer - new method of artificial miRNA design. Acta biochimica Polonica **63**, 71–77.

Miura K, Ikeda M, Matsubara A, Song X-J, Ito M, Asano K, Matsuoka M, Kitano H, Ashikari M. 2010. OsSPL14 promotes panicle branching and higher grain productivity in rice. Nature genetics **42**, 545–549.

Monte I, Franco-Zorrilla JM, García-Casado G, Zamarreño AM, García-Mina JM, Nishihama R, Kohchi T, Solano R. 2019. A Single JAZ Repressor Controls the Jasmonate Pathway in Marchantia polymorpha. Molecular plant **12**, 185–198.

Monte I, Ishida S, Zamarreño AM, *et al. 2018. Ligand-receptor co-evolution shaped the jasmonate pathway in land plants. Nature chemical biology **14**, 480–488.

Montgomery SA, Tanizawa Y, Galik B, *et al. 2020. Chromatin Organization in Early Land Plants Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin. Current biology: CB **30**, 573–588.e7.

Morea EGO, da Silva EM, e Silva GFF, Valente GT, Barrera Rojas CH, Vincentz M, Nogueira FTS. 2016. Functional and evolutionary analyses of the miR156 and miR529 families in land plants. BMC plant biology **16**, 40.

Nagae M, Nakata M, Takahashi Y. 2008. Identification of negative cis-acting elements in response to copper in the chloroplastic iron superoxide dismutase gene of the moss Barbula unguiculata. Plant physiology **146**, 1687–1696.

Naito Y, Hino K, Bono H, Ui-Tei K. 2015. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. Bioinformatics **31**, 1120–1123.

Naramoto S, Hata Y, Fujita T, Kyozuka J. 2022. The bryophytes Physcomitrium patens and Marchantia polymorpha as model systems for studying evolutionary cell and developmental biology in plants. The Plant cell **34**, 228–246.

Nishihama R, Ishida S, Urawa H, Kamei Y, Kohchi T. 2016. Conditional Gene Expression/Deletion Systems for Marchantia polymorpha Using its Own Heat-Shock Promoter and Cre/loxP-Mediated Site-Specific Recombination. Plant & cell physiology **57**, 271–280.

Nishiyama T, Wolf PG, Kugita M, *et al.* 2004. Chloroplast phylogeny indicates that bryophytes are monophyletic. Molecular biology and evolution **21**, 1813–1819.

Noman A, Aqeel M, He S. 2016. CRISPR-Cas9: Tool for Qualitative and Quantitative Plant Genome Editing. Frontiers in plant science **7**, 1740.

Olas JJ, Van Dingenen J, Abel C, Działo MA, Feil R, Krapp A, Schlereth A, Wahl V. 2019. Nitrate acts at the Arabidopsis thaliana shoot apical meristem to regulate flowering time. The New phytologist **223**, 814–827.

Ortiz-Ramírez C, Hernandez-Coronado M, Thamm A, Catarino B, Wang M, Dolan L, Feijó JA, Becker JD. 2016. A Transcriptome Atlas of Physcomitrella patens Provides Insights into the Evolution and Development of Land Plants. Molecular plant **9**, 205–220.

Ossowski S, Schwab R, Weigel D. 2008. Gene silencing in plants using artificial microRNAs and other small RNAs. The Plant journal: for cell and molecular biology **53**, 674–690.

Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF. 2000. B and C floral organ identity functions require SEPALLATA MADS-box genes. Nature **405**, 200–203.

Perea-García A, Andrés-Bordería A, Huijser P, Peñarrubia L. 2021. The Copper-microRNA Pathway Is Integrated with Developmental and Environmental Stress Responses in. International journal of molecular sciences **22**.

Peterson BA, Haak DC, Nishimura MT, Teixeira PJPL, James SR, Dangl JL, Nimchuk ZL. 2016. Genome-Wide Assessment of Efficiency and Specificity in CRISPR/Cas9 Mediated Multiple Site Targeting in Arabidopsis. PloS one **11**, e0162169.

Pietrykowska H, Sierocka I, Zielezinski A, Alisha A, Carrasco-Sanchez JC, Jarmolowski A, Karlowski WM, Szweykowska-Kulinska Z. 2022. Biogenesis, conservation, and function of miRNA in liverworts. Journal of experimental botany **73**, 4528–4545.

