博士論文

Characterization of small RNAs in the liverwort, *Marchantia polymorpha*

（苔類ゼニゴケ（*Marchantia polymorpha*）における small RNA の機能解析）

都筑 正行

Contents

Contents 1

Abbreviations 2

General Introduction 4

Figures for Introduction (I-1 to I-3) 13

Chapter I 16

Introduction 17

Materials and Methods 20

Results 28

Discussion 44

Figures 1-13 51

Tables 1-3 64

Chapter II 69

Introduction 70

Materials and Methods 74

Results 78

Discussion 86

Figures 14-23 90

Table 4 100

Supplementary Figures and Table 102

Concluding remarks 108

References 110

Acknowledgements 127

Abbreviations

AGO	ARGONAUTE
BLAST	basic local alignment search tool
Cas9	CRISPR associated protein 9
CRISPR	clustered regulatory interspaced short palindromic repeats
DCL	DICER-LIKE
DNA	deoxyribonucleic acid
dsRNA	double-stranded RNA
EST	expression sequence tag
gRNA	guide RNA
HSP90	heat shock protein 90
HYL1	HYPONASTIC LEAVES 1
JGI	Joint Genome Institute
MS	Murashige-Skoog
Mya	million years ago
miRNA	microRNA

OX	overexpression
PAM	protospacer adjacent motif
PTGS	posttranscriptional gene silencing
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
RdDM	RNA-dependent DNA methlylation
RDR	RNA-dependent RNA polymerase
RISC	RNA-induced silencing complex
rRNA	ribosomal RNA
SE	SERRATE
siRNA	short interfering RNA
snRNA	small nuclear RNA
ssRNA	single-stranded RNA
TAIR	The Arabidopsis Information Resource
UTR	untranslated region

General Introduction

-Small RNAs

It is essential for multicellular organisms to regulate positional and temporal patterns of gene expression to accomplish their development properly. Small RNA pathway is one of the important gene regulatory mechanisms having significant roles in eukaryote development (Lee *et al*. 1993, Hamilton and Baulcombe 1999, Volpe *et al*. 2002). MicroRNA (miRNA) was the first discovered functional small RNA. By the genetic work using *Caenorhabditis elegans*, it was shown that *lin-4* loss-of-function mutant lost the appropriate developmental timing and displayed up-regulated expression of LIN-14 protein. It was subsequently shown that *C. elegans lin-4* locus encodes miRNA having the complementary sequence to the repeated sequence element in 3'UTR of *lin-14* mRNA (Lee *et al.* 1993).

In plants, the phenomenon of gene suppression by overexpression of transgenes was reported earlier than metazoans. The study using *Petunia hybrida* reported that introduction of chalcone synthese gene, which encodes the key enzyme in flavonoid biosynthesis, under the direction of cauliflower mosaic virus 35S promoter induced unexpected suppression of anthocyanin biosynthesis (Napoli et al. 1990). This phenomenon is called "co-suppression",

and it was suspected that a nucleotide sequence-dependent mechanism is involved. However, the expression of small regulatory RNAs was not confirmed until Hamilton and Baulcombe (1999) reported that small antisense RNAs to the transgenes were induced in tomato and *Nicotiana benthamiana* plants. After the discovery of small RNA expression in plants, the biogenesis pathways and the functional manners of small RNAs have been investigated with *Arabidopsis thaliana*, a model land plant. It is now widely accepted that small RNAs are classified into microRNAs (miRNAs) and short-interfering RNAs (siRNAs). Furthermore, siRNAs are subdivided into those for posttranscriptional gene silencing (PTGS) and for RNA-dependent DNA methylation (RdDM) in plants.

Small RNAs are generally 20-25 nt in length and produced from longer precursor RNAs. RNase III-like enzymes like Dicer proteins are key processors in small RNA biogenesis. In plants, four DICER-LIKE (DCL) proteins named DCL1-4 are involved in small RNA biogenesis. DCL1 and its co-factors, HYPONASTIC LEAVES 1 (HYL1) and SERRATE (SE), recognize the imperfect fold-back structure of the primary transcript of miRNA (pri-miRNA) transcribed from miRNA-encoding region in the genome and cleaves pri-miRNA into the precursor of miRNA (pre-miRNA) forming a stem loop structure

(Kurihara and Watanabe 2004, Kurihara *et al.* 2006) (Fig. I-1). DCL1 sequentially cleaves pre-miRNA into a pair of mature miRNA/miRNA*. On the other hand, DCL2-4 recognize double-stranded RNAs (dsRNAs) as substrates and cleave them into siRNAs with specific sizes (Gasciolli *et al.* 2005). dsRNAs are synthesized from single-stranded RNAs (ssRNAs) recognized by RNA-dependent RNA polymerase 1-6 (RDR1-6) in *A. thaliana*.

Mature small RNAs need to be sorted into ARGONAUTE (AGO) proteins that are effector proteins in small RNA pathways to execute regulatory roles. A pair of AGO protein and small RNA forms the RNA-induced silencing complex (RISC) and works as an effector machinery guided by small RNA in a nucleotide sequence-dependent manner. *A. thaliana* has 10 AGO proteins (AGO1-10). AGO1 is a main effector protein in PTGS by binding to target mRNAs of miRNA or to template RNAs of siRNA. In most cases, AGO1 slices the target RNA in the middle of complementary binding site by the endonuclease activity. AGO4 protein is essential for RdDM pathway to execute properly. AGO4 harbors 24 nt siRNA and mediates *de novo* cytocine methylation by DNA methylase 2 (DRM2) (Cao and Jacobsen 2002, Qi et al. 2006).

-Evolutionally conserved miRNAs and their function in land plants

As with protein-coding genes, some miRNAs are conserved in a certain taxon of plants (Cuperus *et al.* 2011). Until now, miRNAs found in 74 species of land plants (eudicots, *A. thaliana*; monocots, *Oryza sativa*; gymnosperms, *Pinus taeda*; lycopods, *Selaginella moellendorffii*; mosses, *Physcomitrella patens*, etc.) and in a chlorophyte, *Chlamidomonas reinhardii*, were registered in the miRNA database, miRBase (http://www.mirbase.org/). Such collection of miRNA sequences in various species showed that some *MIRNA* genes are shared among land plant species, while no miRNAs are shared between *C. reinhardii* and land plant species (Zhao *et al.* 2007). Eight miRNAs (miR156, 319/159, 160, 166/165, 171, 390, 395, 408) are considered as those conserved in most land plants because their expression was confirmed in both the angiosperm *A. thaliana* and the moss *P. patens* (Axtell *et al.* 2007). Surprisingly, target genes of these miRNAs belong to the common families among different land plant species as well (Axtell *et al.* 2007). These results indicate that miRNA and its target genes have co-evolved, keeping the miRNA/target complementary sequences.

Until now, many studies have shown that conserved miRNAs affect plant development and cell differentiation by repression of the target gene expression. miR156 decreases the

expression of SQUAMOSA PROMOTER BINDING-LIKE (SPL) transcription factor genes and delays the flowering time in *A. thaliana* (Wu and Poethig 2006). miR319/159 family suppresses MYB and TCP transcription factor genes and affects leaf morphogenesis in *A. thaliana* and its relative *Cardamine hirsuta* (Palatnik *et al.* 2003, Allen *et al.* 2007, Rubio-Somoza *et al.* 2014). miR166/165 targets class III homeodomain-leucine zipper (HD-ZIPIII) transcription factor genes and involves both leaf dorsoventral patterning and root cell patterning in *A. thaliana* (McConnel *et al*. 2001, Emily *et al*. 2003). Loss of the miR166/165 complementary region in the HD-ZIPIII gene induces abnormally adaxialized leaves forming a rod-like or trumpet-like structure (McConnel and Barton 1998).

Meanwhile, there are few insights about the roles of gene regulation by miRNAs in diverse morphogenesis and life cycles in land plants. It is considered that land plants and algae diverged at least ~450 Mya (Bowman 2013). Bryophytes including liverworts, mosses and hornworts are first diverged land plants in the lineage (Fig. I-2). Phylogenetic relationship between them is still debated (Qiu *et al*. 2006. Wickett *et al*. 2014). Their main ploidy in the life cycle is haploid and they do not have vascular organs. These characteristics of bryophytes are quite different from those of seed plants. Additionally, even among three bryophyte

taxons, their body plans are different. Whereas moss gametophores have a radial body pattern with non-vascular leaves, most liverworts have flat and prostrate thallus corresponding to leaf and shoot. There are several studies investigating roles of miRNAs in bryophyte development using *P. patens*. Cho *et al.* (2012) revealed that miR156 suppresses expression of *SPL* genes and formation of buds, the primordia of gametophytes, in *P. patens*. The other work showed that miR534a also represses the number of buds (Saleh *et al.* 2011). However, there are no studies about the roles of the other conserved miRNAs in bryophytes. In addition, universal functions of conserved miRNAs and target transcription factor genes in land plant lineages remain unknown.

-*Marchantia polymorpha*

Marchantia polymorpha is a common liverwort and has been frequently used from ancient times as a model organism for plant morphology because of its unique shape (Bowman 2015). *M. polymorpha* has distinct male and female plants and both asexual and sexual reproduction cycles (Fig. I-3). It is experimentally revealed that far-red (FR) light and long day condition induces differentiation of thalli into umbrella-like sexual organs, antheridiophore in male and

archegoniophore in female, which produce sperms and eggs, respectively (Chiyoda *et al*. 2008, Kubota *et al*. 2014).

Recently, *M. polymorpha* is focused as a new model organism of land plants in molecular genetics and comparative genomics. Now the genome sequencing project by Joint Genome Institute (JGI) (http://genome.jgi.doe.gov/Marpha/Marpha.info.html) is in progress and various genetic tools are developed mainly by Prof. Takayuki Kohchi lab (Kyoto University, Japan) and Prof. John L. Bowman lab (Monash University, Australia). It has been revealed that *M. polymorpha* has eight autosomes and one sex chromosome either X (female) or Y (male). Whole size of *M. polymorpha* genome is estimated at about 280 Mb (Yamato and Kohchi 2012). Now assembly of genome sequences was almost completed and gene annotation is being performed by JGI. Now the *M. polymorpha* draft genome sequence is assembled as ver.2.0 and has 2955 scaffolds and 4347 contigs. 93.7% of the main genome is covered by >50 kB scaffolds. EST database were also assembled from transcriptome data obtained by RNA-seq.

As the draft genome and EST sequences were revealed, it has been unclosed that *M. polymorpha* genome has most of the homologous genes found in seed plants like *A. thaliana*,

but in less redundancy, especially as to transcription factor family genes (Zobell *et al*. 2010, Ueda *et al*. 2012, Kanazawa *et al*. 2015, Kato *et al*. 2015, Flores-Sandoval *et al*. 2015b, Eklund *et al*. 2015). It is estimated that *M. polymorpha* has only ~300 transcription factor genes, while *A. thaliana* has ~2000 (personal communication with Dr. Kohchi's lab). Regarding this point, it is expected that use of *M. polymorpha* gives information not only on function of a certain factor through land plant evolution but also on novel gene or protein network, which is universal among land plants. It is known that plant *MIRNA* genes are frequently multiplicated in the genome. Because of this, there are only a few miRNA mutants even in *A. thaliana* mutant lines (Allen *et al*. 2007). Also, miRNA regulation networks in *A. thaliana* are too complex to study in detail for this reason. It is expected that simplicity of gene sets in *M. polymorpha* genome brings new insights in miRNA regulation networks universal in land plants. Molecular genetic experimental systems that use the liverwort *M. polymorpha* as a model organism have been also recently established. These include growth conditions for phase transition, an *Agrobacterium*-mediated transformation method, a homologous recombination method and genome-editing tools (Chiyoda *et al*. 2008, Ishizaki *et al*. 2008, 2013, 2015a, Kubota *et al*. 2013, Sugano et al. 2014). Several studies using these

tools have been reported until now (Ishizaki et al. 2013, Kubota et al. 2014, Kanazawa *et al*. 2015, Kato *et al*. 2015, Flores-Sandoval *et al*. 2015b, Eklund *et al*. 2015).

-Aims of this research

In this research, I analyzed small RNAs of *M. polymorpha*, especially miRNAs, to identify still unknown roles of plant miRNAs in plant evolution. First, using high-throughput sequencing technology, I profiled small RNAs expressed in three different organs of *M. polymorpha* and characterized them. Then, using the profiling data, I investigated developmental roles of conserved miRNAs in *M. polymorpha* by genetic approaches. These results provided new insights into evolution and function of small RNAs in plants.

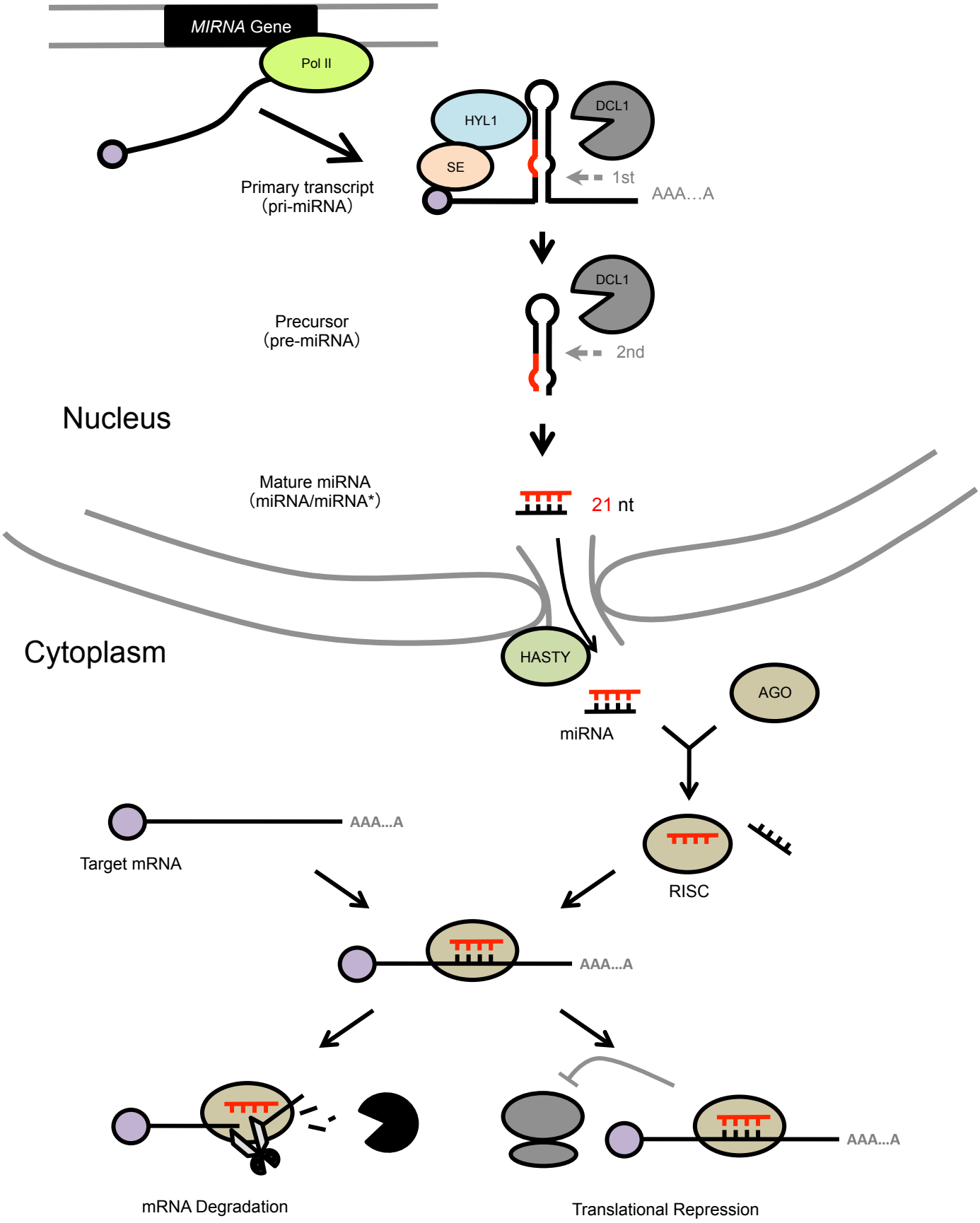


Fig. I-1 MicroRNA pathway in biogenesis and function in plants
Schematic diagram of microRNA pathway from its biogenesis to functioning.

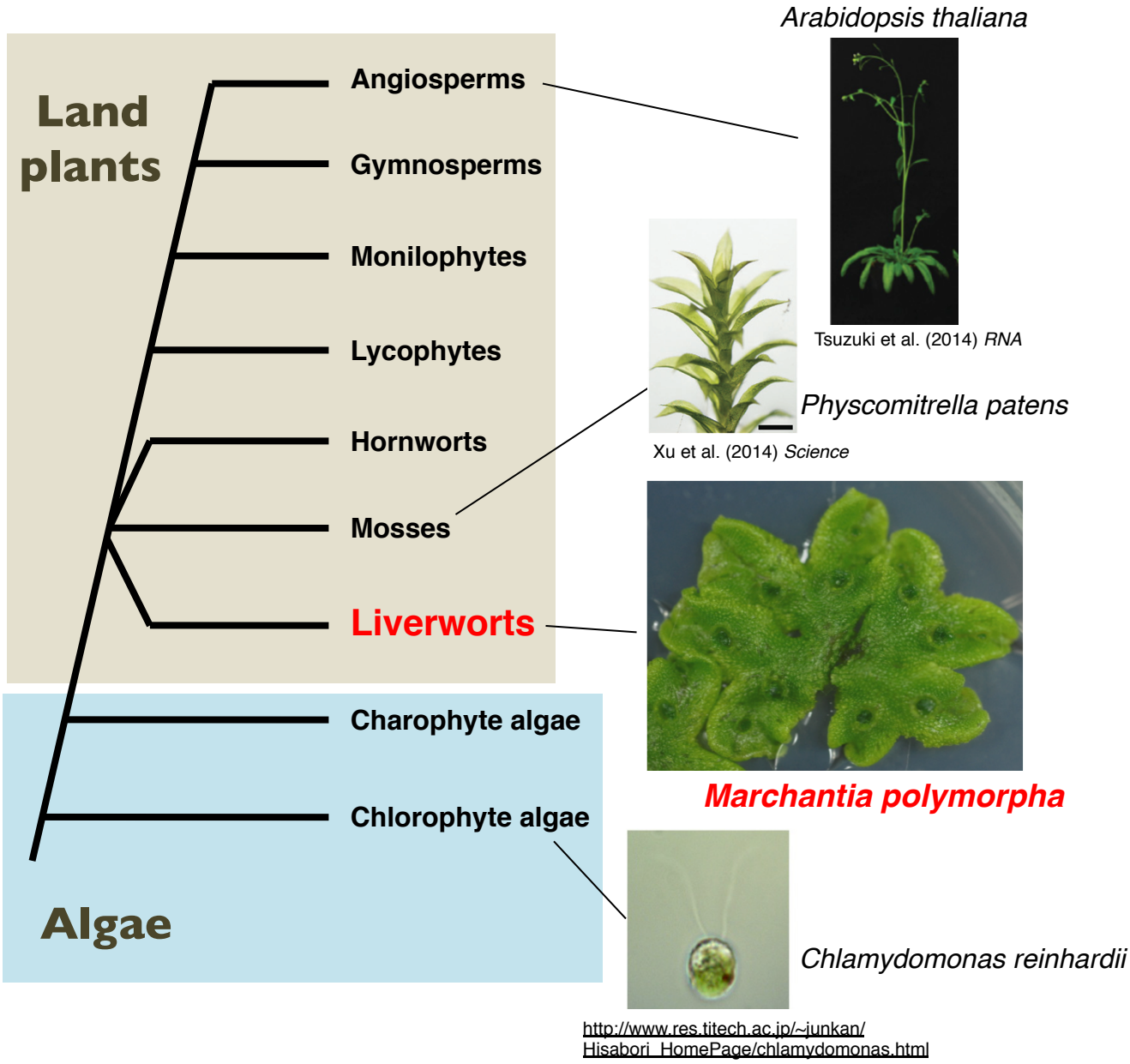


Fig. I-2 Evolution of land plants
A phylogenetic tree showing diversification of plant lineage from algae to land plants. Pictures are the representatives of a certain taxon.

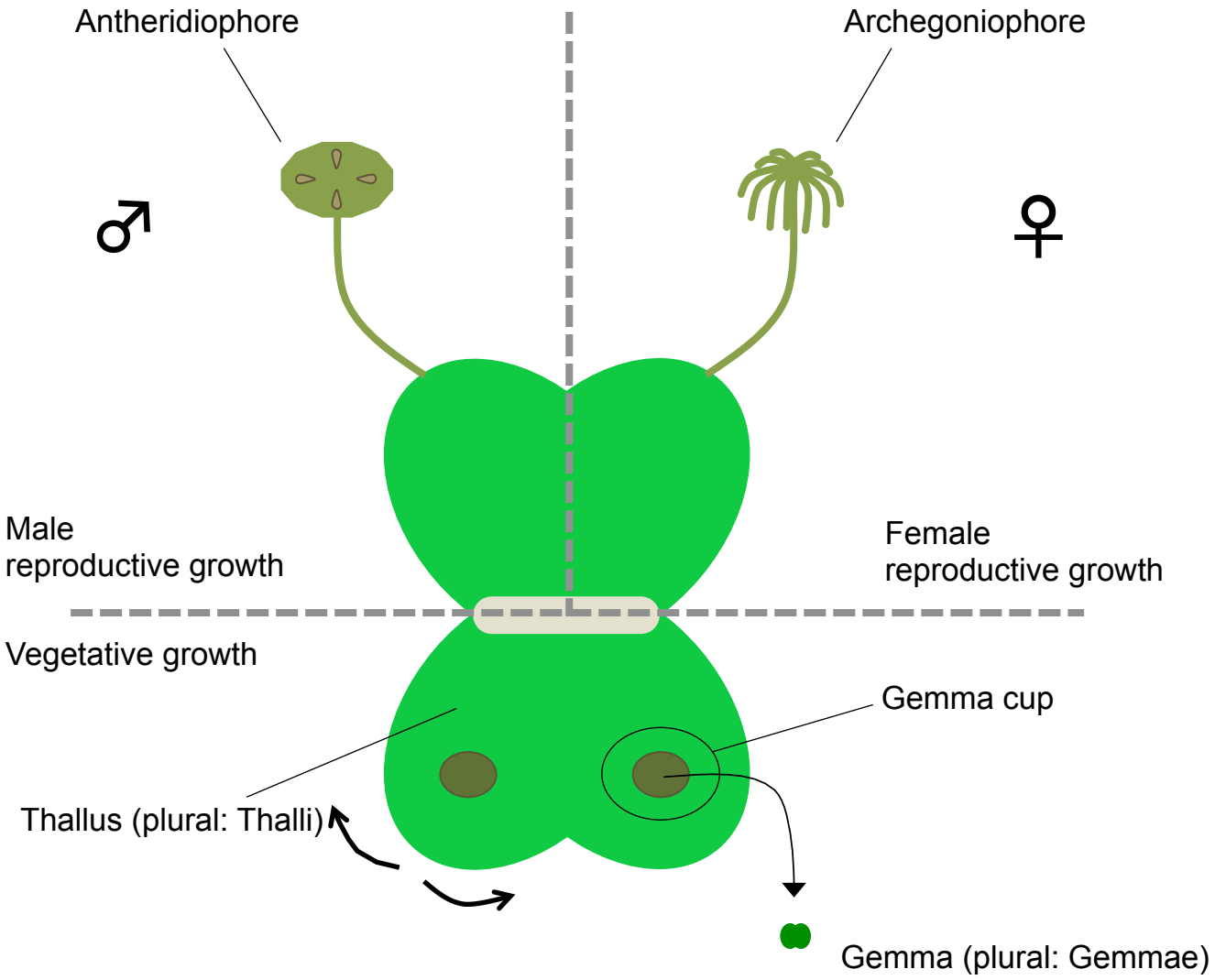


Fig. I-3 Growth phases and body parts of *Marchantia polymorpha*
Representation of the body plan of M. polymorpha in both vegetative and reproductive growth phases.

Chapter I

Profiling and characterization of small RNAs expressed in *M. polymorpha*

Introduction

Information of nucleotide sequences of an organism is fundamental and informative for further molecular biological analyses. High-throughput sequencing technology has been advanced after emerging of next generation sequencers (NGS) (Buerman and den Dunnen 2014). NGS makes it possible to obtain draft genome sequences of a certain organism, population of whole transcripts and RNA/DNA fragments binding to a certain proteins. Sequencing and characterization of small RNAs of short length are quite compatible with NGS, especially when Illumina short-read system is used (http://www.illumina.com/techniques/sequencing.html). When I focus on miRNA of different origins, miRNAs have been identified in various classes of land plants like *A. thaliana* (dicot), *Oryza sativa* (monocot), *Pinus taeda* (gymnosperm), *Sellaginella moellendorffii* (Lycopodiophyta), *Physcomitrella patens* (moss) (Rajagopalan *et al*. 2006, Fahlgren *et al*. 2007, Lu *et al*. 2008, Morin *et al*. 2008, Axtell *et al*. 2007) and registered in publicly available database, miRBase (http://www.mirbase.org).

Based on such sequencing data, we can now compare miRNA sequences between many land plants with evolutionary diversity. It has been suggested that some miRNA sequences

and their binding complementary sequences in target mRNAs are highly conserved in land plant lineages (Axtell 2008, Cuperus *et al*. 2011). One of these conserved miRNA families, miR166/165, is known to be present from bryophytes to angiosperms (Floyd and Bowman 2004). miR166/165 represses the expression of class III homeodomain-leucine zipper (HD-ZIPIII) transcription factor genes, which are quite involved in development, such as leaf dorsoventral patterning in *A. thaliana* (McConnell *et al*. 2001). miR156, another well-known conserved miRNA, targets mRNAs of SQUAMOSA PROMOTER BINDING-LIKE (SPL) transcription factors in various land plants. It was also shown that miR156 modulates plant growth phase transitions in both *A. thaliana* and the moss *P. patens* (Wu and Poethig 2006, Cho *et al*. 2012). Contrary to high conservation level of these miRNAs in land plants, *Chlamydomonas reinhardii*, whose genome sequences were revealed, has its own miRNAs but does not have miRNAs shared with land plants (Zhao *et al*. 2007). Because miRNAs in charophyte algae have not profiled yet, it remains unknown whether conserved miRNAs in land plants emerged before or after terrestrial adaptation.

To refine small RNA sequence data and their expression profile by excluding unrelated sequences, the sequence data of the organism are essential as the reference sequences. Now

that genome information of *M. polymorpha* is available, I performed high-throughput sequencing analysis and profiled small RNAs expressed in *M. polymorpha*. As a result of computational analysis of raw sequence data referring to available genome sequence data, in collaboration with Drs. Kohchi's and Bowman's groups, I found that *M. polymorpha* has eight so called evolutionary conserved miRNAs and hundreds of non-conserved miRNAs. It was also possible that corresponding target genes were also predicted. The results suggested that miRNA/target relationships in conserved miRNAs were also conserved in *M. polymorpha* like other land plants. My results add novel knowledge to discuss the evolution of miRNA-mediated regulation system.

Materials and Methods

-Plant materials and growth conditiosn

Takaragaike-1 (Tak-1: male) and Takaragaike-2 (Tak-2: female) ecotypes were used as wild-type (WT) plants of *M. polymorpha* (Ishizaki *et al*. 2008). They were kindly provided by Dr. Kohchi's group (Kyoto Univ.). Plants were grown asexually on half-strength Gamborg's B5 medium (Gamborg *et al*. 1968) solidified with 1% agar, under continuous white fluorescent light at 22°C. To obtain samples for small RNA sequencing, formation of sexual organs by Tak-1 and Tak-2 were induced by far-red light with FR light-emitting diodes (VBL-TFL600-IR730*, Valore; peak emission at 733 nm) at 22°C (Kubota *et al*. 2014).

Columbia-0 (Col-0) accession plants were used as wild-type *A. thaliana* as controls. They were grown on half-strength Murashige–Skoog (MS) medium containing 1% agar under continuous white fluorescent light at 22°C.

-Homology search of nucleotide and amino acid sequences

To identify *Marchantia* genes, the draft genome sequences of *M. polymorpha* ver.2.0 by Joint Genome Institute (JGI) sequencing project (http://www.jgi.doe.gov/) and EST database were

used as reference sequences. All of these sequence data were originally provided from Prof. Kohchi's and Prof. Bowman's groups. When protein-coding genes were searched, the protein-coding sequences from *A. thaliana* TAIR10 (http://www.arabidopsis.org/) were used as queries. Then, I ran TBLASTN to query sequences and the *M. polymorpha* EST database, and the best scoring sequences were determined as candidates for homologous genes of queries. To search the presence of homologous miRNA sequences, mature sequences of miRNAs in *A. thaliana* from miRBase (http://www.mirbase.org/) were used as reference sequences. Using the sequences of *M. polymorpha* mature miRNAs, which were the most abundant sequences mapped to each *MIRNA* locus identified by ShortStack, as queries, BLASTN with the "blastn-short" option was performed against the data of mature miRNA sequences in all organisms deposited in miRBase.

-*Molecular phylogenetic analysis*

All nucleotide and amino acid sequences of *M. polymorpha* were obtained from both *M. polymorpha* draft genome sequence ver.2.0 and the EST database. Amino acid sequences of *A. thaliana*, *O. sativa*, *S. moellendorffii*, *P. patens* and *C. reinhardii* used for phylogenetic tree

construction are referred from Axtell *et al.* (2007) and obtained from TAIR (https://www.arabidopsis.org) and Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). Gene model names and/or accession numbers used for alignments and phylogenic trees of DCL proteins are as follows: AtDCL1, At1g01040; AtDCL2, At3g03300; AtDCL3, At3g43920; AtDCL4, At5g20320; OsDCL1, Os03g02970; OsDCL2a, Os03g38740; OsDCL2b, Os09g14610; OsDCL3a, Os01g68120; OsDCL3b, Os10g34430; OsDCL4, Os04g43050; SmDCL1, Sm86110; SmDCL3a, Sm74321; SmDCL3b, Sm444049; SmDCL3c, Sm448444; PpDCL1a, Phypa1_1 205895; PpDCL1b, Phypa1_1 163666; PpDCL3, Phypa1_1 130425; PpDCL4, Phypa1_1 234670; CrDCL1, Cre02.g141000. Those of AGO proteins used for phylogenic analysis are as follows: At AGO1, At1g48410; At AGO2, At1g31280; At AGO3, At1g31290; At AGO4, At2g27040; At AGO5, At2g27880; At AGO6, At2g32940; At AGO7, At1g69440; At AGO9, At5g21150; At AGO10, At5g43810; Os01g16870; Os02g07310; Os02g45070; Os02g58490; Os03g33650; Os03g47820; Os03g47830; Os03g57560; Os03g58600; Os04g06770; Os04g47870; Os04g52540; Os04g52550; Os06g39640; Os06g51310; Os07g09020; Os07g16224; Os07g28850; Sm110151; Sm232189; Sm98227; Sm60551; Sm99227; PpAGO1a, Phypa1_1 200513; PpAGO1b, Phypa1_1

134255; PpAGO1c, Phypa1_1 117253; CrAGO1, Cre04.g214250; CrAGO2, Cre04.g214250. Those of SPL proteins are as follows; AtSPL1, AT2G47070; AtSPL2, AT5G43270; AtSPL3, AT2G33810; AtSPL4, AT1G53160; AtSPL5, AT3G15270, AtSPL6, AT1G69170; AtSPL7, AT5G18830; AtSPL8, AT1G02065; AtSPL9, AT2G42200; AtSPL10, AT1G27370; AtSPL11, AT1G27360; AtSPL12, AT3G60030; AtSPL13a, AT5G50570; AtSPL13b, AT5G50670; AtSPL14, AT1G20980; AtSPL15, AT3G57920. For construction of the phylogenetic tree of SPL proteins, 74–amino acid residues of the SBP domain was used (Xing *et al*. 2010).To search DCL and AGO gene candidates in *M. polymorpha*, amino acid sequences of *A. thaliana* DCL and AGO proteins were used as a query in TBLASTN. All phylogenetic analyses were performed by MEGA 6 (Tamura *et al*. 2013). For construction of multiple sequence alignment of amino acids, MUSCLE implemented in MEGA was used with default parameters (Edger 2004). Phylogenetic trees were constructed by neighbor-joining method with 1000 times replication of bootstrapping. The T-coffee server (http://tcoffee.vital-it.ch/apps/tcoffee/index.html) was used to perform a multiple alignment of nucleotide sequences of *TAS3* genes with default parameters. Accession numbers used for constuction of alignment are as follows: AtTAS3a: AT3G17185 (TAIR10), NbTAS3:

FJ804742 (GenBank), OsTAS3: AU100890 (GenBank), PpTAS3b: BK005826 (GenBank).

-*Northern blot analysis*

Total RNA was separated on a denaturing 17.5% polyacrylamide gel (7 M urea) in 0.5× TBE buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA, pH 8.0). Separated RNA was electroblotted onto Hybond-N^+ membranes (GE Healthcare, Little Chalfont, UK). DNA oligonucleotide probes were radiolabeled with [γ-^{32}P] ATP by T4 polynucleotide kinase (Takara). Hybridization was performed at 42°C with PerfectHyb Plus (SIGMA Aldrich, St. Louis, MO, USA). Total RNAs were extracted from mature thalli of Tak-1 and Tak-2, except for detection of miR408. For miR408 detection, 5d-old thalli after planting gemmae were used. Ten micrograms of total RNA were loaded into each lane. U6 snRNA was used as a loading control. All primers used for hybridization are listed in Table S1.

