Post-transcriptional adaptation of the aquatic plant *Spirodela polyrhiza* under stress and hormonal stimuli

Paul Fourounjian, Jie Tang, Bahattin Tanyolac, Yaping Feng, Brian Gelfand, Atul Kakrana, Min Tu, Chris Wakim, Blake C. Meyers, Jiong Ma and Joachim Messing

1 Waksman Institute of Microbiology, Rutgers University, New Brunswick, NJ 08854, USA, 2 Key Laboratory of Coarse Cereal Processing, Ministry of Agriculture and Rural Affairs, School of Pharmacy and Biological Engineering, Chengdu University, Chengdu 610106, China, 3 Donald Danforth Plant Science Center, Saint Louis, MO 63132, USA and 4 School of Environment and Energy, Peking University Shenzhen Graduate School, Shenzhen, Guangdong 518055, China

Received 10 November 2018; accepted 29 January 2019.

**SUMMARY**

The Lemnaceae family comprises aquatic plants of angiosperms gaining attention due to their utility in wastewater treatment, and rapid production of biomass that can be used as feed, fuel, or food. Moreover, it can serve as a model species for neotenous growth and environmental adaptation. The latter properties are subject to post-transcriptional regulation of gene expression, meriting investigation of how miRNAs in *Spirodela polyrhiza*, the most basal and most thoroughly sequenced member of the family, are expressed under different growth conditions. To further scientific understanding of its capacity to adapt to environmental cues, we measured miRNA expression and processing of their target sequences under different temperatures, and in the presence of abscisic acid, copper, kinetin, nitrate, and sucrose. Using two small RNA sequencing experiments and one degradome sequencing experiment, we provide evidence for 108 miRNAs. Sequencing cleaved mRNAs validated 42 conserved miRNAs with 83 targets and 24 novel miRNAs regulating 66 targets and created a list of 575 predicted and verified targets. These analyses revealed condition-induced changes in miRNA expression and cleavage activity, and resulted in the addition of stringently reviewed miRNAs to miRBase. This combination of small RNA and degradome sequencing provided not only high confidence predictions of conserved and novel miRNAs and targets, but also a view of the post-transcriptional regulation of adaptations. A unique aspect is the role of miR156 and miR172 expression and activity in its clonal propagation and neoteny. Additionally, low levels of 24 nt sRNAs were observed, despite the lack of recent retrotransposition.

**Keywords:** Lemnaceae, miRNA, degradome, neoteny, stress response.

**INTRODUCTION**

The Lemnaceae family, commonly called duckweeds, are the smallest, fastest growing, most widely distributed, and highly neotenous family of angiosperms. From an industrial standpoint, they are a promising crop for their ability to recover potentially eutrophying nutrients from wastewater, which could otherwise cause harmful algal blooms, and convert these into biomass that can be used as either animal feed, or biofuel feedstock (Xu et al., 2012). Recently the duckweed species *Spirodela polyrhiza* and *Lemna minor*, along with the marine grass *Zostera marina* have been fully sequenced, increasing scientific understanding of basal monocot evolution (Olsen et al., 2016; Van Hoek et al., 2015; Wang et al., 2014a,b). The Lemnaceae’s highly reduced morphology, preference for asexual budding, and unusual turion dormancy life-cycle permitted us to study plant evolution from a unique perspective.

Due to their roles in development and environmental adaptation, miRNA repertoire and regulation play a large part in understanding transcriptional modification. miRNAs are very short (19–24 nt) strands of RNA that guide the Agronaute (AGO) proteins in plants and animals to cleave mRNA targets, or in some cases inhibit translation (Allen et al., 2004; Lee et al., 2004; Palatnik et al., 2003; Reis et al., 2015). They are transcribed as pri-miRNAs, processed into...
hairpin precursors and subsequently into an miRNA: miRNA* duplex by the action of ribonuclease III-like enzymes known as Dicers, and more specifically, the Dicer-like-1 (DCL1) protein, in plants (Bernstein et al., 2001; Lee et al., 2004; Vazquez, 2006; Xie et al., 2004). In the cytoplasm, miRNA* (also called the passenger strand) is often, but not always degraded, whereas mature miRNA binds an Argonaute family protein to guide RISC to mRNA targets (Calviño et al., 2011; Carrington and Ambros, 2003; Hammond et al., 2000; Hutvagner and Simard, 2008; Noma et al., 2004; Xie et al., 2004). Previous studies have shown that miRNA sequences, targets, and their roles in development, or stress responses are highly conserved in plants, permitting the prediction of miRNAs and their target mRNAs, when a new genome has been sequenced (Berezikov et al., 2006).

