Universidade do Minho
Escola de Ciências

Sara Filipa Pinto Coelho

Study of ancestral genes in *Marchantia polymorpha*

outubro de 2019

Study of ancestral genes in *Marchantia polymorpha*

Sara Coelho

UMinho | 2019

Universidade do Minho
Escola de Ciências

Sara Filipa Pinto Coelho

Study of ancestral genes in *Marchantia polymorpha*

Dissertação de Mestrado
Mestrado em Genética Molecular

Trabalho efetuado sob a orientação da
Professora Doutora Maria Manuela Ribeiro Costa

outubro de 2019

Acknowledgements

This last year was one of the most challenging years of my life, but I cannot describe how grateful I am for the time that I spent working in this thesis.

First of all, I would like to thank Professor Manuela for the opportunity to work with a fascinating theme in her amazing group. I am so thankful for all the dedication, opportunities, advices, learning moments and for always believing in me and in my work.

Sara, I would like to thank for all the help and care that you gave me. You made me part of the team and without your teaching I would never be here.

Rómulo, I do not even know how to express my gratitude to you. I still do not know how you can do your work and help ten people at the same time. You have a giant heart and no matter the situation you are always there. Sorry for all the questions and moments of despair and again, thank you.

Helena, my lab soul mate. Thank you for being so much like me and yet for teaching me so much. All the moments that we shared in the lab will be tenderly remembered.

Ana Maria, I really enjoyed the short period that we shared the bench. Thanks for all the exhaustive change of ideas and for all the precious help in the most difficult moments.

To all the people in the lab, thank you for all the laugh and support. Also, a big thank you for D. Manuela and Sr. Luís for allowing us to perform our work.

Daniela and Helena, a thank you is not enough for everything you did for me. I am not an easy person, but you always dealt with my bad mood and with my existential crises. I hope this is not the end and that we will have more lunches, dinners and deep talks during our national/international trips.

João, Petisco and Tiago, thanks for all the relaxing moments and for never giving up of us when we spent the lunch time talking about the lab.

To my little ones, Adriana, Christina and Rafaela thank you for always support me and for all the times you showed the pride that you have in me. Mass and MJ you were amazing too.

Lastly but not least, I would like to thank to my parents and to my brother. Without your support and love none of this would be possible.

This work was funded by FEDER funds through the Operational Competitiveness Programme-COMPETE and by National Funds through FCT—Fundação para a Ciência e a Tecnologia under the project PTDC/BIA-PLA/1402/2014, "EvoMod- Origin and Evolutionary establishment of a transcriptional module controlling flower asymmetry". Work supported by UID/MULTI/04046/2013 centre grant from FCT, Portugal (to BioISI).

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Resumo

Estudo de genes ancestrais em *Marchantia polymorpha*

A descrição da história evolutiva das plantas é um desafio devido aos constantes eventos de duplicação genómica que ocorreram ao longo da evolução. Estes eventos aumentam a complexidade do genoma, o que por sua vez aumenta a dificuldade em perceber a função de um determinado gene devido à redundância entre membros da mesma família de genes. A análise da função de um determinado gene em plantas basais é um método que permite perceber a função ancestral de um gene devido à presença de um menor número de cópias no genoma destas plantas. Assim, com o estudo de genes ancestrais é possível perceber como módulos reguladores de genes foram estabelecidos e como evoluíram. O módulo DDR, constituído pelas proteínas DIVARICATA (DIV), RADIALIS (RAD) e DIV-and-RAD-Interacting Factors (DRIF), controla diversos processos de desenvolvimento como por exemplo o estabelecimento da assimetria floral em *Antirrhinum majus*. Genes homólogos a *DIV*, *DIVARICATA LIKE* (*DIVL*) e *DRIF* foram encontrados em *Marchantia polymorpha*, uma planta não vascular que tem vindo a ser reafirmada como planta modelo devido à sua posição filogenética, à implementação de diferentes técnicas e ferramentas moleculares e outras características como o curto ciclo de vida que facilita a transformação e propagação. Nesta dissertação, foi estudada a função ancestral dos genes *DIV* e *DRIF* com o objetivo de perceber o papel do módulo DDR em plantas basais e como este módulo foi estabelecido ao longo da evolução.

Para obter o padrão de expressão temporal e espacial dos genes Mp*DIV*, Mp*DIVL* e Mp*DRIF* foi realizada uma reação de polimerização em cadeia quantitativa em tempo real (RT-qPCR) e uma hibridização *in situ*. Fusões GUS::promotor foram iniciadas para confirmar a localização e o nível de expressão destes genes em diferentes tecidos. Linhas *knockout* para os genes Mp*DIV1*, Mp*DIV2* e Mp*DIVL* foram desenvolvidas usando a tecnologia CRISPR/Cas9 e linhas de sobrexpressão para os genes Mp*DIV1* e Mp*DRIF* foram desenvolvidas usando a tecnologia Gateway para obter pistas sobre a função destes genes em *M. polymorpha*.

Palavras chave: Evolução; *Marchantia polymorpha*; Módulo DDR; Tecnologia CRISPR/Cas9

Abstract

Study of ancestral genes in *Marchantia polymorpha*

Exploring the evolutionary path in the plant kingdom is challenging due to the repeated events of duplication that occurred during plant evolution. These events increase the genome complexity, which also increase the difficulty in understanding the function of a given gene due to redundancy within multigene families. The analysis of gene function in early plant species is a process that allows the study of the role of ancestral gene functions due to the existence of fewer gene copies in these plants. Therefore, through the study of ancestral genes is possible to understand the establishment and evolution of regulatory modules. The DDR module, that is constituted by the DIVARICATA (DIV), RADIALIS (RAD) and DIV-and-RAD-Interacting Factors (DRIF) proteins, controls several developmental processes like the flower asymmetry in *Antirrhinum majus*. *DIV*, *DIVARICATA LIKE* (*DIVL*), and *DRIF* homologs were found in *Marchantia polymorpha*, a non-vascular plant reaffirmed as a model species due to its unique phylogenetic position, to the implementation of several different molecular techniques and tools and other characteristics, like the short life cycle that facilitates the transformation and propagation. In this thesis, the ancestral function of *DIV* and *DRIF* was studied to understand the role of the DDR module in early plant species and how this module was established during evolution.

A real-time quantitative polymerase chain reaction (RT-qPCR) and a whole mount *in situ* were performed to obtain the temporal and spatial expression pattern of Mp*DIV*, Mp*DIVL* and Mp*DRIF*. GUS::promoter fusions were initiated to confirm the localization and the expression level of these genes in different tissues. Knockout lines were developed for Mp*DIV1*, Mp*DIV2* and Mp*DIVL* using the CRISPR/Cas9 technology and overexpression lines for Mp*DIV1* and Mp*DRIF* using the Gateway technology to obtain clues about the genes function in *M. polymorpha*.

Keywords: CRISPR/Cas9 technology; DDR module; Evolution; *Marchantia polymorpha*

Table of contents

List of abbreviations and acronyms

A…	Absorbance at *** nanometers
AFBs	AUXIN SIGNALLING F-BOXs
AP	Alkaline phosphatase
ARFs	AUXIN RESPONSE FACTORs
Aux	AUXIN
BCIP	5-bromo-4-chloro-3-indolyl-phosphate
bHLH	Basic-helix-loop-helix
BiFC	Bimolecular fluorescence complementation
bp	Base pairs
BSA	Bovine serum albumin
Cas9	CRISP-associated protein 9
cDNA	Complementary DNA
CRISPR	Clustered regularly interspaced short palindromic repeats
CTAB	Cetrimonium bromide
CYC	*CYCLOIDEA*
DDR	DIV-DRIF-RAD
DEPC	Diethyl pyrocarbonate
dH_2O	Deionised water
DIG	digoxigenin
DIV	*DIVARICATA*
DIVL	*DIVARICATA LIKE*
dNTP	Deoxyribonucleotide
DRIF	*DIV-and-RAD-Interacting Factor*
DSB	Double-strand break
DTT	Dithiothreitol
DUF	Domain of unknown function
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobile shift assay
EtOH	Ethanol
FLC	*FLOWERING LOCUS C*

g	Relative centrifuge force
gDNA	Genomic DNA
gRNA	Guide RNA
GUS	β-glucoronidase
h	Hour
HDR	Homology directed repair
HLS1	*HOOKLESS 1*
IAA	INDOLE ACETIC ACID
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	Kilobase
LB	Lysogeny broth
Mb	Megabase
MEE3	*MATERNAL EFFECT EMBRYO ARREST 3*
MES	2-(N-morpholino)ethanesulfonic acid
min	Minute
MM	Molecular marker
Mp*AN*	*Marchantia polymorpha ANGUSTIFOLIA*
Mp*EF1α*	*Marchantia polymorpha ELONGATION FACTOR1α*
Mp*JAZ*	*Marchantia polymorpha JAZ gene*
mRNA	Messenger RNA
NBT	Nitroblue tetrazolium chloride
NHEJ	Non-homologous end joining
OD	Optical density
Oligod(T)	Oligodeoxythymidylic acid
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PVP	Polyvinylpyrrolidone
RAD	*RADIALIS*
RNase	Ribonuclease
rpm	Revolutions per minute

RSL	*ROOT HAIR DEFECTIVE SIX-LIKE*
RSM1	*RADIALIS-LIKE SANT/MYB 1*
RT-PCR	Reverse transcriptase polymerase chain reaction
RT-qPCR	Real-time quantitative polymerase chain reaction
SDS	Sodium dodecyl sulphate
siPEP	Small interfering peptides
TAE	Tris-acetate-EDTA-buffer
TALENs	Transcription activator-like effector nucleases
TE	Tris-EDTA buffer
TIR1	TRANSPORT INHIBITOR RESISTANT1
Tm	Melting temperature
UTP	Uridine 5'-triphosphate
UTR	Untranslated region
UV	Ultraviolet
WT	Wild type
X-Gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside
X-Gluc	5-bromo-4-chloro-3-indolyl glucuronide
ZFNs	Zinc finger nucleases

1. Introduction

The origin of land plants represents one of the most important evolutionary events that allowed the establishment of complex terrestrial ecosystems (Taylor et al., 2012). The increase in the genome complexity during the evolutionary path of plants hampers the understanding of the function of a given gene due to redundancy within multigene families. Although most of the information about gene functions is available for more recent plant lineages, the study of ancestral gene function in basal land plant species is an important approach to uncover evolutionary events. *Marchantia polymorpha* is an excellent early plant model for the study of ancestral gene function, due to the existence of fewer duplicated gene copies, which allows a better understanding of how early gene regulatory networks were established. Moreover, *M. polymorpha* has practical advantages such as the short life cycle that is predominately haploid and the high susceptibility to *Agrobacterium* transformation (Shimamura, 2016).

1.1. Gene regulatory networks

The establishment of new interactions within a gene regulatory network drives the emergence of new biological functions and morphologies. These new interactions are often associated with repeated events of gene duplication that occur during evolution and increase the genome complexity (Wilson et al., 1974).

Regulatory networks are controlled by different families of transcription factors, that control spatial and temporal expression of several target genes, thus generating a modular structure composed by several sub-circuits (Erwin & Davidson, 2009). Transcription factors, as regulatory proteins, have to interact with DNA and other proteins through different domains within the protein. The DNA-binding domain is responsible for the attachment of the transcription factor to specific sequences of DNA (*cis*-acting elements). The trans-activation or repression domain contains one or more docking sites for other interacting proteins (Mitchell & Tjian, 1989).

Despite the importance of the emergence of new regulatory modules, there are very few studies that explain these molecular events. With the recent rise of high-throughput methods, studies with protein-protein interactions, protein-DNA interactions and RNA-seq have generated large-scale data useful to comprehend some evolutionary mechanisms (Landt et al., 2012; Proost & Mutwil, 2016; Wang et al., 2001; Yu et al., 2011).

The duplication of individual genes, chromosomal segments or whole genomes, and the rearrangement of pre-existing domains by genetic recombination can lead to the creation of novel proteins capable of expanding the regulatory networks (Bagowski et al., 2010; Schmidt & Davies, 2007). In plants, gene duplication is a common process and is widely recognized as the main source of new genes (Shiu et al., 2005).

The rearrangement of pre-existing domains by genetic recombination gives origin to novel proteins in a single event. The rearrangement can occur by domain(s) insertion or deletion, exchange and duplication (Pasek et al., 2006; Schmidt & Davies, 2007). Thus, the acquisition of new domains by genetic recombination can lead to new DNA binding specificities and, consequently, to new interactions (Amoutzias et al., 2008). The rearrangement of pre-existing domains is behind the evolution of the diverse basic-helix-loop-helix (bHLH) transcription factors family (Amoutzias et al., 2004) and the evolution of the transcription factors DIVARICATA (DIV), RADIALIS (RAD) and DIV-and-RAD-Interacting Factor (DRIF) (Raimundo et al., 2018). Pires & Dolan (2010) showed that the major subfamilies of proteins were established in early land plants and conserved during plant evolution.

1.2. DDR regulatory module

The DIV, RAD and DRIF proteins constitute a regulatory module (DDR) that controls several developmental processes across the angiosperms. In *Antirrhinum majus* and *Solanum lycopersicum* (tomato), this regulatory module regulates key developmental traits like flower dorsoventral asymmetry or fruit development, respectively (Machemer et al., 2011; Raimundo et al., 2013).

1.2.1. *Antirrhinum majus* flower asymmetry

Antirrhinum flowers are asymmetric with clear morphologic differences between the upper (dorsal) and lower (ventral and lateral) petals (figure 1A) (Luo et al., 1996). The dorsoventral asymmetry of the *Antirrhinum* flower is controlled by the activity of *CYCLOIDEA* (*CYC*). *CYC* is expressed from early stages of the flower primordia development in the dorsal domain, and it controls the expression of *RAD*. When *RAD* is inactivated, the flowers are partially ventralized, which suggests the important role of this gene in the establishment of the flower asymmetry (figure 1) (Corley et al., 2005; Costa et al., 2005).

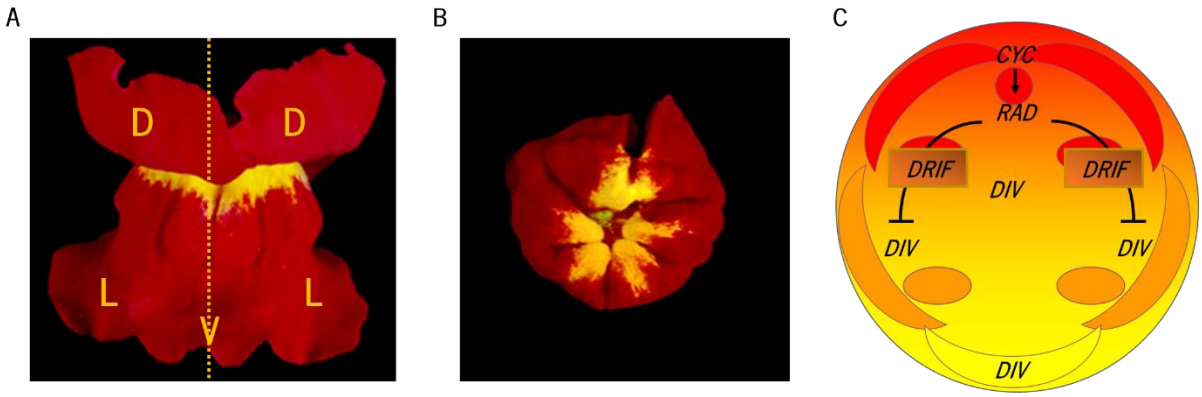

Figure 1: *Antirrhinum majus* flower and the model showing genetic interactions underlying the flower asymmetry establishment. A) *A. majus* structure: five petals with different size and shape along the symmetry axis (indicated by a dashed line). The two upper petals exhibit dorsal identity (D), the two smaller petals exhibit lateral identity (L) and the one even smaller petal exhibits ventral identity (V). **B)** Frontal view of a *rad* mutant line in which is possible to observe the loss of the dorsal identity and the spread of the ventral identity to the dorsal domains. **C)** *A. majus* flower frontal diagram: CYC expression leads to RAD activation in the dorsal domain (red). RAD inhibits DIV activity by interaction with DRIF proteins in the dorsal (red) and lateral (orange) domains. DIV is expressed throughout the flower meristem, but it only has a phenotypic effect in the ventral domain (yellow). Adapted from Corley et al. (2005).

DIV encodes a MYB protein that promotes the ventral identity of *A. majus* flowers (figure 1C) (Galego & Almeida, 2002). *DIV* is expressed throughout the flower, but it only has an effect in the ventral region of the flower meristem (Almeida et al., 1997; Galego & Almeida, 2002). In the dorsal and lateral domains of the flower meristem, RAD inhibits the activity of DIV. Raimundo et al. (2013) showed that RAD does not antagonize DIV directly, instead they compete for proteins named DRIFs. In the ventral domain, DRIF proteins interact with DIV proteins forming a complex that controls regulatory processes involved in ventral identity specificity. The presence of RAD disrupts the interaction between DIV and DRIF either by sequestering DRIF proteins in the cytoplasm, preventing or reducing the formation of DIV-DRIF complexes in the nuclei, or by binding directly to DRIF in the nuclei (figure 2). The small size and the acting mode of the RAD proteins categorises them as small interfering peptides (siPEP) (Seo et al., 2011). These proteins are transcription factors that lost their DNA-binding domain but maintain their protein interaction domain. siPEPS can disrupt the formation of dimers giving origin to non-functional dimers incapable of activating the transcription. The *A. majus* flower asymmetry is, therefore, explained by a molecular antagonistic mechanism of RAD over DIV in the dorsal regions of *Antirrhinum* flowers (figure 2) (Raimundo et al., 2013).

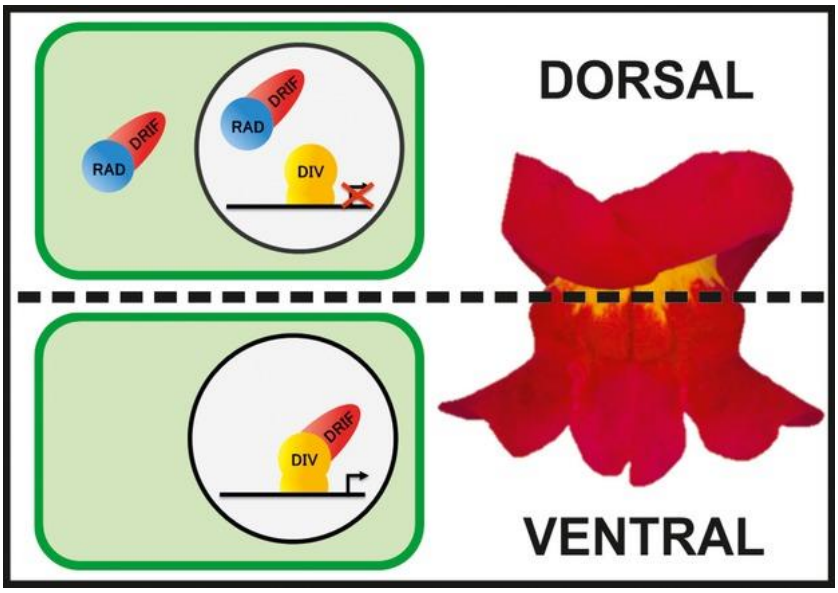

Figure 2: Proposed molecular model for the antagonistic mechanism that RAD exerts over DIV in the dorsal region of *Antirrhinum* flowers. In the ventral region of the *Antirrhinum* flower, DIV interacts with a DRIF protein in the nucleus forming a DIV-DRIF complex that promotes the regulation of genes that specify the ventral identity of the petal. In the dorsal region of the flower, RAD competes with DIV for the DRIF protein. RAD sequesters DRIF in the cytoplasm or binds to DRIF protein in the nucleus preventing the formation of the DIV-DRIF complex. Without the formation of this complex petals acquire dorsal identity and the flower asymmetry of *Antirrhinum* is established (Raimundo et al., 2013).

Homologs of *DIV*, *RAD* and *DRIF* genes were also associated with other different developmental processes in several plant species. In tomato, a homolog of RAD, FSM1, affects negatively the cell expansion in the fruit pericarp. FSM1 interacts with FSB1, a MYB protein that interacts with MYBI, a homolog of DIV. The complex FSM1/FSB1/MYBI controls the cell expansion and consequently the tomato fruit development (Machemer et al., 2011). In *Arabidopsis*, 9 homologs of DIV, 5 homologs of DRIF and 7 homologs of RAD were found, and interactions between these proteins have been reported. However, it is not known whether the antagonistic action within the DDR module occur in *Arabidopsis*. A homolog of *RAD* in *Arabidopsis*, *RADIALIS-LIKE SANT/MYB 1* (RSM1), when inactivated originates plants with no apical hooks and with short hypocotyl. This gene appeared to be related with the *HOOKLESS 1* (*HLS1*) mediated auxin signal in the regulation of early in photomorphogenesis (Hamaguchi et al., 2008). Also in *Arabidopsis*, MATERNAL EFFECT EMBRYO ARREST 3 (MEE3), other RAD homolog acts as a novel repressor of the floral transition by activating *FLOWERING LOCUS C* (*FLC*) (Li et al., 2015). In *Oryza sativa*, MYB genes are involved in the sugar and hormonal regulation of α-amylase gene expression induced by the gibberellins and sugar starvation (Lu et al., 2002).

1.2.2. DDR module evolutionary history

The DDR regulatory module has been recruited during plant evolution to regulate different biological traits. To understand the interactions between the DDR proteins and to obtain new insights about the establishment of the regulatory module, Raimundo et al. (2018) performed an analysis of the DIV, DRIF and RAD domain conformations.

DIV encodes a protein with on average 276 amino acids that is composed by two different MYB domains. The N-terminal domain- MYB domain (MYBI)- is composed by ≈ 44 amino acids (figure 3) and is responsible for protein-protein interactions (Lu et al., 2002; Rose et al., 1999). The C-terminal domain- MYBII- is composed by ≈ 51 amino acids and it is known to bind to DNA (Lu et al., 2002; Rose et al., 1999). A typical MYB domain is characterized by three conserved tryptophan residues regularly spaced (Wang et al., 1997). The MYBI domain is an atypical MYB domain with the third tryptophan residue replaced by a tyrosine (-W-X23-W-X20-Y-). MYBII, the second conserved domain, is denominated as SHAQKYF-type MYB due to the presence of the SHAQKYF amino acid motif that contains the last tyrosine of the MYB motif (-W-X19-W-X22-Y-) (Lu et al., 2002; Lu et al., 2009; Rose et al., 1999).

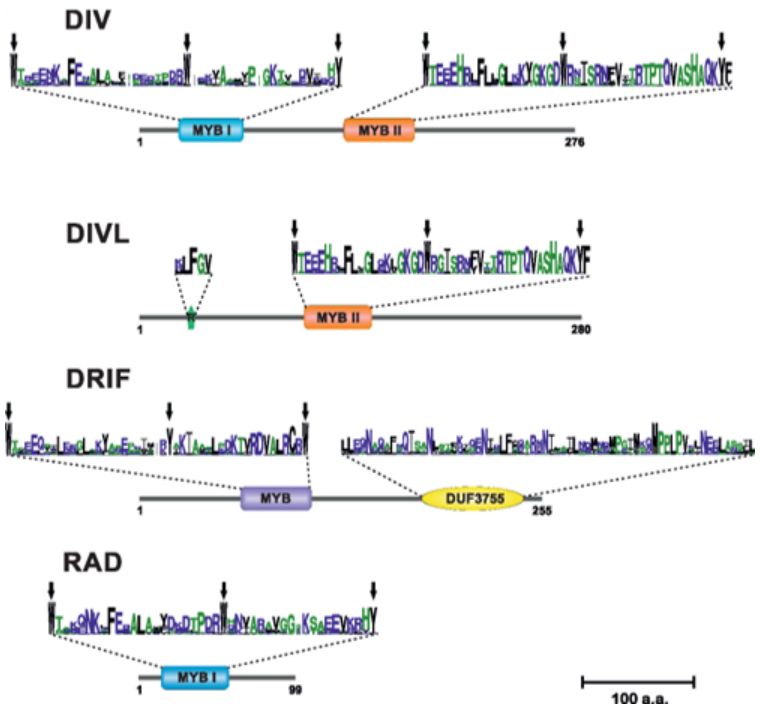

Figure 3: Proteic domains of DIV, DIVL, DRIF and RAD protein families. Schematic representation of the DIV, DIVL, DRIF and RAD proteins from the angiosperms *Amborella trichopoda, Oryza sativa, Solanum lycopersicum, Antirrhinum majus,* and *Arabidopsis thaliana*. The schemes were drawn in a scale of 100 amino acids based on the values of protein length, domain position, and domain length for all of the homologs found in the five angiosperm species. Domains with the same colour have the same topology that is represented by the amino acid sequences. Amino acids symbols with more height represent the degree of certainty of each amino acid position. The conserved aromatic residues typical of the MYB domain are signalled with black arrows (Raimundo et al., 2018).

DIVL proteins are similar to DIV family as they contain in the C-terminal region the SHAQKYF-MYBII domain but lack the atypical MYBI domain in the N-terminal characteristic of the DIV proteins (figure 3). DIVL proteins have on average 280 amino acids and instead of the MYBI domain they contain a very small motif termed R motif (R/KLFGV). The R motif is a plant-specific repression domain that is present in numerous *Arabidopsis* transcription factors (Ikeda & Ohme-Takagi, 2009).

The RAD family is composed by small proteins with on average 99 amino acids (figure 3). These proteins contain only one conserved domain with 44 amino acids similar to the DIV MYBI domain.

The DRIF family is composed by proteins with 255 amino acids on average that contain two conserved domains (figure 3). The first conserved domain is an atypical 46 amino acid long MYB domain that has a tyrosine instead of the central tryptophan characteristic of the MYB motif (-W-X23-Y-X20-W-). The second conserved domain is a domain of unknown function with 66 amino acids that is annotated as DUF3755.

The conserved domains of DIV, RAD and DRIF proteins are unique for each protein family, which makes these proteins easily trackable throughout plant evolution. In order to understand their evolutionary origin, Raimundo et al. (2018) also performed a phylogenetic analysis of the DIV, RAD and DRIF protein families using homologs from different algae and plant species (figure 4). The analysis showed that the MYB domain architecture of DIV, DIVL and DRIF proteins was maintained throughout plant evolution and that these proteins were already present in the green algae lineage. RAD family was only detected in the transition to seed plants and the phylogenetic analysis suggested that it might had its origin on a duplication of the MYBI DIV domain or on the duplication of a *DIV* gene that lost the MYBII domain. Furthermore, the phylogeny results showed that MYBI DIV domain and MYB domain of DRIF and RAD share a common evolutionary origin (figure 4). Thus, it was possible to conclude that the different MYB domains were a result of successive rearrangements and divergence of a single MYB domain that, at different evolutionary points, gave rise to the DIV, DRIF and RAD families.

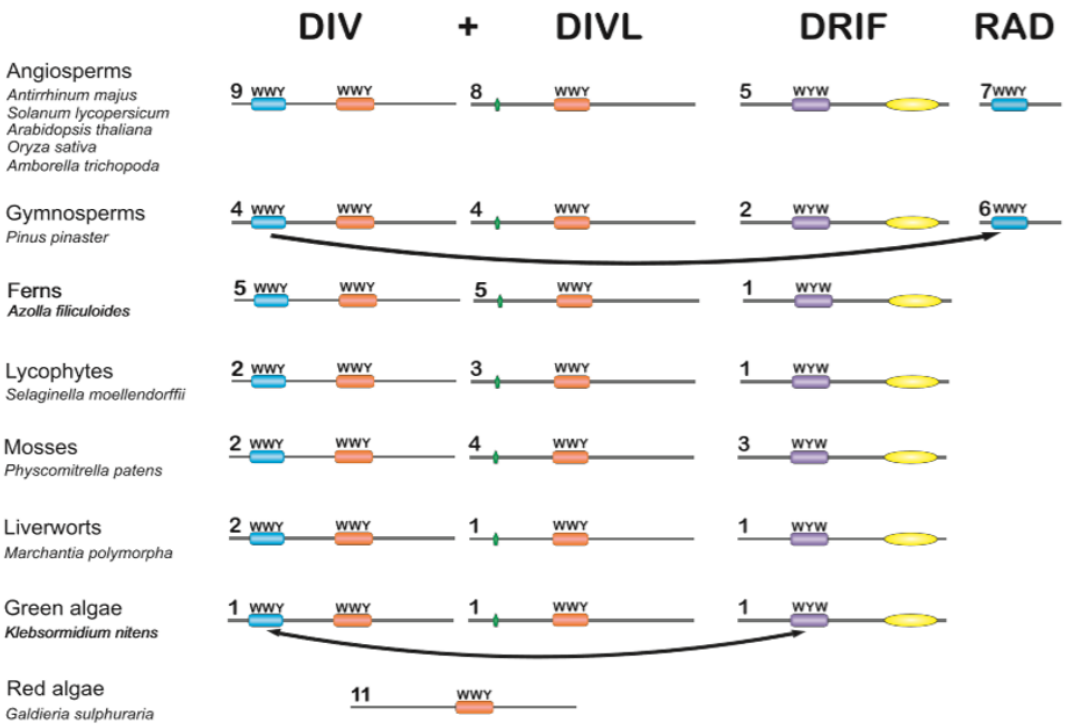

Figure 4: DDR module evolutionary history. Schematic representation of the DIV, DIVL, RAD and DRIF protein domain structure at different evolutionary points, from red algae to angiosperm lineage (*Galdieria sulphuraria, Klebsormidium nitens, Marchantia polymorpha, Physcomitrella patens, Selaginella moellendorffii, Azolla filiculoides, Pinus pinaster, Amborella trichopoda, Oryza sativa, Solanum lycopersicum, Antirrhinum majus,* and *Arabidopsis thaliana*). The average number of homologs from each protein family is represented for each lineage. The conserved protein domains are represented by the following colour code: blue for MYBI domain; orange for MYBII domain; purple for DRIF MYB domain and yellow for the DRIF DUF3375 domain. The typical aromatic residues of the MYB topology are represented next the respective domains. The black arrows represent events of domain duplication (Raimundo et al., 2018).

The DIV-DRIF-RAD antagonism mechanism is involved in the establishment of key developmental traits in *Antirrhinum* and tomato (Machemer et al., 2011; Raimundo et al., 2013). To determine when, during plant evolution, the interactions between the DDR module proteins were established a yeast-two-hybrid assay was performed using the cDNA sequences from different algae and plant species (Raimundo et al., 2018). The yeast-two-hybrid results showed that DIV and DRIF interaction is already established in the green algae lineage and is conserved in the first land plants such as *M. polymorpha*. The interaction DRIF-RAD was first detected in the gymnosperm lineage at the same time that the RAD family emerged (figure 5) (Raimundo et al., 2018).

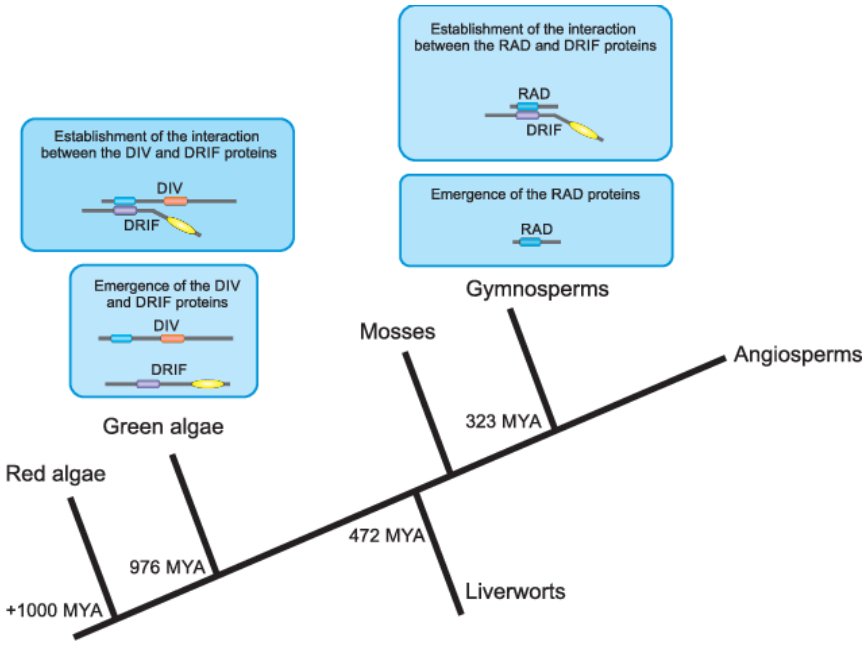

Figure 5: Schematic representation of the origin of DIV, RAD and DRIF protein families and the establishment of the DIV-DRIF and DRIF-RAD interactions. DIV and DRIF protein families emerged in the green algae lineage, and the DIV-DRIF interaction was established in the same lineage before the emergence of the first land plants. RAD protein family emerged later in the gymnosperm lineage as well as the DRIF-RAD interaction forming the DDR regulatory module (Raimundo et al., 2018).

One of the main problems of studying the establishment and evolution of regulatory modules is that during plant evolution repeated events of gene duplication led to an increase in genome complexity, which increases the difficulty in understanding the function of a given gene due to functional redundancy. The analysis of gene function in early plant species is, therefore, a more straightforward process due to the existence of fewer gene copies. Since DIV and DRIF genes, and their interaction, were already present in early plant lineages, the study of the gene ancestral functions in these plants should allow the unravelling of the ancestral role of the DIV and DRIF proteins and to understand how the DDR regulatory module may have been recruited throughout plant evolution.

1.3. Land plant evolution and the importance of studying ancestral genes

Phylogenetic studies showed that the earliest land plants descended from charophyte algae (figure 6) (Graham, 1985; Lewis & McCourt, 2004; Mishleri & Churchill, 1985). Bryophytes (liverworts, mosses, and hornworts), harbour the more ancestral plants and represent the earliest diverging group of terrestrial flora. Between the three divisions, liverworts are considered the most basal lineage of land plants (figure 6) (Mishler & Churchill, 1984; Mishleri & Churchill, 1985; Qiu et al., 2006; Wickett et al., 2014).

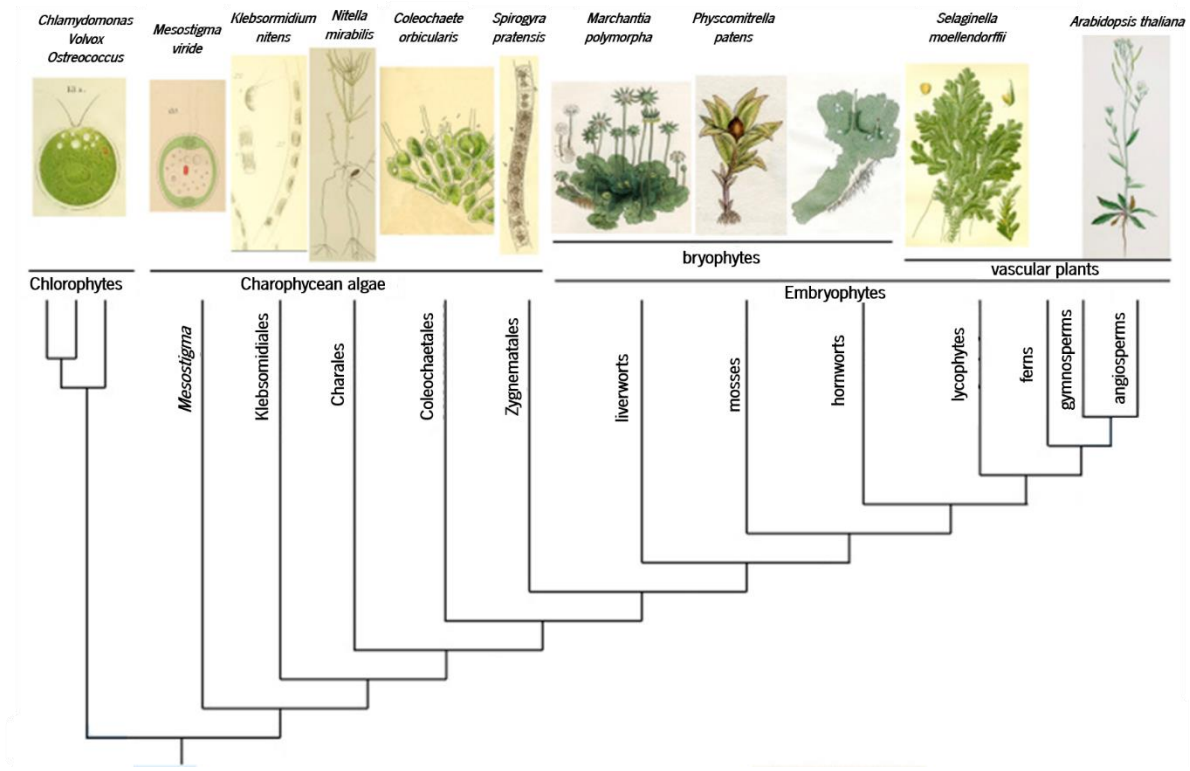

Figure 6: Phylogenetic tree representing the origin of land plants. In the phylogenetic tree are represented green algae, like the charophyte algae, land plants and vascular plants. The phylogenetic analysis revealed that early land plants descended from green algae and that liverworts are the most basal lineage of land plants, since they are phylogenetically closer to green algae. Adapted from Bowman et al. (2017).

Recent advances in phylogenetics and in comparative genomics have produced interesting molecular evolutionary hypotheses about the morphological and physiological evolution of land plants. However, to test these theories, functional studies are indispensable (Ishizaki, 2016; Ligrone et al., 2012). In the last years, several functional studies using both bryophytes and angiosperms showed that during evolution a recruitment of pre-existing gene regulatory networks may have occurred to perform different roles across the different plant lineages (Ishizaki, 2016).

Pires et al. (2013) performed a study with the *ROOT HAIR DEFECTIVE SIX-LIKE* (*RSL*) class I genes. These transcription factors control both the development of root hairs in the multicellular roots of the angiosperm *Arabidopsis thaliana* and the protonema development in *Physcomitrella patens*. The study was performed to understand if the regulatory network that incorporates the *RSL* genes was recruited or remodelled during the evolution of land plants. The phylogenetic analysis revealed that RSL proteins evolved in aquatic charophyte algae or in early land plants and that these proteins were conserved during land plant evolution (Pires et al., 2013). This study also showed that the *RSL* class I genes controlled the differentiation of filamentous structures in mosses after their divergence from other land plants, and that this kernel was recruited and remodelled to regulate the development of root hairs in vascular plants. In

M. polymorpha, the *RSL* class I gene controls the rhizoid formation and the development of gemmae (Proust et al., 2016). These results demonstrate that *RSL* class I genes controlled the generation of adaptive morphologies that allowed the transition of plants from water to land.

Functional studies were also performed to understand how plants acquired auxin-signalling components during evolution. In the angiosperm *A. thaliana* several molecular mechanisms are regulated by auxins (Benková et al., 2003; Friml et al., 2003; Harper et al., 2000; Tatematsu et al., 2004). Auxins act by a mechanism in which auxin receptors TRANSPORT INHIBITOR RESISTANT1/AUXIN SIGNALLING F-BOXs (TIR1/AFBs) target transcriptional co-repressors, AUXIN/INDOLE ACETIC ACIDs (Aux/IAAs), for degradation, thereby releasing AUXIN RESPONSE FACTORs (ARFs) to regulate transcription of auxin responsive genes (Finet & Jaillais, 2012). Orthologs of TIR1/AFBs, Aux/IAAs and ARFs were found in *P. patens* and *M. polymorpha* with lower redundancy than the one observed in more recent plants (Kato et al., 2015; Prigge et al., 2010). Auxin responses were also observed in charophyte, a green algal lineage related with the first land plants (reviewed in Cooke et al., 2002). The TIR1/AFBs, Aux/IAAs and ARFs are not present in the *Klebsormidium nitens* algae genome, despite the presence of an auxin response (Hori et al., 2014). These results suggest that the auxin-mediated transcriptional regulation was established after the divergence of *K. nitens* from the lineage that originate the of land plants (Kato et al., 2015). These two functional studies illustrate the importance of studying the role of ancestral gene functions to understand how molecular mechanisms evolved in plants.

1.4. *Marchantia polymorpha* as model species

M. polymorpha has been reaffirmed as a model species for the study of ancestral functions due to its unique phylogenetic position amongst plants that allows the testing of different evolutionary hypotheses (Bowman et al., 2016).

M. polymorpha is a common non-vascular bryophyte that is characterized by the presence of a dominant plant body in the gametophyte generation. *M. polymorpha* subsists in damp and moist habitats and has reduced dimensions due to the absence of a vascular system (reviewed in Shimamura, 2016).

In early botanic studies, before the appearance of modern plant taxonomy, liverworts like *Marchantia* were referred as lichens. With the arrival of the 16th century, plants that looked like particular body parts were thought to be beneficial and so were used in the treatment of that body part. Particularly, the thalloids of liverworts like *Marchantia* have a shape similar to a human liver hence the adoption of the

term liverwort (Bowman, 2016; Schuster, 1966). The genus *Marchantia* was designated in 1713 by Jean Marchant, a French botanist, in honour of his father Nicolas Marchant (Marchant, 1713).

In the 19th century, *M. polymorpha* was one of the most studied plants and several morphological and physiological studies were performed in this plant species. The appearance of new genetic and genomic tools in the 20th century allowed new advances in plant biology based in model plants like *A. thaliana*. The delay in the creation of these tools for *M. polymorpha* led to a gap in studies using this plant as a model. Nevertheless, in the last few years, *M. polymorpha* has re-emerged as a model species (Shimamura, 2016).

Besides its phylogenetic position, *M. polymorpha* has other characteristics that make it an excellent plant model such as a short life cycle, easy propagation and crossing, a high frequency of transformation, predominantly haploid, a small genome size (approximately 280 Mb) and the fact that the gene networks involved in several biological phenomena are conserved in a simpler form than in other land plants (Shimamura, 2016).

The chloroplastidial (Ohyama et al., 1986) and mitochondrial (Oda et al., 1992) genomes of *Marchantia* cultured cells were fully sequenced as well as the Y chromosome and a part of the X chromosome (Yamato et al., 2007). The whole genome sequencing of *M. polymorpha* was concluded in 2017 (Bowman et al., 2017).

Recently, several molecular techniques and tools were developed for *M. polymorpha* including transformation, overexpression, gene targeting by homologous recombination and genome editing (Ishizaki et al., 2008; Ishizaki et al., 2015; Kubota et al., 2013; Sugano et al., 2014; Sugano et al., 2018; Tsuboyama-Tanaka & Kodama, 2015; Tsuboyama-Tanaka et al., 2015). The implementation of these techniques was essential for the reaffirmation of *M. polymorpha* as a model species.

In *Marchantia*, like in other land plants, a multicellular haploid gametophyte generation alternates with a multicellular diploid sporophyte generation (Ishizaki et al., 2016; Kenrick, 1994; Niklas & Kutschera, 2010). In bryophytes it is common that the haploid gametophyte generation dominates the life cycle (Graham et al., 2000). In the life cycle of *Marchantia*, the gametophytic generation is dominant and leads to the production of the main plant body named thallus (figure 7A). The formation of the haploid gametophyte starts with the germination of a unicellular spore that by multiple asymmetric divisions originates the protonema, an initial short structure that contains rhizoids. The protonema begins to develop by highly regulated cell divisions of an apical cell that initiates the dorsoventral thallus formation.

The complex thalloid plant body grows radially with repeated branching forming a totally developed thallus with a clear dorsoventrality (figure 7A) (reviewed in Shimamura, 2016).

In the dorsal side of the mature thallus occurs the formation of structures called gemma cups that contain numerous cells named gemmae (figure 7A). When propagated, the gemma tissue grows by lateral expansion (figure 8). The four lobes of the gemma start to expand and a process of bifurcation occurs to form two apical notches in the thallus extremities. The two apical notches are separated by a central tissue zone (central lobe) that expands pushing the two notches apart. When completely formed, the central lobe starts to branch giving origin to the thallus (figure 8) (Solly et al., 2017). The gemma cups then develop above the apical cell producing gemmae that allow the thallus to reproduce asexually. For this to happen, it is necessary that a water droplet falls into the gemma cups to initiate a mechanism designated "splash-cup", by which gemmae are splashed out from the parent plant (Brodie, 1951). After this process, the rhizoids start to grow, and the apical cells begin to divide to give origin to a dorsoventral thallus (figure 7A).

The male and female organs of *M. polymorpha* are located in different thallus since this plant is dioecious. The antheridium and the archegonium, the male and female gametangia respectively, develop on umbrella-like sexual structures (gametangiophores) on the male (figure 7B to D) or female (figure 7E to G) thallus.

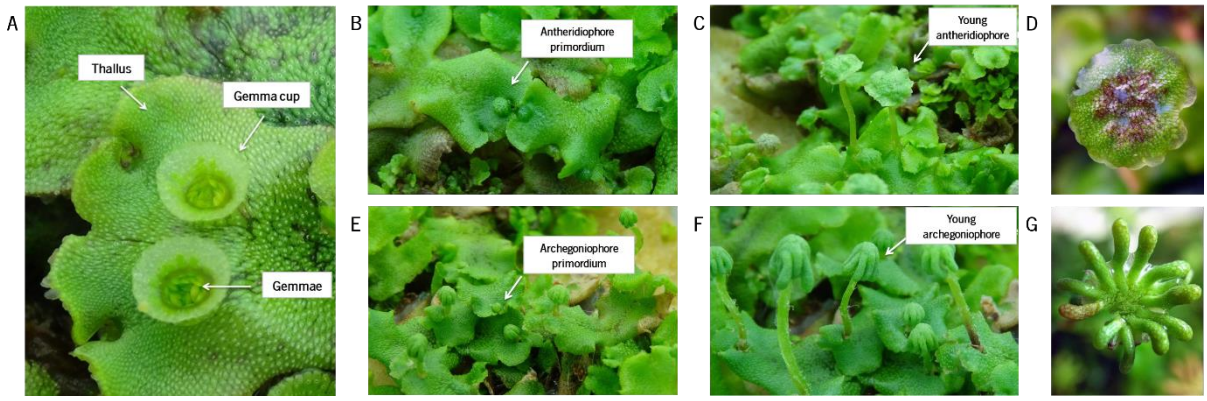

Figure 7: *Marchantia polymorpha* structures. A) A mature *Marchantia polymorpha* thallus showing gemmae cups containing several gemmae. **B)** Anteridiophores primordia (S1 stage). **C)** Young antheridiophores in development (S2 stage). **D)** Mature antheridiophore (S3 stage). **E)** Archegoniophores primordia (S1 stage). **F)** Young archegoniophores in development (S2 stage). **G)** Mature archegoniophores (S3 stage) (https://thenode.biologists.com/a-day-in-the-life-of-a-marchantia-lab/lablife/).

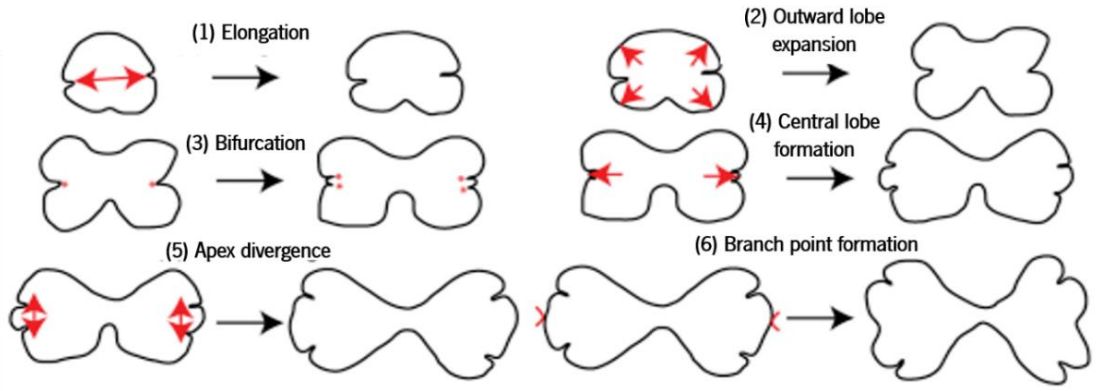

Figure 8: Schematic representation of the key shape transitions during the thallus development. (1) After gemmae propagation, the gemma tissue grows by lateral expansion. **(2)** The four lobes of the gemma start to expand. **(3)** A process of bifurcation occurs to form two apical notches on each one of the thallus extremities. **(4)** The two apical notches are separated by a central tissue zone, the central lobe, that expands pushing the two notches apart. **(5)** The apex starts to diverge leading to the complete central lobe formation. **(6)** When completely formed, the central lobe starts to branch giving origin to the thallus (Solly et al., 2017).

Each antheridium produces numerous motile male gametes named spermatozoids or antherozoids. In each archegonium only one egg is produced. To fertilize the egg, the antherozoids have to reach it using water to be mobilized. Once the fertilization has occurred, the zygote starts to develop into a sporophyte inside the archegonium (figure 9). The sporophyte begins to grow and consequently the inside part of the archegonium, the venter, develops becoming a calyptra, a structure that protects the growing young sporophyte. The sporophytes grow hanged upside down beneath the archegoniophore and after several meiotic divisions within the capsule, numerous spores are released (reviewed in Shimamura, 2016). After being released from the sporophyte, and when the conditions are favourable, the spores swell and germinate in a light dependent process (Nakazato et al., 1999). After germination, the spore starts to divide to originate the protonema. Then, the protonema cells divide irregularly until an apical cell arises to originate a dorsoventral thallus (figure 9) (Crandall-Stotler et al., 2009; Shimamura, 2016).

This predominantly haploid plant species reproduces very easily both asexually and sexually, which is a great advantage because it can be easily propagated under laboratory conditions. Besides, the gametophytes developed from a single spore or a gemma are genetically identical and can be established rapidly and easily propagated facilitating the transformation of this plant. Therefore, genetic analysis in *M. polymorpha* are more rapid and simpler than other model plants.

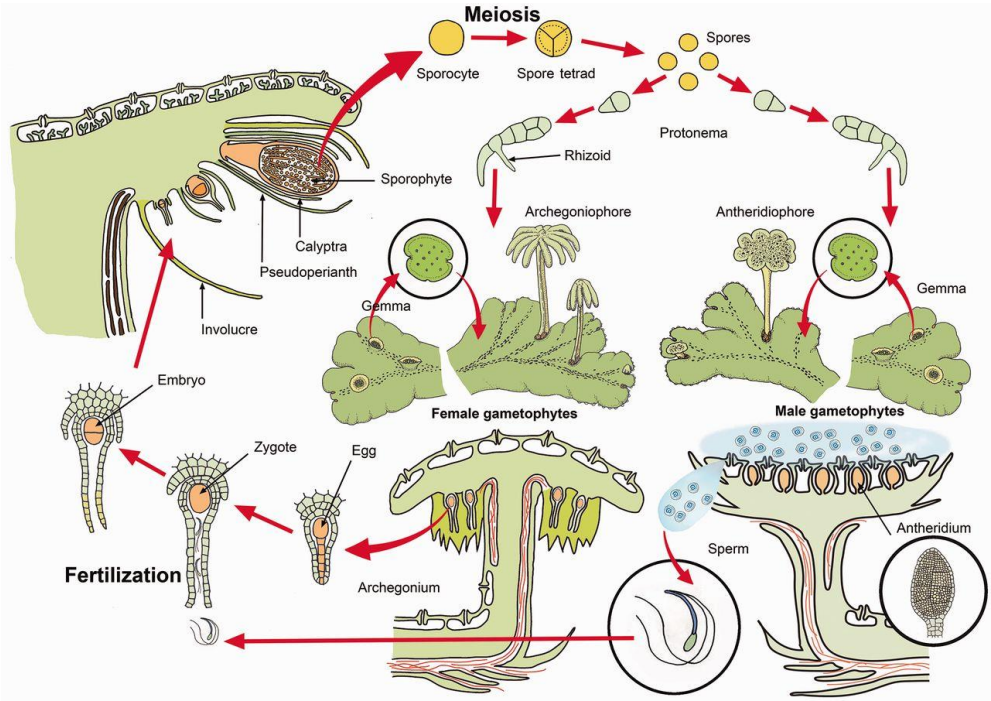

Figure 9: *Marchantia polymorpha* life cycle. The alternation of generations in *M. polymorpha* begins with the formation of a filamentous protonema with short rhizoids from the haploid spore. By mitotic divisions, a new thallus is formed from the protonema. The male and female structures, the antheridia and archegonia respectively, are formed in the young thallus. In each antheridium, numerous male gametes, the antherozoids, are produced, but in each archegonium only one egg cell is produced. To reach the female gametophyte, the antherozoids need water because they are demoted of flagella. After fertilization, the zygote develops inside the archegonium into a sporophyte. Inside the sporangium, meiotic divisions occur, and the product of these divisions are haploid spores (Shimamura, 2016).

1.4.1. CRISPR/Cas9 technology and *Marchantia polymorpha* mutagenesis

In recent years, different targeted genome modification technologies have been successfully used to improve plant breeding and genetic modification (Podevin et al., 2013; Voytas & Gao, 2014). Zinc finger nucleases (ZFNs) (Townsend et al., 2017) and transcription activator-like effector nucleases (TALENs) (Christian et al., 2013) are some of the technologies that have been applied for plant genome editing. In these two techniques, a DNA-binding domain is engineered to recognise a DNA target sequence. However, the use of both ZFNs and TALENs is both expensive and difficult due to the necessity to design and construct binding domains.

The **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats (CRISPR)-associated endonuclease Cas9 system (CRISPR/Cas9) is a key tool that allows the rapid modification of a target gene and is revolutionising targeted genome modification technologies (Cong et al., 2013; Mali et al., 2013). This system is based on the CRISPR adaptive immunity system of *Streptococcus pyogenes* and consists in a short guide RNA (gRNA) and in the CRISP-associated protein 9 (Cas9). The protein Cas9, that has endonuclease activity, can be targeted to a specific DNA sequence by an engineered 20 bp gRNA that binds to the target by base complementarity (figure 10). Target recognition is dependent on a

protospacer adjacent motif (PAM), a NGG sequence adjacent to the 3' end of the target sequence. After recognition, the gRNA anneals with the target sequence and the Cas9 causes a double-strand break (DSB) within it (3-4 nucleotides upstream the PAM sequence) (figure 10).

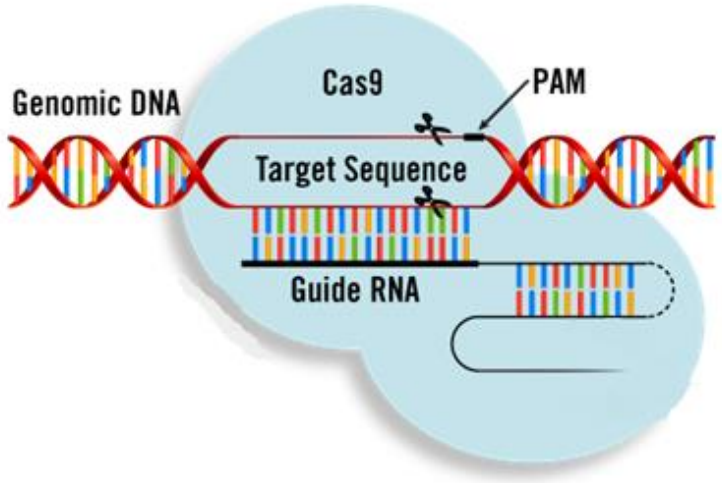

Figure 10: CRISPR/Cas9 technology. The designed guide RNA (gRNA) recognize the DNA target sequence with the help of a protospacer adjacent motif (PAM) sequence. After the recognition, the gRNA anneals with the target sequence and the Cas9 performs a double-strand break (DSB) within the sequence. Adapted from Khodthong et al. (2016).

The resulting DSB is then repaired by the non-homologous end joining (NHEJ) pathway or by the homology directed repair (HDR) pathway. The NHEJ pathway can cause small nucleotide insertions or deletions at the DSB site. If a donor template with homology to the targeted site is provided, the DBS may be repaired by the HDR pathway allowing precise genome mutations. The NHEJ pathway is the most active mechanism and the most common in plants (Belhaj et al., 2015).

The relatively cheap cost and simplicity of this technique allowed it to be successfully applied to several organisms, including plants, to create knockout mutants (Feng et al., 2014; Mao et al., 2013). Sugano et al. (2014) showed the application of this technique in *M. Polymorpha*, which increased the importance of this plant as model system.

1.5. Objectives

During the evolutionary path of land plants, an increase in genome complexity resulted in the establishment of new interactions within gene regulatory networks. The increase in genome complexity was accompanied with gene duplication, what makes the study of gene function more challenging due to the concomitant redundancy within the families of genes. In early plant species, the study of the ancestral gene functions is facilitated due to the existence of fewer gene copies. The DDR regulatory module controls

different developmental processes in angiosperms where, on average, 9 homologs of *DIV*, 8 homologs of *DIVL*, 5 homologs of *DRIF* and 7 homologs of *RAD* can be found in each species genome. On the other hand, in the genome of the basal plant *M. polymorpha*, 2 homologs of *DIV*, 1 homologs of *DIVL* and 1 homologs of *DRIF* can be found. Raimundo et al (2018) described the interaction of the two MpDIV proteins with the MpDRIF protein in *M. polymorpha*. The main aim of this thesis is to unveil the ancestral function of Mp*DIV*, Mp*DIVL* and Mp*DRIF* genes to contribute to the understanding on how their function may have been modulated during plant evolution.

In this work, different genetic tools and techniques were used in the study of the ancestral function of Mp*DIV*, Mp*DIVL* and Mp*DRIF* genes. The temporal and spatial expression pattern of these genes was analysed by RT-qPCR and by whole mount *in situ* hybridization. Constructs of promoter fusions with reporter genes were obtained for future plant transformation. Knockout and overexpression lines were obtained for these genes using the CRISPR/Cas9 and Gateway technologies, and their phenotypes analysed.

2. Materials and methods

2.1. Biological material

2.1.1. Plant material

Marchantia polymorpha ecotype BoGa was obtained from Sabine Zachgo's laboratory (University of Osnabrück, Germany). Plants were grown on half strength Gamborg medium [1.582 g L^{-1} Gamborg B5 medium including vitamins (Duchefa); 0.5 g L^{-1} 2-(N-morpholino)ethanesulfonic acid (MES) (Sigma-Aldrich); 1.4% (w/v) plant agar (Duchefa); pH 5.7] in 100 x 20 mm petri dishes (Greiner bio-one) under long-day conditions (16 h light/ 8 h dark) at 20 °C in a controlled growth room, with light intensity between 40-45 µmol m^{-2} s^{-1}.

2.1.1.1. Gemmae sterilisation protocol

Gemmae were sterilised by collecting 1-3 gemma cups to a sterile microtube. The gemmae were treated with 1 mL of surface sterilisation solution [0.2-0.5% (v/v) sodium hypochlorite and 0.1% (v/v) Triton X-100]. The tube was vortexed and after 1 min the supernatant was removed and the gemmae washed with 1 mL of ultrapure H_2O for 1 min. The H_2O wash was repeated twice and then the gemmae were transferred to half-strength Gamborg B5 medium plates.

2.1.2. Bacterial material

All the bacterial strains used in this work are described in Table 1.

Table 1: Bacterial strains used in the present work.

Specie	Strain	Genotype	Reference
Escherichia coli	DH5α	F- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM15 Δ(lac*ZYA-argF*)U169, hsdR17($r_K^-m_K^+$), λ-	Chen et al., 2018
	DH10β	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ- *rps*L(Str^R) *nup*G	Durfee et al., 2008
	XL-1 blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F- *proAB lacI*^{q} *Z*ΔM15 Tn*10* (Tet)	Bullock et al., 1987
Agrobacterium tumefaciens	C58C1 (GV2260)	pTiB6S3ΔT-DNA	Deblaere et al., 1985

2.2. Bacterial transformation

2.2.1. *Escherichia coli*

2.2.1.1. Preparation of competent cells

A single *E. coli* colony was used to inoculate 5 mL of lysogeny broth (LB) media (10 g L^{-1} NaCl; 10 g L^{-1} Tryptone; 5 g L^{-1} yeast extract) (Bertani, 2013) and incubated overnight at 37 °C and 200 rpm. A flask with 100 mL of LB media was inoculated with 1 mL of the overnight culture and grown at 37 °C and 200 rpm until an OD_{600} (optical density at 600 nm) of 0.25. The cells were pelleted in four tubes by centrifugation at 4 °C and 3000 *g* for 10 min. The supernatant was discarded, and each pellet was resuspended in 8 mL of cold and sterile 0.1 M $MgCl_2$. The four tubes were maintained on ice for 30 min and the cells were then pelleted as before and resuspended gently in 5 mL of cold and sterile TG salts (75 mM $CaCl_2$; 6 mM $MgCl_2$; 15% (v/v) glycerol). Cells were pelleted again as described previously, resuspended in 1.5 mL TG salts and kept on ice for 4-24 h. Cells were dispensed in 100 µL aliquots into pre-chilled microtubes, frozen in liquid nitrogen and stored at -80°C.

2.2.1.2. Transformation

E. coli transformation was initiated by adding 50-100 ng of DNA to an aliquot of 100 µL of competent cells. The mixture was incubated 30 min on ice. Cells were then heat-shocked at 42 °C for 45 s, returned to ice for 2 min and 1 mL of LB was added to the tube. The cells were incubated for 1 h at 37 °C and 200 rpm and then centrifuged for 1 min at 4000 *g*. The supernatant (approximately 1 mL) was discarded and 100 µL of the transformation mix was spread on LB-agar plates (LB medium; 1.5% (w/v) agar) containing the proper antibiotic. Plates were incubated overnight at 37 °C.

2.2.2. *Agrobacterium tumefaciens*

2.2.2.1. Preparation of competent cells

A single *A. tumefaciens* colony was used to inoculate 5 mL of LB media, supplemented with 100 µg mL^{-1} ampicillin and 50 mg mL^{-1} rifampicin, and incubated overnight at 28 °C and 200 rpm. A flask with 50 mL of LB media was inoculated with 50 µL of the initial culture and grown overnight at 28 °C and 200 rpm. When the OD_{600} of the culture was between 0.5-1, the culture was transferred for a tube and cooled on ice for 10 min and centrifuged at 4 °C and 3000 *g* for 6 min. The supernatant was

discarded, and the pellet rinsed with 1 mL of ice cold and sterile 20 mM $CaCl_2$ and spun briefly. The supernatant was discarded again, and the pellet resuspended in 1 mL of cold and sterile 20 mM $CaCl_2$. Cells were dispensed in 100 µL aliquots into pre-chilled microtubes, frozen in liquid nitrogen and stored at -80 °C.

2.2.2.2. Transformation

Agrobacterium transformation was performed by adding 1 µg of plasmid DNA to 100 µL of competent cells. The transformation was mixed by inversion and frozen in liquid nitrogen for 5 min. The tube was then incubated for 10 min at room temperature and 1 mL of LB was added. The tube was incubated for 3 h at 28 °C and 200 rpm. Cells were centrifuged for 1 min at 4000 *g*. Approximately 1 mL was removed from the tube and discarded, and 100 µL of the transformation mix was spread on LB-agar plates with the proper antibiotics. Plates were incubated for 48 h at 28 °C.

2.3. DNA methods

2.3.1. Plant DNA extraction

3-weeks to 1-month plant thallus were sectioned (5 cm^2) and placed in a 1.5 mL microtube together with two metal gridding balls and 400 µL of extraction buffer (200 mM Tris-HCl pH 7.5; 250 mM NaCl; 25 mM EDTA; 0.5% (w/v) SDS). The tissue was homogenized in a mixer mill (RETSCH MM 400) for 45 s at 30 Hz. The tube was then centrifuged at 20000 *g* for 5 min and 300 µL of supernatant was transferred to a new tube. The nucleic acids were precipitated by the addition of 300 µL of isopropanol. After centrifugation, as before, the pellet was rinsed with 500 µL of 70% (v/v) ethanol (EtOH). The pellet was air dried and resuspended in 100 µL of ultrapure H_2O. DNA samples were stored at -20 °C.

2.3.2. Isolation of plasmid DNA

2.3.2.1. *E. coli* miniprep protocol

A single colony of *E. coli* was used to inoculate 5 mL LB supplemented with antibiotic and incubated overnight at 37 °C and 200 rpm. 1.5 mL of the culture were added to a microtube and centrifuged at 15000 *g* for 2 min. The supernatant was discarded, and the pellet was resuspended in 100 µL of GTE solution I (50 mL glucose; 100 mM EDTA; 25 mM Tris-HCl pH 8.0). A GTE solution II was freshly prepared (100 µL 10% (w/v) SDS; 100 µL 1 M NaOH; 800 µL dH_2O) then 200 µL of this solution were mixed by inversion with the suspension. The mix was then supplemented with 150 µL of GTE

solution III (3 M $C_2H_3KO_2$; 5 M CH_3COOH) and mixed by inversion. The tube was incubated on ice for 15 min and then centrifuged for 15 min at 15000 *g*. The supernatant was retrieved and centrifuged as previously. Again, the supernatant was retrieved and mixed with 1 mL of 100% (v/v) EtOH. The tube was immediately centrifuged as previously, the supernatant discarded, and the pellet was rinsed in 50 µL of TE (10 mM Tris-HCl; 1 mM EDTA pH 8.0)- RNase (10 mg mL^{-1}). The mix was incubated for 5 min at 37 °C, vortexed and then incubated again at 37 °C for 10-15 min. The mix was supplemented with 30 µL 20% (v/v) PEG 4000; 2.5 M NaCl, vortexed and incubated on ice for 1-5 h. The mix was then centrifuged as previously, the supernatant discarded, and the tube walls were washed slowly with 500 µL 70% (v/v) EtOH, without agitation. The tube was centrifuged again as previously, the supernatant carefully removed, and when all the EtOH evaporated the DNA pellet was dissolved in 30 µL ultrapure H_2O. The DNA was stored at -20 °C.

2.3.3. DNA concentration estimation

DNA concentration was estimated by UV spectrophotometry using the *Nanodrop* (NanoDrop® ND-1000 UV-Vis Spectrophotometer, Thermo Fisher Scientific) at A_{260}. It was assumed that one unit of absorbance at 260 nm corresponds to 50 µg mL^{-1} of DNA. The DNA sample quality was evaluated by the A_{260}/A_{280} and A_{260}/A_{230} ratios, with the expected values of 1.8 and 2-2.2, respectively, for pure DNA samples.

2.3.4. Agarose gel electrophoresis

DNA samples were separated by electrophoresis when loaded onto an agarose gel, submerged in 0.5x TAE buffer (0.02 M Tris; 95 mM acetic acid; 50 Mm $EDTA.Na_2$, pH 8.0), after addition of 5x loading buffer (20% Ficoll; 0.3% (w/v) Tartrazine; 125 mg mL^{-1} Xylene Cyanol). Polymerase chain reactions (PCR) reactions performed with *NZYTaq II 2× Green Master Mix* (Nzytech) were applied directly onto the gel without addition of loading buffer. *NZYDNA Ladder III* (Nzytech) was used as molecular weight marker. For PCR products and DNA purifications, gels were prepared using 1% (w/v) agarose. Gels were run at approximately 100 V until the yellow colour of the xylene cyanol dye migration reached the end of the gel.

To visualize the DNA, *Green Safe Premium* (Nzytech) (0.5x final concentration) was added to the agarose gel solution before solidification. DNA was visualised on an UV transilluminator and photographed.

2.3.5. Polymerase chain reaction methods for DNA amplification

The PCR amplification was performed in the *MJ Mini Gradient Thermal Cycle* or the *T100™ Thermal Cycler* (Bio-Rad) thermo cyclers. The designed primers are represented in Table B1- Annex B.

2.3.5.1. Amplification of DNA fragments from transgenic plants

To genotype transgenic *Marchantia* plants, total PCR reaction mixtures of 10 µL were set up: *NZYTaq II 2× Green Master Mix*, 0.1-0.5 µM of forward and reverse primer and 50-100 ng genomic DNA (gDNA). PCR was carried out using the following settings: initial denaturation at 95 °C for 3 min; 35 cycles of 95 °C for 45 s, 55-58 °C annealing for 45 s and an extension step of 72 °C for 1-3 min; and one extension cycle at 72 °C for 5 min. The PCR products were stored at -20 °C or at 4 °C.

2.3.5.2. Amplification of DNA fragments for Gateway® cloning technology

The amplification of DNA fragments for Gateway® cloning technology was carried out using the *XpertHighFidelity DNA Polymerase* (GRiSP) or *NZYTaq II 2× Green Master Mix*. A twostep PCR protocol was performed to avoid the use of primers with high number of non-homologous nucleotides. The first reaction was performed to amplify the target sequence and the second reaction was performed to introduce the complete *att*B sequence.

