

Phylogenetic and *in vivo* analysis of MRB proteins in *Physcomitrella patens*

A Major Qualifying Project submitted to the Faculty of
WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the
degree of Bachelor of Science in
Bioinformatics & Computational Biology

by

Jordan Hollands

Date: 18th May 2020

Approved by:

Luis Vidali, PhD – Advisor

Advised by:

Dr. Luis Vidali

Worcester Polytechnic Institute

This report represents work of a WPI undergraduate student submitted to the faculty as evidence of a degree requirement. WPI routinely publishes these reports on its web site without editorial or peer review. For more information about the projects program at WPI, see <http://www.wpi.edu/Academics/Projects>.

Abstract

The uncharacterized myosin binding protein, MyoB, has been identified as possible mechanism through which myosin XI attaches to its cargo. A previous yeast two-hybrid screen using myosin XI cargo binding domain, identified one of six MyoB proteins in the model moss, *Physcomitrella patens*. However, the function of MyoB in *P. patens* remains unknown. In this study, the family of six proteins were analyzed to identify common structural features and construct a phylogenetic tree. In addition, a triple mEGFP fusion protein was constructed to analyze protein localization.

Acknowledgements

Thank you to Professor Luis Vidali, for his unending patience and mentorship from the beginning to end of this project.

Thank you to Prof. Liz Ryder for her time as my advisor, as well as her help and encouragement in graduating from WPI as a BCB major.

I would also like to thank, Liu Boyuan, Guilia Galotto, Rob Orr and the rest of the members of Vidali lab for their advice and assistance with my experiments.

Table of Contents

Abstract.....	ii
Acknowledgements.....	iii
Table of Contents	iv
Introduction.....	1
Materials and Methods.....	9
Results	15
Discussion.....	27
References	30

Introduction

The determining characteristic of bryophytes is their lack of vascular tissues and roots. Bryophytes have neither a true root system or xylem and phloem, to transport water and nutrients throughout the plant (Taiz et al., 2018). Due to divergence as early as 500 millions years ago, bryophytes are an extant species with similar characteristics to the ancestral plants which all land plants evolved from (Rensing et al., 2008). There are three division of bryophytes, liverworts, mosses, and hornworts. Mosses are ideal model for comparative studies of the evolution of biological processes in land plants, because the phylogenetic separation between mosses and vascular plants can provide insight into the mechanisms of this evolution. Mosses and flowering plants share fundamental physiological processes, which are essential for life yet evolved separately from one another. The amount of time since the lineages diverged allowed for each lineage's gene families to expand separate from one another. The moss *Physcomitrella patens* is a model plant that belongs to an early diverging extant land plant lineage (Rensing et al., 2008). By comparing studies done in other extant species, such as the model flowering plant *Arabidopsis thaliana*, with studies done in *P. patens*, we can learn more about important biological processes present in all plants. In this work, I am interested in studying the actin-myosin transport network. This essential system was present in the common ancestor of charophytes and embryophytes and which had a major expansion in conjunction with terrestrial colonization.

P. patens is an ideal plant species to use for reverse genetic studies. Protoplasts, (moss cells stripped of their cell walls) can be easily generated, and these single cells can be regenerated and propagated any stage in the moss's life cycle (Rensing et al., 2020). This results

in the simple production of clonal lines. Under specific conditions, protoplasts can be transformed, allowing for insertion of foreign DNA into the cell, and most importantly, *P. patens* has been shown to incorporate DNA by homologous recombination, allowing for gene deletions and insertions (Schaefer and Zryd, 1997). In bryophytes, the gametophyte generation is dominant over the course of the *P. patens*'s lifecycle, meaning it exists primarily as a multicellular haploid plant (Figure 1). Because a haploid cell contains only a single-copy of each gene, homologous recombination is an efficient gene-targeting method in *P. patens*. (effectively integrating DNA) In addition, the *P. patens* genome has been sequenced based on the wild-type “Gransden” strain (JGI), with extensive annotations by a vibrant international community.

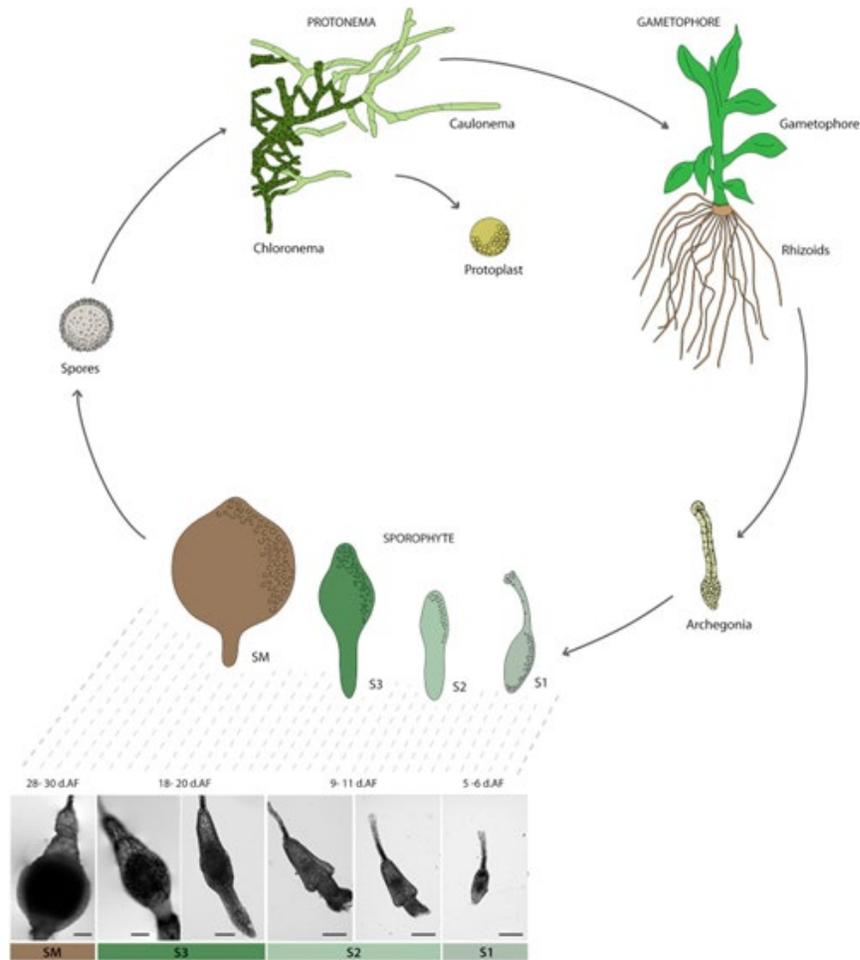


Figure 1. Lifecycle of *P. patens*.

The only time that *P. patens* cells exist in diploid form is during the sporophyte stage, and is haploid during all other stages. Transformation of *P. patens* can occur during the protoplast stage. Genetic alterations would be displayed phenotypically in the protonema which later develop from these protoplasts.

P. patens has been extensively studied as a model for tip growth in plants, because it expands via tip growth of its protonemal cells (Menand et al., 2007). Due to the shared morphological form, research on protonemal cells also applies to understanding the morphogenesis of root hairs and pollen tubes, cells critical for nutrient uptake and sexual reproduction, respectively. The large size of eukaryotic cells necessitates a transport network that

facilitates the movement of macromolecular complexes. Directional transport by cytoskeleton-based motors provides a means to shuttle these cellular components from their origin to their intra-cellular destination. In plants, polarized cell growth is essential for the development of nutrient gathering root hairs, pollen fertilization, and gametophyte development. Tip growth is a ubiquitous process throughout the plant kingdom in which a single cell elongates in one direction in a self-similar manner. In the process of tip growth, vesicles containing the materials required to construct the cell wall are delivered to the apex through the process of exocytosis. This requires a dynamic system, where the transport machinery travels along the newly assembled cytoskeleton as the cell elongates.

