Novel micro-RNAs and intermediates of micro-RNA biogenesis from moss

Mali Talmor-Neiman1, Ran Stav1, Wolfgang Frank2, Bjoern Voss2 and Tzahi Arazi1,*
1Department of Ornamental Horticulture, Agricultural Research Organization, The Volcani Center, PO Box 6, Bet Dagan 50250, Israel, and
2Plant Biotechnology, University of Freiburg, Schauenzlestr. 179104, Freiburg, Germany

Received 18 January 2006; revised 2 March 2006; accepted 3 March 2006.
*For correspondence (fax +972 3 9669583; e-mail tarazi@volcani.agri.gov.il).

Summary

Micro-RNAs (miRNAs) are one class of small non-coding RNAs that have important regulatory roles in higher plants. Much less is known about their prevalence and function in lower land plants. Previously we cloned 100 non-structural small RNAs from the moss
Physcomitrella patens
but could annotate only 11 as miRNAs. To identify additional moss miRNAs among cloned small RNAs we have analyzed their genomic sequences for a characteristic miRNA precursor-like structure. This analysis revealed 19 new moss miRNAs that are predicted to be encoded by 22 putative foldbacks. Northern blot analysis confirmed the expression of 14 new miRNA representatives. Half of these were gametophore specific, the rest were detected at low levels in the protonema. We predicted 12 genes as targets of nine new miRNAs. Three of these show homology to transcription factors and the others appear to play roles in diverse physiological processes including light and cytokine signaling, which have not to date been shown to be regulated by a miRNA in flowering plants. Four target genes, which show homology to ATN1-like protein kinase, NAC transcription factors and a cytokinin receptor, have been validated by miRNA-mediated mRNA cleavage. In addition, our analysis revealed that seven small RNAs represent miRNA* and three represent intermediates of pre-miRNA processing, providing evidence for specific DICER-like cleavage steps during miRNA biogenesis in moss. Our findings suggest that miRNAs are common in mosses and set the stage for the elucidation of their varied biological functions.

Keywords: Physcomitrella patens, moss, miRNA, biogenesis, cytokinin, phytochrome.

Introduction

Micro-RNAs (miRNAs) form an abundant class of 21–24-nucleotide small RNAs that are common to diverse species of multicellular life. In flowering plants they are major regulators of development, but have also been predicted, or confirmed, to regulate other processes such as signal transduction, protein degradation and responses to environmental stress and pathogen invasion (reviewed by
Zhang et al., 2006).

The biogenesis of an miRNA in plants seems to be a multistep process that begins with the transcription of an
MIRNA gene into a large capped and polyadenylated primary transcript (pri-miRNA) that contains an imperfect foldback structure (Chen, 2005; Xie et al., 2005). It has been demonstrated that miRNAs in many higher plants require Dicer-Like1 (DCL1) (Park et al., 2002). It has been established that pri-miR163, pri-miR164b and pri-miR166a are cleaved by DCL1 to a stem-loop precursor or pre-miRNA (Kurihara and Watanabe, 2004; Kurihara et al., 2006). This step is believed to occur in the nucleus (Papp et al., 2003; Park et al., 2005). Pre-miR163 is then processed to short pre-miR163 (Kurihara and Watanabe, 2004). Pre-miR-164b and -166a and short pre-miR163 are further processed by DCL1, to miRNA:miRNA* and a loop remnant (Kurihara and Watanabe, 2004; Kurihara et al., 2006). The miRNA:miRNA* duplex comprises a mature miRNA and a similarly sized complementary fragment called miRNA*; these are offset by two nucleotides because of the staggered cuts of DCL1 (Elbashir et al., 2001). After its release from the pre-miRNA, the miRNA:miRNA* duplex is translocated into the cytoplasm by HASTY, the plant exportin-5 homolog (Park et al., 2005). In the cytoplasm, mature miRNAs are selectively loaded into the RNA-induced silencing complex (RISC) and primarily
guide the cleavage of target mRNAs (Llave et al., 2002; Palatnik et al., 2003; Schwab et al., 2005; Tang et al., 2003).

To date, 872 miRNAs belonging to 42 families have been identified in 71 plant species, most of which are flowering plants (Zhang et al., 2006); the sequence data are accessible at the miRNA registry (http://microrna.sanger.ac.uk/sequences/index.shtml). These miRNAs provided important insights into the functions of miRNAs in land plants. Still, to be able to understand the extent of miRNA regulation throughout the land plant clade, additional information on miRNAs and their biogenesis in more diverse land plants is needed. Mosses are one of the oldest groups of land plants among the Earth’s flora. They originated around 350 million years ago and diverged from flowering plants more than 400 million years ago, although subsequently, as indicated by comparison with the fossil record, they have evolved little (Cove et al., 1997). Information on the molecular biology of mosses has come mainly from studies of the model monocious moss Physcomitrella patens (Cove, 2000) in which efficient homologous recombination allows targeted gene disruption for the study of individual gene function (Schaefer and Zryd, 1997). Although numerous small RNAs have been cloned from the mosses Polytrichum juniperinum (Axtell and Bartel, 2005) and P. patens (Arazi et al., 2005), only 11 were identified as authentic miRNAs in P. patens on the basis of their homology to higher plant miRNAs (Pp-miR319a, Pp-miR319b, Pp-miR319c, miR390 and miR535), their abundance in the library, expression or putative hairpin precursors (miR534, miR538, miR533a, miR536 and miR537) (Arazi et al., 2005). In addition, a homolog of miR160 was identified in Polytrichum juniperinum (Axtell and Bartel, 2005).