-*Sample preparation for small RNA sequencing*

To fractionate small RNAs in *M. polymorpha*, total RNA was extracted from mature thalli of Tak-1 (seq#1), antheridial heads of Tak-1 with developing and mature antheridia (seq#2) and

archegonial heads of Tak-2 before fertilization (seq#3) using RNAisoPlus (Takara) with 2000 μl/100 mg volume. Next, small RNA fraction with the sizes 15-30 nt was gel-purified from the total RNA after electrophoresis in a denaturing 17.5% polyacrylamide gel (7 M urea) in 0.5× TBE buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA, pH 8.0). A TruSeq Small RNA Sample Preparation Kit (Illumina) was used to construct small RNA libraries from the purified small RNA fraction.

-*Small RNA deep sequencing and data analysis*

The MiSeq desktop sequencer (Illumina) was used to sequence each small RNA library with MiSeq reagents kit v2 (Illumina) giving a maximum sequencing size of 40 bases. After sequencing, the R/Bioconducter package 'Shortread' was used for adapter-trimming of the output fastq files (adapter-sequence, TGGAATTCTCGGGTGCCAAGGAACTCCAGTC) (Morgan *et al*. 2009). Sequences shorter than 17 nt were discarded. Next, the remaining sequences were mapped to the *M. polymorpha* draft nuclear genome sequences ver.2.0 with allowing no mismatches. ShortStack (ver.2.0.9) was used with default parameters to identify small RNA clusters (Axtell 2013). Because location of rRNA coding region was not

identified in the draft genome, rRNA sequences were identified by a BLAST (NCBI BLAST ver.2.2.30) search (E-value cutoff of 10.0) to *M. polymorpha* genome sequences using *A. thaliana* 25S and 18S rRNA sequences as queries, and small RNA clusters that mapped to rRNA sequences were discarded. To compare the features of *M. polymorpha* small RNA expression with *A. thaliana*, deposited sequence data set (SRR275588: Liu *et al*. 2012) and TAIR10 genome sequences were processed together. All *MIRNA* genes were predicted as an *MIRNA* locus by ShortStack in at least one sequence data set of three. All sequence data were deposited to DDBJ under the numbers: seq#1 (DRA003824), seq#2 (DRA003825) and seq#3 (DRA003826). All miRNA annotation data newly identified in this study were deposited in miRBase (http://www.mirbase.org/).

-Prediction of pre-miRNA secondary structures

The mfold web server ver.2.3 was used to predict RNA secondary structures, with the folding temperature at 22ºC (http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form2.3).

-5′ RACE

The GeneRacer Kit (Invitrogen; Life Technologies) was used to perform 5′ RNA ligase-mediated rapid amplification of cDNA ends (RACE) (Llave *et al*. 2002). To determine the cleavage sites of target candidate genes, enzyme reaction of calf intestinal phosphatase (CIP) and tobacco acid pyrophosphatase (TAP) were skipped (Fig. 9A). Poly(A)$^+$ RNA was purified by polyATract (Promega). Primers used for PCR are listed in Supplementary Table 1.

Results

-Characterization of DICER-LIKE (DCL) genes and ARGONAUTE (AGO) genes in M. polymorpha

To function as gene regulatory factors, small RNAs need processing proteins, DCLs, and effector proteins, AGOs. In order to assess whether small RNAs really have function and what kinds of small RNAs are expressed in *M. polymopha*, I performed homology search of *M. polymorpha* draft genome ver.2.0 and EST database. I used amino acid sequences of *A. thaliana* 4 DCL and 9 AGO proteins (Fig. 1A) as query sequences, and performed BLAST search against the EST database of *M. polymorpha* with the TBLASTN option.

I obtained three DCL-like and three AGO-like translational products. They have domains characteristic of plant DCL and AGO proteins. Next, I constructed phylogenetic trees of these *M. polymorpha* candidate amino acid sequences and DCL and AGO proteins in various land plants from green algae to eudicots (*A. thaliana* (eudicots), *O. sativa* (monocots), *S. moellendorffii* (Lycopodiophyta), *P. patens* (mosses) and *C. reinhardii* (green algae) as an outgroup) to assess what types of small RNA-related proteins are present in *M. polymorpha*. The phylogenetic tree of DCL proteins showed three DCL candidates located near DCL1,

DCL3 and DCL4 proteins in the other land plants, respectively (Fig. 1C). Then, they were named MpDCL1, MpDCL3, MpDCL4 (Fig. 1A). It was widely accepted that DCL1, DCL2/4 and DCL3 proteins produce miRNAs, siRNA for posttranscriptional gene silencing (PTGS) and that for RNA-directed DNA methylation (RdDM), respectively (Table 1). This suggests that *M. polymorpha* also has these three small RNA pathways.

The phylogenetic tree of plant AGO proteins suggested that three AGO candidates in M. polymorpha are one AGO1 and two AGO4 homologs (Fig. 1B, D). These candidates were named MpAGO1, MpAGO4a and MpAGO4b, respectively (Fig. 1B). As is DCL prtoteins, it is known that AGO1/5/10, AGO4/6/9 and AGO2/3/7 proteins, load miRNA/PTGS-siRNA, hc-siRNA, and specific miRNA/PTGS-siRNA, respectively (Table 1). *M. polymorpha* has no AGO2/3/7-like proteins, but has one AGO1-like and two AGO4-like proteins (Fig. 1B, D, Table 1). MpAGO1 is probably the only effecter protein for miRNA and PTGS siRNA in *M. polymorpha* because there are no other candidates in AGO1/5/10 or AGO2/3/7 clades. In regards to RdDM pathway, the presence of two AGO4-like proteins suggests complexity or functional diversity in *M. polymorpha* RdDM pathway. Comparing to the moss *P. patens*, which has four DCL proteins (PpDCL1a/b, PpDCL3, PpDCL4) and six AGO proteins

(PpAGO1a/b/c, PpAGO4a/b/c) (Axtell *et al*. 2007), both mosses and liverworts have DCL proteins belonging to three clades and AGO proteins belonging to two clades.

-Highly-conserved microRNAs were expressed in M. polymorpha as in other land plants

Based on the result of the phylogenetic analysis, I performed RNA blot analysis to detect miRNAs expressed in *M. polymorpha*. Among various types of small RNAs, I selected several miRNAs (miR156, miR160, miR166, miR171, miR319, miR390, miR395, miR408, miR529, miR536) that are conserved in various land plants. Samples isolated from the thalli of both male and female plants of wild-type *M. polymorpha* (Tak-1: male, Tak-2: female) were analyzed along with the seedlings of wild-type *A. thaliana* (Columbia-0) as a control. I detected mature products of these conserved miRNAs except miR395 in both Tak-1 and Tak-2 by northern blot analysis (Fig. 2A, B). Also, miR536 was detected from Tak-1 and Tak-2 samples (Fig. 2A, B). miR156 and miR529 have similar sequences to each other (Fig. 2C). Northern blotting detected miR156 cross-hybridized with the miR529 probe in the Col-0 lane (Fig. 2B). We cannot conclude whether miR156 is present in *M. polymorpha* (Fig. 2A). These results suggest that *M. polymorpha* has at least 7 conserved miRNAs like other land

plants.

-*Sample preparation for small RNA sequencing*

As a result of homology search and gel blot analysis, *M. polymorpha* is likely to have several types of small RNAs; miRNA, siRNAs for PTGS and siRNAs for RdDM. Next, I aimed to detect and reveal the small RNA profile of *M. polymorpha* by Illumina high-throughput sequencing technology. In most cases, miRNAs are expressed not constitutively but specifically at some developmental stages or in some specific organs (Wu *et al*. 2006, Jung and Park 2007). To identify small RNAs expressed in various developmental stages in *M. polymorpha*, I extracted total RNAs from three different organs; male mature thalli, male antheridiophores and female archegoniophores (Fig. 3A, B). Three times of sequencing were named seq#1-3, respectively. Next, in order to fractionate small RNA fraction from total RNAs, I purified ~15-30 nt RNA fraction from denaturing urea gels (see M. M. section, Fig. 4A-C). Purified small RNA samples were processed with TruSeq Small RNA Sample Preparation Kit (Illumina) (Fig. 4A).

-*Comprehensive small RNA profiling revealed abundant MIRNA loci in M. polymorpha*

I performed high-throughput sequencing for three samples by MiSeq desktop sequencer (Illumina). I obtained over 5 million reads from each of three sequencing (Fig. 5B). Then, adapter-trimming, filtering by length (< 17 nt) and cluster-mapping to the draft genome sequence ver.2.0 by 'ShortStack' were applied to the raw sequence reads, sequentially (Fig. 5A). ShortStack is a stand-alone Perl program (M. Axtell from Penn State Univ.), which performs a series of small RNA characterizations, such as mapping to the reference genome sequences, discovery of small RNA clusters and prediction of *MIRNA* loci based on the criteria (Meyers et al. 2008) and phasing siRNA loci (Axtell 2013). Because degraded rRNA sequences are likely to be included in small RNA sequence data, after discarding rRNA-derived sequences mannually, I obtained 3,620,931 (seq#1), 699,231 (seq#2) and 1,141,550 (seq#3) small RNA reads (Fig. 5B), which were subjected to following analyses.

-*General characteristics of M. polymorpha small RNAs*

The distribution of sequence reads by length shows a major peak at 21 nt, a typical length of plant miRNAs, in all of three sequence data sets (Fig. 6A). The percentages of reads from

predicted *MIRNA* loci in seq#1, 2 and 3 were 53.2, 11.2 and 41.1%, respectively (Fig. 6B). To compare the size distributions of small RNAs with other land plants, we obtained a set of small RNA sequencing reads of *Arabidopsis* leaves deposited in GEO database (SRR275588; Liu *et al*. 2012) and performed sequence mapping against *A. thaliana* TAIR10 genome sequence by ShortStack (Fig. 6C, D). In contrast to *M. polymorpha* case, the small RNA distribution in *Arabidopsis* showed a major peak at 24 nt, which is characteristic of siRNAs involved in RdDM pathway (Fig. 6C). The percentage of miRNAs in *A. thaliana* was 13.4 and was relatively low compared with *M. polymorpha* (Fig. 6D). Whether or not we include siRNAs from two loci (scaffold_52:424482-425362, scaffold_2339:7-3405) that largely affected the population of small RNAs in seq#2 (Fig. 6B, Supplementary Fig. 2A-F), this distribution suggests that miRNAs occupy a major part of *M. polymorpha* small RNA population (Fig. 5B).

-M. polymorpha has nine conserved miRNAs, and their loci are mostly one or two

Based on the small RNA sequence data of three types of samples and the draft genome sequence, ShortStack predicted 233 *MIRNA* loci in the draft genome (Fig. 7A, C, Table 2).

Some of mature miRNAs from distinct multiple loci have similar sequences. By allowing 3 base mismatches between two miRNAs, it was considered that they are in the same family. Then, I identified 222 miRNA families. Based on the miRNA sequence data set obtained from miRBase (http://www.mirbase.org), I found six mature miRNA sequences matched with miR160, miR166, miR171, miR319, miR390 and miR408. These six miRNAs are considered to be conserved miRNAs in most land plants (Fig. 7A, B). Next, compared with the moss *P. patens* miRNAs, I found three miRNAs (miR529, miR536 and miR1030) were conserved between *P. patens* and *M. polymorpha* but not with *A. thaliana* (Fig. 7A, B). Considering that miR529 is also conserved in rice (monocots) and miR536 is present in *Selaginella* (Cuperus *et al*. 2011), only miR1030 was a bryophyte-specific miRNA among miRNAs identified in this study. Prediction of RNA secondary structures suggested that surrounding sequences of these miRNAs formed stem-loop structures characteristic to pre-miRNAs (Fig. 8A, B). These sequences are probably precursors of identified miRNAs.

A previous study of *P. patens* showed that some *MIRNA* loci are located tandemly (Axtell *et al*. 2007). A similar tandem alignment in the *M. polymorpha* genome was observed for some *MIRNA* gene loci, such as *MIR408ab* and *MIR529ab* (Fig. 7C). In the EST database,

I found a tentative transcript forming tandem stem-loop structures included miR529a and miR529b mature sequences (Fig. 8B). These data strongly suggested that each stem-loop functions as pre-miRNA.

Surprisingly, miR156, one of the major miRNAs highly conserved in most land plants (Willmann and Poethig 2007), could not be detected by the sequencing analysis (Fig. 7B), whereas the genomic sequence of mature miR156 and a stem-loop-like structure was found in the draft genome sequence (scaffold_29:515795-515966) (Fig. 8C). One possibility is that the expression of miR156 is limited in specific organs or developmental stages and escaped our experiment. The other possibility is that DCL1 cannot perform proper processing of the pre-miRNA molecule. When we studied its hairpin structure, it appeared to include a large bulge and was likely that the structure is disadvantageous for precise cleavage (Fig. 8C). Also, I could not find the pri-miRNA or pre-miRNA sequences of miR156 in EST database, indicating that tentative *MIR156* locus is not transcribed in *M. polymorpha*.

It is generally accepted that plant *MIRNA* loci have multiplied along the evolutionary divergence of respective species (Reinhart *et al*. 2002). However, I found that the numbers of loci belonging to each conserved *MIRNA* family in *M. polymorpha* is on average less than

those in *P. patens* and *A. thaliana* (Fig. 7B). *M. polymorpha* miRNA families had only one or two loci, except for miR529. This result suggested that the *MIRNA* loci of the conserved miRNA families in *M. polymorpha* remain in a simple genome state of an early land plant ancestor without duplication.

The most abundant sequence detected in three sequence data sets was the mature product of miR166a (UUCGGACCAGGCUUCAUUCCC), which is an ancient miRNA that represses the expression of the cognate HD-ZIPIII transcriptional factor family genes in various land plants (Fig. 7C). Two *MIR166* loci were identified in the *M. polymorpha* genome and designated Mp*MIR166a* and Mp*MIR166b*, respectively. However, ShortStack mapped most reads against one of the two loci, *MIR166a* (scaffold_103:111814-111921) (Fig. 7C). This result suggested that miR166a works in *M. polymorpha* majorly.

-Identification of target genes of conserved miRNAs in M. polymorpha

Compared with animal miRNAs, it is known that plant miRNAs bind to target mRNAs with high complementarity (Bartel 2004). Once a certain combination of miRNA and its target mRNA was established and played an important role, it can be considered that both the

miRNA and the target mRNA were likely to keep complementary sequences unchanged. This could have imposed pressure to keep extensive conservation on such ancient miRNAs throughout long evolutionary history of land plants (Axtell 2008). Thus, it was expected that the complementary sequences in target mRNAs would be similarly conserved if *M. polymorpha* established such tight relationships between miRNA and target mRNA.

To test this assumption, I next searched for candidate target genes with sequences complementary to those of conserved miRNAs in *M. polymorpha* genome from an *M. polymorpha* EST database by computational target prediction with the 'Targetfinder' program (Fahlgren *et al*. 2007). Possible encoded proteins in candidate mRNAs (score=<3) were then subjected to a BLASTP search in the TAIR10 protein database to predict and assign the coded proteins and their possible functions. Consequently, the base-pairings of miRNA/target families, like miR160/Mp*ARF3,* miR166/Mp*HD-ZIPIII,* were predicted to be conserved in *M. polymorpha* as previous studies (Fig. 7C) (Flores-Sandoval *et al*. 2015b, Floyd and Bowman 2004). I also searched for candidate targets of other conserved miRNAs by comparing orthologs between *M. polymorpha* and *A. thaliana*. I identified the conserved pairings of miR319/Mp*MYB33-like* and miR408/Mp*Laccase*-Mp*Uclacyanin* in *M. polymorpha* (Fig. 7C).

I suspected that other conserved miRNAs might target protein-coding sequences of gene families, which are different from those in *Arabidopsis* (Fig. 7C).

Focusing on such predicted pair sets of miRNA/target, I could detect canonical cleavage sites guided *in vivo* by miR160, 166 and 319 by rapid amplification of cDNA ends (RACE) analysis of Mp*ARF3*, Mp*HD-ZIPIII* and Mp*MYB33-like* mRNAs (Llave *et al*. 2002) (Fig. 9A, B). Similar to these three miRNAs, the mature miR529c assisted cleavage of the SPL transcription factor mRNA in *M. polymorpha,* considering its cleavage position (Fig. 9C). miR529 has an overlap sequence of 16–18 nt with that of miR156. The sequence of miR529 is conserved in mosses and monocots, but not in eudicots (Cuperus *et al*. 2011). Small RNA sequencing revealed no expression of miR156 in *M. polymorpha* (Fig. 7B, 9C). miR529c possibly evolved as an alternative miRNA to negatively regulate Mp*SPL2* in *M. polymorpha*.

On the contrary, I could not detect the RACE product of either *Laccase* or *Uclacyanin* mRNAs cleaved at the predicted target sites. The cleavage sites of these miRNAs were consistent with previous notion that plant miRNAs cleave the target mRNA at the site between the 10th and 11th nt from the 5′ end of the mature miRNA sequence (Llave *et al*. 2002). These target sites were almost mapped in the protein coding regions, as in other land

plants (Fig. 9B).

-*Identification of novel miRNAs in M. polymorpha*

Aside from the nine miRNAs described above, I also identified non-conserved miRNAs (Fig. 10A, Table 2). Most of these predicted novel miRNAs were 21 nt long and their precursors form hairpin secondary structures encompassing miRNA/miRNA* pairing as predicted by ShortStack program (Meyers 2008, Axtell 2013) (Fig. 10C, Table 2). I compared these newly-identified miRNAs to those of another liverwort, *Pellia endiviifolia*, identified by Alaba *et al*. (2015). Using these miRNA sequences as queries, I performed BLASTN search with the short-blastn option against our newly identified miRNAs in *M. polymorpha*. Unexpectedly, I found that only pen-miR8163 (Alaba *et al*. 2015) shared a similar sequence with Mpo-MR-27 in my study (Fig. S3). Based on this result, I hypothesized that different classes of liverworts share few miRNAs, if any, and have distinct miRNAs. Rather, it seemed that species-specific miRNAs are abundant in liverworts. A comparison of relative read counts between conserved miRNAs and non-conserved miRNAs revealed no significant differences (Fig. 10B; Mann–Whitney U test, p=0.494).

Target prediction was also performed for novel miRNAs whose RPM were more than 100 in at least one of the three sequencing data sets. Computational prediction revealed that the novel miRNAs were likely to target genes encoding gene regulatory factors like transcription factors, F-box proteins, PPR proteins and leucine-rich repeat receptor-like kinase (LRR-RLK) as reported in other land plants (Jones-Rhoades *et al*. 2006) (Fig.11A). In particular, Mpo-MR-13 was predicted to target one of four *SPL* genes in the *M. polymorpha* EST database, named Mp*SPL1* (Drs. Niwa and Araki in Kyoto Univ., personal communication). My RACE analysis validated this target prediction (Fig. 11B). Together with targeting of Mp*SPL2* by miR529c (Fig. 9C), *M. polymorpha* probably has two types of miRNA-regulation systems for *SPL* genes.

-*Characterization of possible tasiRNA*

Deep sequencing of small RNAs also detected abundant siRNA clusters, using a program in ShortStack that scored a phasing offset to each identified small RNA cluster. Trans-acting siRNAs (tasiRNAs) are one type of endogenous phasing siRNAs that downregulate gene expression at the post-transcriptional level. TasiRNAs are plant-specific, but are conserved

from bryophytes to angiosperms (Axtell *et al*. 2006). TasiRNAs are derived from the direct cleavage of long non-coding RNA from *TAS* loci, under the guidance of miR390, and are further matured through the complicated biogenesis pathway by DCL4, RDR6 and SGS3 (Allen *et al*. 2005, Yoshikawa *et al*. 2005). *TAS3* is a highly conserved non-coding RNA locus and produces mature tasiRNAs that repress *ETTIN/ARF3* and *ARF4* mRNAs both in *P. patens* and *A. thaliana* (Axtell *et al*. 2006). Small RNA sequencing and phylogenetic analysis revealed that *M. polymorpha* has miR390 but no AGO7-like proteins (Fig. 1B, D). Then, I wondered whether or not TAS3-tasiRNA exists. We identified one tentative *TAS3* locus (Mp*TAS3*), having the lowest phase offset score of all small RNA clusters (scaffold_154:327227-327462) and to which phased siRNAs were mapped (Fig. S1A). The sizes of the siRNAs from the Mp*TAS3* locus are predominantly 21 nt, like the *TAS3*-derived tasiRNAs identified so far in other land plants (Fig. S1B). The Mp*TAS3* sequence has two miR390-binding sequences at the 5′ and 3′ sites that are highly conserved among other land plants, such as *A. thaliana*, *Nicotiana benthamiana*, *Oryza sativa* and *P. patens* (Fig. S1C). It was previously reported that *M. polymorpha* has one *TAS3* locus expressing tasiR-AP2, but not tasiR-ARF, yet Pp*TAS3* expresses both tasiRNAs, named PptasiR-AP2 and PptasiR-ARF

(Krasnikova *et al*. 2013). Although I found a sequence similar to PptasiR-AP2 in the Mp*TAS3* locus, neither tasiR-AP2-like nor tasiR-ARF siRNAs were detected in the mapped siRNA sequences (Fig. S1C). Furthermore, based on sequence complementarity, I could not find the genes targeted by any other phased siRNAs. Consequently, it was suggested that *M. polymorpha* has miR390 and TAS3-tasiRNA pathway but whether tasiRNAs have their function or not remains unknown.

-Detection of sex-specific phasiRNAs

I found that, in antheridiophores, two loci express a considerable level of 20 and 21 nt small RNAs from both the positive and negative strands, whereas they were not detected not thalli and were detected only at low levels in archegoniophores (scaffold_52:424482-425362, scaffold_2339:7-3405; Fig. S2A-F, 6A). One of them, scaffold_2339, has tandem repeat sequences of 378 bp long (data not shown). These abundant siRNAs detected in *M. polymorpha* antheridiophores are mainly 21 nt long (Fig. S2B, E) and showed high 21 nt phasing scores (Fig. S2C, F). Recently, it has been reported that abundant tasiRNA-like phasing siRNAs (phasiRNAs) are expressed in various species, such as grasses like

Brachypodium (Fei *et al*. 2013, Axtell 2015). PhasiRNAs are 21 nt long and secondarily processed by DCL4 after cleavage under the guidance of another small RNA. However, I could not identify any miRNAs triggering biogenesis of siRNAs from these two loci, although I did detect phasiRNA-like sequences. Based on these data, I would like to designate these siRNAs as sex-specific phasing siRNAs (SS-phasiRNAs). SS-phasiRNA loci have no obvious ORFs in their nucleotide sequences and are probably processed from certain non-coding transcripts.

Discussion

During adaptation to living on the land, it is assumed that plants have modulated various gene regulatory networks to dictate when and where respective genes are expressed and diversified their body patterning. The data in this chapter indicated that conserved miRNAs comprise a highly conserved gene regulatory system among land plants. Here, I profiled the small RNAs and identified both conserved and novel miRNAs in *M. polymorpha* for the first time.

By characterization of miRNAs identified in this study, it was revealed that *M. polymorpha* has six families of miRNA that are conserved among land plants (Fig. 7A, B). It is commonly accepted that conservation of miRNAs is usually accompanied by conservation of their target genes (Axtell 2008). Among the conserved miRNAs in *M. polymorpha*, miR160, -miR166 and -miR319 guided the mRNA cleavages of the same cognate target gene family as other land plants (Fig. 9B).

In contrast, some target predictions could not be verified experimentally like in other land plants, such as miR171/GRAS transcription factor gene (Fig. 7C). I could not detect cleavage of Laccase and Uclacyanin mRNAs by Mp-miR408 by RACE analysis (data not shown), either. It is unknown whether these conserved miRNAs have different targets or

silencing reaction works but below detection level.

Surprisingly, small RNA sequencing could not detect expression of miR156 although the nucleotide sequence is present in the draft genome sequences (Fig. 7C, 9C). Since there is no sequence from this locus in EST database, *MIR156* gene is probably not expressed in *M. polymorpha*. Alternatively, 5'RACE analysis suggested that Mp*SPL2* mRNA is cleaved by miR529c in *M. polymorpha* instead of miR156. miR529 is conserved from bryophytes to monocots and has similar function to miR156 (Fig. 2D). By phylogenetic analysis, it was revealed that Mp*SPL2* is homologous to seven *SPL* genes (AtSPL2, 6, 10, 11, 13 and 15) targeted by miR156 in *A. thaliana* (Fig. 12B). Amino acid sequences of miR156/529 target site are highly conserved (ALSLLS; Fig. 12A). Considering the last serine and codon nucleotides (TCA) are conserved among At*SPL6, 9, 15* and Mp*SPL2*, in this case, it is suggested that alternating from miR156 to miR529c occurred recently in *M. polymorpha*. This is possibly the case of miRNA disappearing. Also, in *M. polymorpha*, Mp*SPL1*, a homolog of At*SPL8* (Fig. 12), which is not regulated by miR156 in *A. thaliana*, is targeted by a newly identified miRNA, Mpo-MR-13 (Fig. 11B, Table 2). It is not known whether Mpo-MR-13 like miRNAs targeted the Mp*SPL1*/At*SPL8* gene of ancient land plants or *M.*

polymorpha obtained this regulatory system uniquely. Consequently, considering that *SPL* genes are major regulators of developmental phasing in various land plants, regulation of *SPL* genes would be an important factor in land plant development.

In bryophytes, *Physcomitrella patens* has been used as the former model plant to understand land plant evolution by molecular genetic approaches. Its draft genome sequence was published and significant knowledge about plant small RNA evolution was brought by researches using *P. patens* (Axtell *et al*. 2006, 2007, Cho *et al*. 2012, Coruh *et al*. 2015). By small RNA sequence analysis, it was revealed that only nine miRNA families were shared between *M. polymorpha* and *P. patens* (Fig. 7A). We could identify miR1030, which was first reported in a study of small RNA sequencing in *P. patens* (Axtell *et al*. 2007), as the only tentative bryophyte-specific miRNA. Although *P. patens* has several merits as a model plant like high efficiency of homologous recombination, one of difficulties to use is high redundancy of genes. It was suggested that a large-scale duplication happened about 30-60 Mya in mosses (Rensing *et al*. 2007, 2008). *MIRNA* genes coding a precursor and a mature product of miRNA are also multiplicated in the *P. patens* genome (Fig. 6B). The state of multiplication of *MIRNA* loci is reflected in the genome of *A. thaliana* and brings difficulty to

analyze miRNA fuction in detail. Using of *M. polymorpha* has possibility to overcome this difficulty and combining the results of *P. patens* studies, new insights about miRNA evolution in bryophytes will be produced possibly.

Recently, Alaba *et al*. (2015) characterized microRNAs of another liverwort, *Pellia endiviifolia*, which belongs to the Jungermanniopsida, solely based on small RNA sequencing and transcriptome data as reference sequences. The study identified 11 conserved and 42 non-conserved miRNAs in *P. endiviifolia.* By comparing miRNAs in two liverwort species, *M. polymorpha* and *P. endiviifolia*, it was revealed by sequencing in this study that only miR8163 was identified as a tentative liverwort-specific miRNA among 213 novel miRNAs (Fig. S3, Table 2). These data support a model of frequent appearance and loss of *MIRNA* genes over an evolutionary time scale as suggested previously (Axtell *et al*. 2007, Fahlgren *et al*. 2007). Since the genome sequences of *P. endiviifolia* was not revealed, it was not confirmed that miRNAs identified by Alaba *et al*. (2015) are really present in the genome. Also, because molecular genetic tools for analysis of *P. endiviifolia* were not developed, miRNA roles in liverworts have not been reported. In these points, *M. polymorpha* is a valuable model liverwort to understand small RNAs and their function in terms of land plant

evolution.

Low redundancy of gene sets is one of the notable characteristics of *M. polymorpha* as a model organism (Zobell *et al*. 2010, Ueda *et al*. 2012; Kato *et al*. 2015; b Sandoval *et al*. 2015). Although other land plants tend to encode *MIRNA* genes in multiple loci and have some redundancy, the conserved *MIRNA* genes in *M. polymorpha* are encoded in one or few loci (Fig. 7B, C). This suggested that conserved miRNAs in *M. polymorpha* retained a prototype role of the gene regulation by miRNAs that was present before the multiplication of ancient *MIRNA* genes. The low redundancy of *MIRNA* genes offers a good platform where significance of each *MIRNA* locus could be studied using homologous recombination or genome editing technology such as the CRISPR/Cas9 system (Doudna and Charpentier 2014, Sugano *et al*. 2014). Contrary that *MIRNA* genes were limited in one or several loci in *M. polymorpha*, it is wondered that *MpAGO4* genes are duplicated in the genome (Fig. 1B, D). It remains unknown whether each AGO4 protein has specific roles or not.

The sequence data in this study showed that 21 nt small RNAs derived from *MIRNA* loci are expressed as the majority, whereas 24 nt siRNAs are dominant among *A. thaliana* small RNA (Fig. 6A, B). In recent study by Coruh *et al*. (2015), *P. patens* has a number of

heterochromatic siRNA loci with both 23 and 24 nt length. My sequence data also detected both 23 nt long siRNAs as well as 24 nt (Fig. 6A, B). Like *P. patens*, *M. polymorpha* is likely to have heterochromatic siRNA loci, and similarity in bryophytes will be investigated by future studies. Similar to previou studies (Zhai *et al*. 2011, Fei *et al*. 2013), I also found that 20/21 nt long phasing RNAs were abundantly expressed, but only in sexual organs, especially male antheridiophores (Fig. S2A-F). Judging by their length, they were probably produced in a DCL4-dependent manner, but their function or biogenesis machinery remains unknown. One interesting point is that their expression level is much higher in male plants than is in female plants (Fig. 6B). Some miRNAs also showed a difference in expression level between male antheridiophores and female archegoniophores (Fig. 7C, Table 2). There are few small RNA studies of dioecious plants; therefore, *M. polymorpha* could be an important model organism to understand small RNA functions, particularly those depending on differences between sex organs of land plants with X/Y sexual chromosomes.

In this chapter, I could profile and characterize miRNAs in *M. polymorpha* for the first time. To get insights about biological roles of miRNAs in *M. polymorpha*, I next performed molecular genetic analysis by ectopic expression and CRISPR/Cas9-mediated genome editing

of miRNA genes.

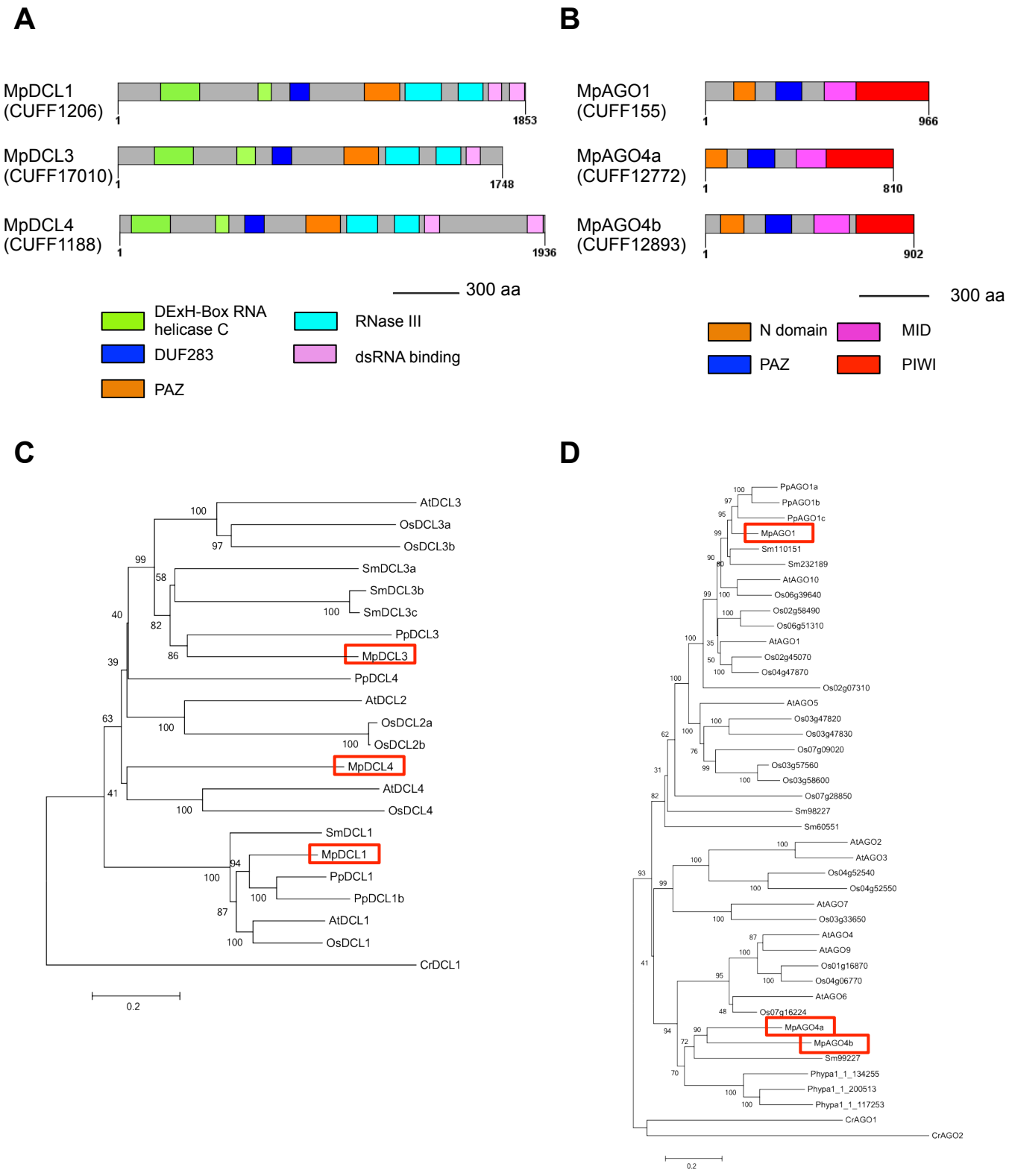


Fig. 1 *M. polymorpha* has three DCL genes and three AGO genes

(A, B) Schematic representation of small RNA related proteins in *M. polymorpha* with predicted domains. Domain models were predicted from those of A. thaliana DCLs and AGOs with alignment data. Tentative numbers from EST database are assigned. Scale bars show the length of 300 amino acids. (A) DCL proteins. (B) AGO proteins. (C, D) Unrooted phylogenetic trees constructing relationships of DCL and AGO proteins in land plants. Numbers are percentages of bootstraps of 1000 replicates. Scale bars represent amino acid substitutions per site. *C. reinhardii* protein sequences are used as outgroups. At, *A. thaliana*; Os, *O. sativa*; Sm, *S. moellendorffii*; Pp, *P. patens*; Mp, *M. polymorpha*; Cr, *C. reinhardii*. Red boxes represent *M. polymorpha* proteins. (C) DCL proteins. (D) AGO proteins.