The two types of molecular motors found within plant cells are myosin and kinesin. kinesin and myosin. Myosin generate force along actin, where as kinesin bind and travel along microtubules. Conventional myosin proteins consist of four polypeptides, comprised of two heavy chain and two light chains. The heavy chains, the largest polypeptide subunits, span the entirety of the protein from terminus to terminus. The N-terminus of each heavy chains is known as the head region of the myosin. The head region contains the binding sites for acting and ATP, which confer myosin with its processive motor qualities. The two smaller light chains bind to the heavy chain upstream from the N-terminus, at what is called the neck domain to provide structural integrity. Past this neck, the two heavy coils spiral around each other in a coil-coil formation. The C-terminal ends of each of the heavy coils are known as the globular tail domain, or GTD, which play a role in binding to myosin's cargo.

Myosin XI, a class of myosin which evolved in plants, plays a key role in tip growth. participates in the organization of actin filaments at the tip, in addition to shuttling its organelle cargo. In flowering plants, myosin XI is also responsible for a rapid indirect transport of

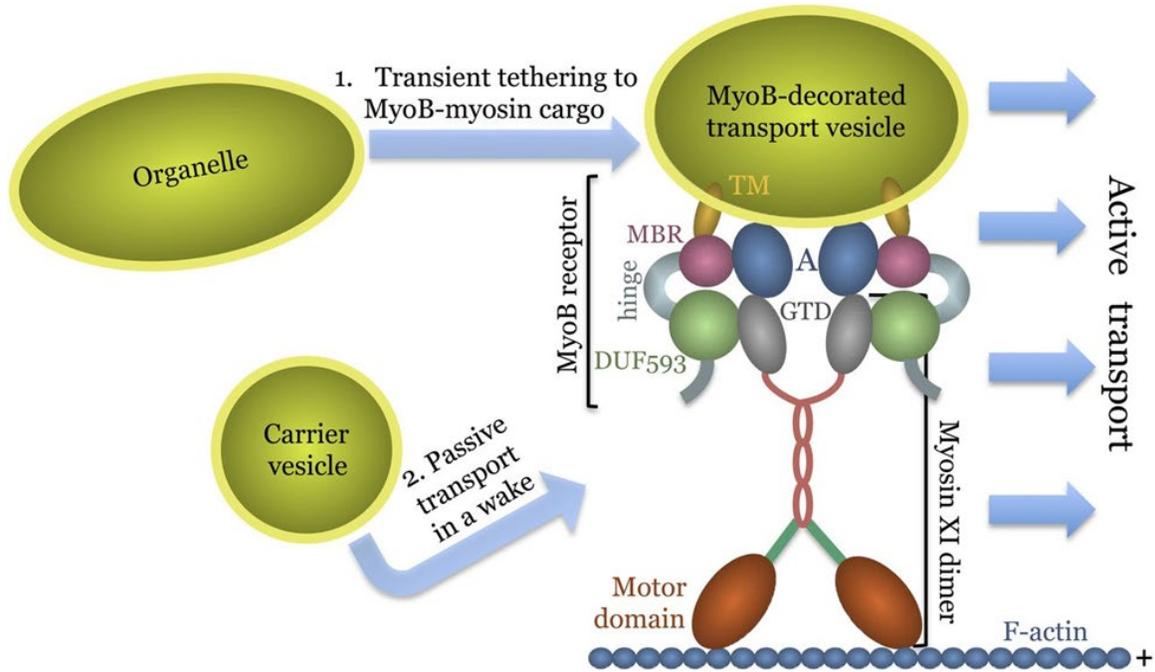


Figure 2. MyoB Myosin XI complex is responsible for cytoplasmic streaming in *Arabidopsis thaliana*
 Structural diagram of Myosin XI MyoB complex which outlines the hypothetical MyoB role could play in vesicle attachment to Myosin XI. The process of cytoplasmic streaming involves the active transport of MyoB attached vesicles by Myosin XI which leads to the passive transport of other vesicles.

organelles, called cytoplasmic streaming (Figure 2). In mosses, cytoplasmic streaming of organelles is not observed, while in the giant characoid algae, cytoplasmic streaming reaches velocities of up to 70 $\mu\text{m}/\text{sec}$ (Tominaga et al., 2013). Because mosses have cells of a similar size to those of higher-order plants, it is not clear why they do not require cytoplasmic streaming.

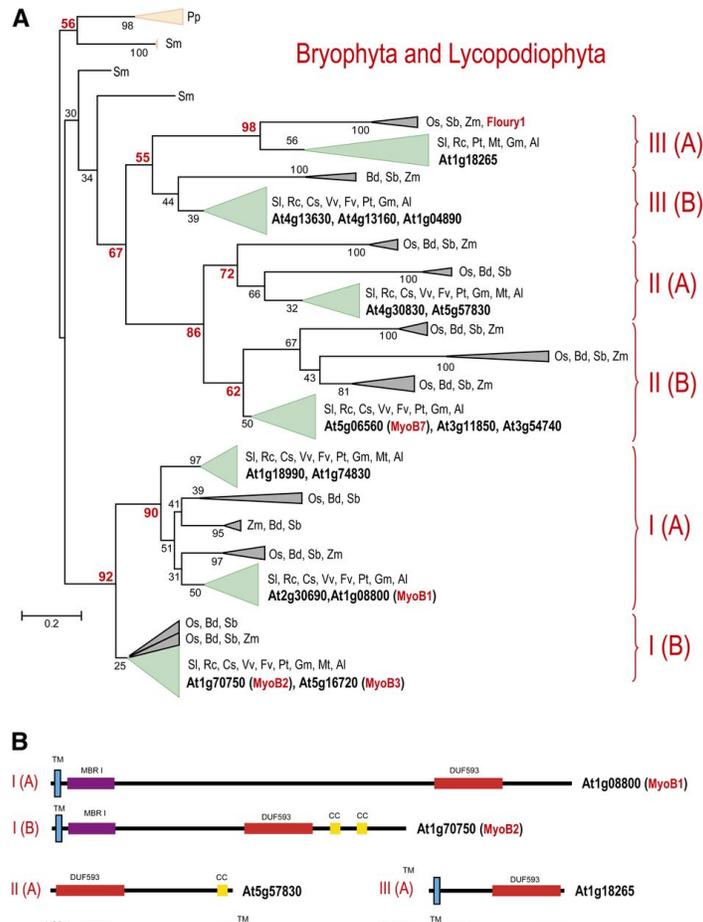


Figure 3. Phylogenetic Tree and Domain Architectures of DUF593-Containing Proteins.
 A Unrooted maximum likelihood tree of DUF593 or MyoB proteins found in land plants.
 B Representation of common domain structures for each subfamily of proteins containing DUF593

Recent studies in *A. thaliana* have shown that the protein MyoB (for myosin binding) may be important for anchoring myosin XI to organelles during cytoplasmic streaming, and that functional cooperation between myosin XI and MyoB is required for proper plant development (Peremyslov, 2013). The number of MyoB genes found in *A. thaliana* is much greater than the 6 unique MyoB genes in *P. patens*. The six genes in *P. patens* share the same general domain architecture, whereas in higher plants, multiple subfamilies of MyoB genes exist which vary greatly in domain structure (Figure 3). Lower land plants such as *P. patens* and *S. moellendorffii* have less complex mechanisms for cellular transport. Recent phylogenetic analysis revealed

congruent evolutionary histories of the myosin XI, MyoB families (Kurth et al., 2017). Both gene families emerged in green algae and show concurrent expansions via serial duplication in flowering plants. Thus, the myosin XI transport network increased in complexity and robustness concomitantly with the land colonization by flowering plants and, by inference, could have been a major contributor to this process.