2.3.5.2.1. Amplification of DNA fragments with *XpertHighFidelity DNA Polymerase*

The *XpertHighFidelity DNA Polymerase* was used to amplify coding sequences from reversed transcribed RNA solutions (dilution 1:20). In the first PCR reaction, a total reaction mixture of 25 µL was set up using 5x *XpertHighFidelity DNA Polymerase* PCR buffer, 0.4 µM of each forward and reverse primers, approximately 100 ng of reversed transcribed RNA and 2 U μL^{-1} *XpertHighFidelity DNA Polymerase* (0.05-0.10 U μL^{-1} final concentration). PCR was carried out with an initial denaturation step at 95 °C for 3 min; 35 cycles of 95 °C for 15 s, 58 °C for 15 s and 72 °C for 15 to 45 s and a final extension step at 72 °C for 3 min. The PCR products were stored at -20 °C or at 4 °C. To perform the second PCR reaction, a total reaction mixture of 25 µL containing 5x *XpertHighFidelity DNA Polymerase* PCR buffer, 0.4 µM of each *att*B adaptor (*att*B1 and *att*B2, Table B1- Annex B), 10 µL of PCR product and 2 U μL^{-1} *XpertHighFidelity DNA Polymerase* (0.05-0.10 U μL^{-1} final concentration). PCR was carried out with an initial denaturation step at 95 °C for 3 min; 5 cycles of 95 °C for 15 s, 45 °C for 15 s and 72 °C for 15 to 45 s; 15 cycles of 95 °C for 15 s, 55 °C for 15 s and 72 °C for 15 to 45 s; and a final

extension step at 72 °C for 5 min. *att*B PCR products were purified as described in sections 2.3.6.2 and 2.3.6.4.

2.3.5.2.2. Amplification of DNA fragments with *NZYTaq II 2× Green Master Mix*

The *NZYTaq II 2× Green Master Mix* was used to amplify target sequences from plant gDNA. In the first PCR reaction, a total reaction mixture was set using *NZYTaq II 2× Green Master Mix*, 0.1-0.5 µM of forward and reverse primer and 50-100 ng gDNA. PCR was carried out at 95 °C of initial denaturation for 3 min; 35 cycles of 95 °C for 45 s, 59-60 °C annealing for 45 s and an extension step of 72 °C for 2-5 min; and one extension cycle at 72 °C for 5 min. The PCR products were stored at -20 °C or at 4 °C. In the second PCR reaction, a total reaction mixture of 50 µL was set up using *NZYTaq II 2× Green Master Mix*, 0.1-0.5 µM of each *att*B adaptor (*att*B1 and *att*B2, Table B1- Annex B) and 10 µL of PCR product. PCR was carried out with an initial denaturation step at 95 °C for 3 min; 5 cycles of 95 °C for 45 s, 45 °C for 45 s and 72 °C for 2 to 5 min; 15 cycles of 95 °C for 45 s, 55 °C for 45 s and 72 °C for 2 to 5 min; and a final extension step at 72 °C for 5 min. *att*B PCR products were purified has described in sections 2.3.6.2 and 2.3.6.4.

2.3.5.3. Colony PCR amplification

Colony PCR amplification was used to confirm the presence of plasmid inserts in transformed *E. coli* or *A. tumefaciens*. A bacterial colony was resuspended in ultrapure H_2O and used as DNA template. A total reaction mixture of 25 µL were set up using 5x *GoTaq® Flexi Buffer* (Promega), 25 mM $MgCl_2$, 10 mM dNTPS, 0.2 µM of each forward and reverse primers and 0.5 µL *Taq* DNA Polymerase. PCR was carried out with an initial denaturation step at 95 °C for 15 min; 35 cycles of 95 °C for 45 s, 55 °C annealing for 45 s and an extension step of 72 °C for 45 s to 5 min; and one final extension cycle at 72 °C for 5 min. For large fragments, the colony PCR was performed with *NZYTaq II 2× Green Master Mix* with the conditions described in section 2.3.5.1.

2.3.6. DNA purification methods

Distinct DNA purification methods were used to purify PCR products.

2.3.6.1. Isopropanol purification

An isopropanol based purification method was used to remove contaminants from DNA samples prior to sequencing. In a microtube, 10 µL PCR product were mixed with 400 µL 75% (v/v) isopropanol.

The tube was incubated at room temperature for 15 min and centrifuged for 30 min at 6500 *g*. The supernatant was removed; the tube was inverted onto absorbent paper and centrifuged for 1 min at 430 *g* to remove the excess liquid. Then, 150 µL 75% (v/v) isopropanol were added to the tube and it was centrifuged as before to precipitate the nucleic acids. The liquid was removed by inverting the tube into an absorbent paper. This last step was repeated one more time. The pellet was air dried for 10 min ant the DNA was dissolved in ultrapure H_2O. The DNA quality and quantity were assessed as described in sections 2.3.3 and 2.3.4.

2.3.6.2. PEG purification protocol

The PEG purification protocol was used to purify the *att*B PCR products (see section 2.3.5.2). To 50 µL of *att*B-PCR product, 150 µL of TE buffer and 100 µL of 30% (w/v) PEG 8000/30 mM $MgCl_2$ were added. All the components were mixed by vortex and centrifuged 15 min at 10000 *g*. The supernatant was carefully removed, and the pellet was dissolved in 50 µL of TE. The DNA quality and quantity were assessed as described in section 2.3.3.

2.3.6.3. Phenol/Chloroform DNA extraction method

The phenol/chloroform purification method was used to remove contaminants from DNA samples. An equal volume of chloroform was added to the DNA sample. The tube was vortexed vigorously and centrifuged for 10 min at 13000 *g*. The aqueous phase was transferred to a new tube with an equal volume of isopropanol and one tenth of a volume of 3 M sodium acetate being added to the mixture. The tube was mixed and incubated for 1 h at -20 °C and centrifuged 15 min at 13000 *g*. The supernatant was discarded, and the pellet washed with 200 µL of 70% (v/v) EtOH. The mixture was centrifuged as described above and the supernatant discarded. The tube was centrifuged to remove the excess EtOH and the pellet was air dried for 10 min and resuspended in the appropriate volume of ultrapure H_2O. The DNA quality and quantity were assessed as described in section 2.3.3.

2.3.6.4. *Wizard® SV Gel and PCR Clean-Up System*

The *Wizard® SV Gel and PCR Clean-Up System* (Promega) kit was used in the extraction and purification of DNA fragments from agarose gels and it was performed according to the manufacturer's instructions.

2.3.7. Cloning procedures

2.3.7.1. Cloning into pMpGWB

The pMpGWBs are a series of Gateway Binary Vectors developed to simplify molecular analyses in *M. polymorpha* (Ishizaki et al., 2015). The Gateway® Technology is a universal cloning method based on the site-specific recombination properties of bacteriophage lambda. This technology provides a rapid and highly efficient way to transfer DNA sequences into multiple vector systems. The BP and LR recombination reactions are the basis of this technology. Two pMpGWB vectors were used- pMpGWB103 (figure C1- Annex C) and pMpGWB104 (figure C2- Annex C). Both vectors were obtained from Justin Goodrich's laboratory at the University of Edinburgh, United Kingdom.

2.3.7.1.1. Cloning into pMpGWB103

Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* coding sequences (Annex D) were cloned using Gateway® technology into pMpGWB103 vector with the aim of overexpressing these genes in *M. polymorpha*. Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* coding sequences of the different transcripts were amplified as described in section 2.3.5.2.1 using primers listed in Table B1- Annex B.

2.3.7.1.1.1. BP reaction

To perform the BP reaction, an *att*B substrate and a donor vector are required to create an entry clone. In this case, an *att*B-PCR product and the pDONR™201 (figure C3- Annex C) were used in a reaction catalised by *BP Clonase* enzyme mix (Invitrogen). To have an equimolar amount of *att*B-PCR product and pDONR™201, the amount of *att*B-PCR product was calculated by the following formula: ng (*att*B-PCR product) = fmol x N (size of the DNA in bp) x (660 fg / fmol) x (1 ng / 10^6 fg). BP reactions of the two Mp*DIV1* (Mp*DIV1.1* and Mp*DIV1.2*) and Mp*DRIF* (Mp*DRIF.1* and Mp*DRIF.*2) transcripts and Mp*DIV2* were performed for a total volume of 10 µL. Each one of the reactions contained pDONR™201 (50 fmol), Mp*DIV1.1,* Mp*DIV1.2*, Mp*DIV2*, Mp*DRIF.1* and Mp*DRIF.2 att*B products (50 fmol), 1 µL of *BP Clonase* enzyme mix and TE buffer to 10 µL. Reactions were incubated at 25 °C for 2 h and the plasmids introduced into *E. coli* DH10β competent cells (see 2.2.1.2). The cells were grown in LB-agar plates with 50 µg mL^{-1} kanamycin. A colony PCR was carried out with pDONOR specific primers. Positive colonies were selected, and the plasmids isolated (see 2.3.2.1) and sent for sequencing.

2.3.7.1.1.2. Cloning of Mp*DIVL* coding sequence into pENTRY/D-TOPO

To generate an entry clone with the Mp*DIVL* coding sequence, pENTRY/D-TOPO (figure C4- Annex C) was used as donor vector because the BP reaction with pDONR™201 did not work.

2.3.7.1.1.2.1. Mp*DIVL* PCR product digestion

Mp*DIVL* coding sequence was amplified as described in 2.3.5.2.2 from reversed transcribed RNA. The PCR fragment was gel purified, as described in section 2.3.6.4, and digested with NotI and AscI endonucleases restriction enzymes in two different steps. First, 10 µL of the gel purified PCR product were mixed with 5 µL of 10x Cutsmart buffer (NEB), 5 U μL^{-1} AscI (NEB) and 34 µL of ultrapure H_2O. This mixture was incubated at 37 °C for 30 min. Then, 1 µL of 5 M NaCl and 10 U μl^{-1} of NotI (FastDigest) were added to the digestion mix and this was incubated at 37 °C for 1 h. The digestion product was purified as described in section 2.3.6.4.

2.3.7.1.1.2.2. pENTRY/D-TOPO digestion

pENTRY/D-TOPO was linearized using NotI and AscI endonucleases restriction enzymes using the digestion protocol described in the previous section. The digestion product was purified as described in section 2.3.6.4.

2.3.7.1.1.2.3. Ligation and transformation

The ligation reaction was carried out using 100 ng of digested pENTRY, 50 ng of digested Mp*DIVL* PCR product, 10x T4 DNA ligase buffer (Invitrogen), 1 U μL^{-1} T4 DNA ligase (Invitrogen) in a 20 µL reaction. The reaction was incubated at 25 °C for 2 h and the constructs introduced into *E. Coli* DH10β competent cells (see section 2.2.1.2). The transformation mixture was plated in LB-agar with 50 µg mL^{-1} kanamycin. A colony PCR was carried out with specific primers. Positive colonies were selected, and the plasmids isolated (see 2.3.2.1).

2.3.7.1.1.3. LR reaction

In the LR reaction, an entry clone is recombined with a destination vector to create an expression clone. This reaction is catalysed by a *LR clonase* enzyme mix (Invitrogen). LR reactions of Mp*DIV1.1*-pDONOR, Mp*DIV1.2*-pDONOR, Mp*DIV2*-pDONOR, Mp*DIVL*-pENTRY, Mp*DRIF.1*-pDONOR and Mp*DRIF.2*-pDONOR with pMpGWB103 destination vector were performed for a total volume of 10 µL. Each one of the reactions contained pMpGWB103 (50 fmol), Mp*DIV1.1*-pDONOR, Mp*DIV1.2*-pDONOR, Mp*DIV2*-

pDONOR, Mp*DIVL*-pENTRY, Mp*DRIF.1*-pDONOR or Mp*DRIF.2*-pDONOR (50 fmol), 1 µL of *LR Clonase* enzyme mix and TE buffer to 10 µL. Reactions were incubated at 25 °C for 2 h and the constructs transformed into *E. coli* DH10β competent cells (see 2.2.1.2). The colonies were grown in plates of LB-agar with 100 µg mL^{-1} spectinomycin. A colony PCR was carried out with pMpGWB103 appropriate primers. Positive colonies were selected, and the plasmids isolated (see 2.3.2.1).

2.3.7.1.2. Cloning into pMpGWB104

Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* predicted promoter sequences (Annex E) were cloned using *Gateway® technology* into pMpGWB104 vector. Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* promoter sequences were amplified as described in section 2.3.5.2.2 using primers listed in Table B1- Annex B.

2.3.7.1.2.1. BP reaction

BP reactions of *pro*Mp*DIV1, pro*Mp*DIV2*, *pro*Mp*DIVL, pro*Mp*DRIF.1* and *pro*Mp*DRIF.2* were performed as described in section 2.3.7.1.1.1 with an overnight incubation of the different reactions.

2.3.7.1.2.2. LR reaction

LR reactions of *pro*Mp*DIV1*-pDONOR, *pro*Mp*DIV2*-pDONOR, *pro*Mp*DIVL*-pDONOR, *pro*Mp*DRIF.1*-pDONOR and *pro*Mp*DRIF.2*-pDONOR with pMpGWB104 destination vector were performed as described in section 2.3.7.1.1.3 with an incubation time of 3 h. The colony PCR was carried out with pMpGWB104 appropriate primers.

2.3.7.2. Cloning for CRISPR/Cas9 technology

The CRISPR/Cas9 technology is based on a gRNA that binds to a target sequence by base complementarity (Sander & Joung, 2014). The target recognition is dependent on a protospacer adjacent motif (PAM), an NGG sequence adjacent to the 3' end of the target sequence. After the recognition, the gRNA anneals with the target sequence and the Cas9 causes a double-strand break.

2.3.7.2.1. gRNA design

For each one of the gene sequences targeted in this study different gRNAs were design with the aim of causing deletions/insertions at different sites of the sequence. To obtain large deletions, insertions or inversions, a strategy with the combination of a two gRNAs was used for each gene. The gRNAs were designed using informatic tools [Web-based CRISPR off-target search for *Marchantia* (marchantia.focas.ayanel.com); CRISPRdirect (crispr.dbcls.jp)] that avoid off -targets in the genome and

that indicate the gRNA specificity. The gRNAs designed in the beginning of the gene are the left gRNAs (L) and the ones designed in the end of the gene are the right gRNAs (R). For each one of the gRNAs two oligos were designed, one sense and other antisense. The gRNAs are described in Table B2- Annex B.

2.3.7.2.1.1. Annealing of oligos

For the annealing of the gRNA oligos, 50 pmol of the sense and antisense oligos were mixed in 10 µL of TE buffer. The annealing was carried out using a thermal cycler with the following program: an initial 95 °C heating step for 5 min; a cooling step from 95 to 85 °C with a decreasing of 2 °C/ s; a step at 85 °C for 1 min; a cooling step from 85 to 25 °C with a decreasing of 0.2 °C/ s and a final hold step at 25 °C.

2.3.7.2.2. Cloning into pMpGE_En04

The pMpGE_En04 (figure C5- Annex C) vector was used as an entry vector to clone the annealed oligos of the different gRNAs. This vector was obtained from Sabine Zachgo's laboratory.

2.3.7.2.2.1. pMpGE_En04 digestion

The pMpGE_En04 digestion was performed with the BsaI-HF (NEB) restriction enzyme. For a reaction with a final volume of 10 µL, 2-3 µg of pMpGE_En04 vector, 10-20 units of BsaI-HF and 2 µL of 10x CutSmart buffer (NEB) were mixed and incubated at 37 °C for 1-2 h. The digestion was separated by electrophoresis, as described in section 2.3.4, to confirm the size of the fragment and then gel purified as described in section 2.3.6.4.

2.3.7.2.2.2. Ligation and transformation

The ligation reaction was set up as described in section 2.3.7.1.1.2.3 using 30-50 ng of the digested vector, 10 pmol of annealed gRNA oligo, T4 ligase buffer (Invitrogen) and 1 U μL^{-1} T4 ligase (Invitrogen) in a final volume of 10 µL.

2.3.7.2.3. Cloning into pMpGE010 and pMpGE011

To express the gRNAs and the Cas9 in the same vector, the pMpGE010 (figure C6- Annex C) and pMpGE011 (figure C7- Annex C) (obtained from Sabine Zachgo's laboratory) were used as destination vectors. These binary vectors have the *Cas9* gene expressed under the strong promoter Mp*EF1α*. The pMpGE010 was used to clone the gRNAs L and the pMpGE011 was used to clone the gRNAs R.

2.3.7.2.3.1. LR reaction

To obtain a final construct with the gRNA, a Gateway LR reaction was performed between the entry vector obtained in the section 2.3.7.2.2 and the each one of the destination vectors. LR reactions of gRNAs L with pMpGE010 and gRNAs R with pMpGE011 were performed in a total volume of 10 µL as described in section 2.3.7.1.1.3. The transformation was grown in plates of LB-agar with 100 µg mL^{-1} spectinomycin and a colony PCR was carried out with appropriate primers. Positive colonies were selected and the plasmids isolated (see 2.3.2.1).

2.4. RNA methods

2.4.1. RNA extraction

For *M. polymorpha* RNA extraction, different protocols were used to increase RNA quality and quantity.

2.4.1.1. TRIzol protocol

M. polymorpha tissues of thallus sections with 5-10 cm^2 and 3-5 sexual structures in different stages of development were grinded with liquid nitrogen and 500 µL *TRIzol* (Invitrogen) were added to the powder. The mixture was grinded again and the liquid was transferred to an microtube. The tube was vortexed and incubated at room temperature for 5 min. Then, 200 µL of chloroform were added, the tube vortexed and incubated at room temperature for 10 min. The tube was centrifuged 10 min at 12000 *g* and 4 °C. The aqueous phase was transferred into a new tube containing 500 µL of isopropanol. The tube was incubated at room temperature for 10 min, inverted several times, and then centrifuged 10 min at 16000 *g* at 4 °C. The supernatant was discarded and 1 mL 75% (v/v) EtOH was added to the pellet. The tube was centrifuged 1 min at 16000 *g* and 4 °C, the supernatant discarded, and the tube was recentrifuged to remove the excess EtOH. This last step was repeated two times, then the pellet was air dried for 5 min and 35 µL diethyl pyrocarbonate (DEPC)-treated H_2O were added. The RNA solution was stored at -80 °C.

2.4.1.2. CTAB protocol

The following protocol was adapted from Azevedo et al. (2003). The tissue of *M. polymorpha* whole plants was grinded with liquid nitrogen and 700 µL of cetrimonium bromide (CTAB) extraction buffer [0.1 M Tris-HCl pH 8.0; 30 mM EDTA; 2 M NaCl; 2% (w/v) CTAB; 2% (w/v) polyvinylpyrrolidone

(PVP-40)] were added to the powder. The mixture was transferred to an microtube, vortexed and incubated at 65 °C for 10 min. One volume (700 µL) of chloroform: isoamilic acid (24:1) was added to the tube and the mixture was vortexed and centrifuged for 15 min at 16900 *g* and 4 °C. The supernatant was transferred for a new tube. One volume (700 µL) of chloroform: isoamilic acid (24:1) was added and the tube vortexed and centrifuged as before. The supernatant was again retrieved and ¼ of volume (175 µL) 8 M LiCl was added. The samples were incubated at 4 °C overnight and centrifuged for 25 min at 16900 *g* at 4 °C. The supernatant was removed, and the pellet washed with one volume (700 µL) of 2 M of LiCl on ice. The samples were again centrifuged 25 min at 16900 *g* at 4 °C. The supernatant was removed, and the pellet resuspended in 30 µL DEPC-treated H_2O. The RNA was stored at -80 °C.

2.4.2. RNA concentration estimation

RNA concentration was estimated by UV spectrophotometry using the *Nanodrop* at A_{260}. It was assumed that one unit of absorbance at 260 nm corresponds to 40 µg mL^{-1} of RNA. The RNA sample quality was evaluated by the A_{260}/A_{280} and A_{260}/A_{230} ratios, with the expected values of 2 and 2-2.2, respectively, for pure RNA samples.

2.4.3. DNase treatment

The RNA solutions were treated with DNase to prevent the sample contamination with genomic DNA. The *TURBO DNA-free™ Kit* (Invitrogen) was used by the following steps: 0.5 µL of TURBO DNase and 2 µL of 10X TURBO DNase Buffer were added to 30 µL of RNA sample. The sample was incubated 30 min at 37 °C. Then, 2 µL of DNase Inactivation Reagent were added to the sample. The sample was incubated 5 min at room temperature and centrifuged 1 min at 10000 *g*. The supernatant was retrieved to another tube and the RNA concentration was estimated as described in section 2.4.2.

2.4.4. cDNA synthesis

cDNA was synthesised using the *SuperScript® IV Reverse Transcriptase* (Invitrogen) kit. In the first step, the $Oligod(T)_{20}$ primer was annealed to the template RNA and for that, a reaction with 200-1000 ng μL^{-1} of RNA, 0.5 µL 10 mM dNTPs and 0.5 µL Oligo $d(T)_{20}$ was performed. This reaction was heated at 65 °C for 5 min and then incubated at 4 °C for 1 min. Then, 3.5 µL of RT reaction mix (2 µL 5× SSIV Buffer; 0.5 µL 100 mM DTT; 0.5 µL RNaseOUT™ Recombinant RNase Inhibitor; 0.5 µL SuperScript® IV Reverse Transcriptase (200 U μL^{-1}) were added to the annealed RNA. This reaction was incubated at 50 °C for 15 min and then at 85 °C for 5 min. The cDNA was stored at -20 °C.

2.4.5. Reverse transcriptase PCR

The Reverse transcriptase PCR (RT-PCR) was used to amplify the transcripts coding sequence of the different genes in study, and the constitutive gene *Elongation Factor 1α* (Mp*EF1α*) from *M. polymorpha* cDNA (see 2.3.5.1), with specific primers listed in Table B1- Annex B. A total reaction mixture of 25 µL was set up using 1 µL cDNA (dilution 1:10), 5x *GoTaq® Flexi Buffer* (Promega), 25 mM $MgCl_2$, 10 mM dNTPS, 0.2 µM of each forward and reverse primers, 0.5 µL *Taq* DNA Polymerase. PCR was carried out with an initial denaturation step at 95 °C for 3 min; 35 cycles of 95 °C for 45 s, 58-60 °C annealing for 45 s and an extension step of 72 °C for 30 s; and one final extension cycle at 72 °C for 5 min. The transcripts abundance was analysed by electrophoresis (see 2.3.4).

2.4.6. Real-time quantitative PCR

RT-qPCR was used to estimate the levels of gene expression. cDNA was synthesized from RNA samples as described in section 2.4.4. cDNA was amplified using *SsoFast™ EvaGreen® Supermix* (Bio-Rad), 250 mM of each gene-specific primer (listed in Table B1- Annex B) and 1 µL of cDNA (1:25 dilution). RT-qPCR reactions were performed in triplicates on the *CFX96 Touch™ Real-Time PCR Detection System* (Bio-Rad) using the following settings: an initial step of 3 min at 95 °C; 40 cycles of a denaturation step of 10 s at 95 °C and an annealing/extension step of 10 s at 60 °C. With each PCR reaction, a melting curve was obtained to check for amplification specificity and reaction contaminations, by heating the amplification products from 60 °C to 95 °C in 5 s intervals. Gene expression analysis was conducted using three biological replicates, and normalized with the reference gene Mp*EF1α* (Saint-Marcoux et al., 2015).

2.5. Whole mount *in situ* hybridization

The whole mount *in situ* hybridization technique was used to analyse the expression pattern of the genes in study.

2.5.1. *In situ* probe cloning

DNA probes were synthetized to hybridize with a known mRNA target sequence. The target sequences were amplified from plasmids that contained the desired sequences using primers listed in Table B1- Annex B. A 10 µL PCR reaction was performed using *NZYTaq II 2× Green Master Mix,* 0.1-0.5 µM of forward and reverse primer and 50-100 ng of plasmid DNA. PCR was carried out at 95 °C of initial

denaturation for 3 min; 35 cycles of 95 °C for 45 s, 59-55 °C annealing for 30 s and an extension step of 72 °C for 1 min; and one extension cycle at 72 °C for 5 min. The PCR products were purified as described in section 2.3.6.3 and stored at -20 °C or at 4 °C. The PCR amplification products were then cloned in the *pGEM®-T easy* (Promega) (figure C8- Annex C) vector. The ligation reaction was performed as described in section 2.3.7.1.1.2.3 using 50-100 ng of purified PCR product, 0.5 µL of *pGEM®-T easy*, 1 µL 10x T4 DNA ligase buffer (Invitrogen), 1 U μL^{-1} T4 DNA ligase (Invitrogen) in a total volume reaction of 10 µL. The constructs were transformed into *E. coli* XL-1 blue as described in section 2.2.1.2. The transformation was grown in plates of LB-agar with 100 µg mL^{-1} ampicillin, 80 µg mL^{-1} 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) and 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The colonies were screened by colony PCR (see 2.3.5.3) using specifics primers for the M13 region and for the gene (Table B1- Annex B).

2.5.2. Probe template synthesis

Plasmids were isolated (see 2.3.2.1) and a PCR was performed with primers for the M13 regions (Table B1- Annex B). Total reaction mixtures of 10 µL were set up using *NZYTaq II 2× Green Master Mix* and 0.1-0.5 µM of forward and reverse primer. PCR was carried out with an initial denaturation step at 95 °C for 15 min; 35 cycles of 95 °C for 45 s, 55 °C annealing for 30 s and an extension step of 72 °C for 1 min; and one final extension cycle at 72 °C for 5 min. The PCR products were purified as described in section 2.3.6.3. The synthetized template probes were stored at -20 °C.

2.5.3. Probe synthesis

A 25 µL transcription reaction mix with 10x transcription buffer (pre-warmed at 37 °C) (Roche); 5 mM ribonucleotides (Roche); 1 mM DIG-UTP (Roche); 1 µL RNase out (40 U μL^{-1}) (Invitrogen); 20 U μL^{-1} T7 or SP6 polymerase (Roche) and 20-100 ng probe was incubated at 37 °C for 2 h 30 min. Then, 1 µL of TURBO DNase (Invitrogen) and 0.2 µL of 100 mg mL^{-1} yeast tRNA were added to the mix and this was incubated again at 37 °C for 30 min. The probes were precipitated with 25 µL 3.8 M ammonium acetate (pH 4.8) and 150 µL 100% (v/v) EtOH and stored at -20 °C overnight. In the next day, the probes were centrifuged for 20 min at 18000 *g* and 4 °C. The supernatant was discarded, and the pellet washed with 200 µL 100% (v/v) EtOH. The tube was centrifuged as described above for 10 min, the supernatant was again discarded, and the pellet dissolved in 50 µL dH_2O. To increase accessibility of the probe to the tissue, the probes were subjected to alkaline hydrolysis. A 50 µL volume of 2x carbonate buffer (80 mM $NaHCO_3$; 120 mM Na_2CO_3; pH 10.2) was added to the resuspended probe and the mixture incubated at

60 °C for 80 min (80-90 min for probes with 800-1200 bp). After hydrolysis, the probes were precipitated by the addition of 10 µL 10% (v/v) acetic acid, 12 µL 3 M sodium acetate (pH 4.8) and 312 µL of ice-cold 100% (v/v) EtOH. The mixture was incubated at -80 °C for 45 min and centrifuged as described above. The supernatant was discarded, and the pellet washed with 70% (v/v) EtOH and centrifuged as described above for 10 min. The pellet was air dried and resuspended in 50 µL DEPC-H_2O. The probes were stored at -20 °C.

2.5.4. Hybridization

The *in situ* hybridization protocol used was adapted from Hejátko et al., (2006). In the first day, a fresh paraformaldehyde solution was prepared by mixing 20 g of paraformaldehyde with 100 mL of dH_2O. To dissolve this fixative, 3 pellets of NaOH were added to 100 mL of solution and this was heated to 60 °C. When dissolved, the pH was adjusted to 7.0 with HCl and the solution was diluted to 4% (w/v) in phosphate buffered saline (PBS) supplemented with 0.1% (v/v) Tween 20 and 0.1% (v/v) Triton X-100. Then, plants with 7 days were immersed in the fixative and vacuumed for 15 min at room temperature. When all the plants sank, 2 washes with 100% (v/v) methanol for 5 min and 2 washes with 100% (v/v) EtOH for 5 min were performed and the samples stored overnight at -20 °C in 100% (v/v) EtOH. In the next day, the EtOH was removed and replaced by a 100% (v/v) EtOH/histoclear (1:1) mixture and the samples were incubated 30 min at room temperature. They were then washed two times with 100% (v/v) EtOH for 5 min at room temperature; 75% (v/v) EtOH in dH_2O for 10 min at room temperature; 50% (v/v) EtOH in PBS for 10 min at room temperature; 25% (v/v) EtOH in PBS for 10 min at room temperature; 4% (w/v) paraformaldehyde (fresh and cooled) for 20 min at room temperature and 2x PBS-T for 10 min at room temperature. The samples were placed in a hybridization solution [50% (v/v) formamide/ 2x SSC (3 M NaCl; 300 mM sodium citrate)] for 10 min at room temperature and then transferred to a new hybridization solution for 1 h at 55 °C with gentle shaking. The samples were placed in a hybridization solution supplemented with 0.1 mg mL^{-1} herring sperm and 20-100 ng mL^{-1} probe (denatured before use at 70 °C) for 16 h at 55 °C. In the third day, samples were washed: three times for 10, 60 and 20 min with 50% (v/v) formamide/ 2x SSC/ 0.1% (v/v) Tween-20 at 55 °C; 2x SSC/ 0.1% (v/v) Tween-20 for 20 min at 55 °C; twice with 0.2x SSC/ 0.1% (v/v) Tween-20 for 20 min at 55 °C; three times with PBS-T for 10 min at room temperature. The samples were then pre-incubated in PBS-T with 1% (w/v) bovine serum albumin (BSA) for 90 min at room temperature and then in PBS-T with 1% (w/v) BSA supplemented with 1:2000 Anti-Digoxigenin-Alkaline Phosphatase (Anti-DIG-AP) (Roche) on dark, overnight at room temperature.

2.5.5. Probe detection

Samples were washed eight times in PBS-T for 20 min at room temperature and incubated twice in APL buffer [0.1 M Tris-HCl pH 9.5; 0.1 M NaCl; 0.1% (v/v) Tween-20] with 50 mM $MgCl_2$ (freshly added) for 10 min at room temperature. The AP staining reaction was performed by mixing the samples with ALP buffer, 2.25 µL mL^{-1} nitro blue tetrazolium (NBT) (Roche) (50 mg mL^{-1}) and 1.75 µL mL^{-1} 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) (Roche) (50 mg mL^{-1}) in dark at room temperature. The material was inspected every 30 min and the staining stopped with 100% (v/v) EtOH.

2.6. Generation of *Marchantia polymorpha* transgenic lines

Two different transformation method were used to obtain *M. polymorpha* transgenic lines. For the plant transformation with the vectors that carried the different gRNAs, an *Agrobacterium*-mediated transformation of spores was performed. To the plant transformation with the constructs obtained in the section 2.3.7.1 a thallus transformation method with *A. tumefaciens* was used.

2.6.1. Plant transformation

2.6.1.1. Spores transformation

The spore transformation was performed in Sabine Zachgo's laboratory. Before the transformation, mature sporangia were collected to obtain the spores. The sporangia surface was sterilized with a solution of 12% (v/v) NaClO, 10% (v/v) Triton X-100 and dried with silica gel and stored at 4 °C until use. Meanwhile, a single colony of *Agrobacterium* harbouring the desired construct was grown in 5 mL LB medium for 48 h at 28 °C and 170 rpm. Then, the spores from the sterilized sporangia were co-cultured with the *Agrobacterium* culture. For that, the spores were resuspended with ultrapure H_2O and 100 µL were transferred to a flask with 25 mL Gamborg B5 liquid medium and cultured under white light (60 µmol photons m^{-2} s^{-1}) at 22 °C for 7 days and 130 rpm. The *Agrobacterium* culture was then added to the liquid spore culture supplemented with 200 µM acetosyringone and co-cultivated at 22 °C and 130 rpm for 3 days. Then, spores were collected, washed three times with half strength Gamborg B5 medium and transferred to half strength Gamborg B5 medium with 1.4% (m/v) plant agar and the selectable markers 10 mg L^{-1} hygromycin, 100 mg L^{-1} cefotaxime and 0.5 µM chlorsulfuron.

2.6.1.2. Thallus transformation

A thallus transformation method adapted from Tsuboyama-Tanaka et al. (2015) was used to generate *M. polymorpha* transgenic plants. *M. polymorpha* wild-type plants were grown in the conditions described in section 2.1.1 for 1 month. A single colony of *A. tumefaciens* harbouring the construct was grown in 5 mL LB medium with the appropriate antibiotics at 30 °C and 200 rpm to an OD_{600} of 2. The 1-month-old plants were cut into 1 mm^2 pieces and placed in a 50 mL falcon with 20 mL of Inoculation Medium (Table 2) and 20 µL of 100 µM acetosyringone for each one of the transformations. Then, 1 mL of *A. tumefaciens* culture was added to the mixture for 30 min. The plant tissue was then transferred to a sterile filter paper and placed to dry for 30 min. The plant tissue was transferred for the co-culture medium (Table 2) and placed in dark and room temperature for 3 days. The plant tissue was transferred to selection medium (Table 2) and grown under white light (60 µmol photons m^{-2} s^{-1}) until differences were visible.

Table 2: Media used in the *Marchantia polymorpha* thallus transformation.

Inoculation medium pH = 5.2	▪ 1.5 g L^{-1} Gamborg B5 medium including vitamins ▪ 0.5 g L^{-1} MES ▪ 10 g L^{-1} sucrose (Fisher Scientific) ▪ 100 µM acetosyringone
Co-culture medium	▪ 1.5 g L^{-1} Gamborg B5 medium including vitamins ▪ 0.5 g L^{-1} MES ▪ 10 g L^{-1} sucrose ▪ 100 µM acetosyringone ▪ 1% (w/v) plant agar
Selection medium	▪ 1.5 g L^{-1} Gamborg B5 medium including vitamins ▪ 0.5 g L^{-1} MES ▪ 10 g L^{-1} sucrose ▪ 100 µM acetosyringone ▪ 1% (w/v) plant agar ▪ 100 mg L^{-1} Cefotaxime (Duchefa) ▪ 10 mg L^{-1} Hygromycin (Duchefa)

2.6.2. Transformants selection

2.6.2.1. Spores selection

Positive transformants were selected by observing their survival in half strength Gamborg medium supplemented with 10 mg L^{-1} hygromycin, 100 mg L^{-1} cefotaxime and 0.5 µM chlorsulfuron. The transformed spores were transferred to square plates (100x100 mm) with selection medium. After approximately 1-month, the positive transformants were asexually propagated for new selection medium.

2.6.2.2. Thallus selection

The small pieces of plant tissue were transferred 3 to 4 times into selection medium [1.582 g L^{-1} Gamborg B5 medium; 0.5 g L^{-1} MES; 10 g L^{-1} sucrose; 1.4% (w/v) plant agar; 100 µM acetosyringone; 10 mg L^{-1} hygromycin and 100 mg L^{-1} cefotaxime]. Positive transformants were selected by observing their survival in half strength Gamborg B5 medium supplemented with 10 mg L^{-1} hygromycin and 100 mg L^{-1} cefotaxime (green small pieces of thallus). Then, the thallus pieces regenerated until the gemma cups development and the positive transformants were asexually propagated in new selection media.

2.7. Fluorescence Microscopy

Bright field and fluorescence observations of gemmae from transgenic plants were performed on a *Leica DM5000B+CTR 5000+ebq 100* (Leica) fluorescence microscope.

3. Results

In *Antirrhinum*, the proteins of the DDR module- DIV, DRIF and RAD- are involved in the establishment of the flower asymmetry (Raimundo et al., 2013). In *M. polymorpha*, two copies of the DIV, one copy of DIVL (a similar DIV protein) and one copy of the DRIF were identified (Raimundo et al., 2018). To understand the ancestral function of the genes that codify these proteins, the spatial and temporal gene expression pattern was characterized by RT-qPCR, whole mount *in situ* hybridization and GUS reporter system. Also, overexpression and knockout lines were obtained using the Gateway and CRISPR/Cas9 technologies.

3.1. Spatial and temporal Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* expression

3.1.1. Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* expression in different tissues

Real-time quantitative polymerase chain reaction (RT-qPCR) is a sensitive and reproducible technique commonly used to measure and compare transcript accumulation levels. The detection of amplification products is based on fluorescence and it is achieved by the use of a DNA-binding dye or a hybridization probe. Samples from different *M. polymorpha* tissues at different developmental stages were collected in the field to analyse the genes expression. RNA was isolated using the TRIzol protocol from male and female thallus, antheridia and archegoniophore primordium (S1 stage), young antheridia and archegoniophore (S2 stage) and mature antheridia (S3 stage). cDNA was synthesised and a PCR reaction was performed using Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* specific RT-PCR primers (Table B1- Annex B). The relative quantification of Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* transcripts was obtained using Mp*EF1α* gene as internal reference (Saint-Marcoux et al., 2015).

Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* were expressed in all the tissues analysed, the exception being the S1 female stage in which Mp*DIV1* was not detected (figure 11). This analysis also suggests that Mp*DIV1*, Mp*DIV2* and Mp*DIVL* are more expressed in the thallus regardless of the sex, and that Mp*DRIF* appears to be more expressed in female sexual structures (figure 11).

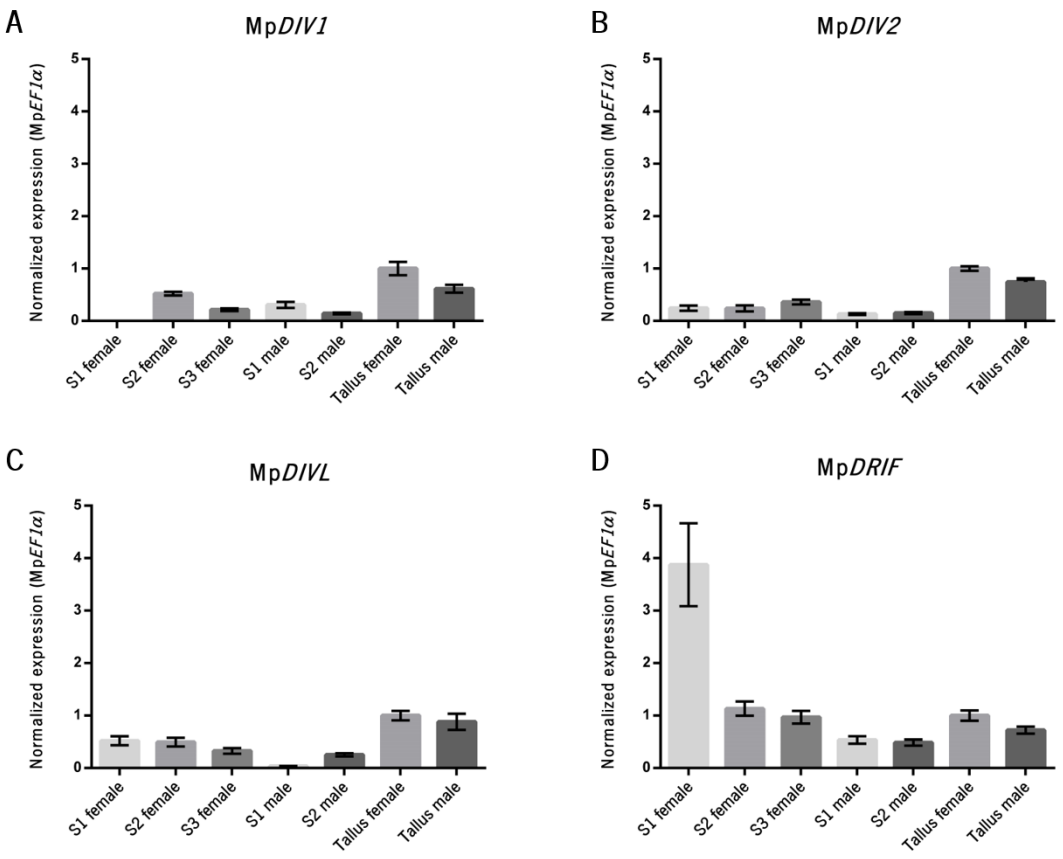

Figure 11: Relative expression of Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* in different female and male tissues. Gene expression analysis by RT-qPCR of *Marchantia polymorpha* **A)** *DIV1*, **B)** *DIV2*, **C)** *DIVL* and **D)** *DRIF* genes in female and male thallus, antheridia (male sexual structure) and archegoniophore (female sexual structure) primordium (S1 stage), young antheridia and archegoniophore (S2 stage) and mature antheridia (S3 stage). Errors bars indicate standart deviation of biological and technincal replicates. Mp*EF1α* was used as reference gene.

3.1.2. Whole mount *in situ* hybridization

In situ hybridization is a technique widely used to analyse gene expression at the tissue level. To obtain the probes, the complete gene coding sequences were amplified by PCR using specific primers (Table B1- Annex B). The PCR products were purified and then cloned in the *pGEM®-T easy* vector. The obtained constructs were transformed into *E. coli* XL-1 blue competent cells, positive colonies were selected by white/blue screening and further screened by colony PCR using a primer for the T7 promoter and a specific transcript primer (figure 12).

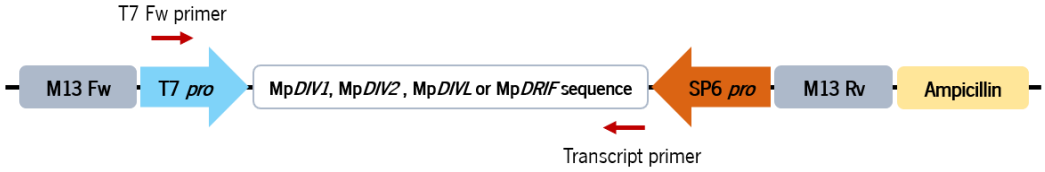

Figure 12: Hybridization localization of the primers used in the screening of Mp*DIV1*, Mp*DIV2*, Mp*DIVL* or Mp*DRIF*pGEM colonies by PCR. The T7 promoter forward (Fw) primer and a specific primer for the transcript sequence were used in the screening.

Mp*DIV1*pGEM clone 75, Mp*DIV2*pGEM clone 29, Mp*DIVL*pGEM clone 37 and Mp*DRIF*pGEM clone 1 showed fragments with the expected sizes (figure 13). The plasmids of these colonies were isolated and sequenced. All the constructs had the correct target sequence inserted.

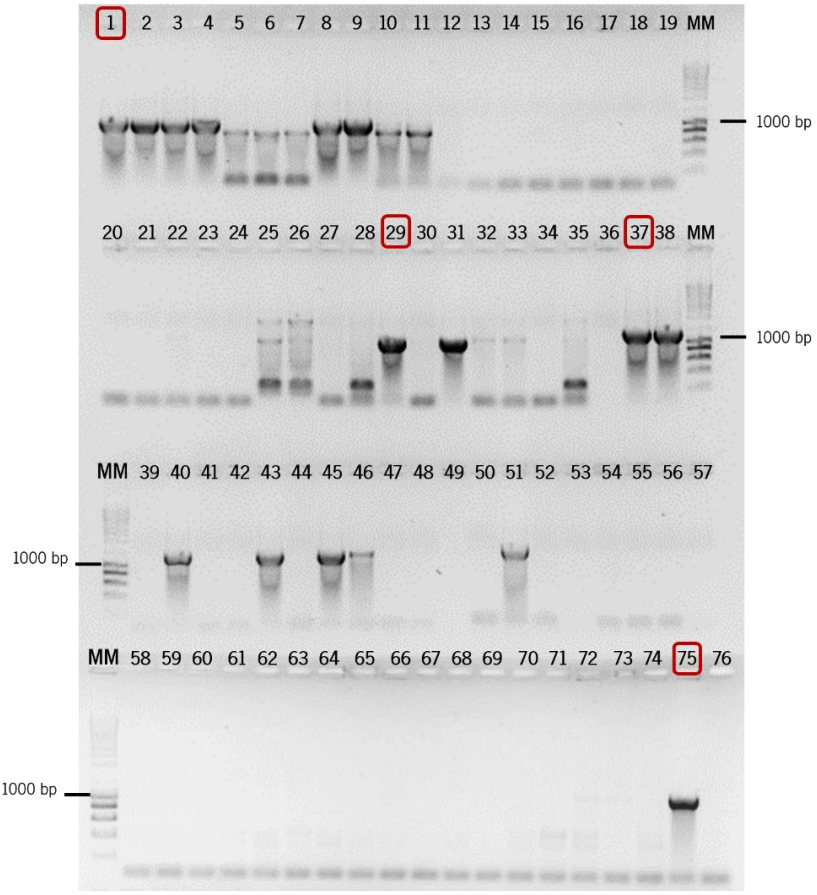

Figure 13: PCR screening of *E. coli* colonies containing constructs with Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* coding sequences in pGEM vector for *in situ* hybridization. The colonies were screened by PCR using specific primers for the T7 promoter and the transcript sequence.
Lanes 1 to 11: Mp*DRIF*pGEM colonies (expected size fragment of 916 bp);
Lanes 13 to 35: Mp*DIV2*pGEM colonies (expected size fragment of 936 bp);
Lanes 37 to 56: Mp*DIVL*pGEM colonies (expected size fragment of 1181 bp);
Lanes 58 to 75: Mp*DIV1*pGEM colonies (expected size fragment of 932 bp);
Lanes 12, 36, 57 and 76: Negative controls;
MM: Molecular marker (NZYDNA Ladder III, Nzytech);
The lanes with red squares correspond to the colonies selected for sequencing.

With the target sequences cloned in *pGEM®-T easy*, the sense and antisense probes, flanked by the T7 and SP6 polymerase-binding sites respectively, were synthesised by PCR amplification using the plasmids described above and specific primers for the M13 regions (figure 14). After confirming the correct size by PCR, the probes were hydrolysed and hybridized with the plant tissue (figure 15).

Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* genes are all expressed in the thallus stage (figure 11) thus, an *in situ* hybridization was carried out on *M. polymorpha* wild-type 7-days-old thallus to evaluate the spatial expression pattern of these genes. The spatial expression pattern was obtained by the deposit

of a purple product derived from the BCIP and NBT enzymatic catabolism. No hybridization signal was detected with the sense probe and some unspecific hybridization occurred with the antisense probe in the agar that was around the tissue (figure 15). Mp*DIV1* and Mp*DIV2* expression was detected in the thallus extremities; no signal was detected with the Mp*DIVL* antisense probe and Mp*DRIF* expression was observed in areas scattered along the thallus (figure 15).

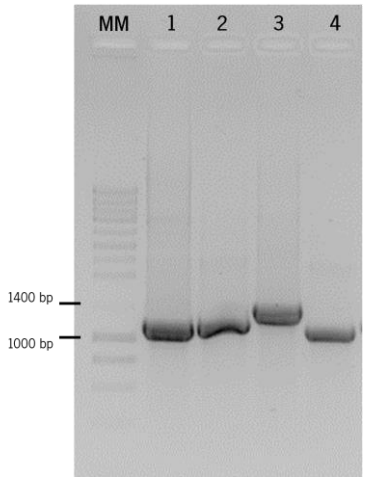

Figure 14: PCR amplification of antisense and sense probes. The *in situ* probes were synthesised using primers for the M13 regions.
Lane 1: Mp*DIV1* probe (expected size fragment of 1148 bp);
Lane 2: Mp*DIV2* probe (expected size fragment of 1152 bp);
Lane 3: Mp*DIVL* probe (expected size fragment of 1397 bp);
Lane 4: Mp*DRIF* probe (expected size fragment of 1132 bp);
MM: Molecular marker (NZYDNA Ladder III, Nzytech).

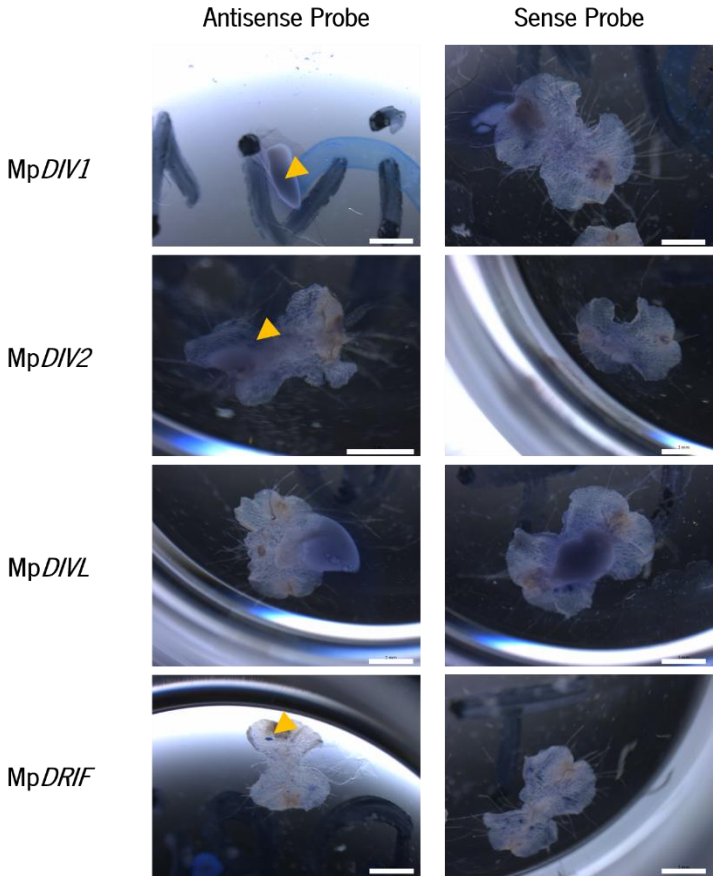

Figure 15: Whole mount *in situ* hybridization of *Marchantia polymorpha* wild-type thallus using DIG-labelled Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* probes. *Marchantia polymorpha* wild-type 7-days-old thallus hybridized with Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* probes. The purple colour (yellow arrow) corresponds to the transcript accumulation. Scale bar: 1 mm.

3.1.3. Spatial expression pattern using the GUS reporter system

The GUS reporter system is a very common tool in molecular biology that allows the analysis of a promoter activity. This system is based on the *E. coli* β-glucoronidase (GUS) enzyme, which transforms a specific substrate into a blue product within cells. The most common substrate in the GUS histochemical staining is the 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc), which is converted by the β-glucuronidase enzyme into a blue product (Jefferson et al., 1987).

3.1.3.1. Promoter sequence prediction

The first step in the study of Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* promoters was to predict the corresponding sequence. This prediction was necessary because the *M. polymorpha* promoters are not well annotated in the genomic sequence. Due to the difficulty that exists in cloning large fragments, an alignment was performed between the Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* genomic sequence (5'UTR, exons, introns and 3'UTR), the 5 kb sequence upstream the genomic sequences annotated in *Phytozome* and the previous gene sequence to exclude the common parts in order to reduce the predicted promoter fragment size that will be cloned.

The alignment showed a distance of: 517 bp between Mp*DIV1* 5'UTR and the previous gene sequence; more than 5 kb from the Mp*DIV2* or Mp*DIVL* genomic sequence to the previous gene sequence and 3575 bp between Mp*DRIF* and the previous gene sequence (figure 16).

Since no large overlaps were detected for Mp*DIV2,* Mp*DIVL* and Mp*DRIF*, the promoter sequences were predicted by selecting the 1000 bp upstream of the 5'UTR (Annex E). The 5'UTR was also included in the promoter sequence since this region can comprise regulatory elements (Hixson et al., 1996; Klein et al., 1998; Reynolds et al., 1996). For Mp*DRIF*, two promoter sequences were predicted, one for each transcript, due to the differences that exist between the 5'UTR sequences (figure 16).

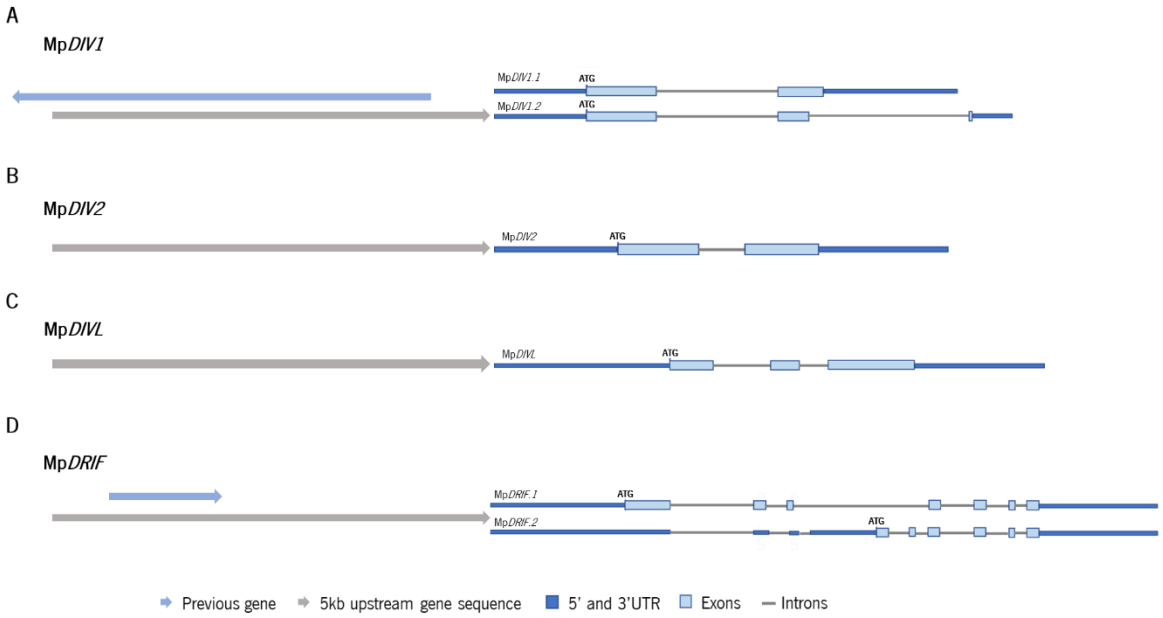

Figure 16: Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* prediction of promoter sequences. The genomic sequence of the different Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* transcripts, the genomic sequence of the previous gene and the 5 kb sequence upstream the genomic sequences annotated in *Phytozome* were aligned to exclude the common parts in order to reduce the promoter fragment size. Schematic representation of promoter sequence predicted for **A)** Mp*DIV1*, **B)** Mp*DIV2*, **C)** Mp*DIVL* and **D)** Mp*DRIF*.

The predicted promoter region of each gene was cloned using the *Gateway® cloning technology*. This universal cloning method is based on site-specific recombination properties of bacteriophage lambda and it allows a rapid and efficient way to move DNA sequences into multiple vector systems for functional analysis and protein expression (Hartley et al., 2000).

The first step of the Gateway technology is to obtain PCR products with *att*B sites (figure 17). The *att*B-PCR products are generated by PCR amplification with primers that incorporate the *att*B sequences. These PCR products are then used as substrate in a recombination reaction with a donor vector. Two recombination reactions are at the basis of this technology: the BP reaction (catalysed by the BP Clonase enzyme) in which occurs the recombination between an *att*B (*att*B-PCR product or a linearized *att*B expression clone) and an *att*P (donor vector) substrate to create an *att*L-containing entry clone; and the LR reaction (catalysed by the LR Clonase enzyme) in which occurs the recombination of an *att*L (entry clone) with an *att*R substrate (destination vector) creating an *att*B-containing expression clone (figure 17) (Hartley et al., 2000).

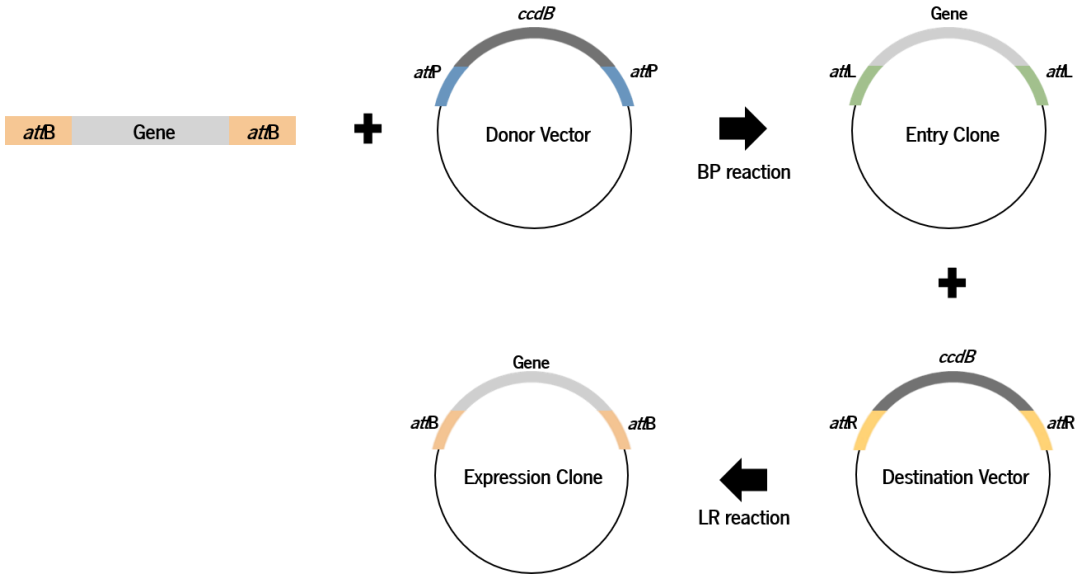

Figure 17: Gateway cloning technology. Two recombination reactions are at the basis of this technology: the BP reaction, catalysed by the BP Clonase enzyme, that enables the recombination of an *attB* (*attB*-PCR product or a linearized *attB* expression clone) and an *attP* (donor vector) substrate to create an *attL*-containing entry clone; and the LR reaction, catalysed by the LR Clonase enzyme, that enables the recombination of an *attL* (entry clone) with an *attR* substrate (destination vector) creating an *attB*-containing expression clone.

The amplification of the Mp*DIV1*, Mp*DIV2*, Mp*DIVL*, Mp*DRIF.1* and Mp*DRIF.2* promoter sequences for Gateway cloning was performed in two distinct PCR reactions to avoid the use of primers with high number of non-homologous nucleotides. The first reaction was performed with specific primers for the target sequence with part of the *attB* sequence incorporated. The second reaction was performed using the PCR products obtained in the first amplification and primers containing the complete *attB* sequence. To avoid non-specific amplifications, a control reaction without *attB* complete primers was set up for each second PCR reaction.

Five independent PCR reactions were performed for each promoter sequence to avoid DNA polymerase induced errors. Concerning the first PCR reaction, no unspecific amplification occurred in any of the five reactions performed for each promoter (figure 18). In the second PCR, five independent reactions were again performed for each promoter sequence and all the PCR reactions produced an amplified fragment when compared to the control reactions.

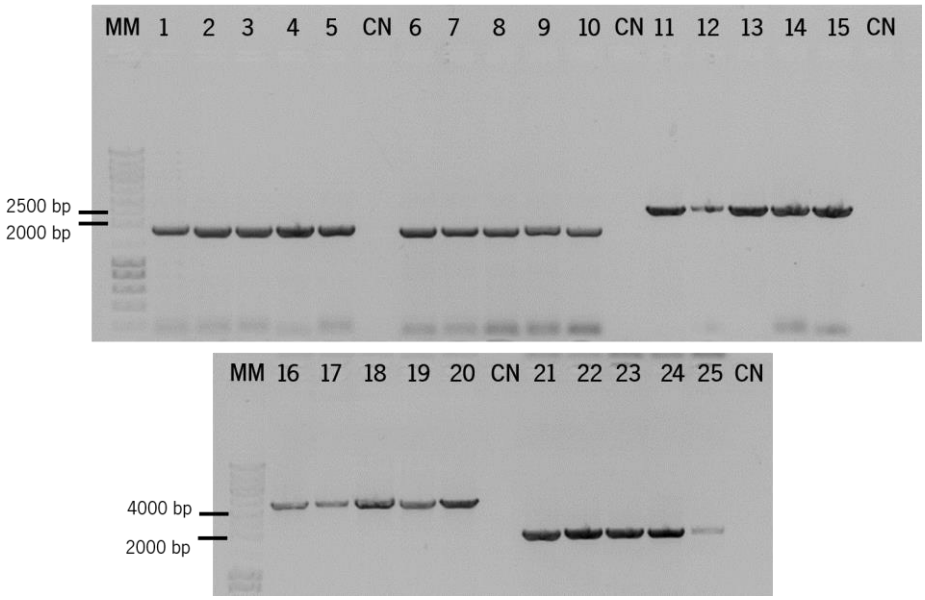

Figure 18: PCR amplification of Mp*DIV1*, Mp*DIV2*, M*pDIVL*, Mp*DRIF.1* and Mp*DRIF.2* predicted promoter sequences for Gateway cloning. In the first amplification reaction, the promoter sequences were amplified with specific primers that had incorporated part of the *att*B sequence.
Lanes 1 to 5: *pro*Mp*DIV1* five independent PCR amplification products (expected size fragment of 1722 bp);
Lanes 6 to 10: *pro*Mp*DIV2* five independent PCR amplification products (expected size fragment of 1656 bp);
Lanes 11 to 15: *pro*Mp*DIVL* five independent PCR amplification products (expected size fragment of 2483 bp);
Lanes 16 to 20: *pro*Mp*DRIF.2* five independent PCR amplification products (expected size fragment of 4554 bp);
Lanes 21 to 25: *pro*Mp*DRIF.1* five independent PCR amplification products (expected size fragment of 1964 bp);
MM: Molecular marker (NZYDNA Ladder III, Nzytech); CN: Negative control.

A BP reaction was performed to recombine the *att*B-PCR products with the pDONOR vector. The obtained constructs were transformed into *E. coli* DH10β competent cells and the colonies were grown in LB-agar medium supplemented with 50 µg mL^{-1} kanamycin. The colonies were screened by PCR using pDONOR specific primers (figure 19; Table B1- Annex B).

Figure 19: Hybridization localization of the primers used in the screening of *pro*Mp*DIV1*, *pro*Mp*DIV2*, *pro*Mp*DIVL* or *pro*Mp*DRIF*pDONOR colonies by PCR. The specific forward (Fw) and reverse (Rv) primers for the pDONOR vector were used for the screening.

The PCR screening showed that *pro*Mp*DIV1.1*$_{pDONOR}$ 28, *pro*Mp*DIV2*$_{pDONOR}$ 42, *pro*Mp*DIVL*$_{pDONOR}$ 3, *pro*Mp*DRIF.1*$_{pDONOR}$ 20 and *pro*Mp*DRIF.2*$_{pDONOR}$ 38 had a PCR amplification product close to the expected size (figure 20). The plasmids of these colonies were isolated and sequenced. All the constructs had the inserted promoter sequence in the desired orientation.

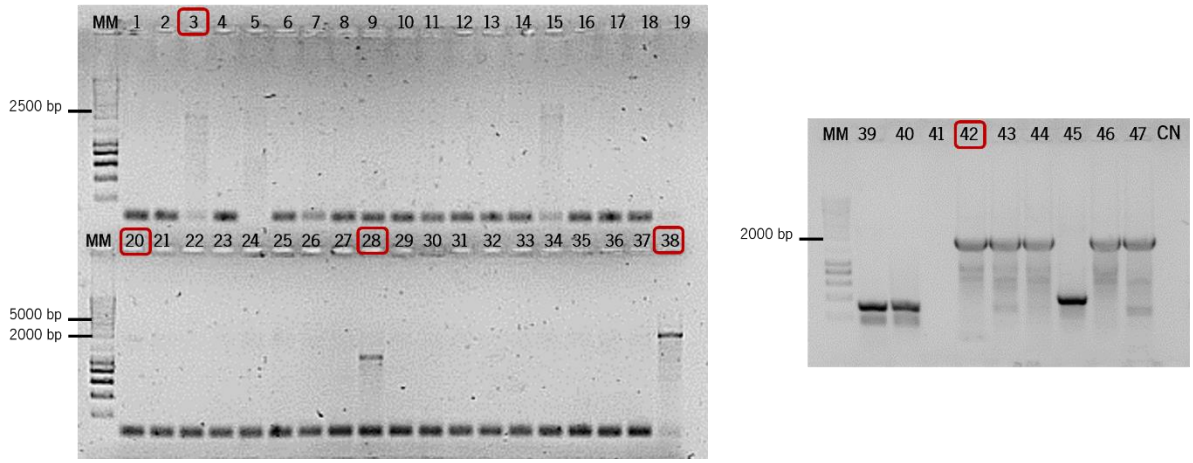

Figure 20: PCR screening of *E. coli* colonies containing constructs with *pro*Mp*DIV1*, *pro*Mp*DIV2*, *pro*Mp*DIVL*, proMpDRIF*.1* and *pro*Mp*DRIF.2* sequences in pDONOR vector for the promoters cloning. The colonies were screened by PCR using pDONOR specific primers.
Lanes 1 to 18: *pro*Mp*DIVL*pDONOR colonies (expected size fragment of 2737 bp);
Lanes 19 to 27: *pro*Mp*DRIF.1*pDONOR colonies (expected size fragment of 2218 bp);
Lanes 28 to 29: *pro*Mp*DIV1*pDONOR colonies (expected size fragment of 2026 bp);
Lanes 30 to 38: *pro*Mp*DRIF.2*pDONOR colonies (expected size fragment of 4808 bp);
Lanes 39 to 47: *pro*Mp*DIV2*pDONOR colonies (expected size fragment of 1910 bp);
MM: Molecular marker (NZYDNA Ladder III, Nzytech); CN: Negative control;
The lanes with red squares correspond to the colonies selected for plasmid isolation.

In the future, to recombine *pro*Mp*DIV1.1*pDONOR, *pro*Mp*DIV2*pDONOR, *pro*Mp*DIVL*pDONOR, *pro*Mp*DRIF.2*pDONOR and *pro*Mp*DRIF.1*pDONOR with the destination vector pMpGWB104 (with the GUS reporter gene), a LR reaction will be performed between each entry clone and the destination vector.

3.2. Overexpression studies

To investigate the function of Mp*DIV1,* Mp*DIV2*, Mp*DIVL* and Mp*DRIF* genes in *M. polymorpha*, overexpression transgenic plants were obtained using *Gateway® cloning technology.*

3.2.1. Cloning of Mp*DIV1,* Mp*DIV2*, Mp*DIVL* and Mp*DRIF* coding sequences

As explained before, the amplification of the coding sequences of the different transcripts (Mp*DIV1.1*, Mp*DIV1.2*, Mp*DRIF.1*, Mp*DRIF.2*, Mp*DIV2 and* Mp*DIVL*) for Gateway cloning was performed in two distinct PCR reactions. Regarding the first reaction, some unspecific amplification occurred but only for Mp*DRIF.2.* The expected PCR fragment of 600 bp was excised from the agarose gel (figure 21D, lane 1). In the second PCR reaction, a control reaction without primers was performed to avoid non-specific amplifications. All the coding sequences were significantly more amplified when compared to the control reactions (figure 21, lane 2).