In the present study, we describe the identification of 19 new P. patens miRNAs from cloned small RNAs following the analysis of the corresponding raw genomic sequence. Target prediction and validation suggest that these miRNAs are involved in various physiological processes. In addition, we identified several miRNA* and intermediates of pre-miRNA processing, shedding light on miRNA biogenesis in lower land plants.

**Results**

**Folding analysis of P. patens raw genomic sequences identifies novel miRNAs**

Previously we cloned a set of 100 non-structural small RNAs from P. patens protonema (Arazi et al., 2005). Further sequencing of that library identified 24 additional distinct small RNA sequences (Table S1). Nevertheless, the majority of cloned small RNAs could not be analyzed for the stem-and-loop gene structure indicative of a miRNA because the moss genomic sequence was not available. However, a shotgun genome sequence of P. patens was recently released including 7,341,086 raw sequences (http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?). Therefore, to identify novel miRNAs, previously unannotated small RNAs were analyzed as illustrated in Figure 1. Our analysis mapped a total of 29 small RNAs to 27 pre-miRNA-like foldback structures (Figure 2). We found that Pp_34, Pp_103, Pp_108 and Pp_111 were complementary with two nucleotides offset to miR533a, miR534, miR537 and Pp-miR319b, respectively. Moreover, their cloning frequency was much lower than the corresponding miRNA (Table S1 and Table S1 in Arazi et al., 2005), suggesting that each represents a miRNA* (Figure 2, gray). Another small RNA, Pp_76, was complementary to Pp_116 with two nucleotides offset (Figure 2, Pp-MIR390c, gray). Because only Pp_116 could be detected by Northern blotting we concluded that Pp_76 is the miRNA* of Pp_116, which by itself represents a new P. patens miRNA (miR1222). We also found that small RNA Pp_29 corresponded to a 3’ arm of a foldback (Figure 2, Pp-MIR390c, gray) with two nucleotides offset to a Pp-miR390a complementary sequence. As Pp-MIR390c was not 100% identical to Pp-miR390a, we suggest that Pp_29 is the miRNA* of an as yet un-cloned member of the miR390 family (Pp_miR390b) in P. patens that might be shorter than miR390 by one nucleotide at its 5’ end (Table 1).

Three foldbacks (Pp-MIR1219a, Pp-MIR1219b and Pp-MIR1219c) are predicted to encode Pp_106a and one (Pp-MIR1219d) is predicted to encode an as yet uncloned family member of Pp_106a (Pp_106b), which is different by one nucleotide at its 5’ end (Table 1). Moreover, folding analysis indicated that Pp_124 was the miRNA* of Pp_106a, which validated its identity as an authentic miRNA (miR1219a; Figure 2, Pp-MIR1219a, gray). Interestingly, putative precursors of miR1219a and miR1219b were found to be organized in two clusters in which precursor pairs are located within

![Figure 1](http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?)

**Figure 1.** Flowchart for the identification of candidate miRNAs in cloned small RNAs. The number of small RNAs passing each stage is shown in parentheses. Identity to a genome sequence was determined by BLASTN against the P. patens genome database (http://moss.nibb.ac.jp/blast/blast.html). Redundant identical genomic sequences were then clustered using the CLUSTALW sequence alignment tool (Thompson et al., 1994). Foldback prediction and stability was done by Mfold (Zuker, 2003) and RANFOLD software (Bonnet et al., 2004b), respectively. The consensus properties of plant pre-miRNAs were as formulated previously by Xie et al. (2005).
approximately 200 bp of each other (Figure 3), suggesting
that they belong to the same pri-miRNA. In contrast to
animal pre-miRNAs, which are commonly clustered, this
arrangement is quite rare in flowering plants and has been
reported to date only for miR399 (Sunkar and Zhu, 2004)
and rice miR395 (*Oryza sativa*; Jones-Rhoades and Bartel, 2004).