In *A. thaliana*, MyoB1 and MyoB2 were identified with a yeast two-hybrid screen using the myosin XI-K tail region as bait. (Peremyslov, 2013). These proteins shared in a common globular tail domain (GTD) binding domain. The GTD-binding domain, previously classified as a myosin binding domain of unknown function 593 (DUF593) confers the ability to bind to the N-terminal end of myosin. The domain DUF593 is conserved in all flowering plants, a PSI-BLAST search in the Refseq database identified 173 DUF593-containing proteins in all sequenced genomes of land plants (Peremyslov, 2013; Stephan et al., 2014). These proteins could have an essential function in the transport network of plants. As a receptor for myosin, these proteins could recruit and attach myosin to their.

It is hypothesized, that in *A. thaliana*, these myosin-binding proteins serve the purpose of linking myosin to vesicle-like endomembrane compartments (Peremyslov, 2013); subsequent movement of myosin and its MyoB linked cargo would then produce a cytosolic flow, which drives the movement of free-floating organelles. This could allow myosin to indirectly transport organelles and secretory vesicles to their destinations throughout the cell via cytoplasmic streaming. Consistent with these, when MyoB activity is knocked out, myosin XI cannot attach to their cargo, halting myosin dependent transportation and cytoplasmic streaming - affecting both polarized and diffuse cell growth (Kurth et al., 2017).

Objective of this work:

The goal of this project was to develop further understanding of the function of myosin-binding proteins present in non-vascular plants where cytoplasmic streaming does not occur. The MyoB gene family in *P. patens* was chosen in order to investigate its unknown function in non-vascular plants. We first studied the genetic sequences of this gene family in order to further develop hypotheses of their function based on their domain structures and phylogenetic relation to other previously studied genes. As a potential interactor with myosin XI we expected MyoB to localize with myosin XI and share the same pattern of expression as myosin XI. Our next objective was to develop a fusion protein of GFP with a MyoB protein. This fusion allowed for the visualization of MyoB expression in-vivo in *P. patens* cells. Visualization of the MyoB offered the other source of information on the possible function of MyoB.

Materials and Methods

Phylogenetic Analysis

MRB Protein Families

To identify all the proteins related to MyoB from *A. thaliana*, we first looked for proteins with homology to the DUF593 domain. To compare the expansion of the gene family we looked for orthologs in other plants. The plant species *Marchantia polymorpha*, *Physcomitrella patens*, *Oryza sativa*, *Selaginella moellendorffii*, and *Arabidopsis thaliana* were chosen in order to include both vascular and non-vascular plants. The genomes for these species were accessed from the Phytozome database (phytozome.jgi.doe.gov). The nucleotide sequences of MyoB 1,2,3,4, and 7 were used to perform multiple BLASTs on each chosen species' genomes. Accession numbers for the MyoB sequences were obtained from the paper (Peremyslov, 2013). The FASTA sequences for the protein homologs identified by BLAST were compiled with the software Geneious.

Protein Analysis and Annotation

In order to visually observe the variation in different myosin-binding proteins, domains were predicted and annotated in Geneious. The plugin Interproscan was used to predict and label transmembrane domains and the GTD-binding domain for each transcript. The Coiled-Coils plugin was used to estimate the likeliness of the sequence from adopting a coiled-coil conformation. Sections of the transcript with high likeliness were labeled manually. An approx. 100 amino acids peptide extraction of the GTD-binding domain was done for each of the transcripts for sequence alignment and tree construction.

Alignment and Trees

The alignment of the extraction representing the GTD-binding domain for each protein was performed using the ClustalW Omega plugin. Phylogenetic trees were created using the MrBayes plugin. The single *M. polymorpha* protein was used to root the tree.

Expression Analysis

To determine which of the six MRB proteins had the greatest expression in differ we looked at the expression of the MRB transcripts in different *P. patens* tissues was found using the Physcomitrella eFP Browser (Ortiz-Ramirez et al., 2016) (http://bar.utoronto.ca/efp_physcomitrella/cgi-bin/efpWeb.cgi). Gene ID's for the proteins were found listed under aliases in the Phytozome gene view. Absolute expression was found for each transcript using a signal threshold determined by the highest expressed transcript MRB1 (4763), to enhance the expression in each specific tissue, a separate set of images was generated thresholded at the highest level of tissue expression of each individual transcript.

MRB1-3mEGF Construct

Primer Design

The primers were designed according to Gateway protocol (Thermo). Each primer consisted of an attachment site (att) specific sequence, and 18-25 bp of gene-specific sequence, and four guanine residues at the 5' end. Two of the four primers contained an ATTT^AAAT swaI digestion site.

Entry Clone Construct. The PCR fragment for the P1P5 entry clone was created through amplifying the Grandsen wild-type DNA using the following primers: Y2HMyoB-SwaI-F and Y2HMyoB-noStop-R. The P3P2 entry clone was created using the following primers: Y2HMyoB-3end-F and Y2HMyoB-3end-SwaI-R. Sequences for these primers can be found in Table I. The PCR was done with a 20 μ l reaction using New England BioLab (NEB) Inc.’s protocol “PCR Protocol for Phusion® High-Fidelity DNA Polymerase (M0530).” The following setting were used according the NEB’s protocol and Tm calculator: 98°C for 30 seconds; 30 cycles of: 98°C for 10 seconds, 64°C for 30 seconds, and 72°C for 2 minutes; 72°C for 2 minutes; hold at 4°C.

TABLE I		
Primer name	Nucleotide sequence	Binding site
Y2HMyoB-SwaI-F	GGGGACAAGTTTGTACA AAAAAGCAGGCTTAATT TAAATATTAGATTCAAG GATACTGGTCAAG	atB1
Y2HMyoB-noStop-R	GGGGACAAC TTTTGTAT ACAAAGTTGTAACAACA CAATGGAGATTATTATTT TCCC	attB5
Y2HMyoB-3end-F	GGGGACAAC TTTTGTATA ATAAAGTTGCTTAGTTA AGGGGTATCAATAGGAC TG	attB3
Y2HMyoB-3end-SwaI-R	GGGGACCACTTTTGTACA AGAAAGCTGGGTAATTT AAATTACTAAA ACTTCA AAAGAAATAACTTCCG	attB2

The amplified PCR products were separated on a 0.8% agarose gel and purified using the Zymoclean Gel DNA recovery kit (Zymo). Successful replication of the PCR amplicons was confirmed through lab sequencing.

BP-reactions were performed to integrate the attB-PCR products into pDONR 221 vectors following manufacturer instructions (Thermo). The BP reaction was performed according to the ThermoFisher protocol. (which binding site was used for each reaction?)

1 ul of each reaction was used to transform chemically competent DH5 α E. coli cells. For each transformation two LB+kanamycin selective plates were prewarmed to 37°C. 20 ul and 100 ul from each transformation were spread onto the plates and incubated overnight at 37°C. Individual colonies were picked and plasmid DNA purified using Zippy Plasmid Miniprep Kit. To screen for positive clones, the plasmid DNA was digested with AflIII and EcorRV restriction enzymes and ran on agarose gel.

Expression Clone Construct

Transformation into E. coli was done using standard protocols. For the four fragment expression clone, an entry clones containing triple-EGFP, and hygromycin resistance were obtained from the lab. An LR reaction was performed with these clones and the two MRB1 entry clones to recombine them into the destination vector. Due to the final destination vector's large size additional steps were taken to maximize recombinational efficiency. A higher efficiency was attained by linearizing the entry clones and destination vector prior to the reaction, as well as incubating the reaction overnight to yield more colonies. DH5 α E. coli were transformed and plated on LB+carbenicillin plates. The selected colonies that grew were picked for overnight growth in 3 mL of LB and carbenicillin. A miniprep was performed on the samples after incubation to prepare DNA for restriction analysis. A BamHI digest was used to confirm all four entry clones had been successfully integrated into the complete expression clone.

A four-fragment LR reaction was performed with the two MRB1 entry clones, an entry clone containing triple-EGFP, and a clone conferring hygromycin resistance. A BamHI

restriction digest was performed to ensure that all fragments were properly integrated into the final targeting construct.