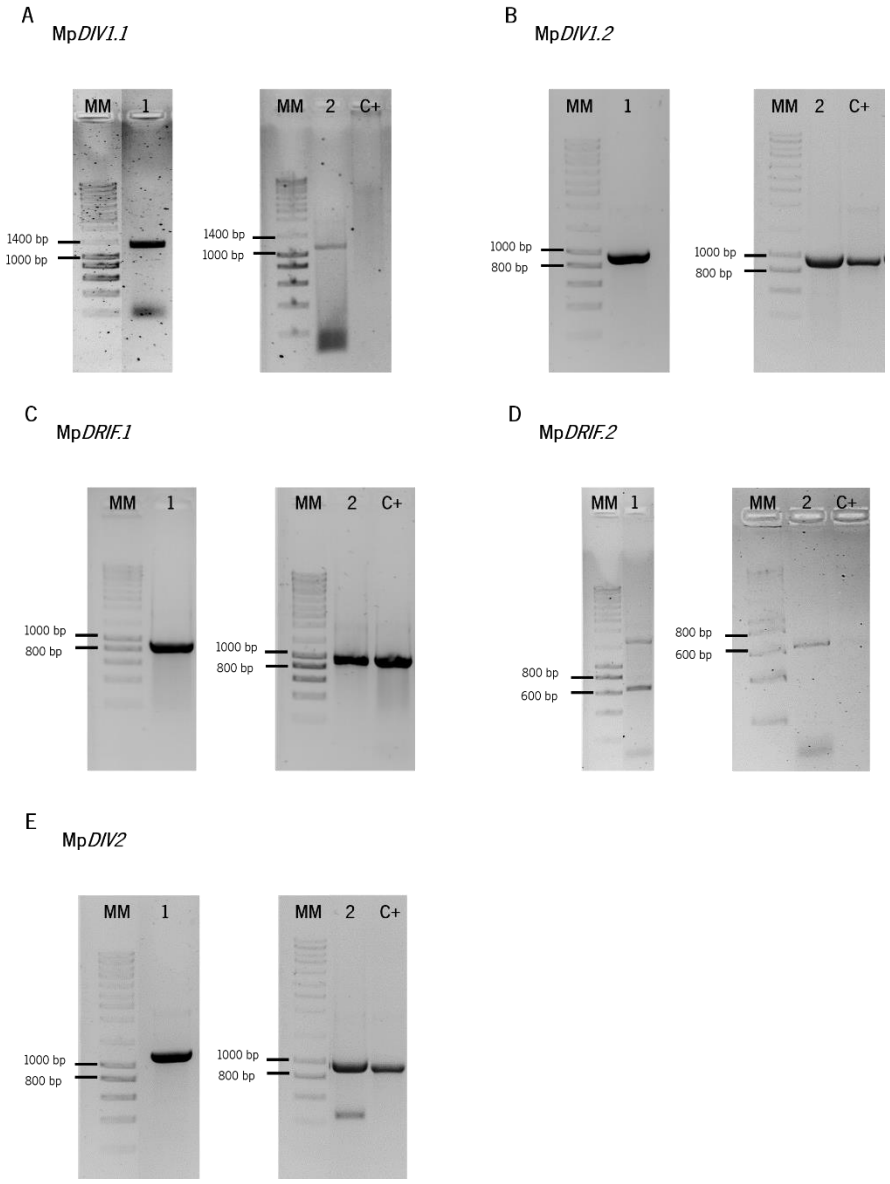

Figure 21: PCR amplification of Mp*DIV1.1*, Mp*DIV1.2*, Mp*DRIF.1*, Mp*DRIF.2* and M*pDIV2* coding sequences for Gateway cloning. The amplification was performed in two different PCR amplification reactions. In the first reaction, the coding sequences were amplified and in the second reaction the *att*B sequences were added to the flanking sites of the coding sequences.

A) Mp*DIV1.1* coding sequence amplification:

Lane 1: First PCR reaction product (expected size fragment of 1095 bp);

Lane 2: Second PCR reaction product (expected size fragment of 1129 bp);

B) Mp*DIV1.2* coding sequence amplification:

Lane 1: First PCR reaction product (expected size fragment of 930 bp);

Lane 2: Second PCR reaction product (expected size fragment of 964 bp);

C) Mp*DRIF.1* coding sequence amplification:

Lane 1: First PCR reaction product (expected size fragment of 912bp);

Lane 2: Second PCR reaction product (expected size fragment of 946 bp);

D) Mp*DRIF.2* coding sequence amplification:

Lane 1: First PCR reaction product (expected size fragment of 600 bp);

Lane 2: Second PCR reaction product (expected size fragment of 634 bp);

E) Mp*DIV2* coding sequence amplification:

Lane 1: First PCR reaction product (expected size fragment of 933 bp);

Lane 2: Second PCR reaction product (expected size fragment of 967 bp).

MM: Molecular marker (NZYDNA Ladder III, Nzytech); C+: Positive control reaction for the second PCR reaction without *att*B universal primers.

The PCR products obtained in the second PCR amplification reaction were purified with a PEG solution and the BP reaction was then performed to recombine the *att*B-PCR products in the pDONOR vector. The constructs were then transformed into competent cells of *E. coli* DH10β and the putative positive colonies were identified in LB-agar medium supplemented with 50 μg mL^{-1} kanamycin. The colonies were screened by PCR using pDONOR specific primers (figure 22; Table B1- Annex B). A control reaction was also performed using the same set of primers and the empty vector as template (expected size fragment of 2493 bp).

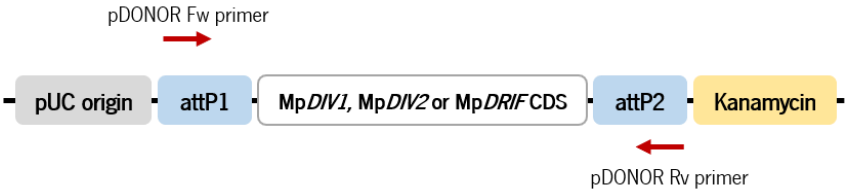

Figure 22: Hybridization localization of the primers used in the screening of Mp*DIV1*, Mp*DIV2* or Mp*DRIF*pDONOR colonies by PCR. The specific forward (Fw) and reverse (Rv) primers for the pDONOR vector were used in the screening.

The Mp*DIV1.1*$_{pDONOR}$ 16, Mp*DIV1.2*$_{pDONOR}$ 5, Mp*DRIF.2*$_{pDONOR}$ 16, Mp*DRIF.1*$_{pDONOR}$ 2 and Mp*DIV2*$_{pDONOR}$ 7 colonies produced a PCR product with the expected size (figure 23). The plasmids of these colonies were isolated and sequenced. All the constructs had the inserted coding sequence in the desired orientation.

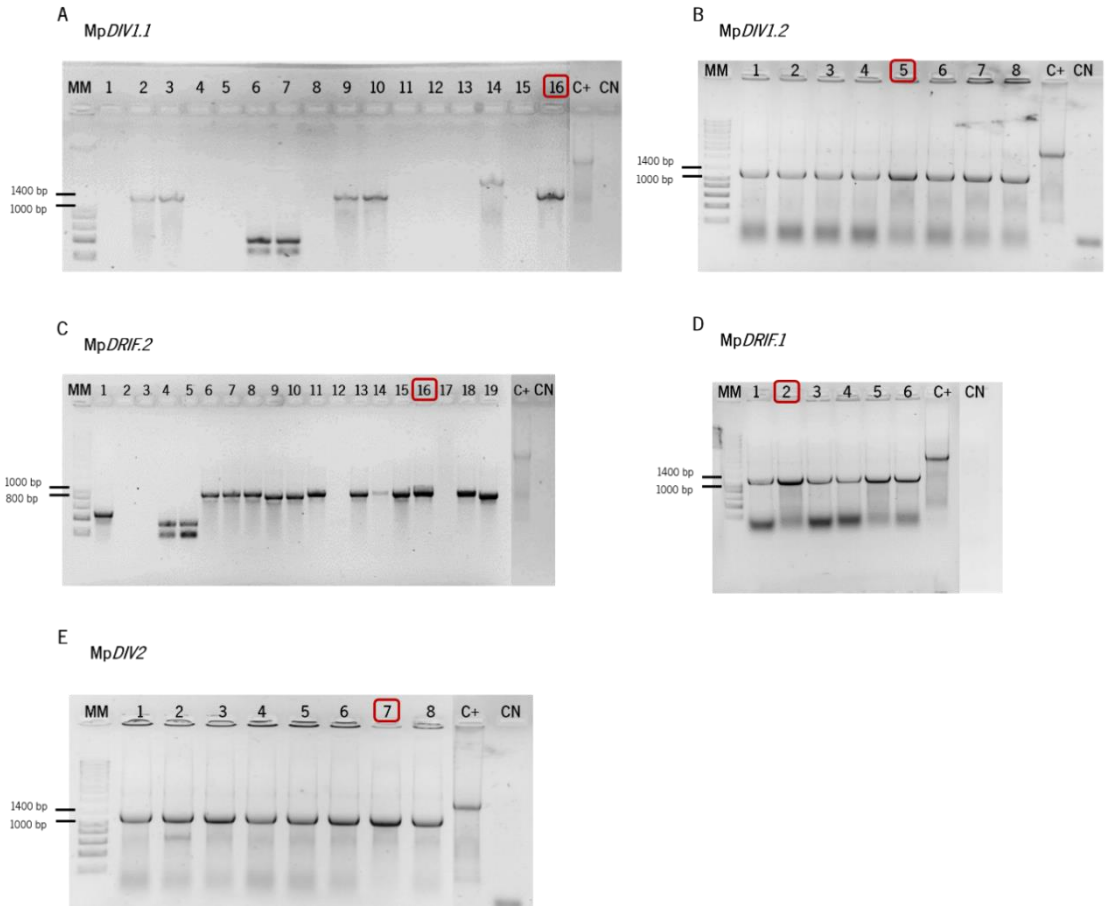

Figure 23: PCR screening of *E. coli* colonies containing constructs with Mp*DIV1.1*, Mp*DIV1.2*, Mp*DRIF.1*, Mp*DRIF.2* and M*pDIV2* coding sequence in pDONOR vector for the overexpression assays. The colonies were screened by PCR using pDONOR specific primers.
A) Mp*DIV1.1*pDONOR colonies, lanes 1 to 16 (expected size fragment of 1349 bp);
B) Mp*DIV1.2*pDONOR colonies, lanes 1 to 8 (expected size fragment of 1184 bp);
C) Mp*DRIF.2*pDONOR colonies, lanes 1 to 19 (expected size fragment of 854 bp);
D) Mp*DRIF.1*pDONOR colonies, lanes 1 to 6 (expected size fragment of 1166 bp);
E) Mp*DIV2*pDONOR colonies, lanes 1 to 8 (expected size fragment of 1187 bp).
MM: Molecular marker (NZYDNA Ladder III, Nzytech); C+: Control reaction with the empty vector (expected size fragment of 2493 bp); CN: Negative control. The lanes with red squares correspond to the colonies selected for sequencing.

The cloning of Mp*DIVL* coding sequence in the pDONOR was not successfull, thus a different cloning strategy was followed in order to overexpress this gene in *M. polymorpha* (figure 24). The pENTRY/D-TOPO was used to obtain an entry clone instead of pDONR™201. First, the Mp*DIVL* coding sequence was amplified with primers that incorporate the recognition sites for the restriction enzymes NotI and AscI (figure 25A; Table B1- Annex B). The purified PCR product and the pENTRY/D-TOPO vector were sequentially digested with the two restriction enzymes NotI and AscI (figure 25B). A ligation reaction was performed to obtain an entry clone. The obtained construct was transformed into competent cells of *E. coli* DH10β and the putative positive colonies were identified in LB-agar medium supplemented with 50 μg mL^{-1} kanamycin. The colonies were screened by colony PCR using vector specific primers (Table B1- Annex B).

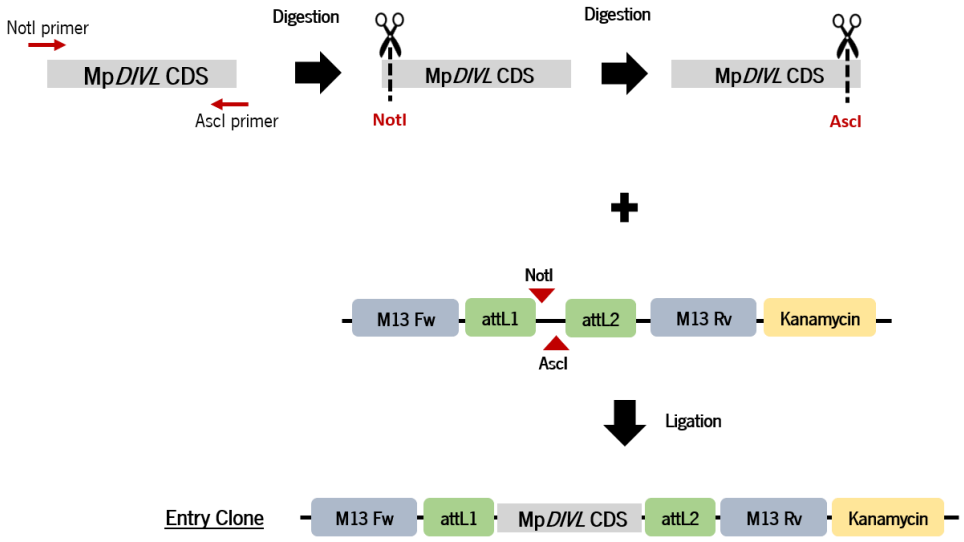

Figure 24: Strategy used for the Mp*DIVL* coding sequence cloning in pENTRY/D-TOPO. Mp*DIVL* coding sequence was amplified with primers containing the recognition sites for the restriction enzymes NotI and AscI. The PCR product and the pENTRY/D-TOPO vector were digested with the restriction enzymes, NotI and AscI, and ligated.

The Mp*DIVL*pENTRY clone 3 showed a PCR amplification product close to the expected size (figure 25C). The plasmid of this clone was isolated and sequenced, confirming that the construct had the coding region of the gene inserted in the desired orientation.

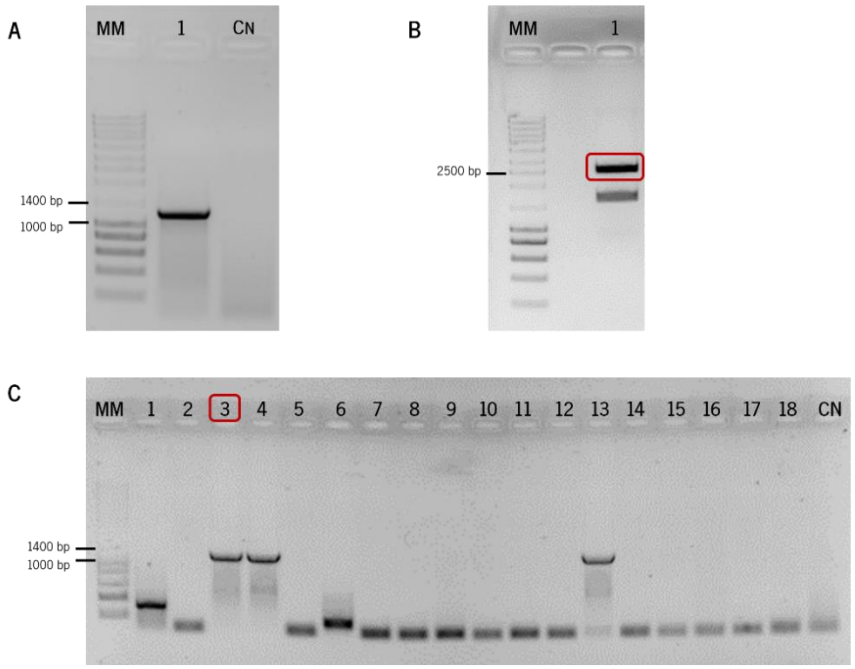

Figure 25: Mp*DIVL* coding sequence cloning in pENTRY/D-TOPO.
A) Mp*DIVL* coding sequence amplification with primers containing the recognition sites of the restriction enzymes NotI and AscI.
Lane 1: PCR reaction product (expected size fragment of 1186 bp);
B) pENTRY/D-TOPO digestion with NotI and AscI restriction enzymes;
Lane 1: Digested product. The lane with the red square corresponds to the expected digestion product (expected size fragment of 2500 bp);
C) PCR screening of *E. coli* colonies containing constructs with Mp*DIVL* coding sequence in pENTRY/D-TOPO for the overexpression assays;
Lanes 1-18: Mp*DIVL*pENTRY colonies (expected size fragment of 1186 bp). The lane with the red square corresponds to the colony selected for sequencing.
MM: Molecular marker (NZYDNA Ladder III, Nzytech); CN: Negative control.

After successfully obtaining the entry clones, the next step was to perform the LR reaction to recombine the Mp*DIV1.1*pDONOR, Mp*DIV1.2*pDONOR, Mp*DRIF.2*pDONOR, Mp*DRIF.1*pDONOR and Mp*DIVL*pENTRY with the destination vector pMpGWB103. This vector has the strong promoter Mp*EF1α* cloned upstream of the Gateway cassette and it allows the overexpression of a gene. The constructs obtained with the LR reaction were transformed into *E. coli* DH10β competent cells and grown in LB-agar medium with 50 μg mL^{-1} spectinomycin. The colonies were screened by PCR using *att*B specific primers (figure 26).

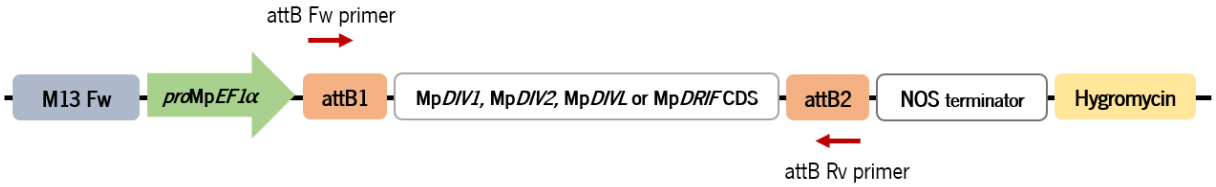

Figure 26: Hybridization localization of the primers used in the screening of Mp*DIV1*, Mp*DIV2*, Mp*DIVL* or Mp*DRIF*pMpGWB103 colonies by PCR. The specific forward (Fw) and reverse (Rv) primers for the *att*B regions were used in the screening.

The screening showed that Mp*DIV1.1*pMpGWB103 clone 8, Mp*DIV1.2*pMpGWB103 clone 19, Mp*DIV2*pMpGWB103 clone 19, Mp*DIVL*pMpGWB103 clone 9, Mp*DRIF.1*pMpGWB103 clone 8 and Mp*DRIF.2*pMpGWB103 clone 8 had a fragment with the expected size (figure 27, 28 and 29). The plasmids of these colonies were isolated to proceed to the *M. polymorpha* transformation.

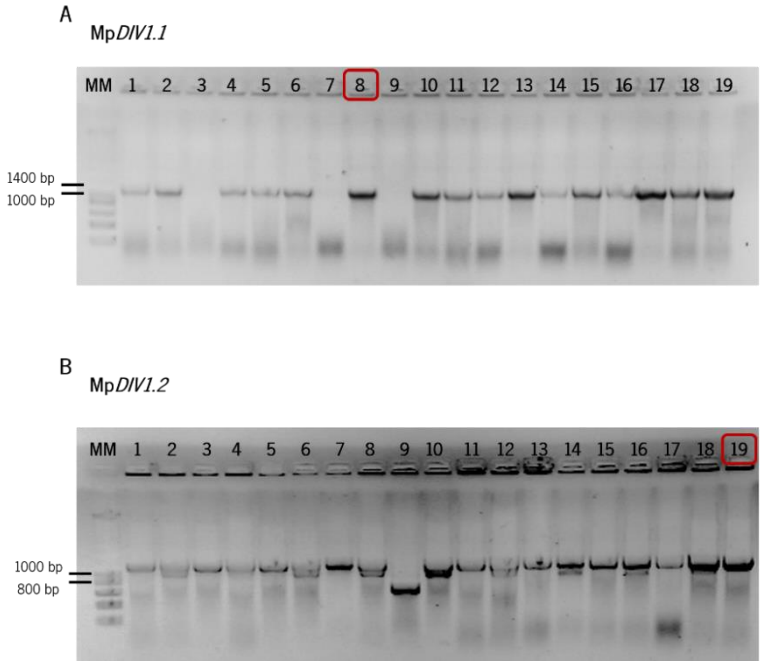

Figure 27: PCR screening of *E. coli* colonies containing constructs with Mp*DIV1.1* and Mp*DIV1.2* coding sequence in pMpGWB103 vector for the overexpression assays. The colonies were screened by PCR using *att*B specific primers.
A) Mp*DIV1.1*pMpGWB103 colonies, lanes 1 to 19 (expected size fragment of 1129 bp);
B) Mp*DIV1.2*pMpGWB103 colonies, lanes 1 to 19 (expected size fragment of 964 bp);
MM: Molecular marker (NZYDNA Ladder III, Nzytech);
The lanes with red squares correspond to the colonies selected for plasmid isolation.

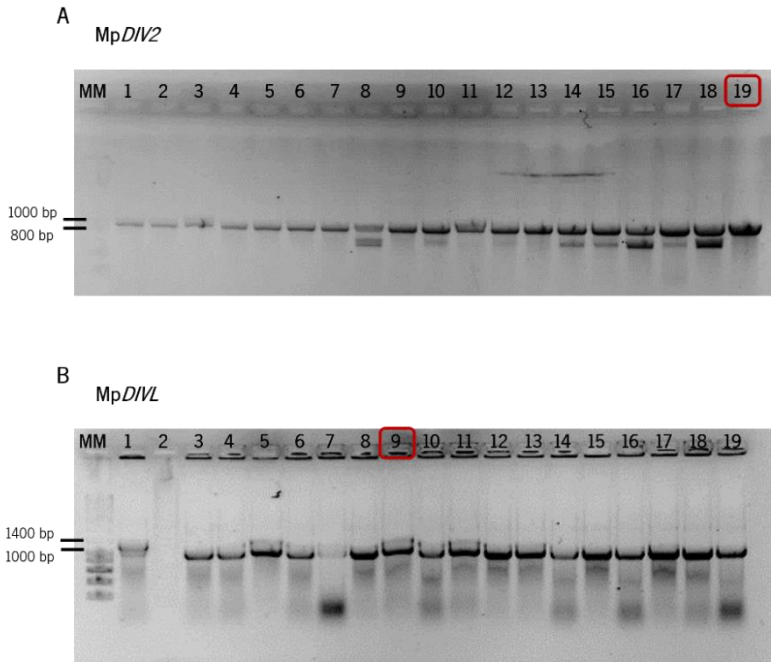

Figure 28: PCR screening of *E. coli* colonies containing constructs with Mp*DIV2* and Mp*DIVL* coding sequence in pMpGWB103 vector for the overexpression assays. The colonies were screened by PCR using *att*B specific primers.
A) Mp*DIV2*pMpGWB103 colonies, lanes 1 to 19 (expected size fragment of 967 bp);
B) Mp*DIVL*pMpGWB103 colonies, lanes 1 to 19 (expected size fragment of 1243 bp);
MM: Molecular marker (NZYDNA Ladder III, Nzytech);
The lanes with red squares correspond to the colonies selected for plasmid isolation.

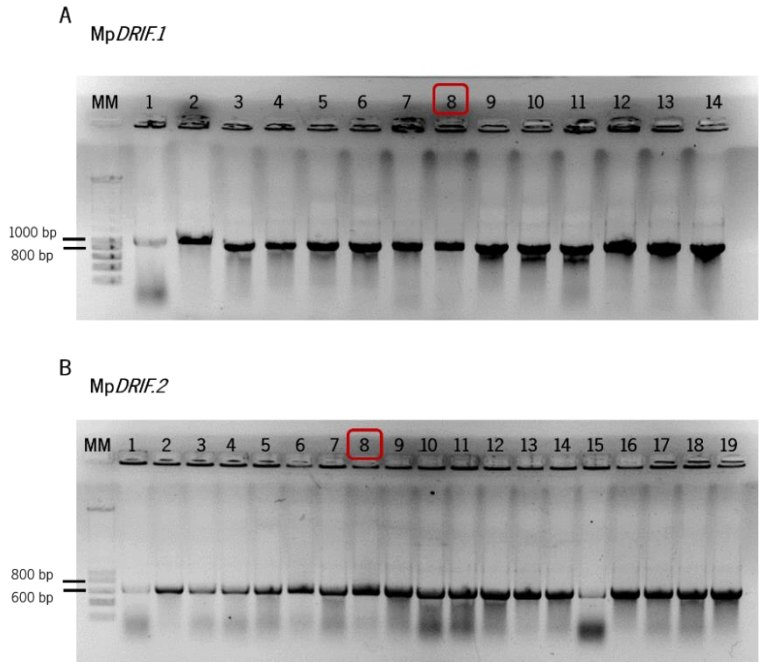

Figure 29: PCR screening of *E. coli* colonies containing constructs with Mp*DRIF.1* and Mp*DRIF.2* coding sequence in pMpGWB103 vector for the overexpression assays. The colonies were screened by PCR using *att*B specific primers.
A) Mp*DRIF.1*pMpGWB103 colonies, lanes 1 to 14 (expected size fragment of 946 bp);
B) Mp*DRIF.2*pMpGWB103 colonies, lanes 1 to 19 (expected size fragment of 634 bp);
MM: Molecular marker (NZYDNA Ladder III, Nzytech);
The lanes with red squares correspond to the colonies selected for plasmid isolation.

3.2.2. *Marchantia polymorpha* thallus transformation

M. polymorpha transgenic lines overexpressing Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* genes were obtained by *Agrobacterium*-mediated transformation. For that, competent cells of the GV2260 *A. tumefaciens* strain were transformed with the Mp*DIV1.1*pMpGWB103, Mp*DIV1.2*pMpGWB103, Mp*DIV2*pMpGWB103, Mp*DIVL*pMpGWB103, Mp*DRIF.1*pMpGWB103 and Mp*DRIF.2*pMpGWB103 constructs. A colony PCR was performed to screen the colonies using *att*B specific primers (figure 26).

Mp*DIV1.1*pMpGWB103 clone 19, Mp*DIV1.2*pMpGWB103 clone 44, Mp*DIV2*pMpGWB103 clone 4, Mp*DIVL*pMpGWB103 clone 31, Mp*DRIF.1*pMpGWB103 clone 60 and Mp*DRIF.2*pMpGWB103 clone 21 in *Agrobacterium* showed a fragment with the expected size (figure 30).

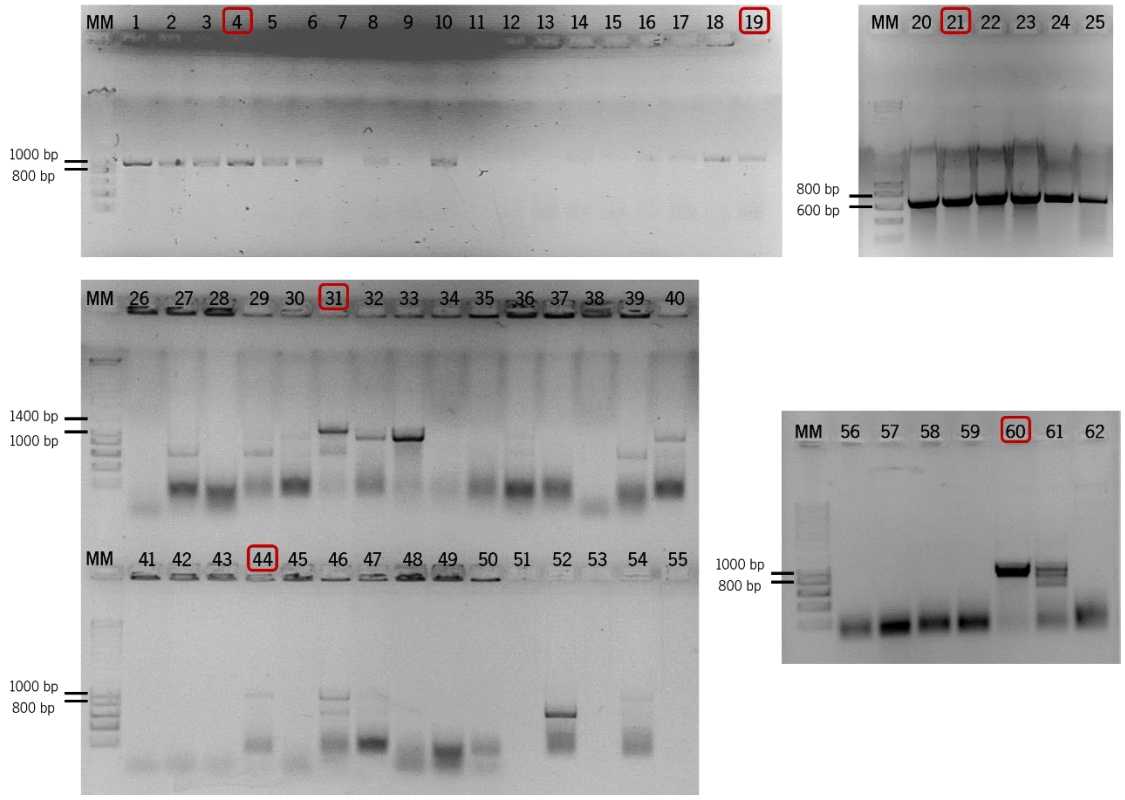

Figure 30: PCR screening of *Agrobacterium* colonies containing constructs with Mp*DIV1.1,* Mp*DIV1.2,* Mp*DIV2,* Mp*DIVL*, Mp*DRIF.1* and Mp*DRIF.2* coding sequence in pMpGWB103 vector for the overexpression assays. The colonies were screened by PCR using *att*B specific primers.

Lanes 1 to 10: Mp*DIV2*pMpGWB103 colonies (expected size fragment of 967 bp);
Lanes 11 to 19: Mp*DIV1.1*pMpGWB103 colonies (expected size fragment of 1129 bp);
Lanes 20 to 25: Mp*DRIF.2*pMpGWB103 colonies (expected size fragment of 634 bp);
Lanes 26 to 40: Mp*DIVL*pMpGWB103 colonies (expected size fragment of 1243 bp);
Lanes 41 to 55: Mp*DIV1.2*pMpGWB103 colonies (expected size fragment of 964 bp);
Lanes 56 to 62: Mp*DRIF.1*pMpGWB103 colonies (expected size fragment of 946 bp);
MM: Molecular marker (NZYDNA Ladder III, Nzytech);
The lanes with red squares correspond to the colonies selected for the thallus transformation.

M. polymorpha wild-type plant thallus from approximately 1-month-old plants were sectioned into 1 mm^2 pieces and transformed with an overnight culture of *A. tumefaciens* GV2260 containing the constructs mentioned above. Positive transformants were selected by observing the plant survival in the selective media with 10 mg L^{-1} hygromycin and 100 mg L^{-1} cefotaxime.

Positive transformants were obtained with the Mp*DIV1.1*$_{pMpGWB103}$ and Mp*DRIF.1*$_{pMpGWB103}$ constructs with, 20 and 5 transformed plants respectively.

3.2.3. Plant genotyping

Approximately 1-month-old plant positive putative transformants of *pro*Mp*EF1α*::Mp*DIV1.1* and *pro*Mp*EF1α*::Mp*DRIF.1* were propagated into new half strength Gamborg B5 medium with 1.4% (w/v) agar with the antibiotics hygromycin and cefotaxime. Wild-type plants were propagated at the same time to serve as control. After 3 to 4 weeks, the DNA was extracted from a thallus section of approximately 5 cm^2 and genotyped by PCR amplification using specific primers (figure 31).

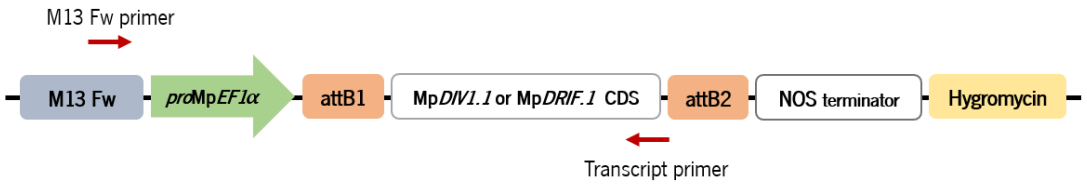

Figure 31: Hybridization localization of the primers used in the genotyping of *pro*Mp*EF1α*::Mp*DIV1.1* and *pro*Mp*EF1α*::Mp*DRIF.1* plants. The M13 forward (Fw) primer and a specific transcript primer were used for the screening.

The expected size of the PCR fragment for a Mp*DIV1.1*$_{pMpGWB103}$ and Mp*DRIF.1*$_{pMpGWB103}$ transformed plant was 2887 and 2704 bp respectively. To confirm the correct size, a control reaction was also performed using the same set of primers and the constructs *pro*Mp*EF1α*::Mp*DIV1.1* and *pro*Mp*EF1α*::Mp*DRIF.1* as template.

The PCR amplification using template gDNA from the *pro*Mp*EF1α*::Mp*DIV1.1* plant lines #2, #3, #4, #5, #7, #8 and #11 produced a fragment with the expected size (figure 32A). The *pro*Mp*EF1α*::Mp*DIV1.1* lines #3, #7, #8 and #11 were chosen to analyse the overexpression phenotype.

The PCR amplification using template gDNA from the *pro*Mp*EF1α*::Mp*DRIF.1* plant lines #2 and #4 produced a fragment with a size close to the expected one (figure 32B). Line #1 showed two fragments, the desired one and one with a bigger size, and for that this plant was discarded. For the plant lines #3 and #5 no fragment was observed, probably due to a low amount of template gDNA (figure 32B). The DNA of these two plants was extracted again, the PCR repeated, and it was possible to observe the desired

fragment. The *pro*Mp*EF1α*::Mp*DRIF.1* lines #2, #3, #4 and #5 were chosen to analyse the overexpression phenotype.

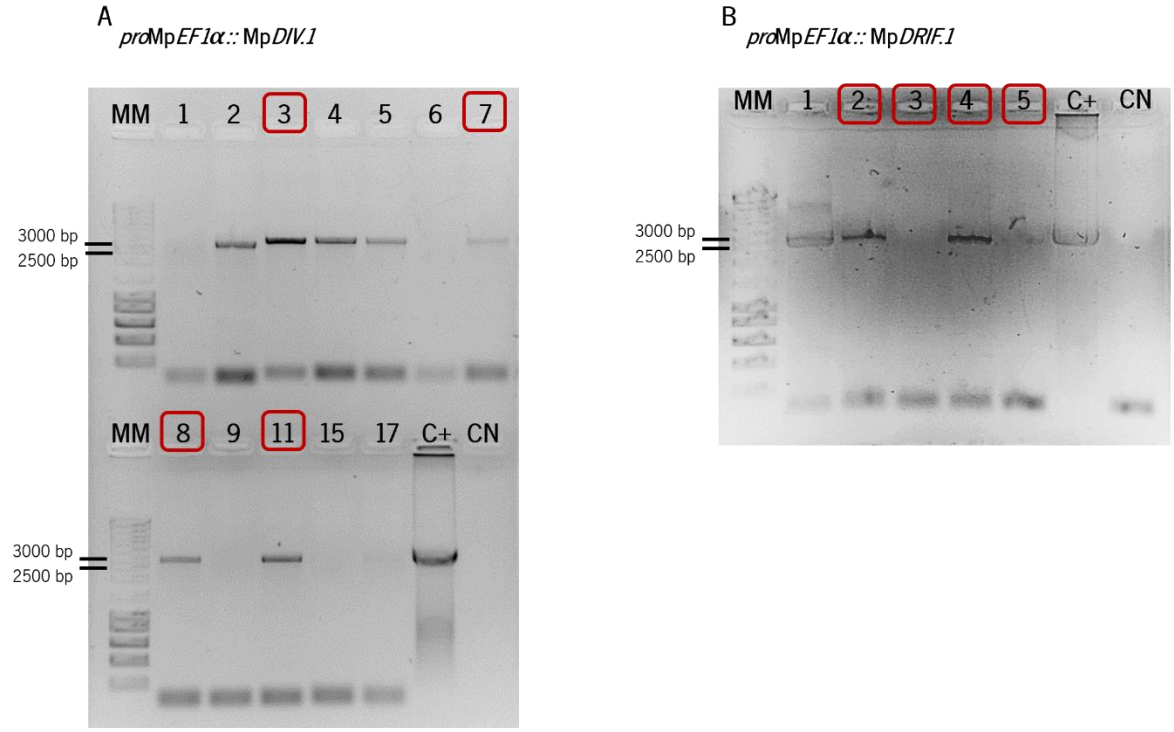

Figure 32: *pro*Mp*EF1α*::Mp*DIV1.1* and *pro*Mp*EF1α*::Mp*DRIF.1* plants genotyping by PCR amplification. The plants were genotyped using specific primers for the M13 region and for the gene.
A) *pro*Mp*EF1α*::Mp*DIV1.1* plant lines 1 to 9, 11, 15 and 17 genotyping (expected size fragment of 2887 bp);
B) *pro*Mp*EF1α*::Mp*DRIF.1* plant lines 1 to 5 genotyping (expected size fragment of 2704 bp);
MM: Molecular marker (NZYDNA Ladder III, Nzytech); C+: proMp*EF1α*::Mp*DIV1.1* and proMp*EF1α*::Mp*DRIF.1* constructs; CN: Negative control. The lanes with the red squares correspond to the plant lines selected for phenotype observation.

3.2.4. Mp*DIV1.1* and Mp*DRIF.1* overexpression

To confirm the overexpression of Mp*DIV1.1* and Mp*DRIF.1* in the respective plant lines, a semi-quantitative RT-PCR was performed to analyse the transcript abundance.

RNA was extracted from the whole tissues of 1-month-old wild-type and Mp*DIV1.1* and Mp*DRIF.1* overexpression plants using the CTAB method and the relative transcript abundance was evaluated by RT-PCR using Mp*DIV1.1* and Mp*DRIF.1* cDNA samples and transcript specific primers. The Mp*EF1α* gene was used as a reference gene.

The Mp*DIV1.1* transcripts abundance in the *pro*Mp*EF1α*::Mp*DIV1.1* lines #3, #7, #8 and #11 is higher than in the wild type (figure 33A). A similar result was observed for the Mp*DRIF.1* transcripts abundance in the *pro*Mp*EF1α*::Mp*DRIF.1* lines #2, #3, #4 and #5 (figure 33B). This suggests that

Mp*DIV1.1* and Mp*DRIF.1* genes are being overexpressed in the respective four independent overexpression lines.

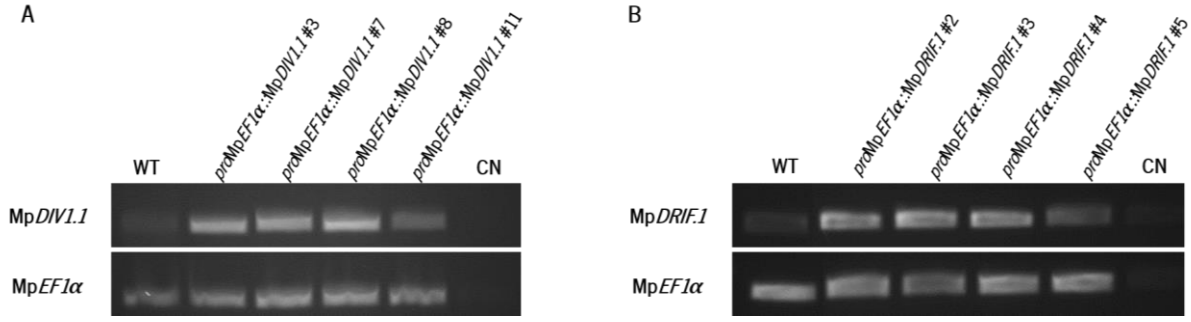

Figure 33: Analysis by RT-PCR of Mp*DIV1.1* and Mp*DRIF.1* expression in overexpression plant lines. A) Mp*DIV1.1* transcripts abundance in wild-type plants and *pro*Mp*EF1α*::Mp*DIV1.1* overexpression plant lines. **B)** Mp*DRIF.1* transcripts abundance in wild-type plants and *pro*Mp*EF1α*::Mp*DRIF.1* overexpression plant lines. Mp*DIV1.1* and Mp*DRIF.1* were amplified in a 35 cycle-PCR. The constitutive gene Mp*EF1α* was used as control. The transcripts abundance was analysed by electrophoresis. CN: Negative control.

3.2.5. Determination of the overexpression plants sex

To perform a complete and correct phenotype analysis in plants that had result from genetic variations, it is necessary to have male and female plants since the genetic alteration can vary with the plant sex.

The transgenic plants obtained by the *M. polymorpha* thallus transformation can either be male or female because the wild-type plants used in the transformation process were randomly selected. The *M. polymorpha* male and female plants can only be morphologically distinguished after displaying the sexual structures. However, to determine the plants sex at early stages of development PCR amplification using pairs of primers specific for male and female sequences can be used (Fujisawa et al., 2001).

DNA was extracted from a thallus section of approximately 5 cm^2 and genotyped by multiplex PCR amplification with two pair of primers to determine the plant sex. It was expected a PCR fragment with 1100 bp for a male plant and a PCR fragment with 516 bp for a female plant. All Mp*EF1α*::Mp*DIV1.1* and Mp*EF1α*::Mp*DRIF.1* plant lines showed a fragment close to 1000 bp meaning they were all male (figure 34). In the future, wild-type female plants will also be transformed since the overexpression of the gene may have phenotype effects in either male and female plants.

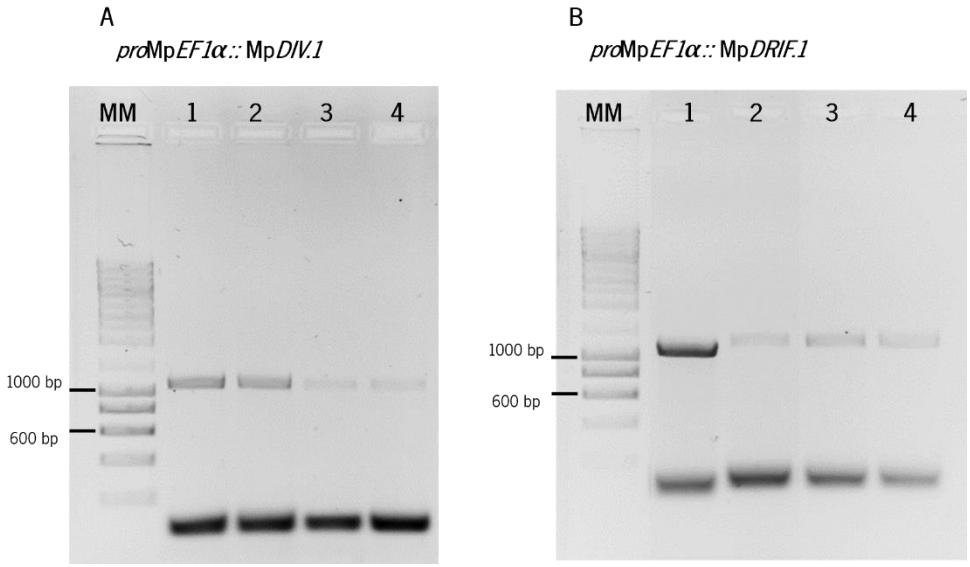

Figure 34: Determination of the *pro*MpEF1α::MpDIV1.1 and *pro*MpEF1α::MpDRIF.1 plants sex by PCR amplification. For a male plant was expected a fragment with a size close to 1100 bp and for a female plant a fragment with a size close to 516 bp.
A) Determination of the *pro*MpEF1α::MpDIV1.1 plants sex.
Lane 1: *pro*MpEF1α::MpDIV1.1 line #3 (≈ 1100 bp fragment);
Lane 2: *pro*MpEF1α::MpDIV1.1 line #7 (≈ 1100 bp fragment);
Lane 3: *pro*MpEF1α::MpDIV1.1 line #8 (≈ 1100 bp fragment);
Lane 4: *pro*MpEF1α::MpDIV1.1 line #11 (≈ 1100 bp fragment);
A) Determination of the *pro*MpEF1α::MpDRIF.1 plants sex.
Lane 1: *pro*MpEF1α::MpDRIF.1 line #2 (≈ 1100 bp fragment);
Lane 2: *pro*MpEF1α::MpDRIF.1 line #3 (≈ 1100 bp fragment);
Lane 3: *pro*MpEF1α::MpDRIF.1 line #4 (≈ 1100 bp fragment);
Lane 4: *pro*MpEF1α::MpDRIF.1 line #5 (≈ 1100 bp fragment);
MM: Molecular marker (NZYDNA Ladder III, Nzytech)

3.2.6. Phenotype observation

*pro*MpEF1α::MpDIV1.1 and *pro*MpEF1α::MpDRIF.1 independent plant lines were followed for approximately 1 month, comprising different stages of thallus development. All the transgenic plants were asexually propagated and plated into half strength Gamborg B5 medium with 1.4% (w/v) plant agar and 10 mg L^{-1} hygromycin and 100 mg L^{-1} cefotaxime. Wild-type plants were also grown during this time to serve as control.

*pro*MpEF1α::MpDIV1.1 overexpression plants #3, #7, #8 and #11 showed an increased tissue proliferation in the central part of the thallus and in the apical notches from 20 days onwards after the gemmae propagation (figure 35).

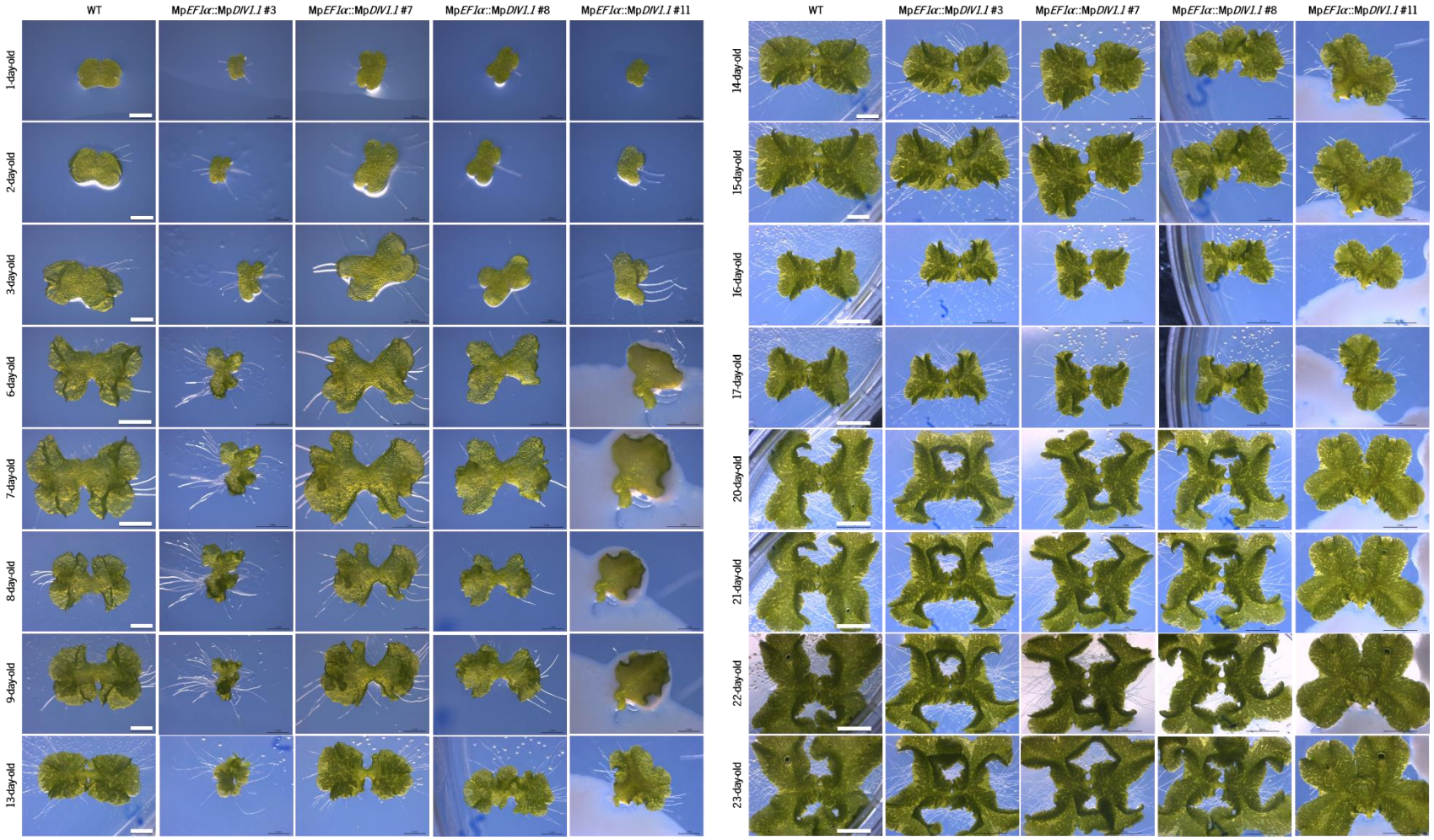

Figure 35: Analysis of *pro*Mp*EF1α::*Mp*DIV1.1* overexpression in *Marchantia polymorpha* plants. Wild-type plants and *pro*Mp*EF1α::*Mp*DIV1.1* overexpression lines #3, #7, #8 and #11 were observed at different stages of development for approximately 20 days after gemmae propagation. The plants were grown on half strength Gamborg B5 medium with the antibiotics hygromycin and cefotaxime under long-day conditions (16 h light/ 8 h dark) at 20 °C and with a light intensity between 40-45 µmol m^{-2} s^{-1}. Scale bar of: 500 µm from day 1 to 3; 1 mm from day 6 to 9; 2 mm from day 13 to 15 and 5 mm from day 16 to 23.

Although no apparent differences were observed between the rhizoids of wild-type and overexpression plants overexpression, gemmae of 1-month-old and wild-type plants were observed under a fluorescence microscope using chlorophyll auto-fluorescence. This fluorescence technique allows the observation of dark areas that correspond to the rhizoid initial cells (Figure 36A). The number of these cells can be associated to the number of rhizoids that the plant will develop (Monte et al., 2019). The number of rhizoid initial cells was divided by gemmae superficial area to observe if significative differences existed between the rhizoids of wild-type and overexpression plants (Figure 36B). *pro*Mp*EF1α*::Mp*DIV1.1* overexpression lines #3, #7, #8 and #11 gemmae have a similar number of rhizoid initial cells per gemmae superficial area when compared to wild type (figure 36B).

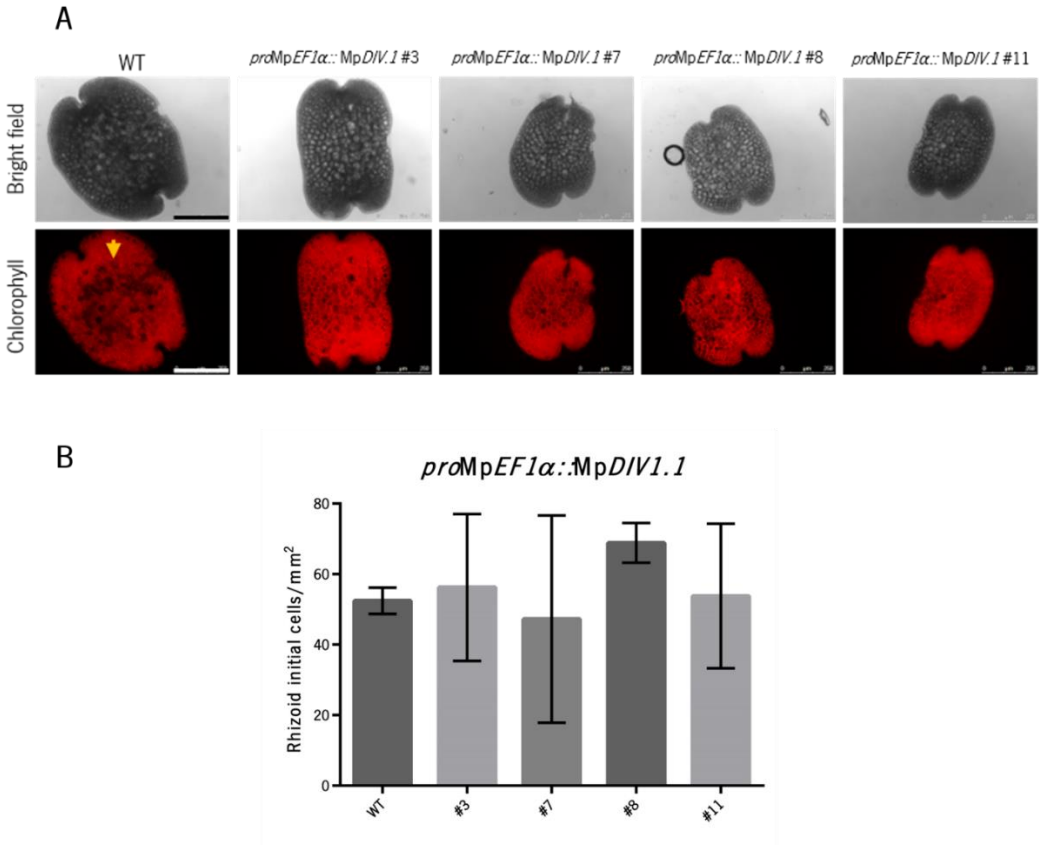

Figure 36: Analysis of *pro*Mp*EF1α*::Mp*DIV1.1* overexpression in *Marchantia polymorpha* rhizoids. A) Gemmae of wild-type plants and *pro*Mp*EF1α*::Mp*DIV1.1* plant lines #3, #7, #8 and #11 were observed under a fluorescence microscope using chlorophyll autofluorescence. The dark areas represent the rhizoid initial cells (yellow arrow). Black and white scale bars: 250 µm. **B)** The number of dark areas was counted in wild type (control) (n=2) and *pro*Mp*EF1α*::Mp*DIV1.1* lines #3 (n=3), #7 (n=2), #8 (n=2) and #11 (n=4), and divided for the gemmae area. Error bars indicate standard deviation.

*pro*Mp*EF1α*::Mp*DRIF.1* overexpression plant lines #2, #3, #4 and #5 appeared to be smaller from 13 days after gemmae propagation and onwards, when compared to wild-type, plants and the bifurcation of the lobes did not happened (figure 37).

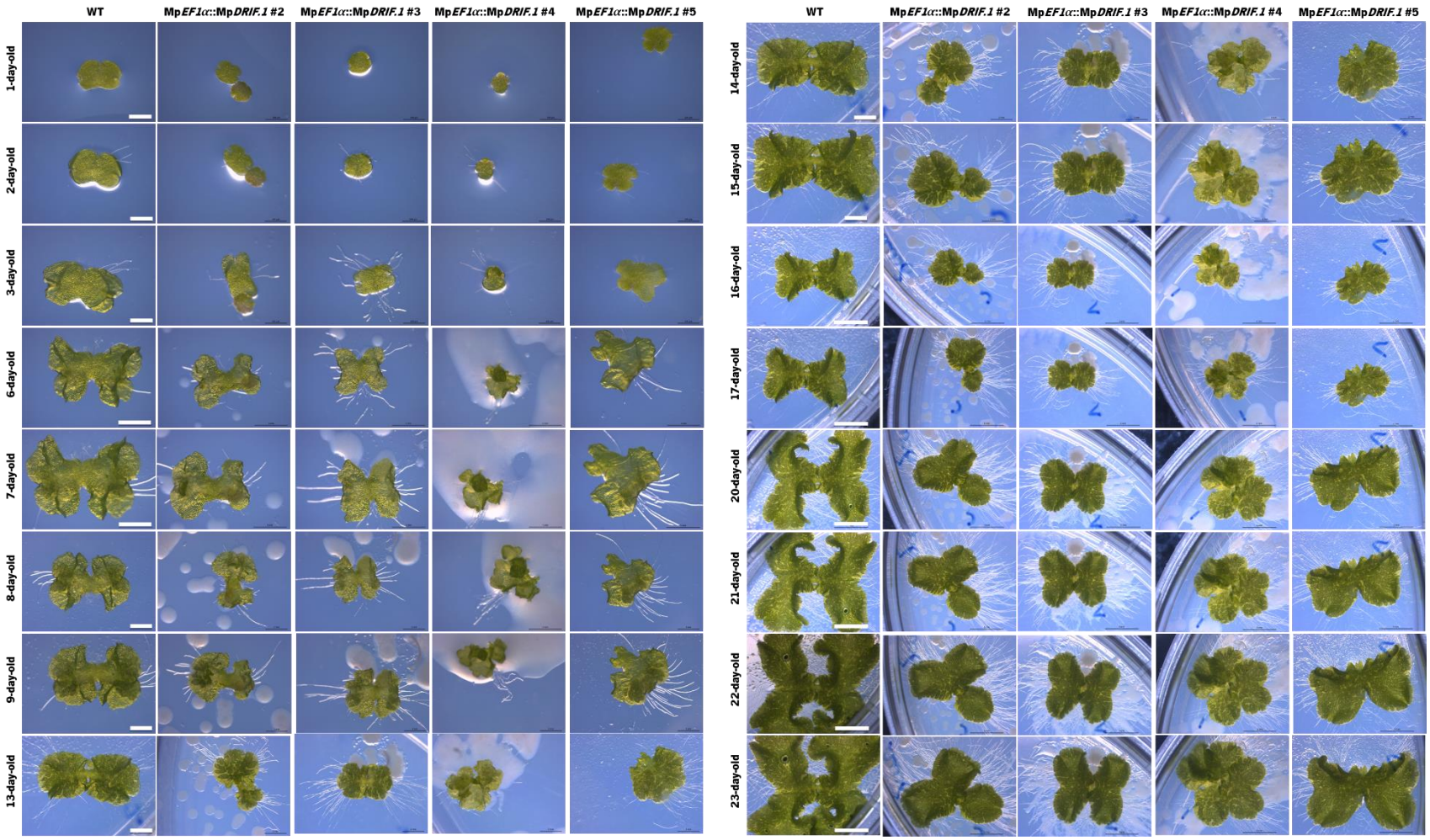

Figure 37: Analysis of *pro*Mp*EF1α::*Mp*DRIF.1* overexpression in *Marchantia polymorpha* plants. Wild-type plants and *pro*Mp*EF1α::*Mp*DRIF.1* overexpression lines #2, #3, #4 and #5 were observed at different stages of development for approximately 20 days after gemmae propagation. The plants were grown on half strength Gamborg B5 medium with the antibiotics hygromycin and cefotaxime under long-day conditions (16 h light/ 8 h dark) at 20 °C and with light intensity between 40-45 µmol m^{-2} s^{-1}. Scale bar of: 500 µm from day 1 to 3; 1 mm from day 6 to 9; 2 mm from day 13 to 15 and 5 mm from day 16 to 23.

In the gemmae images obtained from fluorescence microscopy for the *pro*Mp*EF1α*::Mp*DRIF.1* overexpression lines, a small number of dark areas were visible (figure 38A). When the number of dark areas was counted and divided by the gemmae superficial area, no significative differences were observed in the overexpression lines #2, #3, #4 and #5 in comparison to wild type (figure 38B).

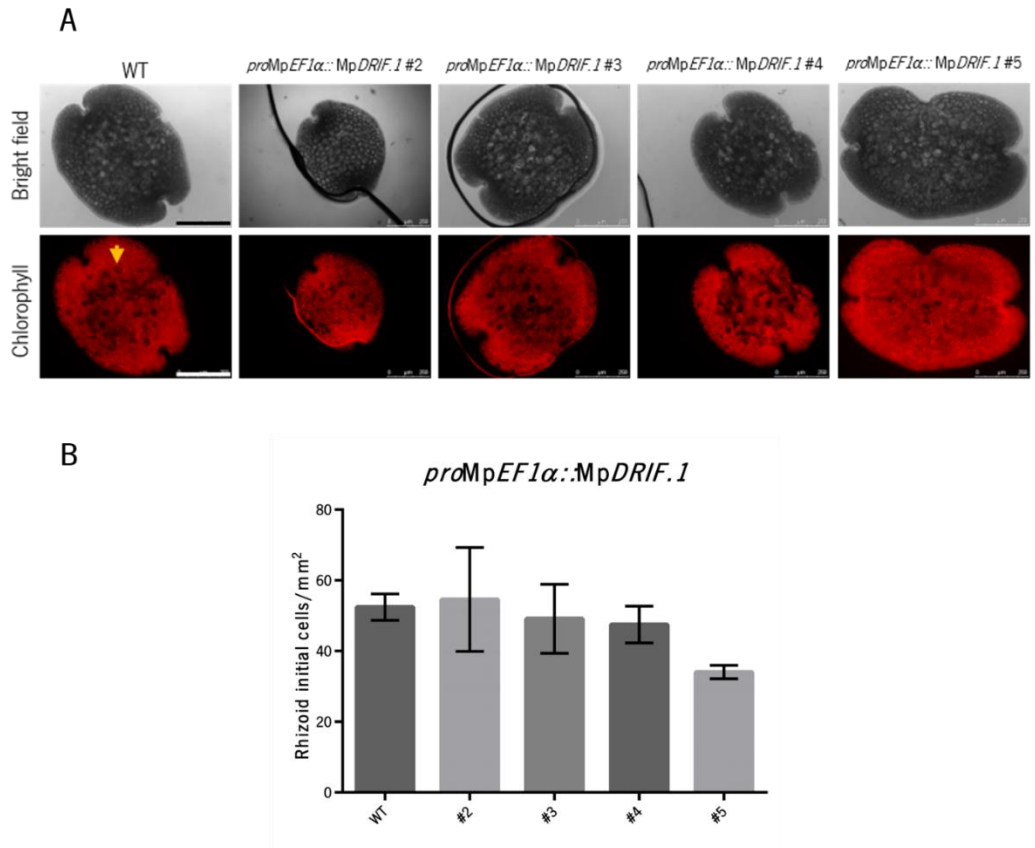

Figure 38: Analysis of *pro*Mp*EF1α*::Mp*DRIF.1* overexpression in *Marchantia polymorpha* rhizoids. A) Gemmae of wild-type plants and *pro*Mp*EF1α*::Mp*DRIF.1* plant lines #2, #3, #4 and #5 were observed under a fluorescence microscope using chlorophyll autofluorescence. The dark areas represent the rhizoid initial cells (yellow arrow). Black and white scale bars: 250 µm. **B)** The number of dark areas was counted in wild type (control) (n=2) and *pro*Mp*EF1α*::Mp*DRIF.1* lines #2 (n=4), #3 (n=2), #4 (n=2) and #5 (n=2) and divided for the gemmae area. Error bars indicate standard deviation.

3.3. Knockout mutants obtained by CRISPR/Cas9 technology

To understand the ancestral function of *DIV1*, *DIV2*, *DIVL* and *DRIF* genes in *M. polymorpha*, the CRISPR/Cas9 genome editing system was used to obtain knockout mutant plants in these genes. To increase the probability of obtaining knockouts, a strategy with the combination of two gRNA was followed (figure 39). In this way, the Cas9 can cause two double strand breaks within the sequence of the target gene leading to knockout mutants with large deletions, insertions or inversions.

Figure 39: CRISPR/Cas9 genome editing system strategy. To obtain knockout mutants, a strategy with the combination of two guide RNASs (gRNAs) was used. Hence, the Cas9 can cause two double strand breaks (DBS) within the gene target sequence. Through the non-homologous end joining (NHEJ) pathway, the DSB are repaired causing large deletions or insertions.

3.3.1. Cloning of gRNAs into pMpGE_En04 vector

The CRIPSR/Cas9 system consists in the Cas9 protein that is directed to the target sequence by a gRNA that binds to the target sequence by complementarity. Target recognition is dependent on a PAM sequence that is adjacent to the 3' end of the target sequence. After the gRNA annealing to the target, the Cas9 will cause a double-strand break within it.

gRNAs L and R with 20 nucleotides were designed with the aim to disrupt the MYBI and SHAQKYF-type MYBII domains of the DIV1 and DIV2 proteins; the R motif (R/KLFGV) and the SHAQKYF-type MYBII domain of DIVL protein and the MYB and DUF3755 domains of DRIF protein. For that, four PAM sequences were selected in the genomic sequences of Mp*DIV1* and Mp*DIVL* and five PAM sequences were selected in the Mp*DIV2* and Mp*DRIF* genomic sequences close to the conserved protein domains (figure 40). All the gRNAs were designed using informatic tools to generate gRNAs with 100% homology and to avoid off -targets in the genome (figure 41).

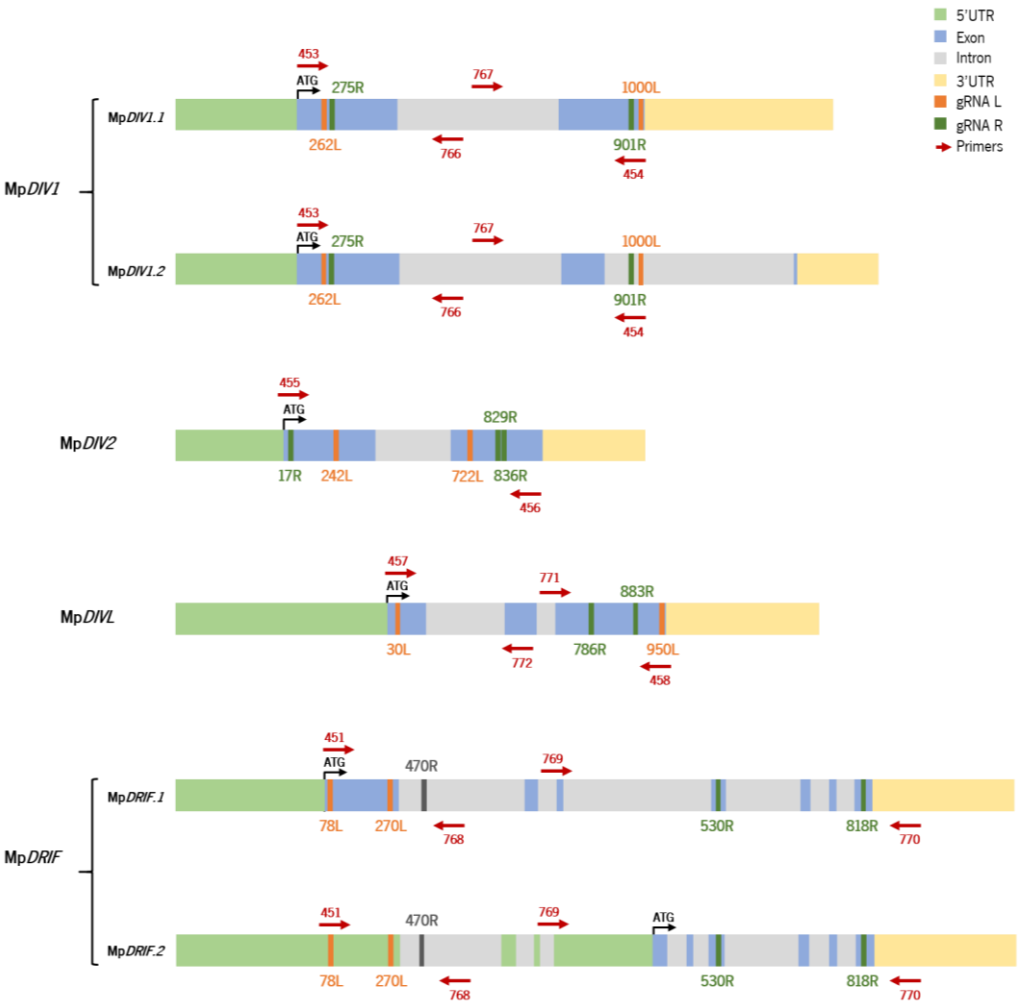

Figure 40: Schematic representation of the Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* genomic sequences. Representation of the 5' UTR, exons, introns and 3'UTR of the possible transcripts for each gene; the gRNA L and R localizations in the genomic sequence and the hybridization localization of the primers used for the plant genotyping.

position		target sequence	sequence information			
start - end	+ −	20mer+PAM (total 23mer)	GC% of 20mer	Tm of 20mer	TTTT in 20mer	restriction sites
273 - 295	+	CACAGCGGATTGGAGCTCTGAGG [gRNA]	60.00 %	75.82 °C	-	MspA1I SacI
279 - 301	+	GGATTGGAGCTCTGAGGAGCAGG [gRNA]	60.00 %	76.48 °C	-	SacI
295 - 317	+	GAGCAGGCCACTCTCGATGATGG [gRNA]	60.00 %	76.59 °C	-	
302 - 324	-	CCACTCTCGATGATGGACTGACA [gRNA]	50.00 %	70.22 °C	-	
331 - 353	+	GTTATTTTAACCATCTGTATTGG [gRNA]	25.00 %	58.58 °C	+	
341 - 363	-	CCATCTGTATTGGTTCCTGAAAC [gRNA]	40.00 %	66.82 °C	-	
356 - 378	-	CCTGAAACTACTAGAACTCTCAA [gRNA]	35.00 %	63.54 °C	-	
363 - 385	+	CTACTAGAACTCTCAAAATCCGG [gRNA]	35.00 %	62.06 °C	-	
382 - 404	-	CCGGCGCTTCTGAGCTCTCGATG [gRNA]	60.00 %	75.06 °C	-	HaeII SacI
408 - 430	+	TCTATACAATCGTCCTTGCCCGG [gRNA]	45.00 %	69.81 °C	-	
421 - 443	+	CCTTGCCCGGTTCCAAGATTTGG [gRNA]	55.00 %	75.91 °C	-	
421 - 443	-	CCTTGCCCGGTTCCAAGATTTGG [gRNA]	55.00 %	75.45 °C	-	
426 - 448	-	CCCGGTTCCAAGATTTGGAATGC [gRNA]	45.00 %	68.90 °C	-	
427 - 449	-	CCGGTTCCAAGATTTGGAATGCC [gRNA]	45.00 %	68.90 °C	-	
433 - 455	-	CCAAGATTTGGAATGCCGTGTTC [gRNA]	45.00 %	69.48 °C	-	
448 - 470	-	CCGTGTTCTACGTCCTCTATCTG [gRNA]	45.00 %	68.69 °C	-	
449 - 471	+	CGTGTTCTACGTCCTCTATCTGG [gRNA]	50.00 %	69.87 °C	-	
461 - 483	-	CCTCTATCTGGTTGCACTTCTCA [gRNA]	45.00 %	69.34 °C	-	
464 - 486	+	CTATCTGGTTGCACTTCTCAAGG [gRNA]	45.00 %	69.34 °C	-	SmlI
465 - 487	+	TATCTGGTTGCACTTCTCAAGGG [gRNA]	40.00 %	67.96 °C	-	SmlI

Figure 41: Mp*DRIF* 470R gRNA design in the *CRISPRdirect* online tool. The Mp*DRIF* sequence was introduced in the tool, different PAM sequences were selected, and the gRNAs were automatically designed. The gRNA positions (that correspond to the sequence base pairs) underlined at green indicate gRNA sequences that are highly specific and that have fewer off-targets. The program further provides information about GC%, the melting temperature (Tm) and possible restriction sites of each gRNA sequence.