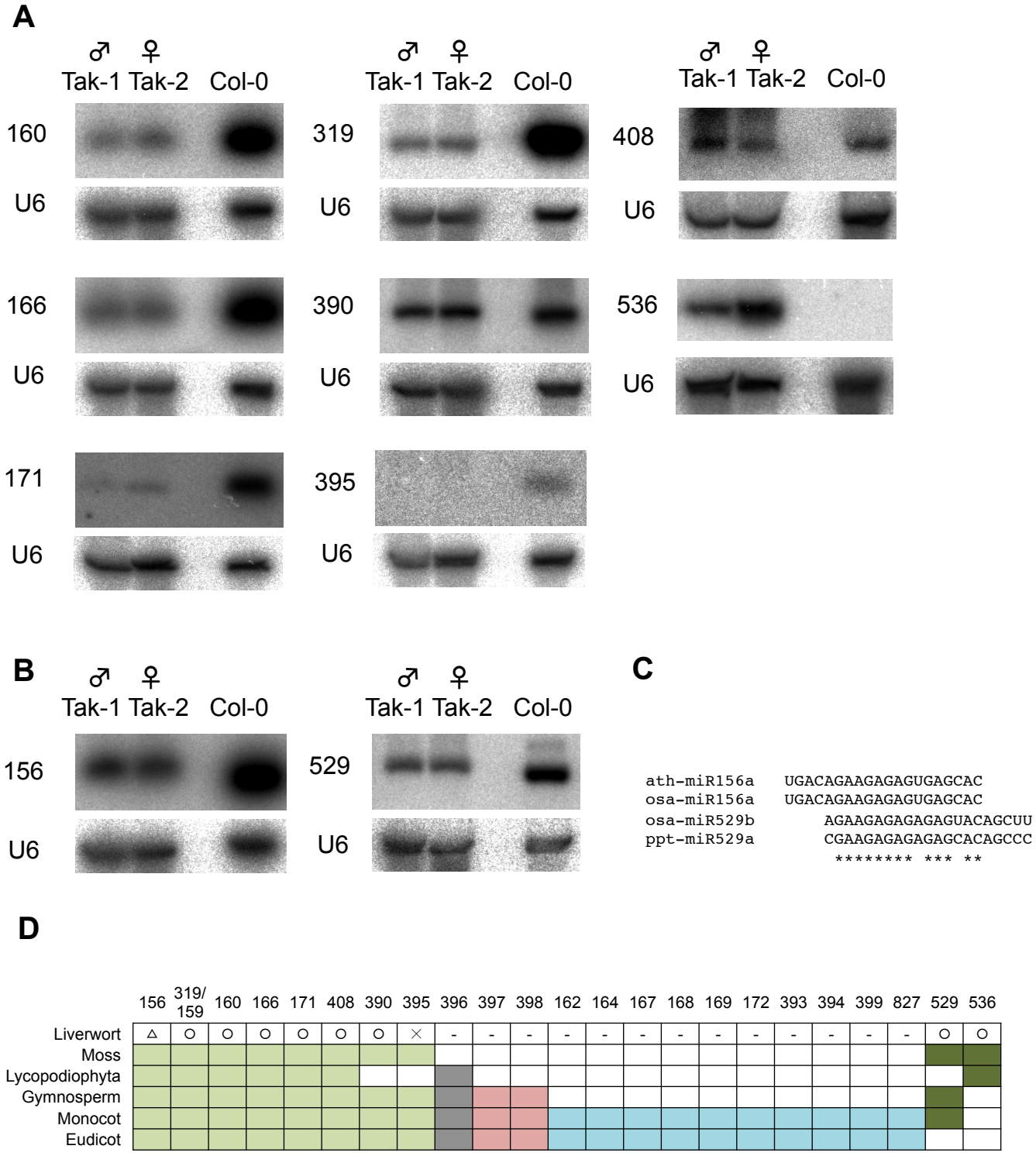


Fig. 2 *M. polymorpha* shares conserved microRNAs with other land plants.
(A) Northern blot analysis to detect conserved miRNAs in *M. polymorpha*. 10 μg of total RNA from mature Tak-1 and Tak-2 thalli was loaded into each lane. Total RNA from Col-0 seedlings was used as a control. U6 snRNA was used as a loading control. (B) Band patterns to miR156 and miR529 in northern blot analysis. (C) Comparison of representative mature sequences of miR156 and miR529 in *A. thaliana* (ath), *O. sativa* (osa) and *P. patens* (ppt). (D) Table representing conservation of miRNAs in land plants. Table was modified from Fig. 1 in Cuperus et al. (2011). miRNAs conserved among all land plants (green), from Lycopodiophyta to eudicots (gray), from gymnosperms to eudicots (red), in monocots and eudicots (blue), not in eudicots (deep green).

A

Number	Plant	Sample	Description
seq#1	Tak-1 (male)	Thallus	~4 weeks from gemma
seq#2	Tak-1 (male)	Antheridial head	Before release of sperms
seq#3	Tak-2 (female)	Archegonial head	Before fertilization

B

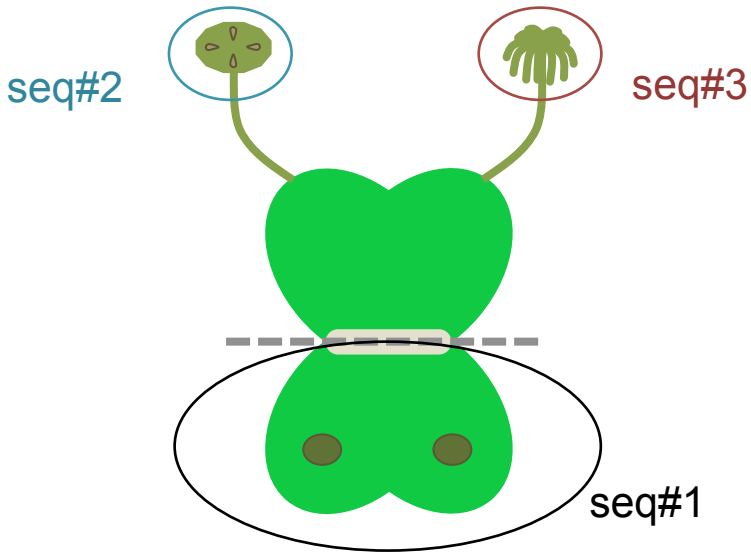


Fig. 3 Samples for small RNA sequencing
(A) The summary of three samples (seq#1, seq#2, seq#3) in this study. (B) A scheme showing body parts for each sequence sample.

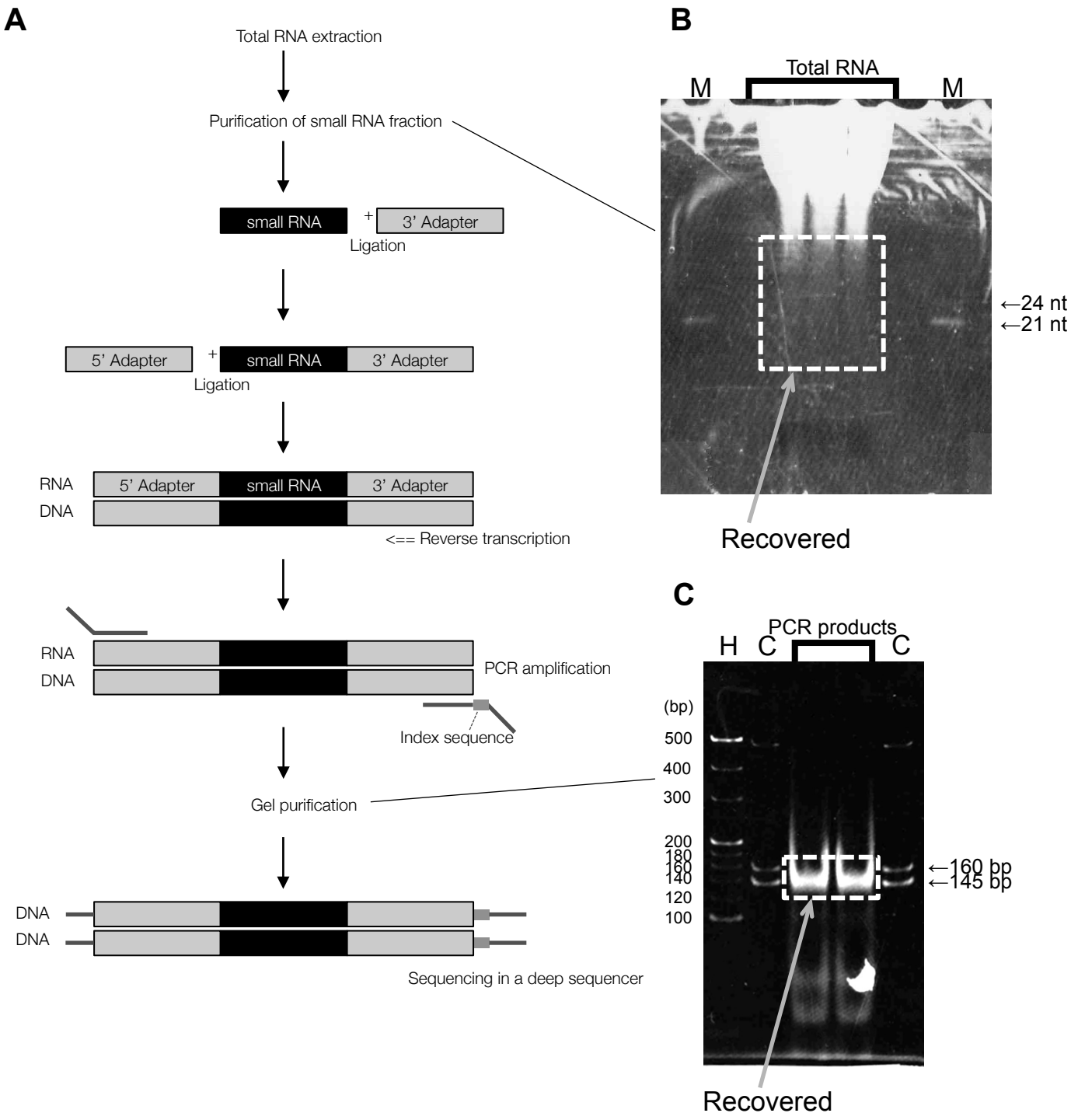
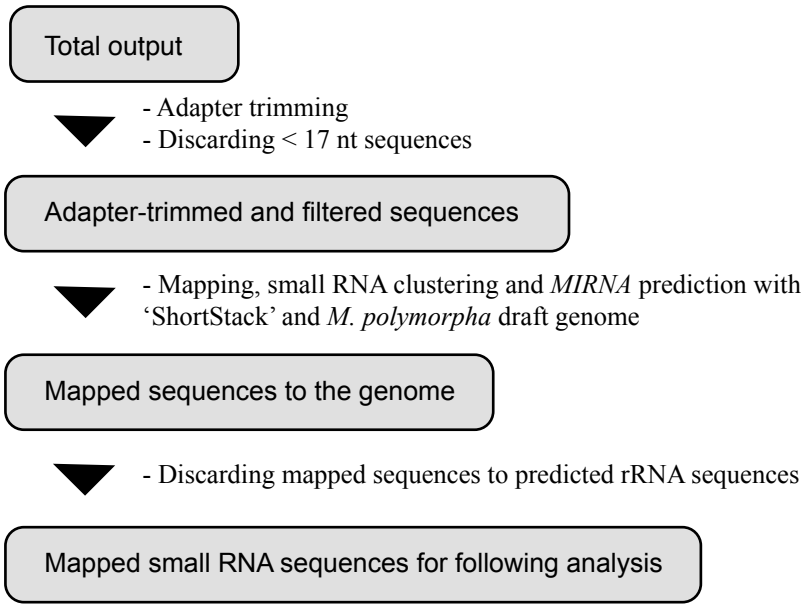


Fig. 4 Sample preparation for high-throughput small RNA sequencing

(A) Schematic diagram of sequencing library preparation. (B, C) Examples of band pattern in gel purification. The boxes in white dashed lines show areas to excise the gel by a razor blade for RNA extraction. (B) 15% acrylamide gel shift pattern after EtBr staining in small RNA purification. M: 21 and 24 nt small RNA marker. (C) Gel shift pattern of PCR products of adapter-ligated cDNA after EtBr staining. Two types of ladder markers were used as ; H, High Resolution Ladder marker; C, Custom Ladder marker (TruSeq small RNA sample kit, Illumina).

A



B

Dataset	seq#1 (DRA003824)		seq#2 (DRA003825)		seq#3 (DRA003826)	
	Total reads	Unique sRNAs	Total reads	Unique sRNAs	Total reads	Unique sRNAs
Total output	16,405,507		7,842,833		9,312,316	
Adapter-trimed and filtered sequences	10,114,331		5,592,286		5,828,557	
Clustered sequences to draft genome	6,006,557	4,529,014	3,382,927	1,964,663	3,008,569	1,978,525
Non-rRNA sequences	3,620,931	2,777,092	699,231	324,682	1,141,550	668,688

Fig. 5 A flow-chart of small RNA sequencing analysis.
(A) Schematic diagram showing the processing of sequence data. (B) Summary of the number of reads of each type of sequence.

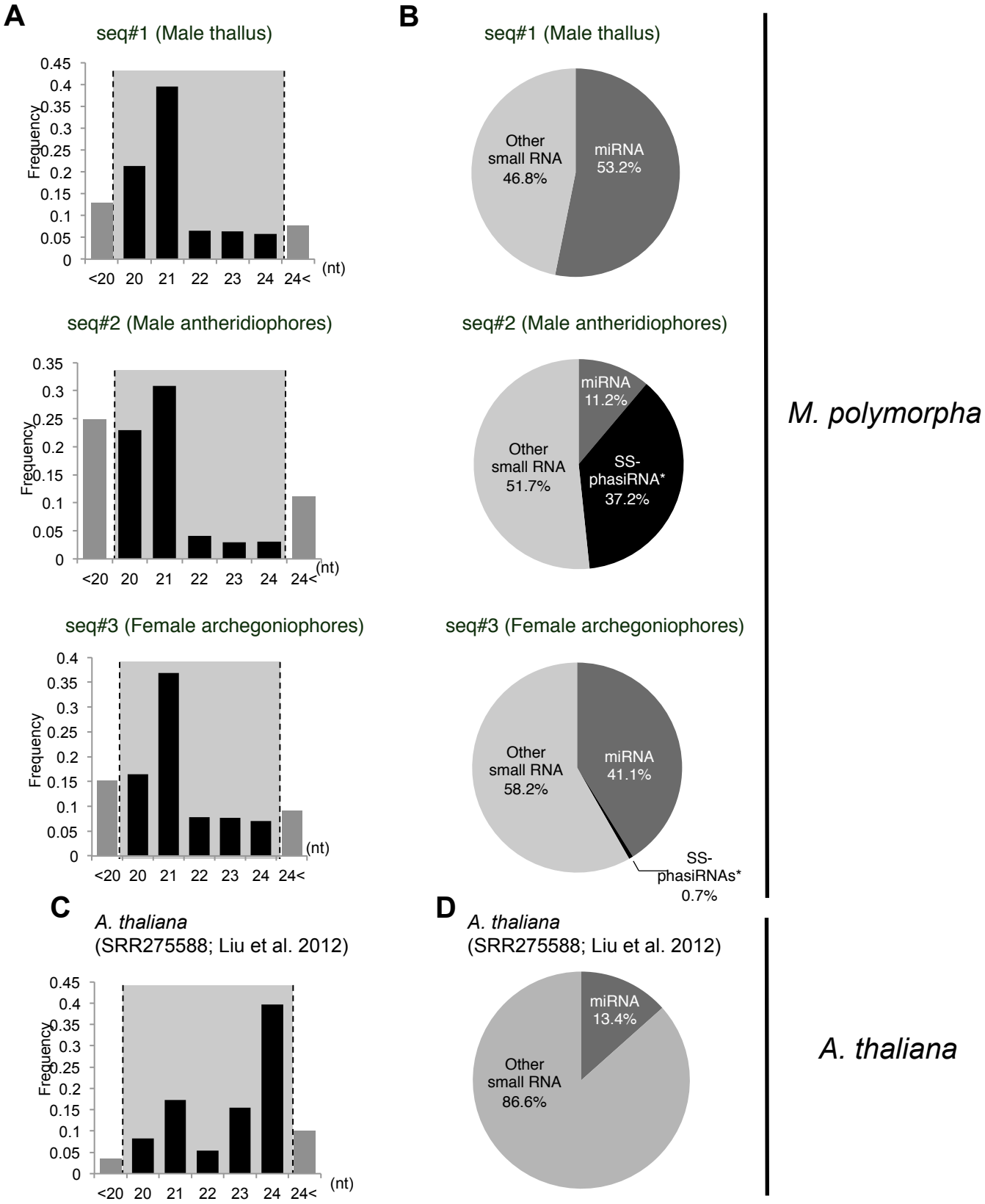
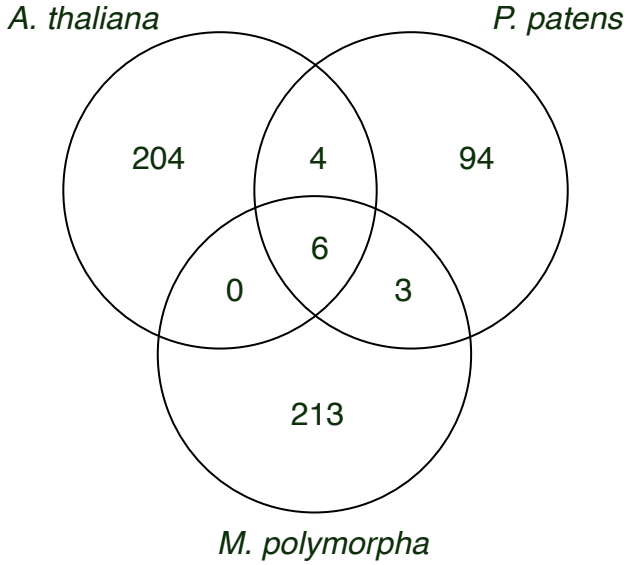


Fig. 6 Distribution of expressed small RNAs in *M. polymorpha.*

(A, B) Distribution of small RNAs detected in seq#1, 2 and3. (A) Distribution by length of reads. The frequency was calculated by determining the number of total clustered non-rRNA reads as 1.0. (B) A pie chart shows the percentage of small RNAs expressed from the *MIRNA* loci and abundant phasing-siRNAs expressed from two loci in scaffold52 and 2339 (SS-phasiRNAs*). (C, D) The same analysis as in A and B using deposited small RNA sequencing data of *Arabidopsis* seedlings (SRR275588; Liu et al. 2012).

A



B

miRNA	*M.polymorpha*	*P. patens*	*A. thaliana*
miR156/157	-	3	12
miR319/159	2	5	6
miR160	1	9	3
miR162	-	-	2
miR164	-	-	3
miR166/165	2	13	9
miR167	-	-	4
miR168	-	-	2
miR171/170	1	2	4
miR172	-	-	5
miR390	1	3	2
miR395	-	1	6
miR408	2	2	1
miR414	-	1	1
miR419	-	1	1
miR529	3	7	-
miR536	1	6	-
miR1030	1	10	-

C

Name	Locus in the draft genome ver.2.0	Sequence	RPM#1	RPM#2	RPM#3	Predicted or validated target genes
miR160	scaffold_2:2177627-2177751	UGCCUGGCUCCCUGUAUGCCA	38	34	99	Auxin response factor (ARF)
miR166a	scaffold_103:111814-111921	UUCGGACCAGGCUUCAUUCCC	160356	21837	64854	Class III homeodomain-leucine zipper (HD-ZIPIII)
miR166b	scaffold_13:1731867-1731968	UCGGACCAGGCUUCAUUCCCC	106	6	43	
miR171	scaffold_30:1375884-1375998	GUGAUGUUGGUGAGGCUCAAU	30	84	191	MATE efflux family
miR319a	scaffold_105:177087-177271	CUUGGACUGAAGGGAGCUCCC	6108	2446	20617	MYB transcription factor
miR319b	scaffold_98:201228-201432	CUUGGACUGAAGGGAGCUCCC	3254	225	13701	
miR390	scaffold_62:46273-46387	AAGCUCAGGAGGGAUAGCGUC	97	23	158	TAS3-like RNA, Leucine rich repeat recepter kinase (LRR-RLK)
miR408a	scaffold_95:208857-208965	UGCACUGCCUCUUCCCUGGCU	2283	7	42	Uclacyanin
miR408b	scaffold_95:205012-205139	UGCACUGCCUCUUCCCUGGCU	9	0	0	
miR529ab	scaffold_6:180634-180738	AGAAGAGAGAGAGCACAGCCC	26	3	4	KANADI family transcription factor
miR529c	scaffold_6:217429-217634	CCAGAAGAGAGAGAGCACAGC	131	6	13	Squamosa promoter binding-like (SPL) family
miR536	scaffold_12:1545887-1546071	CCGUGCCAAGCUGUGUGCAUC	3105	133	865	ABA insensitive 3 (ABI3: AP2/B3 type)
miR1030	scaffold_46:1273737-1273854	UCUGCAUCUGCACCUGCACC	8	521	2123	NB-LRR family

Fig. 7 Small RNA sequencing revealed the presence of conserved miRNAs in *M. polymorpha*.

(A) A Venn diagram showing the number of conserved miRNA families in three model plants: a liverwort (*M. polymorpha*), a moss (*P. patens*) and an angiosperm (*A. thaliana*). miRNAs from *P. pat*ens and *A. thaliana* registered in miRBase were used for homology searching. *M. polymorpha* miRNAs were identified by ShortStack. (B) A table showing the number of loci of each *MIRNA* gene in the three plants. (C) Information about conserved miRNAs identified by three sequencing runs. Sequences are the most abundant in the sequences mapped to the *MIRNA* cluster. Counts were converted to reads per million (RPM) by dividing by total reads of mapped sequences, except rRNA-derived sequences, and multiplying by 1 million. Listed target candidates are those predicted by the TargetFinder program.

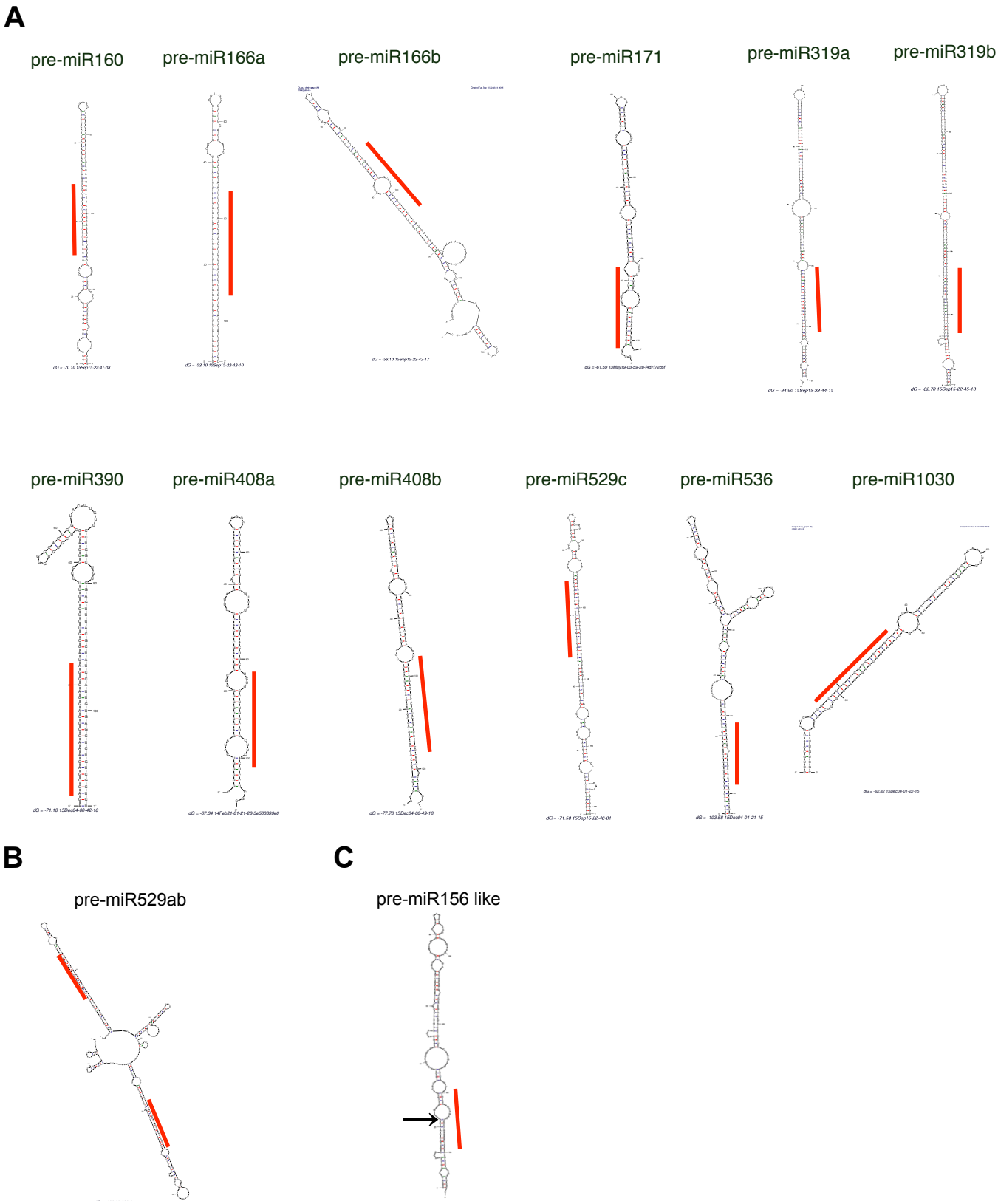


Fig. 8 Predicted hairpin structures of pre-miRNA in *M. polymorpha*

(A-C) Hairpin secondary structure of the predicted precursor conserved miRNAs by RNAfold program. Red lines show mature sequences of miRNA. (A) Twelve precursors of nine conserved miRNAs. (B) Secondary structure of pre-miRNA for both miR529a and b, which are located tandemly in a genome fragment. (C) Predicted secondary structure of sequences containing a miR156-like sequence (red line) in the genome. An arrow shows the bulge structure inside the hairpin structure.

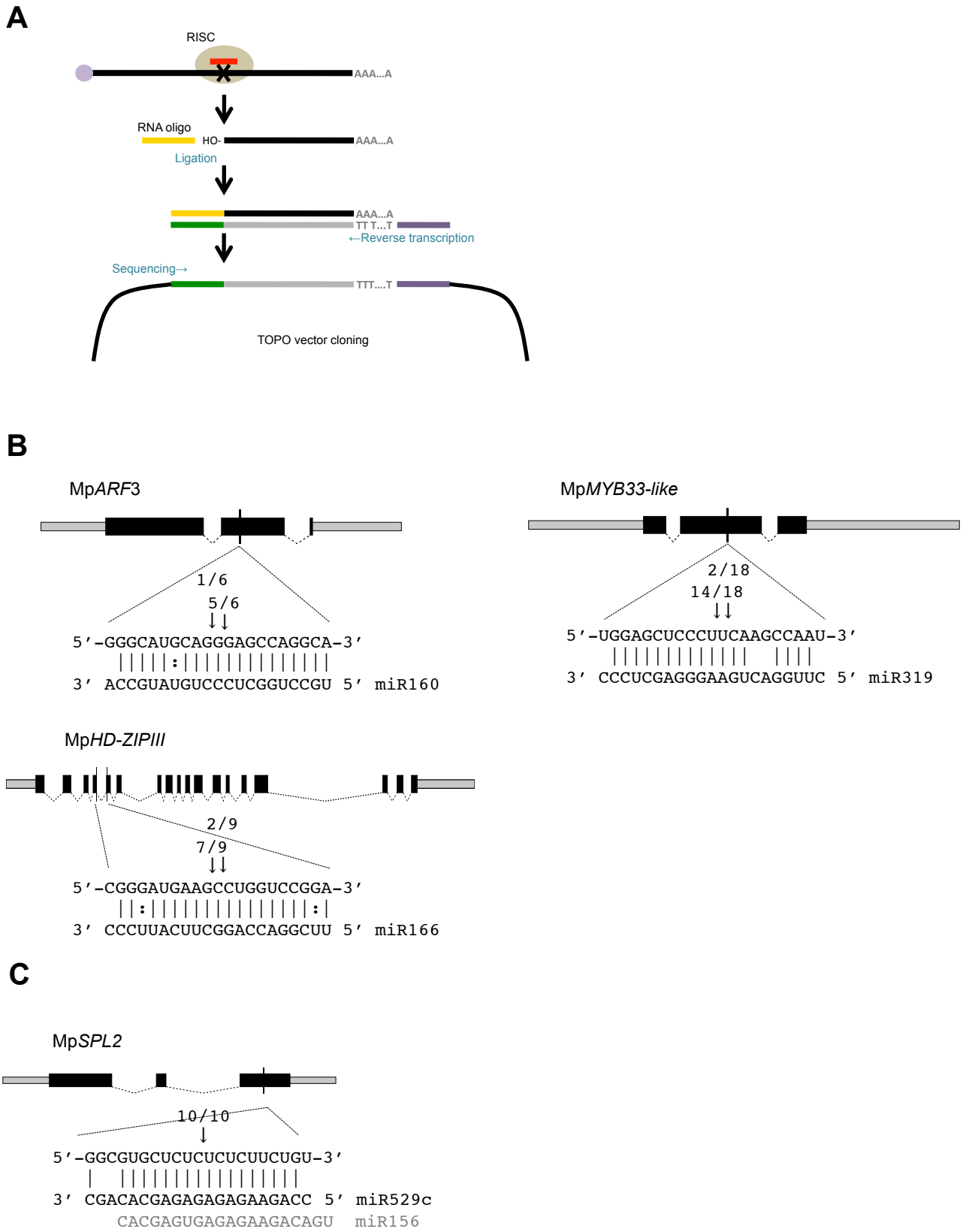


Fig. 9 Conserved miRNAs cleaved mRNA of conserved target genes

(A) Schematic representation of 5' RACE analysis. (B, C) 5′ RACE assay to detect the cleavage sites in the target mRNA by miRNAs. Arrowheads represent the 5′ end position and the number of corresponding RACE-clones out of total clones. Black bars represent exons, and grey bars represent 5′ and 3′ UTRs. (B) miR160/Mp*ARF3*, miR166/Mp*HD-ZIPIII* and miR319/Mp*MYB33-like* pairings. (C) miR529c/Mp*SPL2* pairing and miR156 sequence.