***P. patens* Culture and Stable transformation**

P. patens was cultured following established protocols (Liu and Vidali, 2011). The MRB1-3mEGFP fusion plasmid was re-transformed into DH5 *E. coli* cells. The transformed bacteria was grown in LB and carbenicillin to produce sufficient DNA for a maxi-prep plasmid preparation (Zymo). One hundred and twenty micrograms of the resulting plasmid were digested with *Swa*I to generate a linearized targeting construct. The linearized DNA was precipitated with ethanol and resuspended in 60 μ l of TE. The DNA was transformed into *P. patens* protoplasts using the polyethylene glycol method (Liu and Vidali, 2011). After fourth days of growth on PRMB, Stable transformants were selected by two rounds of hygromycin selection. Each round of selection consisted of a week of growth on PpNH₄ with hygromycin (15 μ g/mL) with one week of growth on normal PpNH₄ between them. The remaining plants were picked and numbered on PpNH₄ master plates.

Fluorescent Microscopy

To determine the phenotype of the picked plants, the master plates were observed and photographed using fluorescent microscopy. The masterplates were imaged using fluorescent microscopy on a Zeiss SteREO Discovery V12 stereofluorescence microscope was used. Photographs were taken using an AxioCamMR3 and the imaging software AxioVision, and viewed using ImageJ processing software. The plants plated on the master plate were grouped by relative fluorescence strength of the plants plated on the. Plants with a range of fluorescence strength were ground and passed to establish new lines on PpNH₄ plates.

Genotyping

The expanded lines were then genotyped through PCR amplification of the construct integrated within the *P. patens* genome. Tissue from each line were harvested for a DNA extract following the Vidali lab rapid genotyping protocol. The mutant lines Y2HMyoB-SwaI-F and Y2HMyoB-3end-SwaI-R. These primers were used for genotyping as they were the external primers used to make the initial entry clones, and will amplify whatever is present at the MRB1 gene locus. PCR using genomes containing the wild-type MRB1, such as the WT Gransden line used as a control should result in an amplicon 2.4 kb in length. This would be the result of the amplification of the WT gene and the trailing 800 bp extraneous fragment of non-coding DNA. If the MRB1-3mEGFP fusion has successfully inserted into the target locus, an 7kb long amplicon would be expected, the size of the entire expression construct.

Confocal Microscopy

Lines chosen for further investigation were screened using confocal microscopy to observe the localization of the 3mGFP signal. Images of the lines were taken using a SP5 confocal microscope expressing the fluorescent MRB1-3mEGFP fusion protein. Single plants were expanded to screen for fluorescent protein expression with a laser scanning confocal microscope and for genotyping by genomic DNA PCR using the external primers used for DNA amplification (Y2HMyoB-SwaI-F and Y2HMyoB-3end-SwaI-R). Using these primers, the targeted locus should either not amplify or amplify 1.25 kb product, while the wild-type untargeted locus should amplify a 7 kb product.

Results

Phylogenetic analysis of MRB proteins

To determine the phylogenetic relationships of proteins in the MRB gene family of *P. patens*, and to make inferences on gene function, we looked at the protein domain architecture compared to homologous proteins whose function has been studied previously by other groups (Peremyslov, 2013; Stephan et al., 2014). We found six homologous proteins, with a transmembrane domain present at the N-terminus, and two coiled-coil domains outside of the GTD binding domain (Figure R1). The MRB proteins found in *P. patens* are most similar to the I (B) subfamily identified previously (Peremyslov, 2013).

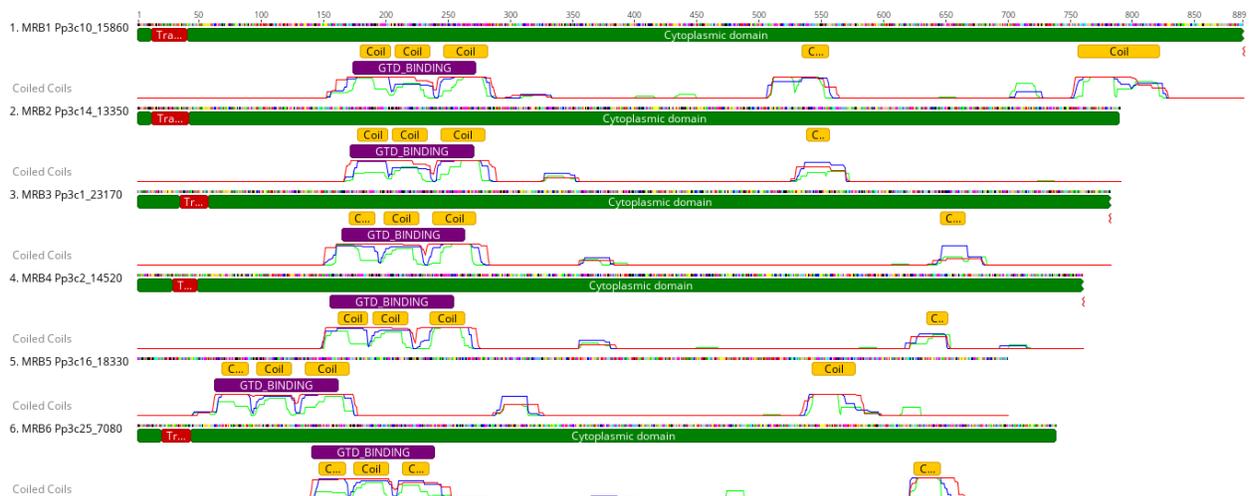


Figure R1. MRB protein with annotations and coiled-coils predictions made in Geneious Prime.

Protein alignment and phylogenetic tree construction show the grouping of the MRB proteins is similar to previously published work (Peremyslov, 2013), we were not able to identify members of the MRB family in Chlorophyte and Charophyte green algae. These results are consistent with the MRB gene family emerging in the common ancestor of all monophyletic terrestrial plants. Bryophytes are an early-diverging lineage of land plants, here we compared the

number of MRB genes in two bryophytes, *Marchantia polymorpha* and *P. patens*. Comparison between the one gene in *M. polymorpha* and the six MRB genes from *P. patens* are consistent with a series of duplications of one original ancestral MRB gene (Figure R2).

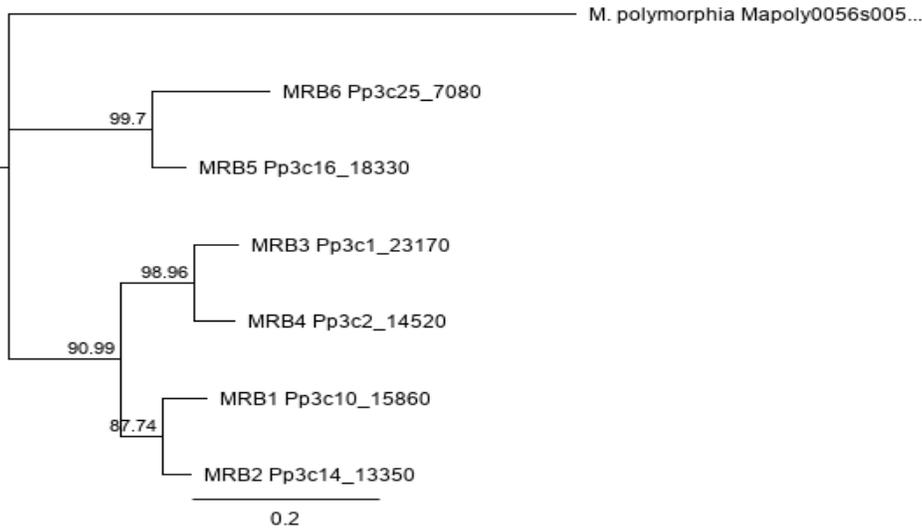


Figure R2. Phylogenetic tree of the MRB gene family in *P. patens*. Sequences were aligned using CLUSTAL and the tree generated using MrBayes in Geneious Prime.