Poole RL. 2007. The TAIR database. Methods in molecular biology **406**, 179–212.

Qi Y. 2019. *Plant Genome Editing with CRISPR Systems: Methods and Protocols*. Springer Nature.

Qiu Y-L, Li L, Wang B, *et al.* 2006. The deepest divergences in land plants inferred from phylogenomic evidence. Proceedings of the National Academy of Sciences of the United States of America **103**, 15511–15516.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**, 841–842.

Ran FA, Hsu PD, Lin C-Y, *et al.* 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell **154**, 1380–1389.

Rensing SA. 2018. Plant Evolution: Phylogenetic Relationships between the Earliest Land Plants. Current biology: CB **28**, R210–R213.

Rensing SA, Goffinet B, Meyberg R, Wu S-Z, Bezanilla M. 2020. *The Moss Physcomitrium (Physcomitrella) Patens: a Model Organism for Non-seed Plants*.

Riese M, Zobell O, Saedler H, Huijser P. 2008. SBP-domain transcription factors as possible effectors of cryptochrome-mediated blue light signalling in the moss Physcomitrella patens. Planta **227**, 505–515.

Sakoh M, Ito K, Akiyama Y. 2005. Proteolytic activity of HtpX, a membrane-bound and stress-controlled protease from Escherichia coli. The Journal of biological chemistry **280**, 33305–33310.

Schwab R, Ossowski S, Riester M, Warthmann N, Weigel D. 2006. Highly Specific Gene Silencing by Artificial MicroRNAs in Arabidopsis. The Plant cell **18**, 1121–1133.

Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D. 2005. Specific effects of microRNAs on the plant transcriptome. Developmental cell **8**, 517–527.

Shaw J, Renzaglia K. 2004. Phylogeny and diversification of bryophytes. American journal of botany **91**, 1557–1581.

Shen B, Zhang W, Zhang J, *et al.* 2014. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. Nature methods **11**, 399–402.

Shimamura M. 2015. Marchantia polymorpha : Taxonomy, Phylogeny and Morphology of a Model System. Plant & cell physiology **57**, 230–256.

Shimono M, Sugano S, Nakayama A, Jiang C-J, Ono K, Toki S, Takatsuji H. 2007. Rice WRKY45 plays a crucial role in benzothiadiazole-inducible blast resistance. The Plant cell **19**, 2064–2076.

Si L, Chen J, Huang X, *et al.* 2016. OsSPL13 controls grain size in cultivated rice. Nature genetics **48**, 447–456.

Strenkert D, Schmollinger S, Sommer F, Schulz-Raffelt M, Schroda M. 2011. Transcription factor-dependent chromatin remodeling at heat shock and copper-responsive promoters in Chlamydomonas reinhardtii. The Plant cell **23**, 2285–2301.

Streubel S, Deiber S, Rötzer J, Mosiolek M, Jandrasits K, Dolan L. 2023. Meristem dormancy in Marchantia polymorpha is regulated by a liverwort-specific miRNA and a clade

III SPL gene. Current biology: CB **33**, 660–674.e4.

Sugano SS, Nishihama R. 2018. CRISPR/Cas9-Based Genome Editing of Transcription Factor Genes in Marchantia polymorpha. Methods in molecular biology **1830**, 109–126.

Sugano SS, Nishihama R, Shirakawa M, Takagi J, Matsuda Y, Ishida S, Shimada T, Hara-Nishimura I, Osakabe K, Kohchi T. 2018. Efficient CRISPR/Cas9-based genome editing and its application to conditional genetic analysis in Marchantia polymorpha. PloS one **13**, e0205117.

Sugano SS, Shirakawa M, Takagi J, Matsuda Y, Shimada T, Hara-Nishimura I, Kohchi T. 2014. CRISPR/Cas9-mediated targeted mutagenesis in the liverwort Marchantia polymorpha L. Plant & cell physiology **55**, 475–481.

Sunkar R. 2012. *MicroRNAs in Plant Development and Stress Responses*. Springer Science & Business Media.