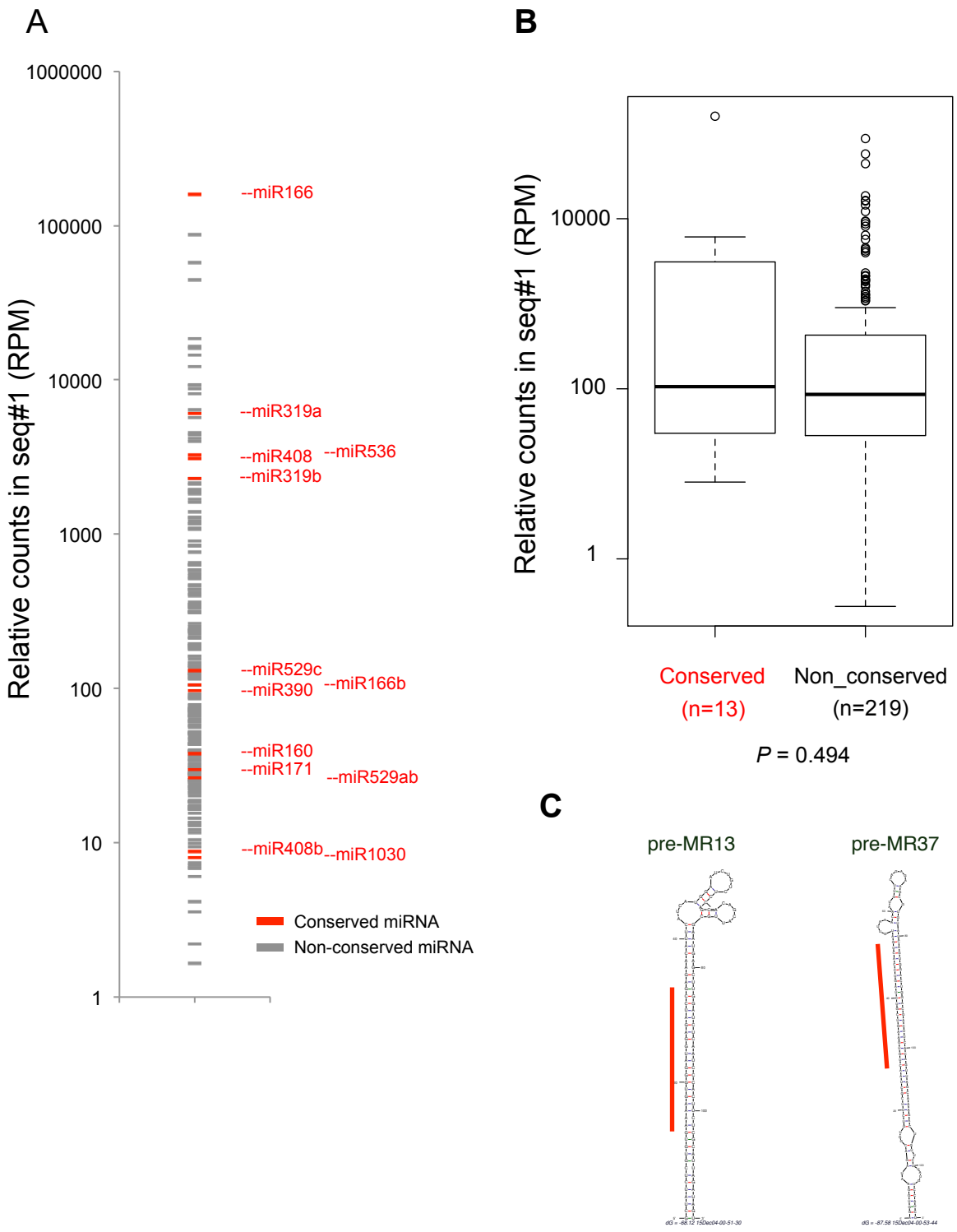


Fig. 10 ***M. polymorpha*** **has a lot of non-conserved miRNAs as well as conserved ones.**
(A) A logarithmic graph showing the normalized expression levels of miRNAs detected in seq#1. Red bars represent conserved miRNAs, and grey bars represent novel miRNAs. The number of reads mapped to a certain locus was converted to reads per million (RPM) by dividing by the total reads of mapped sequences, except rRNA-derived sequences, and multiplying with 1 million. (B) A box plot showing the comparison of relative expression levels between conserved miRNAs and non-conserved miRNAs. The top and bottom of the box show the first and third quartiles, respectively, and the band inside the box shows the second quartile. The ends of the whiskers show the highest datum within the 1.5 interquartile range (IQR) of the upper quartile and the lowest datum within the 1.5 IQR of the lower quartile. There was no significant difference between them (p=0.494, Mann–Whitney U test). (C) Predicted secondary structures of precursors of two novel miRNAs, Mpo-MR13 and Mpo-MR37. Red lines show location of mature miRNA sequences.

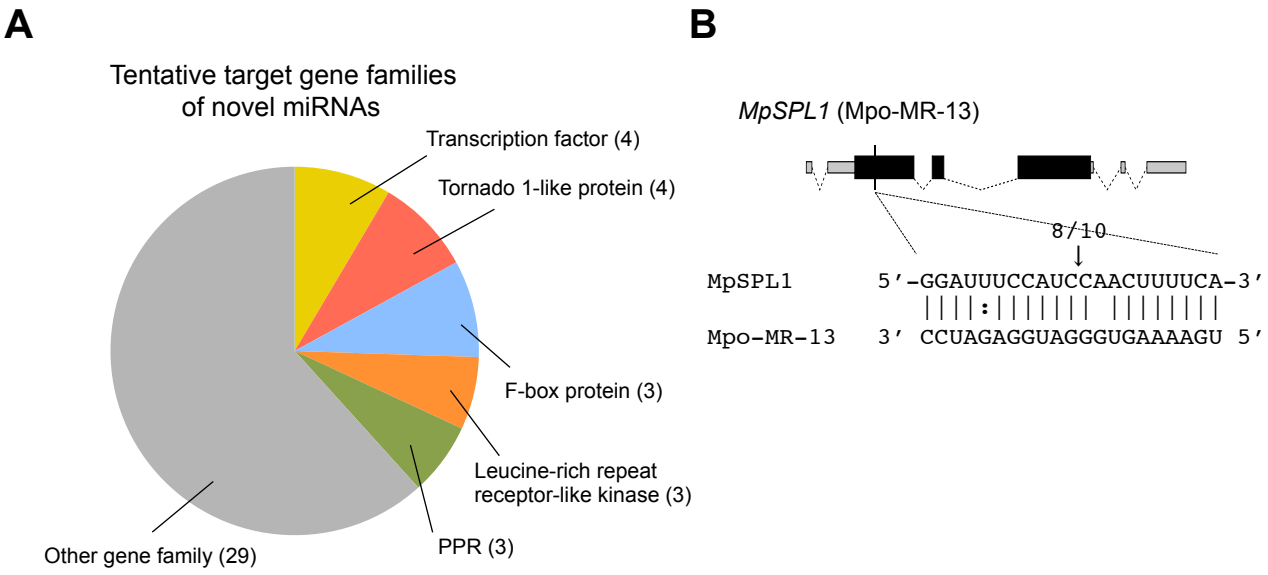


Fig. 11 Prediction of target genes of non-conserved miRNAs.
(A) A pie chart representing the percentage of predicted target gene family of novel miRNAs. The highest scoring gene was counted in the graph. (B) 5′ RACE assay of the Mp*SPL1* gene to detect the cleavage sites in the target mRNA by Mpo-MR-13, one of the novel miRNAs.

A

```
          *   *   *   *   *   *
 T   R   A   L   S   L   L   S   A    aa
ACTCGTGCTCTCTCTTCTGTCAGCT  AtSPL6
 N   C   A   L   S   L   L   S   N    aa
AACTGTGCTCTCTCTTCTGTCAAAT  AtSPL9
 S   C   A   L   S   L   L   S   N    aa
AGCTGTGCTCTCTCTTCTGTCAAAC  AtSPL15
    ||||||  ||||||||||||||
    CACGAGUGAGAGAAGACAGU     At-miR156

 G   R   A   L   S   L   L   S   S    aa
GGGCGTGCTCTCTCTTCTGTCATCT  MpSPL2
 |   ||||||||||||||||||||
 CGACACGAGAGAGAGAAGACC       Mp-miR529c
```

B

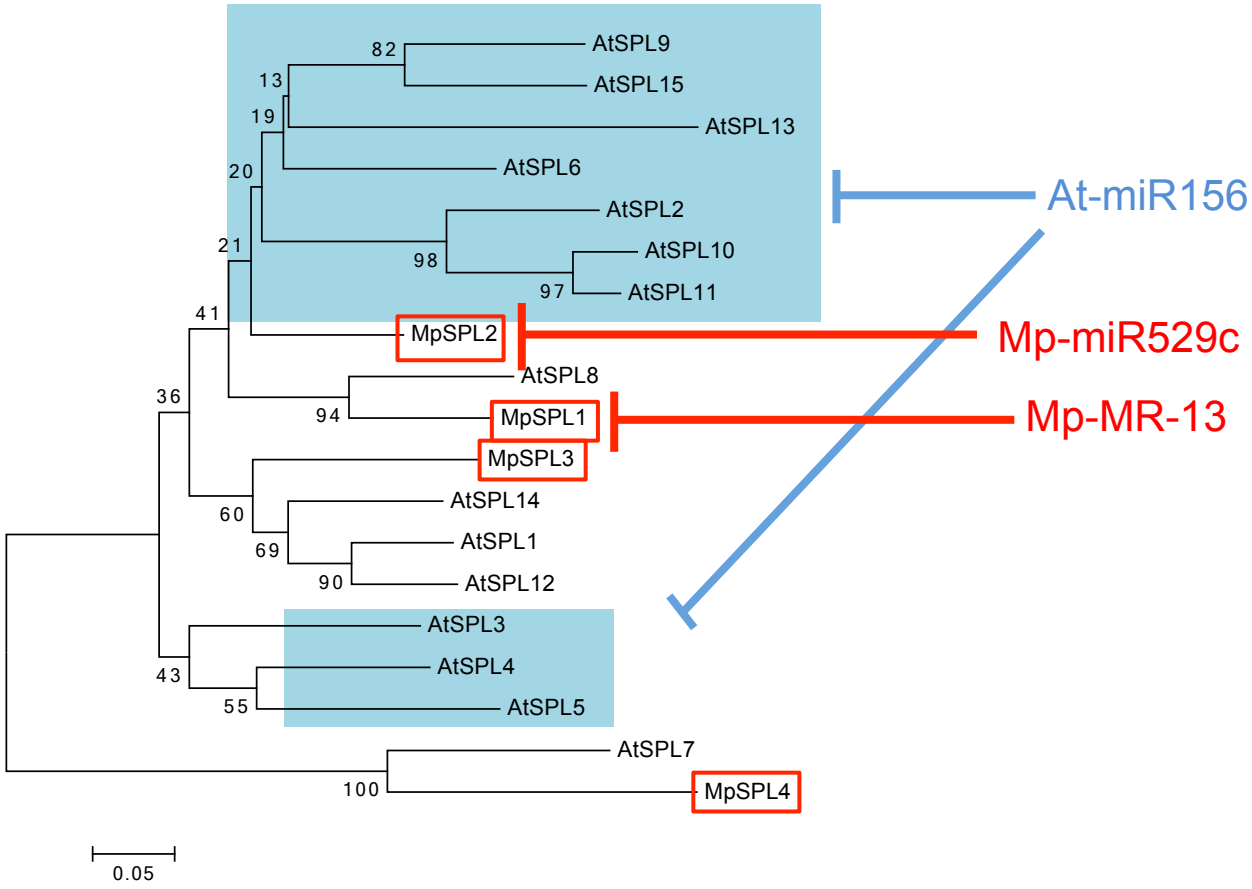


Fig. 12 Multiple regulation system by two distinct miRNAs in *M. polymorpha*.
(A) Comparison of miRNA-target sites of *SPL* genes between *M. polymorpha* and *A. thaliana*. aa: amino acid sequence. (B) An unrooted phylogenetic tree and schematic representation of regulation of SPL family genes in *M. polymorpha* and *A. thaliana*. In *A. thaliana*, miR156 suppresses At*SPL2, 3, 4, 5, 6, 9, 10, 11, 13* and *15* (blue shade box), while miR529c and the novel miRNA, Mp-MR-13, target Mp*SPL2* and Mp*SPL1*, respectively. Numbers are percentages of bootstraps of 1000 replicates. Scale bars represent amino acid substitutions per site. 74 amino acids of SBP domains were used for construction of the phylogenetic tree. Red boxes show four *SPL* genes in *M. polymorpha*.

-Arabidopsis thaliana

	MIRNA loci	Target gene loci	Target family
miR156	10	9	SPL
miR160	3	3	ARF
miR166	9	5	HD-ZIPIII
miR319/159	6	8	MYB

-Marchantia polymorpha

	MIRNA loci	Target gene loci	Target family
miR529c	1	1	SPL
miR160	1	1	ARF
miR166	2	1	HD-ZIPIII
miR319	2	1?	MYB

Fig. 13 Simple regulation networks of conserved miRNAs in *M. polymorpha.*
Comparison of loci numbers of conserved miRNAs and corresponding target genes between *A. thaliana* and *M. polymorpha*.

A

Clade	*A. thaliana*	*P. patens*	*M. polymorpha*	Function	Length of small RNA
DICER-LIKE 1	AtDCL1	PpDCL1a, b	MpDCL1	miRNA	21
DICER-LIKE 2	AtDCL2	-	-	PTGS-siRNA	22
DICER-LIKE 3	AtDCL3	PpDCL3	MpDCL3	RdDM-siRNA	24
DICER-LIKE 4	AtDCL4	PpDCL4	MpDCL4	PTGS-siRNA	21

B

Clade	*A. thaliana*	*P. patens*	*M. polymorpha*	Function
ARGONAUTE 1, 5,10	AtAGO1, 5, 10	PpAGO1a, b, c	MpAGO1	miRNA, PTGS-siRNA
ARGONAUTE 4, 6, 9	AtAGO4, 6, 9	PpAGO4a, b, c	MpAGO4a, b	RdDM-siRNA
ARGONAUTE 2, 3, 7	AtAGO2, 3, 7	-	-	Specific miRNAs, PTGS-siRNA

Table 1 Lists of DCL and AGO proteins in *M. polymorpha*, *P. patens* and *A. thaliana.*

Lists showing DCL and AGO proteins in three land plants, *M. polymorpha*, *P. patens* and *A. thaliana.* Accession numbers of proteins in *A. thaliana* and *P. patens* are listed in the M. M. section.

Name	Locus	Most abundant sequence		RPM#1	RPM#2	RPM#3
Mpo-MR-1	scaffold_3:84045-84285	UCGUAGAACGAUGCUAGACC		87587	12163	41697
Mpo-MR-2	scaffold_7:2690628-2690813	UGCCAUGCCAAUCUUCGGGC		57683	8528	37550
Mpo-MR-3	scaffold_7:2552979-2553185	ACCGGAUUCAUGUCGAGAACG		44359	16768	62489
	scaffold_7:2549779-2549909	ACCGGAUUCAUGUCGAGAACG		3982	2441	15469
Mpo-MR-4	scaffold_120:519599-519688	GACUUGGCCGAACUGUGGAUC		18572	535	23686
Mpo-MR-5	scaffold_12:1546618-1546751	ACUGGAUGUGCGGGGAUCUGA		16405	828	6224
Mpo-MR-6	scaffold_62:989520-989641	CGAACGAACUCCACUGCCCGG		16154	696	3279
Mpo-MR-7	scaffold_48:312105-312288	UUGGAUCCGGAAUUAAUGAAC		14529	732	1310
Mpo-MR-8	scaffold_7:2553739-2553863	UCUCUGAACACCGACAUGUAGG		12167	1989	6035
Mpo-MR-9	scaffold_33:1348679-1348758	AGAGAAUCGAAGUGGAACGCC		9359	2357	7003
Mpo-MR-10	scaffold_103:205958-206125	UCUGGAAUCGUGGAAAGAACA		9222	801	2567
Mpo-MR-11	scaffold_112:186754-186990	CUGUCAAGCUGGAGAAAACCC		8704	7082	17584
Mpo-MR-12	scaffold_26:889832-889984	AAAGUAUCCUGUCGAUGAUA		8143	995	3932
Mpo-MR-13	scaffold_239:43172-43286	UGAAAAGUGGGAUGGAGAUCC		6391	1904	2136
Mpo-MR-14	scaffold_33:328584-328684	UGUCAAGCUUGCGGAUGUCACC		5708	522	5330
Mpo-MR-15	scaffold_95:208994-209107	AUCACUGUUCCUUGCACCGUC		4550	13	95
Mpo-MR-16	scaffold_183:95362-95455	GUGCUUGACUGCAAGAUGGGC		4399	805	1169
Mpo-MR-17	scaffold_327:15843-16005	UCUGGUUCUGCUUUAUGGCCU		4173	3899	6957
Mpo-MR-18	scaffold_39:1337918-1338080	ACCGGUGAGCAUACUCUGAAUG		2303	420	1900
Mpo-MR-19	scaffold_161:61020-61114	UGGUGAGCCAAUUCGGAUCC		2147	1198	1228
Mpo-MR-20	scaffold_55:976306-976388	AGAGGAAACGCAACAGCAUCC		2106	1150	1960
Mpo-MR-21	scaffold_95:390209-390432	CUUUGGCAGCCAAAAGGAGGAC		1949	226	1316
	scaffold_110:110521-110731	CUUUGGCAGCCAAAAGGAGGAG		7	189	1049
Mpo-MR-22	scaffold_49:1026946-1027062	UACUGCGUAGUCGGAUUGCCG		1921	425	1431
Mpo-MR-23	scaffold_50:42319-42436	AAGGACAAGGCCACUGACAAG		1872	548	1321
Mpo-MR-24	scaffold_9:2037097-2037288	UCAAGUUCACUUGCAGGAGCC		1825	1470	6244
Mpo-MR-25	scaffold_6:1906428-1906642	GAGAGAGAGAUUGUGUGCUCC		1679	2487	1941
Mpo-MR-26	scaffold_96:307940-308178	UUUCUGUCCACUACCAUGCCA	GUAUGGUAGUGUACAGGAACC	1601	345	1142
Mpo-MR-27	scaffold_92:147081-147196	UCCGUGGAAGAAGACAAUCUG		1399	1067	6437
Mpo-MR-28	scaffold_7:1962735-1962977	UAGAGGCAACAUAACCGAGAA		1287	1473	2727
Mpo-MR-29	scaffold_152:178959-179087	UAGGAGGUAGAAGAUGCCAGG		1284	466	156
Mpo-MR-30	scaffold_35:587740-588060	UUCCAACGAGUCGGCAGCGCU		1218	302	886
Mpo-MR-31	scaffold_12:905005-905126	GCUGUUUGACGGUUGGGCAAA		1167	124	137
Mpo-MR-32	scaffold_62:46074-46182	UCAUCCAGGAUGAUCAAGAC		1163	957	4429
Mpo-MR-33	scaffold_8:1605213-1605372	UCGUCAUUCGGGCUGCGAAGA		1095	124	475
Mpo-MR-34	scaffold_101:595352-595486	UCCCGUGAGUGUUCCCAGCUG		1080	1750	3996
Mpo-MR-35	scaffold_2:1576816-1576946	GGGGAUGUAGCUCAGAUGGUAGA		900	153	1063
Mpo-MR-36	scaffold_7:1930059-1930185	CGGACUCCAGUUGGCAUGUGC		847	265	138
	scaffold_7:1933152-1933290	CGGACCUCAGUUGGCAUGUGC		71	77	41
Mpo-MR-37	scaffold_56:774050-774180	UAACUGUGAAGUCGUCUAGUC		845	418	1197
Mpo-MR-38	scaffold_53:597649-597740	UCCGAACACCCAUCCAUCCCCC		766	259	560
Mpo-MR-39	scaffold_7:1416128-1416352	ACACGGCGAAGAAGAAUGAUAG		653	536	563
Mpo-MR-40	scaffold_34:1608037-1608137	CGGAUCAUGUGGAGACUCACC		636	216	649
Mpo-MR-41	scaffold_133:150690-150922	CCUGUGUUUGGGAUUGGCCCC		634	646	3872
Mpo-MR-42	scaffold_2:1515661-1515878	UUCCUGUGCAGUCGAUCAAUG		634	262	434
Mpo-MR-43	scaffold_12:1479521-1479630	UCAAUGGUAAACCCCUAGCUG		588	169	1442
	scaffold_12:1479981-1480185	UCAAUGGUAAACCCCUAGCUC		57	19	24
Mpo-MR-44	scaffold_112:421498-421619	CUGGCGAUGAAAGCUGAAGAC		587	310	669
Mpo-MR-45	scaffold_57:59490-59629	UCAGAGAGAAGAGAUGGGACGA		578	140	247
Mpo-MR-46	scaffold_7:2554832-2554955	CGAGUUUAGAUUGGGAAUCUUU		555	265	1178
Mpo-MR-47	scaffold_51:394161-394304	UUUGAUUGUCCACCUUAUGCUU		540	127	301
Mpo-MR-48	scaffold_12:673879-673984	UGAAUUGGACAGAGGUAGAUC		524	17	1
Mpo-MR-49	scaffold_31:157383-157535	UUCCACCCAGUCGAUGCGGGAC		518	2241	4741
Mpo-MR-50	scaffold_75:874061-874167	UCUCGCGAAUCGGAUGGAAGGA		469	30	89
Mpo-MR-51	scaffold_12:1906177-1906265	AUGACCAGGAUUUCAUUGAACAGG		466	87	368
Mpo-MR-52	scaffold_85:248047-248130	UUGGCCAUGAAAGCUGAAGAC		440	539	851
Mpo-MR-53	scaffold_28:154639-154835	UCUCCAAGUUGAUAGAUGCGCA		439	549	1379
Mpo-MR-54	scaffold_47:176529-176617	UCAAAGUCAAAAAUCAUCGAC		439	1565	2844
Mpo-MR-55	scaffold_83:854885-855021	UGUGUUAAUCAAUUGGGGUCC		417	96	492
Mpo-MR-56	scaffold_7:1928748-1928889	CAGGCCAUAGUCGGCAUGUGC		407	107	69
Mpo-MR-57	scaffold_3:1365655-1365793	AUGAUCAGCAAAGUGGAACUU		400	92	178
Mpo-MR-58	scaffold_40:601772-601902	UCAGAGGUAUGAUGAGAACUC		365	144	1261
Mpo-MR-59	scaffold_1:4874562-4874657	UGACAUGCUCUCGAUAUUAUC		352	190	568
Mpo-MR-60	scaffold_4:2788372-2788606	ACGAUUUGAUGAUUCUGGAGA		350	26	104
Mpo-MR-61	scaffold_34:113536-113649	UAAGGUUGGGAUGAAGUAUC		339	189	395
Mpo-MR-62	scaffold_174:272984-273233	UAUCAGGAUCUAACCAAAUC		338	112	413
Mpo-MR-63	scaffold_54:1227155-1227245	UGCCGGAGGCUCUUCAACCUU		333	46	131
Mpo-MR-64	scaffold_39:968078-968194	UCGAGAUUUGAUUUCCCCGAA		316	285	689
Mpo-MR-65	scaffold_38:1510302-1510408	AUCAUAAGAGAUGGUGAGGAU		312	64	89
Mpo-MR-66	scaffold_114:212367-212555	UUGCCAAGCUUGCUAGUGCC		312	89	314
Mpo-MR-67	scaffold_54:346939-347113	ACAGAGGAGAGGGCACGGUUG		311	83	155
Mpo-MR-68	scaffold_5:310930-311011	UCCAUUGUUCAGCAAGAUUCG		263	149	481
Mpo-MR-69	scaffold_108:142150-142347	CUCAAAAUCCACGUUGUCAAG		253	39	56
Mpo-MR-70	scaffold_48:314388-314481	CAUGUUUAGAGAGAAAGAGAG		244	9	7
Mpo-MR-71	scaffold_85:134935-135041	UAACGUCUUCAGAAAGGAGUC		238	152	905
Mpo-MR-72	scaffold_117:211673-211765	UCAGAUCCUCUUGACUAAACU		238	34	167
Mpo-MR-73	scaffold_54:858551-858795	UUCCAGGUUUCGAAUUGUCAGAC		237	7	23
Mpo-MR-74	scaffold_31:572954-573158	ACAAUGCUUUUAGUUCUCCAC		231	41	166
Mpo-MR-75	scaffold_66:841838-841924	UACUCUGAUACUGUGAAGAAUACU		226	77	222
Mpo-MR-76	scaffold_39:964955-965086	GAGUAUCUUGGGCUAUCUUACU		216	137	747
Mpo-MR-77	scaffold_12:468743-468914	CUGUAAGCAGAUUACUCGAC		213	61	34
Mpo-MR-78	scaffold_27:1755729-1755820	AUGCUAGAGCCACUUAAGAGA		211	120	783
Mpo-MR-79	scaffold_83:853391-853543	UAUGCGUCAACCAGCUGGGUU		194	67	378
Mpo-MR-80	scaffold_122:439329-439531	UCGUGGCUGACAUAAUCAACA		194	136	612

Table 2 A list of non-conserved miRNAs in *M. polymorpha*.

Information about non-conserved miRNAs identified by three sequencing runs. Sequences are the most abundant in the sequences mapped to the *MIRNA* cluster. Counts were converted to reads per million (RPM) by dividing by total reads of mapped sequences, except rRNA-derived sequences, and multiplying with 1 million. Listed target candidates are those predicted by the TargetFinder program.

Name	Locus	Most abundant sequence	RPM#1	RPM#2	RPM#3
Mpo-MR-81	scaffold_50:877542-877689	CUCCCAUACUCAGCUAUCGGC	190	213	286
Mpo-MR-82	scaffold_151:68784-68869	UCACAGUACUCUGAUGACAUG	190	66	0
Mpo-MR-83	scaffold_77:725822-726013	UCUAGAGUGCGGUCUUCAUUC	189	84	87
Mpo-MR-84	scaffold_38:684470-684581	UGUUUUUCAAGCUUGUGACAC	186	67	323
Mpo-MR-85	scaffold_9:207020-207212	CGUCCCUGACUCAAUGGUUC	182	56	456
Mpo-MR-86	scaffold_90:47150-47294	ACGACCAAAACUUAGGAUGUC	180	1	1
Mpo-MR-87	scaffold_12:1790513-1790634	UUUCUCUGAAUCUUCUCAUCA	162	30	133
Mpo-MR-88	scaffold_144:386060-386274	ACGGGCAGCUGGUGAGAAGC	162	30	124
Mpo-MR-89	scaffold_55:762278-762718	CAAAGCAGAGUACAUCGGACU	155	44	51
Mpo-MR-90	scaffold_40:1452035-1452145	UCGCAAGUUCGACAGGUUGCC	147	116	222
Mpo-MR-91	scaffold_11:278193-278348	GAAUACCUGGACGUGAACAUCAAC	146	34	62
Mpo-MR-92	scaffold_97:133142-133244	CUAGAUCUCUGUUUGCCUCCC	140	23	50
Mpo-MR-93	scaffold_23:1667684-1667774	CGGAUUGGAGUAAUCACUACC	139	76	239
Mpo-MR-94	scaffold_65:1164983-1165072	CUCAUGACGUAUUUCAUAGCCCCG	137	39	189
Mpo-MR-95	scaffold_125:350234-350321	UUUAGAUCCUGGAUUCCUUCC	129	20	484
	scaffold_50:1363230-1363358	UUUAGAUCCUGGGUUCCUUCC	52	1781	94
	scaffold_50:1353122-1353225	UUUAGAUCCUGGGUUCCUUCC	10	330	28
Mpo-MR-96	scaffold_126:134715-134961	UGAAAUCUUGGGAAGAAAGCGG	126	3	18
Mpo-MR-97	scaffold_54:1241828-1241922	UGUCACACUUGCAAUGAGAGC	125	139	110
Mpo-MR-98	scaffold_42:556285-556470	UCUGUUUGCCAUGGUUUUGCA	122	20	79
Mpo-MR-99	scaffold_170:46306-46476	GACAACUAGAUCCAUUGGAAG	122	87	209
Mpo-MR-100	scaffold_11:1662474-1662587	UCCAGUCGAUUUGUGGGAUCU	118	60	969
Mpo-MR-101	scaffold_73:1001194-1001434	UAGAUCUUUCCCCUCCAUCCA	117	107	149
Mpo-MR-102	scaffold_36:86368-86449	GCGGAUUCUGCUAGAGAGAUC	116	23	65
Mpo-MR-103	scaffold_3:2272922-2273041	CUCCUUUGCUUCUACAAGCUA	106	24	98
Mpo-MR-104	scaffold_125:509868-509955	UGGGGAGCUGGUUCCAUGGAC	105	53	121
Mpo-MR-105	scaffold_221:49990-50217	UUCCAUGAGGUCCUUGAAGGC	92	109	99
Mpo-MR-106	scaffold_33:1412301-1412411	UGAUUCCUCAUAUCGUCCAAC	91	14	23
Mpo-MR-107	scaffold_107:576892-577011	UAACCUGGAACCUUGGGAAGAAG	88	102	191
Mpo-MR-108	scaffold_98:455391-455485	CAGCAAAAGAGAGAAGAGCUG	88	9	72
Mpo-MR-109	scaffold_9:2212657-2212756	CUUUGUCUCCCCUGUGGCAAC	86	300	172
Mpo-MR-110	scaffold_165:215994-216096	ACGGACCGGAGCUCGUGUGUC	78	1	19
Mpo-MR-111	scaffold_13:107096-107272	UGAAUGGGUUACACAGAGGGC	78	23	66
Mpo-MR-112	scaffold_2:1111839-1111934	CAGACCUAGAAUAUGAUGCAC	75	166	571
Mpo-MR-113	scaffold_25:1627721-1627963	UUAGAGUCUCAUUCACAUCCC	74	137	131
Mpo-MR-114	scaffold_51:1287931-1288049	CGCUGGAACUCACAACAGACA	74	50	60
Mpo-MR-115	scaffold_276:20217-20299	UGCUAGAUUGCCCUGUUCACU	74	17	30
Mpo-MR-116	scaffold_161:204258-204349	UAGUGAUCAGUUGGAAGUCAG	73	39	86
Mpo-MR-117	scaffold_152:80051-80293	AGAGAUGAUCAGGUGGGGAGU	72	59	771
Mpo-MR-118	scaffold_69:530409-530491	UGCUAGAUUGCCCUGUUCACU	72	16	42
Mpo-MR-119	scaffold_11:207069-207164	CUUGCAAGUGAUGGGAUUGUC	70	43	140
Mpo-MR-120	scaffold_1:1081384-1081467	CUGGUUUGUUUCUCUAUUCCG	70	14	31
Mpo-MR-121	scaffold_53:513107-513226	GUCGCAGGUUUGGGAAACUAC	67	29	81
Mpo-MR-122	scaffold_64:240018-240247	CCUCUCUUUGUCUGGAACACU	67	39	58
Mpo-MR-123	scaffold_44:1271329-1271445	UGAUAAAGUCCUGUCAGUCCU	66	24	279
Mpo-MR-124	scaffold_13:885900-886028	UCCAUGGAAGAAGACAAUAUG	66	19	67
Mpo-MR-125	scaffold_52:301885-302109	UGAACAGACAUAUUGAAGAGGC	64	44	131
Mpo-MR-126	scaffold_137:195855-195960	UCUGGUGCUAGAAGCUGGUUG	62	36	85
Mpo-MR-127	scaffold_149:264120-264259	GUGGCCUAGACUUGAACUACC	61	10	95
Mpo-MR-128	scaffold_12:382253-382331	CUUUGAUCCAGCACUCGGGUG	61	24	157
Mpo-MR-129	scaffold_68:60269-60407	UGGACUCUCGUAAGCAUGAACU	58	26	18
Mpo-MR-130	scaffold_4:1241202-1241301	UGGAUUCCUGUAAGCAUGAACU	44	33	18
Mpo-MR-131	scaffold_140:115586-115748	GCACCAAUCGGACUGUUCUACG	56	47	134
Mpo-MR-132	scaffold_6:2337054-2337200	CAAUUCCAUCAUGCGGGCAAG	52	9	32
Mpo-MR-133	scaffold_170:38365-38512	UGCCUUCUAGAAUUGAUCGGA	51	16	62
Mpo-MR-134	scaffold_1:401494-401733	UUGAAGGAUGAAACUGAUCU	51	11	27
Mpo-MR-135	scaffold_12:186439-186681	CGCAGAAUUUGGUACGAGGAG	51	40	78
Mpo-MR-136	scaffold_178:225989-226187	UUUGUCGUGGCCUCGAAUUCUGUC	50	27	100
Mpo-MR-137	scaffold_189:49590-49736	UAUCAGCGUUUGGAGCUCGAG	50	30	93
Mpo-MR-138	scaffold_44:533103-533342	UGCACUGACAAGAACGCAGGA	48	50	39
Mpo-MR-139	scaffold_5:372593-372671	AAGUGCAGAUGAGGAUCAGAGC	47	4	19
Mpo-MR-140	scaffold_29:824459-824535	CCUGAACACUUACCUAAACGAG	45	4	9
Mpo-MR-141	scaffold_71:919906-920051	UCUGAAGACGCCACCCAGAAU	44	17	76
Mpo-MR-142	scaffold_7:1021160-1021275	GCGGAUGCUUGGAUGAAUGCG	44	7	29
Mpo-MR-143	scaffold_176:210069-210196	GAGACGUUUUCGUACAGGAAG	44	107	36
Mpo-MR-144	scaffold_113:165810-165912	CGAGGUUCAUGGAAAUUUCUGAGA	40	9	60
Mpo-MR-145	scaffold_91:853948-854052	ACUGACGAUAUCUGAAAAUAU	39	14	67
Mpo-MR-146	scaffold_115:155888-156029	ACUCUCGUCUGCUCUGACCC	39	4	20
Mpo-MR-147	scaffold_5:290076-290177	UCUGUUUGGAUCACAUCGAGC	39	39	92
Mpo-MR-148	scaffold_3:1182026-1182266	UCUGGCUUUUAGAUCCUGGGU	39	6	191
Mpo-MR-149	scaffold_111:56124-56267	AUUUAGAUGUGUGUUUCAGGC	37	13	33
Mpo-MR-150	scaffold_50:1351524-1351605	AACGUAGAGCUCAUGAUCACA	36	92	4
Mpo-MR-151	scaffold_32:1388519-1388751	UUGUGGCUCUUCUGGCGAUUC	35	3	26
Mpo-MR-152	scaffold_100:562985-563196	ACACAUGCGGACAAUAGGUAU	35	3	17
Mpo-MR-153	scaffold_23:661275-661609	CUCUGCUUCUCCACCUAUCAA	34	27	330
Mpo-MR-154	scaffold_15:1440250-1440403	CUUCUACUCUGGACUUAUGCC	33	3	25
Mpo-MR-155	scaffold_8:360610-360713	UCCAUGCUCACGAAAUCCAUC	32	17	12
Mpo-MR-156	scaffold_41:825488-825603	UAUCCAGGAAAAUGACGCCU	31	79	84
Mpo-MR-157	scaffold_54:838606-838715	UGACAAGCUGACAAUCGGGC	31	3	16
Mpo-MR-158	scaffold_27:1757976-1758109	UUGCUUAUACAUUCUCCUUCC	29	86	618
Mpo-MR-159	scaffold_127:65609-65679	CCUCUCGACUGUGAUGGAUCU	28	0	25
Mpo-MR-160	scaffold_170:46010-46166	GAUAACUGAAGGCAUUGCAAG	28	17	42