Because of how early *P. paten* diverged as a lineage, many adaptations of land plants, which emerged more closer to the present time are not present in *P. patens*. Nevertheless, many microRNA's and genes found in angiosperms have homologs in *P. patens*, contributing their function within the gametophyte dominant *P. patens* (Nishiyama et al., 2003; Arazi, 2012). Since the divergence of bryophytes, there has been an extensive expansion of myosins and their putative receptor proteins (MRBs) in vascular plants. This expansion was likely driven by adaptation to new and further growth on land. There is evidence that angiosperms went through two whole-genome duplication events during their evolutionary history (Magadum et al., 2013). To evaluate the phylogenetic relationship of the MRB proteins from *P. patens*, we performed an alignment with representatives from a variety of land plants (Figure R3). We found evidence of

expansion, which can be seen in the larger size of MRB gene families in higher plants. The proteome of *O. sativa* contains 14 proteins expressing a GTD binding domain, and there were 15 proteins found in *A. thaliana*. This is a dramatic increase in number compared to the number of MRB genes found in *M. polymorpha* and the lycopodium *S. moellendorffii* (Figure R3).

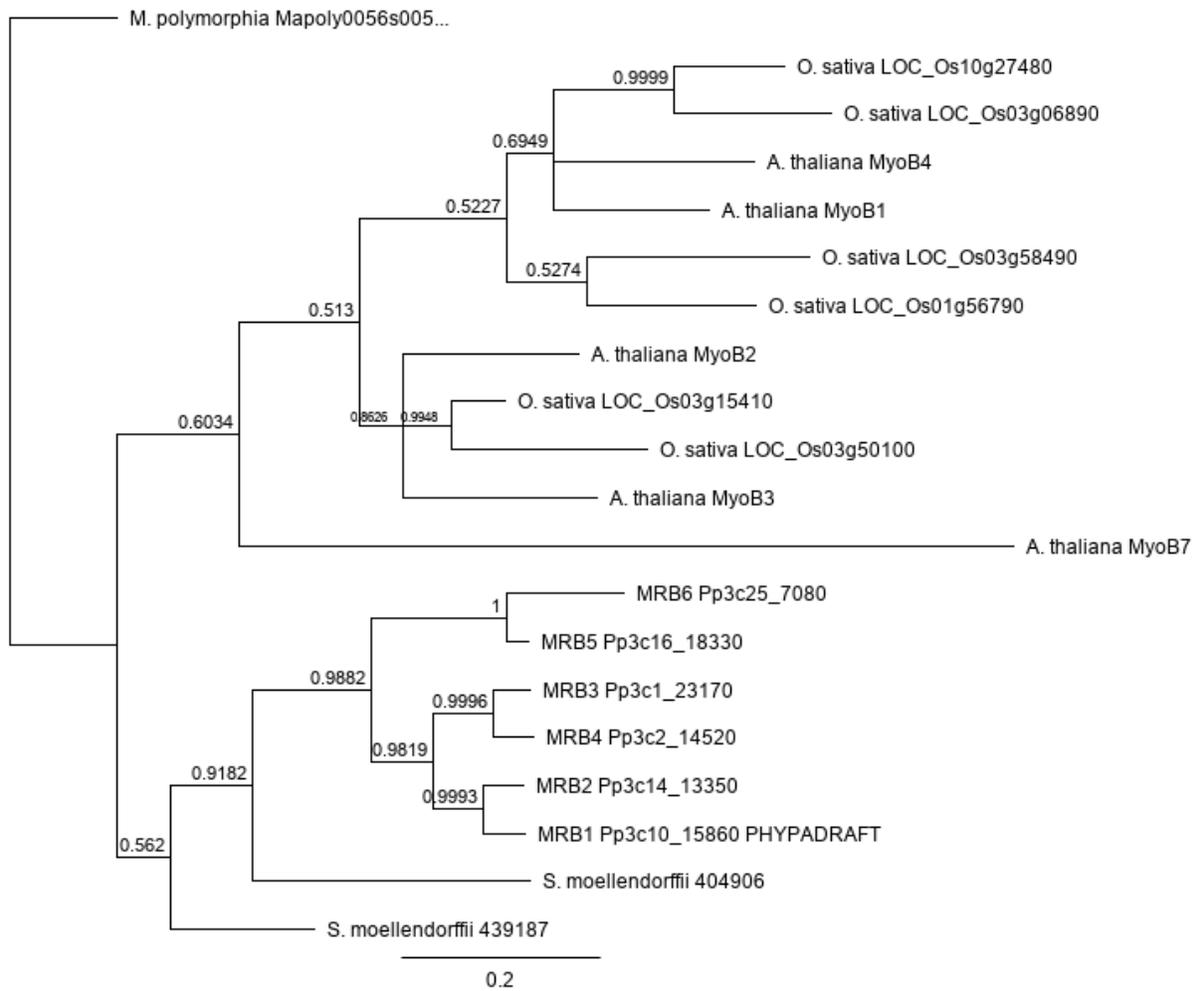


Figure R3. Phylogenetic tree of the MRB gene family in some land plant representatives. Sequences from the DUF593 domain were aligned using CLUSTAL and the tree generated using MrBayes in Geneious Prime.

Analysis of gene expression data

To gain further information about the possible function of the MRB proteins in *P. patens* and to decide what isoform to tag with a green fluorescent protein, we looked at the expression of all six genes in different tissues over the lifecycle of the *P. patens* (Ortiz-Ramirez et al., 2016).

Expression data was obtained using the Physcomitrella eFP Browser from University of Toronto. This website visualizes and interprets the data obtained through microarray gene expression profiling. By inputting the gene ID for the six MyoB proteins, we are able to see the relative expression of these genes over the entire lifecycle of the plant. MRB1, 3, and 4 are have highest expression in caulonemal cells (Figure R4). There is a possibility for functional overlap and redundancy if these three proteins bind to the same myosin-tail region. MRB3 and MRB4 have a similar expression pattern, and are closely related (Figure R3). Caulonemal cells grow on the tips of protonemal filaments within seven days of spore germination. Protoplasts from the moss *P. patens* regenerate directly into filamentous protonema. MRB1-3eGFP should have high expression in the protoplasts derived from stable transformations.

Each gene's greatest expression occurs in caudonema tissues, with the exception of MRB. MRB 6 instead has highest expression in the archegonia, a multicellular structure which contains the ovum of the gametophore



Figure R4: Absolute expression levels in *P. patens* tissue viewed at a signal threshold of 4763 in the *Physcomitrella* eFP Browser. The expression results were determined in Ortiz-Ramírez et al. The different tissue types were isolated from wildtype *P. patens* and sampled in triplicate. Expression data was obtained by hybridization on NimbleGen v1.6.

Fluorescent protein gene targeting and preliminary analysis

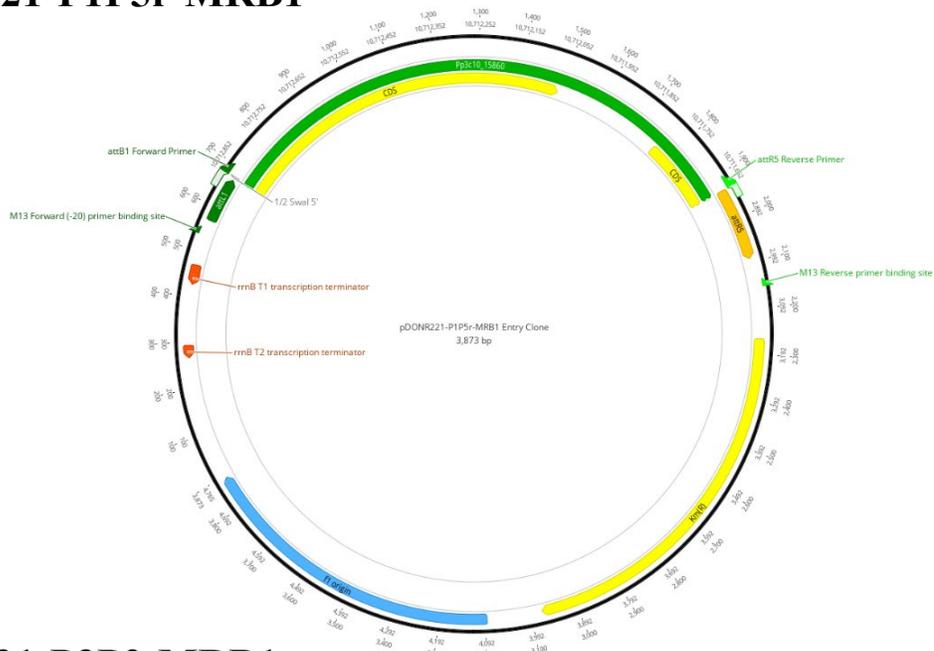


Figure R5: Annotated sequences of the four primers used for the gateway construct

Primer sequences for the amplification of the MRB1 sequence. These sequences code for both a specific attachment site for Gateway reactions, as well as the a 25-bp long sequence of the gene Pp3c10_15860 for primer binding.