Sun M, Shen Y, Yang J, Cai X, Li H, Zhu Y, Jia B, Sun X. 2020. miR535 negatively regulates cold tolerance in rice. Molecular breeding: new strategies in plant improvement **40**, 1–12.

Sun T, Zhou Q, Zhou Z, Song Y, Li Y, Wang H-B, Liu B. 2022. SQUINT Positively Regulates Resistance to the Pathogen Botrytis cinerea via miR156–SPL9 Module in Arabidopsis. Plant & cell physiology **63**, 1414–1432.

Suzuki N, Choi W-G, Rivero RM, Miller G. 2022. *Cellular Signaling Networks in Plant Heat Stress Responses*. Frontiers Media SA.

Suzuki H, Kato H, Iwano M, Nishihama R, Kohchi T. 2023. Auxin signaling is essential for organogenesis but not for cell survival in the liverwort Marchantia polymorpha. The Plant cell **35**, 1058–1075.

Swarbreck D, Wilks C, Lamesch P, *et al.* 2008. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic acids research **36**, D1009–14.

Szövényi P. 2016. The Genome of the Model Species Anthoceros agrestis. Advances in botanical research. Advances in Botanical Research. Elsevier, 189–211.

Takenaka M, Yamaoka S, Hanajiri T, Shimizu-Ueda Y, Yamato KT, Fukuzawa H, Ohyama K. 2000. Direct transformation and plant regeneration of the haploid liverwort Marchantia polymorpha L. Transgenic research **9**, 179–185.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Molecular biology and evolution **38**, 3022–3027.

Tan QW, Lim PK, Chen Z, Pasha A, Provart N, Arend M, Nikoloski Z, Mutwil M. 2023. Cross-stress gene expression atlas of Marchantia polymorpha reveals the hierarchy and regulatory principles of abiotic stress responses. Nature communications **14**, 986.

Terada R, Johzuka-Hisatomi Y, Saitoh M, Asao H, Iida S. 2007. Gene targeting by homologous recombination as a biotechnological tool for rice functional genomics. Plant physiology **144**, 846–856.

Terada R, Urawa H, Inagaki Y, Tsugane K, Iida S. 2002. Efficient gene targeting by homologous recombination in rice. Nature biotechnology **20**, 1030–1034.

Bowman JL. 2022. The liverwort Marchantia polymorpha, a model for all ages. Current Topics in Developmental Biology. Academic Press, 1–32.

Tsuboyama S, Kodama Y. 2014. AgarTrap: a simplified Agrobacterium-mediated transformation method for sporelings of the liverwort Marchantia polymorpha L. Plant & cell physiology **55**, 229–236.

Tsuboyama S, Kodama Y. 2018. AgarTrap Protocols on your Benchtop: Simple Methods for -mediated Genetic Transformation of the Liverwort. Plant biotechnology **35**, 93–99.

Tsuboyama S, Nonaka S, Ezura H, Kodama Y. 2018. Improved G-AgarTrap: A highly efficient transformation method for intact gemmalings of the liverwort Marchantia polymorpha. Scientific reports **8**, 10800.

Tsuboyama-Tanaka S, Kodama Y. 2015. AgarTrap-mediated genetic transformation using intact gemmae/gemmalings of the liverwort Marchantia polymorpha L. Journal of plant research **128**, 337–344.

Tsuzuki M, Futagami K, Shimamura M, *et al.* 2019. An Early Arising Role of the MicroRNA156/529-SPL Module in Reproductive Development Revealed by the Liverwort Marchantia polymorpha. Current biology: CB **29**, 3307–3314.e5.

Tsuzuki M, Nishihama R, Ishizaki K, Kurihara Y, Matsui M, Bowman JL, Kohchi T, Hamada T, Watanabe Y. 2016. Profiling and Characterization of Small RNAs in the Liverwort, Marchantia polymorpha, Belonging to the First Diverged Land Plants. Plant & cell physiology **57**, 359–372.

Verzani J. 2011. *Getting Started with RStudio*. 'O'Reilly Media, Inc.'

Walker JM. 2007. *The Proteomics Protocols Handbook*. Springer Science & Business Media.