As a resource for genomics and functional studies the Spirodela polyrhiza 7498 genome of 158 Mb with 19 623 gene models has previously been sequenced (Wang et al., 2014a,b), plus transcriptomic analysis of its fronds and developing turions has been performed (Wang et al., 2014b). Furthermore, fluorescently labeled bacterial artificial chromosomes (BACs) used for in situ hybridization of chromosomes facilitated the assembly of the ‘7498’ genome into 20 chromosomes (Cao et al., 2016). In total, 413 miRNA genes comprising 93 families were annotated in this sequence. Another ecotype, S. polyrhiza 9509, has also been sequenced and 59 conserved miRNAs from 22 families, 29 novel miRNAs, and 991 target mRNAs were predicted (Michael et al., 2017).

In this study, small RNA-seq experiments were performed with two ecotypes of S. polyrhiza, one with S. polyrhiza ‘LT5a’ under normal axenic conditions, and the other with strain ‘7498’ under eight conditions (control, 0°C, 37°C, abscisic acid, copper chloride, kinetin, potassium nitrate, and sucrose) in biological triplicate (three axenic flasks of clonal populations) to identify differentially expressed miRNAs (Figure 1). In addition, the cleaved fragments of uncapped mRNAs, commonly referred to as the degradome, of strain 7498 under the same eight conditions were sequenced in triplicate. The degradome libraries were used to experimentally validate miRNAs based on their cleavage of a target gene and assess post-transcriptional regulation under different growth conditions.

RESULTS
Generating and collecting miRNA candidates
In total 24 881 393 raw reads were obtained from the S. polyrhiza LT5a small RNA library, with 96.7% remaining after filtering out low quality reads and adapter sequences. The processed reads largely included sRNAs of 21 to 24-nt lengths – 21 nt (26.91%), followed by 22 nt (18.62%), 19–20 nt (14.23%), 23 nt (7.76%) and 24 nt (7.32%) (Figure 2a). Filtered reads were aligned with GenBank and Rfam, and after annotating the sRNAs as rRNAs, tRNAs, snoRNAs, snRNAs, known miRNAs, repeats, and finally mRNAs, there were 3.5 million reads matching known miRNAs and 7.6 million unannotated reads. Using exact matches to miRBase-release-20 (Kozomara and Griffiths-Jones, 2014)
for all plant species, we found 158 conserved miRNAs, and 24 miRNA* sequences. Unannotated sRNAs were mapped back to a Spirodela EST database, inspected in terms of hairpin structure (Figure 3a), and then analyzed by standard criteria (Allen et al., 2004; Schwab et al., 2005). This yielded 41 predicted novel miRNAs.

In a parallel experiment of strain ‘7498’, small RNA sequencing of three flasks of clonal populations serving as biological triplicates and eight conditions generated 32 279 643 clean, perfectly mapped reads. The most common read length was 22 nt, ranging from 60–84% of the sRNAs within the eight conditions, followed by 21 nt, which was between 4.5 and 18% of the sample, with all other lengths never exceeding 10% (Figure 2b). These sRNA reads were converted from SOLiD to FASTA format, combined with the reads from ‘LT5a’, and then all reads were run through miRPlant to identify candidate miRNAs, and hairpin structures in the genome (Figures 1 and 3b) (An et al., 2014). The mature sequences of the hairpins were aligned with conserved miRNA gene families and the ‘LT5a’ novel miRNAs, and the remainder reviewed. Criteria for novel miRNAs were set with a miRPlant score over 3, over 20 reads, and 1 miRNA* per read. The predicted hairpins were further evaluated based on the revised criteria for plant miRNA annotation (Tables S1 and S2) (Axtell and Meyers, 2018).