The sequence of our primers were chosen in order to create a four fragment vector for recombinational cloning. Primers 1 and 2 targeted the 3' UTR, which contains a stop codon, of MRB1 for amplification (Figure R5). This PCR product was flanked by attB2 and attB3 sites. These sites were used to integrate the PCR product into the P3P2 clone (Figure R6). Primers 3 and 4, targeted our gene of interest, MRB1 for amplification (Figure R5). The PCR product resulting from these two primers contained the majority of the coding region for the MRB1 gene, flanked by the attb1 and attB5r sites. These sites were used to integrate the PCR product into the P1P5 clone (Figure R6).

pDONR221-P1P5r-MRB1



pDONR221-P3P2-MRB1

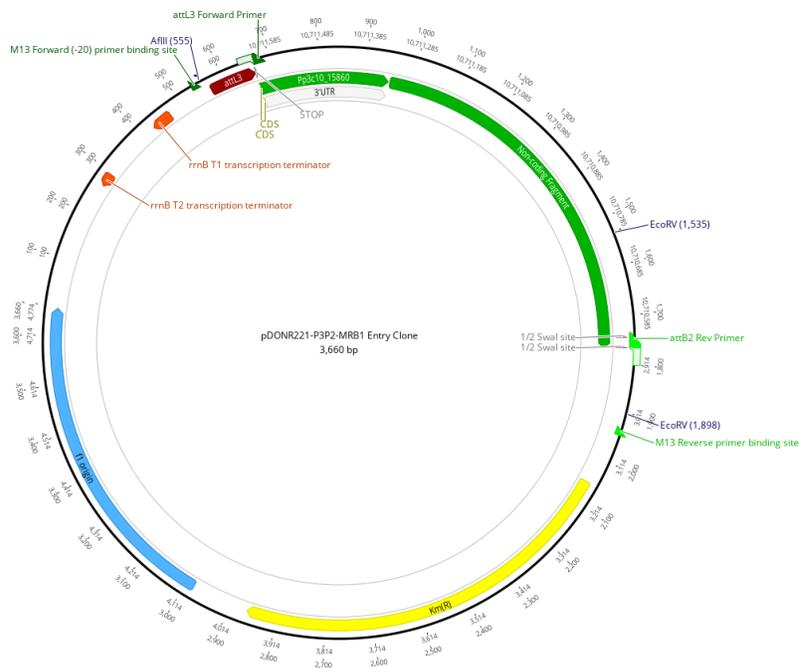
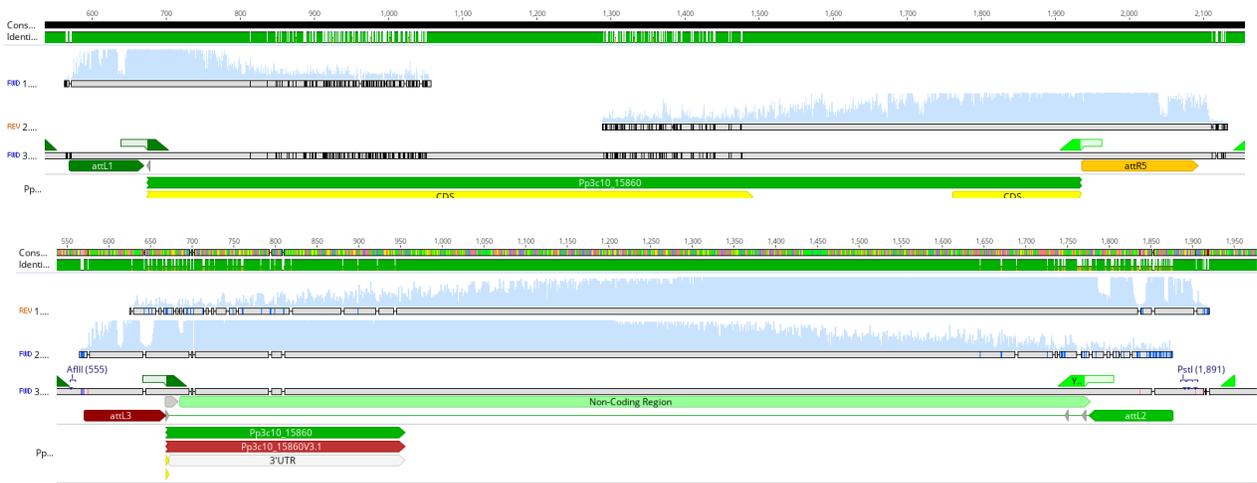


Figure R6: Diagram of entry clone plasmids.

PCR products amplified from wild-type *P. patens* were recombined into Gateway pDONR plasmids.

A.



B.

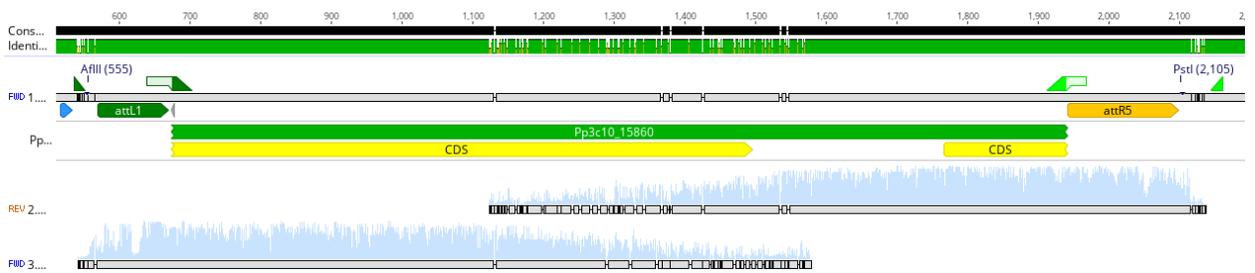


Figure R7: A. Initial sequencing results. B Resequenced P1P5

Sequencing was used to ensure that our gene of interest was successfully amplified from the *P. patens* genome.

DNA sequencing of the P1P5 and P3P2 clones ensured that the correct DNA regions within the *P. patens* genome were targeted (Figure R7).

The sequencing results for the P1P5 pENT plasmid did not overlap, resulting in an area of which's accuracy could not be confirmed. As the pENT clone was sent back to the lab for resequencing, which resulted in a complete read. As the P1P5 contained the majority of the coding sequence for the MRB1 gene, it was essential that it was accurately duplicated.

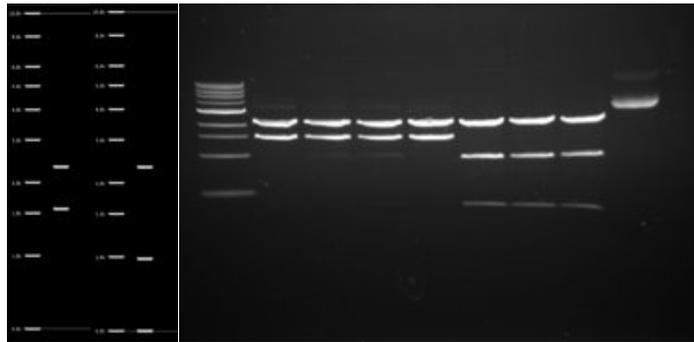


Figure R8: *PCR Results for BP entry clones digested with AfIII and EcoRV. The P1P5 clones are left-most while the P3P2 clones are on the right.*

Our BP entry clones were digested in order to confirm they contained the correct DNA sequences. The gel electrophoresis showed that the resulting bands were of correct length, validating our entry clones validity further (Figure R8).