### Table 1 Novel *P. patens* candidate miRNAs

<table>
<thead>
<tr>
<th>Small RNA</th>
<th>Length</th>
<th>Sequence</th>
<th>Trace/MIR gene</th>
<th>Foldback arm/ nucleotides</th>
<th>RNA blot</th>
<th>miRNA*</th>
<th><em>P. patens</em> EST</th>
<th>Homology*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pp_13</td>
<td>21</td>
<td>AGAGGCCUUUGUGGGAGAGGAG 66715266/MIR1210</td>
<td>3'/99</td>
<td>d</td>
<td></td>
<td>Pj_10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_18</td>
<td>21</td>
<td>AGGGAGGAGGUAUGGAAGCAAG 713855711/MIR1211</td>
<td>5'/112</td>
<td>d</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_26</td>
<td>21</td>
<td>CCAUACAGGGAGUCAGACAGA 784207533/–</td>
<td>3'/62</td>
<td>nd</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_31</td>
<td>21</td>
<td>CGUGGCGACAGCAUGAAGGAG 713871562/MIR1212</td>
<td>3'/99</td>
<td>d</td>
<td></td>
<td>Pj_170</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_44</td>
<td>21</td>
<td>GACUACACCAAGGCAAGCAGACA 717625569/–</td>
<td>3'/62</td>
<td>nd</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_53</td>
<td>21</td>
<td>GUGGAGGCAGCCUGUGGGAGA 828270278/MIR1213</td>
<td>3'/164</td>
<td>d</td>
<td></td>
<td>Pj_201</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_55</td>
<td>21</td>
<td>UACUGAGGAAGCCUACUGAGCA 815622998/MIR1214</td>
<td>5'/93</td>
<td>d</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_63</td>
<td>21</td>
<td>UACUGGAACUAC CCCUGUGGAGA 831635002/MIR1215</td>
<td>5'/127</td>
<td>d</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_75</td>
<td>21</td>
<td>UACUGGGACUGCCUGUGUAUC 774450283/MIR1216</td>
<td>5'/158</td>
<td>d</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_101</td>
<td>21</td>
<td>AAUUGAAGCAUGAUGUCAAG 830441252/MIR1217</td>
<td>3'/111</td>
<td>d</td>
<td></td>
<td>Pp_101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_104</td>
<td>21</td>
<td>CUUCAGAGAGUAGCCUCCUG 774410268/MIR1218</td>
<td>5'/138</td>
<td>d</td>
<td></td>
<td>pph34n07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_106a</td>
<td>21</td>
<td>CUUCCUGCCUCUCACUGAUU 755769908/MIR1219a</td>
<td>5'/135</td>
<td>d</td>
<td></td>
<td>Pp_124</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_106b</td>
<td>21</td>
<td>uucugcccucucucuagcuu 824631517.2/MIR1219c</td>
<td>5'/152</td>
<td>d</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_107</td>
<td>21</td>
<td>GAAGUAGAGAGGUGUAAGA 784303541/MIR1220a</td>
<td>3'/205</td>
<td>d</td>
<td></td>
<td>Pp_106a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_109</td>
<td>21</td>
<td>GACUGGCGACAGCUCUGAGGAA 755815369/MIR533a</td>
<td>5'/108</td>
<td>nt</td>
<td></td>
<td>miR533</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_110</td>
<td>21</td>
<td>GACUGUGCGACGUGAGAGA 85903236/MIR533b</td>
<td>5'/125</td>
<td>nt</td>
<td></td>
<td>miR533</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_111</td>
<td>21</td>
<td>UACGAACGAGAGAGACGAGC 756850258/MIR535d</td>
<td>5'/85</td>
<td>nt</td>
<td></td>
<td>miR535</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_114</td>
<td>21</td>
<td>UAGAUUGGGAGCCUGUACAA 759445026/MIR1221</td>
<td>5'/177</td>
<td>d</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_116</td>
<td>21</td>
<td>UUGAAGGAGGUGCAUGAUGUA 759457106/MIR1222</td>
<td>3'/130</td>
<td>d</td>
<td></td>
<td>Pp_76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_117</td>
<td>21</td>
<td>UUGAAGGAGGUGCAUGUCA 831697289/MIR1223</td>
<td>3'/126</td>
<td>d</td>
<td></td>
<td>Pp_62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pp_miR390b 20</td>
<td>aacucugaggaggaagcgc</td>
<td>78422414430/MIR390c</td>
<td>5'/135</td>
<td>nt</td>
<td></td>
<td>Pp_29</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Lower case nucleotides are predicted. The precise ends of the miRNA are not yet been determined.
*One representative trace for each independent gene is shown. (–) identity as a miRNA was not confirmed.
*Predicted; the actual size of each precursor has not been identified experimentally and may be slightly shorter or longer than represented.
*nd, not detected; d, detected; nt, not tested because of sequence homology.
*Pj, Polytichium juniperinum.*

Sequence alignments of newly identified candidate
miRNAs to known *P. patens* miRNAs (Arazi *et al.*, 2005)
and to small RNAs from the moss *Polytrichum juniperinum* (Axtell and Bartel, 2005) suggest that Pp_109
(miR533b) and Pp_110 (miR533c) are members of the
miR533 family (Figure 4a) and Pp_113 (miR535b) is a
member of the miR535 family (Figure 3b). However, Pp-MiR533a that is predicted to encode the 21-nucleotide miR533a is also predicted to encode the 22-nucleotide miR533b (Figure 2), and thus we cannot exclude the possibility that miR533b is a processing variant of miR533a. In addition, Pp_13 and Pp_31 were found to be identical to Pj_10 and Pj_170, respectively, and Pp_53 and miR1222 had 90.5% identity to Pj_201 (Figure 4c) and Pj_62 (Figure 4d), respectively, raising the possibility that these small RNAs represent conserved miRNAs in mosses (Table 1).

As pri-miRNAs are capped and polyadenylated non-coding RNAs, such pri-miRNAs would appear as non-coding expressed sequence tags (ESTs) (Zhang et al., 2005). A BLASTN search against the P. patens EST database (http://moss.nibb.ac.jp/blast/blast.html) revealed that Pp-MiR1217 and Pp-MiR1218 were identical to the non-coding EST contigs 14150 and pph34n07, respectively. These findings support the characterization of Pp_101 and Pp_104 as miRNAs (miR1217 and miR1218, respectively) that are encoded in capped and polyadenylated pri-miRNA. Thus, the analysis presented so far provides supporting evidence for the identity of 10 new miRNAs (miR1210, miR1212, miR1213, miR1217, miR1218, miR1219a, miR533b, miR533c, miR535b and miR1222), two predicted miRNAs (miR1219b and Pp_miR390b) and seven miRNA* in moss. The rest of the small RNAs (Pp_18/miR1211, Pp_26, Pp_44, Pp_55/miR1214, Pp_63/miR1215, Pp_75/miR1216, Pp_107/miR1220, Pp_114/miR1221 and Pp_117/miR1223), which corresponded to a predicted pre-miRNA foldback structure (except Pp_82, Pp_112 and Pp_115), were characterized as candidate miRNAs (Table 1).
Several cloned small RNAs represent intermediates of pre-miRNA processing

Our folding analysis revealed that Pp_82 matched the root of the putative miR1219a precursor Pp-MIR1219a just three nucleotides upstream of miR1219a (Figure 5a, underlined). In addition, Pp_112 and Pp_115 were found to be identical to the putative Pp-miR319b precursor Pp-MIR319d. They matched a region in its elongated stem between Pp-miR319b:Pp-miR319b* and the loop structure (Figure 5b, single and double underlined). Moreover, Pp_112 and Pp_115 are predicted to complement each other with two nucleotides offset reminiscent of a miRNA:miRNA* duplex (Figure 5b), suggesting that their biogenesis is mediated by a DICER-like protein.