The gRNA design was performed as described by Sugano et al. (2018). After selecting the PAM sequence and the 20 nucleotides of the target sequence, 2 oligos (sense and antisense) were designed for each one of the gRNAs (figure 42; Table B2- Annex B). For that, the PAM sequence was removed, and 4 specific nucleotides were added to the 5' end of each one of the oligos. The oligos were annealed and cloned into the digested gateway vector pMpGE_En04 to obtain each one of the gRNAs (figure 42).

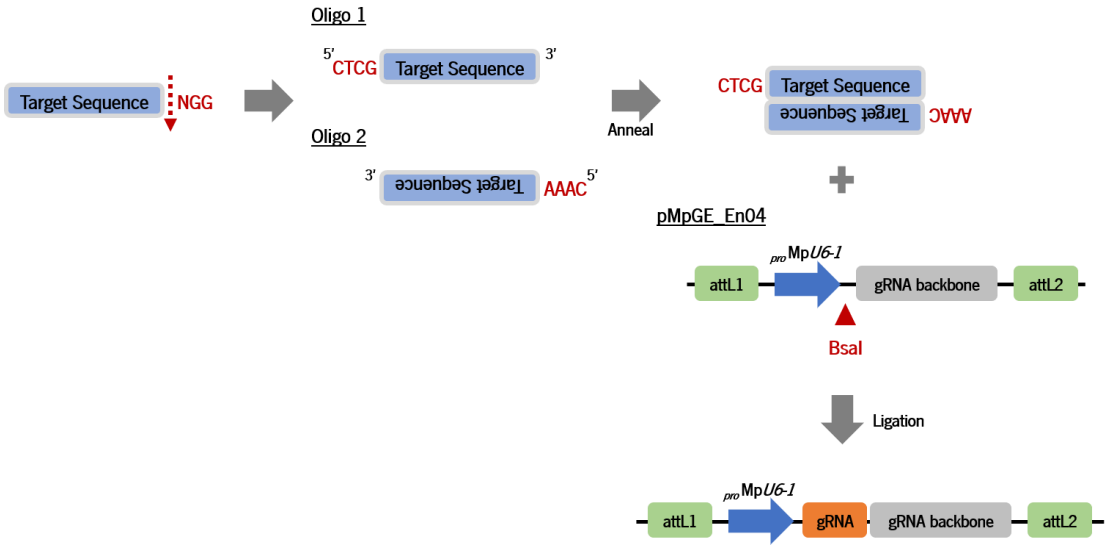

Figure 42: Oligo design and cloning in pMpGE_En04. Two oligos with 20 nucleotides, one sense (Oligo 1) and one antisense (Oligo 2), were designed for each gRNA. After selecting the PAM sequence (NGG) in the target gene sequence, the oligos were designed by removing the PAM sequence and by adding the CTCG nucleotides to the 5'end of the Oligo 1 and the AAAC nucleotides to the 5'end of the Oligo 2. The two oligos were annealed and ligated into the digested pMpGE_En04 vector. This gateway vector has: two *att* recombination sequences that allow the transference of DNA fragments to other gateway vectors; a 0.5 kb fragment of the *pro*Mp*U6-1* for the gRNA expression; the gRNA backbone to express the designed gRNA and a restriction site for the BsaI restriction enzyme. The ligation reaction originates an entry clone with the gRNA sequence.

The pMpGE_En04 gateway vector was linearized with BsaI before ligation with the annealed oligos. The digestion product was run in an agarose gel and a band was observed close to the expected size of 3190 bp (figure 43, lane 1). The DNA fragment was extracted from the agarose gel and purified.

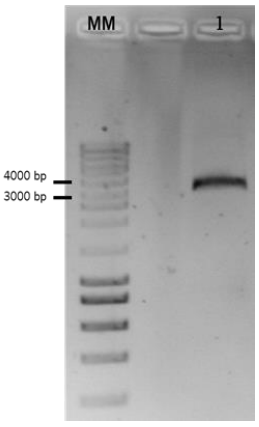

Figure 43: pMpGE_En04 digestion with BsaI.
Lane 1: linearized pMpGE_En04 vector (expected size fragment of 3190 bp);
MM: Molecular marker (NZYDNA Ladder III, Nzytech)

The cloning products were transformed into *E. coli* DH5α or DH10β competent cells. The transformed colonies were capable of growing in LB medium supplemented with 100 μg mL^{-1} spectinomycin. A colony PCR was then conducted to select the positive transformants of each one of the gRNAs cloned in pMpGE_En04. The screening was performed with the M13 primer forward and the specific antisense gRNA oligo (figure 44; Table B1 and B2- Annex B). A fragment of 690 bp was expected as the product of the PCR amplification for a positive clone.

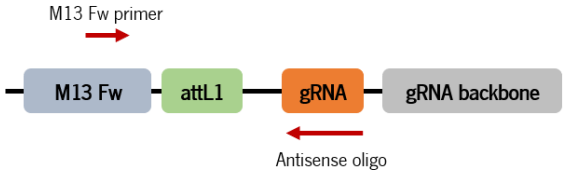

Figure 44: Hybridization localization of the primers used in the screening of gRNApMpGE_En04 colonies by PCR. The M13 forward (Fw) primer and the antisense oligo of each gRNA were used in the screening.

At least one positive clone showed a fragment with of the expected size for each one of the gRNAspMpGE_En04 constructs (figure 45; Figure F1- Annex F).

The plasmids of Mp*DIV1* 1000LpMpGE_En04 clone 6, Mp*DIV2*836RpMpGE_En04 clone 5, Mp*DIVL* 786RpMpGE_En04 clone 4 and Mp*DRIF* 270LpMpGE_En04 clone 6 (figure 45) were isolated and sequenced, confirming that the inserted gRNA was in the desired orientation.

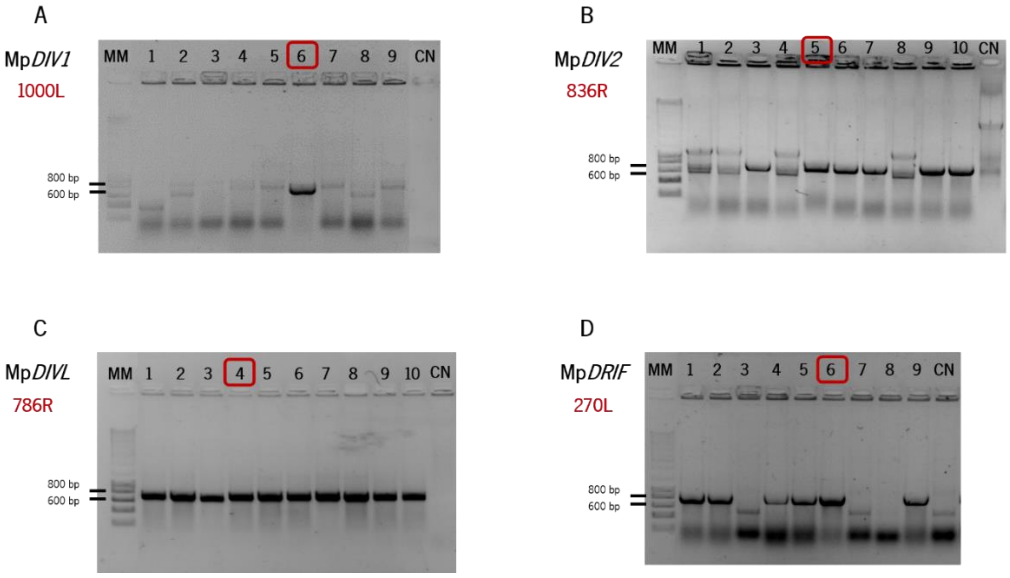

Figure 45: PCR screening of *E. coli* colonies containing constructs with the gRNAs L and R in pMpGE_En04 for CRISPR/Cas9 system.
A) Mp*DIV1* 1000LpMpGE_En04 colonies, lanes 1 to 9 (expected size fragment of 690 bp);
B) Mp*DIV2* 836RpMpGE_En04 colonies, lanes 1 to 10 (expected size fragment of 690 bp);
C) Mp*DIVL* 786RpMpGE_En04 colonies, lanes 1 to 10 (expected size fragment of 690 bp);
D) Mp*DRIF* 270LpMpGE_En04 colonies, lanes 1 to 9 (expected size fragment of 690 bp);
MM: Molecular marker (NZYDNA Ladder III, Nzytech); CN: Negative control;
The lanes with red squares correspond to the colonies selected for sequencing.

3.3.2. Cloning into pMpGE010 and pMpGE011

The gRNAs cloned into the pMpGE_En04 vector were subsequently cloned into two final destination vectors- pMpGE010 and pMpGE011 (figure 46). These destination vectors are binary vectors that harbour the gRNA and the Cas9 expressed under the strong promoter Mp*EF1α*. Also, these two vectors have different plant resistance markers, which allows the selection of a plant transformed with two different gRNAs. Therefore, the gRNAs L were cloned into pMpGE010 and gRNAs R were cloned into pMpGE011. The cloning was achieved by performing an LR reaction (*Gateway® cloning technology*) that consists in the transference of the gRNA from the entry clone to a destination vector.

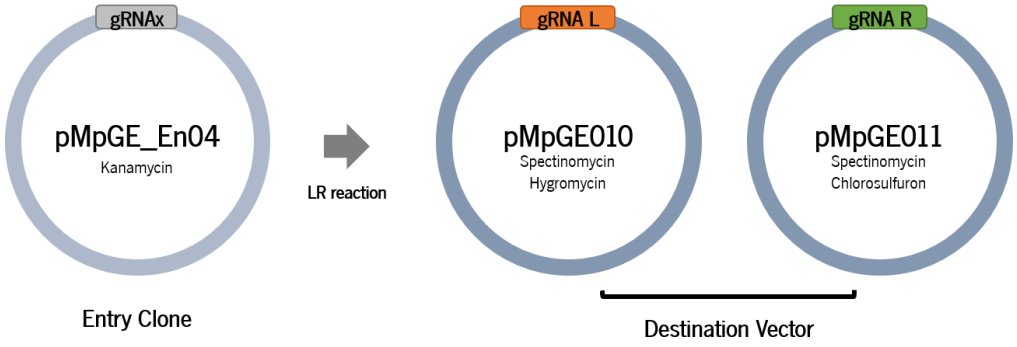

Figure 46: Cloning of gRNA L and R into pMpGE010 and pMpGE011 vectors. The gRNAs cloned into pMpGE_En04 (Entry clone) were transferred by an LR reaction (Gateway® cloning technology) into the two destination vectors pMpGE010 and pMpGE011. The gRNAs L were cloned into pMpGE010 that have the spectinomycin bacterial resistance marker gene and the hygromycin plant resistance marker gene. The gRNAs R were cloned into pMpGE011 that have the spectinomycin bacterial resistance marker gene and the chlorsulfuron plant resistance marker gene.

The cloning products were transformed into *E. coli* DH10β competent cells. The transformed colonies were capable of growing in LB medium supplemented with 100 µg mL^{-1} spectinomycin. A colony PCR was then conducted to select the positive transformants of gRNA L into pMpGE010 and gRNA R into pMpGE011. The transformants screening was performed with the specific sense gRNA oligo and the NOS terminator primer (figure 47; Table B1 and B2- Annex B). A fragment of 274 bp was expected as the product of the PCR amplification for a positive clone.

Figure 47: Hybridization localization of the primers used in the screening of gRNALpMpGE010 and gRNARpMpGE011 colonies by PCR. The specific gRNA sense oligo and the NOS terminator primer were used in the screening.

At least one positive clone showed a fragment with the expected size for the gRNAL$_{pMpGE010}$ and gRNAR$_{pMpGE011}$ constructs (figure 48; Figure F2- Annex F). The plasmids of Mp*DIV1* 1000L$_{pMpGE010}$ clone 1, Mp*DIV2* 836R$_{pMpGE011}$ clone 5, Mp*DIVL* 786R$_{pMpGE011}$ clone 5 and Mp*DRIF* 270L$_{pMpGE010}$ clone 1 (figure 48) were isolated and sequenced. Constructs had the inserted gRNA in the desired orientation.

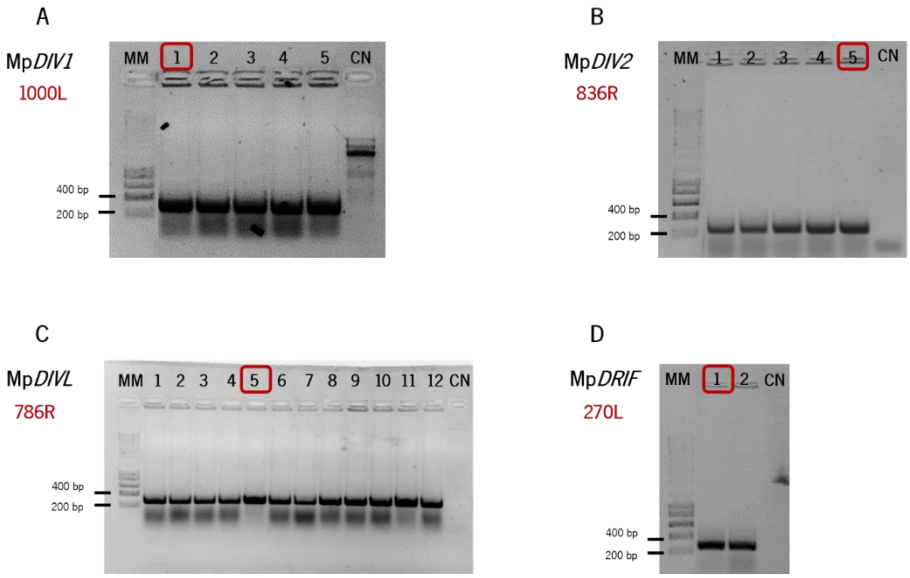

Figure 48: PCR screening of *E. coli* colonies containing constructs with the gRNAs L and R in pMpGE010 and pMpGE011, respectively, for CRISPR/Cas9 system.
A) Mp*DIV1* 1000LpMpGE010 colonies, lanes 1 to 5 (expected size fragment of 274 bp);
B) Mp*DIV2* 836RpMpGE011 colonies, lanes 1 to 5 (expected size fragment of 274 bp);
C) Mp*DIVL* 786RpMpGE011 colonies, lanes 1 to 12 (expected size fragment of 274 bp);
D) Mp*DRIF* 270LpMpGE010 colonies, lanes 1 to 2 (expected size fragment of 274 bp);
MM: Molecular marker (NZYDNA Ladder III, Nzytech); CN: Negative control;
The lanes with red squares correspond to the colonies selected for sequencing.

3.3.3. *Marchantia polymorpha* spore transformation

M. polymorpha knockout lines were obtained by *Agrobacterium*-mediated transformation of spores. Competent cells of the GV2260 *A. tumefaciens* strain were transformed with the constructs described above (gRNAs L into pMpGE010 and gRNAs R into pMpGE011). The colonies were screened by colony PCR using the specific sense gRNA oligo and the NOS terminator primer (figure 47).

Mp*DIV1* 1000L$_{pMpGE010}$ clone 1, Mp*DIV2* 836R$_{pMpGE011}$ clone 2, Mp*DIVL* 883R$_{pMpGE011}$ clone 1 and Mp*DRIF* 270L$_{pMpGE010}$ clone 2 showed a fragment with the expected size (figure 49). A fragment with the same size was observed for the other gRNAs screening.

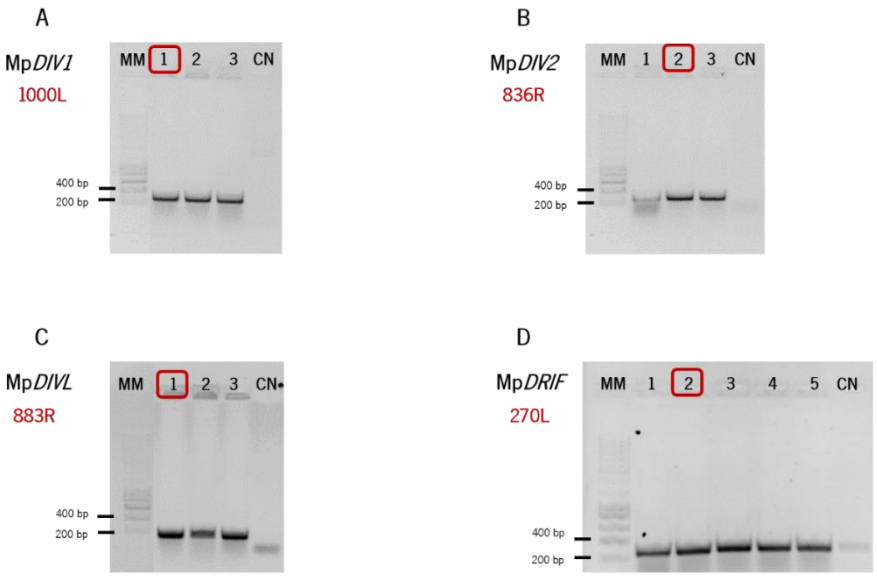

Figure 49: PCR screening of *Agrobacterium* colonies containing constructs with the gRNAs L and R in pMpGE010 and pMpGE011, respectively, for CRISPR/Cas9 system.
A) Mp*DIV1* 1000LpMpGE010 colonies, lanes 1 to 3 (expected size fragment of 274 bp);
B) Mp*DIV2* 836RpMpGE011 colonies, lanes 1 to 3 (expected size fragment of 274 bp);
C) Mp*DIVL* 883RpMpGE011 colonies, lanes 1 to 3 (expected size fragment of 274 bp);
D) Mp*DRIF* 270LpMpGE010 colonies, lanes 1 to 5 (expected size fragment of 274 bp);
MM: Molecular marker (NZYDNA Ladder III, Nzytech); CN: Negative control;
The lanes with red squares correspond to the colonies selected for the spore transformation.

Sterilised spores were transformed with an overnight culture of *A. tumefaciens* GV2260, with the constructs mentioned above in Sabine Zachgo's laboratory (University of Osnabrück, Germany). Plates with the transformants were sent to our research group and screened by observing their survival in the selective media with 10 mg L^{-1} hygromycin, 100 mg L^{-1} cefotaxime and 0.5 µM chlorsulfuron. The possible gRNA combinations successfully transformed in *M. polymorpha* spores are represented in the Table 3. For each combination, approximately 100 transformed plants were obtained.

Table 3: Possible gRNAs combinations for each gene. The combinations in bold are the ones that were successfully transformed in *Marchantia polymorpha* spores.

Possible gRNA combinations			
Mp*DIV1*	**Mp*DIV2***	**Mp*DIVL***	**Mp*DRIF***
	242L + 17R		**78L + 530R**
262L + 275R	**242L + 829R**	**30L + 786R**	**78L + 818R**
262L + 901R	**242L + 836R**	30L + 883R	78L + 470R
1000L + 275R	**722L + 17R**	950L + 786R	270L + 530R
1000L + 901R	722L + 829R	950L + 883R	270L + 818R
	722L + 836R		270L+ 470R

3.3.4. Plant genotyping

Approximately 1-month-old plant positive transformants were plated into half strength Gamborg B5 medium with 1.4% (w/v) plant agar and 10 mg L^{-1} hygromycin, 100 mg L^{-1} cefotaxime and 0.5 µM chlorsulfuron. Wild-type plants were also cultivated during this time to serve as control. After 3 weeks, DNA was extracted from plant tissue and genotyped by PCR amplification using specific primers (Table B1- Annex B). To genotype the putative transformants, three PCR reactions with three different pairs of primers were performed for all genes, except for Mp*DIV2* due to the small gene size (figure 40). The combination of primers and the expected size for wild-type plants are listed in Table 4. Transformed plants with an insertion or deletion in the genomic sequence of each gene were expected to generate a PCR amplification product with bigger or smaller size, respectively.

Table 4: *Marchantia polymorpha* specific gene primers used in the transformants genotyping. The combination of primers (Table B1- Annex B) and the expected size of the PCR fragment amplified in wild-type plants.

Gene	Primer combination	Size (bp)
Mp*DIV1*	453 + 766	1084
	767 + 454	918
	453 + 454	2124
Mp*DIV2*	455 + 456	1159
Mp*DVL*	457 + 772	721
	771 + 458	811
	457 + 458	1731
Mp*DRIF*	451 + 768	731
	769 + 770	1283
	451 + 770	3619

The Mp*DIV1* transformants of the gRNA combination 262L with 901R were genotyped with three pairs of primers (Table 4). Transformants 69 and 87 were found when the plants were genotyped with the 767+454 pair. The expected size of the fragment for a wild-type plant is 918 bp. The plants 69 and 87 showed a smaller size fragment and, therefore, were selected for sequencing (figure 50).

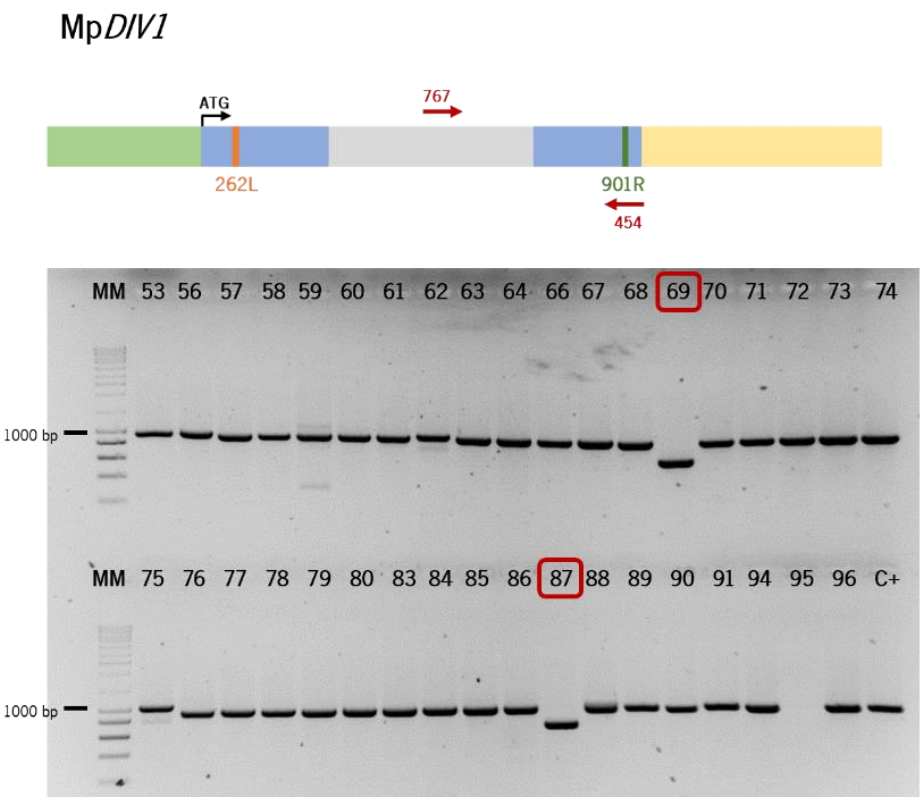

Figure 50: PCR genotyping of *Marchantia polymorpha* Mp*DIV1* transformed plants. Transformed plants 53 to 96 of the gRNA combination 262L+901R were genotyped using the pair of primers 454+767. The expected size of the PCR fragment amplified in wild-type plants was 918 bp. Lane numbers with red squares represent plants whose genomic PCR amplification produced a fragment with a different size from the wild type. MM: Molecular marker (NZYDNA Ladder III, Nzytech); C+: Positive control with wild-type genomic DNA.

The Mp*DIV2* transformants of the three different gRNA combinations (242L with 836R; 722L with 17R and 242L with 829R) were genotyped with only one pair of primers (Table 4). The expected size of the fragment for a wild-type plant with the 455+456 pair of primers is 1159 bp. Fourteen plants showed a smaller size fragment: for the gRNA combination 242L with 836R, the plants 5, 24, 68, 116, 217 and 242 (figure 51; figure G1- Annex G); for the gRNA combination 722L with 17R, the plants 99, 127, 189 and 305 and for the combination 242L with 829R, the plants 135, 169, 257 and 296. Only one plant showed a bigger size fragment: the plant 93 of the gRNA combination 722L with 17R (figure G1- Annex G). The DNA of all these plants was sequenced.

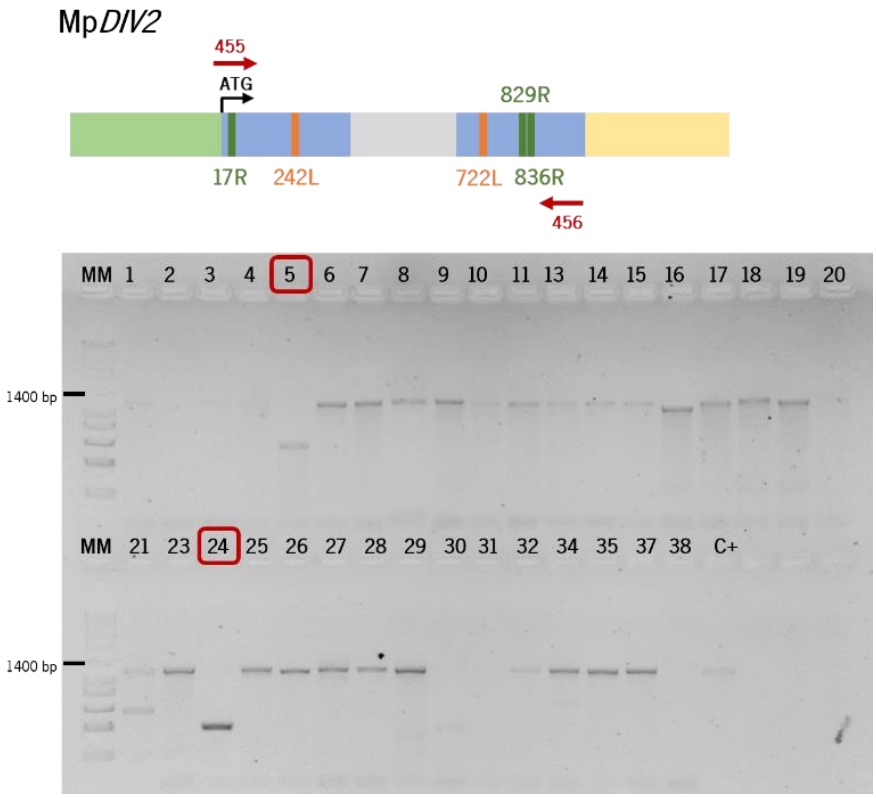

Figure 51: PCR genotyping of *Marchantia polymorpha* Mp*DIV2* transformed plants. Transformed plants 1 to 38 of the gRNA combinations 242L+836R, 722L+17R and 242L+829R were genotyped using the pair of primers 455+456. The expected size of the PCR fragment amplified in wild-type plants was 1159 bp. Lane numbers with red squares represent plants whose genomic PCR amplification produced a fragment with a different size from the wild type. MM: Molecular marker (NZYDNA Ladder III, Nzytech C+: Positive control with wild-type genomic DNA.

The *MpDIVL* transformants of the gRNA combination 30L with 786R were genotyped with three pairs of primers (Table 4). When the plants were genotyped with the 457+772 pair no transformants were found. When genotyped with the 771+458 pair, two transformants were found. The expected size of the fragment for a wild-type plant is 811 bp and the plants 26 and 73 showed a bigger size fragment (figure 52A). One transformant was found when the plants were genotyped with the 457+458 pair. The expected size of the fragment for a wild-type plant is 1731 bp and the plant 87 showed a smaller size fragment (figure 52B).

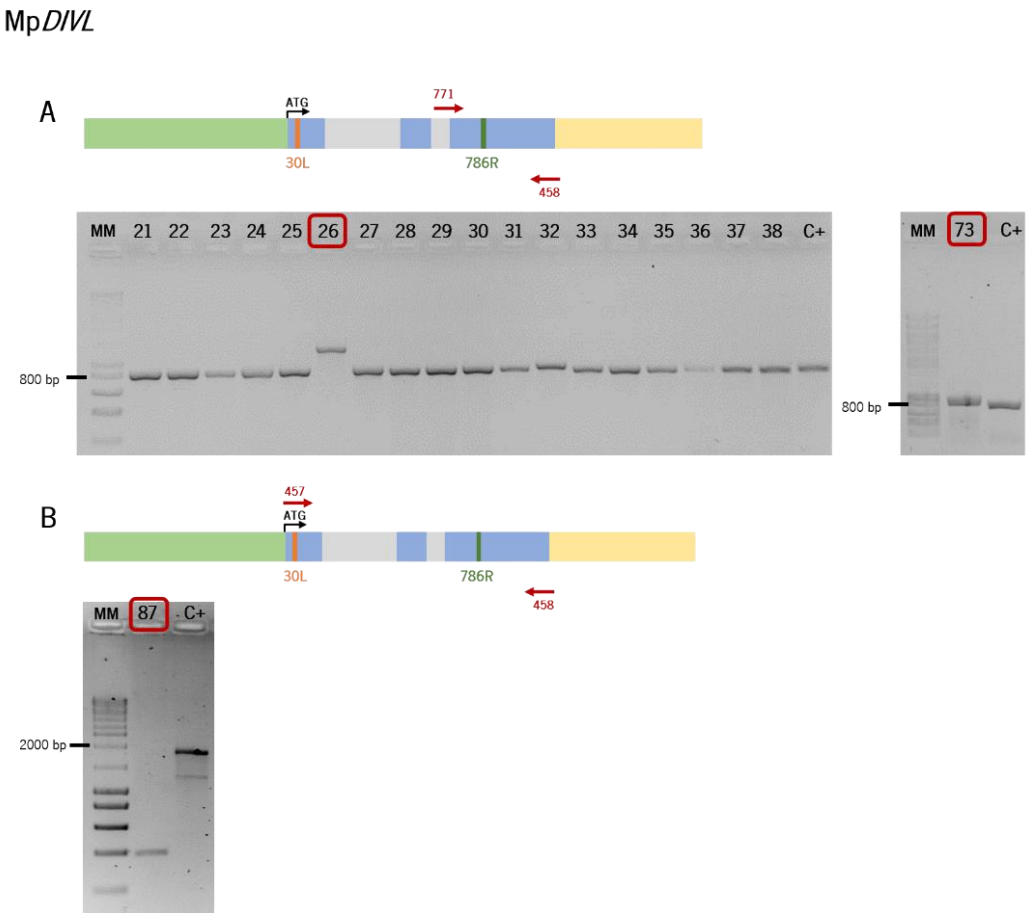

Figure 52: PCR genotyping of *Marchantia polymorpha* Mp*DIVL* transformed plants.
A) Transformed plants 21 to 38 and 73 of the gRNA combination 30L+786R were genotyped with the pair of primers 771+458. The expected size of the PCR fragment amplified in wild-type plants was 811 bp.
B) Transformed plant 87 of the gRNA combination 30L+786R was genotyped with the pair of primers 457+458. The expected size of the PCR fragment amplified in wild-type plants was 1731 bp.
Lane numbers with red squares represent plants whose genomic PCR amplification produced a fragment with a different size from the wild type. MM: Molecular marker (NZYDNA Ladder III, Nzytech); C+: Positive control with wild-type genomic DNA.

For Mp*DRIF*, no transformants were found for gRNA combinations 78L with 530R and 78L with 818R. The plants were genotyped with the three pairs of primers (Table 4) but no differences in the size fragment were observed. This result could be explained by the considerable distance (2000-3500 bp) that exists between the two gRNAs L and R. To reduce this distance and in order to obtain Mp*DRIF* transformants, a new gRNA R- the gRNA 470R - was designed to be combined with the gRNA 78L (figure 40). The new gRNA 470R was successfully cloned into pMpGE_En04 and subsequently into pMpGE011.

3.3.5. Mp*div1*, Mp*div2* and Mp*divl* expression

A semi-quantitative RT-PCR was performed to confirm the Mp*DIV1*, Mp*DIVL* and Mp*DIV2* genes knockout by the analysis of the transcript abundance in the respective mutant plants. The analysis was performed using tissue from completely developed plants (with 1-month-old).

RNA was extracted from whole wild-type and Mp*DIV1*, Mp*DIVL* and Mp*DIV2* mutant plants with 1-month-old by the CTAB method. Due to the amount of Mp*div2* knockout lines, the plants #93, #127, #217 and #296 were selected for the RNA extraction and consequent transcripts abundance analysis. The semi-quantitative PCR was performed using Mp*DIV1*, Mp*DIVL* and Mp*DIV2* transcript specific primers and Mp*EF1α* gene as a reference gene.

The Mp*DIV1* transcript abundance in the plants Mp*div1* #69 and #87 is lower than the wild-type plant suggesting that this gene was knocked down in these two mutant plants (figure 53A). In the plants Mp*divl* #26 the abundance of transcripts is smaller when compared to wild type while for the plant line #73, the abundance of Mp*DIVL* transcripts is similar to wild type which suggests that Mp*DIVL* expression was not affected (figure 53B). In the plant line #87 no Mp*DIVL* expression was detected. In the plants Mp*div2* #127, #217 and #296 the Mp*DIV2* gene had no expression when compared to wild type suggesting the Mp*DIV2* is knocked-out in these plants. Mp*DIV2* expression is normal in the Mp*div2* plant line #93 (figure 53C).

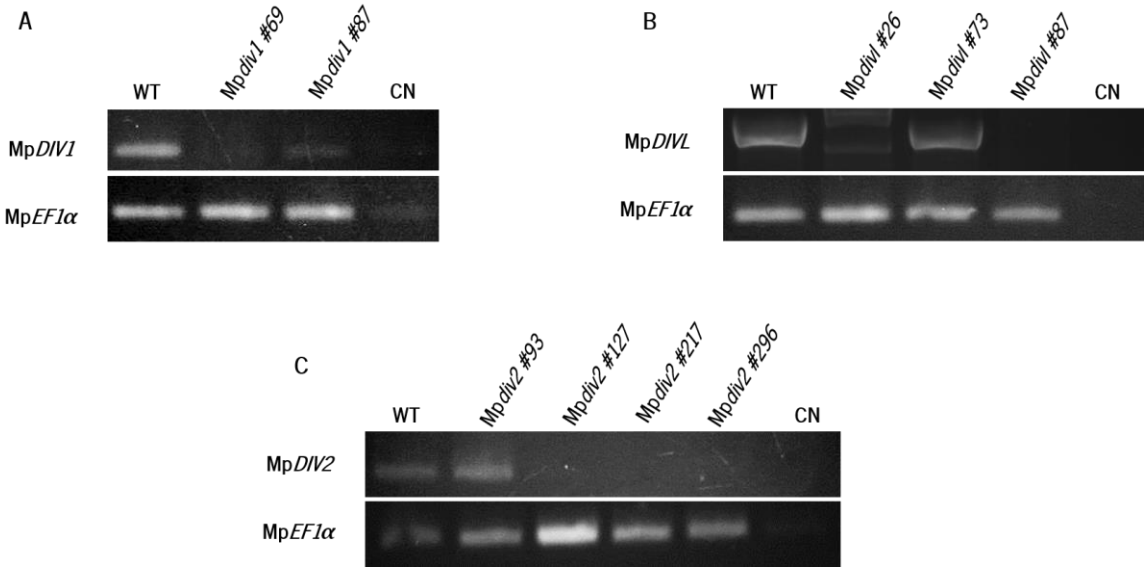

Figure 53: Analysis by RT-PCR of Mp*DIV1*, Mp*DIVL* and Mp*DIV2* expression in knockout mutant plants. A) Mp*DIV1* transcripts abundance in wild-type plants and Mp*div1* mutant plant lines #69 and #87. **B)** Mp*DIVL* transcripts abundance in wild-type plants and Mp*divl* mutant plant lines #26, #73 and #87. **C)** Mp*DIV2* transcripts abundance in wild-type plants and Mp*div2* mutant plant lines #93, #127, #217 and #296. Mp*DIV1*, Mp*DIVL* and Mp*DIV2* were amplified in a 35 cycle-PCR. The constitutive gene Mp*EF1α* was used as control. The transcripts abundance was analysed by electrophoresis. CN: Negative control.

3.3.6. Sequence analysis of mutant plants

The loci of the selected Mp*DIV1*, Mp*DIV2* and Mp*DIVL* transformant plants were sequenced to determine whether the plants had insertions or deletions in the target gene sequences.

The mutant plant Mp*div1* #69 has an insertion of 12 bp and a deletion of 270 bp, while the mutant plant Mp*div1* #87 has a deletion of 229 bp in the second exon (figure 54A). These insertions and deletions gave origin to Mpdiv1 #69 and #87 mutant proteins with a SHAQKYF motif of the MYBII domain replaced by the amino acids GATTG (figure 54B).

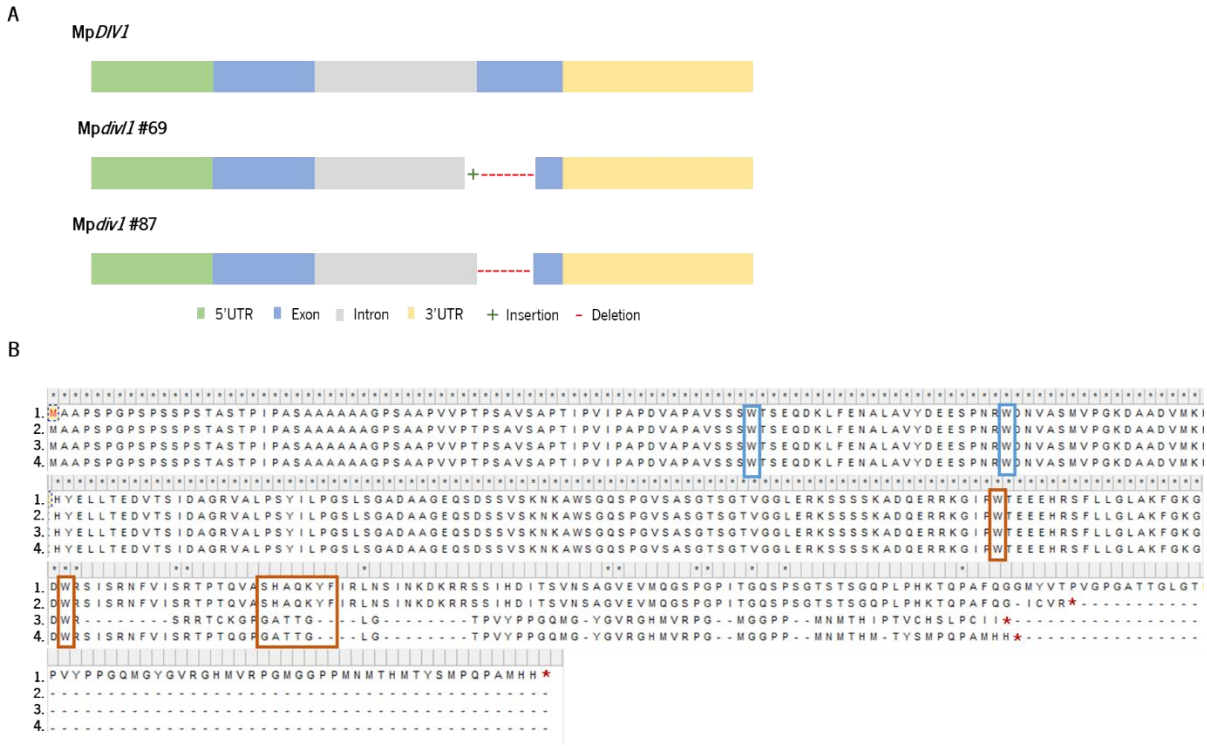

Figure 54: Schematic representation of Mp*DIV1* and Mp*div1* loci in *Marchantia polymorpha* mutant lines, and wild-type and mutant proteins alignment. A) Schematic representation of the Mp*DIV1* genomic sequence and the Mp*div1* lines #69 and #87 insertions and deletions. Mp*div1* #69 has an insertion of 12 bp and a deletion of 270 bp in the second exon. Mp*div1* #87 has a deletion of 229 bp in the second exon. **B)** Wild-type [MpDIV1.1 (1) and MpDIV1.2 (2)] and mutant proteins of the lines #69 (3) and #87 (4) were aligned by the ClustalW algorithm. MpDIV1 conserved MYB domains are represented by the blue- MYBI domain (-W-X$_{23}$-W-X$_{20}$-Y-)- and the orange- MYBII domain (-W-X$_{19}$-W-X$_{22}$-Y-)- colours. The MYBI domain of the Mpdiv1 mutant proteins was not affected while the SHAQKYF motif of the MYBII domain was replaced by the amino acids GATTG. The red asterisks represent the translation termination.

For the Mp*div1* mutant plants, the sequencing results showed that the plant #26 has an insertion of 402 bp and a deletion of 2 bp in the second exon, the plant #73 has an insertion of 102 bp also in the second exon and the plant #87 a deletion of 767 bp that includes the 3 exons and the 2 introns (figure 55A). The insertions of the mutant lines #26 and #73 gave origin to mutant proteins with no changes in the R motif nor in the MYBII conserved domain of the DIVL protein. The deletion that occurred in the mutant line #87 gave origin to a protein with the R motif and the MYBII domain disrupted (figure 55B).

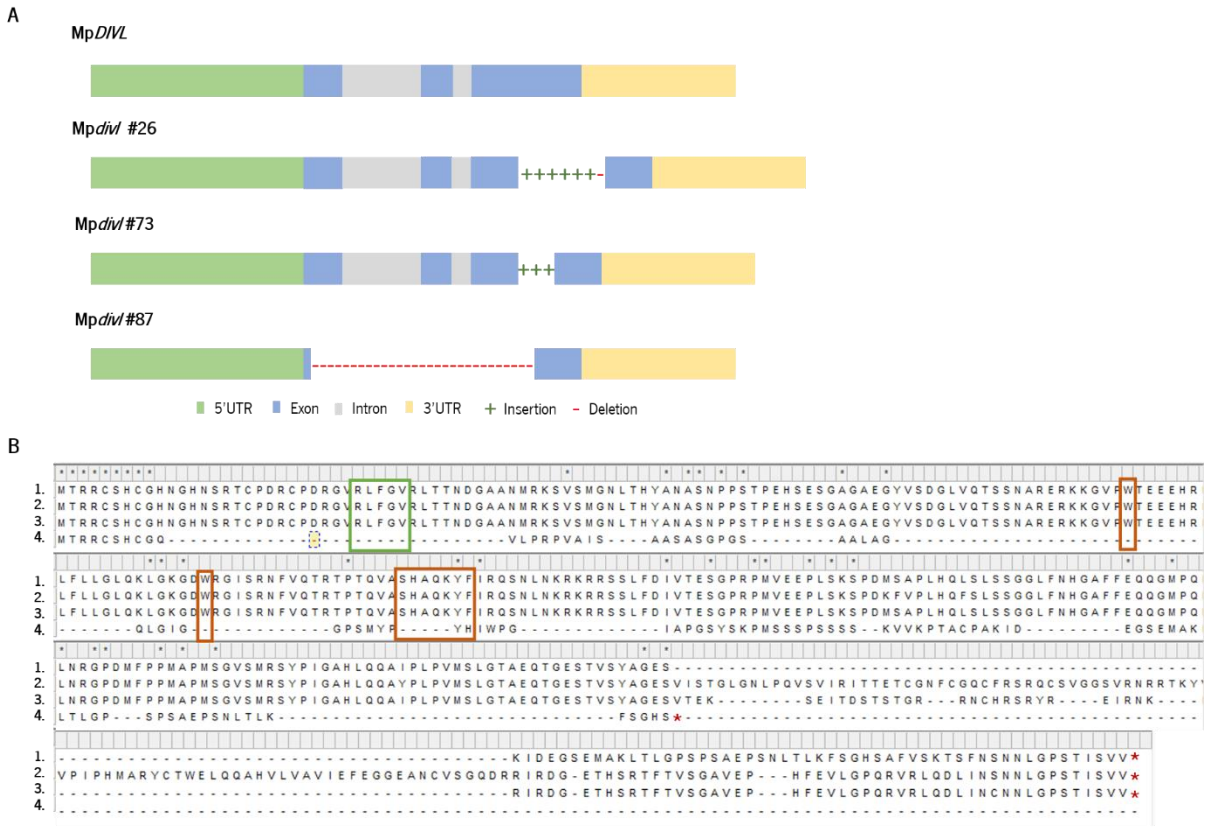

Figure 55: Schematic representation of Mp*DIVL* and Mp*divl* loci in *Marchantia polymorpha* mutant lines, and wild-type and mutant proteins alignment. A) Schematic representation of the Mp*DIVL* genomic sequence and the Mp*divl* lines #26, #73 and #87 insertions and deletions. Mp*divl* #26 has an insertion of 402 bp and a deletion of 2 bp in the second exon. Mp*divl* #73 has an insertion of 102 bp in the second exon. Mp*divl* #87 has a deletion of 767 bp in the 3 exons and the 2 introns. **B)** Wild-type MpDIVL (1) and mutant proteins of the lines #26 (2), #73 (3) and #87 (4) were aligned by the ClustalW algorithm. MpDIVL conserved R motif (KLFGV) and MYBII domains (-W-X19-W-X22-Y-) are represented by the green and orange colours respectively. The R motif and the MYBII domain of the Mpdivl mutant proteins #26 and #73 were not affected by the insertions and deletions while the both domains in the mutant protein #87 were disrupted. The red asterisks represent the translation termination.

The deletions and insertions that occur in the loci of Mp*DIV2* mutant plants were similar, so they were divided in four groups to facilitate the study of the gene function (figure 56A). The first group is composed by 8 independent lines of plant mutants (#5, #24, #68, #116, #135, #217, #242 and #257) that contain a deletion of 300-500 bp in the intron and part of the two exons. The second group is composed for 2 independent lines, #169 and #296, that contain deletions of 351 and 145 bp, respectively, in the first and second exon and a deletion of 21 and 97 bp, respectively, in the intron. The third group has 4 independent lines of plant mutants (#99, #127, #189 and #305) that contain a deletion of ≈ 950 bp in almost all the coding region. The fourth and last group is composed by only one plant mutant #93 that contains an insertion of 75 bp and a deletion of 47 bp in the second exon (figure 56A). The size of all deletions and insertions of all mutants is resumed in Table G1- Annex G.

The deletion that occurred in the loci of the Mp*div2* Group I mutant plant lines gave origin to mutant proteins with the disruption of the MYBII domain (figure 56B; Figure H1- Annex H). In the Mp*div2*

mutant plant lines of Group II the deletions and insertions originated a Mpdiv2 #169 protein with the disruption of the MYBII domain with the SHAQKYF motif replaced by the amino acids SGEPRPK (figure 56B) and a Mpdiv2 #296 with the complete disruption of the MYBII domain (figure H2- Annex H). The large deletion that occurred in the loci of the Mp*div2* Group III mutant plant lines gave origin to proteins with the two conserved MYB domains completely disrupted (figure 56B; figure H3-Annex H). Lastly, the insertion and deletion that occurred in the loci of the Mp*div2* Group IV mutant plant line #93 did not disrupt the MYB domains of the MpDIV protein (Group IV) (figure 56B).

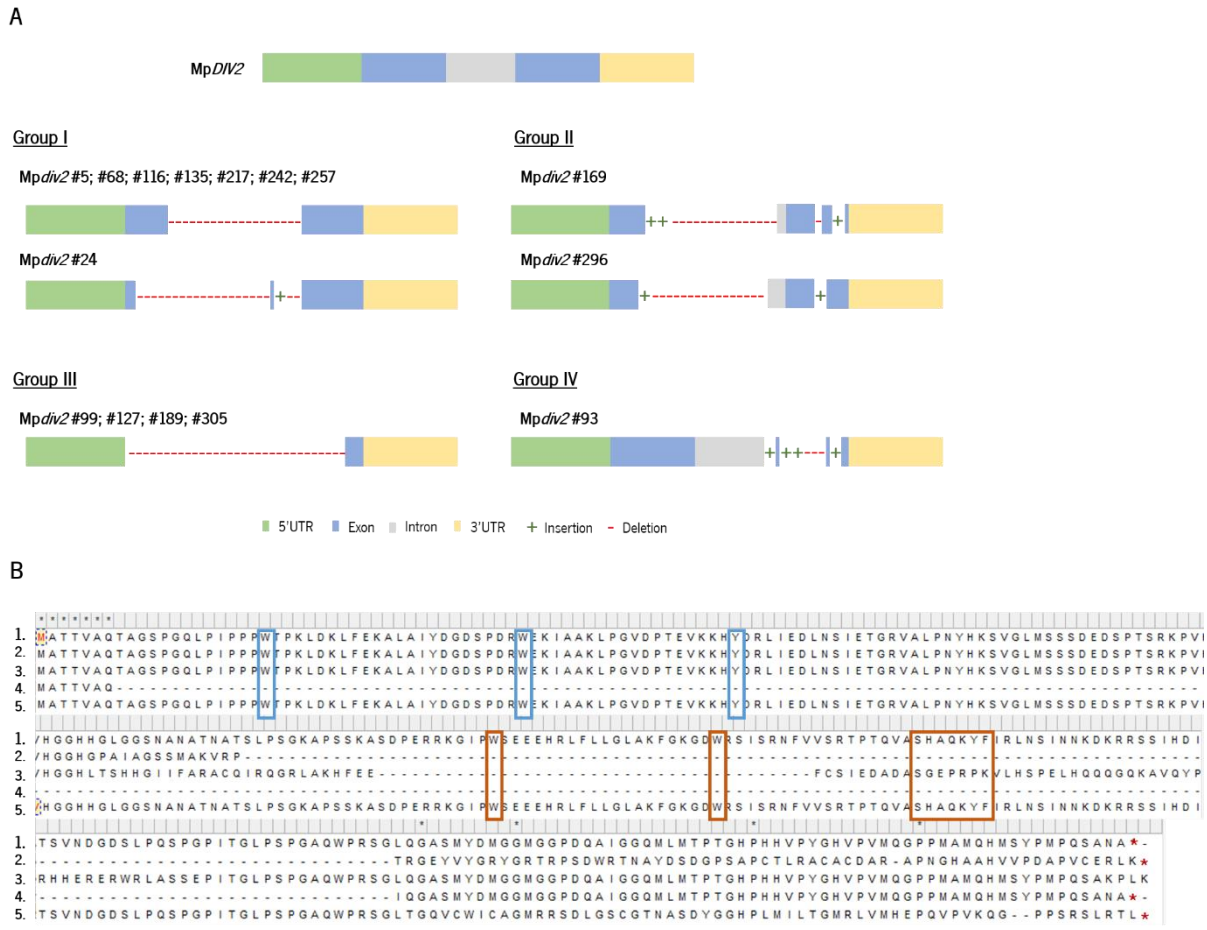

Figure 56: Schematic representation of Mp*DIV2* and Mp*div2* loci in *Marchantia polymorpha* mutant lines, and wild-type and mutant proteins alignment. A) Schematic representation of Mp*DIV2* genomic sequence and the deletions and insertions of the different plant mutant lines. The localizations of the insertions and deletions were similar for some Mp*div2* lines, so they were divided in groups. The first group is composed by the independent mutant lines #5, #24, #68, #116, #135, #217, #242 and #257 in which occurred a deletion of 300-500 bp in the two exons. The second group is composed by 2 independent lines, #169 and #296, that contain deletions of 351 and 145 bp, respectively, in the first and second exon and a deletion of 21 and 97 bp, respectively, in the intron. The third group has the independent mutant lines #99, #127, #189 and #305 in which occurred a deletion of ≈ 950 bp in the first exon and in part of the second exon. The fourth group is composed for the mutant line #93 in which occurred an insertion of 75 bp and a deletion of 47 bp in the second exon. **B)** Wild-type MpDIV2 (1) and mutant proteins of the lines #5 of Group I (2), #169 of Group II (3), #99 of Group II (4) and #93 of Group IV (5) were aligned by the ClustalW algorithm. MpDIV2 conserved MYB domains are represented by the blue- MYBI domain (-W-X23-W-X20-Y-)- and the orange- MYBII domain (-W-X19-W-X22-Y-)- colours. The deletions and insertions gave origin to mutant proteins of: Group I with the disruption of the MYBII domain; Group II with the disruption of the MYBII domain with the SHAQKYF motif replaced by the amino acids SGEPRPK; Group III with the two conserved MYB domains completely disrupted and Group IV with non-disrupted MYB domains. The red asterisks represent the translation termination.

3.3.7. Determination of the knockout plants sex

A multiplex PCR was performed using DNA extracted from a thallus section of approximately 5 cm^2 to distinguish the transgenic plants sex in early stages of development. Plants Mp*div1* #69 and #87; Mp*divl* #73 and #87 and Mp*div2* #24, #68, #116, #169, #189, #217, #242, #257 and #296 are male. Plants Mp*divl* #26 and Mp*div2* #5, #93, #99, #127, #135 and #305 are female (figure 57; figure G2- Annex G). The sex of all mutants is resumed in Table G1- Annex G.

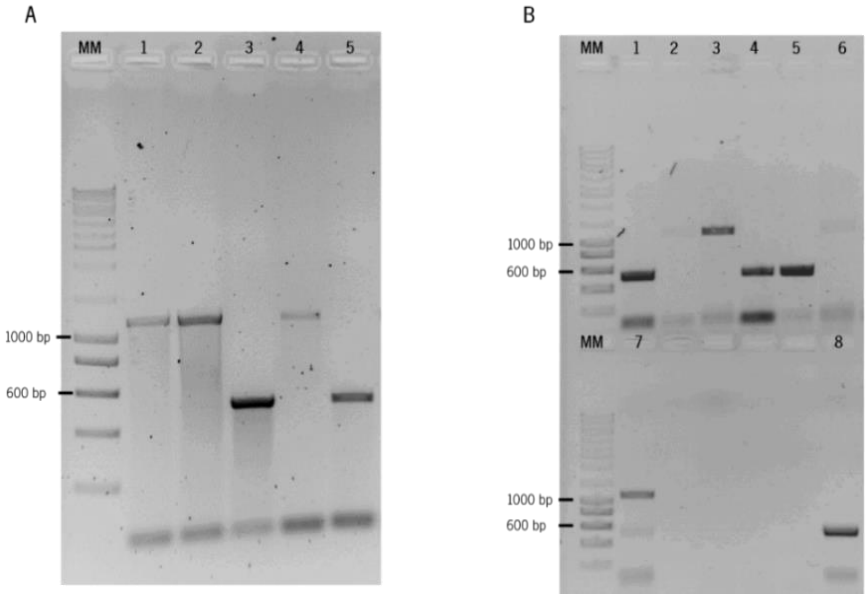

Figure 57: Determination of the Mp*div1*, Mp*divl* and Mp*div2* mutant plants sex by PCR amplification. For a male plant was expected a fragment with a size close to 1100 bp and for a female plant a fragment with a size close to 516 bp.
A) Determination of Mp*div1* and Mp*divl* plants sex.
Lane 1: Mp*div1* line #69 (≈ 1100 bp fragment);
Lane 2: *Mpdiv1* line #87 (≈ 1100 bp fragment);
Lane 3: Female plant control (≈ 516 bp fragment);
Lane 4: Male plant control (≈ 1100 bp fragment);
Lane 5: Mp*divl* line #26 (≈ 516 bp fragment);
B) Determination of Mp*div2* plants sex.
Lane 1: Mp*div2* line #5 (≈ 516 bp fragment);
Lane 2: Mp*div2* line #24 (≈ 1100 bp fragment);
Lane 3: Mp*div2* line #116 (≈ 1100 bp fragment);
Lane 4: Mp*div2* line #127 (≈ 516 bp fragment);
Lane 5: Mp*div2* line #135 (≈ 516 bp fragment);
Lane 6: Mp*div2* line #242 (≈ 1100 bp fragment);
Lane 7: Mp*div2* line #296 (≈ 1100 bp fragment);
Lane 8: Female plant control (≈ 516 bp fragment);
MM: Molecular marker (NZYDNA Ladder III, Nzytech).

3.3.8. Phenotype observation

Mp*div1*, Mp*div2* and Mp*divl* mutant plants were followed and photographed for approximately 1-month at different stages of thallus development. All the mutant plants were asexually propagated and plated into half strength Gamborg B5 medium with 1.4% (w/v) plant agar. Wild-type plants were also grown during this time to serve as control.

3.3.8.1. Analysis of Mp*div1* mutant plants phenotype

Mp*div1* mutant plants #69 and #87 exhibited some phenotypic differences from 17 days to 31 days after the gemmae propagation and appear to have a smaller thallus when compared to the wild-type plant (figure 58A). To confirm these differences, the thallus surface area of the two independent mutant lines was measured for 1-month using the software *ImageJ*. It was possible to observe that the thallus surface areas of the plant lines #69 and #87 are smaller than wild type 10 days after the gemma propagation (figure 58B).

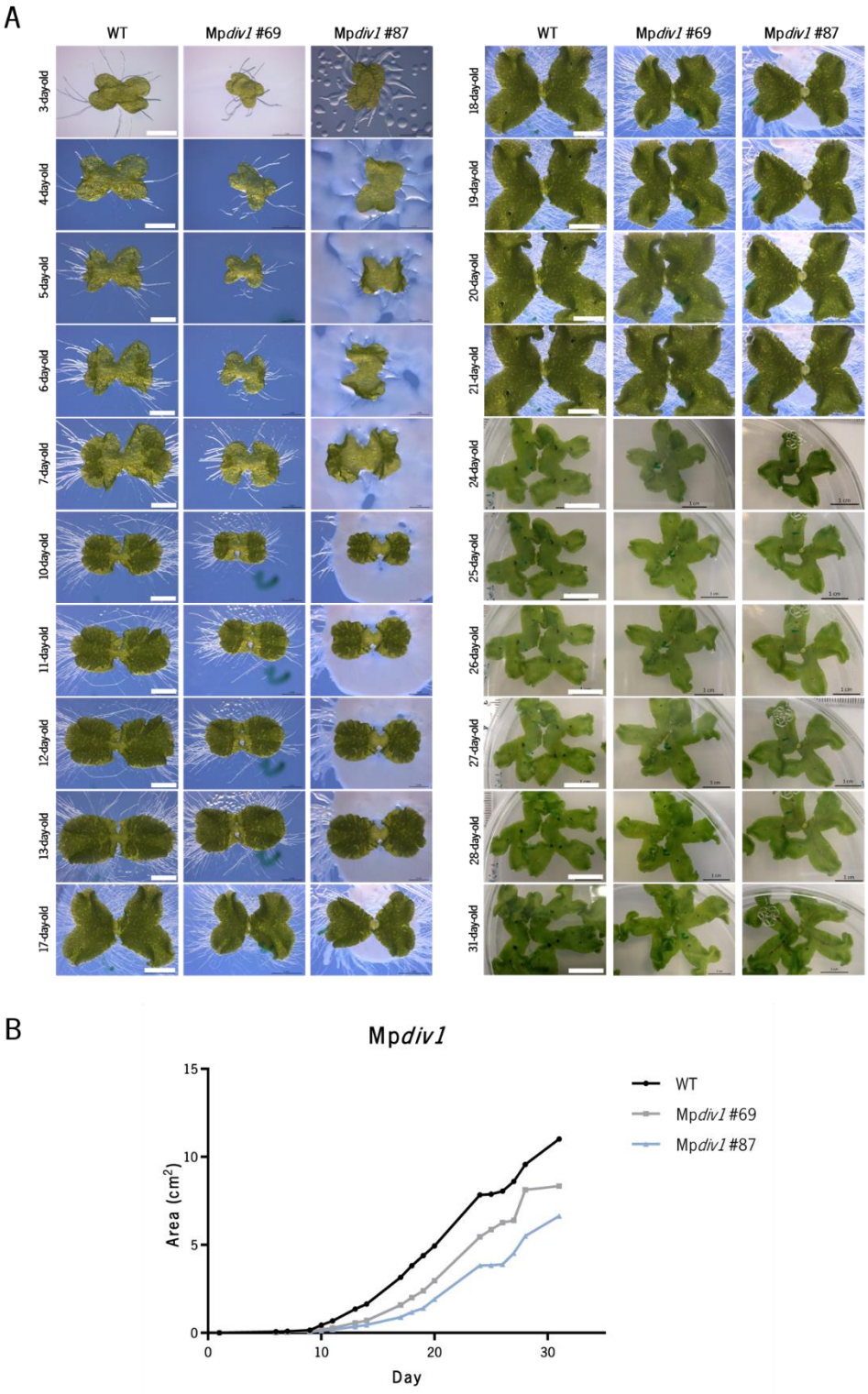

Figure 58: Analysis of Mp*div1* knockout in *Marchantia polymorpha* plants. A) Wild-type plants and Mp*div1* mutant plant lines #69 and #87 were observed during the 31 days after gemmae propagation. The plants were grown on half strength Gamborg B5 medium under long-day conditions (16 h light/ 8 h dark) at 20 °C and with light intensity between 40-45 µmol m^{-2} s^{-1}. Scale bar of: 1 mm from day 3 to 7; 2 mm from day 10 to 13; 5 mm from day 17 to 21 and 1 cm from day 24 to 31. **B)** The thallus superficial area of wild-type and Mp*div1* mutant plants was measured for 1 month. Black line is the control line that represents the wild-type plants thallus superficial area (n=8). The grey line represents the thallus superficial area of the Mp*div1* line #69 plants (n=8) and the light blue line represents the thallus superficial area of Mp*div1* line #87 plants (n=6).

In these mutants, differences in the gemma cups were also observed and to explore these differences, gemma cups of approximately 1-month-old mutant plants were compared with the gemma cups of the wild-type plants (figure 59A). The gemma cups of the mutant plant lines Mp*div1* #69 and #87 appeared to be not completely formed and contained less gemmae than wild type (figure 59A). The gemma cups diameter and the number of gemma cups per thallus surface area were also compared between the mutant and the wild-type plants. The diameter of 3 gemma cups per plant was measured using the software *ImageJ* and it was possible to observe that the gemma cups' diameter of the Mp*div1* #69 and #87 plants is smaller than wild type (figure 59B). On the other hand, Mp*div1* #69 and #87 mutant plants have more gemma cups per thallus superficial area than wild type (figure 59C).

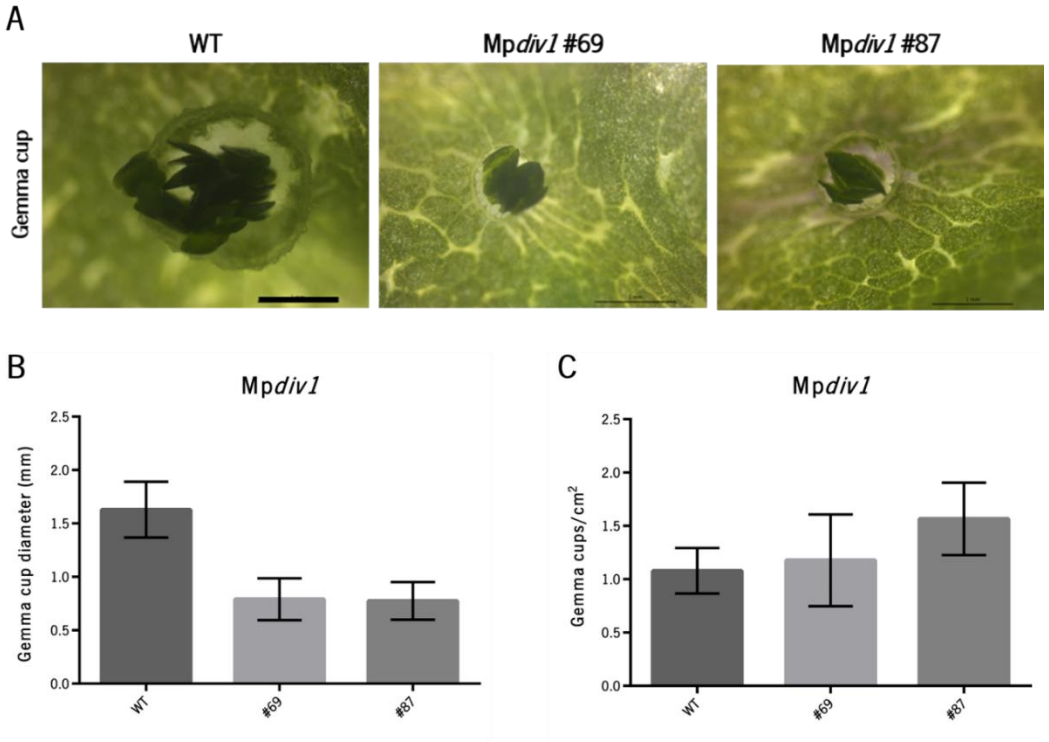

Figure 59: Analysis of Mp*div1* knockout in the gemma cups of *Marchantia polymorpha* plants. A) The gemma cups of wild-type plants and Mp*div1* mutant plant lines #69 and #87 were observed in plants with 1 month. Scale bar: 1 mm. **B)** Gemma cups' diameter of wild type (control) and Mp*div1* lines #69 and #87 plants (n=4). The diameter of three gemma cups per plant was measured using the software *ImageJ*. **C)** Number of gemma cups per thallus superficial area of wild-type (control) and Mp*div1* lines #69 (n=8) and #87 (n=6) plants. The number of gemma cups of wild-type and mutant plants was counted and divided by the thallus superficial area. Error bars indicate standard deviation.

Although no apparent differences were observed between the wild-type and mutant plants' rhizoids, gemmae of 1-month-old mutant and wild-type plants were collected and observed at the fluorescence microscope using chlorophyll autofluorescence (figure 60A). Mp*div1* #69 and #87 gemmae have a small number of rhizoid initial cells per gemmae superficial area when compared to wild type (figure 60B).

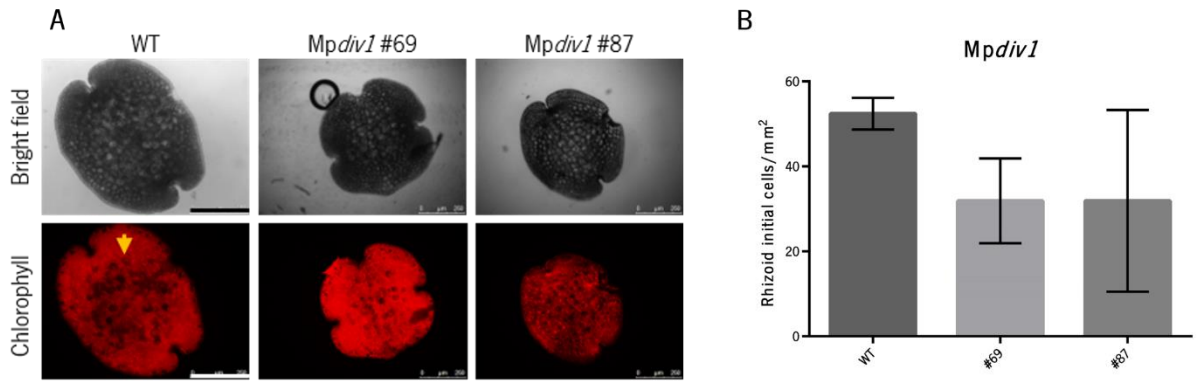

Figure 60: Analysis of Mp*div1* knockout in the rhizoids of *Marchantia polymorpha* plants. A) Gemmae of wild-type plants and Mp*div1* mutant plant lines #69 and #87 were observed under a fluorescence microscope using chlorophyll autofluorescence. The dark areas represent the rhizoid initial cells (yellow arrow). Black and white scale bars: 250 µm. **B)** The number of dark areas was counted in wild type (control) (n=2) and Mp*div1* lines #69 (n=5) and #87 (n=3) and divided for the gemmae area. Error bars indicate standard deviation.