Having validated that our P1P5 and P3P2 clones were accurate, we could then move onto incorporating these into a four-fragment transformation vector along with hygromycin and GFP fragments

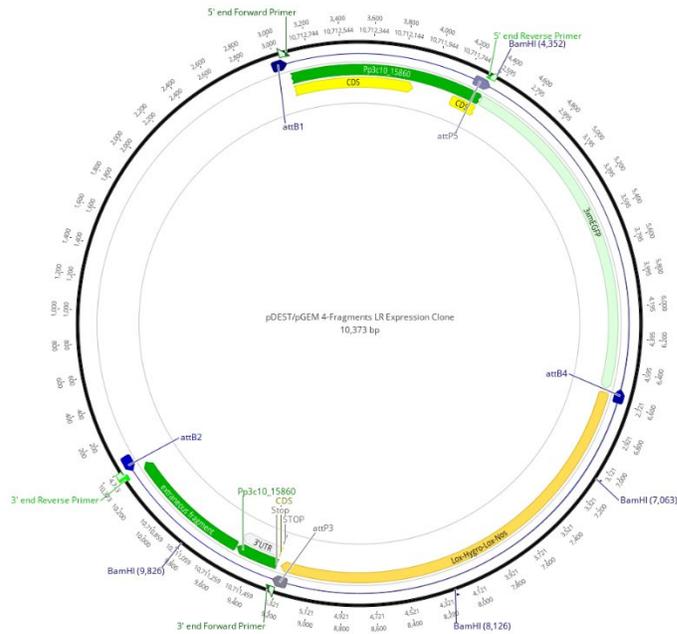


Figure R9: Diagram of the four-fragment pDONR clone.

The expression clone contained our two designed fragments, along with a triple GFP fragment and a fragment containing a hygromycin resistance gene.

The successful construction of our expression clone was confirmed through the use of a bamHI digest. This restriction enzyme was used to cut the MRB1-3EGFP pDONR clone at four specific locuses, producing four fragments of differing length. The length of the fragments from the gel electrophoresis results matched the length and number of fragments we expected (Figure 10) .

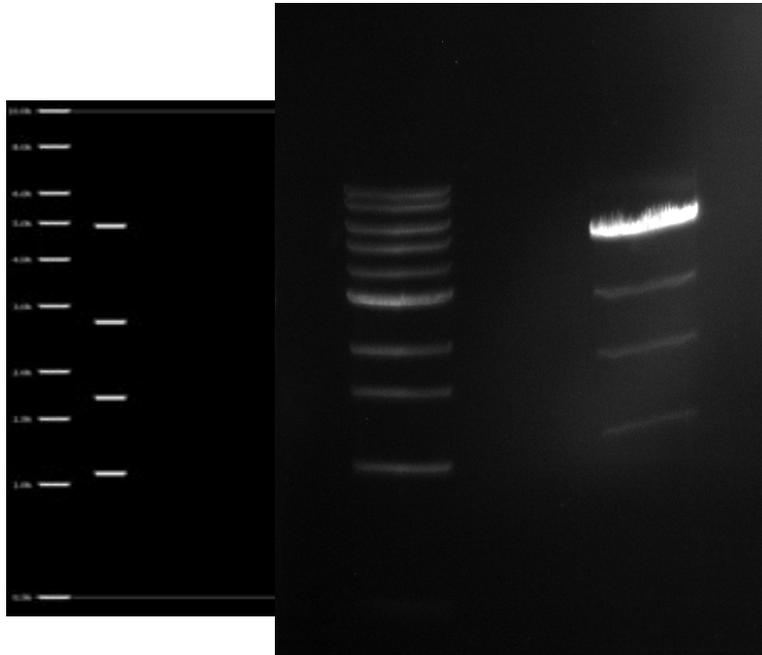


Figure R10: *Gel Electrophoresis results of the BamHI digest of the four-fragment construct.*

The stable moss lines resulting from targeting the 3mEGFP were screened using fluorescence microscopy. The strength of the fluorescence will be used to select the lines to further investigate with confocal microscopy and DNA genotyping (Figure R11).



Figure 6: *Masterplate 3m-EGFP* fluorescence. Images show background fluorescence.

Discussion

The MRB proteins contained the same structural features found in DUF593-containing proteins as reported in by Peremyslov et al. (Peremyslov, 2013). The conserved structures include the coiled-coil domain within the GTD-binding domain, two coiled-coil regions near the C terminus, and a predicted transmembrane helix at the N terminus. Myosin binding is likely to be a common property of the MRB protein family, as all 6 proteins contain a conserved GTD binding domain.

Peremyslov et al. grouped the MyoB family in *A. thaliana* into six different subfamilies based on differing domain architecture (Peremyslov, 2013). While the MyoB family has had notable variability in domain architecture, the domain architecture of the MRB proteins in *P. patens* is identical. The conservation of these domain structures leads us to believe that the six proteins likely perform one common function. This common function, possibly related to myosin binding, is likely essential, as these GTD-binding proteins are highly conserved in plants (Holding et al., 2007). In the liverwort *Marchantia polymorpha* this function is performed by a single protein, as its genome contains only one gene which codes for a GTD-binding domain. The *M. polymorpha* genome has been found to lack redundancy in regulatory genes relative to other sequenced land plants (Bowman et al., 2017). There is also evidence for a lack of ancient whole-genome duplications in *M. polymorpha*, which could have resulted in a greater number of GTD-binding proteins, and could be the case in other liverworts.

The phylogenetic analysis of the MRB genes grouped them into 3 closely related pairs. The 6 genes are likely the result of multiple duplication events, which would explain the phylogenetic closeness of each pair of genes (Rensing et al., 2008). Genome duplication events

allow for gene diversification and the development of new functionalization (Magadum et al., 2013). However, as the domain structures remain the same, we predict that all proteins likely share similar interactions and therefore have a similar molecular function in *P. patens*. The existence of multiple copies of a gene would allow for more sophisticated regulation of the expression of this family of genes. With six different genes coding for a myosin-binding protein, regulatory elements would be able to promote or repress MRB expression resulting, hence controlling when they interact with the same myosin XI at different developmental times and in different tissues.

Evidence for this can be drawn by the lack of overlap of MRB expression in *P. patens* tissue (Figure R4). All six MRB proteins are expressed highest in the haploid generations of *P. patens*. MRB 1, 2, and 6 have high expression in gametophore tissue, while 3, 4, and 5 in protonemal tissue. MRB 6 is uniquely expressed in archegonial gametophore tissue as well. Different versions of the protein performing a putative identical function are found within the same cell type. This suggests that MRB plays an essential developmental role in the growth of developing *P. patens* plants. The relative molecular simplicity of the myosin transport system in bryophytes such as *M. polymorpha* and *P. patens* contrasts with the significant expansion of myosin and its corresponding putative receptors in *A. thaliana*. *P. patens* has only two copies of myosin XI, which have been shown to be functionally redundant (Vidali et al., 2010). *A. thaliana* has 13 class XI myosins (Reddy and Day, 2001; Peremyslov et al., 2011), and 6 subfamilies of MyoB, with additional interactors it comprises what could be a very complex transport network (Kurth et al., 2017). Nevertheless, the presence of MRB representatives in bryophytes, suggests that the essential functions of these molecules existed in the common ancestor of all land plants.