The two highly overlapped sets of miRNAs were aligned to the ‘7498’ genome for consolidation, thereby providing 134 mapped miRNAs, with mature sequence length derived from the most common sRNA reads. These were then compared with the dataset of the ecotype ‘9509’ (Michael et al., 2017). Twenty conserved families were shared between the two ecotypes. All but one of the novel miRNAs were ‘9509’ specific. After reconciling all Spirodela datasets, we checked other species, and each conserved miRNA family found in Arabidopsis and Oryza sativa, which was undetected in Spirodela, was represented by one family member, thereby providing 76 non-Spirodela families to analyze. Therefore, we assembled 220 predicted miRNAs from several sources (Figure 1 and Table S1), and mapped these to the Spirodela polyrhiza ‘7498’ genome, requiring a perfect alignment, and then again allowing one mismatch. This generated 134 aligned miRNAs, only nine of which had a mismatch (Table S3). These were all novel miRNAs for Spirodela, except for osa-miR3979-5p, the only conserved non-Spirodela miRNA that mapped to the Spirodela genome. Certain predicted novel miRNAs and the conserved miRNAs miR395 and miR399 could not be aligned with the Spirodela 7498 genome, possibly due to use of different genome assemblies and analysis methods.

Abundance profiles of known and novel miRNAs

In Spirodela ecotype LT5a under normal growth conditions, miRNA abundances were dominated by families of miR156 (47% of total abundances), miR166 (24%), miR167 (20%) miR168 (5%) and miR169 (1%) (Figure 4a). In strain ‘7498’ sRNA sequencing showed a highly abundant and variable sequence for miR169c, which comprised 33–82% of the reads in each condition (Figure 5a). Given the abundance of this sequence compared with the ‘LT5a’ and

Figure 2. Length distributions of small RNAs from Spirodela polyrhiza. (a) Length distribution in strain LT5a; (b) Length distributions in strain 7498. The error bars indicate the highest and lowest average percentage seen between the eight conditions sequenced.
'9509' results, it was measured along with two other abundant and stably expressed miRNAs miR156d and miR396b via qPCR, confirming its high expression and variability, although the miR169c:miR156d ratio under sucrose conditions was 25 instead of 70 seen in the sequencing data (Figure 5a, b). After the miR169 family, the miR396, miR156, miR168, and miR166 families were the most abundant, with all but miR396 being among the five most abundant in both strains (Figure 4b) (Table S4). There were 15 differentially expressed miRNAs, with four from the miR156 family, and six novel miRNAs (Table 1). Of the 13 instances of differential expression within one condition, only two were classified as overexpressed compared with the control (spo-miRnovel21-3p in abscisic acid (ABA) and miR396a in 37°C), and 13 of the 19 differentially expressed miRNAs were found under sucrose conditions (Table 1). There were also 28 miRNAs, including 20 novel ones, present under stimuli conditions, yet entirely absent under normal growth conditions, probably functioning in stress and hormone responses (Table S4).

Prediction of mRNA targets of Spirodela miRNAs

Predicted targets for the miRNAs were identified with the psRNA target program (Dai and Zhao, 2011). For the 220 predicted miRNAs, we found 205 predicted interactions between 53 miRNAs and 162 targets, (Table S5). Although the possibility of a plant miRNA–target interaction is predictable, the likelihood that a given interaction will occur is subject to miRNA and target co-expression, mRNA secondary structure, and other factors (Dai and Zhao, 2011; Kidner and Martienssen, 2004; Voinnet, 2009). However, an acceptable score from a good prediction program such as psRNA target still has a 50% false-positive rate, requiring the need for degradome-based verification of these predicted targets (Dai et al., 2011).

Degradome sequencing

GMUCT 2.0 (Genome-wide Mapping of UnCleaved Transcripts) sequencing of 24 libraries (eight conditions and three biological replicates) yielded 910 758 343 raw reads, which were mapped to the Spirodela genome and used to find GMUCT-supported targets of miRNAs (Figure 1). The mapping of reads, alignment to predicted miRNAs, and discovery of GMUCT-supported targets from genic and intergenic regions was performed using sRNA PARE Target Analyzer (sPARTA) (Kakrana et al., 2014). For the 220 miRNA candidates, sPARTA yielded 397 genic and 159 intergenic interactions (Table S6 and S7). An interaction in this dataset is documented with an miRNA, target gene, number of reads aligned with the cleavage site, the library of origin, and other information. The intergenic dataset revealed 96 unique loci that would have been
transcribed, poly-adenylated, and precisely cleaved by an miRNA, suggesting they are either unannotated genes or long non-coding RNAs (Table S7) (Kakrana et al., 2014).