Expression analysis of newly identified miRNAs

The miRNAs that we previously characterized in moss were relatively abundant in the protonema small RNA library (Arazi et al., 2005). Conversely, 15 out of the 19 new candidate miRNAs were identified only once in the small RNA library, implying that their abundance in the protonema should be relatively low. Consistent with that, their expression in the moss gametophyte could be detected by Northern blots only on 100 µg samples of total RNA extracted from 8–10-day-old protonema and 45-day-old gametophore or in a low-molecular-weight RNA fraction isolated from 500 µg of protonema total RNA. Northern blots confirmed that 14 out of 16 representative miRNAs tested were stably expressed in the gametophyte, which reinforced their identity as authentic miRNAs (Figure 6). miR1210, miR1213, miR1214, miR1218, miR1221, miR1222 and miR1223 were detected almost exclusively in mature gametophores and were barely detectable in the protonema (Figure 6, lane G), consistent with their low abundance in the protonema library. miR1211, miR1212, miR1215, miR1216, miR1217, miR1219a and miR1220 could be detected only in concentrated protonemal low-molecular-weight RNA (Figure 6, lane C), suggesting that they are expressed at very low steady state levels in the gametophyte. This may be because they are expressed only under particular growth conditions, in specific cell types or mainly in a developmental stage such as the sporophyte that was not tested in this study. Multiple band signals of sizes bigger than 21 nucleotides were detected with probes complementary to Pp_26 and Pp_44 (data not shown), and therefore the identity of these candidates as miRNAs could not be confirmed.

Predicted targets of newly identified miRNAs

To identify the potential targets of newly identified miRNAs, we first used PATSCAN (Dsouza et al., 1997) to search the P. patens public EST database for mRNAs that were

Figure 4. Homology of newly identified miRNAs to known miRNAs and Polytrichum juniperinum cloned small RNAs.
(a–d) Sequences were aligned with the CLUSTALW program. Identical and conserved nucleotides are labeled in upper case letters. Lower case letters represent changes in nucleotide sequences in some group members. Pj, Polytrichum juniperinum.

Expression analysis of newly identified miRNAs

The miRNAs that we previously characterized in moss were relatively abundant in the protonema small RNA library (Arazi et al., 2005). Conversely, 15 out of the 19 new candidate miRNAs were identified only once in the small RNA library, implying that their abundance in the protonema should be relatively low. Consistent with that, their expression in the moss gametophyte could be detected by Northern blots only on 100 µg samples of total RNA extracted from 8–10-day-old protonema and 45-day-old gametophore or in a low-molecular-weight RNA fraction isolated from 500 µg of protonema total RNA. Northern blots confirmed that 14 out of 16 representative miRNAs tested were stably expressed in the gametophyte, which reinforced their identity as authentic miRNAs (Figure 6). miR1210, miR1213, miR1214, miR1218, miR1221, miR1222 and miR1223 were detected almost exclusively in mature gametophores and were barely detectable in the protonema (Figure 6, lane G), consistent with their low abundance in the protonema library. miR1211, miR1212, miR1215, miR1216, miR1217, miR1219a and miR1220 could be detected only in concentrated protonemal low-molecular-weight RNA (Figure 6, lane C), suggesting that they are expressed at very low steady state levels in the gametophyte. This may be because they are expressed only under particular growth conditions, in specific cell types or mainly in a developmental stage such as the sporophyte that was not tested in this study. Multiple band signals of sizes bigger than 21 nucleotides were detected with probes complementary to Pp_26 and Pp_44 (data not shown), and therefore the identity of these candidates as miRNAs could not be confirmed.

Predicted targets of newly identified miRNAs

To identify the potential targets of newly identified miRNAs, we first used PATSCAN (Dsouza et al., 1997) to search the P. patens public EST database for mRNAs that were

Figure 5. Physcomitrella patens cloned small RNAs that represent intermediates of pre-miRNA processing.
(a) Sequence of Pp-MIR1219a stem and loop. The sequences of mir1219a, mir1219a* and Pp_82 are bold-face, gray boxed and underlined, respectively.
(b) Sequence of Pp-MIR319d stem and loop. The sequences of Pp-miR319b, Pp-miR319b*, Pp_112 and Pp_115 are bold-face, gray boxed, underlined and double underlined, respectively.
complementary to them according to the empirical parameters as formulated by Schwab et al. (2005) and described in experimental procedures. In addition, gaps were not allowed, and the G:U non-canonical pair was treated as a mismatch. This analysis identified only one potential target for miR1218 (pshf2k14) that was predicted to encode a NAM/ATAF/CUC (NAC) domain-containing protein (Table 2).