The extant genomes of *P. patens* and other species, which diverged before the appearance of tracheophytes, offer a glimpse into the early evolution of terrestrial plants. These plants diverged when the haploid phase was morphologically more complex than the smaller diploid sporophyte. During the first 100 million years of terrestrial plant evolution, the diploid phase of the plant lifecycle grew to dominance (Bowman et al., 2017). The increase in morphological complexity of the diploid sporophyte body in higher plants has been shown to have resulted in part through the transfer of gametophytic developmental programs to the sporophyte generation (Shaw et al., 2011). The transcription factors which control root hair development in the sporophyte of *A. thaliana* were shown to be the same genes that control the development of rhizoid cells in the gametophyte of *P. patens* (Tam et al., 2015). It is possible that regulatory genes and developmental programs, which allowed for initial adaptation to the terrestrial environment, first appeared in the gametophyte generation and were later transferred to the sporophyte as its function expanded. Because the expression of MRB was found to be primarily in the haploid cycle, and due to its relatively simple and redundant functioning, we can speculate that an earlier form of the myosin transport system in *P. patens* was likely recruited as the basis of the expanded myosin transport system in higher plants.

References

- Arazi, T. (2012). MicroRNAs in the moss *Physcomitrella patens*. *Plant Mol Biol* **80**, 55-65. <https://doi.org/10.1007/s11103-011-9761-5>
- 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., Balasubramanian, S., Barry, K., Bauer, D., Boehm, C.R., Briginshaw, L., Caballero-Perez, J., Catarino, B., Chen, 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., Hiramawa, Y., Hundley, H.N., Ikeda, Y., Inoue, K., Inoue, S.I., 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., Rovekamp, 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., and Schmutz, J. (2017). Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. *Cell* **171**, 287-304 e215. <https://doi.org/10.1016/j.cell.2017.09.030>
- Holding, D.R., Otegui, M.S., Li, B., Meeley, R.B., Dam, T., Hunter, B.G., Jung, R., and Larkins, B.A. (2007). The maize flou1 gene encodes a novel endoplasmic reticulum protein involved in zein protein body formation. *Plant Cell* **19**, 2569-2582. <https://doi.org/10.1105/tpc.107.053538>
- Kurth, E.G., Peremyslov, V.V., Turner, H.L., Makarova, K.S., Iranzo, J., Mekhedov, S.L., Koonin, E.V., and Dolja, V.V. (2017). Myosin-driven transport network in plants. *Proc Natl Acad Sci U S A* **114**, E1385-E1394. <https://doi.org/10.1073/pnas.1620577114>
- Liu, Y.C., and Vidali, L. (2011). Efficient polyethylene glycol (PEG) mediated transformation of the moss *Physcomitrella patens*. *J Vis Exp* **50**, e2560. <https://doi.org/10.3791/2560>
- Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., and Ravikesavan, R. (2013). Gene duplication as a major force in evolution. *Journal of Genetics* **92**, 155-161. <https://doi.org/10.1007/s12041-013-0212-8>
- Menand, B., Calder, G., and Dolan, L. (2007). Both chloronemal and caulonemal cells expand by tip growth in the moss *Physcomitrella patens*. *J Exp Bot* **58**, 1843-1849. <https://doi.org/10.1093/jxb/erm047>

- Nishiyama, T., Fujita, T., Shin-I, T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y., and Hasebe, M. (2003). Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: Implication for land plant evolution. *P Natl Acad Sci USA* **100**, 8007-8012. <https://doi.org/10.1073/pnas.0932694100>
- Ortiz-Ramirez, C., Hernandez-Coronado, M., Thamm, A., Catarino, B., Wang, M., Dolan, L., Feijo, J.A., and Becker, J.D. (2016). A transcriptome atlas of *Physcomitrella patens* provides insights into the evolution and development of land plants. *Mol. Plant* doi: 10.1016/j.molp.2015.12.002. <https://doi.org/10.1016/j.molp.2015.12.002>
- Peremyslov, V.V., Mockler, T.C., Filichkin, S.A., Fox, S.E., Jaiswal, P., Makarova, K.S., Koonin, E.V., and Dolja, V.V. (2011). Expression, splicing, and evolution of the myosin gene family in plants. *Plant Physiol.* **155**, 1191-1204. <https://doi.org/10.1104/pp.110.170720>
- Peremyslov, V.V.M., E.A.; Kurth, E.G.; Makarova, K.S.; Koonin, E.V.; Dolja, V.V. (2013). Identification of myosin XI receptors in *Arabidopsis* defines a distinct class of transport vesicles. *Plant Cell* **25**, 3022-3038.
- Reddy, A.S., and Day, I.S. (2001). Analysis of the myosins encoded in the recently completed *Arabidopsis thaliana* genome sequence. *Genome Biol* **2**, research0024-0024.0017.
- Rensing, S.A., Goffinet, B., Meyberg, R., Wu, S.-Z., and Bezanilla, M. (2020). The Moss *Physcomitrium (Physcomitrella) patens*: A Model Organism for Non-Seed Plants. *The Plant Cell* **32**, 1361-1376.
- Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y., Tanahashi, T., Sakakibara, K., Fujita, T., Oishi, K., Shin, I.T., Kuroki, Y., Toyoda, A., Suzuki, Y., Hashimoto, S., Yamaguchi, K., Sugano, S., Kohara, Y., Fujiyama, A., Anterola, A., Aoki, S., Ashton, N., Barbazuk, W.B., Barker, E., Bennetzen, J.L., Blankenship, R., Cho, S.H., Dutcher, S.K., Estelle, M., Fawcett, J.A., Gundlach, H., Hanada, K., Heyl, A., Hicks, K.A., Hughes, J., Lohr, M., Mayer, K., Melkozernov, A., Murata, T., Nelson, D.R., Pils, B., Prigge, M., Reiss, B., Renner, T., Rombauts, S., Rushton, P.J., Sanderfoot, A., Schween, G., Shiu, S.H., Stueber, K., Theodoulou, F.L., Tu, H., Van de Peer, Y., Verrier, P.J., Waters, E., Wood, A., Yang, L., Cove, D., Cuming, A.C., Hasebe, M., Lucas, S., Mishler, B.D., Reski, R., Grigoriev, I.V., Quatrano, R.S., and Boore, J.L. (2008). The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**, 64-69.
- Schaefer, D.G., and Zryd, J.P. (1997). Efficient gene targeting in the moss *Physcomitrella patens*. *Plant Journal* **11**, 1195-1206. <https://doi.org/10.1046/j.1365-313X.1997.11061195.x>
- Shaw, A.J., Szovenyi, P., and Shaw, B. (2011). Bryophyte diversity and evolution: windows into the early evolution of land plants. *Am J Bot* **98**, 352-369. <https://doi.org/10.3732/ajb.1000316>

- Stephan, O., Cottier, S., Fahlen, S., Montes-Rodriguez, A., Sun, J., Eklund, D.M., Klahre, U., and Kost, B.** (2014). RISAP is a TGN-associated RAC5 effector regulating membrane traffic during polar cell growth in tobacco. *Plant Cell* **26**, 4426-4447. <https://doi.org/10.1105/tpc.114.131078>
- Taiz, L., Zeiger, E., Møller, I.M., and Murphy, A.S.** (2018). *Fundamentals of plant physiology*. (New York, NY: Published in the United States of America by Oxford University Press).
- Tam, T.H., Catarino, B., and Dolan, L.** (2015). Conserved regulatory mechanism controls the development of cells with rooting functions in land plants. *Proc Natl Acad Sci U S A* **112**, E3959-3968. <https://doi.org/10.1073/pnas.1416324112>
- Tominaga, M., Kimura, A., Yokota, E., Haraguchi, T., Shimmen, T., Yamamoto, K., Nakano, A., and Ito, K.** (2013). Cytoplasmic streaming velocity as a plant size determinant. *Dev Cell* **27**, 345-352. <https://doi.org/10.1016/j.devcel.2013.10.005>
- Vidali, L., Burkart, G.M., Augustine, R.C., Kerdavid, E., Tüzel, E., and Bezanilla, M.** (2010). Myosin XI is essential for tip growth in *Physcomitrella patens*. *Plant Cell* **22**, 1868-1882. <https://doi.org/10.1105/tpc.109.073288>