Table 2 A list of non-conserved miRNAs in *M. polymorpha*.
(Continue)

Name	Locus	Most abundant sequence		RPM#1	RPM#2	RPM#3
Mpo-MR-161	scaffold_27:367294-367372	UUACUCACCGAAUCUCUCCUC			28	9	11
Mpo-MR-162	scaffold_95:807851-807957	GACCUAGACAGCAAUCGGGGC			28	10	34
Mpo-MR-163	scaffold_4:718950-719027	AGAUGAAUUCUGCUCUGGUUG			28	10	25
Mpo-MR-164	scaffold_129:300851-300941	UCUUGACGGGGAUUCAUUCGU			28	11	27
Mpo-MR-165	scaffold_81:311442-311548	CAUGAAGGUGGAAGUUUUGGG			27	6	2
Mpo-MR-166	scaffold_2:1433025-1433126	UAUUAUGCUUCCUACAUCGUG			26	14	57
Mpo-MR-167	scaffold_107:577810-577921	CUUUCCAAUUUCUCGGUCAG			26	1	19
Mpo-MR-168	scaffold_36:138165-138335	CUCUUUAAUCCCAACUCCACC			25	3	1
Mpo-MR-169	scaffold_22:194583-194711	GCUCUGAUCAUCAAUGAAGGA			25	1	0
Mpo-MR-170	scaffold_27:297531-297625	UCAUGUAUAGGGAGUCUGGUC			25	10	23
Mpo-MR-171	scaffold_56:826849-826949	GGAGAAAACGAUGGAAAGUGC			24	14	13
Mpo-MR-172	scaffold_1:1935566-1935731	GGUGGCUGUAUGUACCAGUCU	AAGCAGGUGGCUGUAUGUACC		23	9	30
Mpo-MR-173	scaffold_69:173934-174030	UGAUGCGACAUUAUCUACCGAGUA			23	3	18
Mpo-MR-174	scaffold_107:580431-580525	UAGCGUGGAUCUCUAUUCCUUG			23	3	67
Mpo-MR-175	scaffold_82:149888-150011	UGUUAGAAAUGUCCACACUUC			22	14	69
Mpo-MR-176	scaffold_9:1576904-1577046	CAUGAUGACUAUGCAGUGACA			21	1	24
Mpo-MR-177	scaffold_10:2531093-2531199	UAUUUGCCAUUAAUUUCCCUC			21	1	96
Mpo-MR-178	scaffold_216:11644-11881	GAAGGAGAUCGGUUACAGCGA			20	199	87
Mpo-MR-179	scaffold_43:378239-378338	UGUGUGAGAAGAGGCCAAUGU			19	1	51
Mpo-MR-180	scaffold_11:1554045-1554201	UAUAGAAUCUCCUGGGGCAUC			19	31	44
Mpo-MR-181	scaffold_67:1025331-1025415	UCCUUCAUGAGAAUGGCCCAU			19	14	48
Mpo-MR-182	scaffold_69:628976-629067	UGAUAUCCAACCAAUCCUCAGUA			19	16	14
Mpo-MR-183	scaffold_71:592080-592194	GCACUUGCUGUGUGAUGCACG			19	0	11
Mpo-MR-184	scaffold_6:1911348-1911492	CUACGAUCUUGAAACGAGCAG			17	26	33
Mpo-MR-185	scaffold_1:4873549-4873639	UAAUGUCGAGAGCAUGUCACA			17	14	46
Mpo-MR-186	scaffold_23:452063-452162	UUAGCAAGCAGUGGGCAUUCC			17	19	11
Mpo-MR-187	scaffold_40:448112-448214	CACAUAGCAGAACAUGGACUC			17	0	0
Mpo-MR-188	scaffold_46:235688-235854	CCAUGAAAUCAAUCUCGUCGU			17	13	30
Mpo-MR-189	scaffold_62:46806-47029	GAUCCAUCGUCGGCCUCGCGGCA			17	4	10
Mpo-MR-190	scaffold_110:306223-306371	CUCUGCCUCGCUCUCUCAACC			17	33	9
Mpo-MR-191	scaffold_115:261704-261791	UUUGCCUAUCUCUACAAGGGG			17	4	66
Mpo-MR-192	scaffold_51:1229319-1229471	UCCCGGUUGUCUUCACUAUUU			15	9	54
Mpo-MR-193	scaffold_5:247160-247255	AUUUUCCCCUUUAGUAAUCCC			14	3	29
Mpo-MR-194	scaffold_186:185299-185440	CUUUUAAGCUCCAGGAUUGAA			14	6	2
Mpo-MR-195	scaffold_68:61417-61524	UGGAUUCCUGUAAACAUGAACU			14	7	2
Mpo-MR-196	scaffold_89:704453-704550	CUCUGCUUACCUCUCUCAGCU			14	16	87
Mpo-MR-197	scaffold_7:2482194-2482298	UAUAUCAGAACCAUCUAGAUC			13	6	10
Mpo-MR-198	scaffold_114:212873-212958	GCGCUAGCAAGGUUGACCACU			13	7	6
Mpo-MR-199	scaffold_6:2196689-2196802	UGAACCAAAGUGAUACUUAGU			12	3	7
Mpo-MR-200	scaffold_15:369581-369669	UCGCUGGGGAUAUGUUUCGUC			12	13	7
Mpo-MR-201	scaffold_122:514722-514860	CCAAGAUCUCUAGUCGGCACU			10	1	11
Mpo-MR-202	scaffold_75:125482-125615	CCUAUAACAACCUGUGUCGCC			9	3	17
Mpo-MR-203	scaffold_198:121774-121885	UCUUAUGCUUCCAGACUUUGA			7	9	24
Mpo-MR-204	scaffold_31:467344-467466	ACAAAUGUUGGGAUUGUUCC			7	19	25
Mpo-MR-205	scaffold_68:59806-59915	UAGAUUCUCGUAAGCAUGAACU			7	3	8
Mpo-MR-206	scaffold_216:12222-12367	UCUUUGAAGAAAAAUCGAACC			6	94	81
Mpo-MR-207	scaffold_8:1981888-1981987	UGUUUGCUGAUUUUACAUCUC			4	3	40
Mpo-MR-208	scaffold_79:108278-108375	UCCAGAGAUCCGUGAAGGCCU			4	61	12
Mpo-MR-209	scaffold_27:1733862-1734017	UCAUUCGAAGACUGGCUCCUG			2	3	85
Mpo-MR-210	scaffold_52:189547-189791	GUUUGUGUCUAGGCUAAGCUC			2	3113	25
Mpo-MR-211	scaffold_3:892937-893044	CUUGCAAGCGGUUAGCAGCCU			1	1	24
	scaffold_3:893348-893471	CUUGCAAGCGGCUAGCAACC			0	3	46
Mpo-MR-212	scaffold_34:75805-76191	UAGAUGAUGGGUGUUGAUGCA			1	908	8
Mpo-MR-213	scaffold_17:904387-904613	CCGCGAUUCUUACAUGGAAUG			0	0	225

Table 2 A list of non-conserved miRNAs in *M. polymorpha*.
(Continue)

Name	Predicted target genes
Mpo-MR-1	-
Mpo-MR-2	-
Mpo-MR-3a	HOS15 CUFF16522 WD-40 (score4)
Mpo-MR-3b	WD40
Mpo-MR-4	LHCB2
Mpo-MR-5	-
Mpo-MR-6	?
Mpo-MR-7	-
Mpo-MR-8	Tornado1
Mpo-MR-9	-
Mpo-MR-10	Tornado1
Mpo-MR-11	UBP11 CUF1352
Mpo-MR-12	-
Mpo-MR-13	SPL
Mpo-MR-14	-
Mpo-MR-15	-
Mpo-MR-16	unknown function CUFF195 (score4)
Mpo-MR-17	F-BOX CUFF3068
Mpo-MR-18	-
Mpo-MR-19	tornado1
Mpo-MR-20	VPS35 CUFF19354 (score4)
Mpo-MR-21a	receptor like kinase CUFF9033
Mpo-MR-21b	receptor like kinase CUFF9033
Mpo-MR-22	-
Mpo-MR-23	-
Mpo-MR-24	tornado1 CUFF8092
Mpo-MR-25	-
Mpo-MR-26	-
Mpo-MR-27	Acid phosphatase
Mpo-MR-28	TPR PPR CUFF9569
Mpo-MR-29	Galactose oxidase CUFF21989
Mpo-MR-30	Tetratricopeptide repeat, CUFF20450
Mpo-MR-31	-
Mpo-MR-32	F-box CUFF3901, G-box binding CUFF4563
Mpo-MR-33	-
Mpo-MR-34	-
Mpo-MR-35	-
Mpo-MR-36a	zinc-finger CUFF14052
Mpo-MR-36b	zinc-finger CUFF14052
Mpo-MR-37	HD-zip17 CUFF6463
Mpo-MR-38	-
Mpo-MR-39	CUFF13757 PHT2;1 score=4
Mpo-MR-40	-
Mpo-MR-41	GRAS CUFF21052
Mpo-MR-42	-
Mpo-MR-43a	oxidoreductase CUFF466
Mpo-MR-43b	oxidoreductase CUFF466
Mpo-MR-44	-
Mpo-MR-45	-
Mpo-MR-46	-
Mpo-MR-47	-
Mpo-MR-48	O-fucosyltransferase CUFF11467
Mpo-MR-49	Receptor-like CUFF706 (3.5)
Mpo-MR-50	-
Mpo-MR-51	-
Mpo-MR-52	-
Mpo-MR-53	TIR-NBS-LRR CUFF21088, alpha/beta-hydrolases CUFF21881
Mpo-MR-54	-
Mpo-MR-55	-
Mpo-MR-56	-
Mpo-MR-57	-
Mpo-MR-58	-
Mpo-MR-59	PHT2;1 CUFF2887
Mpo-MR-60	-
Mpo-MR-61	CUFF20436 P-loop containing nucleoside triphosphate hydrolases superfamily protein
Mpo-MR-62	-
Mpo-MR-63	-
Mpo-MR-64	-
Mpo-MR-65	-
Mpo-MR-66	-
Mpo-MR-67	-
Mpo-MR-68	-
Mpo-MR-69	-
Mpo-MR-70	-
Mpo-MR-71	--
Mpo-MR-72	-
Mpo-MR-73	-
Mpo-MR-74	-
Mpo-MR-75	-
Mpo-MR-76	-
Mpo-MR-77	-
Mpo-MR-78	-
Mpo-MR-79	-
Mpo-MR-80	-
Mpo-MR-81	-
Mpo-MR-82	-
Mpo-MR-83	-
Mpo-MR-84	-
Mpo-MR-85	-
Mpo-MR-86	-
Mpo-MR-87	-
Mpo-MR-88	-
Mpo-MR-89	-
Mpo-MR-90	-
Mpo-MR-91	-
Mpo-MR-92	-
Mpo-MR-93	-
Mpo-MR-94	-
Mpo-MR-95a	PPR CUFF17540
Mpo-MR-95b	PPR CUFF17540
Mpo-MR-95c	PPR CUFF17540
Mpo-MR-96	-
Mpo-MR-97	-
Mpo-MR-98	-
Mpo-MR-99	-
Mpo-MR-100	-
Mpo-MR-101	-
Mpo-MR-102	-
Mpo-MR-103	-
Mpo-MR-104	-
Mpo-MR-105	-
Mpo-MR-106	-
Mpo-MR-107	-
Mpo-MR-108	-
Mpo-MR-109	-
Mpo-MR-110	-
Mpo-MR-111	Kinesin CUFF13656, CUFF11979
Mpo-MR-112	RNA binding CUFF16357
Mpo-MR-113	RNI-like protein emb2004 CUFF15671 F-box
Mpo-MR-114	-
Mpo-MR-115	-
Mpo-MR-116	-
Mpo-MR-117	-
Mpo-MR-118	-
Mpo-MR-119	-
Mpo-MR-120	-
Mpo-MR-121	-
Mpo-MR-122	-
Mpo-MR-123	-
Mpo-MR-124	-
Mpo-MR-125	-
Mpo-MR-126	-
Mpo-MR-127	-
Mpo-MR-128	PPR CUFF3343, WD40 repeat CUFF5511 (3.5)
Mpo-MR-129	-
Mpo-MR-130	-
Mpo-MR-131	-
Mpo-MR-132	-
Mpo-MR-133	-
Mpo-MR-134	-
Mpo-MR-135	-
Mpo-MR-136	-
Mpo-MR-137	-
Mpo-MR-138	-
Mpo-MR-139	-
Mpo-MR-140	-
Mpo-MR-141	-
Mpo-MR-142	-
Mpo-MR-143	-
Mpo-MR-144	-
Mpo-MR-145	-
Mpo-MR-146	-
Mpo-MR-147	-
Mpo-MR-148	CUFF6835(3.5), ELM2
Mpo-MR-149	-
Mpo-MR-150	-
Mpo-MR-151	-
Mpo-MR-152	-
Mpo-MR-153	-
Mpo-MR-154	-
Mpo-MR-155	-
Mpo-MR-156	-
Mpo-MR-157	-
Mpo-MR-158	F-box CUFF5584 (4)
Mpo-MR-159	-
Mpo-MR-160	-
Mpo-MR-161	-
Mpo-MR-162	receptor like kinase CUFF9033
Mpo-MR-163	-
Mpo-MR-164	-
Mpo-MR-165	receptor like kinase CUFF9033
Mpo-MR-166	-
Mpo-MR-167	-
Mpo-MR-168	-
Mpo-MR-169	-
Mpo-MR-170	-
Mpo-MR-171	-
Mpo-MR-172	-
Mpo-MR-173	-
Mpo-MR-174	-
Mpo-MR-175	-
Mpo-MR-176	-
Mpo-MR-177	-
Mpo-MR-178	-
Mpo-MR-179	-
Mpo-MR-180	-
Mpo-MR-181	-
Mpo-MR-182	-
Mpo-MR-183	-
Mpo-MR-184	-
Mpo-MR-185	-
Mpo-MR-186	-
Mpo-MR-187	-
Mpo-MR-188	-
Mpo-MR-189	-
Mpo-MR-190	-
Mpo-MR-191	-
Mpo-MR-192	-
Mpo-MR-193	-
Mpo-MR-194	-
Mpo-MR-195	-
Mpo-MR-196	-
Mpo-MR-197	-
Mpo-MR-198	-
Mpo-MR-199	-
Mpo-MR-200	-
Mpo-MR-201	-
Mpo-MR-202	-
Mpo-MR-203	-
Mpo-MR-204	-
Mpo-MR-205	-
Mpo-MR-206	-
Mpo-MR-207	-
Mpo-MR-208	-
Mpo-MR-209	-
Mpo-MR-210	phosphoenolpyrvate carboxylase 4
Mpo-MR-211a	-
Mpo-MR-211b	-
Mpo-MR-212	-
Mpo-MR-213	unknown function CUFF6691

Table 3 Non-conserved miRNAs are targeting various genes in *M. polymorpha*.
A list of predicted target genes of novel miRNAs identified in this study. CUFF number is name of the gene in EST database of *M. polymorpha*.

Chapter II

Genetic analysis of conserved microRNAs in *M. polymorpha*

Introduction

Roles of miRNAs in plant development have been reported in many studies mainly using *A. thaliana*, the most popular model land plant. *Arabidopsis* mutants of *DCL1* gene, which encodes a core protein of miRNA processing, show various developmental phenotypes. Null mutants like *dcl1-5* and *dcl1-10* show embryonic lethality and are not able to continue development in the middle of embryogenesis (Nodine and Bartel 2010). Weak mutant alleles including *dcl1-7*, *dcl1-9* and *dcl1-100* also show severe developmental phenotypes like female sterility and a delay of flowering (Ray *et al*. 1996, Jacobsen *et al*. 1999). These data strongly suggest that miRNAs are indispensable for plant proper development and viability. While *DCL1* mutants indicate importance of a bulk of miRNAs, roles of individual miRNA were also characterized by the respective genetic studies. The first report that an individual miRNA has developmental roles was about miR319. miR319 was initially named miR-JAW; its mutants showed serration of leaves by insertion of exogenous enhancer fragment into *JAW* locus (Palatnik *et al*. 2003). miR319 controls correct leaf development through repression of TEOSINTE BRANCHED/CYCLOIDEA/PCF 2, 3 and 4 (TCP2, 3 and 4) transcription factor genes. miR159 has an almost identical sequence as miR319, but slight difference causes

distinct target recognition that represses MYB transcription factors (Palatnik *et al*. 2007).

By characterizing a gene or biological factor in two species belonging to distinct taxa, information on both similarity and diversity of its function are provided. While it is revealed that some miRNAs are conserved in land plant lineages, whether their biological roles are little known except in seed plant species. Concerning bryophytes, only a few studies are present that investigate the roles of each miRNA in the moss, *P. patens* (Saleh *et al.* 2011, Cho *et al.* 2012). However, roles of the other conserved miRNAs are not investigated in bryophytes by now.

As I mentioned in the General Introduction chapter, studies of miRNA-mutants are few. This is because plant miRNA loci are frequently multiplicated and it is rare that T-DNA insertions or fatal nucleotide substitutions occur in short *MIRNA* genes. Function of plant miRNA is tightly related to that of target genes due to high complementarity between miRNA and target mRNA in plants. To analyze the function of individual miRNA, transferring modified target gene constructs having silent nucleotide substitutions in miRNA-target site without changing coding amino acid sequence unchanged has been used as alternative materials for miRNA research. Another way to reveal roles of a certain miRNA is to use

plants overexpressing a specific miRNA.

Recently, several genome editing techniques were invented and drastically improved by a number of researchers (Bibikova *et al*. 2003, Boch *et al*. 2009, Jinek *et al*. 2012). One of them, CRISPR/Cas9 system, originates from immune system of bacteria against DNA virus (Ishino *et al*. 1987, Bolotin *et al*. 2005) and is now becoming a popular technique. Cas9 is a DNA cleaving protein that induces DNA double-stranded break (DSB) with a guide RNA (gRNA) having 18-20 nt complementary sequence to the target DNA sequence (Doudna and Charpentier 2014). After introduction of Cas9 gene and specifically designed gRNA gene, the DNA sequence in the genome would be searched and cleaved at a specific site in the genome. Then, some base modification would result during the process where some DNA repairing occurs without possible template guidance after DSB (Fig. 18A). Such changes would result in phenotypic mutants at a high frequency.

In this chapter, I established *M. polymorpha* transformants of which miRNA expression levels were modified, using miRNA overexpression and CRISPR/Cas9 genome editing transgenes. Then I explored developmental roles of miRNAs in *M. polymorpha*. These mutants showed severe morphological phenotypes and these data suggest that miRNAs are

essential for proper morphogenesis and differentiation patterning in *M. polymorpha* development like other land plants.

Materials and Methods

-Plant materials and growth conditions

Growth conditions were same as Chapter I. Tak-1 (male) was used as WT for transformation.

-Construction of miRNA-OX transgenes

First, pMpGWB100-p35S was constructed by PCR amplification of CaMV35S promoter sequence from pMpGWB102 vector having hygromycin selection marker gene (Ishizaki *et al*. 2015b) and ligated into pMpGWB100 after *Hind*III and *Pst*I digestion. Next, surrounding sequences of mature miRNA were amplified by PCR. Forward and reverse primers were designed at 100 base upstream and downstream from the ends of hairpin structure predicted by RNAfold. Then, PCR fragments were ligated into pMpGWB100-p35S vector after *Sal*I and *Kpn*I digestion. All primers are listed in Table S1.

-Construction of CRISPR-miRNA transgenes

For construction of miRNA-targeting vectors, pMpGE010/011 (binary vector) and pMpGE_En03 (entry vector) were used (provided by Dr. Kohchi's Lab and Dr.

Hara-Nishimura' Lab in Kyoto Univ.) (Fig. 18A). First of all, gRNAs were designed for miRNA targeting. CasOT program (Xiao *et al*. 2014) was used to find PAM sequences and avoid off-target binding. I designed two gRNAs for each miRNA precursor. Every gRNA has at least two mismatches for other off-target candidate sites in seed region of gRNA. Next, annealed oligoes were inserted into the entry vector pMpGE_En03 after cleavage by type II restriction enzyme, *Bsa*I (Fig. 18A). Next, LR reaction was performed between pMpGE_En03-miRNA and pMpGE010/011 (010: miR160, 166a, 319a, 529c; 011: miR166b, 319b) by LR clonase (Invitrogen, Thermo Fisher). Then, these constructs were named CRISPR-miR160, -miR166a, -miR166b, -miR319a, -miR319b and –miR529c, respectively.

-Establishment of transformed plants by Agrobacterium with regenerated thalli

To produce transformed plants, an *Agrobacterium*-mediated transformation method with regenerated thalli was used as reported previously (Kubota *et al*. 2003, Fig. 14A). First, the apical side of thallus grown for 14 d from gemma on 1/2 Gamborg B5 medium containing 1% agar was cut by scalpel, followed by incubation for 3 d on 1/2 Gamborg B5 medium containing 1% agar and 1% sucrose. Then, regenerated thalli were co-incubated with

Agrobacterim tumefaciens (*Rhizobium radiobacter*) GV2260 strain containing transgene plasmids for 3 d in M51C liquid medium with 100 μM acetocyringone (Wako). Thalli were transferred to selective medium containing 10 ng/μl hygromycin or 0.5 μM chlorosulfuron. Independent T1 lines were cultivated after selection, and independent plants from one gemma were called G1 lines. All experiments were performed with G1 lines.

-Northern blot analysis

The protocol was same as that in Chapter I. Probes used for hybridization and detection in northern blot analysis were listed in Table S1.

-Quantitative RT-PCR (qRT-PCR)

qRT-PCR was performed to quantify mRNA accumulation levels. Total RNA was extracted from Tak-1 mature thalli (2-4 weeks after gemma). The PrimeScript RT reagent Kit with gDNA Eraser (TaKaRa) was used to synthesize cDNA according to the manufacturer's protocol. Quantitative PCR was performed with the StepOnePlus Real Time PCR System (Applied Biosystems; Life Technologies) with KAPA SYBR Fast qPCR kit (KAPA

BIOSYSTEMS). In all experiments, Mp*EF1a* gene was used as the internal control. Expression levels were averaged by two technical replicates and three biological replicates. Primer sets for PCR amplification and the amplified sequences are listed in Table S1.

Results

-Construction of miRNA-overexpression plants by Agrobacterium-mediated transformation

Profiling and target prediction/validation showed that miRNAs are expressed and possibly function as regulatory molecules in *M. polymorpha* as in other land plants. To demonstrate that such miRNAs identified in this study have function in *M. polymorpha*, I constructed transformants of *M. polymorpha* expressing one of conserved miRNAs ectopically and observed the effects of them on thallus development. In order to perform genetic analysis, I introduced transgenes into Tak-1 WT male plants by the *Agrobacteirum*-mediated transformation method using regenerated thalli (Kubota *et al*. 2013, Fig. 14A). I selected four conserved miRNAs, miR160, miR166a, miR319b and miR529c in this study. First, the surrounding sequences of mature miRNAs were PCR-amplified and cloned into downstream of CaMV35S promoter in pMpGWB vector (Ishizaki *et al*. 2015b) (Fig. 14B). By *Agrobacterium tumefaciens*-mediated transformation using Tak-1 thalli regeneration system, I established miRNA overexpression (miRNA-OX) lines and named them 35S-Mp*MIR160, -166a, -319b* and *-529c*, respectively (Figs. 14, 16). Northern blot analysis showed overexpression of respective miRNAs compared to WT (Fig. 15A). Consequently, I could

judge that fold-back sequences around miRNAs identified by sequencing were functional to express mature miRNAs and that miRNA-OX lines could be used for further analysis.

-miRNA overexpression induced abnormal morphology in thalli and gemma cups

Most miRNA-OX lines as described above showed obvious morphological changes in thalli (Fig. 16A-O). Thalli of miR160-OX and miR166a-OX lines showed curling to ventral side (Fig. 16B, C, G, H). On the other hand, miR319b-OX line showed curling to dorsal side and elongated rhizoids because of upper curling (Fig. 16D, I). These data suggested these miRNAs have important roles in proper horizontal expansion of thalli. I could not discuss miR529c function on thalli development because no obvious changes were observed in miR529c-OX plants.

Gemma cups of miR166a-OX and miR319b-OX lines were also abnormal, compared to wild-type plants. Thalli of miR166a-OX line showed narrowed gemma cups (Fig. 16M). miR319-OX line showed no visible cups although gemmae were generated (Fig. 16N). The observation suggested that miR160, miR166 and miR319 have important roles in regulation of proper horizontal expansion of thalli and differentiation of gemma cups in *Marchantia*.

-qRT-PCR analysis of predicted target mRNAs

In order to check whether target genes were down-regulated as expected, I analyzed accumulation level of mRNA of predicted target genes in miRNA-OX plants by quantitative RT-PCR (qRT-PCR) (Fig. 17A). We designed primers of sequences upstream and downstream of predicted miRNA-binding sites. As expected, most of the miRNA-OX lines showed a decreased accumulation of mRNA of tentative target genes (Fig. 17B). Compared with WT plant, mRNA accumulation level of Mp*HD-ZIPIII* and Mp*SPL2* was significantly reduced in two lines of miR166a-OX and miR529c-OX as predicted (Fig. 17B). In contrast, expression level of Mp*MYB33-like* gene mRNA was not reduced at a statistically significant level in miR319b-OX plants unexpectedly (Fig. 17B). Considering that miR319 was overexpressed in the transformed lines (Fig. 15A) and the cleavage by miR319 was detected by RACE assay (Fig. 9B), it was suggested that cleavage of Mp*MYB33-like* by miR319 plays a minor role but rather translational repression operates mainly. Detection of mRNA accumulation levels showed that miRNAs really repress target genes as shown by RACE analysis in Chapter I. Construction of miRNA overexpression plants and their observation

provided me a feasible method to tell whether or not identifed miRNAs by small RNA sequencing have biological roles.

-Construction of miRNA-mutant plants by CRISPR/Cas9 technique

Sequencing analysis in Chapter I showed the number of loci of each *MIRNA* gene in *M. polymorpha* is one or two in most cases (Fig. 7C, Table 2). It is considered that low redundancy of *MIRNA* genes is an advantageous character for reverse genetic mutant analysis. Taking this advantage, I aimed to construct miRNA mutants that have defects in down-regulation of target genes and to investigate developmental roles of conserved miRNAs in *M. polymorpha.*

I applied CRISPR/Cas9 genome editing technique (Fig. 18A). While co-injection of Cas9 mRNA and gRNA is frequently used in studies using mammal cells, it was not applicable to plants. In *A. thaliana* and *M. polymorpha*, CRISPR/Cas9 method is successfully applied with *Agrobacterium*-mediated transformation, recently (Feng *et al*. 2013, Sugano *et al*. 2014). Based on such reports I performed *Agrobacterium*-mediated transformation with Cas9/gRNA vector in a same way as miRNA-OX plant establishment. I designed gRNA

sequences to six conserved *MIRNA* loci (*MIR160, 166a, 166b, 319a, 319b* and *529c*) identified (Fig. 7C), and inserted the sequences into pMpGE_En03 vector (Fig. 18B). Two gRNAs were designed at the position having PAM recognition motif (NGG) to each *MIRNA* locus (Fig. 19C). gRNAs were designed by the searching program, CasOT, that searches off-target binding regions of selected gRNAs in the genome sequences (Xiao *et al*. 2014).

If deletion of mature miRNA sequence in the *MIRNA* gene happens in transformed plants, miRNA-KO is successfully achieved (Fig. 18A). Meanwhile, if deletion or insertion of the other sequence in the hairpin fold-back structure occurs, defects in processing of pre-miRNA would happen. By transforming thalli of Tak-1 male plants, I obtained over 10 lines of T1 plants for each Cas9/gRNA construct. Then, the genome sequences of T1 plants were determined around at target *MIRNA* loci and checked whether or not mutations were introduced. Consequently, at least one plant that has mutation was obtained for each Cas9/gRNA construct (Table 4). Interestingly, various mutations were observed among them, one nucleotide insertion, long fragment insertion or deletion (Table 4).

Basically, most all CRISPR-miRNA lines showed severe morphological defects except for CRISPR-miR166b lines (Fig. 20A, B). CRISPR-miR160-1 line showed undulating thalli

(Fig. 20A). Nucleotide sequences of *MIR160a* locus in CRISPR-miR160-1#3, 5-10 lines had an insertion of "T" in the middle of mature sequence of miR160 (Table 4). Northern blot showed shifted band of miR160 with a decreased expression level in CRISPR-miR160-1 line#3, suggesting miR160 was processed in length of 22 nt and have a U insertion in the middle of mature sequence (between 9th and 10th nt) (Fig. 21A). qRT-PCR revealed up-regulated mRNA accumulation of Mp*ARF3* to a significant level (Fig. 22A). These data suggested the "T" insertion caused a defect in Mp*ARF3* repression and led to waved thallus.

CRISPR-miR166a lines show the severe defect in thallus development. Rod-like shaped thallus was shown in both of CRISPR-miR166a-1 and- 2 lines (Fig. 20A). Thallus was narrow and much smaller in size than WT. Nevertheless, gemma cups were observed in the front edges of thalli. Sequencing data and northern blot analysis suggested that miR166 expression was decreased and that knock-out (KO) was successful independently in both CRISPR-miR166a-1 and -2 lines (Fig. 21A). Mp*HD-ZIPIII* mRNA was significantly up-regulated in CRISPR-miR166a-1#21 line but less significantly increased in CRISPR-miR166a-2#3 line. Considering these data together, repression of Mp*HD-ZIPIII* gene expression by miR166 is quite important for *M. polymorpha* thallus development.

Observation of miR319b-OX plants suggested miR319 has roles in thallus and gemma cup development. Consistent with such earlier observation, both CRISPR-miR319a and 319b lines showed depletion of gemmae or gemma cups (Fig. 20A, B). CRISPR-miR319a-1#18 line had a deletion of 13 bp close to mature miRNA coding sequence and showed no gemma phenotype. CRISPR-miR319b-2#4 showed loss of gemma cups and had a deletion of 9 nt sequence upstream of mature miRNA. Although northern blot analysis could not detect mature miR319 expressed from two loci due to gene duplication, miR319 expression was probably decreased in CRISPR-miR319b-2#4, compared to WT (Fig. 21A). Unexpectedly, mRNA level of Mp*MYB33-like* gene did not accumulate in all CRISPR-miR319 lines (Fig. 22C). This result is consistent with that of qRT-PCR analysis of miR319-OX plant (Fig. 17B). These data suggested that both Mp*MIR319a* and *319b* are important but do not affect mRNA level of Mp*MYB33-like* gene in *M. polymorpha*.

By high-throughput sequencing and RACE analysis, it was suggested that miR529c represses Mp*SPL2* gene expression in place of miR156 in other land plants. CRISPR/Cas9 system could induce mutation in Mp*MIR529c* locus and, unexpectedly, CRISPR-miR529c lines showed abnormal thallus differentiation into antheridial head-like form (Fig. 20A).

miR529 also has three loci named Mp*MIR529a*, *529b* and *529c*. But, northern blot analysis revealed that miRNA expression was unchanged (Fig. 21A). qRT-PCR detected significant increase of Mp*SPL2* mRNA accumulation level in CRISPR-miR529c-1#4, which showed abnormally differentiated phenotype, whereas CRISPR-miR529c-2#8 showed no changes in mRNA level with no visible phenotypes (Fig. 20A, 22D). These data suggested that miR529c represses differentiation of thallus into sexual organ through suppression of target Mp*SPL2* gene expression.

Discussion

In this chapter, I tried to establish overexpression lines and CRISPR-mediated defective lines for conserved miRNAs in *M. polymorpha* to reveal the function of conserved miRNAs. Until now, no genetic and physiological analyses of *M. polymorpha* miRNAs have been performed. This research is the first one to study miRNA function in the liverwort, *M. polymorpha*.

By overexpression and reduced mutation in *MIRNA* gene loci, various morphological changes were induced in *M. polymorpha* thalii and gemma cups. As to thallus development, it was suggested that miR160 and miR166 are involved in proper thallus propagation. Wild-type thallus is flat and branches at proper frequency. However, miR160-OX induced curling, meaning perturbation of proper cell proliferation. Also, CRISPR-miR160 reduced lines showed curled edges of thalli. miR160 probably modulates proper cell patterns through repression of Mp*ARF3* expression at posttranscriptional level. CRISPR-miR166a reduced lines showed the most severe phenotype forming small rod-like structure (Fig. 20A, B). This result suggested that miR166 is quite important for thalli propagation of *M. polymorpha*. Small RNA sequence data also suggested the importance of miR166 because miR166 was the most highly expressed miRNA of all (Fig. 10A). These findings indicated that miR160 is

important for proper thallus propagation flatly as well as miR166 has quite essential roles in thallus development (Fig. 23A).

Gemma is an clonal propagule of *M. polymorpha* and is an important organ for propagation. CRISPR-miR319b reduced lines showed no gemma and gemma cup phenotype. In *M. polymorpha* development, every cell originates from one apical cell in the front edge of thallus. Our data suggested that miR319 affects triggering of gemma cup differentiation (Fig. 23A). Meanwhile, expression of Mp*MYB33-like*, a tentative target gene of miR319, was not affected in both overexpression and reduced lines (Fig. 17B, 22C). RACE analysis showed cleavage of the mRNA at target site of miR319 (Fig. 9B). It is generally accepted that, in plants, miRNA usually induced mRNA cleavage but rarely translational repression (Jones-Rhodes and Bartel 2004). Between miR319 and the target site in Mp*MYB33-like* mRNA, mismatches are present at 6th and 7th nt of miR319 (Fig. 9B). Biochemical analysis indicated that mismatches in target site induced translational repression using BY-2 tobacco culture cell lysate (Iwakawa and Tomari 2013). Considering such reports, miR319 possibly down-regulates Mp*MYB33-like* expression through translational repression.