To identify additional targets, a similar search was performed against the P. patens raw genomic sequences deposited at the NCBI trace archive data base (http://www.ncbi.nlm.nih.gov/Traces/traces.cgi?). This search identified numerous genomic sequences with complementarity to all the newly identified miRNAs according to the above rules (data not shown). However, to determine their orientation and thus their relevance as miRNA targets, redundant sequences were first clustered by using the CLUSTALW sequence alignment tool (Thompson et al., 1994). Then a BLASTX search against the NCBI protein database (http://www.ncbi.nlm.nih.gov/BLAST/blast.cgi?) was performed to identify significant similarities to known proteins in the data base. This search identified 12 sequences with similarities to known proteins, which could serve as feasible targets for nine out of the 19 identified miRNAs (Table 2). In nine of these genes the miRNA complementary site was found within a predicted open reading frame (ORF). Five of these genes, which were predicted to be targeted by either miR1215 or miR1217, encoded a protein with a putative kinase domain. These genes showed homology to a protein kinase (756726571), to phytochrome-like kinases (830632667, 816301061, 815676221), and to a disease resistance-like protein (692445811). The remaining targets appeared to encode proteins with homology to an F-BOX protein, a B-box zinc finger, NAC-domain transcription factors, a retrotransposon nucleocapsid protein, an ABC-transporter and a protein with strong homology to Arabidopsis cytokinin response 1 (CRE1)/wooden leg (WOL)/Arabidopsis histidine kinase 4 (AHK4) cytokinin receptor (Table 2).

It was suggested that intermediates of Arabidopsis short pre-miR163 processing might function as miRNAs (Kurihara and Watanabe, 2004). To test this possibility, Pp_82, Pp_112 and Pp_115 target prediction was performed as described above, against the P. patens EST and genomic databases. This search identified one potential target for Pp_115 that showed a significant similarity to an unnamed Arabidopsis protein that contains a conserved domain of unknown function (gnl|CDD|26301) termed DUF647 (Table 2).

**Experimental validation of predicted P. patens miRNA targets**

Most higher plant miRNAs control gene expression post-transcriptionally by targeting cognate mRNAs for degradation (Llave et al., 2002; Palatnik et al., 2003; Tang et al., 2003). To verify whether newly identified P. patens miRNAs can mediate the cleavage of their predicted targets in vivo, we isolated mRNAs from 10-day-old protonema and 45-day-old gametophores and performed an RNA-ligase mediated 5′-rapid amplification of cDNA ends (5′-RACE) (Llave et al., 2002) on five target genes that were predicted to be targeted by an miRNA (Table 2): two genes were representatives of targets of weakly expressed miRNAs (756726571 targeted by miR1215 and 816301061 targeted by miR1217) and three of...
more abundant gametophore-specific miRNAs (pphf2k14 targeted by miR1218, 824685933 targeted by miR1221 and 859476845 targeted by miR1223). A 5’-RACE product could be amplified for 756726571 from the protonema cDNA, and for pphp2k14, 824685933 and 859476845 from the gametophore cDNA. An amplified product could not be detected for 816301061, although it has only one mismatch with miR1217, much lower than the cutoff (3.5 mismatches) for confident target prediction (Jones-Rhoades and Bartel, 2004). One reason for that might be the weak expression of miR1217 in the gametophyte, which suggests a relatively low level of stable 3’ cleavage fragments in that developmental stage. Sequencing of cloned amplified products revealed that 756726571, pphp2k14, 824685933 and 859476845 have specific cleavage sites corresponding to miR1215, miR1218, miR1221 and miR1223 complementary sequences, respectively (Figure 7a–d). In all cases at least 75% of the cloned product 5’ ends terminated at a position that paired with the tenth miRNA nucleotide from their 5’ ends, which suggests that they represented bona fide miRNA mediated cleavage products.

Discussion

By analyzing the folding of P. patens raw genomic sequences flanking cloned small RNAs we identified 19 for which the corresponding foldbacks possessed all the hallmarks of higher-plant miRNA precursors (Xie et al., 2005), suggesting that they represent new moss candidate miRNAs. The identity of 14 candidate miRNAs is further supported by: the discovery of their miRNA* (miR1219a, miR1222, Pp-miR390b), validation of their mRNA targets (miR1215, miR1218, miR1221, miR1223) and their homology to known P. patens miRNAs (miR533b, miR533c, miR535b) or to cloned small RNAs (miR1210, miR1212, miR1213, miR1222) of the moss Polytrichum juniperinum (Axtell and Bartel, 2005). In addition, the stable expression of all representative miRNAs was confirmed by Northern blot hybridizations. Thus, our present findings, together with the 11 miRNAs we identified previously (Arazi et al., 2005), bring the potential number of miRNAs in P. patens to 30. These miRNAs are divided among 23 distinct families of which 19 have no apparent conservation among flowering plants, which indicates that they emerged specifically in the bryophyte lineage or were lost during the evolution of flowering plants. To date, around 25 mRNA families have been discovered in Arabidopsis (Xie et al., 2005), 33 in rice (Sunkar et al., 2005; Wang et al., 2004) and 21 in Populus (Lu et al., 2005). Thus, the extent of miRNA regulation in mosses, one of the oldest groups of land plants among the Earth’s flora, is at least equal to that of flowering plants, reflecting the major role of miRNAs in the evolution of land plants.