Of the 220 miRNA candidates reviewed only 66 were validated by sPARTA with an average of 6.8 and 4.6 targets for conserved and novel miRNAs, respectively (Table S8). Of those 66 miRNAs, 42 were widely conserved, 12 were not found in the *Spirodela* genome, and 12 were novel. Twelve of the 76 conserved miRNAs from *Arabidopsis* and rice not sequenced in the *Spirodela* had degradome support, with 11 sequenced in only one condition (Table S9).

The novel miRNA predictions had an average verification rate of 31%, consistent with the approximately 30% validation rate found in similar studies (Li et al., 2010; Song et al., 2010; Yang et al., 2013).

Of the 220 miRNAs analyzed in the psRNA target, 53 were predicted to have targets in the genome and 32 of those were verified in sPARTA. These sPARTA verified
miRNAs had 120 predicted targets, 20 of which were confirmed in the degradome (Figure 6a and Tables S2, S5 and S10). A closer look at prediction verification rates revealed that the largest factor was not the interaction score from psRNA target, but whether the miRNA was conserved or a novel prediction (Figure 6b). Use of eight different conditions for the degradome enabled us to analyze unique targets within those conditions. Only 15 of the 144 target genes were sequenced in four or more conditions and regulated by six conserved miRNA families, and three novel miRNAs, whereas 71% of the targets were condition specific, whereas 71% of the total, respectively, whereas 24-nt sRNAs were only 7.3% and 1% (Figure 2). The difference in size distribution between the strains may have been due to the PCR amplification of the sRNAs from ‘7498’ before sequencing. This difference is consistent with previous studies of sRNAs in *Brassica juncea*, *Ginkgo biloba*, *Physcomitrella patens*, *Taxus chinensis*, and seven species of conifers (Dolgosheina et al., 2008; Fattash et al., 2007; Qiu et al., 2009; Yang et al., 2013). By contrast, the 24-nt sRNA class was quite abundant in other species, *Arabidopsis thaliana*, *Avicennia marina*, *Citrus trifoliata*, *Medicago truncatula* and *Oryza sativa* (Fahlsgren et al., 2007; Khraiwesh et al., 2013; Song et al., 2010; Sunkar et al., 2008; Szittya et al., 2008). The 24-nt sRNA class mainly comprised small interfering RNAs (siRNA), which are known to guide DNA methylation and heterochromatin formation of repetitive and transposable elements in angiosperms (Mette et al., 2000; Vazquez, 2006). In *Spirodela*, many highly conserved genes involved in DNA methylation and transposon silencing are mutated or transcriptionally silenced, resulting in extremely scarce DNA methylation and heterochromatin formation (A. Harkess, A.J. Bewick, P. Fourounjian, Z. Lu, B.T. Hofmeister, J. Messing, T. Michael, R.J. Schmitz and B.C. Meyers, manuscript in preparation). Although the genome of *S. polyrhiza* has a repeat content consistent with its size it has, in contrast to other sequenced angiosperm genomes, no recent retrotranspositions, suggesting that the function of 24-nt siRNAs to minimize the activities of repetitive and transposable elements may be spatiotemporally specific or that other mechanisms are keeping them in check (Wang et al., 2014a,b). A. Harkess, A.J. Bewick, P. Fourounjian, Z. Lu, B.T. Hofmeister, J. Messing, T. Michael, R.J. Schmitz and B.C. Meyers, manuscript in preparation).