Walther D, Brunnemann R, Selbig J. 2007*a*. The Regulatory Code for Transcriptional Response Diversity and Its Relation to Genome Structural Properties in A. thaliana. PLoS genetics **3**, e11.

Walther D, Brunnemann R, Selbig J. 2007*b*. The Regulatory Code for Transcriptional Response Diversity and Its Relation to Genome Structural Properties in A. thaliana. PLoS genetics **3**, e11.

Wang J-W. 2014. Regulation of flowering time by the miR156-mediated age pathway. Journal of experimental botany **65**, 4723–4730.

Wang J-W, Czech B, Weigel D. 2009. miR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana. Cell **138**, 738–749.

Wang B, Geng S, Wang D, Feng N, Zhang D, Wu L, Hao C, Zhang X, Li A, Mao L. 2015*a*. Characterization of Squamosa Promoter Binding Protein-LIKE genes in wheat. Journal of plant biology = Singmul Hakhoe chi **58**, 220–229.

Wang H, Lu Z, Xu Y, *et al.* 2019. Genome-wide characterization of SPL family in Medicago truncatula reveals the novel roles of miR156/SPL module in spiky pod development. BMC genomics **20**, 552.

Wang W, Rohner N, Wang Y. 2023*a*. *Emerging Model Organisms*. Springer Nature.

Wang L, Sun S, Jin J, Fu D, Yang X, Weng X, Xu C, Li X, Xiao J, Zhang Q. 2015*b*. Coordinated regulation of vegetative and reproductive branching in rice. Proceedings of the National Academy of Sciences of the United States of America **112**, 15504–15509.

Wang X, Tan NWK, Chung FY, Yamaguchi N, Gan E-S, Ito T. 2023*b*. Transcriptional Regulators of Plant Adaptation to Heat Stress. International journal of molecular sciences **24**.

Wang H, Wang H. 2015. The miR156/SPL Module, a Regulatory Hub and Versatile Toolbox, Gears up Crops for Enhanced Agronomic Traits. Molecular plant **8**, 677–688.

Wang Z, Wang Y, Kohalmi SE, Amyot L, Hannoufa A. 2016. SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 2 controls floral organ development and plant fertility by activating ASYMMETRIC LEAVES 2 in Arabidopsis thaliana. Plant molecular biology **92**, 661–674.

Wang S, Wu K, Yuan Q, *et al.* 2012. Control of grain size, shape and quality by OsSPL16 in rice. Nature genetics **44**, 950–954.

Wang L, Yu P, Lyu J, Hu Y, Han C, Bai M-Y, Fan M. 2021. BZR1 Physically Interacts with SPL9 to Regulate the Vegetative Phase Change and Cell Elongation in. International journal of molecular sciences **22**.

Wang Q-H, Zhang J, Liu Y, Jia Y, Jiao Y-N, Xu B, Chen Z-D. 2022. Diversity, phylogeny, and adaptation of bryophytes: insights from genomic and transcriptomic data. Journal of experimental botany **73**, 4306–4322.

Wang J, Zhou L, Shi H, *et al.* 2018. A single transcription factor promotes both yield and immunity in rice. Science **361**, 1026–1028.

Wellman CH, Osterloff PL, Mohiuddin U. 2003. Fragments of the earliest land plants. Nature **425**, 282–285.

Wickett NJ, Mirarab S, Nguyen N, *et al.* 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proceedings of the National Academy of Sciences of the United States of America **111**, E4859–68.

Wood AJ, Oliver MJ, Cove DJ. 2013. *New Frontiers in Bryology: Physiology, Molecular Biology and Functional Genomics*. Springer Science & Business Media.

Xie Y, Liu Y, Wang H, Ma X, Wang B, Wu G, Wang H. 2017. Phytochrome-interacting factors directly suppress MIR156 expression to enhance shade-avoidance syndrome in Arabidopsis. Nature communications **8**, 348.

Xie Q, Wang X, He J, *et al.* 2021. Distinct Evolutionary Profiles and Functions of microRNA156 and microRNA529 in Land Plants. International journal of molecular sciences **22**.

Xie K, Wu C, Xiong L. 2006. Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. Plant physiology **142**, 280–293.