3.3.8.2. Analysis of Mp*divl* mutant plants phenotype

Mp*divl* mutant plants #26, #73 and #87 showed notable differences in the thallus size from day 10 onwards (figure 61A). The thallus shape of line #87 plants was also different from wild-type plants (figure 61A). The thallus surface area of the 3 independent mutant lines measured for 1-month showed that the area of these plants is smaller than wild-type plants from day 10 onwards (figure 61B).

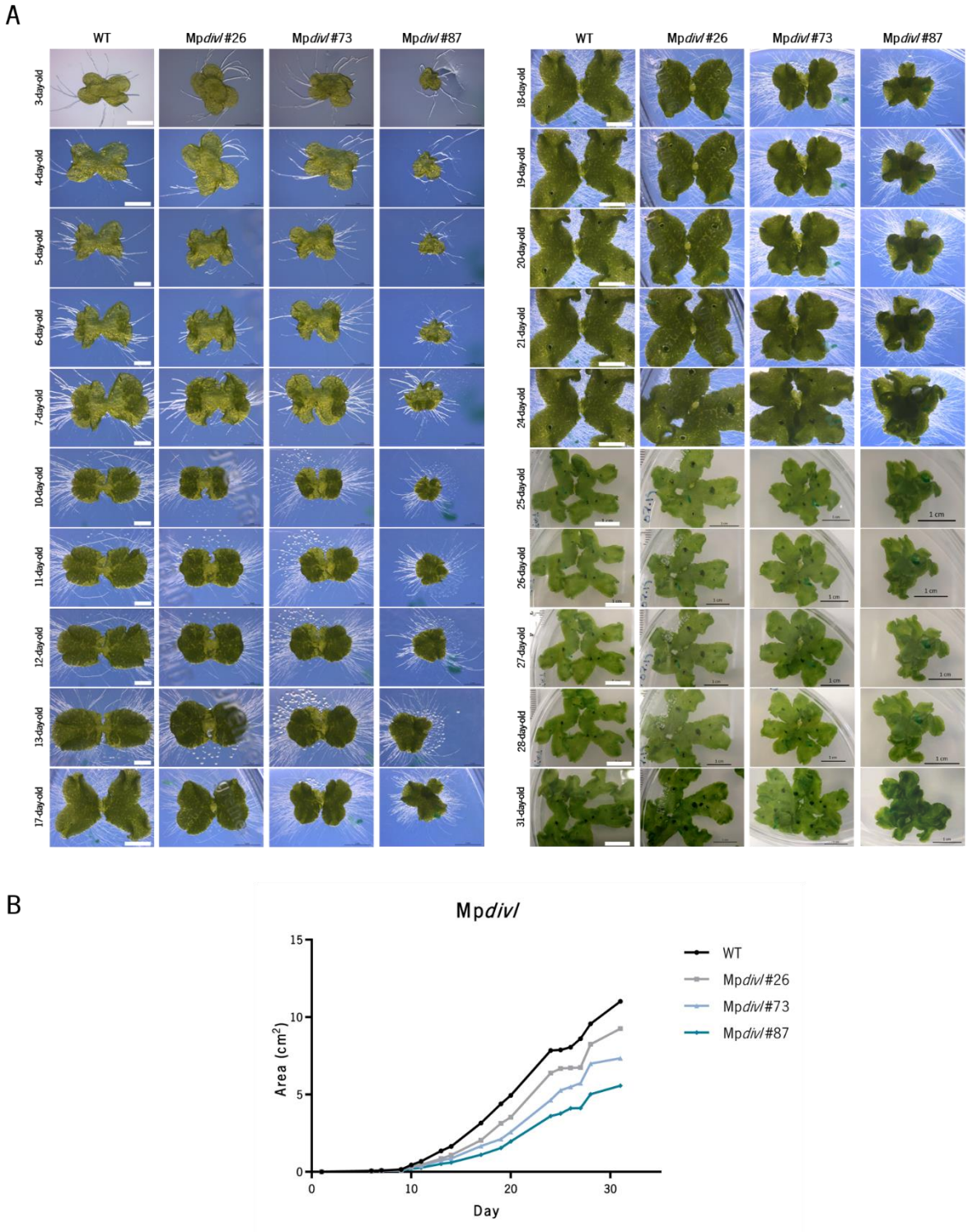

Figure 61: Analysis of Mp*divl* knockout in *Marchantia polymorpha* plants. A) Wild-type plants and Mp*divl* mutant plant lines #26, #73 and #87 were observed during the 31 days after gemmae propagation. The plants were grown on half strength Gamborg B5 medium under long-day conditions (16 h light/ 8 h dark) at 20 °C and with light intensity between 40-45 µmol m^{-2} s^{-1}. Scale bar of: 1 mm from day 3 to 7; 2 mm from day 10 to 13; 5 mm from day 17 to 24 and 1 cm from day 25 to 31. **B)** The thallus superficial area of wild-type and Mp*divl* mutant plants was measured for 1 month. Black line is the control line that represents the wild type thallus superficial area. The grey line represents the thallus superficial area of the Mp*divl* line #26 plants, the light blue line represents the thallus superficial area of Mp*divl* line #73 plants, and the dark blue line represents the thallus superficial area of Mp*divl* line #87 plants (n=8).

The comparison between the gemma cups of Mp*divl* lines #26 and #73 and wild-type plants did not show clear differences from the wild-type structures (figure 62A). The most evident differences were observed in the gemma cups of the mutant line #87. The gemma cups of the line #87 are smaller and contain less gemmae than the wild-type gemma cups which appear to have a halted development (figure 62A). The gemma cups diameter of the Mp*divl* line #26 is slightly bigger than wild type, while the gemma cups diameter of the lines #73 and #87 is smaller than wild type (figure 62B). Mp*divl* lines #26 and #73 have more gemma cups per thallus surface area than wild type (figure 62C). The number of gemma cups per thallus surface area of the Mp*divl* line #87 plants is similar to wild type.

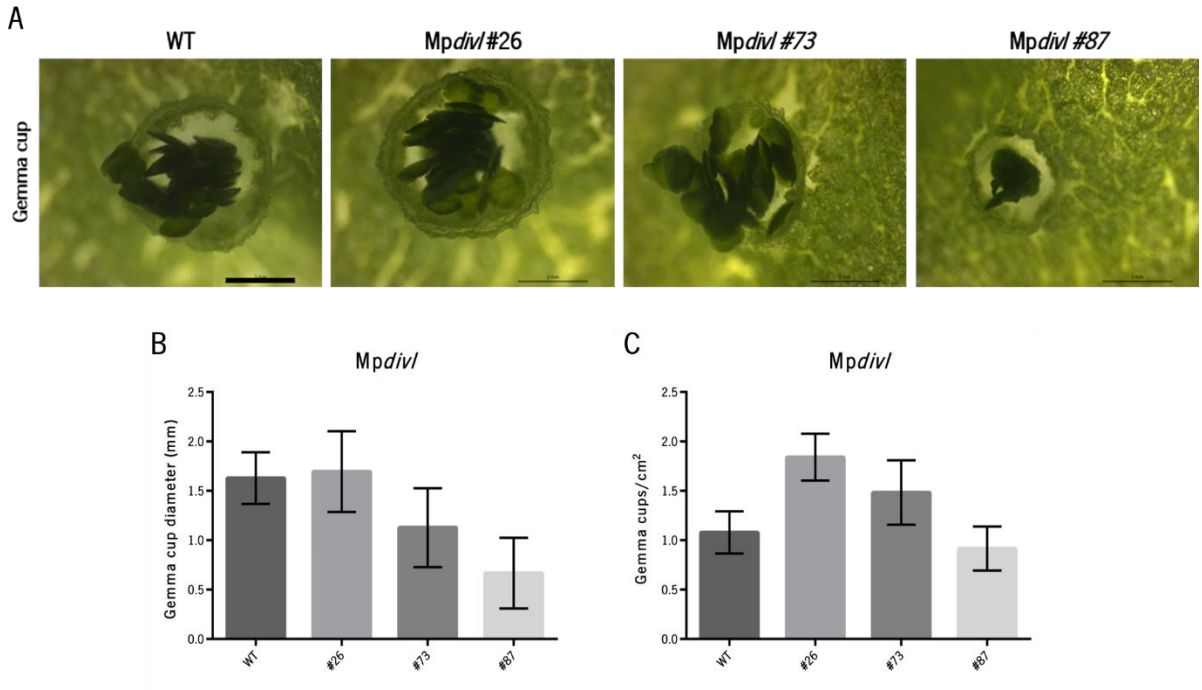

Figure 62: Analysis of Mp*divl* knockout in the gemma cups of *Marchantia polymorpha* plants. A) The gemma cups of wild-type plants and Mp*divl* mutant plant lines #26, #73 and #87 were observed in plants with 1 month. Scale bar: 1 mm. **B)** Gemma cups diameter of wild-type (control) and Mp*divl* lines #26, #73 and #87 plants (n=4). The diameter of three gemma cups per plant was measured using the software *ImageJ*. **C)** Number of gemma cups per thallus superficial area of wild-type (control) and Mp*divl* lines #26, #72 and #87 plants (n=8). The number of gemma cups of wild-type and mutant plants was counted and divided by the thallus superficial area. Error bars indicate standard deviation.

In the gemmae of the Mp*divl*#26, #73 and #87 plants observed into the fluorescence microscope it was possible to observe a smaller number of dark areas (figure 63A). When the number of rhizoid initial cells per gemmae superficial area was counted and compared between the wild-type and mutant gemmae a small number of rhizoid initial cells per gemmae superficial area was obtained for the gemmae of the 3 Mp*divl* mutant lines (figure 63B).

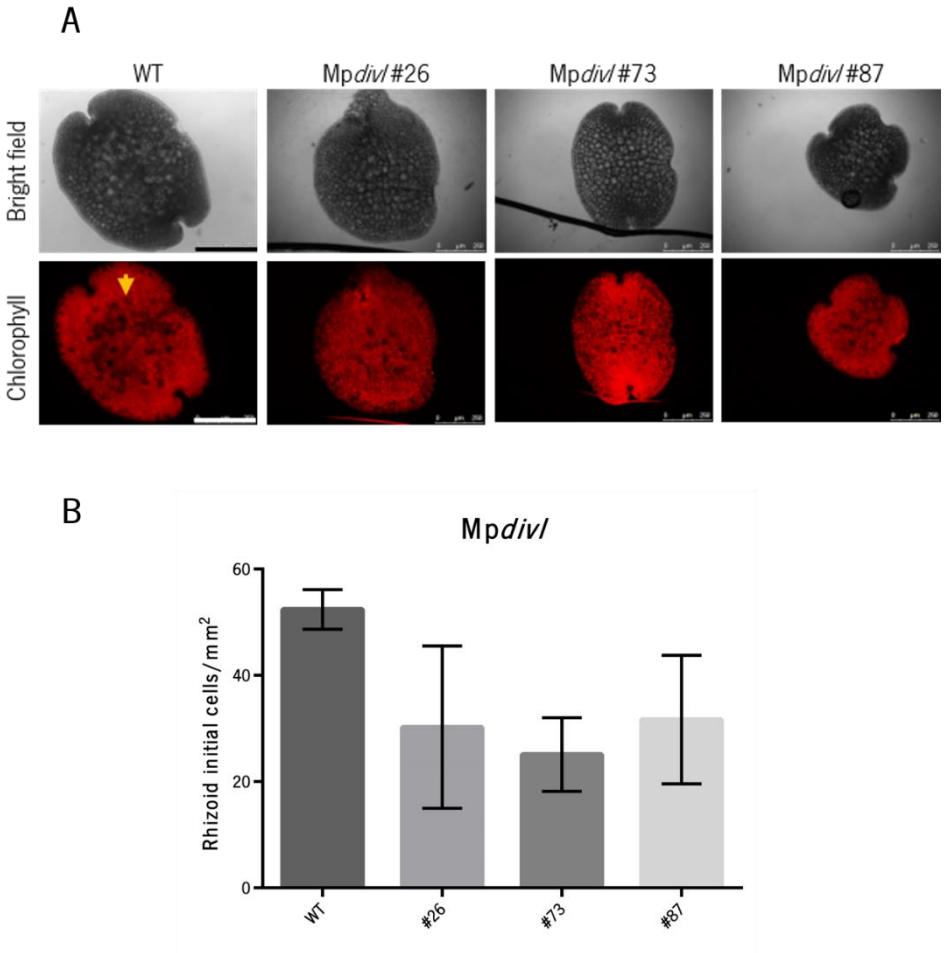

Figure 63: Analysis of Mp*divl* knockout in the rhizoids of *Marchantia polymorpha* plants. A) Gemmae of wild-type plants and Mp*divl* mutant plant lines #26, #73 and #87 were observed under a fluorescence microscope using chlorophyll autofluorescence. The dark areas represent the rhizoid initial cells (yellow arrow). Black and white scale bars: 250 µm. **B)** The number of dark areas was counted in wild-type (control) (n=2) and Mp*divl* lines #26 (n=4), #73 (n=4) and #87 (n=6) and divided for the gemmae area. Error bars indicate standard deviation.

3.3.8.3. Analysis of Mp*div2* mutant plants phenotype

Mp*div2* mutant plant lines are smaller than wild type and have a different thallus shape. Moreover, Mp*div2* independent lines of Group I to III have a smaller thallus with a lighter green colour than wild-type plant with a round shape (figure 64; figure I1 and I2- Annex I). The plant #93 of group IV has a thallus with a similar size or slightly bigger than the wild-type thallus. The thallus shape of the wild-type and #93 starts to be clearly different from day 17 after the gemmae propagation (figure 64). These observations were confirmed by the thallus surface area measurements performed for 1-month old plants. The thallus surface areas of the plants from Group I (lines #5, #24, #68, #116, #135, #217, #242 and #257), Group II (lines #169 and #296) and Group III plants (lines #99, #127, #189 and #305) appeared to be smaller than wild type after 26 days of propagation, and the area of Group IV line #93 similar to the area of the wild-type plants (figure 65; figure I3- Annex I).

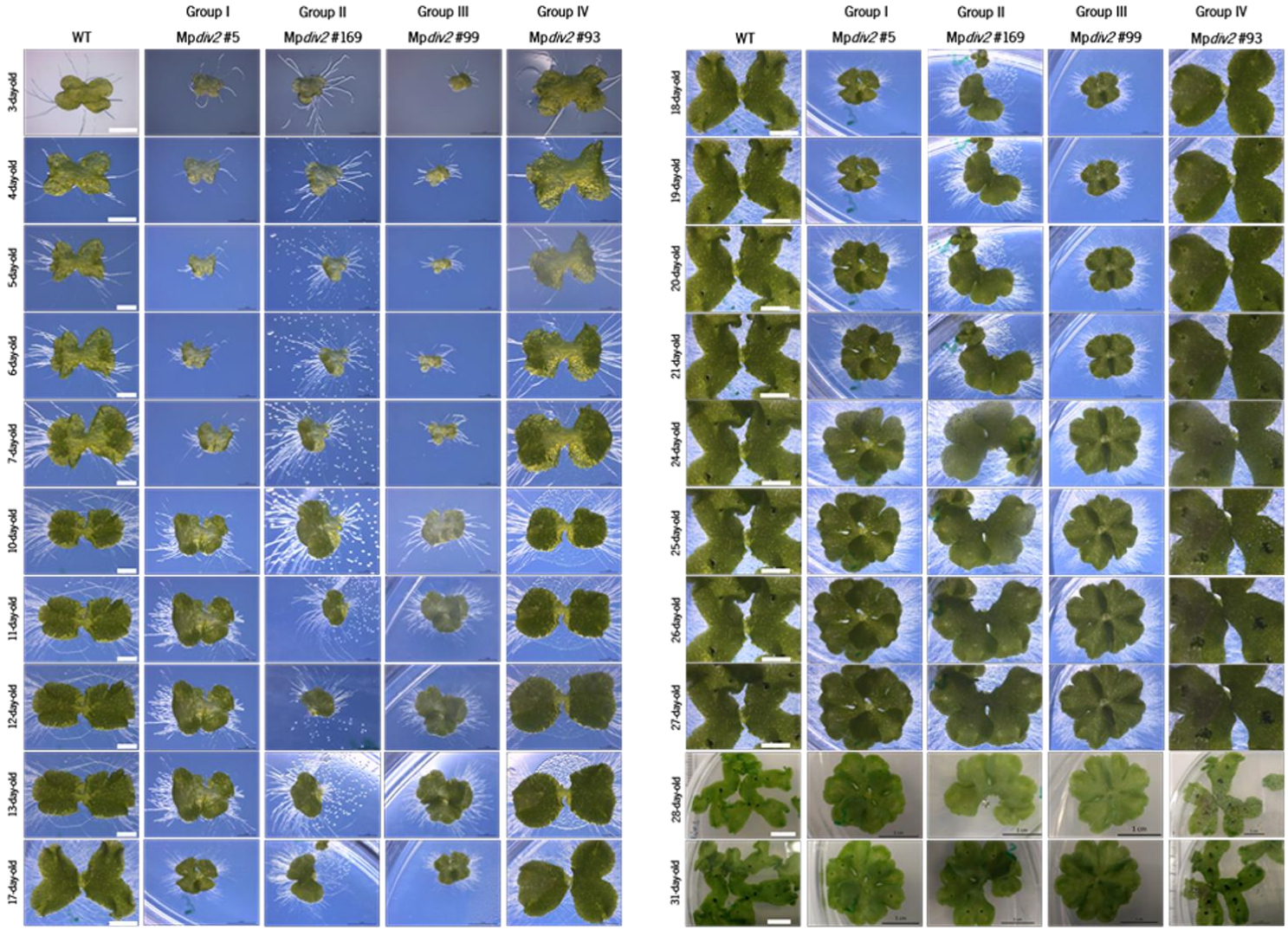

Figure 64: Analysis of Mp*div2* knockout in *Marchantia polymorpha* plants. Wild- type plants and Mp*div2* mutant plant lines #5 (Group I), #169 (Group II), #99 (Group III) and #93 (Group IV) were observed during 31 days after gemmae propagation. The plants were grown on half strength Gamborg B5 medium under long-day conditions (16 h light/ 8 h dark) at 20 °C and with light intensity between 40-45 µmol m^{-2} s^{-1}. Scale bar of: 1 mm from day 3 to 10; 2 mm from day 11 to 13; 5 mm from day 17 to 27 and 1 cm from day 28 to 31.

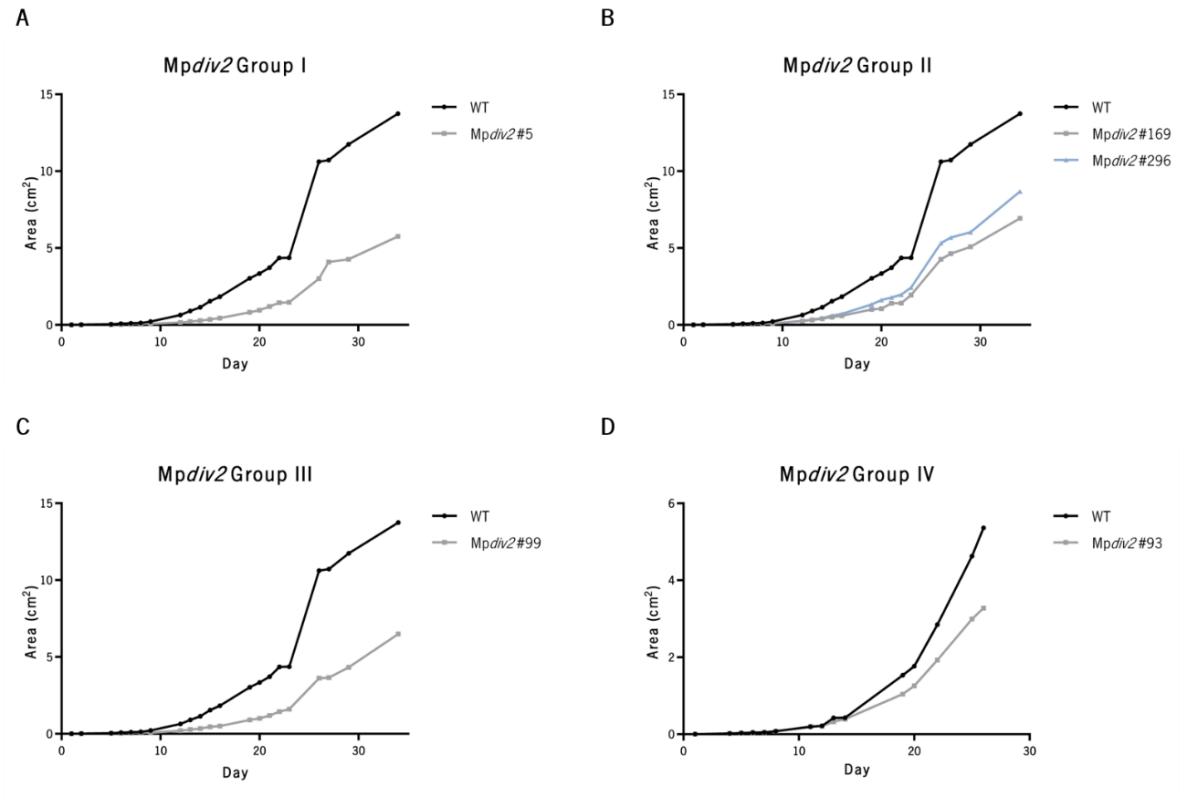

Figure 65: Analysis of Mp*div2* knockout in the thallus superficial area of *Marchantia polymorpha* plants. A) Thallus superficial area of wild-type (black line) and Mp*div2* Group I line #5 plants (grey line) (n=8). **B)** Thallus superficial area of wild-type (black line), Mp*div2* Group II line #169 plants (grey line) and Mp*div2* Group II line #296 plants (light blue line) (n=8). **C)** Thallus superficial area of wild-type (black line) and Mp*div2* Group III line #99 plants (grey line) (n=8). **D)** Thallus superficial area of wild-type (black line) and Mp*div2* Group IV line #93 plants (grey line) (n=4). This experiment was performed in different conditions from the other lines of Groups I, II and III. The analysis was performed using the thallus superficial areas of wild-type and Mp*div2* mutant plants measured for 1 month.

Differences in the number and shape of gemma cups could be observed for all *Mpdiv2* plants. Gemma cups of the Mp*div2* lines of Group I to III are flat and contain less gemmae than wild type (figure 66; figure I4- Annex I). The gemma cups of the line #93 are similar to the wild type.

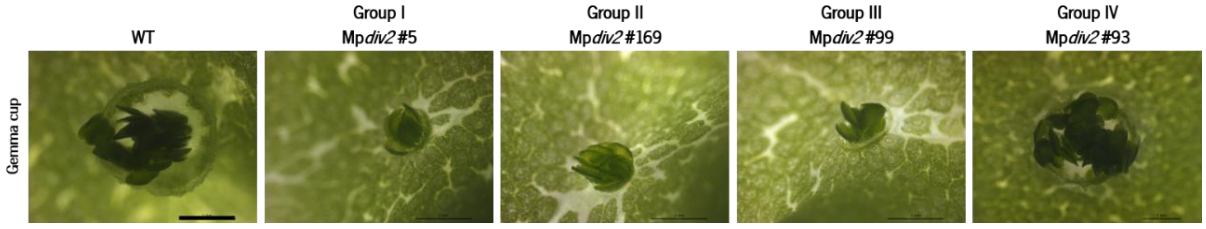

Figure 66: Analysis of Mp*div2* knockout in the gemma cups of *Marchantia polymorpha* plants. The gemma cups of wild-type plants and Mp*div2* mutant plant lines #5 (Group I), #169 (Group II), #99 (Group III) and #93 (Group IV) were observed in plants with 1 month. Scale bar: 1 mm.

The gemma cups diameter of all Mp*div2* mutant plant lines excepting the line #93 of Group IV is smaller than wild type (figure 67). Furthermore, the number of gemma cups per thallus surface area is different and varied between Mp*div2* mutant lines (figure 68). The plants of Group I lines, excepting the line #116, have more gemma cups per cm^2 of thallus than the wild-type plants. The lines #169 and #296

of the Group II also have more gemma cups per cm^2 of thallus than the wild type. An increased number of gemma cups per cm^2 of thallus is also verified for the lines of Group III. The line #93 of Group IV has less gemma cups per area than wild type (figure 68).

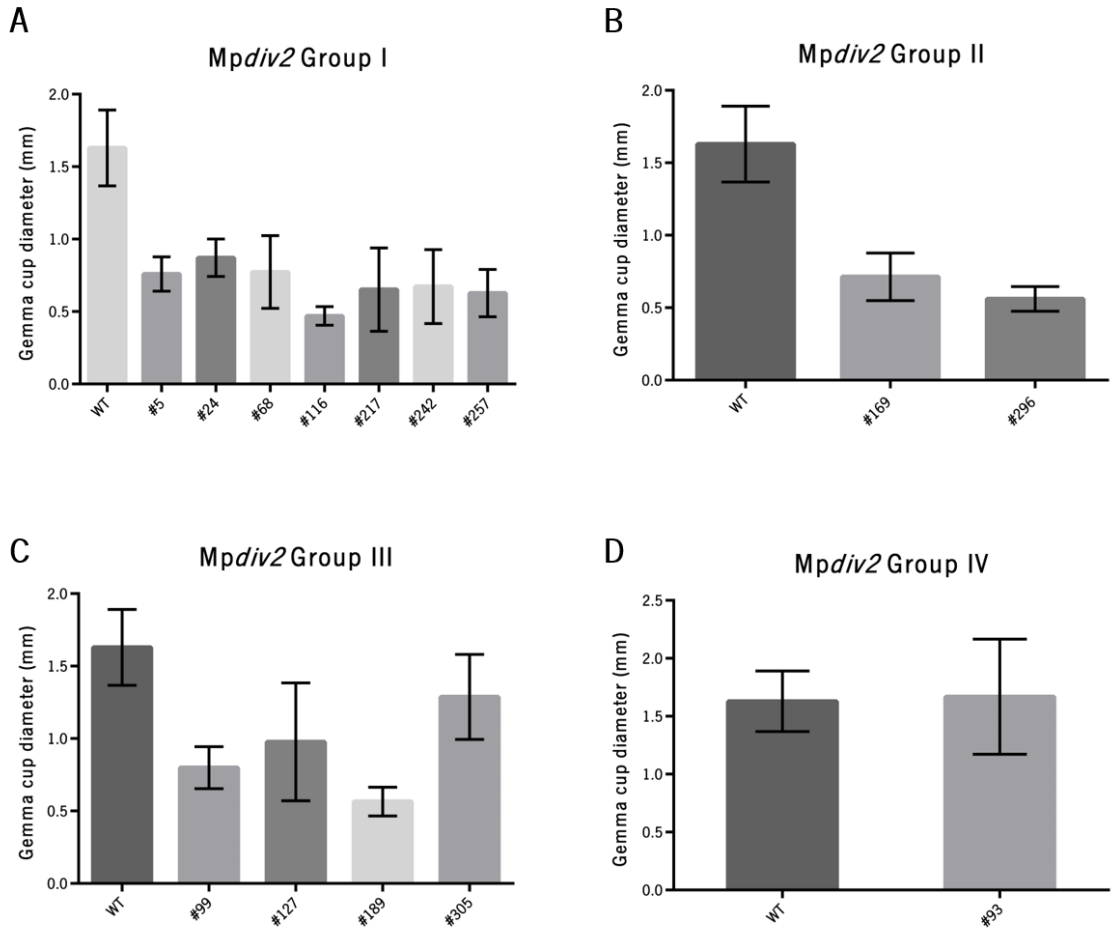

Figure 67: Analysis of Mp*div2* gemma cups diameter. A) Gemma cups diameter of wild-type (control) (n=4) and Mp*div2* Group I lines #5, (n=2), #24 (n=3), #68 (n=4), #116 (n=1), #217 (n=2), #242 (n=2) and #257 (n=2). **B)** Gemma cups diameter of wild-type (control) (n=4) and Mp*div2* Group II lines #169 (n=4) and #296 (n=2). **C)** Gemma cups diameter of wild-type (control) (n=4) and Mp*div2* Group III lines #99 (n=2), #127 (n=2), #189 (n=3) and #305 (n=4). **D)** Gemma cups diameter of wild-type (control) (n=4) and Mp*div2* Group IV lines #93 (n=3). The diameter of three gemma cups per plant was measured using the software *ImageJ*. Error bars indicate standard deviation.

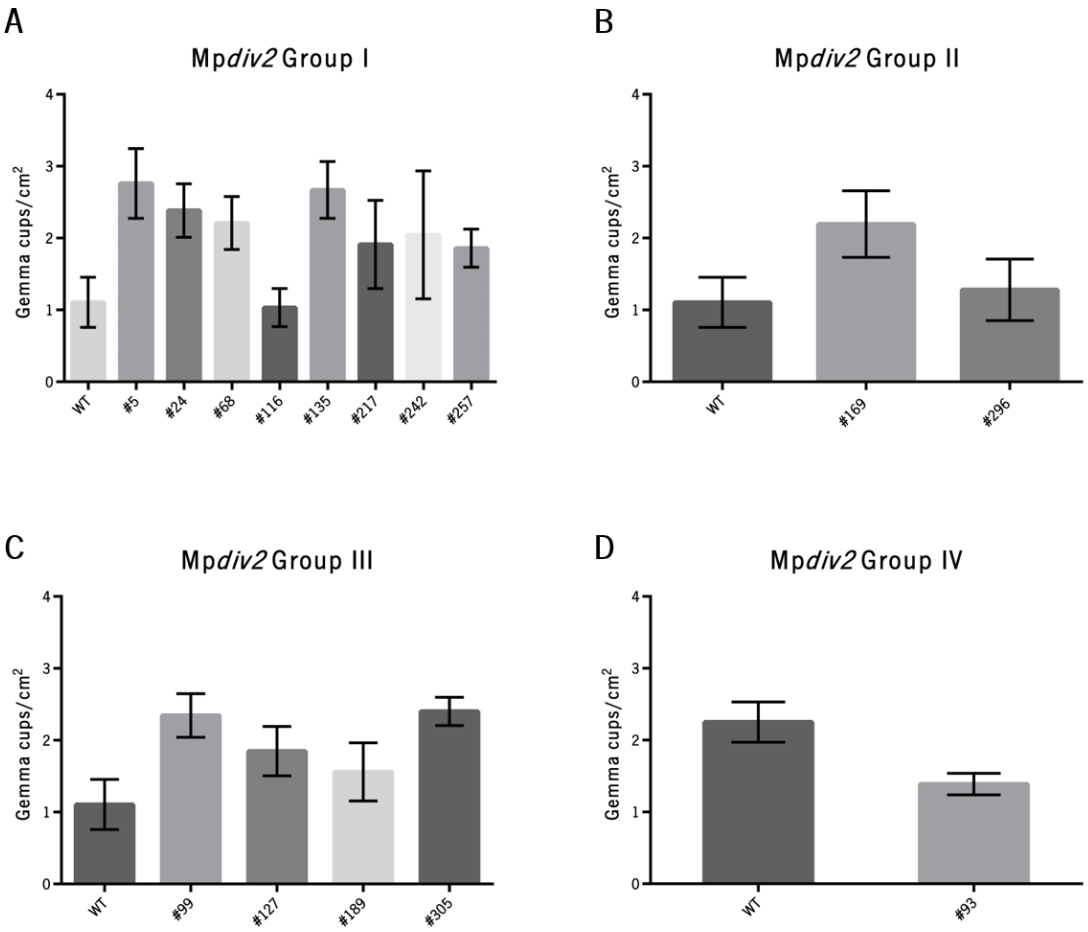

Figure 68: Analysis of Mp*div2* number of gemma cups per thallus superficial area. A) Number of gemma cups per thallus superficial area of wild-type (control) (n=8) and Mp*div2* Group I lines #5, (n=8), #24 (n=8), #68 (n=8), #116 (n=3), #135 (n=4), #217 (n=6), #242 (n=8) and #257 (n=7). **B)** Number of gemma cups per thallus superficial area of wild-type (control) and Mp*div2* Group II lines #169 and #296 (n=8). **C)** Number of gemma cups per thallus superficial area of wild-type (control) and Mp*div2* Group III lines #99, #127, #189 and #305 (n=8). **D)** Number of gemma cups per thallus superficial area of wild-type (control) (n=4) and Mp*div2* Group IV lines #93 (n=4). Gemma cups of wild-type and mutant plants were counted and divided by the thallus superficial area. Error bars indicate standard deviation.

In the gemmae images of the Mp*div2* Group I to III lines a smaller number of dark areas is visible when compared with wild type (figure 69A; figure I5- Annex I). The gemmae of the Mp*div2* Group IV line #93 are the ones that seem to have a similar number of dark areas when compared to wild type. When the number of rhizoid initial cells was divided by the gemmae area, these observations were not confirmed since the number of rhizoid initial cells per gemmae area were similar to wild type excepting for Mp*div2* #127 (figure 69B).

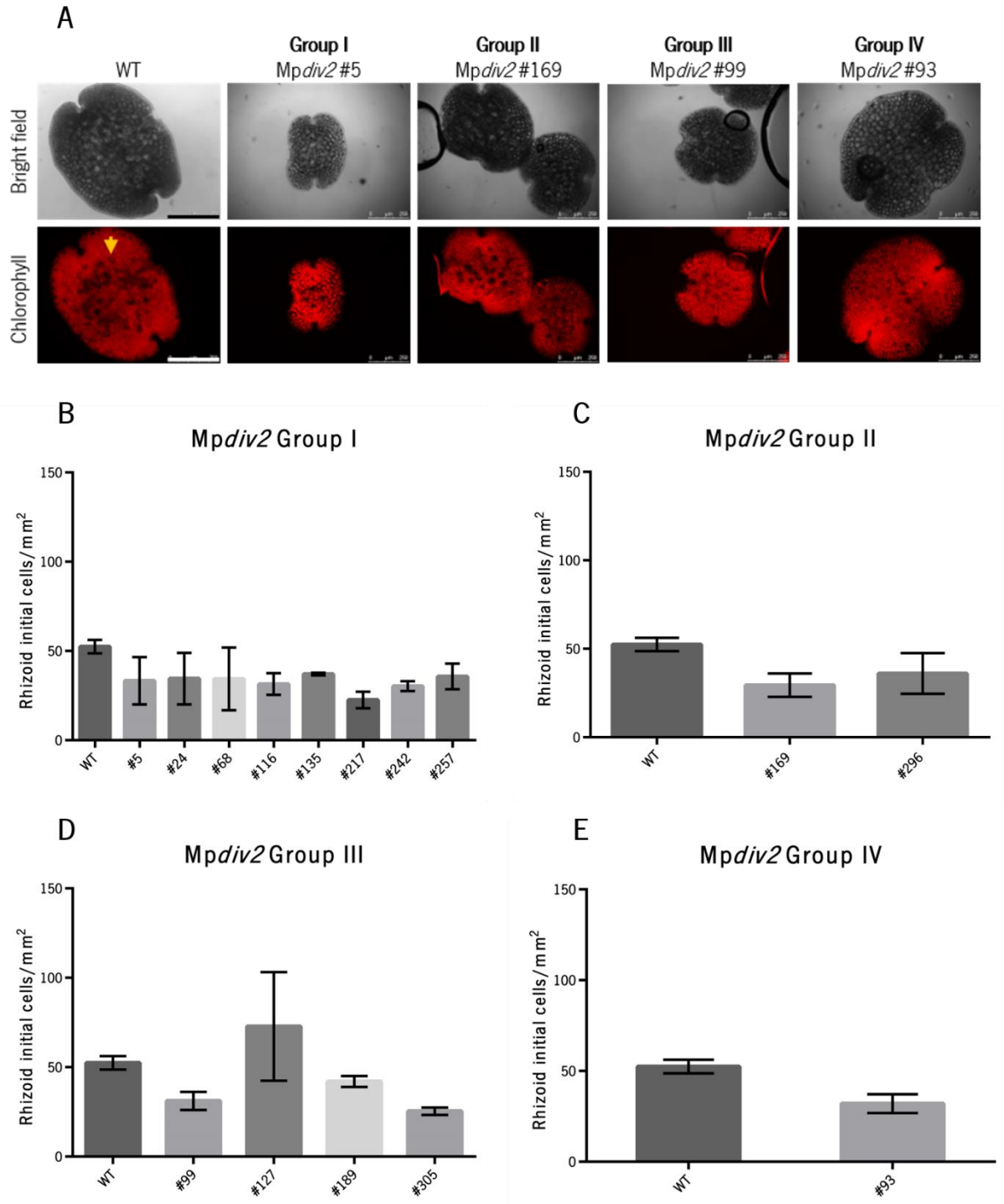

Figure 69: Analysis of Mp*div2* knockout in the rhizoids of *Marchantia polymorpha* plants. A) Gemmae of wild-type plants and Mp*div2* mutant plant lines #5 (Group I), #169 (Group II), #99 (Group III) and #93 (Group IV) were observed under a fluorescence microscope using chlorophyll autofluorescence. The dark areas represent the rhizoid initial cells (yellow arrow). Black and white scale bars: 250 µm. **B)** The number of dark areas in the circle form were counted in wild-type (control) (n=2) and Mp*div2* Group I lines #5 (n=5), #24 (n=3), #68 (n=3), #116 (n=5), #135 (n=4), #217 (n=4), #242 (n=5) and #257 (n=5), and divided by the gemmae area. **C)** The number of dark areas in the circle form were counted in wild-type (control) (n=2) and Mp*div2* Group II lines 169 (n=4) and Mp*div2* #296 (n=6) and divided by the gemmae area. **D)** The number of dark areas in the circle form were counted in wild-type (control) (n=2) and Mp*div2* Group III lines #99 (n=4), #127 (n=7), #189 (n=4) and #305 (n=4) and divided by the gemmae area. **E)** The number of dark areas in the circle form were counted in wild-type (control) (n=2) and Mp*div2* Group IV line #93 (n=4) and divided by the gemmae area. Error bars indicate standard deviation.

4. Discussion

During evolution, the establishment of new interactions within a conserved gene regulatory network can lead to the emergence of new biological functions and morphologies. New interactions are often related to repeated events of gene duplication that increase the genome complexity. Extensive sub-functionalisation of gene function hinders the understanding of a given gene function due to redundancy within multigene families. The study of ancestral gene functions in early plants enables a better understanding on the establishment of gene regulatory networks due to the existence of fewer duplicated gene copies. The DDR transcriptional module is composed by three MYB proteins- DIV, RAD and DRIF, and it controls several development processes in angiosperms, such as flower asymmetry in *Antirrhinum majus*. DIV and DRIF emerged in the lineage that gave rise to green algae while RAD and its interaction with DRIF only occurs in gymnosperms and angiosperms (Raimundo et al., 2018). Homologs for DIV and DRIF proteins were found in *M. polymorpha*, a basal plant model, but their function is still unknown (Raimundo et al., 2018). This work focused on the unravelling the function of Mp*DIV* and Mp*DRIF*.

4.1. Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF* are expressed in different tissues

The function of *DIV1*, *DIV2*, *DIVL* and *DRIF* genes in *M. polymorpha* is still unknown but, a recent transcriptomic work performed by Flores-Sandoval et al. (2018) provided some clues about the possible function of Mp*DIV1*, Mp*DIV2* and Mp*DIVL* genes. Mp*DRIF* is not considered a transcription factor in the available databases and, therefore, it was not included in the transcriptomic analysis. In this study, Mp*DIV1* and Mp*DIV2* are upregulated in spore germination, where several physiological and developmental processes are being activated. Mp*DIVL* appeared upregulated in sporelings with 72 h, a stage in which the tissue polarity is being properly established. In the current thesis, a RT-qPCR and whole mount *in situ* performed to analyse the gene expression in different tissues suggested that Mp*DIV1*, Mp*DIV2* and Mp*DIVL* may be controlling thallus development since the expression of these genes was higher in this structure. The expression results also showed that Mp*DRIF* is more expressed in the thallus and female structures, which suggests that Mp*DRIF* may be involved in thallus development but also in the development of the female sexual structures. To validate this information, a new RT-qPCR analysis should be performed with tissues collected daily during more than one plant developmental cycle to observe with more accuracy how gene expression varies at different stages of development. A whole

mount *in situ* should also be conducted using different stages of thallus and sexual structures development. In parallel with the whole mount *in situ*, the study of the spatial gene expression pattern with the GUS::promoter fusions, developed in this work, will be continued to analyse the promoter activity of these genes in different tissues. These expression studies will allow a better understanding about the ancestral function of these genes and which developmental processes they might be controlling.

4.2. Mp*DIV1* might be involved in cellular expansion or cellular proliferation

DIV homologs were found to be involved in the establishment of the flower asymmetry in *A. majus* and in the cellular expansion of the tomato fruit (Machemer et al., 2011; Raimundo et al., 2013). In *M. polymorpha*, the overexpression of one of the Mp*DIV1* transcripts (Mp*DIV1.1*) originates plants with more cell proliferation or expansion in the apical notches and in the apex of the thallus. On the contrary, the knockout of Mp*DIV1* leads to plants with smaller thallus and with less gemma cups. These results together with the expression of Mp*DIV1* in initial stages of plant development (Flores-Sandoval et al., 2018) suggests that Mp*DIV1* may also be involved in organ expansion processes, similarly to homologs of more recent plants. Other studies performed using *M. polymorpha* as model species also showed defects in the cellular growth and expansion. Monte et al. (2019) described the effect of Mp*JAZ*, a gene that represses the jasmonate biosynthesis, in the promotion of cell growth. Mp*jaz* plant mutant lines exhibited development effects in the thallus development that lead to plants with a smaller plant area. Furuya et al. (2018) showed that *ANGUSTIFOLIA* (*AN*), a gene that controls different developmental traits in *Arabidopsis*, is involved in the morphogenesis regulation of the *M. polymorpha* plant. In Mp*an* knockout mutants was possible to observe an abnormal growth of the thallus due to an irregular cell expansion. Therefore, further experiments should involve the use of histologic analysis to observe the number and the size of the cells in knockout and overexpression plants to understand if they are in fact expanding or having more proliferating activity. Moreover, to validate the phenotype observed in the Mp*DIV1.1* overexpression plants, the second Mp*DIV1* transcript- Mp*DIV1.2*- should also be overexpressed. The observation of all the phenotypes exhibited by the Mp*div1* knockout and overexpression lines were performed in the thallus stage since we were unable to obtain sexual structures in the conditions used to grow *M. polymorpha*. In the future, to understand if the exhibited phenotypes are sex related with Mp*DIV1* function, female Mp*DIV1* knockout lines and female Mp*DIV1.1* overexpression lines should be obtained since the lines observed in this work were of the male sex. Complementation assays should also be

performed in the knockout lines to ensure that the phenotypes observed are not caused by off-targets mutations derived from the use CRISPR-Cas9 methodology.

When the number of rhizoid initial cells was counted in gemmae of knockout and overexpression lines and divided for the gemmae area to analyse possible differences between the wild-type and transgenic plant rhizoids, no significative differences were observed. In the future, gemmae should be observed under a confocal microscope to better distinguish the dark areas that represent the rhizoid initial cells and to understand if Mp*DIV1* is involved in rhizoid's development.

DIV proteins are composed by two functionally conserved MYB domains: the MYBI domain, in the N-terminal, that is responsible for the protein-protein interactions and the MYBII domain, in the C-terminal, that binds to DNA. In the obtained mutants, the MYBII domain is deleted from the mature MpDIV protein. Thus, in these plants, MpDIV1 protein may still be able to bind MpDRIF but it is unable to bind DNA. This can explain the phenotype observed in the Mp*div1* mutant lines. The interaction between DIV and DRIF in *M. polymorpha* was confirmed by yeast-two-hybrid by Raimundo et al. (2018). To confirm if the same interaction exists *in planta*, the two proteins can be subcellularly-localized using translational fusions with reporter genes. This will unravel how and where the MpDIV and MpDRIF proteins interact. The bimolecular fluorescence complementation (BiFC) technology, a fluorescence based technique, can also be used to validate the interaction between the MpDIV and MpDRIF proteins. The proteins are fused with the C and N-fragments of a fluorescent reporter protein and expressed in the cells. The proteins interaction will bring closer the fluorescent fragments which allow to reform the native fluorescent reporter protein structure that will emit a fluorescent signal. For this to happen, MpDIV and MpDRIF proteins have to be expressed in the same organ and cell compartment of the plant. The RT-qPCR showed that both Mp*DIV1* and Mp*DRIF* genes are expressed in the thallus, therefore, the interaction between them can be tested in this organ. In *Antirrhinum*, the interaction between these proteins leads to the formation of a complex that activates the expression of genes that confer the ventral identity of the flower. Thus, if the DIV-DRIF interaction is happening in *M. polymorpha* is expected that knockout lines for both Mp*DIV1* and Mp*DRIF* genes originate plants with similar phenotypes. Also, a transcriptomic assay could be performed to identify DIV target genes in *M. polymorpha*. With the abolition of the MYBII domain in the two Mp*div1* mutant lines it will be possible to identify the genes that are downregulated in comparison with wild type. This will allow to identify which genes are regulated by DIV.

4.3. Mp*DIV2* might be controlling plant size and shape

Mp*DIV2*, the second *DIV* homolog of *M. polymorpha*, seems to have a similar expression pattern when compared to Mp*DIV1*. The RT-qPCR performed in this work showed that both Mp*DIV* copies are more expressed in female and male thallus than in the female and male sexual structures. The transcriptomic work performed by Flores-Sandoval et al. (2018) showed that both Mp*DIV* copies are upregulated in spore germination. The fifteen Mp*div2* knockout lines obtained with the CRISPR/Cas9 technology in this thesis were divided in groups according to the type of deletion or insertion. The fourteen lines of the Groups I, II and III showed a similar phenotype with differences from the wild type in size, shape and colour. The wild-type thallus develops by consecutive shape transitions that are initiated in early stages of the plant development by divisions of an apical cell that leads to a dorsoventral thallus structure (figure 8) (Shimamura, 2016). When propagated, the gemma tissue grows by lateral expansion. Then, the four lobes of the gemma start to expand and a process of bifurcation occurs to form two apical notches on each one of the thallus extremities. The two apical notches are separated by a central tissue zone (central lobe) that expands pushing the two notches apart. When completely formed, the central lobe starts to branch (figure 8) (Solly et al., 2017). The Mp*div2* knockout lines from the Groups I, II and III do not show the same wild-type shape transitions during gemmae development since they are smaller and with a rounder shape. This suggests that Mp*DIV2* might be involved in controlling thallus shape and size. Since the differences are visible a few days after gemmae propagation, and since the structure of the thallus in Mp*div2* plants is different from the wild type, this gene may be controlling the initial cell divisions involved in the gemmae development. Arai et al. (2019) observed some differences in the gemma cups of plants that overexpress the MpBHLH12 transcription factor. The gemma cups exhibited an arrested development characterized by a smaller cup area. This phenotype was similar to the one observed in the Mp*div2* knockout lines of Groups I, II and III in which the gemma cups appeared less developed that wild type and with a smaller dimeter of the cup. Altogether this information supports the idea that Mp*DIV2* might be controlling different development processes of the thallus.

Another phenotypic difference observed in the Mp*div2* plant lines from Groups I, II and III is the lighter green colour of the thallus. Therefore, in the future, the levels of chlorophyll should be measured. Also, the number of rhizoid initial cells should also be counted in gemmae observed under a confocal microscope to better distinguish the dark areas that indicate the presence of these cells.

No Mp*DIV2* expression was detected in the semi-quantitative RT-PCR for Mp*div2* plant lines of Group I, II or III, using RNA extracted from whole plants with 1 month. This indicates that these plants are probably knockouts and the phenotype is due to a mutation in the gene. However, phenotype complementation assays should be performed to discard potential phenotypes derived from off-targets. The plant transformation with the constructs for the Mp*DIV2* overexpression should be accomplished to observe the phenotype exhibited by plant lines and to obtain more clues about the function of this gene.

Mp*div2* mutants possess different alterations in the MpDIV2 protein. Proteins from Mp*div2* Group I and II do not have the MYBII domain, whereas proteins derived from Mp*div2* mutants of Group III do not have the two MYB domains. Once the phenotype exhibited by these plants is very similar, it is possible to conclude that the MYBII domain of the MpDIV2 protein is a key domain to the protein function. Without the MYBII domain, these proteins have the same domain of RAD. This protein it is not present in the basal plant *M. polymorpha*, thus, in the future it will be interesting to understand if the mutant Mpdiv2 proteins are potentially playing the same role as RAD in more recent plant species. In *Antirrhinum*, it is known that RAD disrupts the formation of DIV-DRIF complex (Raimundo et al., 2013) and in *Arabidopsis*, RAD acts as a repressor of the floral transition and in the it regulates the early morphogenesis (Hamaguchi et al., 2008; Li et al., 2015). In *M. polymorpha*, the Mpdiv2 mutant proteins equivalents to RAD can be disrupting the DIV-DRIF complex, therefore, the sub localization of these proteins can be performed to observe if Mpdiv2 mutant proteins are interacting with MpDRIF preventing the DIV-DRIF interaction.

The Mp*div2* mutant plant line #93 (Group IV), with two functional MYB domains, presents a phenotype with differences from wild type and Mp*div2* knockout lines of Group I, II and III in the thallus shape and in the gemma cups structure. The semi-quantitative RT-PCR showed a similar Mp*DIV2* transcript accumulation in the wild-type and Mp*div2* #93 plants. In future, the semi-quantitative RT-PCR should be repeated with RNA extracted from tissues of different stages of the plant development for a better comparison of the transcript accumulation in the wild-type and mutant plant.

Despite the differences observed in the phenotypes of Mp*div1* and Mp*div2* knockout mutant lines, it was possible to conclude that both Mp*DIV* copies are controlling thallus development processes. This, together with the similar expression patterns observed in the expression studies, suggests that these two gene copies have similar functions to confirm this, a crossing between the Mp*div1* and Mp*div2* mutant plants should be performed to analyse the combined function of the DIV genes.

4.4. Mp*DIVL* might be regulating the plant shape

The role *DIVL* genes is still unknown. DIVL protein has a MYBII SHAQKYF-type domain identical to the MYBII domain present in the DIV family, but instead of the MYBI domain they contain a small R motif. This R motif was identified as plant-specific active repression domain that is present in some *Arabidopsis* transcription factors (Ikeda & Ohme-Takagi, 2009). Active transcriptional repressors are related to several aspects of plant growth and development. In this work, a RT-qPCR analysis showed that this gene is more expressed in the thallus of *M. polymorpha*. In the whole mount *in situ* no expression was observed. Three Mp*divl* mutant lines were obtained using the CRISPR/Cas9 technology. Mp*divl* line #87 was the one that showed a more prominent phenotype with a smaller thallus with a different shape. The gemma cups of this plant also exhibited differences from wild type since they appeared less developed and with a smaller diameter. In the transcriptomics work performed by Flores-Sandoval et al. (2018), Mp*DIVL* appeared upregulated in spores with 72 h, an important stage for the plant shape development. This, together with the phenotype observed for the mutant plant line #87 suggests that this gene might be controlling thallus shape. As for the Mp*divl* lines #26 and #73, no differences were observed in the thallus shape and size. This might be related to the fact that the proteins originated by the mutant DNA sequences of the lines #26 and #73 maintained the two conserved domains, and consequently the DIVL protein function, despite the insertions that occurred in the genomic sequence. In future, new thallus superficial area measurements should be performed to confirm if the thallus size differences are significant when compared with the wild type. Mp*DIVL* overexpression lines will also be helpful to understand the phenotypes observed in the knockout mutant lines and to obtain clues about the function of this gene.

As mentioned above, DIVL has a repression domain that is present in several other *Arabidopsis* transcription factors. In *Antirrhinum*, RAD, a siPEP, antagonizes the function of DIV preventing the formation of a dimer between DIV and DRIF. Since no RAD protein was found in *M. polymorpha*, it will be interesting to understand if MpDIVL is somehow repressing the function of MpDIV protein as described in *Antirrhinum*. An electrophoretic mobility shift assay (EMSA) could be performed to test if MpDIV and MpDIVL are competing for the same DNA target sequence.

4.5. Mp*DRIF*: the unknow function

In *A. majus*, DRIF interacts with DIV, through the MYBI domain, in the nuclei, leading to the formation of DIV-DRIF complexes promoting the regulation of genes that confer the ventral identity to the flower petal (Raimundo et al., 2013). In *M. polymorpha*, one DRIF homolog was found, and to understand its ancestral function different molecular approaches were used. The CRISPR/Cas9 technology was used to knockout the Mp*DRIF* gene but no mutant lines were obtained with the designed gRNAs. This can be related to the method used to genotype the plants that only allows the identification of large deletions or insertions. A new gRNA was designed and in future *M. polymorpha* spores will be transformed with a new combination of gRNAs. Overexpression lines of one of the Mp*DRIF* transcripts (Mp*DRIF.1*) were obtained and despite the confirmation of the gene overexpression, the phenotype analysis was not conclusive due to severe plant contaminations that affected the plant growth. Further experiments should involve a detailed analysis of the Mp*DRIF.1* overexpression phenotype and the development of Mp*DRIF.*2 (second Mp*DRIF* transcript) overexpression lines. The performed RT-qPCR showed that Mp*DRIF* is more expressed in the thallus and female sexual structures. In the future, it will be interesting to understand if this gene is involved with the development of female or male sexual structures by the phenotype analysis in female or male knockout and overexpression lines. Experiments performed in the laboratory suggests that the *Arabidopsis DRIF* family have an effect in the photoperiod pathway that regulate flowering time (Cunha, 2018; Raimundo, 2015). The photoperiod is related with the circadian clock, a key element for plants to adjust it activities with surrounding environment. To understand if Mp*DRIF* is circadian regulated as it happens in *Arabidopsis*, the expression of this gene should be analysed during the day and night periods.

4.6. Concluding remarks

M. polymorpha is an excellent plant model for the study of ancestral gene function due to the low number of gene copies present in its genome. The presence of *DIV* and *DRIF* genes and DIV-DRIF interaction in this plant was already described, however the ancestral function of these genes is still unknown. The phenotypes observed in the knockout and overexpression lines showed that members of *DIV* family might be controlling development process in the thallus and gemma cups of *M. polymorpha*. The function of Mp*DIVL* and Mp*DRIF* it is still unclear, but overexpression lines and knockout mutant plants of these genes are being developed. This work will help to unravel the ancestral function of the DIV and DRIF proteins in early plant species which will provide new insights for the study of these genes in other plant species.

5. Bibliography

Almeida, J., Rocheta, M., & Galego, L. (1997). Genetic control of flower shape in *Antirrhinum majus*. Development, **124**, 1387–1392.

Amoutzias, G. D., Robertson, D. L., Oliver, S. G., & Bornberg-Bauer, E. (2004). Convergent evolution of gene networks by single-gene duplications in higher eukaryotes. EMBO Reports, **5**, 274–279.

Amoutzias, G. D., Robertson, D. L., Van de Peer, Y., & Oliver, S. G. (2008). Choose your partners: dimerization in eukaryotic transcription factors. Trends in Biochemical Sciences, **33**, 220–229.

Arai, H., Yanagiura, K., Toyama, Y., & Morohashi, K. (2019). Genome-wide analysis of MpBHLH12, a IIIf basic helix-loop-helix transcription factor of *Marchantia polymorpha*. Journal of Plant Research, **132**, 197–209.

Azevedo, H., Lino-Neto, T., & Tavares, R. M. (2003). RNA isolation from adult maritime pine needles. Plant Molecular Biology Reporter, **21**, 333–338.

Bagowski, C. P., Bruins, W., & Te Velthuis, A. J. W. (2010). The nature of protein domain evolution: shaping the interaction network. Current Genomics, **11**, 368–376.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N. J., & Nekrasov, V. (2015). Editing plant genomes with CRISPR/Cas9. Current Opinion in Biotechnology, **32**, 76–84.

Benková, E., Michniewicz, M., Sauer, M., Teichmann, T., Seifertová, D., Jürgens, G., & Friml, J. (2003). Local, efflux-dependent auxin gradients as a common module for plant organ formation. Cell, **115**, 591–602.

Bertani, G. (2013). Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. Journal of Bacteriology, **186**, 595-600.

Bowman, J. L. (2016). A brief history of *Marchantia* from Greece to genomics. Plant and Cell Physiology, **57**, 210–229.

Bowman, J. L., Araki, T., & Kohchi, T. (2016). *Marchantia*: Past, present and future. Plant and Cell Physiology, **57**, 205–209.

Bowman, J. L., Kohchi, T., Yamato, K. T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., Adam, C., Aki, S. S., Althoff, F., Araki, T., Arteaga-Vazquez, M. A., Balasubrmanian, S., Barry, K., Bauer, D., Boehm, C. R., Briginshaw, L., Caballero-Perez, J., Catarino, B., Cheng, F., Chiyoda, S., Chovatia, M., Davies, K. M., Delmans, M., Demura, T., Dierschke, T., Dolan, L., Dorantes-Acosta, A. E., Eklund, D. M., Florent, S. N., Flores-Sandoval, E., Fujiyama, A., Fukuzawa, H., Galik, B., Grimanelli, D., Grimwood, J., Grossniklaus, U., Hamada, T., Haseloff, J., Hetherington, A. J., Higo, A., Hirakawa, Y., Hundley, H. N., Ikeda, Y., Inoue, K., Inoue, S., Ishida, S., Jia, Q., Kakita, M.,

Kanazawa, T., Kawai, Y., Kawashima, T., Kennedy, M., Kinose, K., Kinoshita, T., Kohara, Y., Koide, E., Komatsu, K., Kopischke, S., Kubo, M., Kyozuka, J., Lagercrantz, U., Lin, S. S., Lindquist, E., Lipzen, A. M., Lu, C. W., De Luna, E., Martienssen, R. A., Minamino, N., Mizutani, M., Mizutani, M., Mochizuki, N., Monte, I., Mosher, R., Nagasaki, H., Nakagami, H., Naramoto, S., Nishitani, K., Ohtani, M., Okamoto, T., Okumura, M., Phillips, J., Pollak, B., Reinders, A., Rövekamp, M., Sano, R., Sawa, S., Schmid, M. W., Shirakawa, M., Solano, R., Spunde, A., Suetsugu, N., Sugano, S., Sugiyama, A., Sun, R., Suzuki, Y., Takenaka, M., Takezawa, D., Tomogane, H., Tsuzuki, M., Ueda, T., Umeda, M., Ward, J. M., Watanabe, Y., Yazaki, K., Yokoyama, R., Yoshitake, Y., Yotsui, I., Zachgo, S., & Schmutz, J. (2017). Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. Cell, **171**, 287–304.

Brodie, H. J. (1951). The splash-cup dispersal mechanism in plants. Canadian Journal of Botany, **29**, 224–234.

Bullock, W. O., Fernandez, J. M., & Short, J. M. (1987). XL1-Blue: a high-efficiency plasmid transforming recA *Escherichia coli* strain with β-galactosidase selection. Biotechniques, **5**, 376–39.

Chen, J., Li, Y., Zhang, K., & Wanga, H. (2018). Whole-genome sequence of phage-resistant strain *Escherichia coli* DH5α. American Society for Microbiology, **6**, 5–6.

Christian, M., Qi, Y., Zhang, Y., & Voytas, D. F. (2013). Targeted mutagenesis of *Arabidopsis thaliana* using engineered TAL effector nucleases. G3: Genes, Genomes, Genetics, **3**, 1697–1705.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Hsu, P. D., Wu, X., Jiang, W., & Marraffini, L. A. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science, **339**, 819–823.

Cooke, T. J., Poli, D. B., Sztein, A. E., & Cohen, J. D. (2002). Evolutionary patterns in auxin action. Plant Molecular Biology, **49**, 319–338.

Corley, S. B., Carpenter, R., Copsey, L., & Coen, E. (2005). Floral asymmetry involves an interplay between TCP and MYB transcription factors in *Antirrhinum*. Proceedings of the National Academy of Sciences of the United States of America, **102**, 5068–5073.

Costa, M. M. R., Fox, S., Hanna, A. I., Baxter, C., & Coen, E. (2005). Evolution of regulatory interactions controlling floral asymmetry. Development, **132**, 5093–5101.

Crandall-Stotler, B., Storler, R. E., & Long, D. G. (2009). Phylogeny and classification of the Marchantiophyta. Edinburgh Journal of Botany, **66**, 155–198.

Cunha, A. M. (2018). Functional characterisation of *AtDRIFs* in *Arabidopsis thaliana*. Master thesis.

Deblaere, R., Bytebier, B., De Greve, H., Deboeck, F., Schell, J., Van Montagu, M., & Leemans, J. (1985). Efficient octopine Ti plasmid-derive vectors for *Agrobacterium*-mediated gene transfer to plants. Nucleic Acids Research, **13**, 4777–4788.

Durfee, T., Nelson, R., Baldwin, S., Plunkett, G., Burland, V., Mau, B., Petrosino J. F., Qin X., Muzny D. M., Ayele M., Gibbs R. A., Csörgo B., Pósfai G., Weinstock G. M., & Blattner, F. R. (2008). The complete genome sequence of *Escherichia coli* DH10β: insights into the biology of a laboratory workhorse. Journal of Bacteriology, **190**, 2597–2606.

Erwin, D. H., & Davidson, E. (2009). The evolution of hierarchical gene regulatory networks. Nature Reviews Genetics, **10**, 141–148.

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D. L., Wang, Z., Zhang, Z., Zheng, R., Yang, L., Zeng, L., Liu, X., & Zhu, J. K. (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.

Finet, C., & Jaillais, Y. (2012). AUXOLOGY: When auxin meets plant evo-devo. Developmental Biology, **369**, 19–31.

Flores-Sandoval, E., Romani, F., & Bowman, J. L. (2018). Co-expression and transcriptome analysis of *Marchantia polymorpha* transcription factors supports class C ARFs as independent actors of an ancient auxin regulatory module. Frontiers in Plant Science, **9**, 1–21.

Friml, J., Vieten, A., Sauer, M., Weijers, D., Schwarz, H., Hamann, T., Offringa, R., & Jürgens, G. (2003). Efflux-dependent auxin gradients establish the apical-basal axis of *Arabidopsis*. Nature, **426**, 147–153.

Fujisawa, M., Hayashi, K., Nishio, T., Bando, T., Okada, S., Yamato, K. T., Fukuzawa, H., & Ohyania, K. (2001). Isolation of X and Y chromosome-specific DNA markers from a liverwort, *Marchantia polymorpha*, by representational difference analysis. Genetics, **159**, 981–985.

Furuya, T., Hattori, K., Kimori, Y., Ishida, S., Nishihama, R., Kohchi, T., & Tsukaya, H. (2018). ANGUSTIFOLIA contributes to the regulation of three-dimensional morphogenesis in the liverwort *Marchantia polymorpha*. Development, **145**, 1–10.

Galego, L., & Almeida, J. (2002). Role of *DIVARICATA* in the control of dorsoventral asymmetry in *Antirrhinum* flowers. Genes and Development, **16**, 880–891.

Graham, L. (1985). The origin of the life cycle of land plants. American Scientist, **73**, 178–186.

Graham, L. E., Cook, M. E., & Busse, J. S. (2000). The origin of plants: body plan changes contributing to a major evolutionary radiation. Proceedings of the National Academy of Sciences of the United States of America, **97**, 4535–4540.

Hamaguchi, A., Yamashino, T., Koizumi, N., Kiba, T., Kojima, M., Sakakibara, H., & Mizuno, T. (2008). A small subfamily of *Arabidopsis RADIALIS-LIKE SANT/MYB* genes: a link to HOOKLESS1-mediated signal transduction during early morphogenesis. Bioscience, Biotechnology, and Biochemistry, **72**, 2687–2696.

Harper, R. M., Stowe-Evans, E. L., Luesse, D. R., Muto, H., Tatematsu, K., Watahiki, M. K., Yamamoto, K., & Liscum, E. (2000). The *NPH4* locus encodes the auxin response factor ARF7, a conditional regulator of differential growth in aerial *Arabidopsis* tissue. The Plant Cell, *12*, 757–770.

Hartley, J. L., Temple, G. F., & Brasch, M. A. (2000). DNA cloning using in vitro site-specific recombination. Genome Research, **10**, 1788–1795.

Hejátko, J., Blilou, I., Brewer, P. B., Friml, J., Scheres, B., & Benková, E. (2006). *In situ* hybridization technique for mRNA detection in whole mount *Arabidopsis* samples. Nature Protocols, **1**, 1939–1946.

Hixson, J. E., Jett, C., & Birnbaum, S. (1996). Identification of promoter sequences in the 5' untranslated region of the baboon apolipoprotein[a] gene. Journal of Lipid Research, **37**, 2324–2331.

Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., Sato, S., Yamada, T., Mori, H., Tajima, N., Moriyama, T., Ikeuchi, M., Watanabe, M., Wada, H., Kobayashi, K., Saito, M., Masuda, T., Sasaki-Sekimoto, Y., Mashiguchi, K., Awai, K., Shimojima, M., Masuda, S., Iwai, M., Nobusawa, T., Narise, T., Kondo, S., Saito, H., Sato, R., Murakawa, M., Ihara, Y., Oshima-Yamada, Y., Ohtaka, K., Satoh, M., Sonobe, K., Ishii, M., Ohtani, R., Kanamori-Sato, M., Honoki, R., Miyazaki, D., Mochizuki, H., Umetsu, J., Higashi, K., Shibata, D., Kamiya, Y., Sato, N., Nakamura, Y., Tabata, S., Ida, S., Kurokawa, K., & Ohta, H. (2014). *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. Nature Communications, **5**, 1–9.

Ikeda, M., & Ohme-Takagi, M. (2009). A novel group of transcriptional repressors in *Arabidopsis*. Plant and Cell Physiology, **50**, 970–975.

Ishizaki, K. (2016). Evolution of land plants: insights from molecular studies on basal lineages. Bioscience, Biotechnology and Biochemistry, **81**, 73–80.

Ishizaki, K., Chiyoda, S., Yamato, K. T., & Kohchi, T. (2008). *Agrobacterium*-mediated transformation of the haploid liverwort *Marchantia polymorpha* L., an emerging model for plant biology. Plant and Cell Physiology, **49**, 1084–1091.

Ishizaki, K., Nishihama, R., Ueda, M., Inoue, K., Ishida, S., Nishimura, Y., Shikanai, Toshiharu, K., & T. (2015). Development of gateway binary vector series with four different selection markers for the liverwort *Marchantia polymorpha*. PLoS ONE, **10**, 1–13.

Ishizaki, K., Nishihama, R., Yamato, K. T., & Kohchi, T. (2016). Molecular genetic tools and techniques for *Marchantia polymorpha* research. Plant and Cell Physiology, **57**, 262–270.

Jefferson, R. A., Kavanagh, T. A., & Bevan, M. W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. The EMBO Journal, **6**, 3901–3907.

Kato, H., Ishizaki, K., Kouno, M., Shirakawa, M., Bowman, J. L., Nishihama, R., & Kohchi, T. (2015). Auxin-mediated transcriptional system with a minimal set of components is critical for morphogenesis through the life cycle in *Marchantia polymorpha*. PLoS Genetics, **11**, 1–27.

Kenrick, P. R. (1994). Alternation of generations in land plants: new phylogenetic and palaeobotanical evidence. Biological Reviews, **69**, 293–330.

Khodthong, C., Snow, J., & Juckem, L. (2016). Optimization of DNA, RNA and RNP delivery for efficient mammalian cell engineering. Madison, Wisconsin USA.

Klein, M., Pieri, I., Uhlmann, F., Pfizenmaier, K., & Eisel, U. (1998). Cloning and characterization of promoter and 5'-UTR of the NMDA receptor subunit ε2: evidence for alternative splicing of 5'-non-coding exon. Gene, **208**, 259 – 269.

Kubota, A., Ishizaki, K., Hosaka, M., & Kohchi, T. (2013). Efficient *Agrobacterium*-mediated transformation of the liverwort *Marchantia polymorpha* using regenerating thalli. Bioscience, Biotechnology, and Biochemistry, **77**, 167–172.

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B. E., Bickel, P., Brown, J. B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K. I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A. J., Hoffman, M. M., Iyer, V. R., Jung, Y. L., Karmakar, S., Kellis, M., Kharchenko, P. V., Kharchenko, P. V., Li, Q., Liu, T., Liu, X. S., Ma, L., Milosavljevic, A., Myers, R. M., Park, P. J., Pazin, M. J., Perry, M. D., Raha, D., Reddy, T. E., Rozowsky, J., Shoresh, N., Sidow, A., Slattery, M., Stamatoyannopoulos, J. A., Tolstorukov, M. Y., White, K. P., Xi, S., Farnham, P. J., Lieb, J. D., Wold, B. J., & Snyder, M. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Research, **22**, 1813–1831.