### DISCUSSION

**sRNA size distribution**

*Spirodela ‘LT5a’* and ‘7498’ exhibited an abundance of miRNA sized 20- to 22-nt sRNA comprising 60% and 89% of the total, respectively, whereas 24-nt sRNAs were only 7.3% and 1% (Figure 2). The difference in size distribution between the strains may have been due to the PCR amplification of the sRNAs from ‘7498’ before sequencing. This size distribution is consistent with previous studies of sRNAs in *Brassica juncea*, *Ginkgo biloba*, *Physcomitrella patens*, *Taxus chinensis*, and seven species of conifers (Dolgosheina et al., 2008; Fattash et al., 2007; Qiu et al., 2009; Yang et al., 2013). By contrast, the 24-nt sRNA class was quite abundant in other species, *Arabidopsis thaliana*, *Avicennia marina*, *Citrus trifoliata*, *Medicago truncatula* and *Oryza sativa* (Fahlsgren et al., 2007; Khraiwesh et al., 2013; Song et al., 2010; Sunkar et al., 2008; Szittya et al., 2008). The 24-nt sRNA class mainly comprised small interfering RNAs (siRNA), which are known to guide DNA methylation and heterochromatin formation of repetitive and transposable elements in angiosperms (Mette et al., 2000; Vazquez, 2006). In *Spirodela*, many highly conserved genes involved in DNA methylation and transposon silencing are mutated or transcriptionally silenced, resulting in extremely scarce DNA methylation and heterochromatin formation (A. Harkess, A.J. Bewick, P. Fourounjian, Z. Lu, B.T. Hofmeister, J. Messing, T. Michael, R.J. Schmitz and B.C. Meyers, manuscript in preparation). Although the genome of *S. polyrhiza* has a repeat content consistent with its size it has, in contrast to other sequenced angiosperm genomes, no recent retrotranspositions, suggesting that the function of 24-nt siRNAs to minimize the activities of repetitive and transposable elements may be spatiotemporally specific or that other mechanisms are keeping them in check (Wang et al., 2014a,b). A. Harkess, A.J. Bewick, P. Fourounjian, Z. Lu, B.T. Hofmeister, J. Messing, T. Michael, R.J. Schmitz and B.C. Meyers, manuscript in preparation).

### Table 1 Differentially expressed miRNAs

<table>
<thead>
<tr>
<th>Condition</th>
<th>miRNA name</th>
<th>P-value adjusted</th>
<th>RPM in control</th>
<th>Relative expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABA</td>
<td>spo-miR156d-3p</td>
<td>0.03886</td>
<td>5900.0</td>
<td>15%</td>
</tr>
<tr>
<td>ABA</td>
<td>spo-miR171</td>
<td>0.04690</td>
<td>4919.6</td>
<td>16%</td>
</tr>
<tr>
<td>ABA</td>
<td>spo-miRnovel14-5p</td>
<td>0.03886</td>
<td>173.1</td>
<td>5%</td>
</tr>
<tr>
<td>ABA</td>
<td>spo-miRnovel21-3p</td>
<td>0.03886</td>
<td>68.7</td>
<td>480%</td>
</tr>
<tr>
<td>ABA</td>
<td>spo-miRnovel48-3p</td>
<td>0.00022</td>
<td>2451.4</td>
<td>7%</td>
</tr>
<tr>
<td>37°C</td>
<td>spo-miR396a</td>
<td>0.00925</td>
<td>30.3</td>
<td>698%</td>
</tr>
<tr>
<td>Sucrose</td>
<td>spo-miR156c</td>
<td>0.00114</td>
<td>2819.7</td>
<td>4%</td>
</tr>
<tr>
<td>Sucrose</td>
<td>spo-miR156c-3p</td>
<td>0.00002</td>
<td>198.6</td>
<td>1%</td>
</tr>
<tr>
<td>Sucrose</td>
<td>spo-miR156d-3p</td>
<td>0.00012</td>
<td>5900.0</td>
<td>6%</td>
</tr>
<tr>
<td>Sucrose</td>
<td>spo-miR171</td>
<td>0.00013</td>
<td>4919.6</td>
<td>8%</td>
</tr>
<tr>
<td>Sucrose</td>
<td>spo-miR224</td>
<td>0.00803</td>
<td>37.0</td>
<td>0%</td>
</tr>
<tr>
<td>Sucrose</td>
<td>spo-miR396b-3p</td>
<td>0.04752</td>
<td>254.9</td>
<td>5%</td>
</tr>
<tr>
<td>Sucrose</td>
<td>spo-miR396c</td>
<td>0.00658</td>
<td>383.0</td>
<td>14%</td>
</tr>
<tr>
<td>Sucrose</td>
<td>spo-miR8005</td>
<td>0.04052</td>
<td>104.5</td>
<td>0%</td>
</tr>
<tr>
<td>Sucrose</td>
<td>spo-miRnovel3-5p</td>
<td>0.01754</td>
<td>72.2</td>
<td>2%</td>
</tr>
<tr>
<td>Sucrose</td>
<td>spo-miRnovel14-5p</td>
<td>0.00114</td>
<td>173.1</td>
<td>5%</td>
</tr>
<tr>
<td>Sucrose</td>
<td>spo-miRnovel37</td>
<td>0.01232</td>
<td>554.0</td>
<td>11%</td>
</tr>
<tr>
<td>Sucrose</td>
<td>spo-miRnovel40</td>
<td>0.00800</td>
<td>85.4</td>
<td>0%</td>
</tr>
<tr>
<td>Sucrose</td>
<td>spo-miRnovel48-3p</td>
<td>0.00004</td>
<td>2451.4</td>
<td>5%</td>
</tr>
</tbody>
</table>