It has been shown that the DCL1 enzymes of Arabidopsis (Park et al., 2002; Reinhart et al., 2002) and rice (Liu et al., 2005) are involved in accumulation of miRNA. We have identified seven miRNA*s and thus provide evidence for the involvement of an as yet uncharacterized DICER-like enzyme (Elbashir et al., 2001) in the biogenesis of moss miRNAs. It has been demonstrated that biogenesis of Arabidopsis miR163 requires at least three cleavage steps and that DCL1 catalyzes at least the first and the second of these (Kurihara and Watanabe, 2004). The second cleavage step, which converts long pre-miR163 to short pre-miR163, was suggested to occur only during the processing of long

---

**Table 2** Predicted and validated targets of newly identified P. patens miRNAs

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Predicted target</th>
<th>Contig/trace coordinates</th>
<th>Target site</th>
<th>Target protein family</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR1210</td>
<td>862817171 (4/1)</td>
<td>698–719</td>
<td>3’-UTR</td>
<td>F-box protein-like (Oryza sativa)</td>
</tr>
<tr>
<td>miR1212</td>
<td>75567308 (4/0)</td>
<td>744–724</td>
<td>3’-UTR</td>
<td>B-box zinc finger protein (Arabidopsis thaliana)</td>
</tr>
<tr>
<td>miR1215</td>
<td>756726571 (4/2)</td>
<td>530–510</td>
<td>ORF</td>
<td>Putative protein kinase/ATN1-like protein kinase (Arabidopsis thaliana)</td>
</tr>
<tr>
<td>miR1216</td>
<td>830632667 (4/2)</td>
<td>194–174</td>
<td>ORF</td>
<td>Photoreceptor (Ceratodon purpureus)</td>
</tr>
<tr>
<td>miR1217</td>
<td>755689191 (4/1)</td>
<td>632–612</td>
<td>ORF</td>
<td>NAM/ATAF/CUC3 (NAC3) protein (Glycine max)</td>
</tr>
<tr>
<td>miR1218</td>
<td>863137111 (3/1)</td>
<td>160–140</td>
<td>5’-UTR</td>
<td>ABC transporter family protein (Arabidopsis thaliana)</td>
</tr>
<tr>
<td>miR1221</td>
<td>824685933 (4/1)</td>
<td>498–478</td>
<td>ORF</td>
<td>CRE1/WOL/AHK4 histidine kinase</td>
</tr>
<tr>
<td>miR1223</td>
<td>859476845 (3/0)</td>
<td>336–316</td>
<td>5’-UTR</td>
<td>NAC domain protein NAC2 (Glycine max)</td>
</tr>
<tr>
<td>Pp-miR</td>
<td>759809697 (4/0)</td>
<td>325–305</td>
<td>ORF</td>
<td>Unnamed protein (Arabidopsis thaliana)</td>
</tr>
</tbody>
</table>

In parenthesis, total number of mismatches/G:U wobbles. Validated targets are in bold.

The BLASTX search was done against the NCBI non-redundant data base.
pre-miRNAs that possess stems larger than 42 nucleotides, including the miRNA. The outcome of such a processing was shown to be two 20–21-nucleotide small RNAs designated UL and LL (Kurihara and Watanabe, 2004). Pp_82, which mapped to \textit{Pp-MIR1219a} stem proximal to miR1219a, is reminiscent of such small RNAs. Moreover, \textit{Pp-MIR1219a} stem is predicted to be 49 nucleotides long including the miR1219a (Figure 5). Thus, we conclude that a similar step probably occurs during the processing of \textit{MIR1219a}. We have also found that small RNAs Pp_112 and Pp_115 formed an miRNA:miRNA*–like duplex that mapped to \textit{Pp-MIR319d} stem between Pp-miR319b:Pp-miR319b* and the loop structure (Figure 5b). This finding presents evidence of an unreported processing step of a plant pre-miRNA stem found between the miRNA:miRNA* and the loop structure that is likely to be catalyzed by a DICER-like enzyme. An interesting question is whether this processing step occurs after or before the release of the miRNA:miRNA* duplex. We predict that such a cleavage step may occur in other pre-miRNAs that have elongated double-stranded stems upstream of miRNA:miRNA* duplex, because such a folding takes a typical structure of a DICER substrate. Interestingly, Li \textit{et al.} (2005) found that Arabidopsis and rice pre-miRNAs from the \textit{MIR319} and \textit{MIR159} families had had elongated double-stranded stems that were atypically conserved in addition to their miRNA:miRNA* region, which is typically conserved in most pre-miRNAs. They hypothesized that such a conserved elongated stem might be a functional segment or might even encode for another mature miRNA. The identification of the small RNAs Pp_112 and Pp_115 and a putative target for Pp_115 supports such a hypothesis. Conversely, because \textit{Pp-MIR319d} is predicted to encode Pp-miR319b, which was the most abundant small RNA in our library (Arazi \textit{et al.}, 2005), we cannot exclude the possibility that Pp_112 and Pp_115 represent non-functional remnants from the processing of an abundant precursor.

To date, targets of miR156 (Arazi \textit{et al.}, 2005), miR160 (Axtell and Bartel, 2005) and miR166 (Floyd and Bowman, 2004) have been validated in \textit{P. patens}. The data presented in this study raise the number of \textit{P. patens} miRNA validated targets to seven. Our target prediction suggests that, as in higher plants (Zhang \textit{et al.}, 2006), the miRNAs in \textit{P. patens} are involved in the regulation of a variety of gene families. In spite of the specificity of corresponding miRNAs to mosses, several targets belong to gene families that have been already validated as miRNA targets in flowering plants, highlighting the importance of miRNA regulation for the proper function of these genes in plants. For instance, our results support that pphf2k14 and 859476845 putative NAC-domain transcription factors are genuine targets of miR1218 and miR1223, respectively. In Arabidopsis, genes encoding NAC-domain transcription factors were demonstrated to be regulated by miR164 and this regulation was shown to be required for organ separation (Laufs \textit{et al.}, 2004; Mallory \textit{et al.}, 2004) and lateral root development (Guo \textit{et al.}, 2005). miR1210 is predicted to target a gene that encodes an F-box-like protein. Genes belonging to this class were validated as miR393 and miR394 targets in Arabidopsis (Jones-Rhoades and Bartel, 2004). miR1217 is predicted to target a disease resistance-like protein, which might suggest a role for this miRNAs in a defense pathway in moss. In \textit{Populus}, several disease resistance genes were validated as targets of miR482 (Lu \textit{et al.}, 2005).