It was revealed that miR529c targets Mp*SPL2* gene in this study. In *A. thaliana*, *SPL*

family genes, homologs of Mp*SPL2*, are regulated by miR156 (Fig. 12B). It is known that miR156 suppresses growth phase transition in *A. thaliana* (Aukerman and Sakai 2003). CRISPR-miR529c lines ectopically differentiated thallus into male sex organ antheridial head-like form (Fig. 20A). Normally, sex organs in *M. polymorpha* are induced only in the presence of far-red light and long day condition (Kubota *et al*. 2014). Phenotype of CRISPR-miR529c suggested that miR529c suppresses differentiation of vegetative thallus into a sexual organ and suppression of *SPL* genes is important for growth phase transition in both *A. thaliana* and *M. polymorpha* (Fig. 23A, B). I constructed both miRNA-OX and CRISPR-miRNA plants to know the function of four conserved miRNAs. However, I coud not identify the relationships between gain-of-function and loss-of-function phenotypes. Currently, only a few studies about *M. polymorph*a development are present. To describe these phenotypes properly, further studies about positional and temporal patterns of the expression of miRNAs and target genes are needed.

CRISPR/Cas9 technique is a developing tool and many modifications are added. In this study, I transformed WT plants by CRISPR-miRNA vectors that target *MIRNA* genes. Since there are no studies in plants that create miRNA mutants by genome editing tools previously,

this study offered the first example of genome editing to establish miRNA mutants. In the case of protein coding genes, nonsense mutation will happen by various types of mutation leading to knock-out of the gene. On the other hand, miRNA is a functional non-coding RNA but its sequence is very short; candidate sequences of gRNA are limited and it is hard to design them. In this study, CRISPR-miR160-1 transgene did work well not to induce deletion of sequences but to induce insertion of heterogenous sequences. Nucleotide insertions in precursor sequence like CRISPR-miR166a-1#21 also caused miRNA-KO (Fig. 21A, Table 4). These lines did not have miRNA deletion but had defects in miRNA function. Although there are few candidate PAM sequences in miRNA mature sequences, mutations in precursor sequences possibly can induce functional defects.

In this chapter, I revealed that conserved miRNAs have crucial roles in *M. polymorpha* development. Especially, morphogenesis of thalli was quite affected by miRNAs, miR160, miR166 and miR319. In addition, it was suggested that thallus differentiation into reproductive phase is suppressed by miR529c in *M. polymorpha* as well as miR156 in *A. thaliana*. These results revealed that conserved miRNAs are essential for development even in *M. polymorpha*.

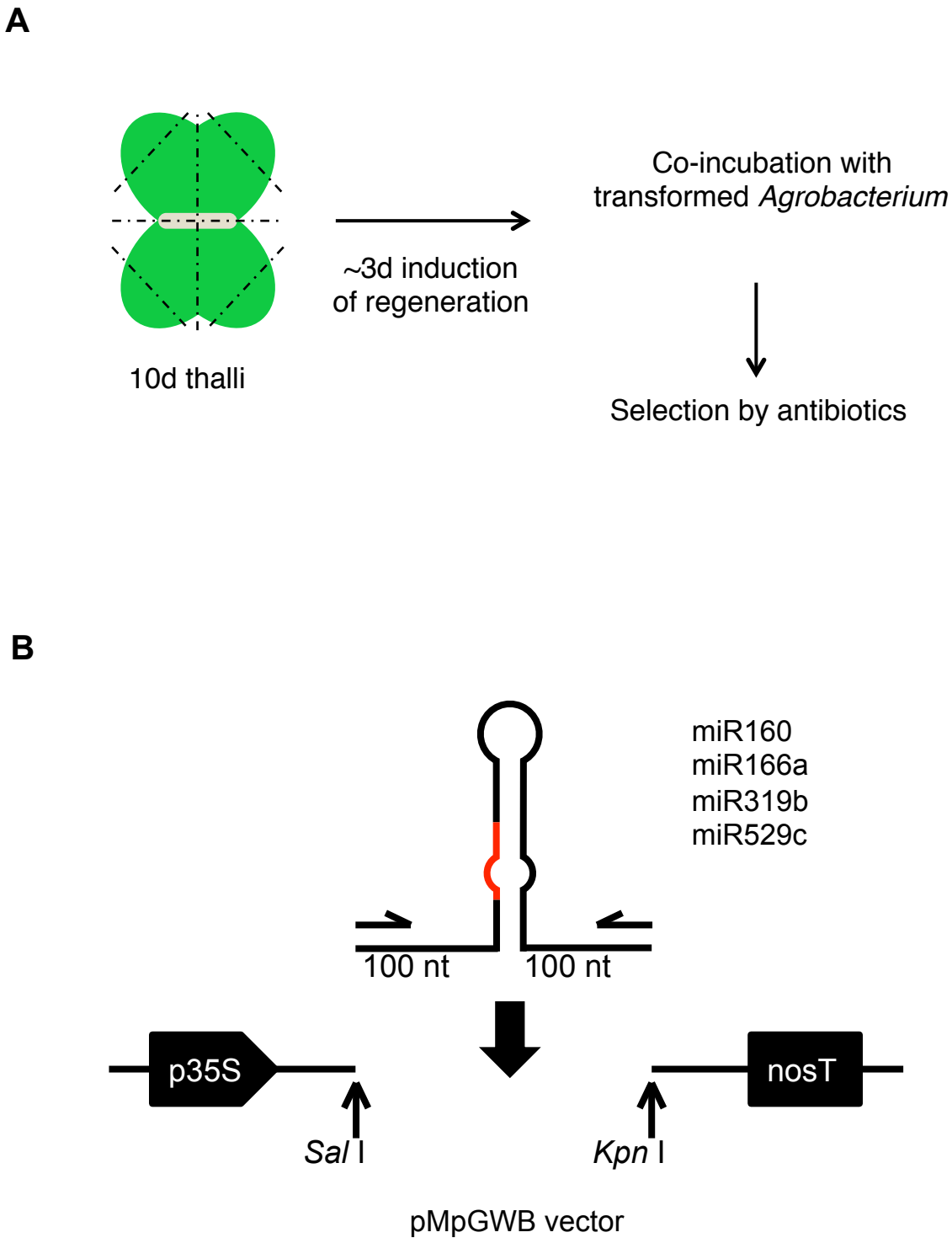


Fig. 14 *Agrobacterium*-mediated transformation and construction of miRNA-OX plants.
(A) A procedure of regenerated thalli transformation with *Agrobacterium*. (B) Schematic procedure of vector construction for miRNA overexpession. p35S: promoter sequence of CaMV35S, nosT: terminator sequence of nopaline syntase

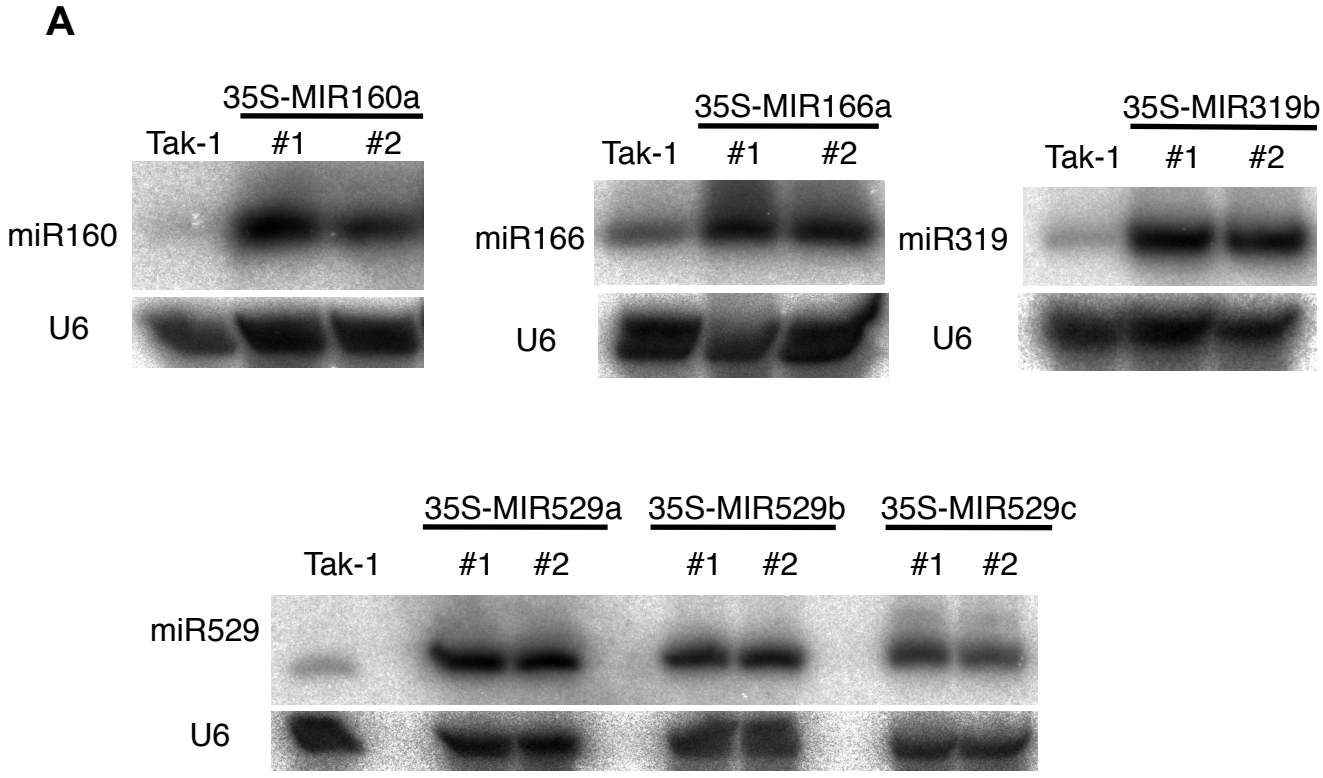


Fig. 15 miRNA-OX plants showed successful overexpression of each miRNA
(A) Detection of miRNA overexpression by northern blot analysis. #1 and #2 represent distinct transformed lines. U6 snRNA was used as a loading control.

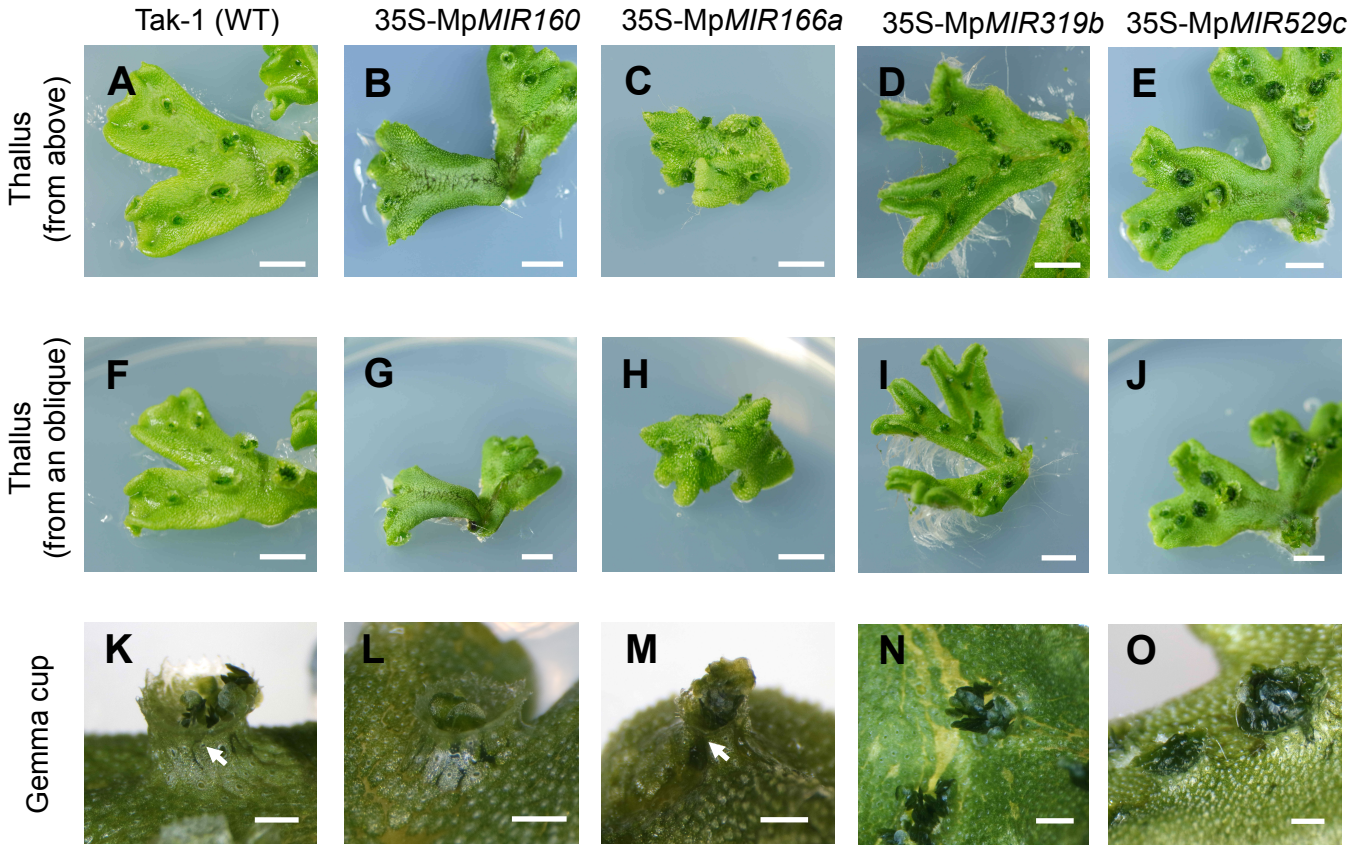
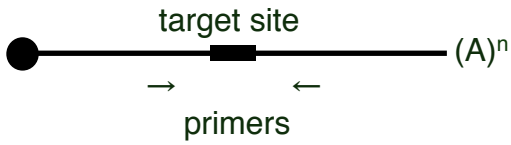


Fig. 16 miRNA-OX plants showed morphological phenotypes in their thalii and gemma cups

(A) Bright images of thallis and gemma cups of Tak-1 (A, F and K), 35S-MpMIR160 (B, G and L), 30S-MIR166a (C, H and M), 35S-MpMIR319b (D, I and N) and 35S-MpMIR529c (E, J and O) lines. (A-E) Bright images of mature thalli from above. (F-J) Bright images of mature thalli from an oblique. (K-O) Bright images of gemma cups. White scale bars show 5 mm (A-J) and 1 mm (K-O).

A



B

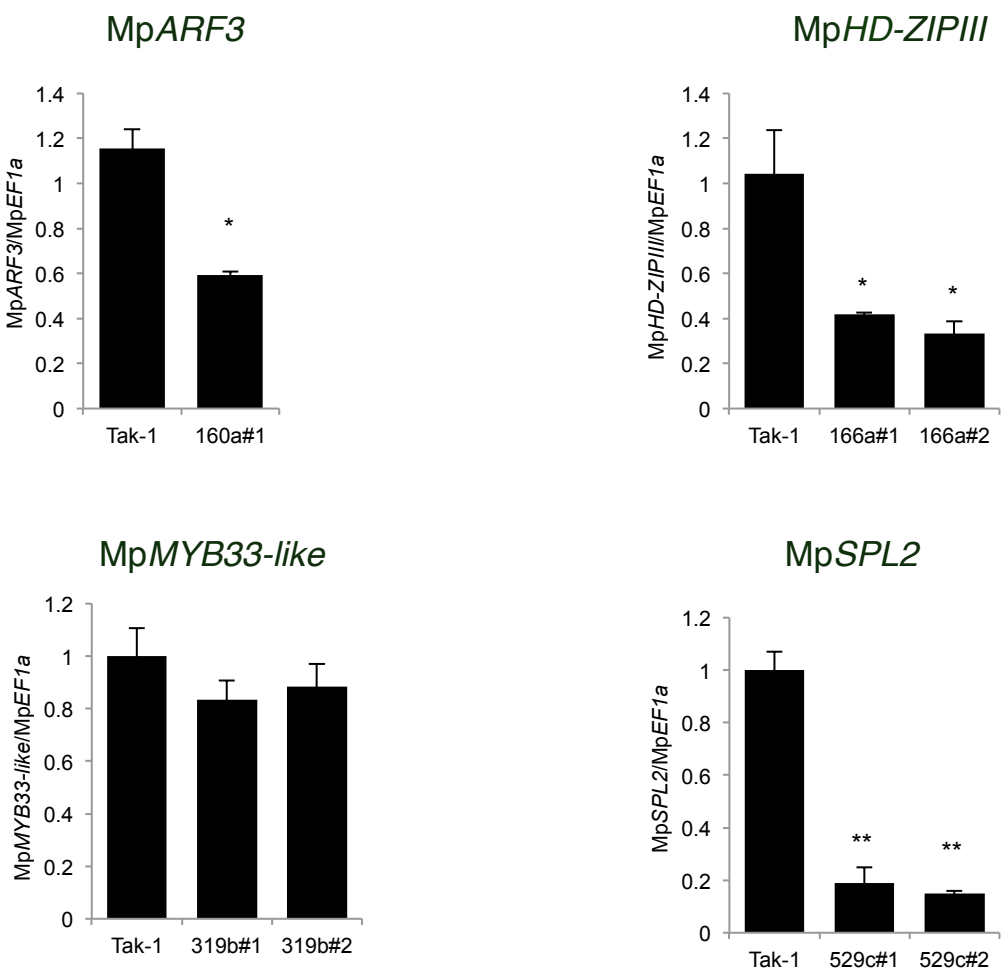


Fig. 17 mRNA accumulation levels of target genes were down-regulated

(A) Schematic design represents location of primers for qRT-PCR detecting miRNA cleavage of target mRNAs. (B) Detection of target mRNA accumulation level in miRNA-OX plants by qRT-PCR. Data represent relative expression level where those in Tak-1 were converted into 1. Data are relative expression level normalized by that of Mp*EF1a*. Error bars show standard error (SE). (*) $P < 0.05$, (**) $P < 0.01$. #1 and 2 mean independent lines.

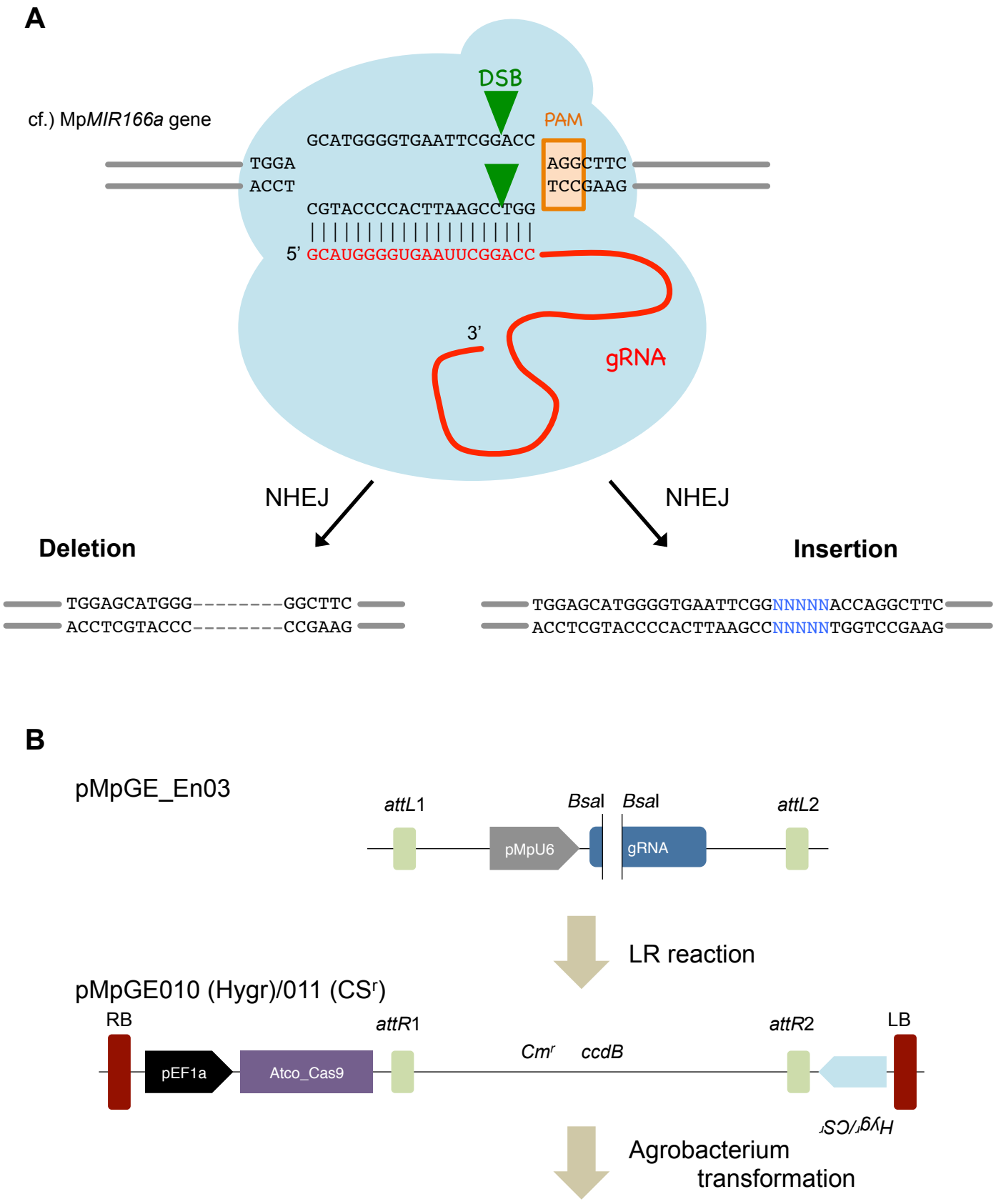


Fig. 18 Design of CRISPR/Cas9 construct to establish miRNA reduced mutants.
(A) Schematic representation of CRISPR/Cas9 induced mutation system. PAM: protospacer adjacent motif. DSB: double stranded break. NHEJ: non-homologous end joining. (B) Schematic design of CRISPR/Cas9 vector used in this study. *Bsa*I is a type II restriction enzyme that cuts not recognition site but near sequence.

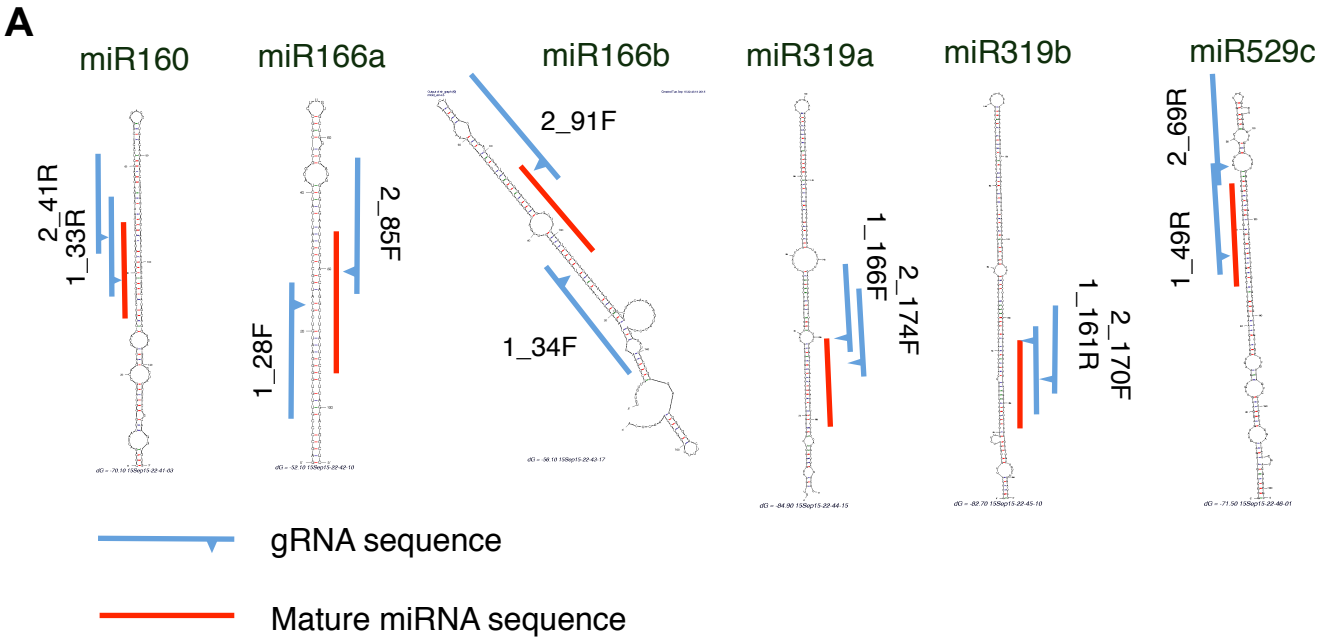
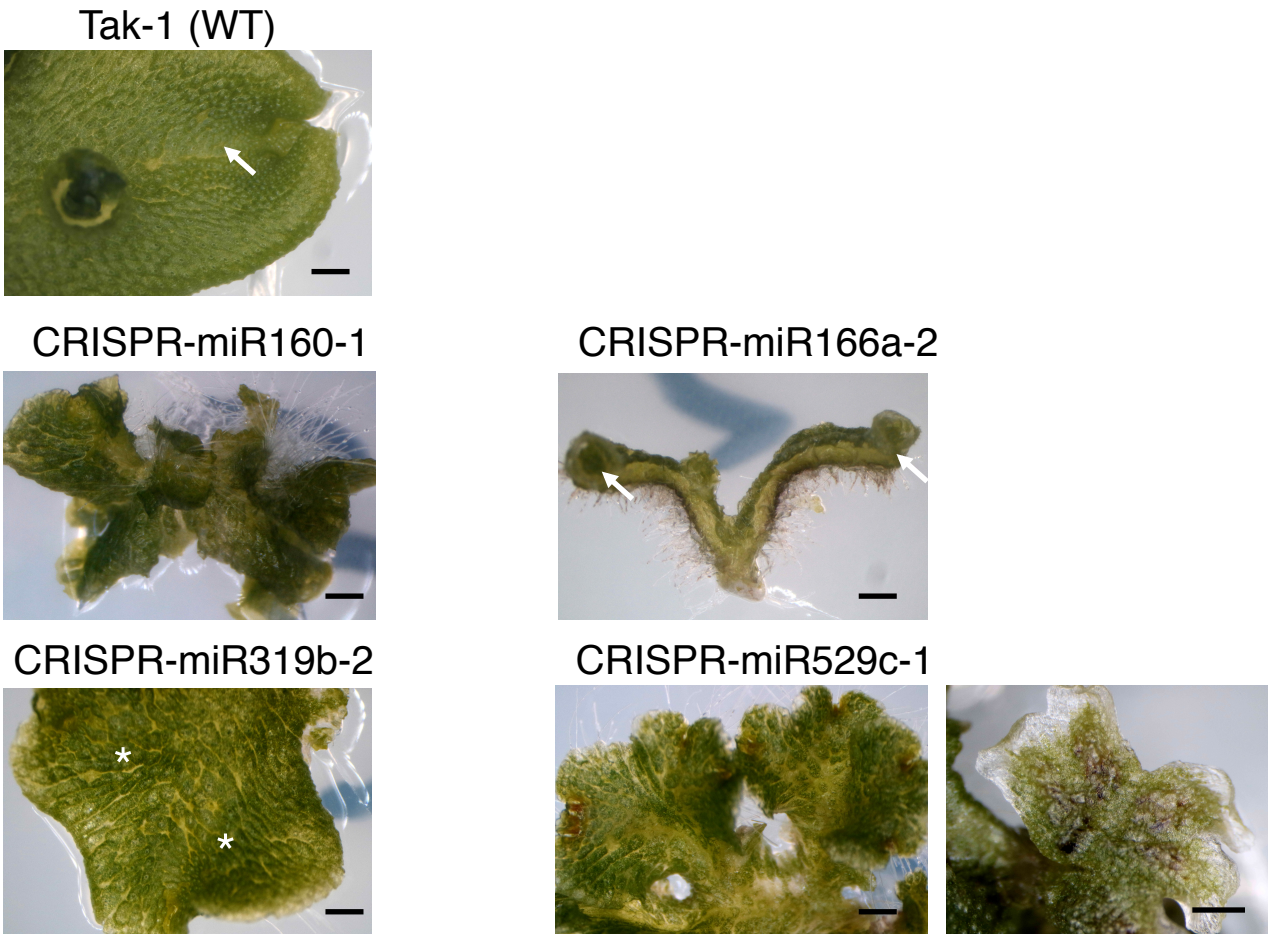


Fig. 19 Design of CRISPR/Cas9 construct to establish miRNA deletion mutants.
(A) Schematic pictures represent location of respective guide RNA (gRNA) sequences. Blue lines show gRNAs and spines are located at dsDNA cleavage sites by Cas9 protein. Red lines show mature miRNA sequences in each precursor.

A



B

-CRISPR-miR160

Line	PAM	Normal	Undulating thalli
1	33R	2	8
2	41R	9	1

-CRISPR-miR319a

#	PAM	Normal	No gemmae
1	166F	9	1
2	174F	8	2

-CRISPR-miR166a

#	PAM	Normal	Rod-like
1	28F	8	2
2	85F	1	9

-CRISPR-miR319b

#	PAM	Normal	No gemmae/cups
1	161R	9	1
2	170F	9	1

-CRISPR-miR166b

#	PAM	Normal
1	34F	20
2	91F	20

-CRISPR-miR529c

#	PAM	Normal	Antheridiophore-like
1	49R	6	4
2	69F	9	1

Fig. 20 CRISPR-miRNA transgenes induced severe morphological phenotypes in *M. polymorpha*

(A) Images of thalli of T1 plants of CRISPR/Cas9-miRNA transformed lines. White arrows show gemma cups. Astarisks show the position where gemma cups are normally locating. Black scale bars represent 1 mm. (B) Summary of phenotypes of CRISPR-miRNA transformed lines. Total number of each line is 10.

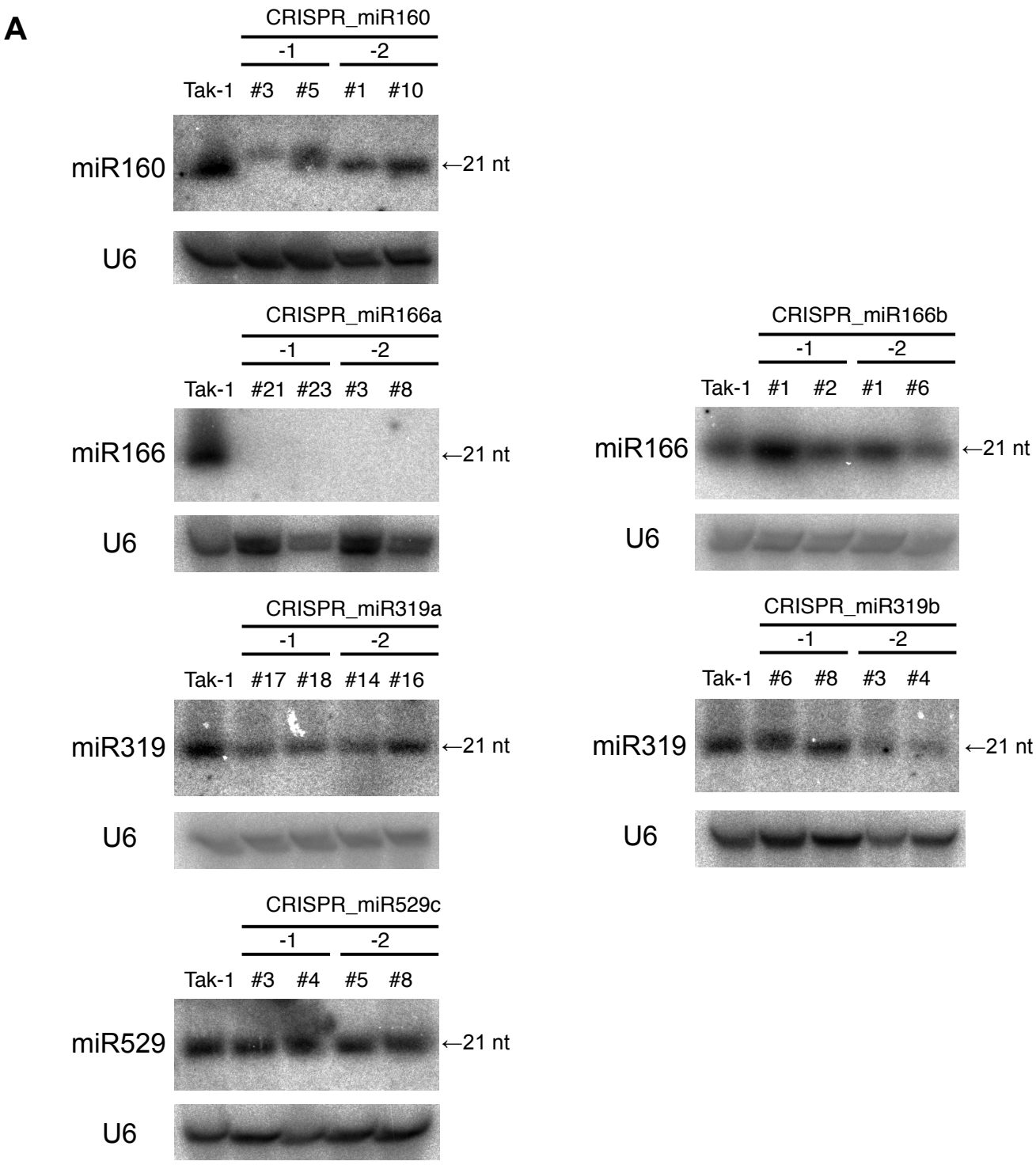


Fig. 21 The level of miRNAs in CRISPR-miRNA lines

(A) Northern blot analysis detecting miRNAs that was targeted by CRISPR-miRNA in each line. Two independent T1 lines were used at each CRISPR-miRNA transgene. miR166, miR319 and miR529 are located in the multiple loci and cross-hybridized by the same probe. U6 snRNA was used as a loading control.