Expression under control conditions was measured as reads per million miRNA reads (Table S4). All genes required an adjusted $P < 0.05$ from the DESeq2 program to be considered differentially expressed. ABA, abscisic acid.
and the Chinese yew, whereas the other group consists only of angiosperms. Therefore, the connection between the 24-nt siRNA-mediated regulatory pathway, the evolutionary relations between gymnosperms and angiosperms, and alternative transposon control mechanisms could be an interesting subject for future studies.

### Assembling and verifying miRNA candidates

In this study, filtered reads from *Spirodela* ‘LT5a’ were aligned to known non-coding RNAs and analyzed using Mireap and Mfold to identify 158 conserved and 34 novel miRNAs (Figure 1). In addition to the standard alignment, 32 million filtered reads from ecotype 7498 grown under eight conditions were analyzed in miRPlant to detect 35 conserved and nine novel miRNAs. A previous study of a different ecotype of *Spirodela* discovered two conserved miRNAs and 29 novel miRNAs, which were included in our analysis (Michael et al., 2017). Finally, 76 representatives of conserved miRNA families in *Arabidopsis* and rice in the miRBase-release-21 (Kozomara and Griffiths-Jones, 2014) that were not expressed under the tested conditions were added. These miRNAs were aligned to the *S. polyrhiza* ‘7498’ genome to remove repeats, establish length, and organize conserved miRNAs, which yielded 134 aligned miRNAs and 47 hairpin structures (Tables S2 and S3). These aligned miRNAs were obtained from the three *Spirodela* sequencing experiments with the exception of osa-3979-5p, which was found in three different degradome libraries with three different targets (Tables S6 and S9).

The small RNA sequencing experiments revealed that the miR156, miR166, miR168, and miR169 families were among the most abundant in both *Spirodela* ‘LT5a’ and ‘7498’ under control conditions (Figure 4). These highly conserved families functioned in maintenance of a juvenile state, leaf vertical polarity, the miRNA binding Argonaute-1 protein, and nuclear transcription factor Y related to stomatal closure, respectively. Among 19 instances of differential expression, 13 occurred under sucrose conditions, commonly used when growing Lemnaceae species in a laboratory (Tables 1 and 2). All 19 of the miRNAs absent under control condition were rare miRNAs with an average expression under 10 RPM (Table S4).

To interpret the function of the miRNAs, these were analyzed via psRNA target to identify 205 predicted target miRNAs (Table S5). These miRNAs were then utilized in a degradome sequencing experiment with *Spirodela* 7498 grown under eight conditions (Figure 1). The GMUCT 2.0 library construction protocol requires isolation of uncapped mRNAs through two poly-A selections and a 5’ adapter ligation. After sequencing and mapping cleaved miRNAs the sPARTA program aligns predicted miRNAs to the genome to identify cleavage sites. For filtered genic reads we chose a stringent complementarity filter of three, with a minimum of 10 aligned reads comprising 75% or more of the total nearby reads, and a corrected $P < 0.05$. Applying these parameters, sequencing of cleaved targets validated 171 interactions of 42 conserved miRNAs with 83 targets and 24 novel miRNAs regulating 66 targets (Table S8). Although 22 interactions were found between different conserved miRNA family members and the same gene, there was no overlap between novel miRNAs.

The degradome sequencing and sPARTA validation proved highly valuable in confirming predicted miRNAs and targets. Novel miRNAs are difficult to predict, and often hard to verify due to their low expression in specific tissues paired with a smaller number of less abundant targets compared with conserved miRNAs, leading to a near 30% degradome validation rate (Li et al., 2010; Song et al., 2010; Yang et al., 2013). This makes degradome sequencing effective at accurately predicting novel miRNAs. When predicting mRNA targets, 78% of the unverified predictions come from novel miRNAs, while those from conserved miRNAs had a 40% validation rate (Figure 6b and Table S10).