Lewis, L. A., & McCourt, R. M. (2004). Green algae and the origin of land plants. American Journal of Botany, **91**, 1535–1556.

Li, C., Zhou, Y., & Fan, L. M. (2015). A novel repressor of floral transition, MEE3, an abiotic stress regulated protein, functions as an activator of *FLC* by binding to its promoter in *Arabidopsis*. Environmental and Experimental Botany, **113**, 1–10.

Ligrone, R., Duckett, J. G., & Renzaglia, K. S. (2012). Major transitions in the evolution of early land plants: a bryological perspective. Annals of Botany, **109**, 851–871.

Lu, C. A., Ho, T. D., Ho, S. L., & Yu, S. M. (2002). Three novel MYB proteins with one DNA binding repeat mediate sugar and hormone regulation of α-amylase gene expression. The Plant Cell, **14**, 1963 – 1980.

Lu, S. X., Knowles, S. M., Andronis, C., Ong, M. S., & Tobin, E. M. (2009). CIRCADIAN CLOCK ASSOCIATED1 and LATE ELONGATED HYPOCOTYL function synergistically in the circadian clock of *Arabidopsis*. Plant Physiology, **150**, 834–843.

Luo, D., Carpenter, R., Vincent, C., Copsey, L., & Coen, E. (1996). Origin of floral asymmetry in *Antirrhinum*. Nature, **383**, 794–799.

Machemer, K., Shaiman, O., Salts, Y., Shabtai, S., Sobolev, I., Belausov, E., Grotewold, E., & Barg, R. (2011). Interplay of MYB factors in differential cell expansion, and consequences for tomato fruit development. The Plant Journal, **68**, 337–350.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. Science, **339**, 823–826.

Mao, Y., Zhang, H., Xu, N., Zhang, B., Gou, F., & Zhu, J. K. (2013). Application of the CRISPR-Cas system for efficient genome engineering in plants. Molecular Plant, **6**, 2008–2011.

Marchant, J. (1713). Nouvelles découverte des Fleurs et des Graines d'une Plante rangée par les Botanistes sous le genre du Lichen. *In* Histoire de l'Académie royale des sciences. Band 35, pp. 229–234, Académie des sciences, Paris.

Mishler, B. D., & Churchill, S. P. (1984). A cladistic approach to the phylogeny of the "bryophytes". Brittonia, **36**, 406–424.

Mishler, B. D., & Churchill, S. P. (1985). Transition to a land flora: phylogenetic relationships of the green algae and bryophytes. Cladistics, **1**, 305–328.

Mitchell, P. J., & Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science, **245**, 371–378.

Monte, I., Franco-Zorrilla, J., García-Casado, G., Zamarren, A. M., García-Mina, J. M., Nishihama, R., Kohchi, T., & Solano, R. (2019). A single JAZ repressor controls the jasmonate pathway in *Marchantia polymorpha*. CellPress, **12**, 185–198.

Nakazato, T., Kadota, A., & Wada, M. (1999). Photoinduction of spore germination in *Marchantia polymorpha* L. is mediated by photosynthesis. Plant and Cell Physiology, **40**, 1014–1020.

Niklas, K. J., & Kutschera, U. (2010). The evolution of the land plant life cycle. New Phytologist, **185**, 27–41.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T., & Ohyama, K. (1992). Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. Journal of Molecular Biology, **223**, 1–7.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S. I., Inokuchi, H., & Ozeki, H. (1986). Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature, **322**, 572–574.

Pasek, S., Risler, J. L., & Brézellec, P. (2006). Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. Bioinformatics, **22**, 1418–1423.

Pires, N. D., Yi, K., Breuninger, H., Catarino, B., Menand, B., & Dolan, L. (2013). Recruitment and remodeling of an ancient gene regulatory network during land plant evolution. Proceedings of the National Academy of Sciences of the United States of America, **110**, 9571–9576.

Pires, N., & Dolan, L. (2010). Origin and diversification of basic-helix-loop-helix proteins in plants. Molecular Biology and Evolution, **27**, 862–874.

Podevin, N., Davies, H. V., Hartung, F., Nogué, F., & Casacuberta, J. M. (2013). Site-directed nucleases: a paradigm shift in predictable, knowledge-based plant breeding. Trends in Biotechnology, **31**, 375–383.

Prigge, M. J., Lavy, M., Ashton, N. W., & Estelle, M. (2010). *Physcomitrella patens* auxin-resistant mutants affect conserved elements of an auxin-signaling pathway. Current Biology, **20**, 1907–1912.

Proost, S., & Mutwil, M. (2016). Tools of the trade: studying molecular networks in plants. Current Opinion in Plant Biology, **30**, 130–140.

Proust, H., Honkanen, S., Jones, V. A. S., Morieri, G., Prescott, H., Kelly, S., Ishizaki, K., Kohchi, T., & Dolan, L. (2016). *RSL* Class I genes controlled the development of epidermal structures in the common ancestor of land plants. Current Biology, **26**, 93–99.

Qiu, Y.L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J., Estabrook, G. F., Hendry, T. A., Taylor, D. W., Testa, C. M., Ambros, M., Crandall-Stotler, B., Duff, R. J., Stech, M., Frey, W., Quandt, D., & Davis, C. C. (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.

Raimundo, J. (2015). Evolution of the gene regulatory network controlling flower dorsoventral asymmetry. Doctoral thesis. Available at http://hdl.handle.net/1822/40439.

Raimundo, J., Sobral, R., Bailey, P., Azevedo, H., Galego, L., Almeida, J., Coen, E., & Costa, M. M. R. (2013). A subcellular tug of war involving three MYB-like proteins underlies a molecular antagonism in *Antirrhinum* flower asymmetry. The Plant Journal, **75**, 527–538.

Raimundo, J., Sobral, R., Laranjeira, S., & Costa, M. M. R. (2018). Successive domain rearrangements underlie the evolution of a regulatory module controlled by a small interfering peptide. Molecular Biology and Evolution, **35**, 2873–2885.

Reynolds, K., Zimmer, A. M., & Zimmer, A. (1996). Regulation of RARβ2 mRNA expression: evidence for an inhibitory peptide encoded in the 5'-untranslated region. Journal of Cell Biology, **134**, 827 – 835.

Rose, A., Meier, I., & Wienand, U. (1999). The tomato I-box binding factor LeMYBI is a member of a novel class of Myb-like proteins. The Plant Journal, **20**, 641–652.

Saint-Marcoux, D., Proust, H., Dolan, L., & Langdale, J. A. (2015). Identification of reference genes for real-time quantitative PCR experiments in the liverwort *Marchantia polymorpha*. PLoS ONE, **10**, 1–14.

Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for genome editing, regulation and targeting. Nature Biotechnology, **32**, 347–355.

Schmidt, E. E., & Davies, C. J. (2007). The origins of polypeptide domains. BioEssays, **29**, 262–270.

Schuster, R. M. (1966). The Hepaticae and Anthocerotae of North America vol. 1. East of the hundredth meridian. New York: Columbia University Press.

Seo, P. J., Hong, S. Y., Kim, S. G., & Park, C. M. (2011). Competitive inhibition of transcription factors by small interfering peptides. Trends in Plant Science, **16**, 541–549.

Shimamura, M. (2016). *Marchantia polymorpha*: taxonomy, phylogeny and morphology of a model system. Plant and Cell Physiology, **57**, 230–256.

Shiu, S. H., Shih, M.C., & Li, W. H. (2005). Transcription factor families have much higher expansion rates in plants than in animals. Plant Physiology, **139**, 18–26.

Solly, J. E., Cunniffe, N. J., & Harrison, C. J. (2017). Regional growth rate differences specified by apical notch activities regulate liverwort thallus shape. Current Biology, **27**, 16–26.

Sugano, S. S., 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**, 1–22.

Sugano, S. S., 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 and Cell Physiology, **55**, 475–481.

Tatematsu, K., Kumagai, S., Muto, H., Sato, A., Watahiki, M. K., Harper, R. M., Liscum, E., & Yamamoto, K. T. (2004). *MASSUGU2* encodes Aux/IAA19, an auxin-regulated protein that functions together with the transcriptional activator NPH4/ARF7 to regulate differential growth responses of hypocotyl and formation of lateral roots in *Arabidopsis thaliana*. The Plant Cell, **16**, 379–393.

Taylor, L., Banwart, S. A., Valdes, P. J., Leake, J. R., & Beerling, D. J. (2012). Evaluating the effects of terrestrial ecosystems, climate and carbon dioxide on weathering over geological time: a global-scale process-based approach. Philosophical Transactions of the Royal Society B: Biological Sciences, **367**, 565–582.

Townsend, J., Wright, D., Winfrey, R., Fu, F., Maeder, M., Joung, J., & Voytas, D. (2017). High frequency modification of plant genes using engineered zinc finger nucleases. Nature, **25**, 1032–1057.

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.

Tsuboyama-Tanaka, S., Nonaka, S., & Kodama, Y. (2015). A highly efficient AgarTrap method for genetic transformation of mature thalli of the liverwort *Marchantia polymorpha* L. Plant Biotechnology, **32**, 333–336.

Voytas, D. F., & Gao, C. (2014). Precision genome engineering and agriculture: opportunities and regulatory challenges. PLoS Biology, **12**, 1–6.

Wang, Z. Y., Kenigsbuch, D., Sun, L., Harel, E., Ong, M. S., Tobin, E. M., & Gan, R. (1997). A Myb-related transcription factor is involved in the phytochrome regulation of an *Arabidopsis Lhcb* gene. The Plant Cell, **9**, 491–507.

Wang, Z., Gerstein, M., & Snyder, M. (2001). RNA-seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics, **10**, 57–63.

Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M. S., Burleigh, J. G., Gitzendanner, M. A., Ruhfel, B. R., Wafula, E., Der, J. P., Graham, S. W., Mathews, S., Melkonian, M., Soltis, D. E., Soltis, P. S., Miles, N. W., Rothfels, C. J., Pokorny, L., Shaw, A. J., DeGironimo, L., Stevenson, D. W., Surek, B., Villarreal, J. C., Roure, B., Philippe, H., DePamphilis, C. W., Chen, T., Deyholos, M. K., Baucom, R. S., Kutchan, T. M., Augustin, M. M., Wang, J., Zhang, Y., Tian, Z., Yan, Z., Wu, X., Sun, X., Wong, G. K.-S., & Leebens-Mack, J. (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**, 4859–4868.

Wilson, A. C., Maxson, L. R., & Sarich, V. M. (1974). Two types of molecular evolution: evidence from studies of interspecific hybridization. Proceedings of the National Academy of Sciences of the United States of America, **71**, 2843–2847.

Yamato, K. T., Ishizaki, K., Fujisawa, M., Okada, S., Nakayama, S., Fujishita, M., Bando, H., Yodoya, K., Hayashi, K., Bando, T., Hasumi, A., Nishio, T., Sakata, R., Yamamoto, M., Yamaki, A., Kajikawa, M., Yamano, T., Nishide, T., Choi, S. H., Shimizu-Ueda, Y., Hanajiri, T., Sakaida, M., Kono, K., Takenaka, M., Yamaoka, S., Kuriyama, C., Kohzu, Y., Nishida, H., Brennicke, A., Shin, T., Kohara, Y., Kohchi, T., Fukuzawa, H., & Ohyama, K. (2007). Gene organization of the liverwort Y chromosome reveals distinct sex chromosome evolution in a haploid system. Proceedings of the National Academy of Sciences of the United States of America, **104,** 6472–6477.

Yu, H., Tardivo, L., Tam, S., Weiner, E., Gebreab, F., Fan, C., Svrzikapa, N., Hirozane-Kishikawa, T., Rietman, E., Yang, X., Sahalie, J., Salehi-Ashtiani, K., Hao, T., Cusick, M. E., Hill, D. E., Roth, F. P., Braun, P., Vidal, M. (2011). Leveraging the power of next-generation sequencing to generate interactome datasets. Nature Methods, **8**, 478–480.

Supplementary information

Annex A: Gene nomenclature

Table A1: Gene nomenclature.

Gene	Transcripts	Accession code (*Phytozome*)	Name
Mp*DIV1*	Mp*DIV1.1* Mp*DIV1.2*	Mapoly0026s0070.1 Mapoly0026s0070.2	*Marchantia polymorpha DIVARICATA 1*
Mp*DIV2*	Mp*DIV2.1*	Mapoly0143s0011.1	*Marchantia polymorpha DIVARICATA 2*
Mp*DIVL*	Mp*DIVL.1*	Mapoly0020s0013.1	*Marchantia polymorpha DIVARICATA Like*
Mp*DRIF*	Mp*DRIF.1* Mp*DRIF.2*	Mapoly0127s0027.1 Mapoly0127s0027.2	*Marchantia polymorpha DIV-and-RAD-Interacting Factor*

Annex B: Primers used for cloning

Table B1: Oligonucleotide sequences.

Primer name	Oligonucleotide sequence (5' to 3')	Comments
Mp*DIV1* (453) Fw	ATGGCAGCACCCTCTCCG	Mp*div1* plants genotyping
Mp*DIV1* (454) Rv	CTAATGATGCATGGCAGGCT	
Mp*DIV1* Intron (767) Fw	GGGCGAATCCCTTGGAGTTT	
Mp*DIV1* Intron (766) Rv	AGGGTGGTAAGCAACGAGTG	
Mp*DIV2* (455) Fw	ATGGCAACAACCGTCGC	Mp*div2* plants genotyping
Mp*DIV2* (456) Rv	TTAAGCGTTCGCAGACTGGG	
Mp*DIVL* (457) Fw	ATGACTCGCCGTTGCTCC	Mp*divl* plants genotyping / *in situ* probes
Mp*DIVL* (458) Rv	CTACACCACACTAATAGTGCTC	
Mp*DIVL* Intron (771) Fw	TTGAGTTTGGGTGCTGTCC	
Mp*DIVL* Intron (772) Rv	CAAGAAGAAAGAGGCGGTGT	
Mp*DRIF* (451) Fw	ATGGCGGGCTCCGTCGG	Mp*drif* plants genotyping
Mp*DRIF* Intron (769) Fw	CCTTCCACTGTTCTCCGGTT	
Mp*DRIF* Intron (768) Rv	ACCTTCCGTGCCTACGAATCAT	
Mp*DRIF* 3'UTR (770) Rv	AGGCATAGCACCCCGTTTAC	
Mp*DIV1* attb Fw	AAAAAGCAGGCTTGCGGATGGCAGCACCCTC	Mp*DIV1.1* Gateway cloning (1stPCR)
Mp*DIV1.1* attb Rv	AGAAAGCTGGGTGCTAATGATGCATGGCAGGCT	
Mp*DIV1.2* attb Rv	AGAAAGCTGGGTGTTATCGTACACATATTCCTTGAAATG	Mp*DIV1.2* Gateway cloning (1stPCR)
Mp*DIV2* attb Fw	AAAAAGCAGGCTCACTGATGGCAACAACCGTC	Mp*DIV2* Gateway cloning (1stPCR)
Mp*DIV2* attb Rv	AGAAAGCTGGGTTTTAAGCGTTCGCAGACTGGG	
Mp*DRIF* attb Fw	AAAAAGCAGGCTTCAGAATGGCGGGCTCCG	Mp*DRIF.1* Gateway cloning (1stPCR)
Mp*DRIF.1* attb Rv	AGAAAGCTGGGTGTTACGTTTGTGAAGTTGGAGAC	
Mp*DRIF.2*attb Fw	AAAAAGCAGGCTCAGGCATGTTTGTGATTAGGTTTG	Mp*DRIF.2* Gateway cloning (1stPCR)
Mp*DIVL* CDS Fw	AAGCGGCCGCCATGACTCGCCGTTGCTCCCA	Cloning Mp*DIVL* CDS in pENTRY (NotI)
Mp*DIVL* CDS Rv	AAGGCGCGCCCCTACACCACACTAATAGTGCT	Cloning Mp*DIVL* CDS in pENTRY (AscI)
*pro*Mp*EF1α* Fw	AAAAAGCAGGCTTGCAAATGAGTCACACACATTG	Mp*EF1α* promoter Gateway cloning (1stPCR)
*pro*Mp*EF1α* Rv	AGAAAGCTGGGTTGGTGACAACCTTTCTGCAGG	
*pro*Mp*DIV1* Fw	AAAAAGCAGGCTGCCACTCGCCAAGAACAGAA	Mp*DIV1* promoter Gateway cloning (1st PCR)
*pro*Mp*DIV1* Rv	AGAAAGCTGGGTCACCACTCAGACAGGCACCA	
*pro*Mp*DIV2* Fw	AAAAAGCAGGCTTCTCCGAACATAAAGTGCACG	Mp*DIV2* promoter Gateway cloning (1st PCR)
*pro*Mp*DIV2* Rv	AGAAAGCTGGGTAGTGATCTCCGATTCACCTCA	
*pro*Mp*DIVL* Fw	AAAAAGCAGGCTAAACTGCCTAATCACCGGGAA	Mp*DIVL* promoter Gateway cloning (1st PCR)
*pro*Mp*DIVL* Rv	AGAAAGCTGGGTTCCAGTTCTTCTGGCCCTGA	

Table B1 (continuation): Oligonucleotide sequences.

Primer name	Oligonucleotide sequence (5' to 3')	Comments
*pro*Mp*DRIF* Fw	AAAAAGCAGGCTAGATGTCCCGACTCGTGATAG	Mp*DRIF* promoter Gateway cloning (1st PCR)
*pro*Mp*DRIF.1* Rv	AGAAAGCTGGGTTGAGCGGGCCAATCACTCA	Mp*DRIF.1* promoter Gateway cloning (1st PCR)
*pro*Mp*DRIF.2* Rv	AGAAAGCTGGGTACCAGTGGGAGCCTTGGTAGA	Mp*DRIF.2* promoter Gateway cloning (1st PCR)
attB1 Fw attB2 Rv	GGGGACAAGTTTGTACAAAAAAGCAGGCT GGGGACCACTTTGTACAAGAAAGCTGGGT	Gateway 2nd PCR
pDONOR Fw pDONOR Rv	TCGCGTTAACGCTAGCATGGATCT GTAACATCAGAGATTTTGAGACAC	pDONOR specific primers for PCR colony
M13 Fw M13 Rv	GTTGTAAAACGACGGCCAGT GGAAACAGCTATGACCAT	Specific primers for the M13 region
NOS terminator Rv	GATAATCATCGCAAGACC	Specific primer for the NOS terminator region
T7 promoter Fw	TAATACGACTCACTATAGGG	Specific primer for the T7 promoter region
Mp female marker Fw Mp female marker Rv Mp male marker Fw Mp male marker Rv	CACCATGGGCCTACTTGTTTCAGTCGCTGGTGG TCAAAGGCTAGTGTTTCCATTACTTGGAC GCAGCTGTGTTTTGTGCAGATCGTC ATTCTGACCTTACAAGAAATCCTCC	Sex determining specific primers
Mp*EF1α* RT Fw Mp*EF1α* RT Rv	CCGAGATCCTGACCAAGG GAGGTGGGTACTCAGCGAAG	Mp*EF1α* specific primers for RT-PCR
Mp*DIV1* RT Fw Mp*DIV1* RT Rv	CGAAGAAGAGCACAGGTCATT ATCGGCGCTTATCTTTGTTG	Mp*DIV1* specific primers for RT-PCR
Mp*DIV2* RT Fw	GTCCGAAGAAGAGCACAGATTA	Mp*DIV2* specific primers for RT-PCR
Mp*DIV2* RT Rv	CGTGATGTCGTGGATACTGG	
Mp*DIVL* RT Fw	CGAAAGAAAGAAAGGTGTACCG	Mp*DIVL* specific primers for RT-PCR
Mp*DIVL* RT Rv	GACGCTTCCTTTTGTTCAGG	
Mp*DRIF* RT Fw	TCAGACCCGGTTTGTCTAGC	Mp*DRIF* specific primers for RT-PCR
Mp*DRIF* RT Rv	CGGTGTTCTCTTGCAGCTT	
Mp*EF1α* Fw	ATGGGAAAGGAGAAATTCC	*In situ* probes
Mp*EF1α* Rv	CACTTTTTCTTGGCGGC	
Mp*DIV1* is Fw	AGGATCCATGGCAGCACCCTCTC	
Mp*DIV1* is Rv	AACTCGAGTTATCGTACACATATTCCTTGAAA	
Mp*DIV2* is Fw	AGGATCCATGGCAACAACCGTCGC	
Mp*DIV2* is Rv	AAAGTCGACTTAAGCGTTCGCAGACTGG	
Mp*DRIF* is Fw	TCGGATCCATATGGCGGGCTCCGTCG	
Mp*DRIF* is Rv	AGGATCCTTACGTTTGTGAAGTTGGAGAG	

Table B2: gRNAs oligonucleotide sequences.

Oligo	Oligonucleotide sequence
Mp*DIV1*	
Mp*DIV1* 262 L1	CTCGCGAAGAATCGCCTAATCGCT
Mp*DIV1* 262 L2	AAACAGCGATTAGGCGATTCTTCG
Mp*DIV1* 275 R1	CTCGCGCGTCCTTGCCAGGAACCA
Mp*DIV1* 275 R2	AAACTGGTTCCTGGCAAGGACGCG
Mp*DIV1* 901 R1	CTCGGTTGTTGCACCGGGACCTAC
Mp*DIV1* 901 R2	AAACGTAGGTCCCGGTGCAACAAC
Mp*DIV1* 1000 L1	CTCGGCACATGGTGAGGCCCGGAA
Mp*DIV1* 1000 L2	AAACTTCCGGGCCTCACCATGTGC
Mp*DIV2*	
Mp*DIV2* 17 R1	CTCGTCCAGGAGACCCCGCTGTTT
Mp*DIV2* 17 R2	AAACAAACAGCGGGGTCTCCTGGA
Mp*DIV2* 242 L1	CTCGCCCGTTCACGGAGGACACCA
Mp*DIV2* 242 L2	AAACTGGTGTCCTCCGTGAACGGG
Mp*DIV2* 829 R1	CTCGGCAGGCCCGTGATGGGCCCT
Mp*DIV2* 829 R2	AAACAGGGCCCATCACGGGCCTGC
Mp*DIV2* 836 R1	CTCGGGCGATGGCAGGCCCGTGAT
Mp*DIV2* 836 R2	AAACATCACGGGCCTGCCATCGCC
Mp*DIV2* 722 L1	CTCGTGGCCAAGGTCAGGCCTACA
Mp*DIV2* 722 L2	AAACTGTAGGCCTGACCTTGGCCA
Mp*DIVL*	
Mp*DIVL* 30 L1	CTCGTGCTCCCACTGTGGTCATAA
Mp*DIVL* 30 L2	AAACTTATGACCACAGTGGGAGCA
Mp*DIVL* 786 R1	CTCGAGGTCTCGGTAGTACCGTTA
Mp*DIVL* 786 R2	AAACTAACGGTACTACCGAGACCT
Mp*DIVL* 883 R1	CTCGTCTTGGAGACGAACGCGCTG
Mp*DIVL* 883 R2	AAACCAGCGCGTTCGTCTCCAAGA
Mp*DIVL* 950 L1	CTCGCGGTCCGAGCACTATTAGTG
Mp*DIVL* 950 L2	AAACCACTAATAGTGCTCGGACCG

Table B2 (continuation): gRNAs oligonucleotide sequences.

Oligo	Oligonucleotide sequence
Mp*DRIF*	
Mp*DRIF* 78 L1	CTCGCTGCGTCGCCGGCCGTCAAT
Mp*DRIF* 78 L2	AAACATTGACGGCCGGCGACGCAG
Mp*DRIF* 270 L1	CTCGTCCAGTTGTTACATGACCCA
Mp*DRIF* 270 L2	AAACTGGGTCATGTAACAACTGGA
Mp*DRIF* 470 R1	CTCGCAGATAGAGGACGTAGAACA
Mp*DRIF* 470 R2	AAACTGTTCTACGTCCTCTATCTG
Mp*DRIF* 530 R1	CTCGACCGGGTCTGATATGACCAG
Mp*DRIF* 530 R2	AAACCTGGTCATATCAGACCCGGT
Mp*DRIF* 818 R1	CTCGAAGGATCGTATCTGCCAATT
Mp*DRIF* 818 R2	AAACAATTGGCAGATACGATCCTT

Annex C: Vector maps

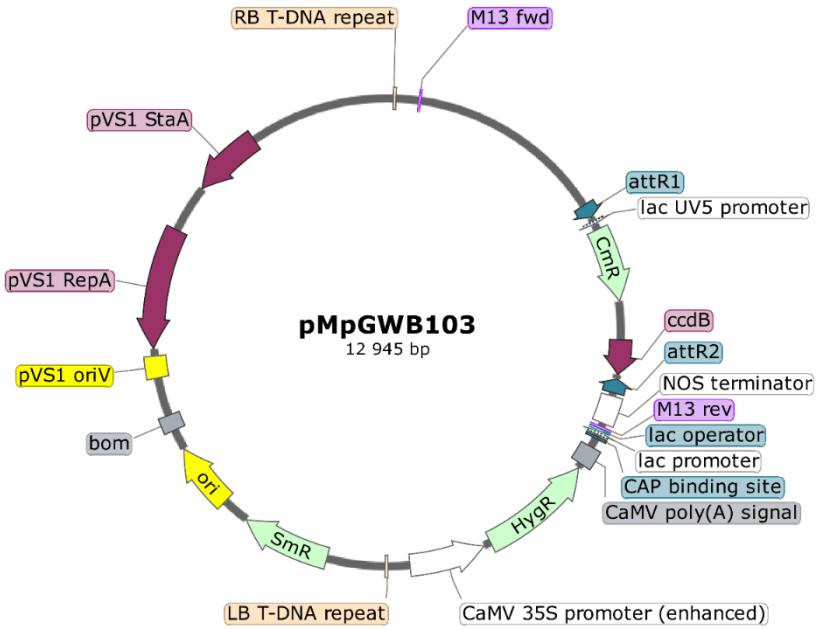

Figure C1: pMpGWB103 vector map. Gateway destination vector with the strong promoter Mp*EF1α* and with the hygromycin, chloramphenicol and spectinomycin resistance marker genes. This vector was used for overexpression assays.

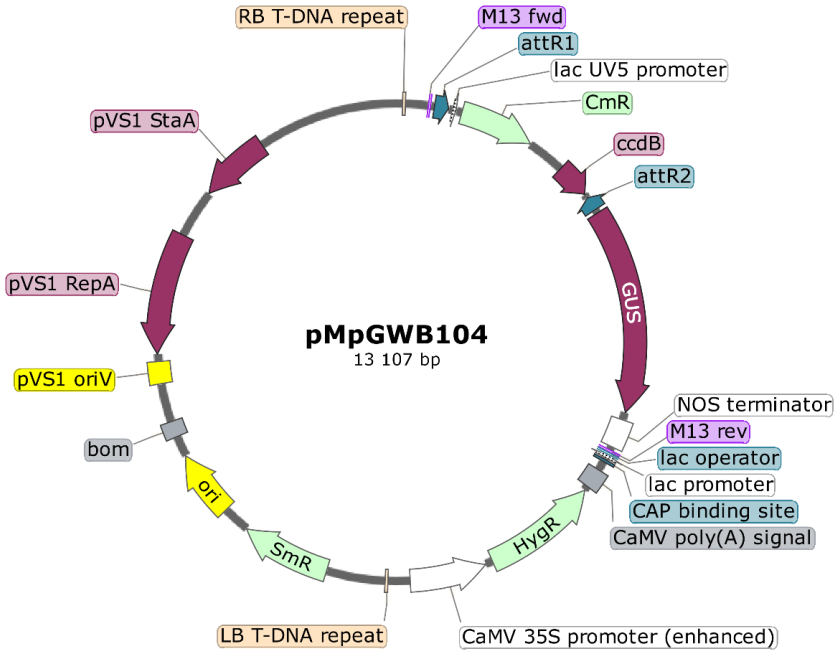

Figure C2: pMpGWB104 vector map. Gateway destination vector for *GUS* fusion, under the regulation of the 35S promoter and with the hygromycin, chloramphenicol and spectinomycin resistance marker genes. This vector was used for reporter gene fusions.

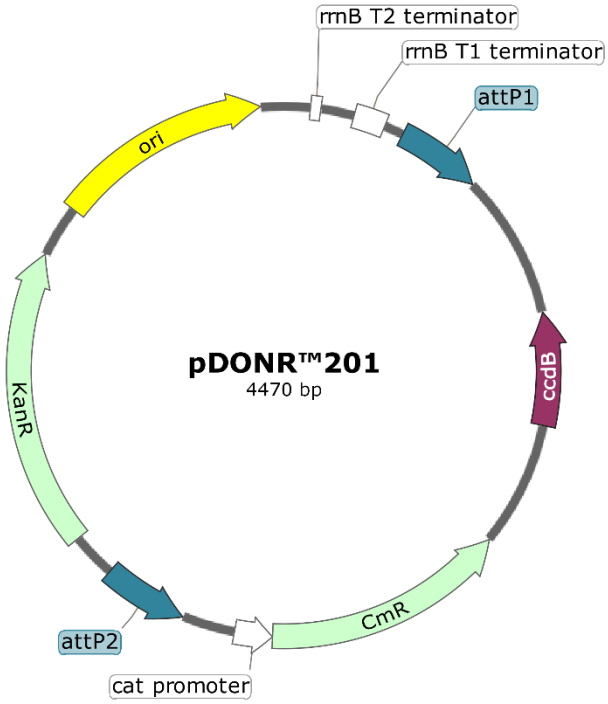

Figure C3: pDONR™201 vector map. Gateway donor vector for the cloning of DNA fragments with the kanamycin resistance marker gene. This vector was used to create entry clones.

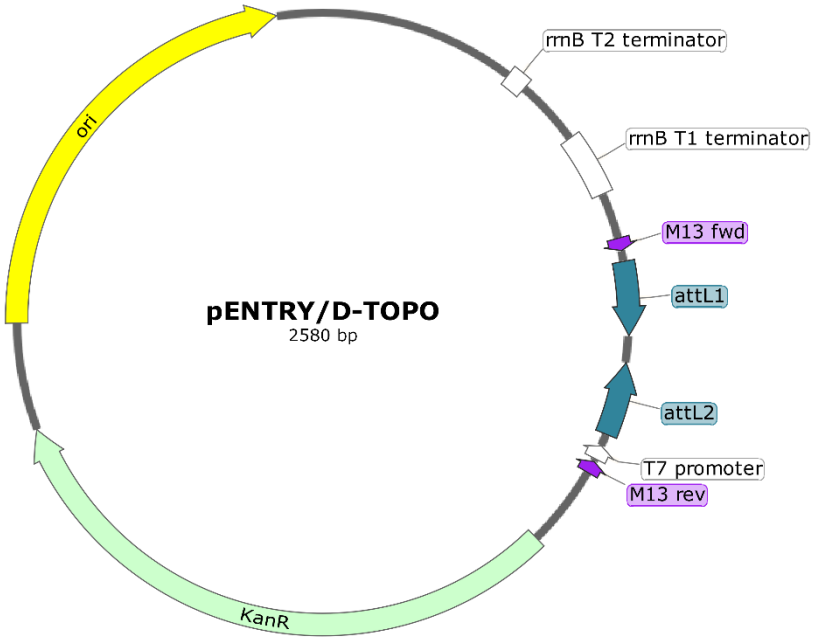

Figure C4: pENTRY/D-TOPO vector map. Gateway donor vector for the cloning of DNA fragments with the kanamycin resistance marker gene. This vector was used to create entry clones.

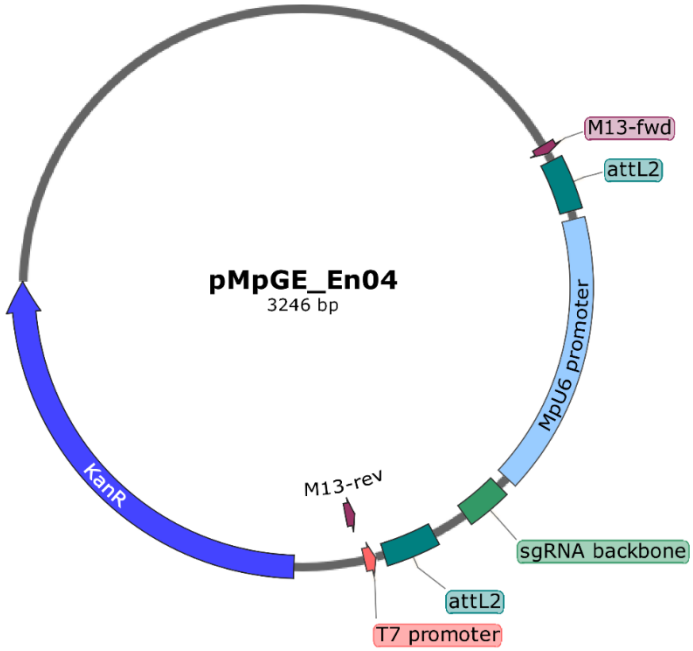

Figure C5: pMpGe_En04 vector map. Gateway vector for the creation of gRNAs under the regulation of the promoter Mp*U6-1* and with the kanamycin resistance marker gene. This vector was used for CRISPR-Cas9 mutation of Mp*DIV1*, Mp*DIV2*, Mp*DIVL* and Mp*DRIF*.

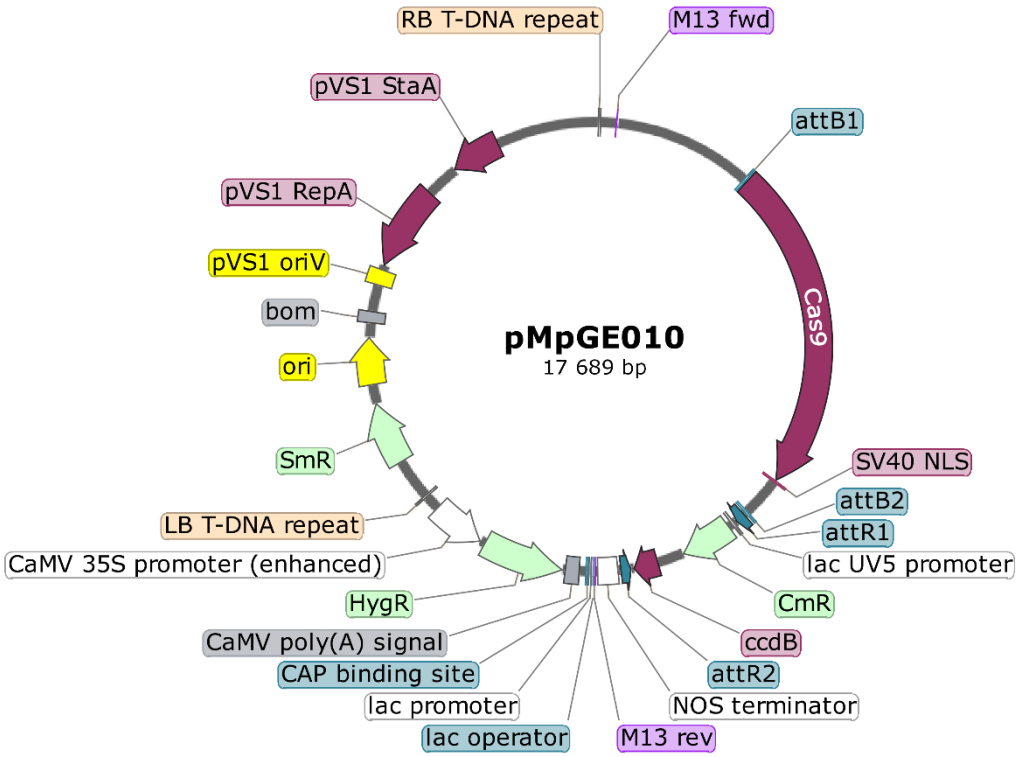

Figure C6: pMpGE010 vector map. Gateway destination vector with the Cas9 and with the hygromycin, chloramphenicol and spectinomycin resistance marker genes. This vector was used for the CRISPR-Cas9 transformation of *Marchantia polymorpha*.

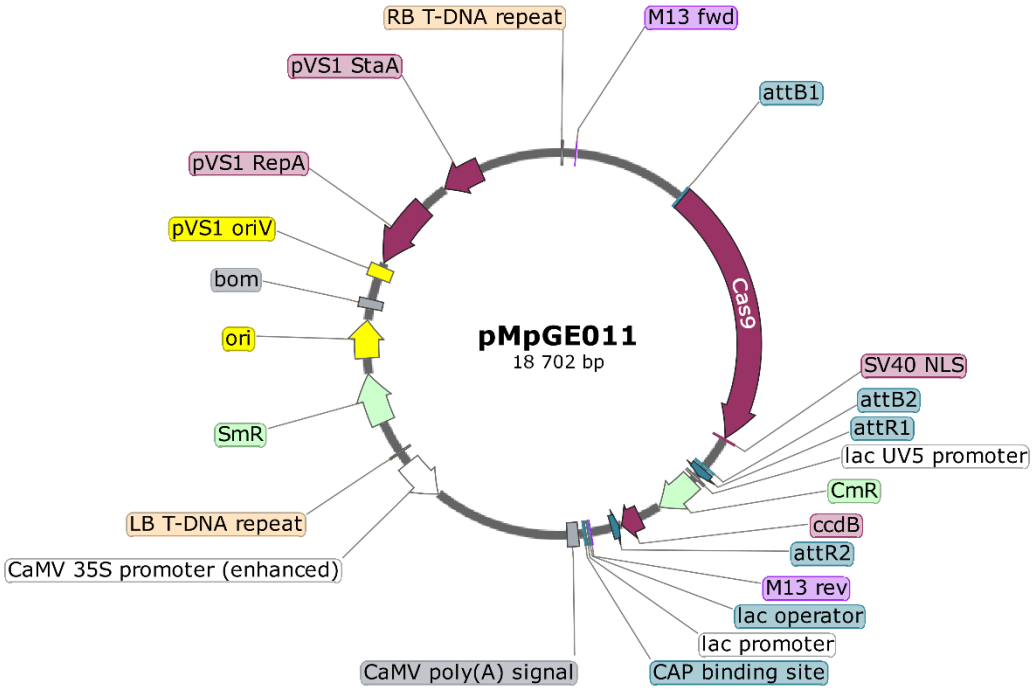

Figure C7: pMpGE011 vector map. Gateway destination vector with the Cas9 and with the chlorsulfuron, chloramphenicol and spectinomycin resistance marker genes. This vector was used for the CRISPR-Cas9 transformation of *Marchantia polymorpha*.

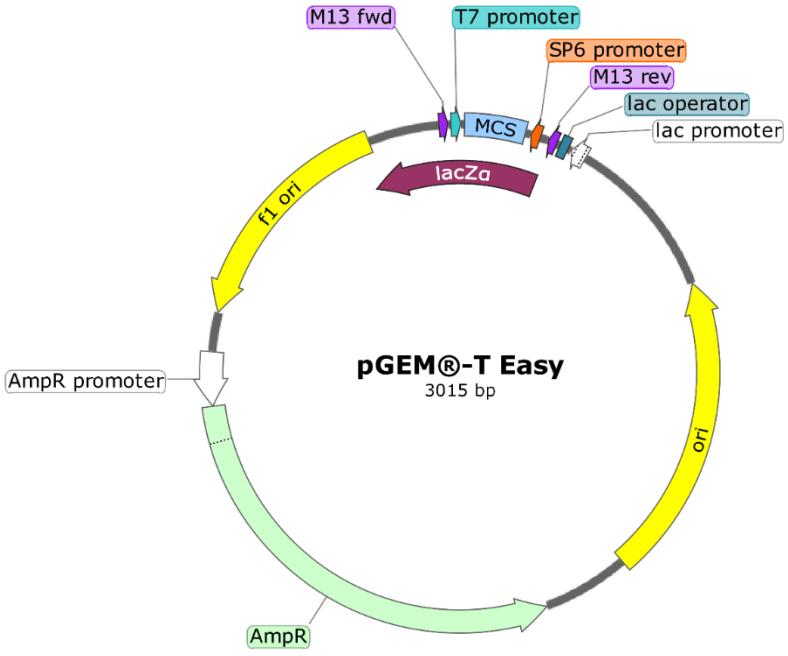

Figure C8: pGEM®-T easy vector map. Cloning vector with the T7 and SP6 promoters and with the ampicillin resistance marker gene. This vector was used for the *in situ* probes synthesis.

Annex D: Nucleotide and protein sequences

In the genomic sequences, the introns are not underlined, the exons underlined at blue and the insertions in the mutant genomic sequences at green. The MpDIV protein sequences have the conserved MYB motifs at bold; the MpDIVL protein sequence have the R motif marked at red and the MYB motif at bold and the MpDRIF protein sequences have the MYB motif at bold and the DUF3755 at bold and underlined.

Mp*DIV1*

>Mp*DIV1.1* genomic sequence

ATGGCAGCACCCTCTCCGGGGCCCTCACCATCCTCCCCATCCACTGCCAGTACTCCGATTCCTGCGAGCGCTGCT
GCCGCCGCGGCAGGACCGAGTGCCGCGCCTGTCGTGCCCACCCCTTCTGCTGTCTCAGCTCCGACGATTCCCGTC
ATTCCGGCCCCCGATGTTGCTCCGGCAGTCTCTTCTTCATGGACCTCTGAGCAAGACAAGCTCTTCGAGAATGCA
TTGGCGGTGTACGACGAAGAATCGCCTAATCGCTGGGACAATGTGGCCTCCATGGTTCCTGGCAAGGACGCGGCG
GATGTGATGAAGCACTACGAGCTCCTCACGGAAGATGTGACCTCGATTGATGCGGGCAGAGTTGCCCTTCCGAGC
TACATTCTTCCGGGCAGCTTATCAGGTGCAGATGCTGCTGGCGAACAGTCCGATAGCTCGGTGTCCAAGAACAAA
GCGTGGTCTGGCCAGTCTCCGGGCGTCTCGGCCAGTGGAACCAGTGGAACGGTGGGTGGTCTTGAGAGGAAGTCC
AGTAGTAGCAAAGCGGACCAAGAAAGACGCAAGGGCATCCCGTGGACCGAAGAAGAGCACAGGTCTGCCGCTACA
CTCAGTTGTACTCTTTCACGTTCAGCTCATAGATTTGGTCTCTGGTTGATCACTTTAATCACCTTCAGGGATGTG
GGTGGTTTCTGACCCGGAGTATAAGTGTGGGCCACCAGTGCACATCCCTGGATTATGTCGTCAATAGTAAACCCC
GGTGTTCACCATTTATCGGTTGGTTCCATTTACTTTGTTGTGTGCTTCGATCCATCTCGATTAGTGTGTTGACAT
ACTATCATCAAGTGAGGGTGCTGAGTGAACGCCCAAGCTGATACGATCTGTATAATCCAATCCCTAGTGTTCAAT
GGGAAATAATAAACCCTCTGCGATGAAGCTGAGTTGCAAATTCTTGGAATCATGTGCCTCAGTGGAGCCAGAACT
GTTGATTTGTTCACCCTCTTGGATTATAGAGGATGAACATCGCAAGTGTAGCCGACTTGGGCGGACCTCGGAAGA
AGAGAATGGCCGCTCACTCGTTGCTTACCACCCTTGAGAGCTAGTTATCGTGATAACGACGGGATTACCTACAGT
TCAAGACTCTTATGTAGGTTGCATTCACTGAACTCACCACCTTCGTTGAACAAAATCAGAGTCTTCTTTCACCGA
CGACCTGGGCGAATCCCTTGGAGTTTCACGGAGGCTGACATATGACGCAGTGACATGGATGTGACATACTTTGAT
GTTGCTCTTGCCCGTGCTTTTATGCGCACATCATAGAAGCCAGTCATGGGGATGGAGTTTATTCGGTACAAGCAT
AGGGTGCTTTGTACCTACTAAAACGTATTTTACGAATGCACAATCTATGCAAGAGCTGGTTTGTGATAACACATT
ATCACAAACGAGTGGTAGTGGACCTTCATTCTCTTGAATCGGAAGGGTGTTCCTGGATGAGATTCTTACTCAAAC
TTCGACCTTTCAGCGGAAGACTTTCGCTTGATCTTGCTGACCTCATTGCCTTTTACACTGAAGTGTATAGGTAGA
GTAATACCTTCAAATTTCACCTAGATTTGTTGTTAGTGACATCTGTTGTTGGTGTGTGTCGTGTGCCGCAGGTCA
TTTTTACTTGGCCTTGCAAAGTTCGGCAAAGGTGACTGGAGGAGCATATCTCGAAATTTTGTAATATCGCGTACA
CCAACGCAAGTTGCGAGTCACGCGCAGAAGTATTTCATCAGATTGAACTCTATCAACAAAGATAAGCGCCGATCC
TCCATCCACGATATCACTAGCGTCAACTCAGCCGGAGTGGAAGTTATGCAGGGCAGTCCTGGTCCCATCACTGGT
CAATCACCTTCAGGAACGTCGACGTCCGGGCAACCTCTTCCTCACAAAACGCAACCTGCATTTCAAGGTGGGATG
TACGTTACTCCTGTAGGTCCCGGTGCAACAACAGGGTTGGGTACCCCGGTCTATCCTCCAGGACAGATGGGTTAT
GGAGTGAGAGGGCACATGGTGAGGCCCGGAATGGGAGGTCCCCCAATGAATATGACTCACATGACCTACAGTATG
CCACAGCCTGCCATGCATCATTAG

>Mp*DIV1.2* genomic sequence

ATGGCAGCACCCTCTCCGGGGCCCTCACCATCCTCCCCATCCACTGCCAGTACTCCGATTCCTGCGAGCGCTGCT
GCCGCCGCGGCAGGACCGAGTGCCGCGCCTGTCGTGCCCACCCCTTCTGCTGTCTCAGCTCCGACGATTCCCGTC
ATTCCGGCCCCCGATGTTGCTCCGGCAGTCTCTTCTTCATGGACCTCTGAGCAAGACAAGCTCTTCGAGAATGCA
TTGGCGGTGTACGACGAAGAATCGCCTAATCGCTGGGACAATGTGGCCTCCATGGTTCCTGGCAAGGACGCGGCG
GATGTGATGAAGCACTACGAGCTCCTCACGGAAGATGTGACCTCGATTGATGCGGGCAGAGTTGCCCTTCCGAGC
TACATTCTTCCGGGCAGCTTATCAGGTGCAGATGCTGCTGGCGAACAGTCCGATAGCTCGGTGTCCAAGAACAAA
GCGTGGTCTGGCCAGTCTCCGGGCGTCTCGGCCAGTGGAACCAGTGGAACGGTGGGTGGTCTTGAGAGGAAGTCC
AGTAGTAGCAAAGCGGACCAAGAAAGACGCAAGGGCATCCCGTGGACCGAAGAAGAGCACAGGTCTGCCGCTACA
CTCAGTTGTACTCTTTCACGTTCAGCTCATAGATTTGGTCTCTGGTTGATCACTTTAATCACCTTCAGGGATGTG

GGTGGTTTCTGACCCGGAGTATAAGTGTGGGCCACCAGTGCACATCCCTGGATTATGTCGTCAATAGTAAACCCC
GGTGTTCACCATTTATCGGTTGGTTCCATTTACTTTGTTGTGTGCTTCGATCCATCTCGATTAGTGTGTTGACAT
ACTATCATCAAGTGAGGGTGCTGAGTGAACGCCCAAGCTGATACGATCTGTATAATCCAATCCCTAGTGTTCAAT
GGGAAATAATAAACCCTCTGCGATGAAGCTGAGTTGCAAATTCTTGGAATCATGTGCCTCAGTGGAGCCAGAACT
GTTGATTTGTTCACCCTCTTGGATTATAGAGGATGAACATCGCAAGTGTAGCCGACTTGGGCGGACCTCGGAAGA
AGAGAATGGCCGCTCACTCGTTGCTTACCACCCTTGAGAGCTAGTTATCGTGATAACGACGGGATTACCTACAGT
TCAAGACTCTTATGTAGGTTGCATTCACTGAACTCACCACCTTCGTTGAACAAAATCAGAGTCTTCTTTCACCGA
CGACCTGGGCGAATCCCTTGGAGTTTCACGGAGGCTGACATATGACGCAGTGACATGGATGTGACATACTTTGAT
GTTGCTCTTGCCCGTGCTTTTATGCGCACATCATAGAAGCCAGTCATGGGGATGGAGTTTATTCGGTACAAGCAT
AGGGTGCTTTGTACCTACTAAAACGTATTTTACGAATGCACAATCTATGCAAGAGCTGGTTTGTGATAACACATT
ATCACAAACGAGTGGTAGTGGACCTTCATTCTCTTGAATCGGAAGGGTGTTCCTGGATGAGATTCTTACTCAAAC
TTCGACCTTTCAGCGGAAGACTTTCGCTTGATCTTGCTGACCTCATTGCCTTTTACACTGAAGTGTATAGGTAGA
GTAATACCTTCAAATTTCACCTAGATTTGTTGTTAGTGACATCTGTTGTTGGTGTGTGTCGTGTGCCGCAGGTCA
TTTTTACTTGGCCTTGCAAAGTTCGGCAAAGGTGACTGGAGGAGCATATCTCGAAATTTTGTAATATCGCGTACA
CCAACGCAAGTTGCGAGTCACGCGCAGAAGTATTTCATCAGATTGAACTCTATCAACAAAGATAAGCGCCGATCC
TCCATCCACGATATCACTAGCGTCAACTCAGCCGGAGTGGAAGTTATGCAGGGCAGTCCTGGTCCCATCACTGGT
CAATCACCTTCAGGAACGTCGACGTCCGGGCAACCTCTTCCTCACAAAACGCAACCTGCATTTCAAGGTGGGATG
TACGTTACTCCTGTAGGTCCCGGTGCAACAACAGGGTTGGGTACCCCGGTCTATCCTCCAGGACAGATGGGTTAT
GGAGTGAGAGGGCACATGGTGAGGCCCGGAATGGGAGGTCCCCCAATGAATATGACTCACATGACCTACAGTATG
CCACAGCCTGCCATGCATCATTAGCTGCTCGATACGATCTACCCCCTTTTATCTTGTCTCGTATACCAGAATTCT
CTTTTTGACCTAGTGTCCCCCATGTTGCACTATCTTCACCATCTTCCTTGAGCAAACTCATACTTGTGAACTGCA
CTGTATACCCTGGAAGGGAGGTCTTGATCTGAAAAGACGTCAGCAGACGTACCTTTCGGAGAAGTCATCTCCAGA
AAGATTATTGAGGGTTTGCAACAGTTTTCCTGTTTCCATGGTAAGCATGCCCTGCTGAGAATTTCATCCTGCCAC
ACGCATCCAATCCATGAAAATCTTTGGGAGCTCTTTTGTAATTCCTCTATGAACCCAGTACCTCTCATGTCCAAT
CTGTACACCTCAAACTTTGAGTGTCACAAATCTCAAGTACAGTCTTTGTTTCTCTCTTCCACCAATTCAACGTAG
TATTTCCAAGCGCCTTGCCTCCACGCTGAGGAGATGGCGGGCTCTGCTGTACAGTACACTCCTTCCCTGGAATCC
CTCCCTCTGGCTTCATGCAGGCACCTACTGGGTAGTCTACCTTGGGATGGCTGCAATTCATCTTCCTAGGTGACT
AGGGCTGTTTGCAAAGATTGGGTTGAGGAGAAAGAGAAGCAACGAACCACAGGACTCGTGGTGCTGTTGGTTCGG
ATGGTGATGTAAGGACGACTTTTGTGCAGTATGGTACTAAAATGGATACTAACGTTACTCCTGATAAGATTTGTA
TTCTAGAGTTAATCTGCCGGAAGTACTTGAAAGCTACACAACACACTCCACTCATTTTTCCCCTTCCTCAGGCGT
AAGCATGAGGATAGCTTCTAGTTTACCTGCTTTGCCGTAGGCTCTGAGAGTTGGAGAGGATGTGATGCCAGGACT
TGAGTAAAGAGCCCGATACTACGTGCTATACTTGATTTAGTTCGAGCTTGCCACACCAATAGCACTACTGCAGAG
TTGGTATACTAAATGCCACGGGAAACAACTATTCTCTACGTGATCCAAACTGGGGGCACGTAAGGGAAGTTGAAG
TTCTGATGTCAAATTGTAATAAGTCTGGGTGGCTTACGATGTCCATTTGTACATACACAGTGGCCCAATTAAATG
GTGTCACTCTTTTGGAGCTCAAAGGAGAAGATTGTAGCCATGCTATGTTTTTACATATAATGCCTCTGTCTCAAG
CGCGGAAGACGAATTGCCCAGAACAGTGTCAATGGCACTGACCTCAGTTCAATCCATTCACCTGTCTTATCATGA
CCTGACCAAAATATCTGCCATCAAACGCTTGTGCCGTATTAAAGAGCTTGTCTCTACGTCATCGTCGAACTACAA
AGAGTAAAGCTTAACGTCGGATTCATTAGTGAAATGGCTGCATTTGACTTGGAAGTAACGTGGGAAACTTGTGAA
TGTAGGAATATGTGTACGATAA

>Mp*DIV1.1* coding sequence

ATGGCAGCACCCTCTCCGGGGCCCTCACCATCCTCCCCATCCACTGCCAGTACTCCGATTCCTGCGAGCGCTGCT
GCCGCCGCGGCAGGACCGAGTGCCGCGCCTGTCGTGCCCACCCCTTCTGCTGTCTCAGCTCCGACGATTCCCGTC
ATTCCGGCCCCCGATGTTGCTCCGGCAGTCTCTTCTTCATGGACCTCTGAGCAAGACAAGCTCTTCGAGAATGCA
TTGGCGGTGTACGACGAAGAATCGCCTAATCGCTGGGACAATGTGGCCTCCATGGTTCCTGGCAAGGACGCGGCG
GATGTGATGAAGCACTACGAGCTCCTCACGGAAGATGTGACCTCGATTGATGCGGGCAGAGTTGCCCTTCCGAGC
TACATTCTTCCGGGCAGCTTATCAGGTGCAGATGCTGCTGGCGAACAGTCCGATAGCTCGGTGTCCAAGAACAAA
GCGTGGTCTGGCCAGTCTCCGGGCGTCTCGGCCAGTGGAACCAGTGGAACGGTGGGTGGTCTTGAGAGGAAGTCC
AGTAGTAGCAAAGCGGACCAAGAAAGACGCAAGGGCATCCCGTGGACCGAAGAAGAGCACAGGTCATTTTTACTT
GGCCTTGCAAAGTTCGGCAAAGGTGACTGGAGGAGCATATCTCGAAATTTTGTAATATCGCGTACACCAACGCAA
GTTGCGAGTCACGCGCAGAAGTATTTCATCAGATTGAACTCTATCAACAAAGATAAGCGCCGATCCTCCATCCAC
GATATCACTAGCGTCAACTCAGCCGGAGTGGAAGTTATGCAGGGCAGTCCTGGTCCCATCACTGGTCAATCACCT
TCAGGAACGTCGACGTCCGGGCAACCTCTTCCTCACAAAACGCAACCTGCATTTCAAGGTGGGATGTACGTTACT
CCTGTAGGTCCCGGTGCAACAACAGGGTTGGGTACCCCGGTCTATCCTCCAGGACAGATGGGTTATGGAGTGAGA
GGGCACATGGTGAGGCCCGGAATGGGAGGTCCCCCAATGAATATGACTCACATGACCTACAGTATGCCACAGCCT
GCCATGCATCATTAG

>Mp*DIV1.2* coding sequence

ATGGCAGCACCCTCTCCGGGGCCCTCACCATCCTCCCCATCCACTGCCAGTACTCCGATTCCTGCGAGCGCTGCT
GCCGCCGCGGCAGGACCGAGTGCCGCGCCTGTCGTGCCCACCCCTTCTGCTGTCTCAGCTCCGACGATTCCCGTC
ATTCCGGCCCCCGATGTTGCTCCGGCAGTCTCTTCTTCATGGACCTCTGAGCAAGACAAGCTCTTCGAGAATGCA
TTGGCGGTGTACGACGAAGAATCGCCTAATCGCTGGGACAATGTGGCCTCCATGGTTCCTGGCAAGGACGCGGCG
GATGTGATGAAGCACTACGAGCTCCTCACGGAAGATGTGACCTCGATTGATGCGGGCAGAGTTGCCCTTCCGAGC
TACATTCTTCCGGGCAGCTTATCAGGTGCAGATGCTGCTGGCGAACAGTCCGATAGCTCGGTGTCCAAGAACAAA
GCGTGGTCTGGCCAGTCTCCGGGCGTCTCGGCCAGTGGAACCAGTGGAACGGTGGGTGGTCTTGAGAGGAAGTCC
AGTAGTAGCAAAGCGGACCAAGAAAGACGCAAGGGCATCCCGTGGACCGAAGAAGAGCACAGGTCATTTTTACTT
GGCCTTGCAAAGTTCGGCAAAGGTGACTGGAGGAGCATATCTCGAAATTTTGTAATATCGCGTACACCAACGCAA
GTTGCGAGTCACGCGCAGAAGTATTTCATCAGATTGAACTCTATCAACAAAGATAAGCGCCGATCCTCCATCCAC
GATATCACTAGCGTCAACTCAGCCGGAGTGGAAGTTATGCAGGGCAGTCCTGGTCCCATCACTGGTCAATCACCT
TCAGGAACGTCGACGTCCGGGCAACCTCTTCCTCACAAAACGCAACCTGCATTTCAAGGAATATGTGTACGATAA

>Mp*div1* #69 genomic sequence

ATGGCAGCACCCTCTCCGGGGCCCTCACCATCCTCCCCATCCACTGCCAGTACTCCGATTCCTGCGAGCGCTGCT
GCCGCCGCGGCAGGACCGAGTGCCGCGCCTGTCGTGCCCACCCCTTCTGCTGTCTCAGCTCCGACGATTCCCGTC
ATTCCGGCCCCCGATGTTGCTCCGGCAGTCTCTTCTTCATGGACCTCTGAGCAAGACAAGCTCTTCGAGAATGCA
TTGGCGGTGTACGACGAAGAATCGCCTAATCGCTGGGACAATGTGGCCTCCATGGTTCCTGGCAAGGACGCGGCG
GATGTGATGAAGCACTACGAGCTCCTCACGGAAGATGTGACCTCGATTGATGCGGGCAGAGTTGCCCTTCCGAGC
TACATTCTTCCGGGCAGCTTATCAGGTGCAGATGCTGCTGGCGAACAGTCCGATAGCTCGGTGTCCAAGAACAAA
GCGTGGTCTGGCCAGTCTCCGGGCGTCTCGGCCAGTGGAACCAGTGGAACGGTGGGTGGTCTTGAGAGGAAGTCC
AGTAGTAGCAAAGCGGACCAAGAAAGACGCAAGGGCATCCCGTGGACCGAAGAAGAGCACAGGTCTGCCGCTACA
CTCAGTTGTACTCTTTCACGTTCAGCTCATAGATTTGGTCTCTGGTTGATCACTTTAATCACCTTCAGGGATGTG
GGTGGTTTCTGACCCGGAGTATAAGTGTGGGCCACCAGTGCACATCCCTGGATTATGTCGTCAATAGTAAACCCC
GGTGTTCACCATTTATCGGTTGGTTCCATTTACTTTGTTGTGTGCTTCGATCCATCTCGATTAGTGTGTTGACAT
ACTATCATCAAGTGAGGGTGCTGAGTGAACGCCCAAGCTGATACGATCTGTATAATCCAATCCCTAGTGTTCAAT
GGGAAATAATAAACCCTCTGCGATGAAGCTGAGTTGCAAATTCTTGGAATCATGTGCCTCAGTGGAGCCAGAACT
GTTGATTTGTTCACCCTCTTGGATTATAGAGGATGAACATCGCAAGTGTAGCCGACTTGGGCGGACCTCGGAAGA
AGAGAATGGCCGCTCACTCGTTGCTTACCACCCTTGAGAGCTAGTTATCGTGATAACGACGGGATTACCTACAGT
TCAAGACTCTTATGTAGGTTGCATTCACTGAACTCACCACCTTCGTTGAACAAAATCAGAGTCTTCTTTCACCGA
CGACCTGGGCGAATCCCTTGGAGTTTCACGGAGGCTGACATATGACGCAGTGACATGGATGTGACATACTTTGAT
GTTGCTCTTGCCCGTGCTTTTATGCGCACATCATAGAAGCCAGTCATGGGGATGGAGTTTATTCGGTACAAGCAT
AGGGTGCTTTGTACCTACTAAAACGTATTTTACGAATGCACAATCTATGCAAGAGCTGGTTTGTGATAACACATT
ATCACAAACGAGTGGTAGTGGACCTTCATTCTCTTGAATCGGAAGGGTGTTCCTGGATGAGATTCTTACTCAAAC
TTCGACCTTTCAGCGGAAGACTTTCGCTTGATCTTGCTGACCTCATTGCCTTTTACACTGAAGTGTATAGGTAGA
GTAATACCTTCAAATTTCACCTAGATTTGTTGTTAGTGACATCTGTTGTTGGTGTGTGTCGTGTGCCGCAGGTCA
TTTTTACTTGGCCTTGCAAAGTTCGGCAAAGGTGACTGGAGGAGCA**GGAGGACCTGCAA**AGGTCCCGGTGCAACA
ACAGGGTTGGGTACCCCGGTCTATCCTCCAGGACAGATGGGTTATGGAGTGAGAGGGCACATGGTGAGGCCCGGA
ATGGGAGGTCCCCCAATGAATATGACTCACATACCTACAGTATGCCACAGCCTGCCATGCATCATTAG

>Mp*div1* #87 genomic sequence

ATGGCAGCACCCTCTCCGGGGCCCTCACCATCCTCCCCATCCACTGCCAGTACTCCGATTCCTGCGAGCGCTGCT
GCCGCCGCGGCAGGACCGAGTGCCGCGCCTGTCGTGCCCACCCCTTCTGCTGTCTCAGCTCCGACGATTCCCGTC
ATTCCGGCCCCCGATGTTGCTCCGGCAGTCTCTTCTTCATGGACCTCTGAGCAAGACAAGCTCTTCGAGAATGCA
TTGGCGGTGTACGACGAAGAATCGCCTAATCGCTGGGACAATGTGGCCTCCATGGTTCCTGGCAAGGACGCGGCG
GATGTGATGAAGCACTACGAGCTCCTCACGGAAGATGTGACCTCGATTGATGCGGGCAGAGTTGCCCTTCCGAGC
TACATTCTTCCGGGCAGCTTATCAGGTGCAGATGCTGCTGGCGAACAGTCCGATAGCTCGGTGTCCAAGAACAAA
GCGTGGTCTGGCCAGTCTCCGGGCGTCTCGGCCAGTGGAACCAGTGGAACGGTGGGTGGTCTTGAGAGGAAGTCC
AGTAGTAGCAAAGCGGACCAAGAAAGACGCAAGGGCATCCCGTGGACCGAAGAAGAGCACAGGTCTGCCGCTACA
CTCAGTTGTACTCTTTCACGTTCAGCTCATAGATTTGGTCTCTGGTTGATCACTTTAATCACCTTCAGGGATGTG
GGTGGTTTCTGACCCGGAGTATAAGTGTGGGCCACCAGTGCACATCCCTGGATTATGTCGTCAATAGTAAACCCC
GGTGTTCACCATTTATCGGTTGGTTCCATTTACTTTGTTGTGTGCTTCGATCCATCTCGATTAGTGTGTTGACAT

ACTATCATCAAGTGAGGGTGCTGAGTGAACGCCCAAGCTGATACGATCTGTATAATCCAATCCCTAGTGTTCAAT
GGGAAATAATAAACCCTCTGCGATGAAGCTGAGTTGCAAATTCTTGGAATCATGTGCCTCAGTGGAGCCAGAACT
GTTGATTTGTTCACCCTCTTGGATTATAGAGGATGAACATCGCAAGTGTAGCCGACTTGGGCGGACCTCGGAAGA
AGAGAATGGCCGCTCACTCGTTGCTTACCACCCTTGAGAGCTAGTTATCGTGATAACGACGGGATTACCTACAGT
TCAAGACTCTTATGTAGGTTGCATTCACTGAACTCACCACCTTCGTTGAACAAAATCAGAGTCTTCTTTCACCGA
CGACCTGGGCGAATCCCTTGGAGTTTCACGGAGGCTGACATATGACGCAGTGACATGGATGTGACATACTTTGAT
GTTGCTCTTGCCCGTGCTTTTATGCGCACATCATAGAAGCCAGTCATGGGGATGGAGTTTATTCGGTACAAGCAT
AGGGTGCTTTGTACCTACTAAAACGTATTTTACGAATGCACAATCTATGCAAGAGCTGGTTTGTGATAACACATT
ATCACAAACGAGTGGTAGTGGACCTTCATTCTCTTGAATCGGAAGGGTGTTCCTGGATGAGATTCTTACTCAAAC
TTCGACCTTTCAGCGGAAGACTTTCGCTTGATCTTGCTGACCTCATTGCCTTTTACACTGAAGTGTATAGGTAGA
GTAATACCTTCAAATTTCACCTAGATTTGTTGTTAGTGACATCTGTTGTTGGTGTGTGTCGTGTGCCGCAGGTCA
TTTTTACTTGGCCTTGCAAAGTTCGGCAAAGGTGACTGGAGGAGCATATCTCGAAATTTTGTAATATCGCGTACA
CCAACGCAAGGTCCCGGTGCAACAACAGGGTTGGGTACCCCGGTCTATCCTCCAGGACAGATGGGTTATGGAGTG
AGAGGGCACATGGTGAGGCCCGGAATGGGAGGTCCCCCAATGAATATGACTCACATGACCTACAGTATGCCACAG
CCTGCCATGCATCATTAG

>MpDIV1.1 protein sequence

MAAPSPGPSPSSPSTASTPIPASAAAAAAGPSAAPVVPTPSAVSAPTIPVIPAPDVAPAVSSS**W**TSEQDKLFENA
LAVYDEESPNR**W**DNVASMVPGKDAADVMKH**Y**ELLTEDVTSIDAGRVALPSYILPGSLSGADAAGEQSDSSVSKNK
AWSGQSPGVSASGTSGTVGGLERKSSSSKADQERRKGIP**W**TEEEHRSFLLGLAKFGKGD**W**RSISRNFVISRTPTQ
VA**SHAQKYF**IRLNSINKDKRRSSIHDITSVNSAGVEVMQGSPGPITGQSPSGTSTSGQPLPHKTQPAFQGGMYVT
PVGPGATTGLGTPVYPPGQMGYGVRGHMVRPGMGGPPMNMTHMTYSMPQPAMHH*

>MpDIV1.2 protein sequence

MAAPSPGPSPSSPSTASTPIPASAAAAAAGPSAAPVVPTPSAVSAPTIPVIPAPDVAPAVSSS**W**TSEQDKLFENA
LAVYDEESPNR**W**DNVASMVPGKDAADVMKH**Y**ELLTEDVTSIDAGRVALPSYILPGSLSGADAAGEQSDSSVSKNK
AWSGQSPGVSASGTSGTVGGLERKSSSSKADQERRKGIP**W**TEEEHRSFLLGLAKFGKGD**W**RSISRNFVISRTPTQ
VA**SHAQKYF**IRLNSINKDKRRSSIHDITSVNSAGVEVMQGSPGPITGQSPSGTSTSGQPLPHKTQPAFQGICVR*

>Mp*div1* #69 protein sequence

MAAPSPGPSPSSPSTASTPIPASAAAAAAGPSAAPVVPTPSAVSAPTIPVIPAPDVAPAVSSS**W**TSEQDKLFENA
LAVYDEESPNR**W**DNVASMVPGKDAADVMKH**Y**ELLTEDVTSIDAGRVALPSYILPGSLSGADAAGEQSDSSVSKNK
AWSGQSPGVSASGTSGTVGGLERKSSSSKADQERRKGIP**W**TEEEHRSFLLGLAKFGKGD**W**RSRRTCKGPGATTGL
GTPVYPPGQMGYGVRGHMVRPGMGGPPMNMTHIPTVCHSLPCII

>Mp*div1* #87 protein sequence

MAAPSPGPSPSSPSTASTPIPASAAAAAAGPSAAPVVPTPSAVSAPTIPVIPAPDVAPAVSSS**W**TSEQDKLFENA
LAVYDEESPNR**W**DNVASMVPGKDAADVMKH**Y**ELLTEDVTSIDAGRVALPSYILPGSLSGADAAGEQSDSSVSKNK
AWSGQSPGVSASGTSGTVGGLERKSSSSKADQERRKGIP**W**TEEEHRSFLLGLAKFGKGD**W**RSISRNFVISRTPTQ
GPGATTGLGTPVYPPGQMGYGVRGHMVRPGMGGPPMNMTHMTYSMPQPAMHH-