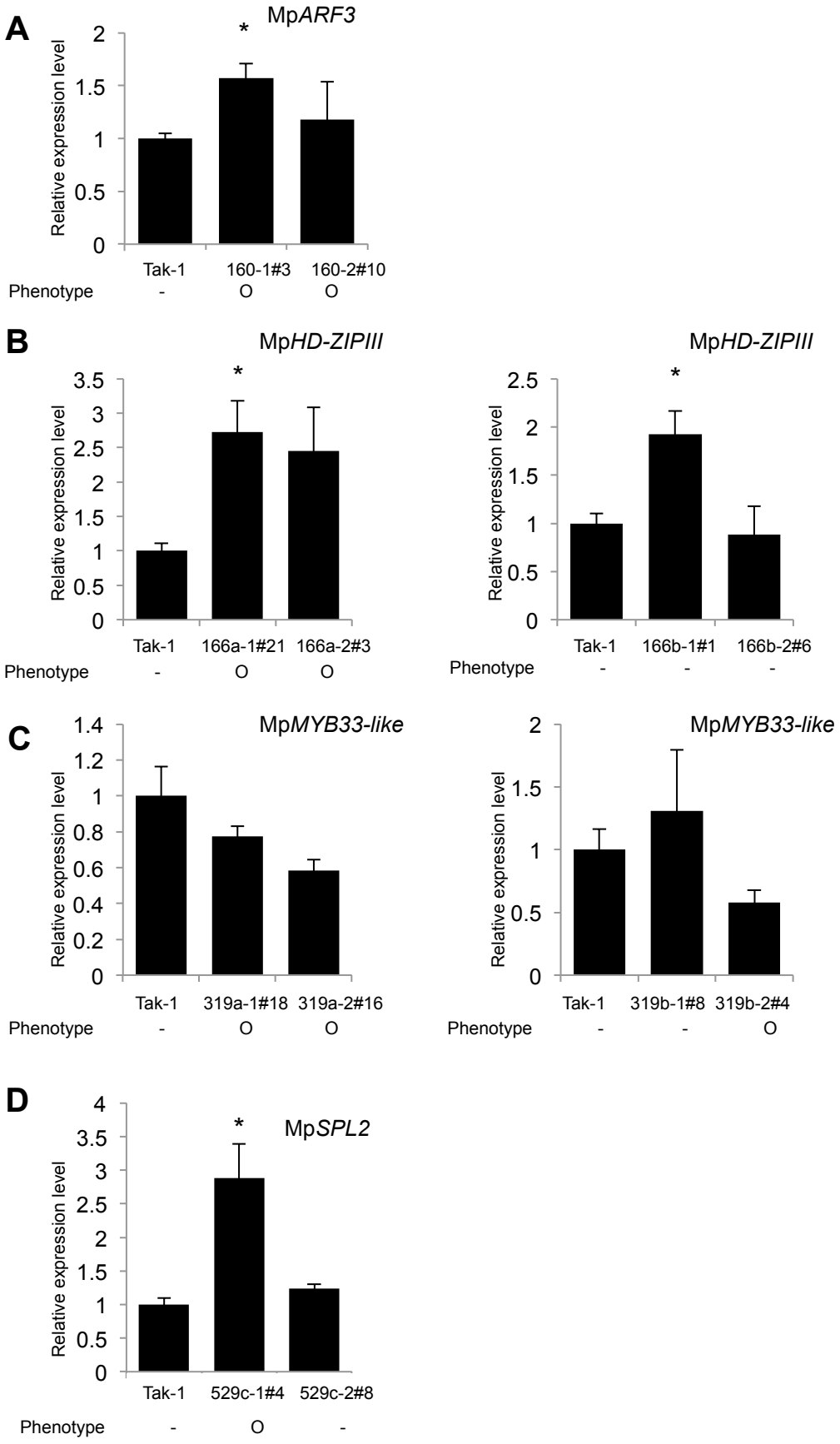


Fig. 22 Detection of mRNA accumulation levels in CRISPR-miRNA lines

(A-D) qRT-PCR analysis to detect mRNA accumulation levels of target genes. Data represent relative expression levels that those in Tak-1 were converted into 1. Data are relative expression levels normalized by those of Mp*EF1a*. Error bars show standard error (SE). (*) $P < 0.05$, Student's t-test. (A) Mp*ARF3* in CRISPR-miR160 lines. (B) Mp*HD-ZIPIII* in CRISPR-miR166 lines. (C) Mp*MYB33-like* in CRISPR-miR160 lines. (D) Mp*SPL2* in CRISPR-miR529c lines.

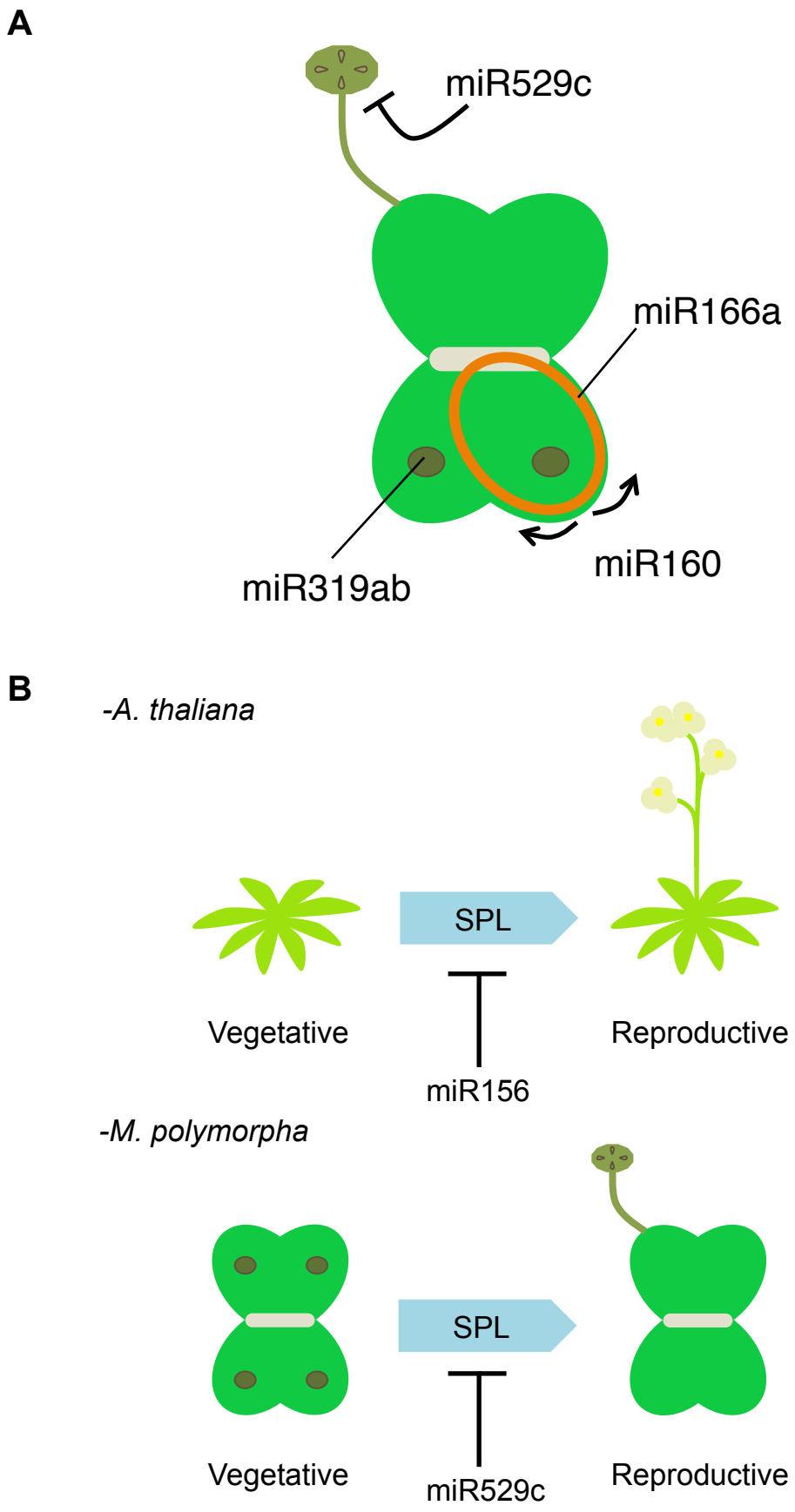


Fig. 23 Developmental roles of conserved miRNAs in *M. polymorpha*

(A) Developmental roles of each conserved miRNA suggested in this study.

(B) Similar function of miR156 and miR529c repressing *SPL* genes on suppression of the growth phase transition in *A. thaliana* and *M. polymorpha*.

Genotype	#		Phenotype
160-1	1	-	Undulated thalli
	2	Chimeric	Normal
	3	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Undulated thalli
	4	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Normal
	5	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Undulated thalli
	6	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Undulated thalli
	7	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Undulated thalli
	8	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Undulated thalli
	9	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Undulated thalli
	10	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Undulated thalli
160-2	1	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Normal
	2	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Normal
	3	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Normal
	4	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Normal
	5	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Normal
	6	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Normal
	7	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Normal
	8	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Normal
	9	CGCGACTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTCCCTGTATGCCAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGAAGTCGCTACCTCCCGCAAGGTGCG	Normal
	10	Chimeric	Normal
166a-1	1	GAGCCCCGTTGTGGGGAATGTCTGCATGGCCCGAATCCATCTACGGCTCGGCCTTGTGCTGGAGCATGGGGTGAATTCGGACCAGGCTTCATTCCCCTCAGCACGGCTC	Normal
	2	GAGCCCCGTTGTGGGGAATGTCGCATGGCCCGAATCCATCTACGGCTCGGCCTTGTGCTGGAGCATGGGGTGAATTCGGACCAGGCTTCATTCCCCTCAGCACGGCTC	Normal
	3	GAGCCCCGTTGTGGGGAATGTCCGCATGGCCCGAATCCATCTACGGCTCGGCCTTGTGCTGGAGCATGGGGTGAATTCGGACCAGGCTTCATTCCCCTCAGCACGGCTC	Normal
	4	GAGCCCCGTTGTGGGGAATGTCGCATGGCCCGAATCCATCTACGGCTCGGCCTTGTGCTGGAGCATGGGGTGAATTCGGACCAGGCTTCATTCCCCTCAGCACGGCTC	Normal
	5	GAGCCCCGTTGTGGGGAATGTCGGCATGGCCCGAATCCATCTACGGCTCGGCCTTGTGCTGGAGCATGGGGTGAATTCGGACCAGGCTTCATTCCCCTCAGCACGGCTC	Normal
	6	GAGCCCCGTTGTGGGGAATGTCTGCATGGCCCGAATCCATCTACGGCTCGGCCTTGTGCTGGAGCATGGGGTGAATTCGGACCAGGCTTCATTCCCCTCAGCACGGCTC	Normal
	7	GAGCCCCGTTGTGGGGAATGTCTGCATGGCCCGAATCCATCTACGGCTCGGCCTTGTGCTGGAGCATGGGGTGAATTCGGACCAGGCTTCATTCCCCTCAGCACGGCTC	Normal
	8	GAGCCCCGTTGTGGGGAATGTCGGGACATTCCCCACAACGGGGCTTTGCGAACCCATGGCAGGGTACGGCTCGGCCTTGTGCTGGAGCATGGGGTGAATTCGGACCAGGCTTCATTCCCCTCAGCACGGCTC	Rod-like
	9	Chimeric	Normal
	10	GAGCCCCGTTGTGGCATGGCCCGAATCCATCTACGGCTCGGCCTTGTGCTGGAGCATGGGGTGAATTCGGACCAGGCTTCATTCCCCTCAGCACGGCTC	Rod-like
166a-2	1	GAGCCCCGTTGTGGGGAATGTCGCATGGCCCGAATCCATCTACGGCTCGGCCTTGTGCT AAATGAAGCCAGGCTGCATTGGCATGCATCATTGGCA TTCGACCAGGCTTCATTCCCCTCAGCACGGCTC	Rod-like
	2	GAGCCCCGTTGTGGGGAATGTCGCATGGCCCGAATCCATCTACGGCTCGGCCTTGTGCTGGAGCATGGGGTGAATTC T ACCAGGCTTCATTCCCCTCAGCACGGCTC	Rod-like
	3	GAGCCCCGTTGTGGGGAATGTCGCATGGCCCGAATCCATCTACGGCTCGGCCTTGTGCTGGAGCATGGGGTGAA CCGATGTC ACCAGGCTTCATTCCCCTCAGCACGGCTC	Rod-like
	4	-	Rod-like
	5	GAGCCCCGTTGTGGGGAATGTCGCATGGCCCGAATCCATCTACGGCTCGGCCTTGTGCTGGAGCATGGGGTGAATTCGGACCAGGCTTCATTCCCCTCAGCACGGCTC	Normal
	6	-	Rod-like
	7	-	Rod-like
	8	GAGCCCCGTTGTGGGGAATGTCGCATGGCCCGAATCCATCTACGGCTCGGCCTTGTGCTGGAGCATGGGGTGAA ACCAGGCTTCATTCCCCTCAGCACGGCTC	Rod-like
	9	-	Rod-like
	10	GAGCCCCGTTGTGGGGAATGTCGCATGGCCCGAATCCATCTACGGCTCGGCCTTGTGCTGGAGCATGGGGTGAATTCGG CCAA ACCAGGCTTCATTCCCCTCAGCACGGCTC	Rod-like
166b-1	1	TGTAGTTGCGGAATTCTAGGGTCCAGTTGTGGGGAATGTTGACTGGTTCGGTAGCGACTGAACTTCACCTTGAACCAGATTGATATCGGACCAGGCTTCATTCCCCTCAGCATGACTCTCTTCAGAAGTACCCAGATATTTCCGTGATGGAGCACTTTTTGGTCCAAACTCA	Normal
	2	TGTAGTTGCGGAATTCTAGGGTCCAGTTGTGGGGAATGTTGACTGGTTCGGTAGCGACTGAACTTCACCTTGAACCAGATTGATATCGGACCAGGCTTCATTCCCCTCAGCATGACTCTCTTCAGAAGTACCCAGATATTTCCGTGATGGAGCACTTTTTGGTCCAAACTCA	Normal
	3	TGTAGTTGCGGAATTCTAGGGTCCAGTTGTGGGGAATGTTGACTGGTTCGGTAGCGACTGAACTTCACCTTGAACCAGATTGATATCGGACCAGGCTTCATTCCCCTCAGCATGACTCTCTTCAGAAGTACCCAGATATTTCCGTGATGGAGCACTTTTTGGTCCAAACTCA	Normal
	4	TGTAGTTGCGGAATTCTAGGGTCCAGTTGTGGGGAATGTTGACTGGTTCGGTAGCGACTGAACTTCACCTTGAACCAGATTGATATCGGACCAGGCTTCATTCCCCTCAGCATGACTCTCTTCAGAAGTACCCAGATATTTCCGTGATGGAGCACTTTTTGGTCCAAACTCA	Normal
	5	TGTAGTTGCGGAATTCTAGGGTCCAGTTGTGGGGAATGTTGACTGGTTCGGTAGCGACTGAACTTCACCTTGAACCAGATTGATATCGGACCAGGCTTCATTCCCCTCAGCATGACTCTCTTCAGAAGTACCCAGATATTTCCGTGATGGAGCACTTTTTGGTCCAAACTCA	Normal
	6	TGTAGTTGCGGAATTCTAGGGTCCAGTTGTGGGGAATGTTGACTGGTTCGGTAGCGACTGAACTTCACCTTGAACCA GGCTTCATTCCCCTCAGCATGACTCTCTTCAGAAGTACCCAGATATTTCCGTGATGGAGCACTTTTTGGTCCAAACTCA	Normal
	7	TGTAGTTGCGGAATTCTAGGGTCCAGTTGTGGGGAATGTTGACTGGTTCGGTAGCGACTGAACTTCACCTTGAACCAGATTGATATCGGAACCAGGCTTCATTCCCCTCAGCATGACTCTCTTCAGAAGTACCCAGATATTTCCGTGATGGAGCACTTTTTGGTCCAAACTCA	Normal
	8	TGTAGTTGCGGAATTCTAGGGTCCAGTTGTGGGGAATGTTGACTGGTTCGGTAGCGACTGAACTTCACCTTGAACCAGATTGATATCGG CTTCATTCCCCTCAGCATGACTCTCTTCAGAAGTACCCAGATATTTCCGTGATGGAGCACTTTTTGGTCCAAACTCA	Normal
	9	TGTAGTTGCGGAATTCTAGGGTCCAGTTGTGGGGAATGTTGACTGGTTCGGTAGCGACTGAACTTCACCTTGAACCAGATTGATATCGGTTTCATACCAGGCTTCATTCCCCTCAGCATGACTCTCTTCAGAAGTACCCAGATATTTCCGTGATGGAGCACTTTTTGGTCCAAACTCA	Normal
	10	TGTAGTTGCGGAATTCTAGGGTCCAGTTGTGGGGAATGTTGACTGGTTCGGTAGCGACTGAACTTCACCTTGAACCAGATTGATATCGGAACCAGGCTTCATTCCCCTCAGCATGACTCTCTTCAGAAGTACCCAGATATTTCCGTGATGGAGCACTTTTTGGTCCAAACTCA	Normal

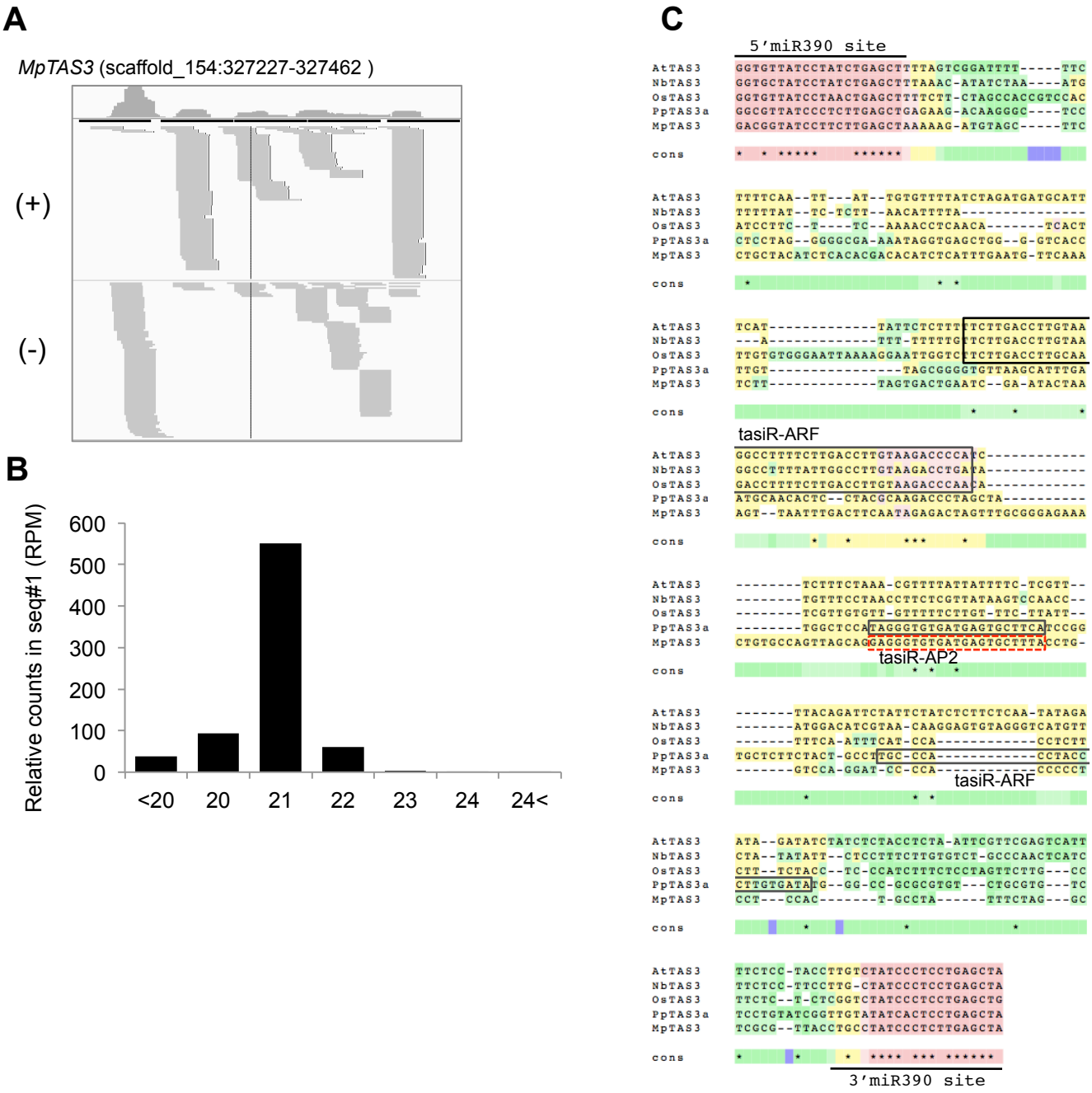
Table 4 Various mutations in miRNA precursors by CRISPR/Cas9 genome editing.

List of sequence results at CRISPR-target miRNA precursor loci. Red characters represent mature miRNA sequences. Blue characters show insertion of foreign sequences. Blank show deletion of native sequences. (-) Unsequenced clone.

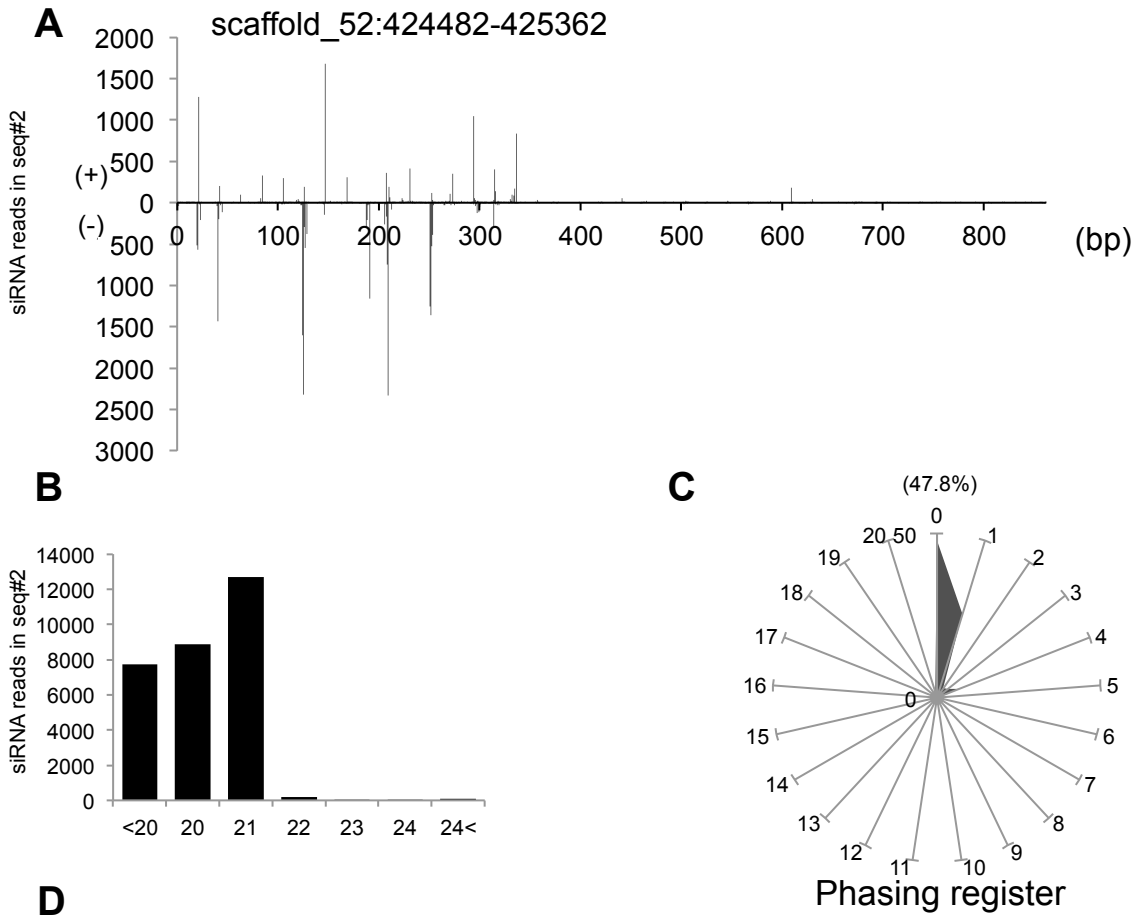
Genotype	#		Phenotype
319a-1	11	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
	12	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
	13	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
	14	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
	15	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
	16	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCTGTCCAAGGACGGATTGTGCTTTCGATCTGTACAATCCTGTCCTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	No gemmae
	17	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
	18	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
	19	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
	20	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
319a-2	11	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
	12	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
	13	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
	14	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGAC GGGATCGAAAGATCCCGTCCTGAATCCTTCCCTTCGATCGATCAGAATCCTTCCCTTCGCCTTCGTCAATCCGTCC TGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	No gemmae
	15	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
	16	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCC TGTCCAAGGACGGATTGTGCTTTCGATCTGTACAATCCTGTCCTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	No gemmae
	17	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
	18	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
	19	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
	20	GTAGAAAGAAGTTCGGATGGAACTCCCTTCAGTTCAACAGTGGGCTGTTTGGGTCCATATCTGCTGCTGAATCATTCATTCGCCCGTTCAATCCTACGAGAAAAGGATTGATTGGGCGTGTGATTGAGGCAGGAGCTGCTTTCGATCTGTACAATCCGTCCTTGGACTGAAGGGAGCTCCCTCTCAGGCTTCGTCTC	Normal
319b-1	1	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCCGACTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
	2	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCCGACTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
	3	-	Normal
	4	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCCGACTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
	5	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCCGACTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
	6	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCCGACTTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	No gemma cups
	7	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCCGACTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
	8	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCCGACT GGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
	9	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCCGACTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
	10	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCCGACTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
319b-2	1	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCC ACTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
	2	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCCGACTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
	3	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCC ACTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
	4	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTA CTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	No gemma cups
	5	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCC ACTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
	6	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCCGACTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
	7	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCCGACTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
	8	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCCGACTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
	9	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCCGACTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
	10	GGATCATTGCTGGCTCTTGATGGAGCTTTTTTTCAGTTCAACTCTGGCTGGACTGGAGCTGGAGCTGCTGCTGAATCATGCATTCGCCTACTCTGTCCTAGATTACAAGGGCAAAGTAGTTGCATGCCTGAAGCAGGAGTTGCACCTACCTCCTATTTGGCCCGACTTGGACTGAAGGGAGCTCCCTCGCCTGCTGTCTTGATCT	Normal
529c-1	1	GCGCATGTAGATGAAAGAGGAAGTAGATGAAGAGCACCAAAGCCGTTACCAGAA GAGCACAGCCTTTGCCGTCTCTCCCGCTTCAGCCTGGGGCTTGACAACTGGGGCTGTGCTATCCCTCTTCTGCTGCCGGATTTGTTGCATGTTAGTCCTTCCTCCGAATCTCGTACGTGCGT	Normal
	2	-	Normal
	3	GCGCATGTAGATGAAAGAGGAAGTAGATGAAGAGCACCAAAGCCGTTACCAGAAGAG -deletion	Antheridiophore-like
	4	GCGCATGTAGATGAAAGAGGAAGTAGATGAAGAGCACCAAAGCCGTTA GAGAGAGAGCACAGCCTTTGCCGTCTCTCCCGCTTCAGCCTGGGGCTTGACAACTGGGGCTGTGCTATCCCTCTTCTGCTGCCGGATTTGTTGCATGTTAGTCCTTCCTCCGAATCTCGTACGTGCGT	Antheridiophore-like
	5	GCGCATGTAGATGAAAGAGGAAGTAGATGAAGAGCACCAAAGCCGTTACCAGAA AGAGAGAGCACAGCCTTTGCCGTCTCTCCCGCTTCAGCCTGGGGCTTGACAACTGGGGCTGTGCTATCCCTCTTCTGCTGCCGGATTTGTTGCATGTTAGTCCTTCCTCCGAATCTCGTACGTGCGT	Antheridiophore-like
	6	GCGCATGTAGATGAAAGAGGAAGTAGATGAAGAGCACCAAAGCCGTTACCAGAAGAGAGAGAGCACAGCCTTTGCCGTCTCTCCCGCTTCAGCCTGGGGCTTGACAACTGGGGCTGTGCTATCCCTCTTCTGCTGCCGGATTTGTTGCATGTTAGTCCTTCCTCCGAATCTCGTACGTGCGT	Normal
	7	-	Normal
	8	GCGCATGTAGATGAAAGAGGAAGTAGATGAAGAGCACCAAAGCC CCAGGCTGTGCTCTCTC GAGAGAGAGCACAGCCTTTGCCGTCTCTCCCGCTTCAGCCTGGGGCTTGACAACTGGGGCTGTGCTATCCCTCTTCTGCTGCCGGATTTGTTGCATGTTAGTCCTTCCTCCGAATCTCGTACGTGCGT	Normal
	9	GCGCATGTAGATGAAAGAGGAAGTAGATGAAGAGCACCAAAGCCGTTACCAGAAGAGAGAGAGAGCACAGCCTTTGCCGTCTCTCCCGCTTCAGCCTGGGGCTTGACAACTGGGGCTGTGCTATCCCTCTTCTGCTGCCGGATTTGTTGCATGTTAGTCCTTCCTCCGAATCTCGTACGTGCGT	Antheridiophore-like
	10	GCGCATGTAGATGAAAGAGGAAGTAGATGAAGAGCACCAAAGCCGTTAC AGAGAGAGAGCACAGCCTTTGCCGTCTCTCCCGCTTCAGCCTGGGGCTTGACAACTGGGGCTGTGCTATCCCTCTTCTGCTGCCGGATTTGTTGCATGTTAGTCCTTCCTCCGAATCTCGTACGTGCGT	Normal
529c-2	1	-	Normal
	2	GCGCATGTAGATGAAAGAGGAAGTAGATGAAGAGCACCAAAGCCGTTACCAGAAGAGAGAGAGAGCACAGCCTTTGC CCTTTGCACA GTCTCTCCCGCTTCAGCCTGGGGCTTGACAACTGGGGCTGTGCTATCCCTCTTCTGCTGCCGGATTTGTTGCATGTTAGTCCTTCCTCCGAATCTCGTACGTGCGT	Normal
	3	-	Normal
	4	GCGCATGTAGATGAAAGAGGAAGTAGATGAAGAGCACCAAAGCCGTTACCAGAAGAGAGAGAGAGCACAGCCTTTGC CCTTTGCACA GTCTCTCCCGCTTCAGCCTGGGGCTTGACAACTGGGGCTGTGCTATCCCTCTTCTGCTGCCGGATTTGTTGCATGTTAGTCCTTCCTCCGAATCTCGTACGTGCGT	Normal
	5	GCGCATGTAGATGAAAGAGGAAGTAGATGAAGAGCACCAAAGCCGTTACCAGAAGAGAGAGAGAGCACAGCCTTTG -467bp deletion-	Antheridiophore-like
	6	-	Normal
	7	GCGCATGTAGATGAAAGAGGAAGTAGATGAAGAGCACCAAAGCCGTTACCAGAAGAGAGAGAGCACAGCC GTCTCTCCCGCTTCAGCCTGGGGCTTGACAACTGGGGCTGTGCTATCCCTCTTCTGCTGCCGGATTTGTTGCATGTTAGTCCTTCCTCCGAATCTCGTACGTGCGT	Normal
	8	GCGCATGTAGATGAAAGAGGAAGTAGATGAAGAGCACCAAAGCCGTTACCAGAAGAGAGAGAGAGCACAGCC GTCTCTCCCGCTTCAGCCTGGGGCTTGACAACTGGGGCTGTGCTATCCCTCTTCTGCTGCCGGATTTGTTGCATGTTAGTCCTTCCTCCGAATCTCGTACGTGCGT	Normal
	9	GCGCATGTAGATGAAAGAGGAAGTAGATGAAGAGCACCAAAGCCGTTACCAGAAGAGAGAGAGAGCACAGCCTTTGC CCTTTGCACA GTCTCTCCCGCTTCAGCCTGGGGCTTGACAACTGGGGCTGTGCTATCCCTCTTCTGCTGCCGGATTTGTTGCATGTTAGTCCTTCCTCCGAATCTCGTACGTGCGT	Normal
	10	GCGCATGTAGATGAAAGAGGAAGTAGATGAAGAGCACCAAAGCCGTTACCAGAAGAGAGAGAGAGCACAGCC ACAA GCCGTCTCTCCCGCTTCAGCCTGGGGCTTGACAACTGGGGCTGTGCTATCCCTCTTCTGCTGCCGGATTTGTTGCATGTTAGTCCTTCCTCCGAATCTCGTACGTGCGT	Normal

(Cont.) Table 4 Various mutations in miRNA precursors by CRISPR/Cas9 genome editing.