It is important to note that many interactions not supported in this study may await future degradome validation in an experiment, when both the miRNA and target
are spatiotemporally expressed, the RISC induces cleavage instead of inhibiting translation, and the degradome reads are abundant enough to withstand the data filtering steps (Dai et al., 2011; Kakrana et al., 2014; Kidner and Martienssen, 2004; Voinnet, 2009). Given the stringency of degradome interactions and their differences from predicted datasets, the massively parallel signature sequencing databases at mpss.danforthcenter.org clearly distinguished degradome results from predictions (Kakrana et al., 2014; Nakano et al., 2006), whereas the miRTarBase shows the experimental evidence behind miRNA target interactions, allowing the user to judge their confidence (Chou et al., 2016; Hsu et al., 2011, 2014).

**miR156 and miR172 families in Spirodela**

More than half of the conserved miRNA families (13 out of 24) were represented by more than 1000 reads in ‘LT5a’ (Figure 4). Among these, the miR156, miR166, miR167 and miR168 families were the most abundant. This accumulation profile is consistent with equally pooled tissues from other plant species. The highest expression levels were achieved with miR156, miR166 and miR168 in peanut (Zhao et al., 2010), and miR166, miR319 and miR396 in mangrove (Khraiwesh et al., 2013).

The maintenance of juvenile characteristics is known for expression of the miR156 family and suppression of miR172 family, which gradually change their antagonistic abundances during plant development (Wu et al., 2009). In Arabidopsis and maize, overexpression of miR156 results in prolonged expression of juvenile characteristics and extremely delayed flowering, affecting the phase transition from vegetative growth to reproductive growth and the rise in miR172 expression levels (Aukerman, 2003; Chuck et al., 2007; Schwab et al., 2005). Interestingly, miR156 is highly expressed in *Spirodela* LT5a with 633,065 reads, more than 62-fold that of miR172 (10,101 reads) (Figure 4). Given that, under regular conditions, duckweed undergoes predominantly clonal propagation without flowering, we suggested that the extremely high ratio of miR156:miR172 plays a key role in maintaining such a vegetative state of growth and reproduction. In strain ‘7498’ we see that the four miR156 family members are 171 times more abundant than four miR172 members under normal growth conditions, whereas four of the other growth conditions exhibited a lower ratio (Figure 7a).

Although the miR156:miR172 expression ratio ranged from 408 in kinetin exposure to 71 under heavy metal stress, it does not reflect the degree of target mRNA cleavage activity (Figure 7a, b). When measuring the number of filtered cleaved mRNA targets within three replicate libraries, the comparatively scarce miR172 is more active under three of eight conditions (heat, copper, and sucrose exposure) (Figure 7b). By contrast the high abundance, but relative low *trans*-acting activity of the miR156 family could be explained by the lack of transcription of some of its squamosa promoter-like family targets. Similar to the miR156 family, we see that the extremely abundant spo-miR169c sequence only cleaved one target in three different conditions, and that the two most abundant novel miRNAs were only found to be active in one condition (Tables S4 and S9). Overall, these results demonstrated that degradome sequencing not only allowed *in vivo* validation of miRNAs but also the comparison of miRNA activity to correctly perceive the post-transcriptional response to environmental adaptation.

**Common and condition-specific cleaved targets**

Considering that miRNAs function in plant development and environmental adaptation, various growth conditions were used for sequencing libraries to broadly survey post-transcriptional regulation. There were 15 commonly expressed targets cleaved within four or more growth conditions. These targets were often transcription factors regulated by conserved microRNA families such as miR156, miR172, and miR396 (Table S9). Spo-miRnovel21, Spo-miRnovel23-5p, and Spo-miRnovel67 were also found to regulate a mix of transcription factors, other, and unknown proteins under four or more growth conditions, suggesting important functions under multiple environmental adaptations.

Although small RNA sequencing yielded a few condition-specific miRNAs, 71% of cleaved targets identified in the degradome were condition-specific (Tables S4 and S9). Although these targets included several transcription factors from conserved families, the bulk of targets were a variety of metabolic enzymes and unknown proteins, cleaved mainly by novel miRNAs and well conserved miRNA families (Table S9). Of the 66 miRNAs supported by degradome sequencing only 24 were found under control condition, and 35 were condition-specific (Table S9). Surprisingly, whereas spo-miRnovel32 and spo-miRnovel41 comprised between 66 and 93% of the total novel miRNA expression depending on the condition, the only degradome interaction seen for these was spo-miRnovel32 cleaving miRNAs, encoding a DNA repair protein under normal growth conditions, indicating that miRNA expression and activity are not necessarily correlated.