Mp*DIV2*

>Mp*DIV2* genomic sequence

ATGGCAACAACCGTCGCCCAAACAGCGGGGTCTCCTGGACAGTTACCGATTCCTCCTCCTTGGACGCCAAAATTG
GACAAGTTGTTTGAGAAGGCGTTAGCCATCTATGACGGAGATTCTCCCGATCGATGGGAGAAGATTGCAGCCAAG
CTGCCTGGCGTCGATCCCACCGAGGTGAAGAAGCATTACGATAGGTTGATAGAAGATCTGAATTCAATAGAGACT
GGACGTGTGGCCCTGCCAAACTACCACAAGTCCGTCGGCCTAATGTCGTCTTCGGACGAGGATTCCCCCACCTCG
AGGAAGCCCGTTCACGGAGGACACCACGGTCTCGGTGGCTCCAATGCAAATGCCACGAATGCCACCAGTCTGCCT
TCGGGGAAGGCTCCTAGCTCCAAAGCCAGTGATCCCGAGCGGCGCAAAGGAATTCCGTGGTCCGAAGAAGAGCAC
AGGTTTGTCACTCTCCCGCTCGTTCTCCATTCTTTCGAGGCTCAAGGAGAGCCATTTGAACTGTACCTAAGGTTT
TATTCAAACACAATGAGGCACAAGGATCCTCCTGATTTCGAATCCGCTAACTGTCAGTGTCGAGACCTAAGTGAA
ACGAGGTCGGTCCGAGTGAGGTAGAGGACCCGTAGCTGTCTCAGAAAATGATTCGAGCTAGCCCGTGCCGGACTG

ACGATGATGCATCTGTGGGTATGGTATGCACAGATTATTTTTGCTAGGGCTTGCCAAATTCGGCAAGGGAGACTG
GCGAAGCATTTCGAGGAATTTTGTAGTATCGAGGACGCCGACGCAAGTGGCGAGCCACGCCCAAAAGTACTTCAT
TCGCCTGAACTCCATCAACAACAAGGACAAAAGGCGGTCCAGTATCCACGACATCACGAGCGTGAACGATGGAGA
CTCGCTTCCTCAGAGCCCAGGGCCCATCACGGGCCTGCCATCGCCGGGAGCTCAATGGCCAAGGTCAGGCCTACA
AGGGGCGAGTATGTATGATATGGGAGGTATGGGAGGACCCGACCAAGCGATTGGAGGACAAATGCTTATGACTCC
GACGGGCCATCCGCACCATGTACCTTACGGGCATGTGCCTGTGATGCAAGGGCCCCCAATGGCCATGCAGCACAT
GTCGTACCCGATGCCCCAGTCTGCGAACGCTTAA

>Mp*DIV2* coding sequence

ATGGCAACAACCGTCGCCCAAACAGCGGGGTCTCCTGGACAGTTACCGATTCCTCCTCCTTGGACGCCAAAATTG
GACAAGTTGTTTGAGAAGGCGTTAGCCATCTATGACGGAGATTCTCCCGATCGATGGGAGAAGATTGCAGCCAAG
CTGCCTGGCGTCGATCCCACCGAGGTGAAGAAGCATTACGATAGGTTGATAGAAGATCTGAATTCAATAGAGACT
GGACGTGTGGCCCTGCCAAACTACCACAAGTCCGTCGGCCTAATGTCGTCTTCGGACGAGGATTCCCCCACCTCG
AGGAAGCCCGTTCACGGAGGACACCACGGTCTCGGTGGCTCCAATGCAAATGCCACGAATGCCACCAGTCTGCCT
TCGGGGAAGGCTCCTAGCTCCAAAGCCAGTGATCCCGAGCGGCGCAAAGGAATTCCGTGGTCCGAAGAAGAGCAC
AGATTATTTTTGCTAGGGCTTGCCAAATTCGGCAAGGGAGACTGGCGAAGCATTTCGAGGAATTTTGTAGTATCG
AGGACGCCGACGCAAGTGGCGAGCCACGCCCAAAAGTACTTCATTCGCCTGAACTCCATCAACAACAAGGACAAA
AGGCGGTCCAGTATCCACGACATCACGAGCGTGAACGATGGAGACTCGCTTCCTCAGAGCCCAGGGCCCATCACG
GGCCTGCCATCGCCGGGAGCTCAATGGCCAAGGTCAGGCCTACAAGGGGCGAGTATGTATGATATGGGAGGTATG
GGAGGACCCGACCAAGCGATTGGAGGACAAATGCTTATGACTCCGACGGGCCATCCGCACCATGTACCTTACGGG
CATGTGCCTGTGATGCAAGGGCCCCCAATGGCCATGCAGCACATGTCGTACCCGATGCCCCAGTCTGCGAACGCT
TAA

>Mp*div2* #5 genomic sequence

ATGGCAACAACCGTCGCCCAAACAGCGGGGTCTCCTGGACAGTTACCGATTCCTCCTCCTTGGACGCCAAAATTG
GACAAGTTGTTTGAGAAGGCGTTAGCCATCTATGACGGAGATTCTCCCGATCGATGGGAGAAGATTGCAGCCAAG
CTGCCTGGCGTCGATCCCACCGAGGTGAAGAAGCATTACGATAGGTTGATAGAAGATCTGAATTCAATAGAGACT
GGACGTGTGGCCCTGCCAAACTACCACAAGTCCGTCGGCCTAATGTCGTCTTCGGACGAGGATTCCCCCACCTCG
AGGAAGCCCGTTCACGGAGGACACGGGCCTGCCATCGCCGGGAGCTCAATGGCCAAGGTCAGGCCTACAAGGGGC
GAGTATGTATGATATGGGAGGTATGGGAGGACCCGACCAAGCGATTGGAGGACAAATGCTTATGACTCCGACGGG
CCATCCGCACCATGTACCTTACGGGCATGTGCCTGTGATGCAAGGGCCCCCAATGGCCATGCAGCACATGTCGTA
CCCGATGCCCCAGTCTGCGAACGCTTAAAAC

>Mp*div2* #24 genomic sequence

ATGGCAACAACCGTCGCCCAAACAGCGGGGTCTCCTGGACAGTTACCGATTCCTCCTCCTTGGACGCCAAAATTG
GACAAGTTGTTTGAGAAGGCGTTAGCCATCTATGACGGAGATTCTCCCGATCGATGGGAGAAGATTGCAGCCAAG
CTGCCTGGCGTCGCTTCCTCAGAGCACTCGGGCCTGCCATCGCCGGGAGCTCAATGGCCAAGGTCAGGCCTACAA
GGGGCGAGTATGTATGATATGGGAGGTATGGGAGGACCCGACCAAGCGATTGGAGGACAAATGCTTATGACTCCG
ACGGGCCATCCGCACCATGTACCTTACGGGCATGTGCCTGTGATGCAAGGGCCCCCAATGGCCATGCAGCACATG
TCGTACCCGATGCCCCAGTCTGCGAACGCTTAA

>Mp*div2* #68 genomic sequence

ATGGCAACAACCGTCGCCCAAACAGCGGGGTCTCCTGGACAGTTACCGATTCCTCCTCCTTGGACGCCAAAATTG
GACAAGTTGTTTGAGAAGGCGTTAGCCATCTATGACGGAGATTCTCCCGATCGATGGGAGAAGATTGCAGCCAAG
CTGCCTGGCGTCGATCCCACCGAGGTGAAGAAGCATTACGATAGGTTGATAGAAGATCTGAATTCAATAGAGACT
GGACGTGTGGCCCTGCCAAACTACCACAAGTCCGTCGGCCTAATGTCGTCTTCGGACGAGGATTCCCCCACCTCG
AGGAAGCCCGTTCACGGAGGACAACGGGCCTGCCATCGCCGGGAGCTCAATGGCCAAGGTCAGGCCTACAAGGGG
CGAGTATGTATGATATGGGAGGTATGGGAGGACCCGACCAAGCGATTGGAGGACAAATGCTTATGACTCCGACGG
GCCATCCGCACCATGTACCTTACGGGCATGTGCCTGTGATGCAAGGGCCCCCAATGGCCATGCAGCACATGTCGT
ACCCGATGCCCCAGTCTGCGAACGCTTAAAAT

>Mp*div2* #93 genomic sequence

ATGGCAACAACCGTCGCCCAAACAGCGGGGTCTCCTGGACAGTTACCGATTCCTCCTCCTTGGACGCCAAAATTG
GACAAGTTGTTTGAGAAGGCGTTAGCCATCTATGACGGAGATTCTCCCGATCGATGGGAGAAGATTGCAGCCAAG
CTGCCTGGCGTCGATCCCACCGAGGTGAAGAAGCATTACGATAGGTTGATAGAAGATCTGAATTCAATAGAGACT

GGACGTGTGGCCCTGCCAAACTACCACAAGTCCGTCGGCCTAATGTCGTCTTCGGACGAGGATTCCCCCACCTCG
AGGAAGCCCGTTCACGGAGGACACCACGGTCTCGGTGGCTCCAATGCAAATGCCACGAATGCCACCAGTCTGCCT
TCGGGGAAGGCTCCTAGCTCCAAAGCCAGTGATCCCGAGCGGCGCAAAGGAATTCCGTGGTCCGAAGAAGAGCAC
AGGTTTGTCACTCTCCCGCTCGTTCTCCATTCTTTCGAGGCTCAAGGAGAGCCATTTGAACTGTACCTAAGGTTT
TATTCAAACACAATGAGGCACAAGGATCCTCCTGATTTCGAATCCGCTAACTGTCAGTGTCGAGACCTAAGTGAA
ACGAGGTCGGTCCGAGTGAGGTAGAGGACCCGTAGCTGTCTCAGAAAATGATTCGAGCTAGCCCGTGCCGGACTG
ACGATGATGCATCTGTGGGTATGGTATGCACAGATTATTTTTGCTAGGGCTTGCCAAATTCGGCAAGGGAGACTG
GCGAAGCATTTCGAGGAATTTTGTAGTATCGAGGACGCCGACGCAAGTGGCGAGCCACGCCCAAAAGTACTTCAT
TCGCCTGAACTCCATCAACAACAAGGACAAAAGGCGGTCCAGTATCCACGACATCACGAGCGTGAACGATGGAGA
CTCGCTTCCTCAGAGCCCAGGGCCCATCACGGGCCTGCCATCGCCGGGAGCTCAATGGCCAAGGTCAGGCCTTAC
GGGGCAAGTATGTTGGATATGTGCAGGTATGAGA**AGGTCTGATCTGGGAAGTTGTGGGACGAACGCATCTGACTA**
CTAGGGGGGACATCCGCTCATGATCCTTACGGGCATGCGCCTCGTGATGCATGAGCCGCAAGTGCCCGTGAAGCA
AGGTCCTCCCTGATCCCGCAGTCTGCGAACGCTTAA

>Mp*div2* #99 genomic sequence

ATGGCAACAACCGTCGCCCAAATACAAGGGGCGAGTATGTATGATATGGGAGGTATGGGAGGACCCGACCAAGCG
ATTGGAGGACAAATGCTTATGACTCCGACGGGCCATCCGCACCATGTACCTTACGGGCATGTGCCTGTGATGCAA
GGGCCCCCAATGGCCATGCAGCACATGTCGTACCCGATGCCCCAGTCTGCGAACGCTTAAA

>Mp*div2* #116 genomic sequence

ATGGCAACAACCGTCGCCCAAACAGCGGGGTCTCCTGGACAGTTACCGATTCCTCCTCCTTGGACGCCAAAATTG
GACAAGTTGTTTGAGAAGGCGTTAGCCATCTATGACGGAGATTCTCCCGATCGATGGGAGAAGATTGCAGCCAAG
CTGCCTGGCGTCGATCCCACCGAGGTGAAGAAGCATTACGATAGGTTGATAGAAGATCTGAATTCAATAGAGACT
GGACGTGTGGCCCTGCCAAACTACCACAAGTCCGTCGGCCTAATGTCGTCTTCGGACGAGGATTCCCCCACCTCG
AGGAAGCCCGTTCACGGAGGA**GG**CACGGGCCTGCCATCGCCGGGAGCTCAATGGCCAAGGTCAGGCCTACAAGGG
GCGAGTATGTATGATATGGGAGGTATGGGAGGACCCGACCAAGCGATTGGAGGACAAATGCTTATGACTCCGACG
GGCCATCCGCACCATGTACCTTACGGGCATGTGCCTGTGATGCAAGGGCCCCCAATGGCCATGCAGCACATGTCG
TACCCGATGCCCCAGTCTGCGAACGCTTAAAAC

>Mp*div2* #127 genomic sequence

ATGGCAACAACCGTCGCCCAAAACAAGGGGCGAGTATGTATGATATGGGAGGTATGGGAGGACCCGACCAAGCGA
TTGGAGGACAAATGCTTATGACTCCGACGGGCCATCCGCACCATGTACCTTACGGGCATGTGCCTGTGATGCAAG
GGCCCCCAATGGCCATGCAGCACATGTCGTACCCGATGCCCCAGTCTGCGAACGCTTAAA

>Mp*div2* #135 genomic sequence

ATGGCAACAACCGTCGCCCAAACAGCGGGGTCTCCTGGACAGTTACCGATTCCTCCTCCTTGGACGCCAAAATTG
GACAAGTTGTTTGAGAAGGCGTTAGCCATCTATGACGGAGATTCTCCCGATCGATGGGAGAAGATTGCAGCCAAG
CTGCCTGGCGTCGATCCCACCGAGGTGAAGAAGCATTACGATAGGTTGATAGAAGATCTGAATTCAATAGAGACT
GGACGTGTGGCCCTGCCAAACTACCACAAGTCCGTCGGCCTAATGTCGTCTTCGGACGAGGATTCCCCCACCTCG
AGGAAGCCCGTTCACGGAGGACACCCATCACGGGCCTGCCATCGCCGGGAGCTCAATGGCCAAGGTCAGGCCTAC
AAGGGGCGAGTATGTATGATATGGGAGGTATGGGAGGACCCGACCAAGCGATTGGAGGACAAATGCTTATGACTC
CGACGGGCCATCCGCACCATGTACCTTACGGGCATGTGCCTGTGATGCAAGGGCCCCCAATGGCCATGCAGCACA
TGTCGTACCCGATGCCCCAGTCTGCGAACGCTTAAAAAG

>Mp*div2* #169 genomic sequence

ATGGCAACAACCGTCGCCCAAACAGCGGGGTCTCCTGGACAGTTACCGATTCCTCCTCCTTGGACGCCAAAATTG
GACAAGTTGTTTGAGAAGGCGTTAGCCATCTATGACGGAGATTCTCCCGATCGATGGGAGAAGATTGCAGCCAAG
CTGCCTGGCGTCGATCCCACCGAGGTGAAGAAGCATTACGATAGGTTGATAGAAGATCTGAATTCAATAGAGACT
GGACGTGTGGCCCTGCCAAACTACCACAAGTCCGTCGGCCTAATGTCGTCTTCGGACGAGGATTCCCCCACCTCG
AGGAAGCCCGTTCACGGAGGACATCTGACATCACATCACGGACCGGACTGACGATGATGCATCTGTGGGTATGGT
ATGCACAGATTATTTTTGCTAGGGCTTGCCAAATTCGGCAAGGGAGACTGGCGAAGCATTTCGAGGAATTTTGTA
GTATCGAGGACGCCGACGCAAGTGGCGAGCCACGCCCAAAAGTACTTCATTCGCCTGAACTCCATCAACAACAAG
GACAAAAGGCGGTCCAGTATCCACGACATCACGAGCGTGAACGATGGAGACTCGCTTCCTCAGAGCCCATCACGG
GCCTGCCATCGCCGGGAGCTCAATGGCCAAGGTCAGGCCTACAAGGGGCGAGTATGTATGATATGGGAGGTATGG

GAGGACCCGACCAAGCGATTGGAGGACAAATGCTTATGACTCCGACGGGCCATCCGCACCATGTACCTTACGGGC
ATGTGCCTGTGATGCAAGGGCCCCCAATGGCCATGCAGCACATGTCGTACCCGATGCCCCAGTCTGCGAAACCGC
TTAAA

>Mp*div2* #189 genomic sequence

ATGGCAACAACCGTCGCCCAAATACAAGGGGCGAGTATGTATGATATGGGAGGTATGGGAGGACCCGACCAAGCG
ATTGGAGGACAAATGCTTATGACTCCGACGGGCCATCCGCACCATGTACCTTACGGGCATGTGCCTGTGATGCAA
GGCCCCTCAATGGCCATGCAGCATGTCGTACCCGATGCCCCAGTCTGCGAACGCTTAAA

>Mp*div2* #217 genomic sequence

ATGGCAACAACCGTCGCCCAAACAGCGGGGTCTCCTGGACAGTTACCGATTCCTCCTCCTTGGACGCCAAAATTG
GACAAGTTGTTTGAGAAGGCGTTAGCCATCTATGACGGAGATTCTCCCGATCGATGGGAGAAGATTGCAGCCAAG
CTGCCTGGCGTCGATCCCACCGAGGTGAAGAAGCATTACGATAGGTTGATAGAAGATCTGAATTCAATAGAGACT
GGACGTGTGGCCCTGCCAAACTACCACAAGTCCGTCGGCCTAATGTCGTCTTCGGACGAGGATTCCCCCACCTCG
AGGAAGCCCGTTCACGGAGGACACGGGCCTGCCATCGCCGGGAGCTCAATGGCCAAGGTCAGGCCTACAAGGGGC
GAGTATGTATGATATGGGAGGTATGGGAGGACCCGACCAAGCGATTGGAGGACAAATGCTTATGACTCCGACGGG
CCATCCGCACCATGTACCTTACGGGCATGTGCCTGTGATGCAAGGGCCCCCAATGGCCATGCAGCACATGTCGTA
CCCGATGCCCCAGTCTGCGAACGCTTAAAAC

>Mp*div2* #242 genomic sequence

ATGGCAACAACCGTCGCCCAAACAGCGGGGTCTCCTGGACAGTTACCGATTCCTCCTCCTTGGACGCCAAAATTG
GACAAGTTGTTTGAGAAGGCGTTAGCCATCTATGACGGAGATTCTCCCGATCGATGGGAGAAGATTGCAGCCAAG
CTGCCTGGCGTCGATCCCACCGAGGTGAAGAAGCATTACGATAGGTTGATAGAAGATCTGAATTCAATAGAGACT
GGACGTGTGGCCCTGCCAAACTACCACAAGTCCGTCGGCCTAATGTCGTCTTCGGACGAGGATTCCCCCACCTCG
AGGAAGCCCGTTCACGGAGGACAACGGGCCTGCCATCGCCGGGAGCTCAATGGCCAAGGTCAGGCCTACAAGGGG
CGAGTATGTATGATATGGGAGGTATGGGAGGACCCGACCAAGCGATTGGAGGACAAATGCTTATGACTCCGACGG
GCCATCCGCACCATGTACCTTACGGGCATGTGCCTGTGATGCAAGGGCCCCCAATGGCCATGCAGCACATGTCGT
ACCCGATGCCCCAGTCTGCGAACGCTTAAAATG

>Mp*div2* #257 genomic sequence

ATGGCAACAACCGTCGCCCAAACAGCGGGGTCTCCTGGACAGTTACCGATTCCTCCTCCTTGGACGCCAAAATTG
GACAAGTTGTTTGAGAAGGCGTTAGCCATCTATGACGGAGATTCTCCCGATCGATGGGAGAAGATTGCAGCCAAG
CTGCCTGGCGTCGATCCCACCGAGGTGAAGAAGCATTACGATAGGTTGATAGAAGATCTGAATTCAATAGAGACT
GGACGTGTGGCCCTGCCAAACTACCACAAGTCCGTCGGCCTAATGTCGTCTTCGGACGAGGATTCCCCCACCTCG
AGGAAGCCCGTTCACGGAGGACGGTCTCGGTGGCTCCAATGCAAATG**ACGGA**GCCCATCACGGGCCTGCCATCGC
CGGGAGCTCAATGGCCAAGGTCAGGCCTACAAGGGGCGAGTATGTATGATATGGGAGGTATGGGAGGACCCGACC
AAGCGATTGGAGGACAAATGCTTATGACTCCGACGGGCCATCCGCACCATGTACCTTACGGGCATGTGCCTGTGA
TGCAAGGGCCCCCAATGGCCATGCAGCACATGTCGTACCCGATGCCCCAGTCTGCGAACGCTTAAAA

>Mp*div2* #296 genomic sequence

ATGGCAACAACCGTCGCCCAAACAGCGGGGTCTCCTGGACAGTTACCGATTCCTCCTCCTTGGACGCCAAAATTG
GACAAGTTGTTTGAGAAGGCGTTAGCCATCTATGACGGAGATTCTCCCGATCGATGGGAGAAGATTGCAGCCAAG
CTGCCTGGCGTCGATCCCACCGAGGTGAAGAAGCATTACGATAGGTTGATAGAAGATCTGAATTCAATAGAGACT
GGACGTGTGGCCCTGCCAAACTACCACAAGTCCGTCGGCCTAATGTCGTCTTCGGACGAGGATTCCCCCACCTCG
A**ACGGTCCT**CCCGTAGCTGTCTCAGAAAATGATTCGAGCTAGCCCGTGCCGGACTGACGATGATGCATCTGTGGG
TATGGTATGCACAGATTATTTTTGCTAGGGCTTGCCAAATTCGGCAAGGGAGACTGGCGAAGCATTTCGAGGAAT
TTTGTAGTATCGAGGACGCCGACGCAAGTGGCGAGCCACGCCCAAAAGTACTTCATTCGCCTGAACTCCATCAAC
AACAAGGACAAAAGGCGGTCCAGTATCCACGACATCACGAGCGTGAACGATGGAGACTCGCTTCCTCAGAGCCCA
GGCCTCAGAGCCTCAGAGCCCATCACGGGCCTGCCATCGCCGGGAGCTCAATGGCCAAGGTCAGGCCTACAAGGG
GCGAGTATGTATGATATGGGAGGTATGGGAGGACCCGACCAAGCGATTGGAGGACAAATGCTTATGACTCCGACG
GGCCATCCGCACCATGTACCTTACGGGCATGTGCCTGTGATGCAAGGGCCCCCAATGGCCATGCAGCACATGTCG
TACCCGATGCCCCAGTCTGGAAAAGGCTTAAAAAAC

>Mp*div2* #305 genomic sequence

ATGGCAACAACCGTCGCCCAAATACAAGGGGCGAGTATGTATGATATGGGAGGTATGGGAGGACCCGACCAAGCG
ATTGGAGGACAAATGCTTATGACTCCGACGGGCCATCCGCACCATGTACCTTACGGGCATGTGCCTGTGATGCAA
GGGCCCCCAATGGCCATGCAGCACATGTCGTACCCGATGCCCCAGTCTGCGAACGCTTAA

>MpDIV2 protein sequence

MATTVAQTAGSPGQLPIPPP**W**TPKLDKLFEKALAIYDGDSPDR**W**EKIAAKLPGVDPTEVKKH**Y**DRLIEDLNSIET
GRVALPNYHKSVGLMSSSDEDSPTSRKPVHGGHHGLGGSNANATNATSLPSGKAPSSKASDPERRKGIP**W**SEEEH
RLFLLGLAKFGKGD**W**RSISRNFVVSRTPTQVA**SHAQKYF**IRLNSINNKDKRRSSIHDITSVNDGDSLPQSPGPIT
GLPSPGAQWPRSGLQGASMYDMGGMGGPDQAIGGQMLMTPTGHPHHVPYGHVPVMQGPPMAMQHMSYPMPQSANA
*

>Mpdiv2 #5 protein sequence

MATTVAQTAGSPGQLPIPPP**W**TPKLDKLFEKALAIYDGDSPDR**W**EKIAAKLPGVDPTEVKKH**Y**DRLIEDLNSIET
GRVALPNYHKSVGLMSSSDEDSPTSRKPVHGGHGPAIAGSSMAKVRPTRGEYVYGRYGRTRPSDWRTNAYDSDGP
SAPCTLRACACDARAPNGHAAHVVPDAPVCERLK

>Mpdiv2 #24 protein sequence

MATTVAQTAGSPGQLPIPPP**W**TPKLDKLFEKALAIYDGDSPDR**W**EKIAAKLPGVASSEHSGLPSPGAQWPRSGLQ
GASMYDMGGMGGPDQAIGGQMLMTPTGHPHHVPYGHVPVMQGPPMAMQHMSYPMPQSANA

>Mpdiv2 #68 protein sequence

MATTVAQTAGSPGQLPIPPP**W**TPKLDKLFEKALAIYDGDSPDR**W**EKIAAKLPGVDPTEVKKH**Y**DRLIEDLNSIET
GRVALPNYHKSVGLMSSSDEDSPTSRKPVHGGQRACHRRELNGQGQAYKGRVCMIWEVWEDPTKRLEDKCL-
LRRAIRTMYLTGMCL-CKGPQWPCSTCRTRCPSLRTLK

>Mpdiv2 #93 protein sequence

MATTVAQTAGSPGQLPIPPP**W**TPKLDKLFEKALAIYDGDSPDR**W**EKIAAKLPGVDPTEVKKH**Y**DRLIEDLNSIET
GRVALPNYHKSVGLMSSSDEDSPTSRKPVHGGHHGLGGSNANATNATSLPSGKAPSSKASDPERRKGIP**W**SEEEH
RLFLLGLAKFGKGD**W**RSISRNFVVSRTPTQVA**SHAQKYF**IRLNSINNKDKRRSSIHDITSVNDGDSLPQSPGPIT
GLPSPGAQWPRSGLTGQVCWICAGMRRSDLGSCGTNASDYGGHPLMILTGMRLVMHEPQVPVKQGPP-SRSLRTL

>Mpdiv2 #99 protein sequence

MATTVAQIQGASMYDMGGMGGPDQAIGGQMLMTPTGHPHHVPYGHVPVMQGPPMAMQHMSYPMPQSANA-

>Mpdiv2 #116 protein sequence

MATTVAQTAGSPGQLPIPPP**W**TPKLDKLFEKALAIYDGDSPDR**W**EKIAAKLPGVDPTEVKKH**Y**DRLIEDLNSIET
GRVALPNYHKSVGLMSSSDEDSPTSRKPVHGGGTGLPSPGAQWPRSGLQGASMYDMGGMGGPDQAIGGQMLMTPT
GHPHHVPYGHVPVMQGPPMAMQHMSYPMPQSANA-N

>Mpdiv2 #127 protein sequence

MATTVAQNKGRVCMIWEVWEDPTKRLEDKCL-LRRAIRTMYLTGMCL-CKGPQWPCSTCRTRCPSLRTLK

>Mpdiv2 #135 protein sequence

MATTVAQTAGSPGQLPIPPP**W**TPKLDKLFEKALAIYDGDSPDR**W**EKIAAKLPGVDPTEVKKH**Y**DRLIEDLNSIET
GRVALPNYHKSVGLMSSSDEDSPTSRKPVHGGHPSRACHRRELNGQGQAYKGRVCMIWEVWEDPTKRLEDKCL-
LRRAIRTMYLTGMCL-CKGPQWPCSTCRTRCPSLRTLKK

>Mpdiv2 #169 protein sequence

MATTVAQTAGSPGQLPIPPP**W**TPKLDKLFEKALAIYDGDSPDR**W**EKIAAKLPGVDPTEVKKH**Y**DRLIEDLNSIET
GRVALPNYHKSVGLMSSSDEDSPTSRKPVHGGHLTSHHGIIFARACQIRQGRLAKHFEEFCSIEDADASGEPRPK

VLHSPELHQQQGQKAVQYPRHHERERWRLASSEPITGLPSPGAQWPRSGLQGASMYDMGGMGGPDQAIGGQMLMT
PTGHPHHVPYGHVPVMQGPPMAMQHMSYPMPQSAKPLK

>Mpdiv2 #189 protein sequence

MATTVAQIQGASMYDMGGMGGPDQAIGGQMLMTPTGHPHHVPYGHVPVMQGPSMAMQHVVPDAPVCERL

>Mpdiv2 #217 protein sequence

MATTVAQTAGSPGQLPIPPP**W**TPKLDKLFEKALAIYDGDSPDR**W**EKIAAKLPGVDPTEVKKH**Y**DRLIEDLNSIET
GRVALPNYHKSVGLMSSSDEDSPTSRKPVHGGHGPAIAGSSMAKVRPTRGEYVYGRYGRTRPSDWRTNAYDSDGP
SAPCTLRACACDARAPNGHAAHVVPDAPVCERLK

>Mpdiv2 #242 protein sequence

MATTVAQTAGSPGQLPIPPP**W**TPKLDKLFEKALAIYDGDSPDR**W**EKIAAKLPGVDPTEVKKH**Y**DRLIEDLNSIET
GRVALPNYHKSVGLMSSSDEDSPTSRKPVHGGQRACHRRELNGQGQAYKGRVCMIWEVWEDPTKRLEDKCLLRRA
IRTMYLTGMCL-CKGPQWPCSTCRTRCPSLRTLKM

>Mpdiv2 #257 protein sequence

MATTVAQTAGSPGQLPIPPP**W**TPKLDKLFEKALAIYDGDSPDR**W**EKIAAKLPGVDPTEVKKH**Y**DRLIEDLNSIET
GRVALPNYHKSVGLMSSSDEDSPTSRKPVHGGRSRWLQCKRSPSRACHRRELNGQGQAYKGRVCMIWEVWEDPTK
RLEDKCL-LRRAIRTMYLTGMCL-CKGPQWPCSTCRTRCPSLRTLK

>Mpdiv2 #296 protein sequence

MATTVAQTAGSPGQLPIPPP**W**TPKLDKLFEKALAIYDGDSPDR**W**EKIAAKLPGVDPTEVKKH**Y**DRLIEDLNSIET
GRVALPNYHKSVGLMSSSDEDSPTSNGPIIFARACQIRQGRLAKHFEEFCSIEDADASGEPRPKVLHSPELHQQQ
GQKAVQYPRHHERERWRLASSEPRPQSLRAHHGPAIAGSSMAKVRPTRGEYVYGRYGRTRPSDWRTNAYDSDGPS
APCTLRACACDARAPNGHAAHVVPDAPVWKRLKK

>Mpdiv2 #305 protein sequence

MATTVAQIQGASMYDMGGMGGPDQAIGGQMLMTPTGHPHHVPYGHVPVMQGPPMAMQHMSYPMPQSANA-

Mp*DIVL*

>Mp*DIVL* genomic sequence

ATGACTCGCCGTTGCTCCCACTGTGGTCATAATGGGCACAATTCTCGTACGTGTCCCGATCGTTGTCCTGATCGG
GGGGTACGGTTATTTGGTGTTCGACTCACAACGAACGATGGTGCAGCGAACATGCGTAAGAGTGTCAGCATGGGC
AATTTGACGCATTATGCCAACGCCAGCAACCCTCCATCCACCCCCGAGCATTCGGAGTCCGGAGCTGGTGCTGAG
GGATACGTTTCAGACGGCCTCGTGCAGACTTCCAGCAACGCTCGCGAAAGAAAGAAAGGTGGGTTTCCATGGCCC
TTTTGATGTTTCAATTTCCAGCCAAGGACATTTCTCGAGAGTAGACGAAGAGCCTTCAACCTAAGACCGCCAGCA
TGTTTTCCCCTTCTTCTTGCATGAGAGAGTGATTCTTTGGAATCAGTTAATAGCCAGCGTGAGGTCGCCCTCATG
CACGGCCCTCGTAAAGGAAGCGGGGTGGTCATGGTTGGTTGGTGCTGCTGCTTTTGAGCACTCACCTGCATCGGA
AACCGAGCCCTTGCCCTAGATGAAGCGGTCCACTTCCAGTGAAAGCCACATTCACGTCTTTGCCGTGCATGATCG
TTTTGAAATTCTGACATGCCGTCCATGGCGGTTGAAGTCTGTTGTCCCTCTACTAACTCGTGTTCATTCCGCTCG
ACAGGTGTACCGTGGACTGAAGAAGAACACCGCCTCTTTCTTCTTGGGCTCCAGAAACTTGGTAAAGGTGATTGG
AGGGGCATATCCAGGAATTTTGTGCAGACTAGGACTCCTACTCAGGTGGCCAGTCACGCTCAAAAATATTTCATC
AGACAGAGTAACCTGAACAAAAGGAAGCGTCGTTCAAGCCTCTTCGACATTGTTACGGAGTCGGTGAGTCTTGGT
GCACTTTCTCTACGTAGATCTTGAGTTTGGGTGCTGTCCATGGATAGCTAGCCTTTCTTGGGTCAGTTGTTTTCG
CTCGTGAGAAATATATTATCTAGAGGTGAGCACTCGTCGATGCTGGCTTTCTTACATGGCGTGCGCTTCTCTTTA
TGTCCGCAGGGTCCGAGGCCAATGGTTGAAGAGCCCCTCTCAAAGAGCCCTGACATGTCTGCTCCCCTCCACCAA
CTCAGCTTGAGCTCCGGAGGATTGTTCAACCACGGTGCGTTCTTCGAACAACAGGGAATGCCGCAACTGAACAGA
GGGCCAGACATGTTTCCTCCGATGGCTCCCATGTCTGGAGTGTCGATGAGAAGTTATCCGATTGGAGCCCATCTC
CAACAAGCTATTCCTCTGCCAGTGATGTCTCTGGGAACGGCCGAGCAGACAGGTGAGTCGACTGTCAGCTACGCC
GGCGAATCCGTAACGGTACTACCGAGACCTGTGGCAATTTCTGCGGCCAGTGCTTCAGGTCCCGGCAGTGCAGCG

TTGGCGGGTCAGTTAGGAATCGGAGGACCAAGTATGTACCCATACCACATATGGCCAGGTATTGCACCTGGGAGC
TACAGCAAGCCCATGTCCTCGTCGCCGTCATCGAGTTCGAAGGTGGTGAAGCCAACTGCGTGTCCGGCCAAGATC
GATGAAGGATCCGAGATGGCGAAACTCACTCTCGGACCTTCACCGTCAGCGGAGCCGTCGAACCTCACTTTGAAG
TTCTCGGGCCACAGCGCGTTCGTCTCCAAGACCTCATTCAACAGCAACAATCTCGGTCCGAGCACTATTAGTGTG
GTGTAG

>Mp*DIVL* coding sequence

ATGACTCGCCGTTGCTCCCACTGTGGTCATAATGGGCACAATTCTCGTACGTGTCCCGATCGTTGTCCTGATCGG
GGGGTACGGTTATTTGGTGTTCGACTCACAACGAACGATGGTGCAGCGAACATGCGTAAGAGTGTCAGCATGGGC
AATTTGACGCATTATGCCAACGCCAGCAACCCTCCATCCACCCCCGAGCATTCGGAGTCCGGAGCTGGTGCTGAG
GGATACGTTTCAGACGGCCTCGTGCAGACTTCCAGCAACGCTCGCGAAAGAAAGAAAGGTGTACCGTGGACTGAA
GAAGAACACCGCCTCTTTCTTCTTGGGCTCCAGAAACTTGGTAAAGGTGATTGGAGGGGCATATCCAGGAATTTT
GTGCAGACTAGGACTCCTACTCAGGTGGCCAGTCACGCTCAAAAATATTTCATCAGACAGAGTAACCTGAACAAA
AGGAAGCGTCGTTCAAGCCTCTTCGACATTGTTACGGAGTCGGGTCCGAGGCCAATGGTTGAAGAGCCCCTCTCA
AAGAGCCCTGACATGTCTGCTCCCCTCCACCAACTCAGCTTGAGCTCCGGAGGATTGTTCAACCACGGTGCGTTC
TTCGAACAACAGGGAATGCCGCAACTGAACAGAGGGCCAGACATGTTTCCTCCGATGGCTCCCATGTCTGGAGTG
TCGATGAGAAGTTATCCGATTGGAGCCCATCTCCAACAAGCTATTCCTCTGCCAGTGATGTCTCTGGGAACGGCC
GAGCAGACAGGTGAGTCGACTGTCAGCTACGCCGGCGAATCCGTAACGGTACTACCGAGACCTGTGGCAATTTCT
GCGGCCAGTGCTTCAGGTCCCGGCAGTGCAGCGTTGGCGGGTCAGTTAGGAATCGGAGGACCAAGTATGTACCCA
TACCACATATGGCCAGGTATTGCACCTGGGAGCTACAGCAAGCCCATGTCCTCGTCGCCGTCATCGAGTTCGAAG
GTGGTGAAGCCAACTGCGTGTCCGGCCAAGATCGATGAAGGATCCGAGATGGCGAAACTCACTCTCGGACCTTCA
CCGTCAGCGGAGCCGTCGAACCTCACTTTGAAGTTCTCGGGCCACAGCGCGTTCGTCTCCAAGACCTCATTCAAC
AGCAACAATCTCGGTCCGAGCACTATTAGTGTGGTGTAG

>Mp*div*#26 genomic sequence

ATGACTCGCCGTTGCTCCCACTGTGGTCATAATGGGCACAATTCTCGTACGTGTCCCGATCGTTGTCCTGATCGG
GGGGTACGGTTATTTGGTGTTCGACTCACAACGAACGATGGTGCAGCGAACATGCGTAAGAGTGTCAGCATGGGC
AATTTGACGCATTATGCCAACGCCAGCAACCCTCCATCCACCCCCGAGCATTCGGAGTCCGGAGCTGGTGCTGAG
GGATACGTTTCAGACGGCCTCGTGCAGACTTCCAGCAACGCTCGCGAAAGAAAGAAAGGTGGGTTTCCATGGCCC
TTTTGATGTTTCAATTTCCAGCCAAGGACATTTCTCGAGAGTAGACGAAGAGCCTTCAACCTAAGACCGCCAGCA
TGTTTTCCCCTTCTTCTTGCATGAGAGAGTGATTCTTTGGAATCAGTTAATAGCCAGCGTGAGGTCGCCCTCATG
CACGGCCCTCGTAAAGGAAGCGGGGTGGTCATGGTTGGTTGGTGCTGCTGCTTTTGAGCACTCACCTGCATCGGA
AACCGAGCCCTTGCCCTAGATGAAGCGGTCCACTTCCAGTGAAAGCCACATTCACGTCTTTGCCGTGCATGATCG
TTTTGAAATTCTGACATGCCGTCCATGGCGGTTGAAGTCTGTTGTCCCTCTACTAACTCGTGTTCATTCCGCTCG
ACAGGTGTACCGTGGACTGAAGAAGAACACCGCCTCTTTCTTCTTGGGCTCCAGAAACTTGGTAAAGGTGATTGG
AGGGGCATATCCAGGAATTTTGTGCAGACTAGGACTCCTACTCAGGTGGCCAGTCACGCTCAAAAATATTTCATC
AGACAGAGTAACCTGAACAAAAGGAAGCGTCGTTCAAGCCTCTTCGACATTGTTACGGAGTCGGTGAGTCTTGGT
GCACTTTCTCTACGTAGATCTTGAGTTTGGGTGCTGTCCATGGATAGCTAGCCTTTCTTGGGTCAGTTGTTTTCG
CTCGTGAGAAATATATTATCTAGAGGTGAGCACTCGTCGATGCTGGCTTTCTTACATGGCGTGCGCTTCTCTTTA
TGTCCGCAGGGTCCGAGGCCAATGGTCGAAGAGCCCCTCTCAAAGAGCCCTGACAAGTTTGTTCCCCTCCACCAA
TTCAGCTTGAGCTCCGGAGGATTGTTCAACCACGGTGCGTTCTTCGAACAACAGGGAATGCCGCAACTGAACAGA
GGGCCGGACATGTTTCCTCCGATGGCTCCCATGTCGGGAGTTTCGATGAGAAGTTATCCGATTGGAGCCCATCTC
CAACAAGCTTATCCTCTGCCAGTGATGTCTCTGGGAACGGCCGAGCAGACAGGTGAGTCGACTGTCAGCTACGCC
GGCGAATCCGTAA**TTTCCACAGGTCTCGGTAATTAATTGCCACAGGTCTCGGTAATACGGATTACTACCGAGACC**
TGTGGCAATTTCTGCGGCCAGTGCTTCAGGTCCCGGCAGTGCAGCGTTGGCGGGTCAGTTAGGAATCGGAGGACC
AAGTATGTACCCATACCACATATGGCCAGGTATTGCACCTGGGAGCTACAGCAAGCCCATGTCCTCGTCGCCGTG
ATCGAGTTCGAAGGTGGTGAAGCCAACTGCGTGTCCGGCCAAGATCGATGAAGGATCCGAGATGGCGAAACTCAC
TCTCGGACCTTCACCTTCACCGTCTCGGTAGTAATCCGTATTACCGAGACCTGTGGCAATTAATCGTAATCGTAA
CCGTAATCTCGGTAGTAATCCGTAATCTCGGTAGTAATCCGTACTACCGAGACCTGTGGCAATTTCTGCGGCCAG
TGCTTCAGGTCCCGGCAGTGCAGCGTTGGCGGGTCAGTTAGGAATCGGAGGACCAAGTATGTACCCATACCACAT
ATGGCCAGGTATTGCACCTGGGAGCTACAGCAAGCCCATGTCCTCGTCGCCGTCATCGAGTTCGAAGGTGGTGAA
GCCAACTGCGTGTCCGGCCAAGATCGATGAAGGATCCGAGATGGCGAAACTCACTCTCGGACCTTCACCGTCAGC
GGAGCCGTCGAACCTCACTTTGAAGTTCTCGGGCCACAGCGCGTTCGTCTCCAAGACCTCATTCAACAGCAACAA
TCTCGGTCCGAGCACTATTAGTGTGGTGTAG

>Mp*divl*#73 genomic sequence

ATGACTCGCCGTTGCTCCCACTGTGGTCATAATGGGCACAATTCTCGTACGTGTCCCGATCGTTGTCCTGATCGG
GGGGTACGGTTATTTGGTGTTCGACTCACAACGAACGATGGTGCAGCGAACATGCGTAAGAGTGTCAGCATGGGC
AATTTGACGCATTATGCCAACGCCAGCAACCCTCCATCCACCCCCGAGCATTCGGAGTCCGGAGCTGGTGCTGAG
GGATACGTTTCAGACGGCCTCGTGCAGACTTCCAGCAACGCTCGCGAAAGAAAGAAAGGTGGGTTTCCATGGCCC
TTTTGATGTTTCAATTTCCAGCCAAGGACATTTCTCGAGAGTAGACGAAGAGCCTTCAACCTAAGACCGCCAGCA
TGTTTTCCCCTTCTTCTTGCATGAGAGAGTGATTCTTTGGAATCAGTTAATAGCCAGCGTGAGGTCGCCCTCATG
CACGGCCCTCGTAAAGGAAGCGGGGTGGTCATGGTTGGTTGGTGCTGCTGCTTTTGAGCACTCACCTGCATCGGA
AACCGAGCCCTTGCCCTAGATGAAGCGGTCCACTTCCAGTGAAAGCCACATTCACGTCTTTGCCGTGCATGATCG
TTTTGAAATTCTGACATGCCGTCCATGGCGGTTGAAGTCTGTTGTCCCTCTACTAACTCGTGTTCATTCCGCTCG
ACAGGTGTACCGTGGACTGAAGAAGAACACCGCCTCTTTCTTCTTGGGCTCCAGAAACTTGGTAAAGGTGATTGG
AGGGGCATATCCAGGAATTTTGTGCAGACTAGGACTCCTACTCAGGTGGCCAGTCACGCTCAAAAATATTTCATC
AGACAGAGTAACCTGAACAAAAGGAAGCGTCGTTCAAGCCTCTTCGACATTGTTACGGAGTCGTGGGGTAGTTGT
TTTCGCTCGTGAGAATATATTATCTAGAGGTGAGCACTCGTCGATGCTGGCTTTCTTACATGGCGTGCGCTTCTC
TTTATGTCCGCAGGGTCCGAGGCCAATGGTCGAAGAGCCCCTCTCAAAGAGCCCTGACATGTCTGCTCCCCTCCA
CCAACTCAGCTTGAGCTCCGGAGGATTGTTCAACCACGGTGCGTTCTTCGAACAACAGGGAATGCCGCAACTGAA
CAGAGGGCCAGACATGTTTCCTCCGATGGCTCCCATGTCGGGAGTGTCGATGAGAAGTTATCCGATTGGAGCCCA
TCTCCAACAAGCTATTCCTCTTCCAGTGATGTCTCTGGGAACGGCCGAGCAGACAGGTGAGTCGACTGTCAGCTA
CGCCGGCGAATCCGTAAC**CGAGAAATCGGAAATTACGGATTCGACCTGAAGCACTGGCCGCAGAAATTGCCACAG**
GTCTCGGTAGTACCGAGAAATCAGAAATTAGAAATTATTAATCTCGGTACTACCGAGACCTGTGGCAATTTCTGC
GGCCAGTGCTTCAGGTCCCGGCAGTGCAGCGTTGGCGGGTCAGTTAGGAATCGGAGGACCAAGTATGTACCCATA
CCACATATGGCCAGGTATTGCACCTGGGAGCTACAGCAAGCCCATGTCCTCGTCGCCGTCATCGAGTTCGAAGGT
GGTGAAGCCAACTGCGTGTCCGGCCAAGATCGATGAAGGATCCGAGATGGCGAAACTCACTCTCGGACCTTCACC
GTCAGCGGAGCCGTCGAACCTCACTTTGAAGTTCTCGGGCCACAGCGCGTTCGTCTCCAAGACCTCATTCAACAG
CAACAATCTCGGTCCGAGCACTATTAGTGTGGTGTAG

>Mp*divl*#87 genomic sequence

ATGACTCGCCGTTGCTCCCACTGTGGTCAGGTACTACCGAGACCTGTGGCAATTTCTGCGGCCAGTGCTTCAGGT
CCCGGCAGTGCAGCGTTGGCGGGTCAGTTAGGAATCGGAGGACCAAGTATGTACCCATACCACATATGGCCAGGT
ATTGCACCTGGGAGCTACAGCAAGCCCATGTCCTCGTCGCCGTCATCGAGTTCGAAGGTGGTGAAGCCAACTGCG
TGTCCGGCCAAGATCGATGAAGGATCCGAGATGGCGAAACTCACTCTCGGACCTTCACCGTCAGCGGAGCCGTCG
AACCTCACTTTGAAGTTCTCGGGCCACAGCGCGTTCGTCTCCAAGACCTCATTCAACAGCAACAATCTCGGTCCG
AGCACTATTAGTGTGGTGTAG

>MpDIVL protein sequence

MTRRCSHCGHNGHNSRTCPDRCPDRGV**RLFGV**RLTTNDGAANMRKSVSMGNLTHYANASNPPSTPEHSESGAGAE
GYVSDGLVQTSSNARERKKGVP**W**TEEEHRLFLLGLQKLGKGD**W**RGISRNFVQTRTPTQVA**SHAQKYF**IRQSNLNK
RKRRSSLFDIVTESGPRPMVEEPLSKSPDMSAPLHQLSLSSGGLFNHGAFFEQQGMPQLNRGPDMFPPMAPMSGV
SMRSYPIGAHLQQAIPLPVMSLGTAEQTGESTVSYAGESVTVLPRPVAISAASASGPGSAALAGQLGIGGPSMYP
YHIWPGIAPGSYSKPMSSSPSSSSKVVKPTACPAKIDEGSEMAKLTLGPSPSAEPSNLTLKFSGHSAFVSKTSFN
SNNLGPSTISVV*

>Mpdivl #26 protein sequence

MTRRCSHCGHNGHNSRTCPDRCPDRGV**RLFGV**RLTTNDGAANMRKSVSMGNLTHYANASNPPSTPEHSESGAGAE
GYVSDGLVQTSSNARERKKGVP**W**TEEEHRLFLLGLQKLGKGD**W**RGISRNFVQTRTPTQVA**SHAQKYF**IRQSNLNK
RKRRSSLFDIVTESGPRPMVEEPLSKSPDKFVPLHQFSLSSGGLFNHGAFFEQQGMPQLNRGPDMFPPMAPMSGV
SMRSYPIGAHLQQAYPLPVMSLGTAEQTGESTVSYAGESVISTGLGNLPQVSVIRITTETCGNFCGQCFRSRQCS
VGGSVRNRRTKYVPIPHMARYCTWELQQAHVLVAVIEFEGGEANCVSGQDR-
RIRDGETHSRTFTFTVSVVIRITETCGN-S-S-P-SR-
SVISVVIRTTETCGNFCGQCFRSRQCSVGGSVRNRRTKYVPIPHMARYCTWELQQAHVLVAVIEFEGGEANCVSG
QDR-RIRDGETHSRTFTVSGAVEPHFEVLGPQRVRLQDLINSNNLGPSTISVV-

>Mpdivl #73 protein sequence

MTRRCSHCGHNGHNSRTCPDRCPDRGV**RLFGV**RLTTNDGAANMRKSVSMGNLTHYANASNPPSTPEHSESGAGAE
GYVSDGLVQTSSNARERKKGVP**W**TEEEHRLFLLGLQKLGKGD**W**RGISRNFVQTRTPTQVA**SHAQKYF**IRQSNLNK
RKRRSSLFDIVTESGPRPMVEEPLSKSPDMSAPLHQLSLSSGGLFNHGAFFEQQGMPQLNRGPDMFPPMAPMSGV
SMRSYPIGAHLQQAIPLPVMSLGTAEQTGESTVSYAGESVTEKSEITDST-STGRRNCHRSR-YREIRN-

>Mpdivl #87 protein sequence

MTRRCSHCGQVLPRPVAISAASASGPGSAALAGQLGIGGPSMYPYHIWPGIAPGSYSKPMSSSPSSSSKVVKPTA
CPAKIDEGSEMAKLTLGPSPSAEPSNLTLKFSGHSAFVSKTSSTATISVRALLVWC

Mp*DRIF*

>Mp*DRIF.1* genomic sequence

ATGGCGGGCTCCGTCGGAAACAATTCCACCACGAATTCTAGTGCTGCAGCAACCTCTGCGTCGCCGGCCGTCAAT
GGGAATCATAGCAGTATGTATAACAGCAATGCACAGGGAGCGTCGACTCAGTCGACAACCACGACGATAAACAGT
GGCAACAATGGTATTAGTCGACCAAACGGCCCCGCCACTAATGGAAGCGGGAATGGGACCAACAGTGTCGCTTCG
GACCAACCCCCTCCTCTTCAGCTCCAGTTGTTACATGACCCAGGGATCACAGCGGATTGGAGCTCTGAGGAGCAG
GCCACTCTCGATGATGGACTGACAAAGTGAGTTATTTTAACCATCTGTATTGGTTCCTGAAACTACTAGAACTCT
CAAAATCCGGCGCTTCTGAGCTCTCGATGAACTCTATACAATCGTCCTTGCCCGGTTCCAAGATTTGGAATGCCG
TGTTCTACGTCCTCTATCTGGTTGCACTTCTCAAGGGCGAATCATATCGAAAATAAATAACTATATTTGAGAGAG
AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGGCTGACCTAATTGT
TGCATGATATCACATCCAGTTTAGAGGATGTGATATGATCTGGAGCTCCCATACCATATCAGCTTTCGACATCTT
ACTGGCCTATTTTTCAATGTGAGATTGTATTTAAATGATTCGTAGGCACGGAAGGTACCAAGCTCAAATCTATCC
AGTCAGTGGTAGGAATAGGTGTTTGTTTCTGTTAACAGTCAGTATAAACATGTTGTGATATGCCGATGTAGCGTA
GCTCTTAGCTGGTTGTAAAAGCACGAAGCATATCAGAGGACCCTGCTCCAGGACAATAAACTTCAACAGCTCGTC
GCATTTCTCCTTCTCTCCTCGTCACAGGTCTAGATAGTGTATCTTAAAATAGTTTCATGTTAGAGAGGCTTACCT
TATTCAGGTTCTTAGCGTTGTCTCTCATACGGCTGCGTCAGTTGAGGTTCCGTGGAGTTGTCGTAAAGTTGATTC
ATACAACTTATTGATAGAGGCATGTTTGTGATTAGGTTTGCCGGAGAAACAAGCAATCTGGCTAAATACATCAAA
ATAGCCAATCTGTTACCGGAGAAGACAGTCCGAGATGTCGCCATGCGTTGCAGATGGATGACGGTATGTTCCCGC
TTTCACTAACTTTTGCGTTTTAAGTTTCTTAGTCGAGTTCTGTTGAATCAGTTTATAATAAGTACCGTCGTGGAG
GATTTCCAAGGGGAAATGATGTATCAGATGTTTCTCGCTTGGCAATTATTTTAATATTTAGACCAGTGAACTTAG
GATCGTACTGATGACGGCAATGCTTTGTGGTACAGAAAAAGGAGATTGGCAAGAGGCGTAAACCTGAGGATCAAA
ATGCAAGCAAGAAGAATAAAGACAAAAAGGTTAAATAACAACTCCACGCCCGTGTATGCCGACAGTAGCCTTCCA
CTGTTCTCCGGTTGTACTGACTAACATTAATATGGTTCAATGGATTCAGGACAAATCCGACTCGATGGAGAGAGA
GAGAGATCCGGGCTGACCTAATTGTTGCATGATATCACATCCAGTTTAGAGGATGTGATATGATCTGGAGCTCCC
ATACCATATCAGCTTTCGACATCTTACTGGCCTATTTTTCAATGTGAGATTGTATTTAAATGATTCGTAGGCACG
GAAGGTACCAAGCTCAAATCTATCCAGTCAGTGGTAGGAATAGGTGTTTGTTTCTGTTAACAGTCAGTATAAACA
TGTTGTGATATGCCGATGTAGCGTAGCTCTTAGCTGGTTGTAAAAGCACGAAGCATATCAGAGGACCCTGCTCCA
GGACAATAAACTTCAACAGCTCGTCGCATTTCTCCTTCTCTCCTCGTCACAGGTCTAGATAGTGTATCTTAAAAT
AGTTTCATGTTAGAGAGGCTTACCTTATTCAGGTTCTTAGCGTTGTCTCTCATACGGCTGCGTCAGTTGAGGTTC
CGTGGAGTTGTCGTAAAGTTGATTCATACAACTTATTGATAGAGGCATGTTTGTGATTAGGTTTGCCGGAGAAAC
AAGCAATCTGGCTAAATACATCAAAATAGCCAATCTGTTACCGGAGAAGACAGTCCGAGATGTCGCCATGCGTTG
CAGATGGATGACGGTATGTTCCCGCTTTCACTAACTTTTGCGTTTTAAGTTTCTTAGTCGAGTTCTGTTGAATCA
GTTTATAATAAGTACCGTCGTGGAGGATTTCCAAGGGGAAATGATGTATCAGATGTTTCTCGCTTGGCAATTATT
TTAATATTTAGACCAGTGAACTTAGGATCGTACTGATGACGGCAATGCTTTGTGGTACAGAAAAAGGAGATTGGC
AAGAGGCGTAAACCTGAGGATCAAAATGCAAGCAAGAAGAATAAAGACAAAAAGGTTAAATAACAACTCCACGCC
CGTGTATGCCGACAGTAGCCTTCCACTGTTCTCCGGTTGTACTGACTAACATTAATATGGTTCAATGGATTCAGG
ACAAATCCGACTCGATGTCTACCAAGGCTCCCACTGGTCATATCAGACCCGGTTTGTCTAGCTACACCGCTCCTA
CACCGAACGTCGACAACGACGACGGCATCTCCAACGACGGTTGGTTACTTTGCTCTCTTATTCGATCTTAGCTCT
AGTCTGCCGTGCTTGTCCACTTGGCTATACTATGTCCTTCACTTTTCTGATTTTTCATGTCTCCCTTATGTCCAG
GAGTGAAACCTAACCGTTGAGCTCCTCCTCTTTAAATTATATACTTTCCTGGCGCAAGCTTCAGGGTATCGGTTC
TAGTCTGGCCAGAACACCACTTAGATCGTGCTTACTCATATCTCTCCTCCTTGTTTGACCGGGAGGTTTTGCGTC
CTGCCAGTGCTTGTGCCGATGCCTAGGTTGTGATGTTACTAATTCGATTTTGTCGTCGCACTGAATGTACAGCAA
TTGGAGGCACAACTGGCCAGCTGTTGGAGCAAAATTCTCACGTGATATTGCAAATCAGGTCAAATCTTGCTGCCA

TGAAGGTACGTGGAAGATTGAAGTAGAGAGTCCTCTGGTCACCTAAAGCAAGTCGAGGTCCTCTGGAATTCAAAC
TATCTCTGTTTATGAACGCTGTCTTTCTCAAAGCCGGTTGTATGTTGTAAATTTGGATGTGCAGCTGCAAGAGAA
CACCGAGCTTCTCGTAAGATTCAGGGATAACATCTGTGCCATCTTAAATGGGTGAGTTAGAACAACTTGACCGGC
GGACTTTGAATCCATTAAATACAGTTGATTTCCAGCGCAGGCAGTTGGCTTTGTTATTTTATCATTTGATCTTTT
GTTCTAACCTGGTGTTTTGCATATTCAGGATGACAAACATGCCGGGGATTATGAGTCAAATGCCACCTCTCCCTG
TGAAACTCAATACCGAATTGGCAGATACGATCCTTCCCAAATCTTTACCACAAGCGTCTCCAACTTCACAAACGT
AA

>Mp*DRIF.2* genomic sequence

ATGTTTGTGATTAGGTTTGCCGGAGAAACAAGCAATCTGGCTAAATACATCAAAATAGCCAATCTGTTACCGGAG
AAGACAGTCCGAGATGTCGCCATGCGTTGCAGATGGATGACGGTATGTTCCCGCTTTCACTAACTTTTGCGTTTT
AAGTTTCTTAGTCGAGTTCTGTTGAATCAGTTTATAATAAGTACCGTCGTGGAGGATTTCCAAGGGGAAATGATG
TATCAGATGTTTCTCGCTTGGCAATTATTTTAATATTTAGACCAGTGAACTTAGGATCGTACTGATGACGGCAAT
GCTTTGTGGTACAGAAAAAGGAGATTGGCAAGAGGCGTAAACCTGAGGATCAAAATGCAAGCAAGAAGAATAAAG
ACAAAAAGGTTAAATAACAACTCCACGCCCGTGTATGCCGACAGTAGCCTTCCACTGTTCTCCGGTTGTACTGAC
TAACATTAATATGGTTCAATGGATTCAGGACAAATCCGACTCGATGTCTACCAAGGCTCCCACTGGTCATATCAG
ACCCGGTTTGTCTAGCTACACCGCTCCTACACCGAACGTCGACAACGACGACGGCATCTCCAACGACGGTTGGTT
ACTTTGCTCTCTTATTCGATCTTAGCTCTAGTCTGCCGTGCTTGTCCACTTGGCTATACTATGTCCTTCACTTTT
CTGATTTTTCATGTCTCCCTTATGTCCAGGAGTGAAACCTAACCGTTGAGCTCCTCCTCTTTAAATTATATACTT
TCCTGGCGCAAGCTTCAGGGTATCGGTTCTAGTCTGGCCAGAACACCACTTAGATCGTGCTTACTCATATCTCTC
CTCCTTGTTTGACCGGGAGGTTTTGCGTCCTGCCAGTGCTTGTGCCGATGCCTAGGTTGTGATGTTACTAATTCG
ATTTTGTCGTCGCACTGAATGTACAGCAATTGGAGGCACAACTGGCCAGCTGTTGGAGCAAAATTCTCACGTGAT
ATTGCAAATCAGGTCAAATCTTGCTGCCATGAAGGTACGTGGAAGATTGAAGTAGAGAGTCCTCTGGTCACCTAA
AGCAAGTCGAGGTCCTCTGGAATTCAAACTATCTCTGTTTATGAACGCTGTCTTTCTCAAAGCCGGTTGTATGTT
GTAAATTTGGATGTGCAGCTGCAAGAGAACACCGAGCTTCTCGTAAGATTCAGGGATAACATCTGTGCCATCTTA
AATGGGTGAGTTAGAACAACTTGACCGGCGGACTTTGAATCCATTAAATACAGTTGATTTCCAGCGCAGGCAGTT
GGCTTTGTTATTTTATCATTTGATCTTTTGTTCTAACCTGGTGTTTTGCATATTCAGGATGACAAACATGCCGGG
GATTATGAGTCAAATGCCACCTCTCCCTGTGAAACTCAATACCGAATTGGCAGATACGATCCTTCCCAAATCTTT
ACCACAAGCGTCTCCAACTTCACAAACGTAA

>Mp*DRIF.1* coding sequence

ATGGCGGGCTCCGTCGGAAACAATTCCACCACGAATTCTAGTGCTGCAGCAACCTCTGCGTCGCCGGCCGTCAAT
GGGAATCATAGCAGTATGTATAACAGCAATGCACAGGGAGCGTCGACTCAGTCGACAACCACGACGATAAACAGT
GGCAACAATGGTATTAGTCGACCAAACGGCCCCGCCACTAATGGAAGCGGGAATGGGACCAACAGTGTCGCTTCG
GACCAACCCCCTCCTCTTCAGCTCCAGTTGTTACATGACCCAGGGATCACAGCGGATTGGAGCTCTGAGGAGCAG
GCCACTCTCGATGATGGACTGACAAAGTTTGCCGGAGAAACAAGCAATCTGGCTAAATACATCAAAATAGCCAAT
CTGTTACCGGAGAAGACAGTCCGAGATGTCGCCATGCGTTGCAGATGGATGACGAAAAAGGAGATTGGCAAGAGG
CGTAAACCTGAGGATCAAAATGCAAGCAAGAAGAATAAAGACAAAAAGGACAAATCCGACTCGATGTCTACCAAG
GCTCCCACTGGTCATATCAGACCCGGTTTGTCTAGCTACACCGCTCCTACACCGAACGTCGACAACGACGACGGC
ATCTCCAACGACGCAATTGGAGGCACAACTGGCCAGCTGTTGGAGCAAAATTCTCACGTGATATTGCAAATCAGG
TCAAATCTTGCTGCCATGAAGCTGCAAGAGAACACCGAGCTTCTCGTAAGATTCAGGGATAACATCTGTGCCATC
TTAAATGGGATGACAAACATGCCGGGGATTATGAGTCAAATGCCACCTCTCCCTGTGAAACTCAATACCGAATTG
GCAGATACGATCCTTCCCAAATCTTTACCACAAGCGTCTCCAACTTCACAAACGTAA

>Mp*DRIF.2* coding sequence

ATGTTTGTGATTAGGTTTGCCGGAGAAACAAGCAATCTGGCTAAATACATCAAAATAGCCAATCTGTTACCGGAG
AAGACAGTCCGAGATGTCGCCATGCGTTGCAGATGGATGACGAAAAAGGAGATTGGCAAGAGGCGTAAACCTGAG
GATCAAAATGCAAGCAAGAAGAATAAAGACAAAAAGGACAAATCCGACTCGATGTCTACCAAGGCTCCCACTGGT
CATATCAGACCCGGTTTGTCTAGCTACACCGCTCCTACACCGAACGTCGACAACGACGACGGCATCTCCAACGAC
GCAATTGGAGGCACAACTGGCCAGCTGTTGGAGCAAAATTCTCACGTGATATTGCAAATCAGGTCAAATCTTGCT
GCCATGAAGCTGCAAGAGAACACCGAGCTTCTCGTAAGATTCAGGGATAACATCTGTGCCATCTTAAATGGGATG
ACAAACATGCCGGGGATTATGAGTCAAATGCCACCTCTCCCTGTGAAACTCAATACCGAATTGGCAGATACGATC
CTTCCCAAATCTTTACCACAAGCGTCTCCAACTTCACAAACGTAA

>MpDRIF.1 protein sequence

MAGSVGNNSTTNSSAAATSASPAVNGNHSSMYNSNAQGASTQSTTTTINSGNNGISRPNGPATNGSGNGTNSVAS
DQPPPLQLQLLHDPGITAD**W**SSEEQATLDDGLTKFAGETSNLAK**Y**IKIANLLPEKTVRDVAMRCR**W**MTKKEIGKR
RKPEDQNASKKNKDKKDKSDSMSTKAPTGHIRPGLSSYTAPTPNVDNDDGISNDAIGGTTGQLLEQNSHVILQIR
SNLA**AMKLQENTELLVRFRDNICAILNGMTNMPGIMSQ**MPPLPVKLNTELADTILPKSLPQASPTSQT*

>MpDRIF.2 protein sequence

MFVIRFAGETSNLAKYIKIANLLPEKTVRDVAMRCR**W**MTKKEIGKRRKPEDQNASKKNKDKKDKSDSMSTKAPTG
HIRPGLSS**Y**TAPTPNVDNDDGISNDAIGGTTGQLLEQNSHVILQIRSNLA**AMKLQENTELLVRFRDNICAILNGM
TNMPGIMSQ**MPPLPVKLNTELADTILPKSLPQASPTSQT*

Annex E: Promoter sequences

The predicted promoter sequences have the ≈1000 bp of the promoter sequence available in the *Phytozome* underlined at grey and the 5'UTR underlined at green.