Xu M, Hu T, Zhao J, Park M-Y, Earley KW, Wu G, Yang L, Poethig RS. 2016. Developmental Functions of miR156-Regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) Genes in Arabidopsis thaliana. PLoS genetics **12**, e1006263.

Yadav S, Basu S, Srivastava A, Biswas S, Mondal R, Jha VK, Singh SK, Mishra Y. 2023. Bryophytes as Modern Model Plants: An Overview of Their Development, Contributions, and Future Prospects. Journal of plant growth regulation, 1–18.

Yamasaki H, Hayashi M, Fukazawa M, Kobayashi Y, Shikanai T. 2009. SQUAMOSA Promoter Binding Protein-Like7 Is a Central Regulator for Copper Homeostasis in Arabidopsis. The Plant cell **21**, 347–361.

Yamasaki K, Kigawa T, Inoue M, *et al.* 2004. A novel zinc-binding motif revealed by solution structures of DNA-binding domains of Arabidopsis SBP-family transcription factors. Journal of molecular biology **337**, 49–63.

Yanagisawa S. 1998. Transcription factors in plants: Physiological functions and regulation of expression. Journal of plant research **111**, 363–371.

Yanagisawa S. 2004. Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants. Plant & cell physiology **45**, 386–391.

Yin H, Hong G, Li L, Zhang X, Kong Y, Sun Z, Li J, Chen J, He Y. 2019. miR156/SPL9 Regulates Reactive Oxygen Species Accumulation and Immune Response in Arabidopsis thaliana. Phytopathology **109**, 632–642.

Yu N, Cai W-J, Wang S, Shan C-M, Wang L-J, Chen X-Y. 2010. Temporal control of trichome distribution by microRNA156-targeted SPL genes in Arabidopsis thaliana. The Plant cell **22**, 2322–2335.

Yue E, Li C, Li Y, Liu Z, Xu J-H. 2017. MiR529a modulates panicle architecture through regulating SQUAMOSA PROMOTER BINDING-LIKE genes in rice (Oryza sativa). Plant molecular biology **94**, 469–480.

Yu S, Galvão VC, Zhang Y-C, Horrer D, Zhang T-Q, Hao Y-H, Feng Y-Q, Wang S, Schmid M, Wang J-W. 2012. Gibberellin regulates the Arabidopsis floral transition through miR156-targeted SQUAMOSA promoter binding-like transcription factors. The Plant cell **24**, 3320–3332.

Zhang J, Fu X-X, Li R-Q, *et al.* 2020. The hornwort genome and early land plant evolution. Nature plants **6**, 107–118.

Zhang L-L, Huang Y-Y, Zheng Y-P, *et al.* 2022. Osa-miR535 targets SQUAMOSA promoter binding protein-like 4 to regulate blast disease resistance in rice. The Plant journal: for cell and molecular biology **110**, 166–178.

Zhang H, Li L. 2013. SQUAMOSA promoter binding protein-like7 regulated microRNA408 is required for vegetative development in Arabidopsis. The Plant journal: for cell and molecular

biology **74**, 98–109.

Zhang S-D, Ling L-Z, Yi T-S. 2015. Evolution and divergence of SBP-box genes in land plants. BMC genomics **16**, 787.

Zhang H, Zhang L, Han J, Qian Z, Zhou B, Xu Y, Wu G. 2019. The nuclear localization signal is required for the function of squamosa promoter binding protein-like gene 9 to promote vegetative phase change in Arabidopsis. Plant molecular biology **100**, 571–578.

Zhao H, Liu X, Wang J, Qian Q, Zhang G. 2022. The coordinated regulation mechanism of rice plant architecture and its tolerance to stress. Frontiers in plant science **13**, 1087378.

Zhao J, Lu Z, Wang L, Jin B. 2020. Plant Responses to Heat Stress: Physiology, Transcription, Noncoding RNAs, and Epigenetics. International journal of molecular sciences **22**.

Zheng C, Ye M, Sang M, Wu R. 2019. A Regulatory Network for miR156-SPL Module in. International journal of molecular sciences **20**.

Zolkiewski M. 2006. A camel passes through the eye of a needle: protein unfolding activity of Clp ATPases. Molecular microbiology **61**, 1094–1100.