Interestingly, several targets belong to gene families that, in flowering plants, do not include members that are predicted to serve as miRNA targets. miR1215 and miR1217 are predicted to target several phytochrome-like protein coding genes. Genes belonging to this class were validated as miR164 targets in Arabidopsis (Guo \textit{et al.}, 2005). miR1210 is predicted to target a gene that encodes an F-box-like protein. Genes belonging to this class were validated as miR393 and miR394 targets in Arabidopsis (Jones-Rhoades and Bartel, 2004). miR1217 is predicted to target a disease resistance-like protein, which might suggest a role for this miRNAs in a defense pathway in moss. In \textit{Populus}, several disease resistance genes were validated as targets of miR482 (Lu \textit{et al.}, 2005).

Figure 7. Experimental validation of predicted mRNA targets of miRNAs \textit{miR1215}, \textit{miR1218}, \textit{miR1221} and \textit{miR1223}.
(a) mRNA 756726571, (b) mRNA pphf2k14, (c) mRNA 824685933, (d) mRNA 859476845. The mRNA cleavage sites were determined by modified RNA ligase-mediated 5'-RACE. The miRNA corresponding positions within the ORF or predicted untranslated regions (UTR) are indicated. The miRNA sequences and partial sequence of the corresponding mRNAs are shown. The arrows indicate the positions of inferred cleavage sites and the numbers above indicate the fraction of cloned PCR products terminating at different positions.
proteins that contain a putative serine/threonine/tyrosine kinase domain in their C-termini. A homologous phytochrome (CpPHY1) has been characterized previously from the moss Ceratodon purpureus (Pasentis et al., 1998; Thummler et al., 1992). The CpPHY1 recombinant C-terminal domain was demonstrated to phosphorylate serine and threonine residues (Thummler et al., 1995). On the basis of its predicted structure it was suggested that it might represent a soluble light sensor protein kinase that functions in a cellular phosphorylating cascade (Thummler et al., 1992). In this regard our results support that 756726571, which is predicted to encode a putative kinase-like protein, is a genuine target of miR1215. Thus, there is a possibility that both targets of miR1215, a phytochrome-kinase and a protein kinase, function in the same signal transduction pathway. In Arabidopsis, miR395 was demonstrated to target three different genes all involved in the sulfur assimilation pathway (Bonnet et al., 2004a; Jones-Rhoades and Bartel, 2004). An mRNA product, consistent with miR1221-mediated cleavage, has been detected by us for 824685933 that encodes for a moss CRE1/WOL/AHK4 homolog. CRE1/WOL/AHK4 has been identified as a cytokinin receptor in Arabidopsis (Inoue et al., 2001). In P. patens, cytokinin stimulates the differentiation of a protonema-tip growing cell into a three-faced apical cell, the so-called bud, which later produces the leafy gametophore (Reski and Abel, 1985). Our data indicate that miR1221 is relatively abundant in differentiated gametophores but almost absent from cytokinin responsive non-differentiated protonema. From that we speculate that by clearing the mRNA that encodes for a cytokinin receptor, miR1221 may reduce the response of a given gametophore cell to cytokinin.

Finally, we have identified numerous genomic sequences with complementarity to the remaining identified miRNAs. However, since the P. patens genome is raw and not annotated, homology to known proteins was used as a means of verifying their orientation. This constraint probably prevented us from identifying many additional miRNA targets, some of which probably encode for novel moss proteins. The anticipated annotation of the moss genome will soon enable their identification. Such data will enrich our knowledge of the different functions of miRNAs in land plants. Future functional studies of identified miRNAs and their targets are expected to be facilitated in P. patens in which efficient homologous recombination allows targeted gene disruption (Schaefer and Zryd, 1997).

Experimental procedures

Plant material and culture conditions

Protonemata of P. patens subspecies patens (Ashton and Cove, 1977) were cultured on a solid minimal medium described by Ashton et al. (1979) supplemented with 2.7 mM NH₄ tartrate. Cultures were grown in 9-cm Petri dishes on medium solidified with 0.7% agar (Merck 1614) and overlaid with 8-cm diameter cellophane disks (type 325P; A. A. Packaging Limited, Preston, UK). Petri dishes were kept in a culture room at 24 ± 1°C. Light was provided from above by two to four fluorescent tubes (Osrám L 18W/10; Osrám, Munich, Germany) under a regime of 16-h light/8-h darkness. Moss protonema was subcultured every 7 days with a Polytron homogenizer (type PT 1600E; Kinematica, Lucerne, Switzerland). For isolation of adult gametophores, small pieces of 1-week-old healthy protonemal tissue were inoculated for 45 days on solid minimal medium in 9-cm diameter Petri dishes, under the above temperature and light conditions. Adult gametophores were then isolated by cutting their stems a few millimeters above the agar with fine scissors.