> *pro*Mp*DIV1*

AGAACAGAAAGACTCGATCCCTTCGGCACTCCGCTCACCCTCACCTGCTTTCCTCGCCTTCTCAGAAAAAAAAAATACAGCAA
ACGTACATACAAAAGCGATCCGTCAGTGCCGATGGAAAAGCTTCTCAAGAAGGAAAGAGTAAGTGGCTGACTGACTGAATCAA
TCAATCAATCAATCAATCAATCGATCCATTCATCAATCAAAAGCGCAGCGCAGCATAGCACAGAGCTGCCACCCACCTCGCGC
ACAGAGCGAGTCCTTTAGGTCAGGGCGGAGGGTTTTAGCGAGAGAGAGGAGAGCAGATTCATGACCGTCGAACCGAACAGGAC
GAGAGAATAAAAGCCCTAATTTGGCAAAACCTGGCACTCATTCGACCGACGACATGCTCGCACCAGCCGGTCAGGGGAGTGCG
GGGGGTCATTCGTCGGGGCGACGCGGGGCCGCGAGTGGCCCATGGCACGCCCATCGCATTCAATCGCGCCCGGTGGGGCCGAC
CTCCGCAGCGGACGTACGACGTACCTACTCTAGGGCCGGTCTGACTGTGCGACCTGGTGTGACCCCAGTGCCCGTCCCAGGCT
CTGGCACCGTACGTAGTACCGCCTGTAGAAGTAGAAGCAGAAGGATCGGGACTCGGAGGAGGTGCGCTACGACCCGTGCAATA
CAGTCCGGCCCCGCCCGAGTAGCGACGGGCGGACGGGCGGGCTGGCGGGCGGGCTAACTAACTACTAACTACAGTGTACCGTT
GCAGAGGTGAGCTGAGGCTGAGGCTGCTGAGCGGTGCAGCGCCAAGCTGATCGGTTGCAGAAAGAGGCATCCACTTTCCGAAG
GCCGGAAAGCTCCTTGCTGGCCTGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCGAGGGACGGGCAGGGGCAGGGGCAGGGCA
GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGGATAAGCATGGGGCGCCTTATCCCGCAGATGATATTCACGAACCGCAGG
CTCCTCTTTCGGGGGCTGGATAAACTTCCGCCAGGTCCGGAGGTTTGGCTGGGGCTCGGCCAGTCGAGGAGCCCCCGGCGAGA
GAAGGCGAAGAAGGCTGTGGAGGGAAAAGGCGGGTGATTTTCCGCTGCTCGTTTCACTGACTTCTTGCTTGTGGGTGATGTTG
GCGGGGGCCGCGGGCTGCCAATTGTTGTGGTCGAGAGCAGACAGGCGGTGCGCGAGTGGGTGGGCGCGCTCCTGTTGCCTCCG
GACAAAGATCCTTGAGCGCGAGCACGAGAGAGAGAGAGGTGTGGAGTCGAGAGCTGCCTGCGGCAGGCGGTGCCTTGGTTGCT
CGGCTCGGCTCGGCTCGGAATGTGGAAGGCGAGACGAGAAGGCCCGTCAGCATCGAGACGAGGTTGTTGGTGCGGGATCGCGG
GAGTGATCGAATGAGAGGCCAAGGGAGCCGATGACAGTTCGCAGGTCGTATCGGGAACTGAATGGTGGGATCGGTGGGTAGAG
GCGAGGCGAGGAGAGTATGTTGCACGAGGTGGTGGGTACTTTGTGTGTGCTTCTGGTGATCGTCTACACGCTGCACCGATTGC
GCTTCTCTCTGCCTTGCAGCAGCAGCTCCTTATGGACTGACATGCGAGCATTATGGAAGGGACGCGATGATCTGGGACGAGAC
TCACATGCTGGCCAAAGTCTCGTGTAGCTTTCGGTGACATCGTGTGAGTCTGTTTTCTGGTGCCTGTCTGAGTGGTGCGG

> *pro*Mp*DIV2*

GGCGAAATATATAAATCTCCGAACATAAAGTGCACGGGCGCCCGCGAGAAAACAAAACCCCAAGGCCGCGCTTGCAAGCCTCC
AACATCAAGCCAGCTCTAAATCAAGGCCCGCGAGACGCGCGCGAGAGCGAGACGAGGGGCGACCTGCCGAGCTAAATGCCGAT
GGACTTAAAGGCCGGCTTTCCGACACTTGAATCCAGGCCCTCCGATGAGCCATCCGACGGCATTGTCGCCAGATTCTTGACAT
ATTCAAACGCTCTCGAGAGCACGCCGTGGGACGGATATAGCTCGTTGCGGAGGATCGCCGAGAAAAAAAATGGCAATATGCGAC
AAGACCGCGCCCGATCTCCGATCCGACGGAGGGGAGCGCGCGACCGTGCTCGGACGCCATTGGCCGGCCCAAAGAGGAAAATA
TTTAGGCCCGTATCGCTGCCCCCCCCGGCCCAGCCGAGCCGAGCCTGAGCTCTGCCTGCGAGTGGGCGGGCCGGGAGGATAGT
CCGGTAGTGCACTGTTAGCAGTAATGAAGCGAGGATATCCGACAAGTCGCTGACGCCGCTGGTCGCGATCGCTCGACCGATCG
GTGGCTGATACACAGAGCAAGGATCCGGAGGAAAGTTTAGGAGAATGACAGCCGAGCCCGTCGAGGACCGTTTAGGAGAAGGG
CCGAAGGAAGGCCGGCGGGCGGGCGAGAGAGAGAGAGAGAGAGAGAGAGATGACGAGGCTGGTATTTACTGTGTCGTAGGATA
GCCCGCCCGCCCGGTCGCTCGGGGCTTGCCCTCTCGCCGGGCCTGGTAGTGCACTGTGAAGGACACGGCGGGCGGGCGGGCAA
GGGTGTTGGTTATTAGTGCCCTGCCCTTCGCTTTCGCGAGCCTGCCTACTTGTCTGCCCGCCCGCCCCCGCCCGCCCCCGCAG
TCCGGTTGCGACTGCCGAAGGGGACCGCGGGCGAGAGCGAGCGAGCCCATAAAGCCAGAGCCCTCCACATTGCTCAGCCCCTT
TGCCGCACCGAGAGCTCTCTCTCACACCTGCTTCCGCCCGAGGCCCGTTTCTCTCCTCTCCTTTCATCTCCTCTCTCTCCT
CGAGCAGCAGCAGCAGCTGCTGCTGCAGCGCAGAGATCTCCTCTTCAGAGTTCCTGTGCATCCAACAGCACCGCCTTTCGCCC
CTTTCCATTCGCTTTCCGCCGCTGCTTCCTTCTCTTCTCGTTCGTACCGACTGCGTCTGCTGCCGCGGGGTTCCGTTCAGCTG
CGACTGTCACTTCCACAGCTCGTTCCACCATCTTCCTGCCCCGCCTCGCGAGCGAGCGAGTATTTTTCTATATTTTTCATCCG
GAAGACCGTCCCGTCCGGGCCTTTAAGCTGCGACATCTTTCAGACCAGCTCGAACGGGTTCCTCAATTCTCGTTCATCTCGGC
CGCGGGATTGATCGATTGATCGATTGACCGAGGGGTGGTGGAGTTGCCGAGTTGGAGTCGAAGAAAGAAGGTTTTTATGGGTG
GGAGGTCATTTGATTTTGAGCCCGCCGTTTGCCAACGAGGCGCCGAGTGAAGCAATCCCGTCGATTGCCGAGCCTCATTTCAC
CGTCCCGCAATCGAATTTCTTCCGATTTTAAATCGCTGGGTGCTGGGCTTGAGGTGAATCGGAGATCACTG

> *pro*Mp*DIVL*

AAACTGCCTAATCACCGGGAAAAAATCGAAGATGCGGGTCCTTCCCACAGTGGTTACCCCCCTCGGCCACGCACCTAACGAGC
TAACGGTTCGGAATTAATTTGCATGCAGCGAACAGGACGCATCGACTTAAATTAATGAGCCCCGGAGTTACCAACTAAGCTAC
CGAATTTAAACTCGCGGTGGATGAGAAGGACCAGTGGGGAACTCGACGACGTGAGGTCGAGGAGGTACCGTCACGGACTAAAT
AAAGTCGAACCATTTTATAATTTTCGCAACGTCGCAGTTTCGCCATGTCCCTCTCGGAGTAACCATCGCGCTCGCTCTTCCTT

ACCGCGGACGGCTTGTAGGTTGACCGACCGATACGCAAAAGGAACCGTAGTGCAGTGTAGTTACCTGGTAGGGATTATTTGGA
AGCGGACCCGACGGTGACGGATGGGGGTAGCGCATTGGACTTATTGGAGATATAGTTGCCGGGCGCAGTCCTGTTAGCTCACC
GGCATCCATCACGAGAAGGAGAAGGTTGCACACCGTTCAAAACTCCACGGTGGAGGAGCGGGAGGAGGAGACAAAGCCGTGGG
CGAGTGAAAGGGAGAGGCCGAAAAGGTGAAAGCCGCGGGAGTCATGAATGGGAATATTCGCGGGCGCGGGGGAGAGGCGGACA
CGAGGGCGCGCCGCGGTTAAACAGAAGACGGTAATAGTGACAGCGCGAGCTAACCACGCGGTTGGTTGGTTGGTTAAGTGTGA
AGGAAGGACAGTCTGTACCTGCCCCCGCTTCAGCCCAGCAGGACGGAGGAGGAAGAGGAGGAGGAGGAGGAAGAGGATGGTGG
TGGGCTGAAACGAACGGATCAGGATGGATGCATGGATGCATGGCGGATGCGGGCGCGGGTCAGGCATGTCCGGCACCGTCCGT
TGCCTGTCATCTGGTCGCTGGCCATCCACACATCCGGGCTCAACAGGCTGCCGACCTCCGCTCAGCTCCCATTCTTCCCCGCC
CCATCATCATCGTCCGCTCTCCCCTTCTTCCTCTCGCCCTCGACCGATCCAGTTTCATTCGTTCGCTCTCTCGCCCTCCTCCC
TTCGTACCCTCGCCCTCCTCCTCCTCCCCTCCTCCTCCTTCATAAGGCCCTCCCTCCCGTCGTGTGTGCGCGGAACGCTCACT
CTGCTTGTGCACTGCTGTCGTAGTCGCTGCTGTGCAGGGCCCAGTTTGCGCGTCGAGTCGAGTAGTAGTTGTTCTCTGTTCAA
CCATTCCTTCTCTATCTCTCTCTTTCTTTCTCCTCCTCGATCTCTCTCTCTCTCTCTCTCGCGCGCGCTCTCTCTCTTTCTCG
GTCTCTTTGCTTGCGCGATAGCATCTTGCTTCGATCTGTAATTTCTCTTGACATCTTTCTCGTTCGATACCATCATTATTATC
AACCCACAACACCATCACCACCAACCATCCAGTCGTGGGCTTGGGCCGTCGTGTCGTGACGAGCTCGGTCGGTCGGGTTCGTT
CGAGGTCTGTGATAAAAGATTCTTCCGGCTCTCGAGATCGTTCGAAGGTCGCGCGAGTGAGTCTTGCTACGACTCCTGGTGGT
CAGATCCGTGGGGTTTGAGTCGAGGCTTGGACAAAGAGAAGCGTAGTTGAGGCGTCGGTAGTGGAGGTTGGAGGTTATCAGAG
CGTGGACGGGAGGAAGAGGACTCACTCAATAGGTTGCAAGGATATCGGGAAAGGAATAAAGGAGGAAGAGAAGACGAGATAGG
AAGAGAACGAACGAAGCCCGGTCCTGTCGTGGTCTGCTGTCGGTGGGGCTTCGAGTCTTGTTCGTTCGGTCGGTCGGTTGCTC
GAAAGGCGGTAAGGAAGTGCAACAAAAGGGAGAAAGGAAAGGAAAAGTAGCGACGAGTAGACGAGGGAGATAGCGCGAGAAGG
AAGTCCCGCTCCGCATCGTTTTCTTGTGCTGTCTTGAGAGAGGGAGCATCGGTTGTGATCGGTTTCTGGCTTCGGGCTTTATA
CCACGTGGTTGCTTCTTCGAGTTCTCTCTCAGATCCTACGTCTTTCTTGTCCTGCGACTTGGAAGCCACGGATTTCGGTTTCA
TCGGCGGGCACGTGTTTGCTCTCTTTCTCTCCCTCTCTCACTGCTCCAAGGGGCATATGCTCTCGCCTGCTGTCCTCCGTCCT
CGTGGTGACATCTAATTCGAATCGGGTTCCTCTCGGTGTCCCACTCTTCGCGGATCTGCTCTCTCGTTGCCAGTGTGAACAGG
CGCCGGTGCCCTCATCCATCTGCGTTCACAACAGACGACTCGTGAACCGGAAGAGAACGGAGCCGCACCGACCTTGTGCAGTT
CGCTCCTCGTGCTCTACATCACACGATTTCTCCATCGTCTTCCCAGTCAGTAAAAATAGACCCATTCAAGTTACCACCGGTCG
CTGTGGCTAACTTACGTATTCCTTCGGGGCGATCAGGGCCAGAAGAACTGGAGACCGAA

> *pro*Mp*DRIF.1*

GTTAAAGGCTAAAATTTGAGATGTCCCGACTCGTGATAGATTCCTGTGATGAAATGCTGGCGAGCCCATTTCTGAGGGAGGTT
GAATTAACTTATTTTCACGGTTTTTTTCCTCTAGAAATAGCCCTGGAGGACTCTACAGCCCATACAAAATGTGGATATTTCGT
CGGGCACATTTCACCTCACGAGCAGTGTCGGGCCGCTAAGTTTGAGGTCACACCTCTACCCTACGACAGACAAAGATGCTCTT
ATCCACACCAAATATCTGAAAGTGGCCTTGGACACCAAAATGGCCCCCGAAAAATCTAAGCACCTCCAACAATTCGAATGGTT
CGGAAATGTGGCGCGACCGATCCGATAATGCGAATGCGGAAGACCTGGGCACCCCACATCTCATCCGGAACCCGATTTTTCTC
TCTCGCCGCTTAGTCGAATTAATGCGCCCATACTCCTCCGCCTGCGGCCCCGAAGGAACCGAGCTCGCTCTCACTCCGTCCCT
CGGGGCGTAAAGTCCGTAAGACGGGAGGAGTTTTAGGTCGGGCAGAGAGAGAGAAGGCCGAGCTAGATCCGGTGGGGACTGAA
CATGAGGCAGAATTTGGGCAGCATGTTCTTATACGTGATTGCAGGAGAAGTGAGACCGACGCTTATTTTAAGGTTTGTTGGAA
AGTCTCGAGGTTCGCTGCCCTGCCCTGCGTGAGACTTAACGAAGGCGGCCGGGCAGTTTACATTTCGCAAATCGGAGGTGGAA
GTCCGTCCAGAAGAGGTACAGAACCTCGCGGCGTAAAGAATCGAAAGAAACCCAACACGGAACCCGAGGCCGGAGCAGGGCAG
CGCAGCGCAGCGTTGGGCAGGGCAGTGAAGGGGAGGGGAGGGGAGGGGAGGAGGAGGCGGAGGGCAAGGCCGGTGCTCGTGCT
GCTGCTGCTGCCGGTGTGGGGTGGGGTTGGGGTATGTATATAGGCCGGGAAATCTCGGGTCACATCTTTTGCTCTCGTCCTTC
GCTGCAGGTCTTCGACCACAGCAACGACAACAACAACAACAAGGACAGCCAGCAGCAGACGCTGCTCTGAGCTGCTCTGCTCT
GCTCGGCCACCCTTCGCTTTCGCTCGTTTCGCTCCGCTTGGCTCGGCTGCTGCTGCTGATGATGCTGCAGTTGCTGGTGGCGT
TCGAATGAAGGATCGCGACGACGGCTGTTCTGCGGGCCTGGCATCGACGTCCGCCGCGGCATCATCATCGCACCCGCGCGGGG
TGTAATTTTGGAGGCGTTGCGGCTGCTGCTGCAGCAAGCAGCTGAGGTCCGAGCTGCGCGCGTTGGCGGATAGATAGATTTGA
GGAGAGGCGTTTGATTTGGAAGGGGTGTGGTGCTCGCTGTGAAGTCGGACGACAGCTGCCGAGCTGCCGCCTCGTCGTTGGCT
CGGCCTTGCGGTTTCAGGCGTAGGCAGACGCCGACATAGAGGGGTGTTTTTTTGGGGGCCTTTTGAGCATGTGGATAATAGAG
TCGGAAGGTTGAAGGCGAAAGAGCCGGAATGGAGGCGACGCGACGACGAACGAGGTAGAAAGAGCTGGAGACGTCGGACGAGA
AGGGTTTTTTTGTCATGCCCGGATGTTCGCAATGCGCGCATCTCGTGACGAGCTGTTAGCGTTTAGAGCAACCCCGTGCAGAG
GCGGAGGTCGAGACAGGAGAAGAGTGGGAGGGTTGTAGCGTTTGATCTCTCGGGAGCTAAATCTTGGAGTGAATCGAAACGAG
GTGGACCGATTTGGGTGAATGTTGTGAGGACAGTGCCGCGCGAACTGTGCAACTTCCTCATCGTGAATCCAACTATTTGGGGA
GAAAATTGGGATGTCGATGCACGCCCGCTCGTGTTAGCCAGTAGAGTGGCTTTCTCTGGCATCGAGGGAGGTTCTAAGGAGAG
GATGAAAGAGGAATGGGACCAGAGTAAACGTGAGTGATTGGCCCGCTCAGA

> *pro*Mp*DRIF.2*

GTTAAAGGCTAAAATTTGAGATGTCCCGACTCGTGATAGATTCCTGTGATGAAATGCTGGCGAGCCCATTTCTGAGGGAGGTT
GAATTAACTTATTTTCACGGTTTTTTTCCTCTAGAAATAGCCCTGGAGGACTCTACAGCCCATACAAAATGTGGATATTTCGT
CGGGCACATTTCACCTCACGAGCAGTGTCGGGCCGCTAAGTTTGAGGTCACACCTCTACCCTACGACAGACAAAGATGCTCTT
ATCCACACCAAATATCTGAAAGTGGCCTTGGACACCAAAATGGCCCCCGAAAAATCTAAGCACCTCCAACAATTCGAATGGTT
CGGAAATGTGGCGCGACCGATCCGATAATGCGAATGCGGAAGACCTGGGCACCCCACATCTCATCCGGAACCCGATTTTTCTC

TCTCGCCGCTTAGTCGAATTAATGCGCCCATACTCCTCCGCCTGCGGCCCCGAAGGAACCGAGCTCGCTCTCACTCCGTCCCT
CGGGGCGTAAAGTCCGTAAGACGGGAGGAGTTTTAGGTCGGGCAGAGAGAGAGAAGGCCGAGCTAGATCCGGTGGGGACTGAA
CATGAGGCAGAATTTGGGCAGCATGTTCTTATACGTGATTGCAGGAGAAGTGAGACCGACGCTTATTTTAAGGTTTGTTGGAA
AGTCTCGAGGTTCGCTGCCCTGCCCTGCGTGAGACTTAACGAAGGCGGCCGGGCAGTTTACATTTCGCAAATCGGAGGTGGAA
GTCCGTCCAGAAGAGGTACAGAACCTCGCGGCGTAAAGAATCGAAAGAAACCCAACACGGAACCCGAGGCCGGAGCAGGGCAG
CGCAGCGCAGCGTTGGGCAGGGCAGTGAAGGGGAGGGGAGGGGAGGGGAGGAGGAGGCGGAGGGCAAGGCCGGTGCTCGTGCT
GCTGCTGCTGCCGGTGTGGTTAAAGGCTAAAATTTGAGATGTCCCGACTCGTGATAGATTCCTGTGATGAAATGCTGGCGAGC
CCATTTCTGAGGGAGGTTGAATTAACTTATTTTCACGGTTTTTTTCCTCTAGAAATAGCCCTGGAGGACTCTACAGCCCATAC
AAAATGTGGATATTTCGTCGGGCACATTTCACCTCACGAGCAGTGTCGGGCCGCTAAGTTTGAGGTCACACCTCTACCCTACG
ACAGACAAAGATGCTCTTATCCACACCAAATATCTGAAAGTGGCCTTGGACACCAAAATGGCCCCCGAAAAATCTAAGCACCT
CCAACAATTCGAATGGTTCGGAAATGTGGCGCGACCGATCCGATAATGCGAATGCGGAAGACCTGGGCACCCCACATCTCATC
CGGAACCCGATTTTTCTCTCTCGCCGCTTAGTCGAATTAATGCGCCCATACTCCTCCGCCTGCGGCCCCGAAGGAACCGAGCT
CGCTCTCACTCCGTCCCTCGGGGCGTAAAGTCCGTAAGACGGGAGGAGTTTTAGGTCGGGCAGAGAGAGAGAAGGCCGAGCTA
GATCCGGTGGGGACTGAACATGAGGCAGAATTTGGGCAGCATGTTCTTATACGTGATTGCAGGAGAAGTGAGACCGACGCTTA
TTTTAAGGTTTGTTGGAAAGTCTCGAGGTTCGCTGCCCTGCCCTGCGTGAGACTTAACGAAGGCGGCCGGGCAGTTTACATTT
CGCAAATCGGAGGTGGAAGTCCGTCCAGAAGAGGTACAGAACCTCGCGGCGTAAAGAATCGAAAGAAACCCAACACGGAACCC
GAGGCCGGAGCAGGGCAGCGCAGCGCAGCGTTGGGCAGGGCAGTGAAGGGGAGGGGAGGGGAGGGGAGGAGGAGGCGGAGGGC
AAGGCCGGTGCTCGTGCTGCTGCTGCCGGTGTGGGGTGGGGTTGGGGTATGTATATAGGCCGGGAAATCTCGGGTCACAT
CTTTTGCTCTCGTCCTTCGCTGCAGGTCTTCGACCACAGCAACGACAACAACAACAACAAGGACAGCCAGCAGCAGACGCTGC
TCTGAGCTGCTCTGCTCTGCTCGGCCACCCTTCGCTTTCGCTCGTTTCGCTCCGCTTGGCTCGGCTGCTGCTGCTGATGATGC
TGCAGTTGCTGGTGGCGTTCGAATGAAGGATCGCGACGACGGCTGTTCTGCGGGCCTGGCATCGACGTCCGCCGCGGCATCAT
CATCGCACCCGCGCGGGGTGTAATTTTGGAGGCGTTGCGGCTGCTGCTGCAGCAAGCAGCTGAGGTCCGAGCTGCGCGCGTTG
GCGGATAGATAGATTTGAGGAGAGGCGTTTGATTTGGAAGGGGTGTGGTGCTCGCTGTGAAGTCGGACGACAGCTGCCGAGCT
GCCGCCTCGTCGTTGGCTCGGCCTTGCGGTTTCAGGCGTAGGCAGACGCCGACATAGAGGGGTGTTTTTTTGGGGGCCTTTTG
AGCATGTGGATAATAGAGTCGGAAGGTTGAAGGCGAAAGAGCCGGAATGGAGGCGACGCGACGACGAACGAGGTAGAAAGAGC
TGGAGACGTCGGACGAGAAGGGTTTTTTTGTCATGCCCGGATGTTCGCAATGCGCGCATCTCGTGACGAGCTGTTAGCGTTTA
GAGCAACCCCGTGCAGAGGCGGAGGTCGAGACAGGAGAAGAGTGGGAGGGTTGTAGCGTTTGATCTCTCGGGAGCTAAATCTT
GGAGTGAATCGAAACGAGGTGGACCGATTTGGGTGAATGTTGTGAGGACAGTGCCGCGCGAACTGTGCAACTTCCTCATCGTG
AATCCAACTATTTGGGGAGAAAATTGGGATGTCGATGCACGCCCGCTCGTGTTAGCCAGTAGAGTGGCTTTCTCTGGCATCGA
GGGAGGTTCTAAGGAGAGGATGAAAGAGGAATGGGACCAGAGTAAACGTGAGTGATTGGCCCGCTCAGAATGGCGGGCTCCGT
CGGAAACAATTCCACCACGAATTCTAGTGCTGCAGCAACCTCTGCGTCGCCGGCCGTCAATGGGAATCATAGCAGTATGTATA
ACAGCAATGCACAGGGAGCGTCGACTCAGTCGACAACCACGACGATAAACAGTGGCAACAATGGTATTAGTCGACCAAACGGC
CCCGCCACTAATGGAAGCGGGAATGGGACCAACAGTGTCGCTTCGGACCAACCCCCTCCTCTTCAGCTCCAGTTGTTACATGA
CCCAGGGATCACAGCGGATTGGAGCTCTGAGGAGCAGGCCACTCTCGATGATGGACTGACAAAGTGAGTTATTTTAACCATCT
GTATTGGTTCCTGAAACTACTAGAACTCTCAAAATCCGGCGCTTCTGAGCTCTCGATGAACTCTATACAATCGTCCTTGCCCG
GTTCCAAGATTTGGAATGCCGTGTTCTACGTCCTCTATCTGGTTGCACTTCTCAAGGGCGAATCATATCGAAAATAAATAACT
ATATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGGCTGACCTA
ATTGTTGCATGATATCACATCCAGTTTAGAGGATGTGATATGATCTGGAGCTCCCATACCATATCAGCTTTCGACATCTTACT
GGCCTATTTTTCAATGTGAGATTGTATTTAAATGATTCGTAGGCACGGAAGGTACCAAGCTCAAATCTATCCAGTCAGTGGTA
GGAATAGGTGTTTGTTTCTGTTAACAGTCAGTATAAACATGTTGTGATATGCCGATGTAGCGTAGCTCTTAGCTGGTTGTAAA
AGCACGAAGCATATCAGAGGACCCTGCTCCAGGACAATAAACTTCAACAGCTCGTCGCATTTCTCCTTCTCTCCTCGTCACAG
GTCTAGATAGTGTATCTTAAAATAGTTTCATGTTAGAGAGGCTTACCTTATTCAGGTTCTTAGCGTTGTCTCTCATACGGCTG
CGTCAGTTGAGGTTCCGTGGAGTTGTCGTAAAGTTGATTCATACAACTTATTGATAGAGGCATGTTTGTGATTAGGTTTGCCG
GAGAAACAAGCAATCTGGCTAAATACATCAAAATAGCCAATCTGTTACCGGAGAAGACAGTCCGAGATGTCGCCATGCGTTGC
AGATGGATGACGGTATGTTCCCGCTTTCACTAACTTTTGCGTTTTAAGTTTCTTAGTCGAGTTCTGTTGAATCAGTTTATAAT
AAGTACCGTCGTGGAGGATTTCCAAGGGGAAATGATGTATCAGATGTTTCTCGCTTGGCAATTATTTTAATATTTAGACCAGT
GAACTTAGGATCGTACTGATGACGGCAATGCTTTGTGGTACAGAAAAAGGAGATTGGCAAGAGGCGTAAACCTGAGGATCAAA
ATGCAAGCAAGAAGAATAAAGACAAAAAGGTTAAATAACAACTCCACGCCCGTGTATGCCGACAGTAGCCTTCCACTGTTCTC
CGGTTGTACTGACTAACATTAATATGGTTCAATGGATTCAGGACAAATCCGACTCGATGGAGAGAGAGAGAGATCCGGGCTGA
CCTAATTGTTGCATGATATCACATCCAGTTTAGAGGATGTGATATGATCTGGAGCTCCCATACCATATCAGCTTTCGACATCT
TACTGGCCTATTTTTCAATGTGAGATTGTATTTAAATGATTCGTAGGCACGGAAGGTACCAAGCTCAAATCTATCCAGTCAGT
GGTAGGAATAGGTGTTTGTTTCTGTTAACAGTCAGTATAAACATGTTGTGATATGCCGATGTAGCGTAGCTCTTAGCTGGTTG
TAAAAGCACGAAGCATATCAGAGGACCCTGCTCCAGGACAATAAACTTCAACAGCTCGTCGCATTTCTCCTTCTCTCCTCGTC
ACAGGTCTAGATAGTGTATCTTAAAATAGTTTCATGTTAGAGAGGCTTACCTTATTCAGGTTCTTAGCGTTGTCTCTCATACG
GCTGCGTCAGTTGAGGTTCCGTGGAGTTGTCGTAAAGTTGATTCATACAACTTATTGATAGAGGC

Annex F: gRNAs cloning into pMpGe_En04 and pMpGE010/11

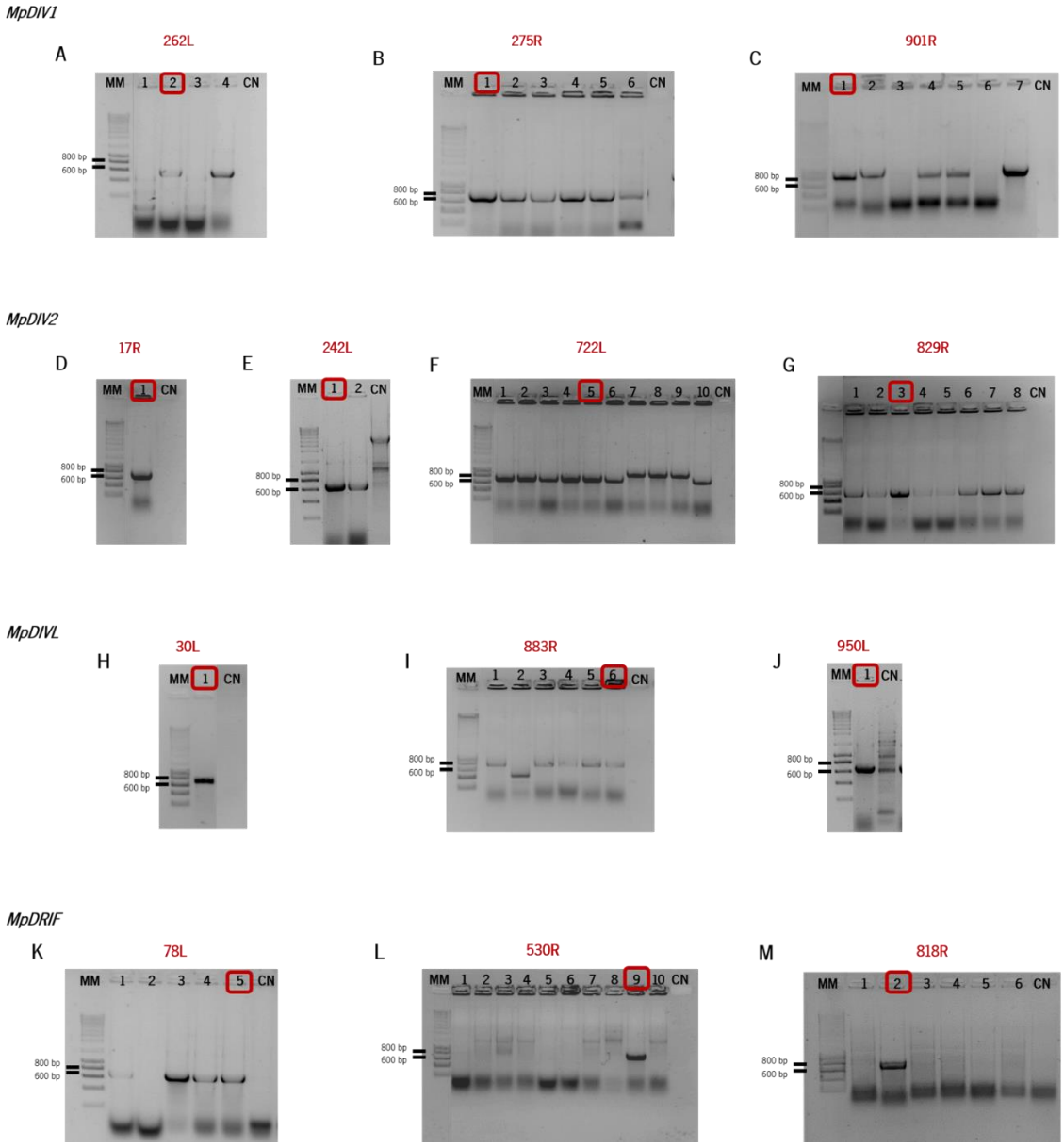

Figure F1: PCR screening of *E. coli* colonies containing constructs with the gRNAs in pMpGE_En04 vector for CRISPR/Cas9 system.
A) Mp*DIV1* 262LpMpGE_En04 colonies, lanes 1 to 4 (expected size fragment of 690 bp);
B) Mp*DIV1* 275RpMpGE_En04 colonies, lanes 1 to 6 (expected size fragment of 690 bp);
C) Mp*DIV1* 901RpMpGE_En04 colonies, lanes 1 to 7 (expected size fragment of 690 bp);
D) Mp*DIV2* 17RpMpGE_En04 colonies, lane 1 (expected size fragment of 690 bp);
E) Mp*DIV2* 242LpMpGE_En04 colonies, lanes 1 to 2 (expected size fragment of 690 bp);
F) Mp*DIV2* 722LpMpGE_En04 colonies, lanes 1 to 10 (expected size fragment of 690 bp);
G) Mp*DIV2* 829RpMpGE_En04 colonies, lanes 1 to 8 (expected size fragment of 690 bp);
H) Mp*DIVL* 30LpMpGE_En04 colonies, lane 1 (expected size fragment of 690 bp);
I) Mp*DIVL* 883RpMpGE_En04 colonies, lanes 1 to 6 (expected size fragment of 690 bp);
J) Mp*DIVL* 950LpMpGE_En04 colonies, lane 1 (expected size fragment of 690 bp);
K) Mp*DRIF* 78LpMpGE_En04 colonies, lanes 1 to 5 (expected size fragment of 690 bp);
L) Mp*DRIF* 530RpMpGE_En04 colonies, lanes 1 to 10 (expected size fragment of 690 bp);
M) Mp*DRIF* 818RpMpGE_En04 colonies, lanes 1 to 6 (expected size fragment of 690 bp);
MM: Molecular marker (NZYDNA Ladder III, Nzytech); CN: Negative control;
The lanes with red squares correspond to the colonies selected for sequencing.

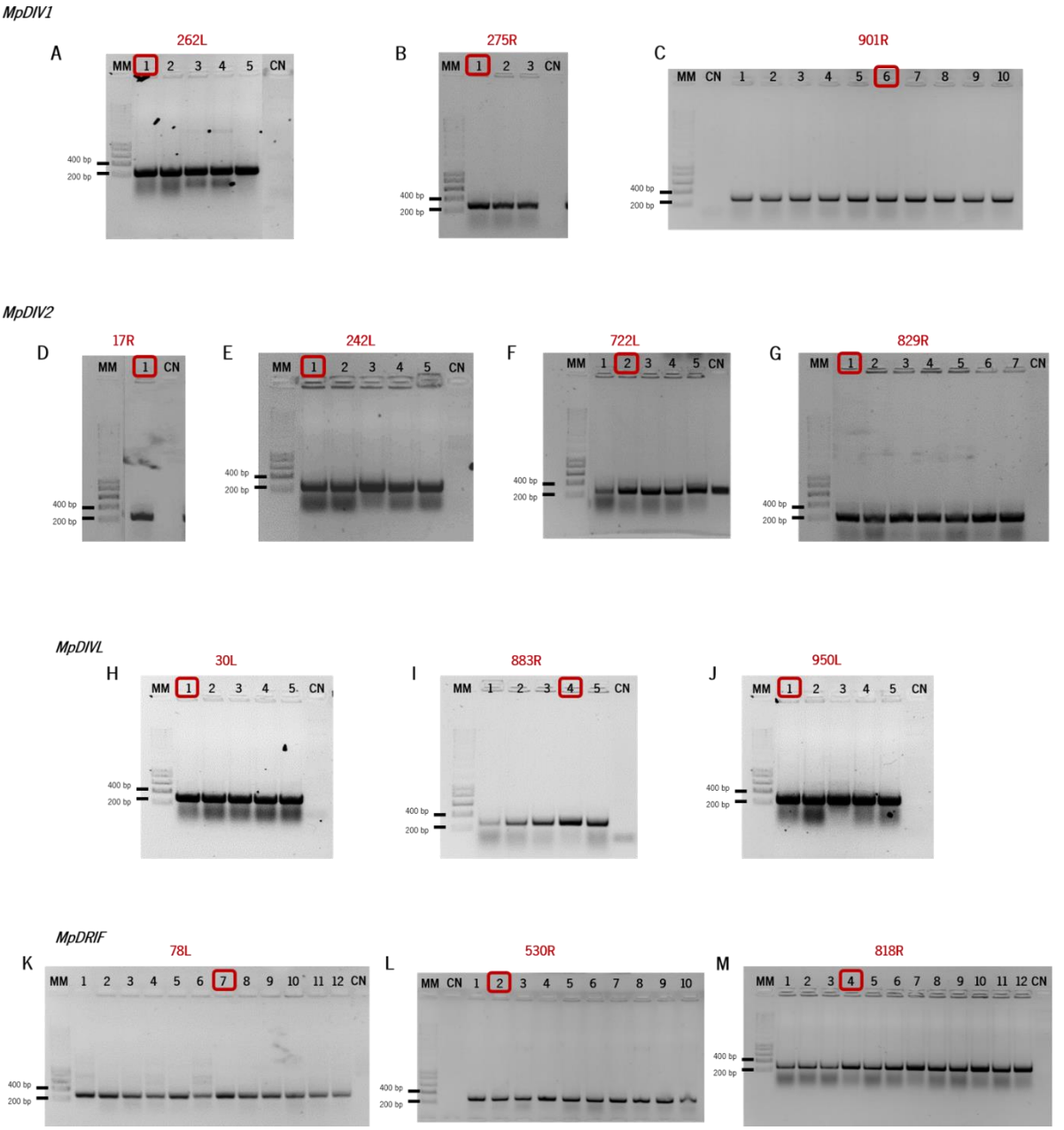

Figure F2: PCR screening of *E. coli* colonies containing constructs with the gRNAs L and R in pMpGE010 and pMpGE011 vectors, respectively, for CRISPR/Cas9 system.

A) Mp*DIV1* 262LpMpGE010 colonies, lanes 1 to 5 (expected size fragment of 274 bp);
B) Mp*DIV1* 275RpMpGE011 colonies, lanes 1 to 3 (expected size fragment of 274 bp);
C) Mp*DIV1* 901RpMpGE011 colonies, lanes 1 to 10 (expected size fragment of 274 bp);
D) Mp*DIV2* 17RpMpGE011 colonies, lane 1 (expected size fragment of 274 bp);
E) Mp*DIV2* 242LpMpGE010 colonies, lanes 1 to 5 (expected size fragment of 274 bp);
F) Mp*DIV2* 722LpMpGE010 colonies, lanes 1 to 5 (expected size fragment of 274 bp);
G) Mp*DIV2* 829RpMpGE011 colonies, lanes 1 to 7 (expected size fragment of 274 bp);
H) Mp*DIVL* 30LpMpGE010 colonies, lanes 1 to 5 (expected size fragment of 274 bp);
I) Mp*DIVL* 883RpMpGE011 colonies, lanes 1 to 5 (expected size fragment of 274 bp);
J) Mp*DIVL* 950LpMpGE010 colonies, lanes 1 to 5 (expected size fragment of 274 bp);
K) Mp*DRIF* 78LpMpGE010 colonies, lanes 1 to 12 (expected size fragment of 274 bp);
J) Mp*DRIF* 530RpMpGE011 colonies, lanes 1 to 10 (expected size fragment of 274 bp);
M) Mp*DRIF* 818RpMpGE011 colonies, lanes 1 to 12 (expected size fragment of 274 bp);
MM: Molecular marker (NZYDNA Ladder III, Nzytech); CN: Negative control;
The lanes with red squares correspond to the colonies selected for sequencing.

Annex G: CRISPR/Cas9 mutant plants genotyping

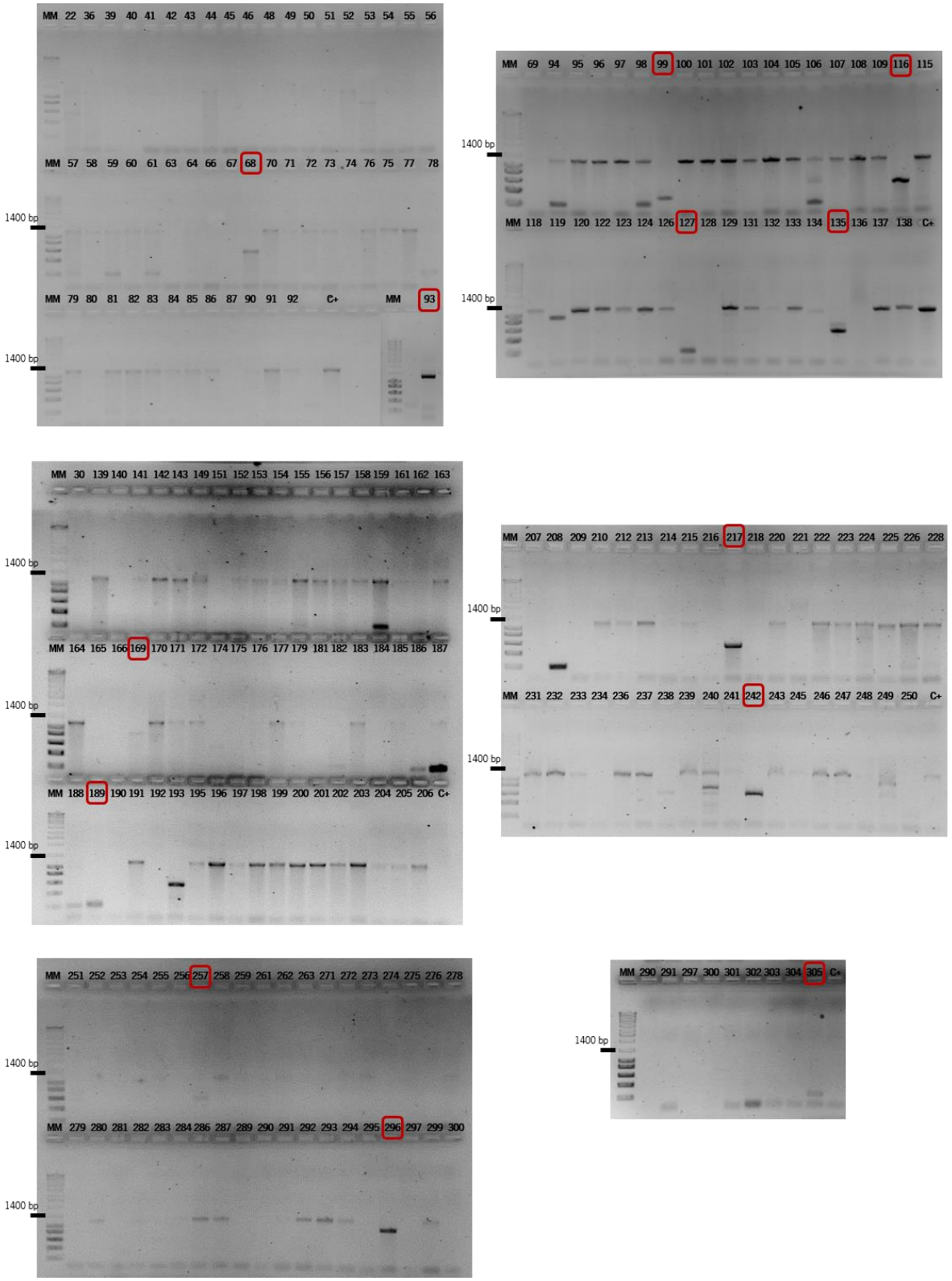

Figure G1: PCR genotyping of *Marchantia polymorpha* Mp*DIV2* transformed plants. Transformed plants 22 to 305 of the gRNA combinations 242L+836R, 722L+17R and 242L+829R were genotyped using the pair of primers 455+456. The expected size of the PCR fragment amplified in wild-type plants was 1159 bp. Lane numbers with red squares (68, 93, 99, 116, 127, 135, 169, 189, 217, 242, 257, 296 and 305) represent plants whose genomic PCR produced a fragment with a different size from the wild type. MM: Molecular marker (NZYDNA Ladder III, Nzytech); C+: Positive control with wild-type genomic DNA

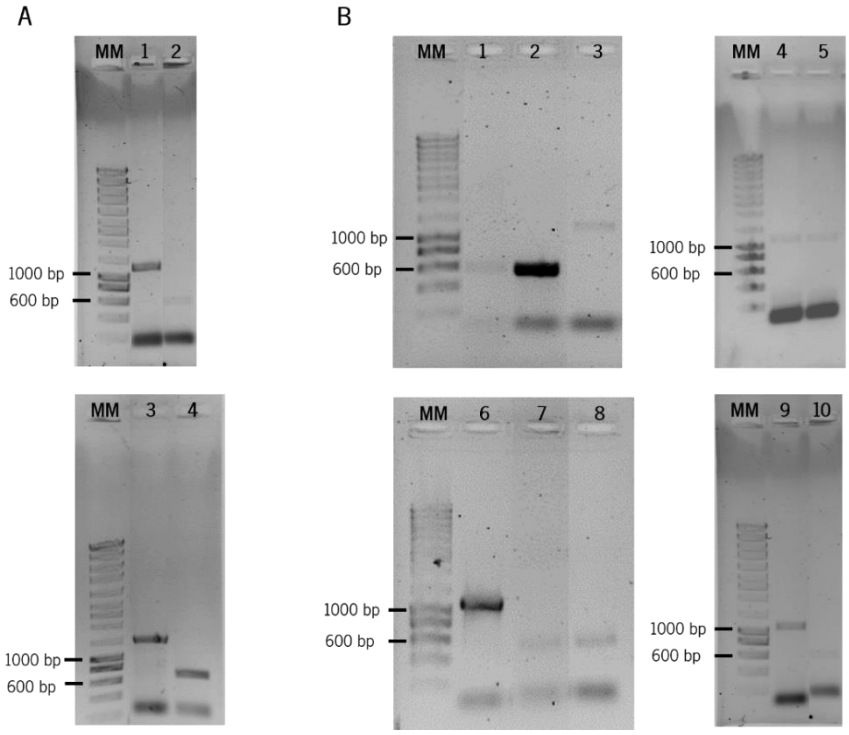

Figure G2: Determination of the Mp*divl* and Mp*div2* mutant plants sex by PCR amplification. For a male plant was expected a fragment with a size close to 1100 bp and for a female plant a fragment with a size close to 516 bp.

A) Determination of Mp*divl* plants sex.

Lane 1: Mp*divl* line #73 (≈ 1100 bp fragment);

Lane 2: Female plant control (≈ 516 bp fragment);

Lane 3: Mp*divl* line #87 (≈ 1100 bp fragment);

Lane 4: Female plant control (≈ 516 bp fragment);

B) Determination of Mp*div2* plants sex.

Lane 1: Mp*div2* line #93 (≈ 516 bp fragment);

Lane 2: Mp*div2* line #99 (≈ 516 bp fragment);

Lane 3: Mp*div2* line #189 (≈ 1100 bp fragment);

Lane 4: Mp*div2* line #169 (≈ 1100 bp fragment);

Lane 5: Mp*div2* line #257 (≈ 1100 bp fragment);

Lane 6: Mp*div2* line #217 (≈ 1100 bp fragment);

Lane 7: Mp*div2* line #305 (≈ 516 bp fragment);

Lane 8: Female plant control (≈ 516 bp fragment);

Lane 9: Mp*div2* line #68 (≈ 1100 bp fragment);

Lane 10: Female plant control (≈ 516 bp fragment);

MM: Molecular marker (NZYDNA Ladder III, Nzytech)

Table G1: Mp*div1*, Mp*div*l and Mp*div2* mutant plants insertions, deletions and sex.

Gene	Mutant	gRNA combination	Insertion (+bp)	Deletion (-bp)	Sex
Mp*div1*	#69	262F/901R	12	270	Male
	#87	262F/901R	———	229	Male
Mp*div*l	#26	30F/786R	402	2bp	Female
	#73	30F/786R	102	———	Male
	#87	30F/786R	———	767	Male
Mp*div2*	#5	242F/836R	———	350	Female
	#24	242F/836R	3	485	Male
	#68	242F/836R	———	349	Male
	#93	722F /17R	75	47	Female
	#99	722F /17R	———	954	Female
	#116	242F/836R	2	350	Male
	#127	722F /17R	———	950	Female
	#135	242F/829R	———	343	Female
	#169	242F/829R	21	351	Male
	#189	722F /17R	———	949	Male
	#217	242F/836R	———	350	Male
	#242	242F/836R	———	349	Male
	#257	242F/829R	4	313	Male
	#296	242F/829R	97	145	Male
	#305	722F /17R	———	950	Female

Annex H: Mpdiv2 protein alignments

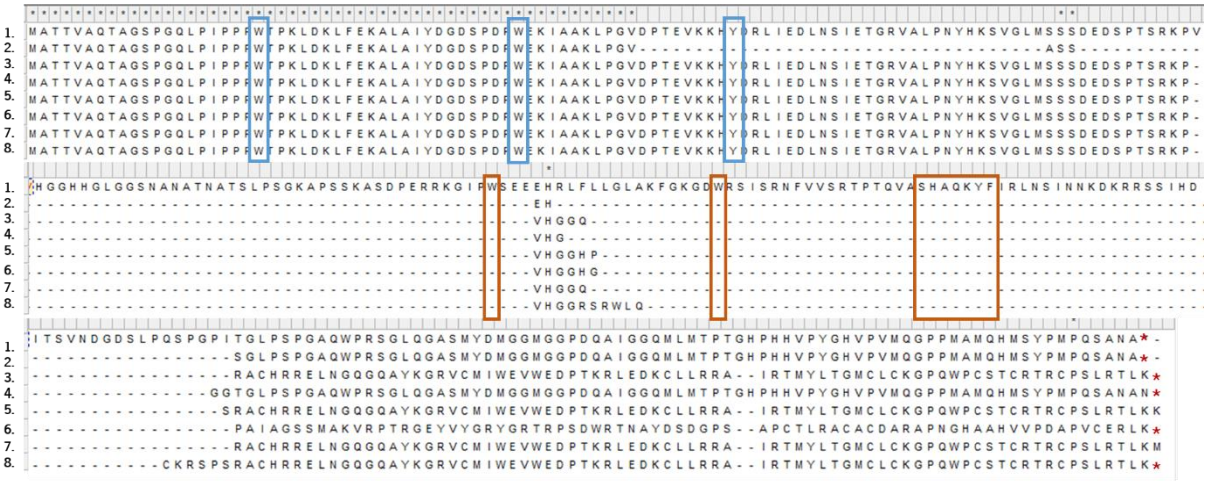

Figure H1: MpDIV2 and Mpdiv2 Group I proteins alignment. Wild-type MpDIV2 (1) and Mpdiv2 mutant proteins of Group I #24 (2), #68 (3), #116 (4), #135 (5), #217 (6), #242 (7) and #257 (8) were aligned by the ClustalW algorithm. MpDIV2 conserved MYB domains are represented by the blue- MYBI domain (-W-X23-W-X20-Y-)- and the orange- MYBII domain (-W-X19-W-X22-Y-)- colours. The deletions gave origin to mutant proteins with the disruption of the MYB II domain. The red asterisks represent the translation termination.

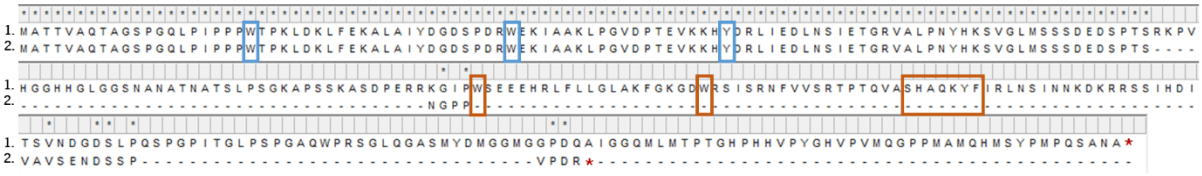

Figure H2: MpDIV2 and Mpdiv2 Group II proteins alignment. Wild-type MpDIV2 (1) and Mpdiv2 mutant protein of Group II #296 (2) were aligned by the ClustalW algorithm. MpDIV2 conserved MYB domains are represented by the blue- MYBI domain (-W-X23-W-X20-Y-)- and the orange- MYBII domain (-W-X19-W-X22-Y-)- colours. The deletions and insertions gave origin to mutant proteins with the disruption of the MYB II domain. The red asterisks represent the translation termination.

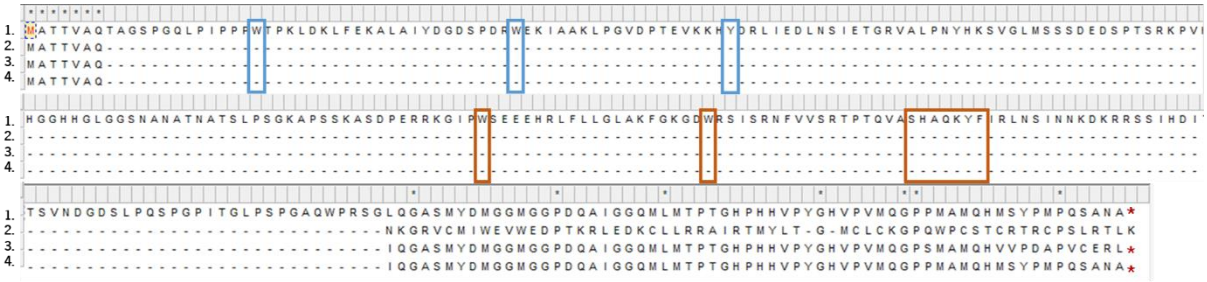

Figure H3: MpDIV2 and Mpdiv2 Group III proteins alignment. Wild-type MpDIV2 (1) and Mpdiv2 mutant protein of Group III #127 (2, #189 (3) and #305 (4) were aligned by the ClustalW algorithm. MpDIV2 conserved MYB domains are represented by the blue- MYBI domain (-W-X23-W-X20-Y-)- and the orange- MYBII domain (-W-X19-W-X22-Y-)- colours. The deletions and insertions gave origin to mutant proteins with the disruption of the MYB I and MYB II domains. The red asterisks represent the translation termination.

Annex I: Mp*div2* phenotype analysis

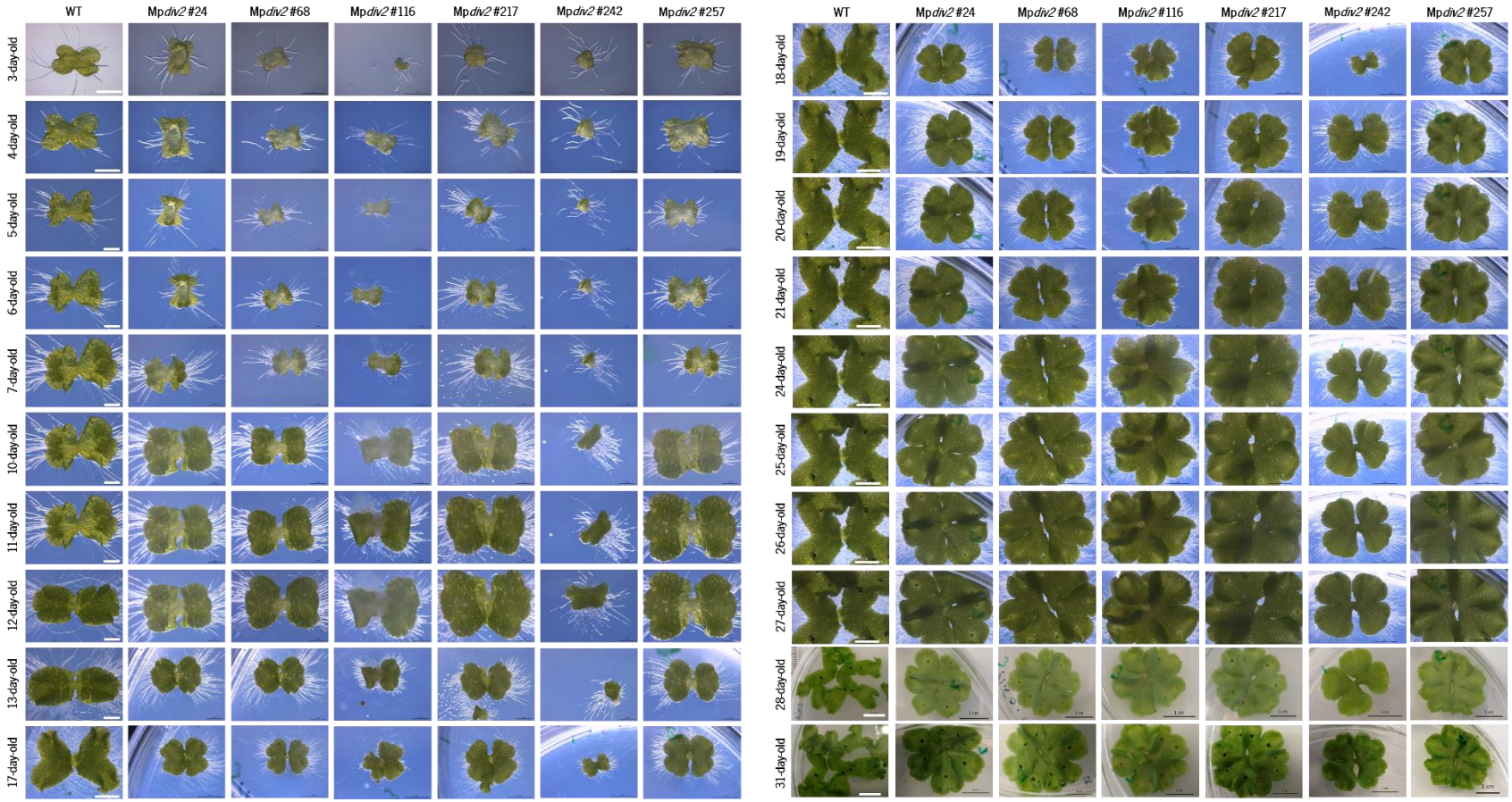

Figure I1: Analysis of Mp*div2* Group I knockout in *Marchantia polymorpha* plants. Wild- type plants and Mp*div2* mutant plant lines #24, #68, #116, #217, #242 and #257 of Group I were observed during 31 days after gemmae propagation. The plants were grown on half strength Gamborg B5 medium under long-day conditions (16 h light/ 8 h dark) at 20 °C and with light intensity between 40-45 µmol m^{-2} s^{-1}. Scale bar of: 1 mm from day 3 to 12; 2 mm for day 13; 5 mm from day 17 to 27 and 1 cm from day 28 to 31.

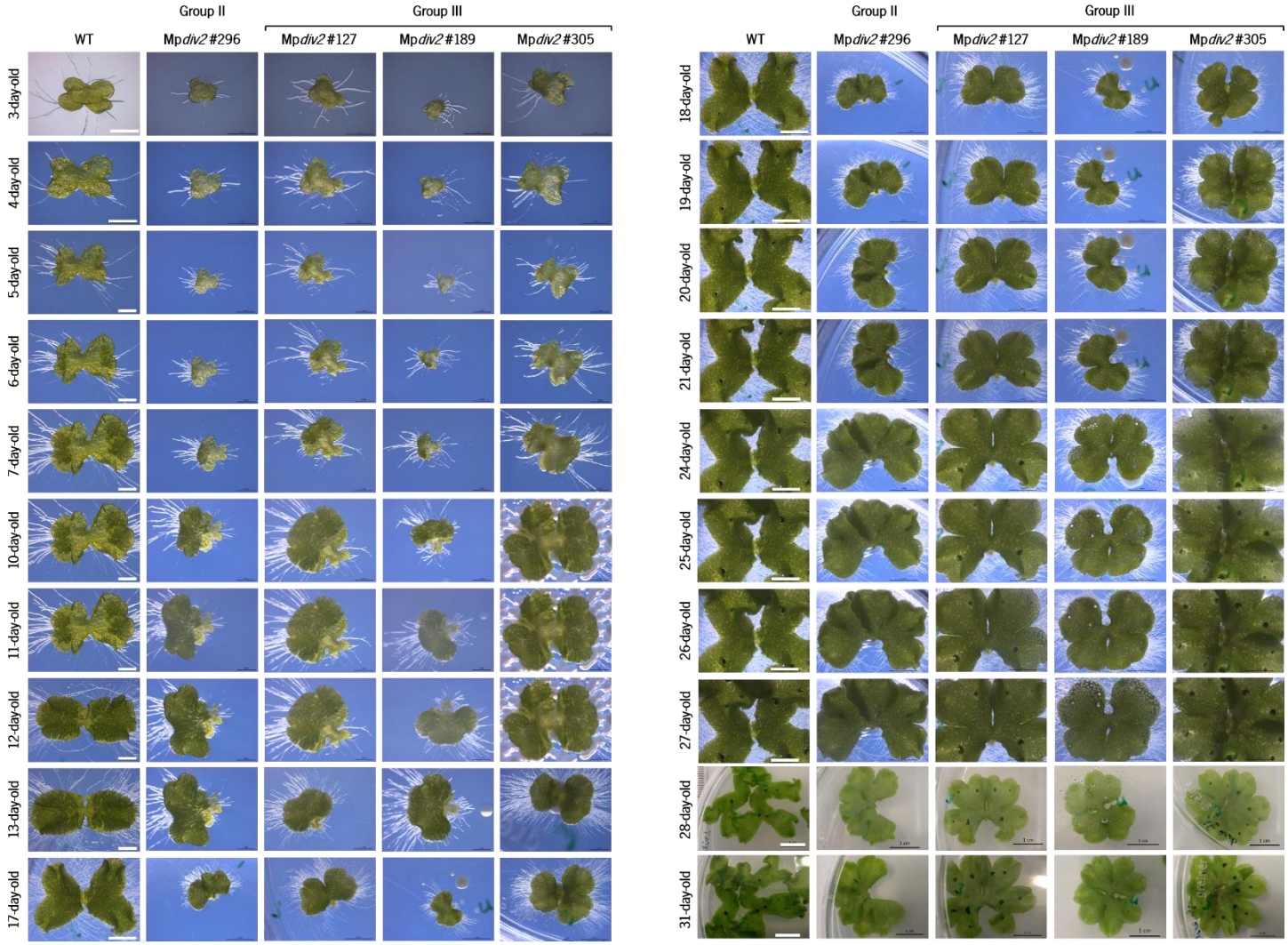

Figure I2: Analysis of Mp*div2* Group II and III knockout in *Marchantia polymorpha* plants. Wild- type plants and Mp*div2* mutant plant lines #296 of Group II and #127, #189 and #305 of Group III were observed during 31 days after gemmae propagation. The plants were grown on half strength Gamborg B5 medium under long-day conditions (16 h light/ 8 h dark) at 20 °C and with light intensity between 40-45 µmol m^{-2} s^{-1}. Scale bar of: 1 mm from day 3 to 12; 2 mm for day 13; 5 mm from day 17 to 27 and 1 cm from day 28 to 31.

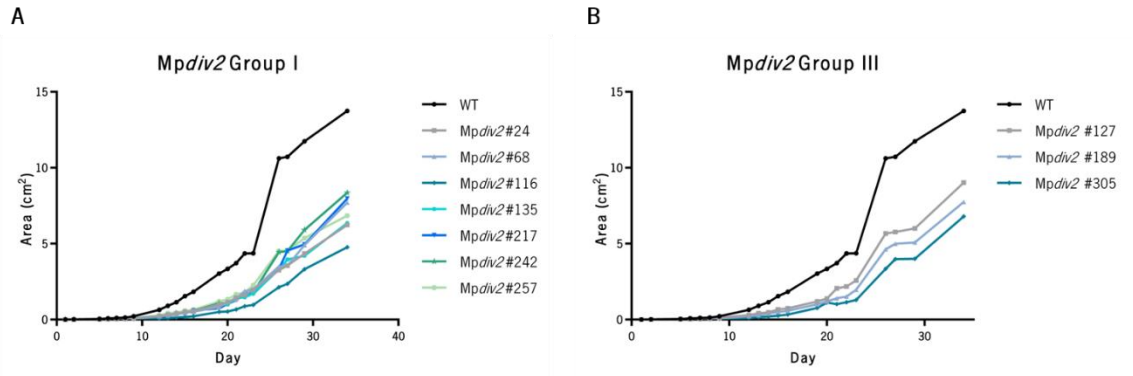

Figure I3: Analysis of Mp*div2* knockout in the thallus superficial area of *Marchantia polymorpha* plants. A) Thallus superficial area of wild-type (black line) (n=8) and Mp*div2* Group I line #24 (n=8), #68 (n=8), #116 (n=3), #135 (n=4), #217 (n=6), #242 (n=8) and #257 (n=7) plants (grey line). **B)** Thallus superficial area of wild-type (black line), Mp*div2* Group III line #127, #189 and #305 plants (n=8). The analysis was performed using the thallus superficial areas of wild-type and Mp*div2* mutant plants measured for 1 month.

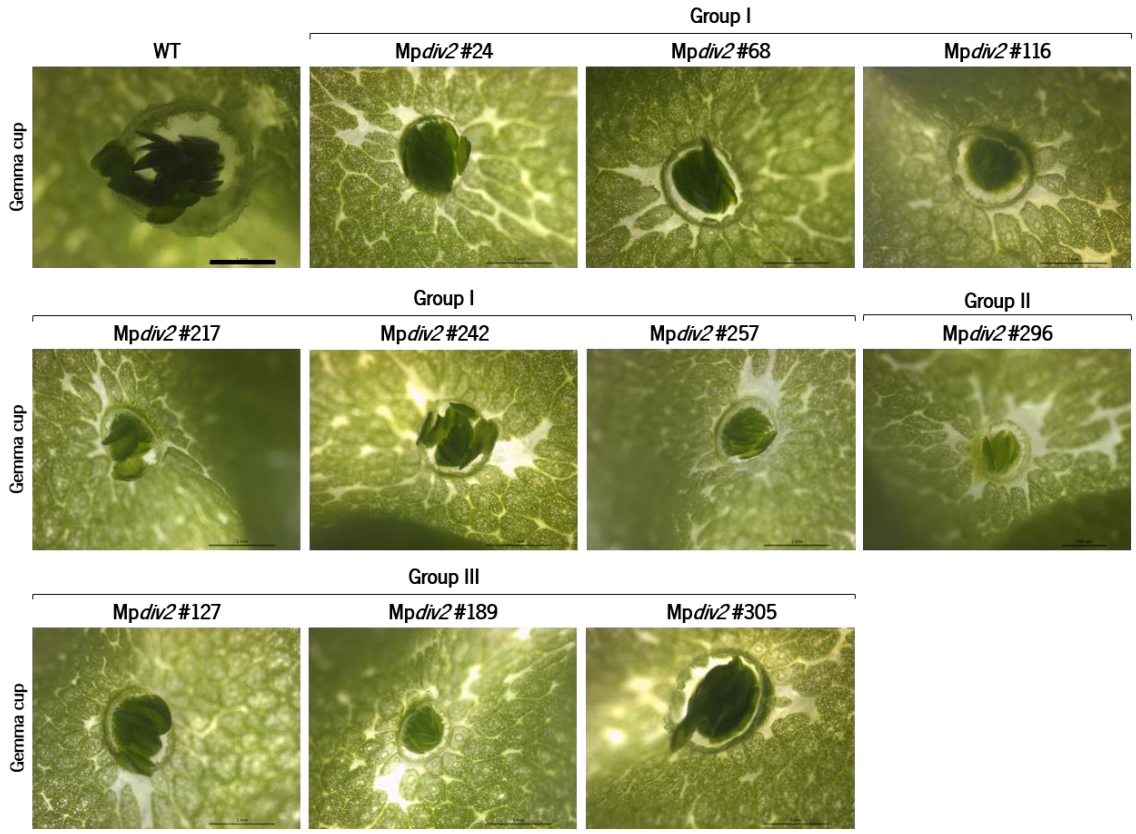

Figure I4: Analysis of Mp*div2* knockout in the gemma cups of *Marchantia polymorpha* plants. The gemma cups of wild-type plants and Mp*div2* mutant plant lines #24, #58, #116, #217, #242, #257 (Group I), #296 (Group II), #127, #189 and #305 (Group III) were observed in plants with 1 month. Scale bar: 1 mm.

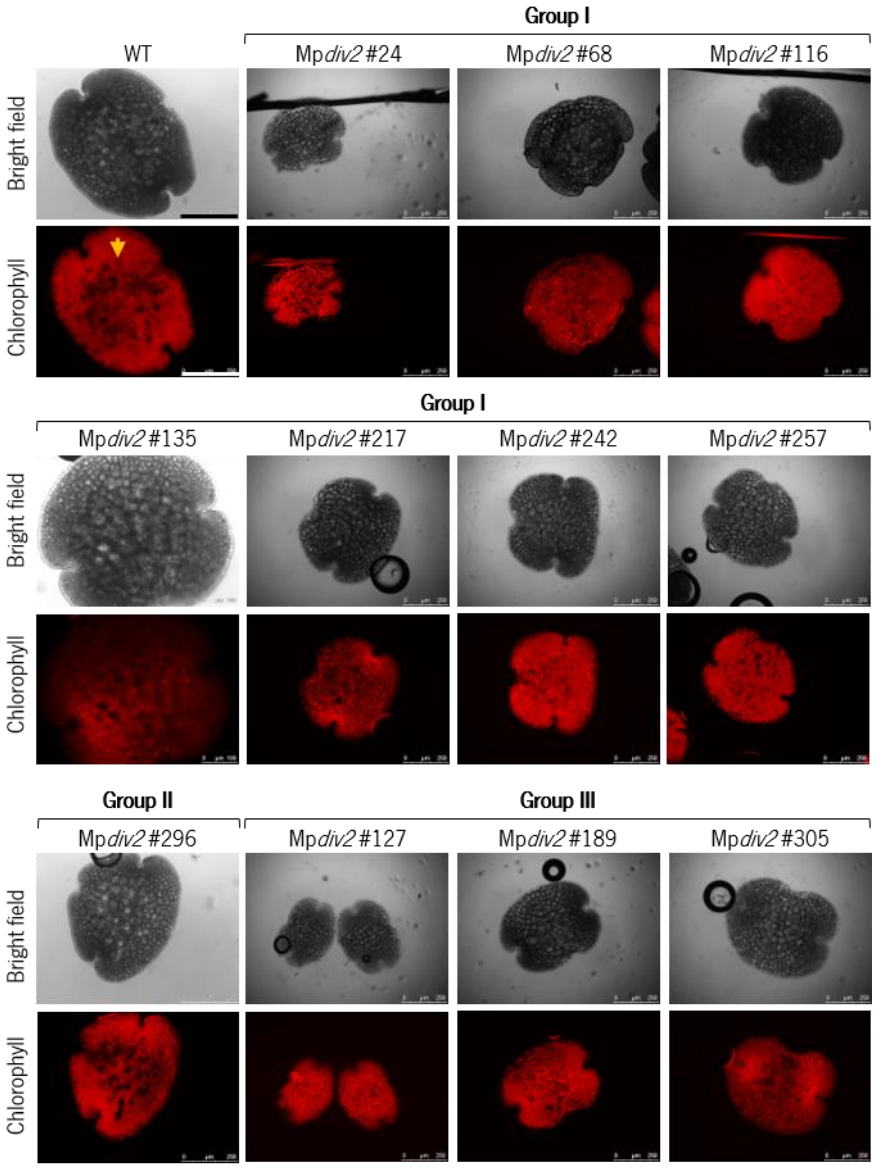

Figure I5: Analysis of Mp*div2* knockout in the rhizoids of *Marchantia polymorpha* plants. Gemmae of wild-type plants and Mp*div2* mutant plant lines #24, #58, #116, #135, #217, #242, #257 (Group I), #296 (Group II), #127, #189 and #305 (Group III) were observed under a fluorescence microscope using chlorophyll autofluorescence. The dark areas represent the rhizoid initial cells (yellow arrow). Black and white scale bars: 250 µm.