List of sequence results at CRISPR-target miRNA precursor loci. Red characters represent mature miRNA sequences. Blue characters show insertion of foreign sequences. Blank show deletion of native sequences. (-) Unsequenced clone.

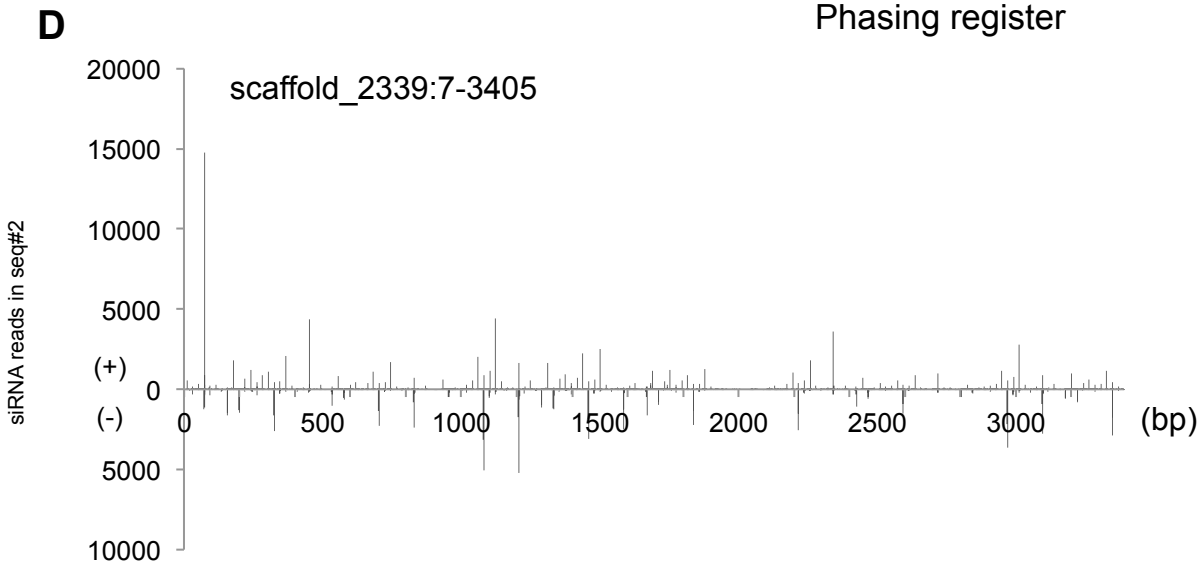


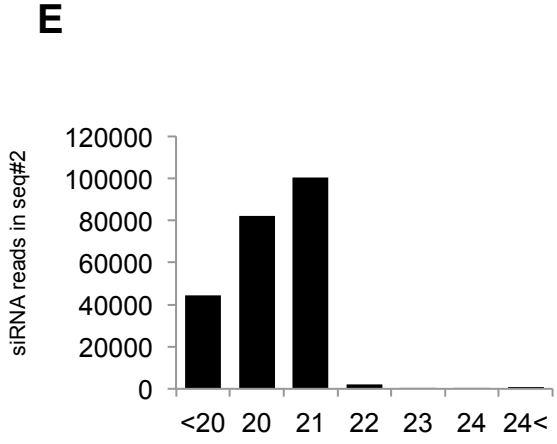
Supplementary figure 1 The TAS3 locus and tasiRNA in *M. polymorpha.*
(A) Genome browser view (Integrative Genomics Viewer; Robinson et al. 2011) of small RNA reads mapped to the MpTAS3 locus. The blue box represents tentative miR390-target sites. (B) Distribution of small RNA lengths mapped to the MpTAS3 locus in seq#1. Counts were converted to relative expression level (RPM). (C) Results of clustalW analysis of TAS3 nucleotide sequences in land plants (At: Arabidopsis thaliana, Nb: Nicotiana benthamiana, Os: Oryza sativa, Pp: Physcomitrella patens, Mp: Marchantia polymorpha). Black boxes indicate previously reported tasiRNA sequences (Axtell et al. 2006, 2007). A red broken box shows the previously reported tentative tasi-AP2 sequence (Krasnikova et al. 2013).

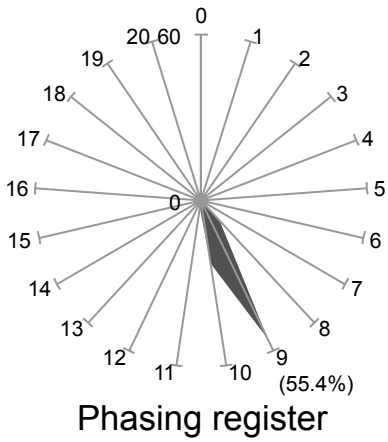

A
scaffold_52:424482-425362
siRNA reads in seq#2
(+)
(-)
2000
1500
1000
500
0
500
1000
1500
2000
2500
3000
0
100
200
300
400
500
600
700
800
(bp)

B
siRNA reads in seq#2
14000
12000
10000
8000
6000
4000
2000
0
<20
20
21
22
23
24
24<

C
(47.8%)
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
50
Phasing register

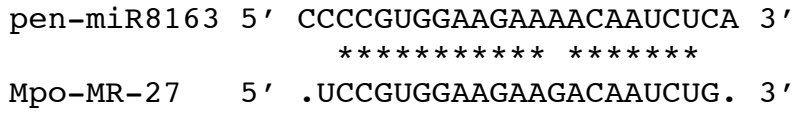

D
scaffold_2339:7-3405
siRNA reads in seq#2
(+)
(-)
20000
15000
10000
5000
0
5000
10000
0
500
1000
1500
2000
2500
3000
(bp)


E
siRNA reads in seq#2
120000
100000
80000
60000
40000
20000
0
<20
20
21
22
23
24
24<


F
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
60
(55.4%)
Phasing register

Supplementary figure 2　Description of the phasing siRNAs in seq#2.

(A–C) scaffold_52:424482-425362. (D–F) scaffold_2339:7-3405. (A, D) Genomic location of the 1st nucleotide of mapped siRNA sequences. Number of reads is shown. (B, E) Read count distribution of siRNA length mapped to each locus. (C, F) Rader plots showing the percentages of reads in each of the 21 nt registers (Axtell et al. 2006, Montgomery et al. 2008). The 5′ end of the siRNA sequences mapped to the negative strand of the genome was displaced to the +2 position of each location to reflect the two base overhang of DCL4 activity.



Supplementary Figure 3
Sequence homology between Mpo-MR-27 and pen-miR8163.

name	sequence	use
miR156probe	TGTGCTCACACTCTTCTGTCA	Northern blot
miR160probe	TGGCATACAGGGAGCCAGGCA	Northern blot
miR166probe	GGGGAATGAAGCCTGGTCCGA	Northern blot
miR171probe	ATGTGATATTGGAGCGGCTCA	Northern blot
miR319probe	GGAGCTCCCTTCAGTCCAAG	Northern blot
miR390probe	GACGCTATCCCTCCTGAGCTT	Northern blot
miR408probe	GCCAGGGAAGAGGCAGTGCAT	Northern blot
miR529probe	GGGCTGTGCTCTCTCTCTTCT	Northern blot
miR536probe	GATGCACACAGCTTGGCACGG	Northern blot
U6probe	AGGGGCCATGCTAATCTTCTC	Northern blot
GeneRacer-5'primer	CGACTGGAGCACGAGGACACTGA	RACE
GeneRacer-5'nestprimer	GGACACTGACATGGACTGAAGGAGTA	RACE
GeneRacer-3'primer	GCTGTCAACGATACGCTACGTAACG	RACE
GeneRacer-3'nestprimer	CGCTACGTAACGGCATGACAGTG	RACE
MpARF3-RACE-R1	GGGGTCGACACGGATCACCAGGCGCC	RACE
MpARF3-RACE-nestR1	ACGGATCACCAGGCGCCTTGCGGAC	RACE
MpHD-ZIP-RACE-R1	GCGACGTCAACCGGAGCGAACACGAGT	RACE
MpHD-ZIP-RACE-nestR1	GCGAACACGAGTTGCGCACATGCACCC	RACE
MpMYB33-RACE-R1	TCCGAGCTGGCACACACCCGGCATGTT	RACE
MpMYB33-RACE-nR2	CGCCACAGGATGTCCTGCCGCTCCCAA	RACE
MpSPL2-RACE-R1	CGCAAGGGTCAGACAGAACGACGG	RACE
MpSPL2-RACE-nestR1	CAGACCGACCCGTCGTTCTGTCTG	RACE
MpSPL1-RACE-R1	AGCATCACTGCATGGGGAGGCCGC	RACE
MpSPL1-RACE-nestR1	GGGAGGCCGCTCTTCATTTCGGCC	RACE
MpARF3-qRT-F	GCCCATCACACTACCTTCGGT	qRT-PCR
MpARF3-qRT-R	GGGAGTTTACCAGGCCGGA	qRT-PCR
MpHD-ZIP-qRT-F	GGCGATTGCAGAAGAAACCTTGA	qRT-PCR
MpHD-ZIP-qRT-R	ACAACCGTGTGAAATGGCTACA	qRT-PCR
MpMYB33-qRT-F	CTCTTGTTCCAGGCCGAAT	qRT-PCR
MpMYB33-qRT-R	AAAATGCCGTGTAAGGGCTA	qRT-PCR
MpSPL3-qRT-F1	GACTTCCGGTACCACGAGTGT	qRT-PCR
MpSPL3-qRT-R1	CCAGTGATACTGAGCCCGGG	qRT-PCR
MpEF1_qPCR_F	AAGCCGTCGAAAAGAAGGAG	qRT-PCR
MpEF1_qPCR_R	TTCAGGATCGTCCGTTATCC	qRT-PCR

Supplementary Table 1 Primers used for this study.

name	sequence	use
Hind-35SP-F	GTTAAGCTTAGATTAGCCTTTTCAATTTCAGAAAG	Vector construction
Pst-35SP-R	GTTCTGCAGCGTGTTCTCTCCAAATGAAATGA	Vector construction
Sal-MpMIR160-F	GTTGTCGACTCGGTTGGGGTCGCGAGC	Vector construction
Kpn-MpMIR160-R	GTTGGTACCGAAAAATTCGACACTAACGCACGA	Vector construction
Sal-MpMIR166a-F	GTTGTCGACGAATTCACTCATGAAAGAGTCTTTG	Vector construction
Kpn-MpMIR166a-R	GTTGGTACCGTAACTTGGAAGCAGCGCA	Vector construction
Sal-MIR319b-F	GTTGTCGACGTGAACTCAGCGCTCATG	Vector construction
Kpn-MIR319b-R	GTTGGTACCTACAATCTTTCAAATCAGTACGGAT	Vector construction
Sal-MIR529c-F	GTTGTCGACTTGCGCCCGCTAGACCAC	Vector construction
Kpn-MIR529c-R	GTTGGTACCCTCCATAAACTCATGTGATGTTCAAATAGC	Vector construction
PAM_MIR160-33r-F	CTCGCAGTTGGCATACAGGGAGCC	Vector construction
PAM_MIR160-33r-R	AAACGGCTCCCTGTATGCCAACTG	Vector construction
PAM_MIR160-41r-F	CTCGGAGCTCCTCAGTTGGCATAC	Vector construction
PAM_MIR160-41r-R	AAACGTATGCCAACTGAGGAGCTC	Vector construction
PAM_MIR166a-28f-F	CTCGCCGTTGTGGGGAATGTCGCA	Vector construction
PAM_MIR166a-28f-R	AAACTGCGACATTCCCCACAACGG	Vector construction
PAM_MIR166a-85f-F	CTCGGCATGGGGTGAATTCGGACC	Vector construction
PAM_MIR166a-85f-R	AAACGGTCCGAATTCACCCCATGC	Vector construction
PAM_MIR166b-34f-F	CTCGAATTCTAGGGTCCAGTTGTG	Vector construction
PAM_MIR166b-34f-R	AAACCACAACTGGACCCTAGAATT	Vector construction
PAM_MIR166b-95f-F	CTCGAACCAGATTGATATCGGACC	Vector construction
PAM_MIR166b-95f-R	AAACGGTCCGATATCAATCTGGTT	Vector construction
PAM_MIR319a-166f-F	CTCGCGATCTGTACAATCCGTCCT	Vector construction
PAM_MIR319a-166f-R	AAACAGGACGGATTGTACAGATCG	Vector construction
PAM_MIR319a-174f-F	CTCGACAATCCGTCCTTGGACTGA	Vector construction
PAM_MIR319a-174f-R	AAACTCAGTCCAAGGACGGATTGT	Vector construction
PAM_MIR319b-161r-F	CTCGAGCTCCCTTCAGTCCAAGTC	Vector construction
PAM_MIR319b-161r-R	AAACGACTTGGACTGAAGGGAGCT	Vector construction
PAM_MIR319b-170f-F	CTCGACCTCCTATTTGGCCCGACT	Vector construction
PAM_MIR319b-170f-R	AAACAGTCGGGCCAAATAGGAGGT	Vector construction
PAM_MIR529c-49r-F	CTCGAGGCTGTGCTCTCTCTCTTC	Vector construction
PAM_MIR529c-49r-R	AAACGAAGAGAGAGAGCACAGCCT	Vector construction
PAM_MIR529c-69r-F	CTCGTGAAGCGGGAGAGACGGCAA	Vector construction
PAM_MIR529c-69r-R	AAACTTGCCGTCTCTCCCGCTTCA	Vector construction

Supplementary Table 1 Primers used for this study (Cont.)

Concluding remarks

In Chapter I of this study, I profiled small RNAs in the liverwort, *M. polymorpha*, with the draft genome sequences for the first time. Small RNA sequencing revealed that *M. polymorpha* shared 9 miRNAs (miR160, 166, 171, 319, 390, 408, 529, 536 and 1030) with other land plants. Computational prediction and RACE analysis revealed that miRNA/target gene pairs of three miRNAs, miR160/Mp*ARF3*, miR166/MpHD-ZIPIII and miR319/Mp*MYB33-like*, were also conserved in *M. polymorpha*. Small RNA sequence analysis could not detect miR156 though its nucleotide sequence is present in the genome. Alternatively, RACE analysis suggested Mp*SPL2* gene, which is a homolog gene of *SPL* genes regulated by miR156 in *A. thaliana*, is targeted by miR529c. I also identified over 200 novel miRNAs expressed in *M. polymorpha*. RACE analysis showed Mpo-MR-13, one of the novel miRNAs, cleaved mRNA of Mp*SPL1* gene. These results suggested that expression levels of two *SPL* genes are controlled by two distinct miRNAs in *M. polymorpha*.

Next, in Chapter II, I performed molecular genetic analysis of conserved miRNAs by establishment of miRNA-OX and CRISPR/Cas9-targeted miRNA mutants. Most of these plants showed morphological abnormality and changed expression level of target candidate

genes. CRISPR-miRNA mutants of miR160 showed undulated thallus phenotypes. miR166 mutants by CRISPR/Cas9-mediated genome editing showed the severe phenotype with rod-like thalli. These result showed miR160 and miR166 are important for proper thallus propagation. Also, mutations in miR319 gene induced depletion of gemmae or gemma cups. CRISPR-miR529c mutants showed abnormally differentiated thalli into antheridiophore-like structure. This result suggested that miR529c suppresses growth phase transition from vegetative growth to reproductive growth in *M. polymorpha* as well as miR156 in *A. thaliana*.

This study revealed small RNAs in *M. polymorpha* have both specific and universal characters compared with other land plants. I hope new insights provided by this study will expand understanding of the evolutionary origin and expansion of negative gene expression mediated by miRNA and small noncoding RNAs.

References

Allen, E., Xie, Z., Gustafson, AM. and Carrington, J. C. (2005) MicroRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* 121: 207-221.

Allen, R. S., Li, J., Stahle, M. I., Dubroué, A., Gubler, F. and Millar, A. A. (2007) Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis miR159 family. *Proc. Nat. Aca. Sci.* 104: 16371-16376.

Aukerman, M. J. and Sakai, H. (2003) Regulation of flowering time and floral organ identity by a microRNA and its AP*ET AL*A2-like target genes. *Plant Cell* 15: 2730-2741.

Axtell, M. J., Jan, C., Rajagopalan, R. and Bartel, D. P. (2006).A two-hit trigger for siRNA biogenesis in plants. *Cell* 127: 565–77.

Axtell, M. J., Snyder, J. A. and Bartel, D. P. (2007) Common functions for diverse small RNAs of land plants. *Plant Cell* 19: 1750–1769.

Axtell, M. J. (2008) Evolution of microRNAs and their targets: are all microRNAs biologically relevant? *Biochim. Biophys. Acta* 1779: 725-734.

Axtell, M. J. (2013) ShortStack: Comprehensive annotation and quantification of small RNA

genes. *RNA* 19: 740-751.

Bartel, D. P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281-297.

Bernstein, E., Caudy, A. A., Hammond, S. M. and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409: 363–366.

Bibikova, M., Beumer, K., Trautman, J. K. and Carroll, D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science* 300: 764.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326: 1509–1512.

Bollman, K. M., Aukerman, M. J., Park, M. Y., Hunter, C., Berardini, T. Z. and Poethig, R. S. (2003) HASTY, the Arabidopsis ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. *Development* 130: 1493–1504.

Bolotin, A., Quinquis, B., Sorokin, A. and Ehrlich, S. D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151: 2551–2561.

Bouché, N., Lauressergues, D., Gasciolli, V. and Vaucheret, H. (2006) An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO J* 25: 3347–3356.

Bowman, J. L. (2013) Walkabout on the long branches of plant evolution. *Curr. Opin. in Plant Biol. 16*: 70–77.

Bowman, J. L. (2015) A Brief History of Marchantia from Greece to Genomics. *Plant and Cell Physiol. 0*: 1–20. doi:10.1093/pcp/pcv04

Buermans, H. P. J. and den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochim. Biophys. Acta*. 1842: 1932-1940.

Cao, X. and Jacobsen, S. E. (2002). Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Current Biology 12*: 1138–44.

Chen, X. (2004). A microRNA as a translational repressor of AP*ET AL*A2 in Arabidopsis

flower development. *Science* 303: 2022–2025.

Chiyoda, S., Ishizaki, K., Kataoka, H., Yamato, K. T. and Kohchi, T. (2008) Direct transformation of the liverwort Marchantia polymorpha L. by particle bombardment using immature thalli developing from spores. *Plant Cell Rep.* 27: 1467–1473.

Cho, S. H., Coruh, C. and Axtell, M. J. (2012) miR156 and miR390 regulate tasiRNA accumulation and developmental timing in Physcomitrella patens. *Plant Cell* 3: 1-14.

Coruh, C., Shahid, S. and Axtell, M. J. (2014) Seeing the forest for the trees: annotating small RNA producing genes in plants. *Curr. Opin. Plant Biol.* 18: 87–95.

Coruh, C., Cho, S. H., Shahid, S., Liu, Q., Wierzbicki, A. and Axtell, M. J. (2015) Comprehensive Annotation of *Physcomitrella patens* Small RNA Loci Reveals That the Heterochromatic Short Interfering RNA Pathway Is Largely Conserved in Land Plants. *Plant Cell* 27: 2148–2162

Cuperus, J. T., Fahlgren, N. and Carrington, J. C. (2011) Evolution and functional diversification of MIRNA Genes. *Plant Cell* 23: 431-442.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and

space complexity. *BMC Bioinformatics 5*, 113-121.

Eklund, D. M., Ishizaki, K., Flores-sandoval, E., Kikuchi, S., Takebayashi, Y., Tsukamoto, S., *et al*. (2015) Auxin produced by the indole-3-pyruvic acid pathway regulates development and gemmae dormancy in the liverwort *Marchantia polymorpha*. *Plant Cell* 27: 1650-1669.

Emily, J. F., Floyd, S. K., Alvarez, J., Eshed, Y., Hawker, N. P., Izhaki, A., Baum, S. F. and Bowman, J. L. (2003) Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. *Curr. Biol.* 13: 1768-74.

Fahlgren, N., Howell, M. D., Kasschau, K. D., Chapman, E. J., Sullivan, C. M., Cumbie, J. S., Givan, S. A., Law, T. F., Grant, S. R., Dangl, J. L. and Carrington, J. C. (2007) High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. *PloS One 2*: e219.

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.L., Wei, P. *et al*. (2013) Efficient genome editing in plants using a CRISPR/Cas system. *Cell Res*. 23: 1229–1232.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. and Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.

Nature 391: 806–811.

Flores-Sandoval, E., Eklund, D. M. and Bowman, J. L. (2015) A Simple Auxin Transcriptional Response System Regulates Multiple Morphogenetic Processes in the Liverwort Marchantia polymorpha. *Plos Gen.* 11: e1005207.

Gamborg, O. L., Miller, R. A and Ojima, K. (1968) Nutrient requirements of suspension cultures of soybean root cells. *Exp. Cell Res.*: 50: 151-158.

Gasciolli, V., Mallory, A. C., Bartel, D. P. and Vaucheret, H. (2005). Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Current Biology* 15: 1494–500.

Golden, T. A., Schauer, S. E., Lang, J. D., Mushegian, A. R., Grossniklaus, U., Meinke, D. W. and Ray, A. (2002) FACTORY , a Dicer Homolog , Is a Maternal Effect Gene Required for Embryo Development in Arabidopsis 1. *Plant Physiol.* 130: 808–822.

Hamilton, A. J. and Baulcombe, D. C. (1999) A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants. *Science* 286: 950–952.

Herr, A. J., Jensen, M. B., Dalmay, T. and Baulcombe, D. C. (2005) RNA polymerase IV

directs silencing of endogenous DNA. *Science* 308: 118–120.

Iki, T., Yoshikawa, M., Nishikiori, M., Jaudal, M. C., Matsumoto-Yokoyama, E., Mitsuhara, I., *et al*. (2010) In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. *Mol. Cell* 39: 282-291.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology 169*: 5429–33.

Ishizaki, K., Mizutani, M., Shimamura, M., Masuda, a., Nishihama, R. and Kohchi, T. (2013). Essential Role of the E3 Ubiquitin Ligase NOPPERABO1 in Schizogenous Intercellular Space Formation in the Liverwort Marchantia polymorpha. *Plant Cell* 25: 4075-4084.

Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T. and Tomari, Y. (2010) Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol. Cell* 39: 292-299.

Jacobsen, S. E., Running, M. P. and Meyerowitz, E. M. (1999) Disruption of an RNA

helicase/RNAse III gene in Arabidopsis causes unregulated cell division in floral meristems. *Development* 126: 5231–43.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* 337: 816–821.

Jones-rhoades, M. W. and Bartel, D. P. (2004) Computational Identification of Plant MicroRNAs and Their Targets , Including a Stress-Induced miRNA. Mol. Cell 14: 787–799.

Jung, J.-H. and Park, C.-M. (2007) MIR166/165 genes exhibit dynamic expression patterns in regulating shoot apical meristem and floral development in Arabidopsis. *Planta* 225: 1327–1338.

Kanazawa, T., Era, A., Minamino, N., Shikano, Y., Fujimoto, M., Uemura, T., Nishihama, R., Yamato, K. T., Ishizaki, K., Nishiyama, T., Kohchi, T., Nakano, A. and Ueda, T. (2015) SNARE Molecules in *Marchantia polymorpha* : Unique and Conserved Features of the Membrane Fusion Machinery. *Plant and Cell Physiol.* 12*,* pcv076.

Kato, H., Ishizaki, K., Kouno, M., Shirakawa, M., Bowman, J. L., Nishihama, R., *et al*. (2015) Auxin-Mediated Transcriptional System with a Minimal Set of Components Is Critical for Morphogenesis through the Life Cycle in Marchantia polymorpha. *Plos Gen.* 11: e1005084.

Kawamata, T. and Tomari, Y. (2010) Making RISC. *Trends Biochem. Sci*. 35: 368–76.

Kubota, A, Ishizaki, K, Hosaka, M, and Kohchi, T. (2013) Efficient Agrobacterium-mediated transformation of the liverwort Marchantia polymorpha using regenerating thalli. *Biosci, Biotech, and Biochem.* 77: 167-72.

Kubota, A., Kita, S., Ishizaki, K., Nishihama, R., Yamato, K. T. and Kohchi, T. (2014) Co-option of a photoperiodic growth-phase transition system during land plant evolution. *Nature Commun. 5*: 1–9.

Kurihara, Y. and Watanabe, Y. (2004) Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Nat. Aca. Sci*. 101: 12753-12758.

Kurihara, Y., Takashi, Y. and Watanabe, Y. (2006) The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis.

RNA 12: 206–12.

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75: 843–54.

Li, S., Liu, L., Zhuang, X., Yu, Y., Liu, X., Cui, X., *et al*. (2013) MicroRNAs inhibit the translation of target mRNAs on the endoplasmic reticulum in Arabidopsis. *Cell* 153: 562-574.

Liu, C, Axtell, M. J, and Fedoroff, N. V. (2012). The helicase and RNaseIIIa domains of Arabidopsis DCL1 modulate catalytic parameters during microRNA biogenesis. *Plant Physiol.* 159: 748-758.

Llave, C., Xie, Z., Kasschau, K. D., Carrington, J. C. (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* 297: 2053-2056.

Lu, C., Jeong, D.-H., Kulkarni, K., Pillay, M., Nobuta, K., German, R., Thatcher, S. R., Maher, C., Zhang, L., Ware, D., Liu, B., Cao, X., Meyers, B. C. and Green, P. J. (2008) Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). *Proc. Nat. Aca. Sci.* 105: 4951–4956.

McConnell, J. R. and Barton, M. K. (1998) Leaf polarity and meristem formation in

Arabidopsis. *Development* 125: 2935–2942.

Mcconnell, J. R., Emery, J., Eshed, Y., Bao, N., Bowman, J. and Barton, M. K. (2001) Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. *Nature* 411: 709–713.

Meyers, B. C., Axtell, M. J., Bartel, B., Bartel, D. P., Baulcombe, D., Bowman, J. L., Cao, X., Carrington, J. C., Chen, X., Green, P. J., Griffiths-Jones, S., Jacobsen, S. E., Mallory, A. C., Martienssen, R. A., Poethig, R. S., Qi, Y., Vaucheret, H., Voinnet, O., Watanabe, Y., Weigel, D. and Zhu, J.-K. (2008). Criteria for annotation of plant MicroRNAs. *Plant Cell 20*: 3186–3190.

Morgan, M., Anders, S., Lawrence, M., Aboyoun, P., Pagès, H., and Gentleman, R. (2009). ShortRead: A bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics* 25: 2607-2608.

Morin, R. D., Aksay, G., Dolgosheina, E., Ebhardt, H. A., Magrini, V., Mardis, E. R., Sahinalp, S. C., Unrau, P. J. and Va, B. (2008). Comparative analysis of the small RNA transcriptomes of Pinus contorta and Oryza sativa. *Genome Res.* 18: 571–584.

Napoli, C., Lemieux, C. and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *Plant Cell* 2: 279–289.

Nodine, M. D. and Bartel, D. P. (2010) MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Gene. Dev.* 24: 2678–2692.

Qi, Y., He, X., Wang, X.-J., Kohany, O., Jurka, J. and Hannon, G. J. (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* 443: 1008–1012.

Qiu, Y.L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., *et al*. (2006) The deepest divergences in land plants inferred from phylogenomic evidence. *Proc. Natl Acad. Sci*. 103: 15511–15516.

Rajagopalan, R., Vaucheret, H., Trejo, J. and Bartel, D. P. (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Gene.Dev.* 20: 3407–3425.

Ray, A., Lang, J. D., Golden, T. and Ray, S. (1996) SHORT INTEGUMENT (SIN1), a gene required for ovule development in Arabidopsis, also controls flowering time. *Development*

122: 2631–2638.

Rensing, S. a, Ick, J., Fawcett, J. a, Lang, D., Zimmer, A., Van de Peer, Y., & Reski, R. (2007). An ancient genome duplication contributed to the abundance of metabolic genes in the moss Physcomitrella patens. *BMC Evol. Biol.* 7*:* 130.

Rensing, S. A. *et al*. (2008). The Physcomitrella genome reveals evolutionary insights into conqest of land by plants. *Science* 319: 64–69.

Rubio-Somoza, I., Zhou, C.-M., Confraria, A., Martinho, C., von Born, P., Baena-Gonzalez, E., Wang, J.-W. and Weigel, D. (2014) Temporal Control of Leaf Complexity by miRNA-Regulated Licensing of Protein Complexes. *Current Biology* 24: 2714–2719.

Saleh, O., Issman, N., Seumel, G. I., Stav, R., Samach, A., Reski, R., Frank, W. and Arazi, T. (2011). MicroRNA534a control of BLADE-ON-PETIOLE 1 and 2 mediates juvenile-to-adult gametophyte transition in Physcomitrella patens. *Plant J* 65: 661–74.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, Dinshaw J. and Jacobsen, S. E. (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat. Stru. Mol. Biol.* 21: 64–72.

Sugano, S. S., Shirakawa, M., Takagi, J., Matsuda, Y., Shimada, T., Hara-Nishimura, I. and Kohchi, T. (2014) CRISPR/Cas9-mediated targeted mutagenesis in the liverwort Marchantia polymorpha L. *Plant and Cell Physiol.* 55: 475-481.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. and Evol.* 30: 2725-2729.

Tsuzuki, M., Takeda, A., and Watanabe, Y. (2014). Recovery of dicer-like 1-late flowering phenotype by miR172 expressed by the noncanonical DCL4-dependent biogenesis pathway. *RNA* 20: 1320-1327.

Ueda, M., Kuniyoshi, T., Yamamoto, H., Sugimoto, K., Ishizaki, K., Kohchi, T., *et al*. (2012) Composition and physiological function of the chloroplast NADH dehydrogenase-like complex in Marchantia polymorpha. *Plant J* 72: 683-693.

Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I. S. and Martienssen, R. A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297: 1833–1837.

Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N.,

Ayyampalayam, S., *et al*. (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl. Acad. Sci.* 111: 4859–4868.

Wu, G. and Poethig, R. S. (2006). Temporal regulation of shoot development in Arabidopsis thaliana by miR156 and its target SPL3. *Development* 133: 3539-3547.

Wu, G., Park, M. Y., Conway, S. R., Wang, J., Weigel, D. and Poethig, R. S. (2009) The Sequential Action of miR156 and miR172 Regulates Developmental Timing in Arabidopsis. *Cell* 138: 750-759.

Wu, M.-F., Tian, Q. and Reed, J. W. (2006) Arabidopsis microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction. *Development 133*: 4211–4218.

Xiao, A., Cheng, Z., Kong, L., Zhu, Z., Lin, S., Gao, G. and Zhang, B. (2014) CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics* 30: 1180–1182.

Xie, Z., Johansen, L. K., Gustafson, A. M., Kasschau, K. D., Lellis, A. D., Zilberman, D., Jacobsen, S. E. and Carrington, J. C. (2004). Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol*. 2: E104.

Xie, Z., Allen, E., Wilken, A. and Carrington, J. C. (2005) DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana. *Proc. Natl Acad. Sci.* 102: 12984–12989.

Xing, S., Salinas, M., Höhmann, S., Berndtgen, R. and Huijser, P. (2010). miR156-targeted and nontargeted SBP-box transcription factors act in concert to secure male fertility in Arabidopsis. *Plant Cell* 22: 3935–50.

Xu, B., Ohtani, M., Yamaguchi, M., Toyooka, K., Wakazaki, M., Sato, M., Kubo, M., Nakano, Y., Sano, R., Hiwatashi, Y., Murata, T., Kurata, T., Yoneda, A., Kato, K., Hasebe, M. and Demura, T. (2014). Contribution of NAC Transcription Factors to Plant Adaptation to Land. *Science* 343: 1505-1508.

Yamato, K.T., Ishizaki, K., Fujisawa, M., Okada, S., Nakayama, S., Fujishita, M., *et al*. (2007) Gene organization of the liverwort Y chromosome reveals distinct sex chromosome evolution in a haploid system. *Proc. Natl. Acad. Sci*. 104: 6472–6477.

Yamato, K. T. and Kohchi, T. (2012) The “sneak” preview of the *Marchantia polymorpha* genome. *BSJ review* 3: 71-83.

Yoshikawa, M., Peragine, A., Park, M. Y. and Poethig, R. S. (2005) A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Gene. Dev.* 19: 2164-2175.

Zhao, T., Li, G., Mi, S., Li, S., Hannon, G. J., Wang, X. J. and Qi, Y. (2007). A complex system of small RNAs in the unicellular green alga. *Gene. Dev.* 21: 1190–1203.

Zobell, O., Faigl, W., Saedler, H. and Münster, T. (2010) MIKC* MADS-Box proteins: Conserved regulators of the gametophytic generation of land plants. *Mol. Biol. Evol.* 27: 1201-1211.

Acknowledgements

First of all, I would like to express the best appreciation to Prof. Yuichiro Watanabe for leading me to greatness of and letting me know respect for scientific researches and nature throughout my doctoral period.

I am very grateful to Prof. Takayuki Kohchi in Kyoto Univ. and Prof. John L. Bowman in Monash Univ. for establishing and developing all circumstances to use *Marchantia polymorpha* as a great model plant on molecular and evolutionary genetics.

I also thank Drs. Ryuichi Nishihama in Kyoto Univ., Kimitsune Ishizaki in Kobe Univ., Yukio Kurihara and Minami Matsui in RIKEN for providing various tools and kind advice.

I express my thanks to Assis. Prof. Takahiro Hamada for giving me kind advice and having interesting and ambitious discussion together.

I extend my gratitude to Assoc. Prof. Atsushi Takeda for teaching me all of beginnings about researches.

I thank all of the past and present members in Watanabe Lab for spending time at lab together.

Finally, I am obliged to my parents for supporting my student life in every aspect.