Nucleic acid isolation

Total RNA was extracted from protonemata or gametophores with TRI reagent (Sigma, Rehovot, Israel), according to the manufacturer’s protocol, except that the upper phase containing the RNA was re-extracted two or three more times with 25:24:1 phenol:chloroform:isoamyl alcohol to remove small-RNA-binding proteins. After addition of isopropanol, the RNA extract was incubated overnight at −20°C, instead of 5 min at room temperature, to enhance the precipitation of low-molecular-weight RNAs. Following an ethanol wash, RNA was resuspended in 50% formamide (Sigma) and kept at −80°C until use. To prepare concentrated low-molecular-weight RNA from 10-day-old protonema, 500 µg of total protonema RNA was filtered through a Microcon YM100 column (Millipore, Bedford, MA, USA) according to the manufacturer’s protocol. The low-molecular-weight RNA filtrate was then collected; ethanol precipitated and resuspended in 20 µl of 50% formamide.

RNA gel blot analysis

Total RNA or low-molecular-weight RNA were resolved by electrophoresis on denaturing 15% polyacrylamide gel containing 7 M urea in TBE buffer (45 mM Tris-borate, pH 8.0 and 1.0 mM EDTA), electroblotted to a Zeta-Prob membrane (Bio-Rad, Hercules, CA, USA) by means of a trans-blot transfer cell (Bio-Rad) for 1 h at 500 mA. Following transformation, the membrane was UV cross-linked and dried overnight. Radiolabeled probes were made by end-labeling DNA oligonucleotides complementary to miRNA sequences with γ²⁵-ATP by means of T4 polynucleotide kinase (NEB). Blots were pre-hybridized and hybridized with EZ-hybridization solution (Biological Industries, Beit-Haemek, Israel). Hybridization was performed at 35–45°C overnight. Blots were washed two to three times at 45–55°C with washing buffer (2 x SSC, 0.1% SDS) and autoradiographed using a phosphor-imager (Fuji, Tokyo, Japan).

Folding analysis and target prediction

Foldback structures were predicted and their stability validated with the MFOLD program (http://www.bioinfo.rpi.edu/applications/mfold/old/ma/form1.cgi) for default parameters (Zuker, 2003) and the RANDFOLD program for the 999 randomizations and dinucleotide shuffling (Bonnet et al., 2004b). Stable foldback structures (0.005 ≥ P ≥ 0.001) were then selected only if they conformed to the consensus properties of plant pre-miRNAs as formulated previously by Xie et al. (2005): (i) a minimum of 16 paired bases within the
miRNA:miRNA* duplex, three or fewer contiguous non-paired bases and a perfect complementary not allowed; (ii) the miRNA:miRNA* duplex has to be predicted within a single foldback stem.

Targets were predicted from the *P. patens* EST database (http://moss.nibb.ac.jp/blast/blast.html) or the *P. patens* raw genomic sequences (ftp://ftp.ncbi.nih.gov/pub/TraceDB/). *PATSCAN* (Douza et al., 1997) was first used to identify predicted targets that contained sequences complementary to any of the newly identified miRNAs according to the empirical parameters as formulated by Schwab *et al.* (2005): (i) no mismatch as positions 10–11; (ii) no more then one mismatch at positions 2–12; (iii) no more than two consecutive mismatches downstream of position 13. Only predicted targets that had at least 72% of free hybridization energy with the miRNA (Markham and Zuker, 2005) compared than two consecutive mismatches downstream of position 13. (ii) no more then one mismatch at positions 2–12; (iii) no more than two consecutive mismatches downstream of position 13. Predicted targets were assigned manually according to the function of the best hit from the BLASTX homology search (Altschul *et al.*, 1990) against the NCBI non-redundant data base.

**Target validation by cleavage site mapping**

A modified procedure for RNA ligase-mediated 5' -RACE was performed with the GeneRacer Kit (Invitrogen, Carlsbad, CA, USA) according to Llave *et al.* (2002). Total RNA was isolated from 10-day-old protonema or 45-day-old gametophores grown on a minimal medium with NH₄tartrate. Poly(A)-mRNA was prepared with an Oligotex mRNA Mini Kit (Qiagen, Valencia, CA, USA) and directly ligated to GeneRacer RNA Oligo adaptor without further modification. The GeneRacer Oligo dT-18 primer was used to prime cDNA synthesis with reverse transcriptase superscript II (Invitrogen). This cDNA was subjected to an amplification procedure with the GeneRacer 5’ primer (5'-CGACTGGAGCACGAGGAGACCTGA-3') and 3’ gene-specific primers (5'-TGAAACAAGTCTTCCAGGAAATACCCCTTAGAATA-3'), (5'-GCCGCTCTGCTGTTGGAAGAGAAGA-3'), (5'-GTTACGACACTGCCACACTTACGCTACACACAC-3') and (5'-CCGGCAAATTACGGGAAATACCCCTTAGAATA-3'). Amplified 756726571, pphf2k14, 824685933 and 859476845 were subjected to nested PCR with GeneRacer 5’ nested primer (5'-GGACTCCAGTCTGGAAGAGAAGAAGA-3') and 3’ gene-specific nested primers (5'-CTGCACCAGCTGCTGAAATACACAGACACAC-3') and (5'-TACGCAGGAAATACCCCTTAGAATA-3'). The conditions used for both amplification steps were the same as those recommended by the manufacturer. The amplified products were gel purified, cloned and sequenced.

**Acknowledgements**

We thank Dr Amit Gal-On for critical reading of the manuscript. This work was supported by the Israel Science Foundation Grant 825/05 to TA. This is contribution no. 104/2006 series from the Agricultural Research Organization, the Volcani Center, Bet Dagan, Israel.

**Supplementary Material**

The following supplementary material is available for this article online:

**Table S1** Sequences of small RNAs cloned from *P. patens*

This material is available as part of the online article from http://www.blackwell-synergy.com

**References**