The heat stress, copper, and sucrose growth conditions yielded the highest number of specific mRNA targets, with 15, 18, and 19, respectively (Tables 2 and S9). Under the mixotrophic condition, which also had the most differentially expressed miRNAs, specific mRNA targets encoded a mix of metabolic, signaling, and unknown proteins, cleaved by well conserved miRNAs and eight different novel miRNAs. This change in post-transcriptional regulation is of importance when extrapolating laboratory results to duckweed growing outdoors because most laboratories use sucrose in their growth media. Copper stress caused
the cleavage of mRNAs encoding two polyphenol oxidases, reminiscent of the decrease in protein levels observed with non-tolerant Agrostis capilaris (Hego et al., 2016). The mRNA encoding respiratory burst oxidase homolog protein E was also cleaved, demonstrating a strong enrichment for ROS regulation under this growth condition. The heat/stress condition induced the degradation of several energy generating proteins and other enzymes. Finally, kinetin exposure induced the highest number of condition-specific putative transcription factors, with five of these cleaved by four conserved miRNA families (miR156, miR159, miR160, and miR394), supporting the link between cytokinin exposure and differential expression of transcription factors (Brenner et al., 2005; Heyl et al., 2008).

CONCLUSION

In summary, this work predicted 108 miRNAs with 140 targets through small RNA sequencing experiments of two Spirodela ecotypes, included previously predicted and identified miRNAs, and validated 66 miRNAs and 149 targets by degradome sequencing. Based on this support and stringent miRNA hairpin criteria 35 miRNAs were added to miRBase (Table S2). A catalog of 171 verified and 404 lower confidence miRNA-target interactions was established (Table S11). Viewing the network in Cytoscape shows large numbers of targets for several miRNAs (Figure 8) (Shannon et al., 2003). The study of seven different hormonal and stress responses, facilitated by a rapidly growing aquatic plant, provided us with a broad aspect of the diverse roles of miRNAs, on which future work on environmental adaptation can be built.

EXPERIMENTAL PROCEDURES

Methods

In brief, sRNA sequencing and miRNA prediction was performed in Spirodela polyrhiza ‘LT5a’ under control conditions and ‘7498’ under eight conditions (Figure 1). After miRNA prediction, miRNAs from ‘9509’, Arabidopsis and rice were added, and these were mapped to the ‘7498’ genome, analyzed for hairpin loci, measured for expression patterns, and predicted targets. Then the degradome of ‘7498’ in the same eight conditions was sequenced to verify miRNAs by cleavage activity in total and across the conditions.

‘LT5a’ small RNA experiment

Growth conditions. Spirodela polyrhiza LT5a collected from Lake Tai, China in a previous study (Tang et al., 2014) was cultured in Schenk and Hildebrandt liquid medium at pH 5.8 and grown with 16 h of light 100 μmol m⁻² sec⁻¹, light intensity at 23°C and 8 h of dark (15°C). Whole plants were collected at 1 day, 3 days and 5 days after inoculation during the light period. All samples were immediately frozen in liquid nitrogen and stored at −80°C.

RNA extraction. Total RNA was extracted from the frozen samples using TRIzol reagent (Invitrogen, Carlsbad, CA, USA)
following the manufacturer’s instructions, and pooled in an equal fraction ratio. Small RNAs (sRNAs) of 18–30 nt in length were size fractionated using 15% denaturing polyacrylamide gel electrophoresis. After purification, 5' and 3' adaptors were ligated to the sRNAs using T4 RNA ligase (Promega, Madison, WI), which were again purified and used as templates for reverse-transcription (RT) reaction. The cDNA was further purified by 15% denaturing polyacrylamide gel electrophoresis. Finally, the cDNA library was sequenced on an Illumina sequencer (HiSeq 2000, Illumina) by the Beijing Genomics Institute (BGI, Shenzhen, China).

Data processing and annotation of sRNAs. The raw reads were filtered to remove low quality reads based on SQ value, size, and adaptor presence. The remaining reads (clean reads) were used for further analyses. Note that the following analyses were performed first before the Spirodela genome was available. After the filtering steps, sRNA clean reads were first aligned against the sequences of non-coding RNA (rRNA, tRNA, snRNA, and snoRNA, etc.) deposited in the GenBank non-coding RNA database and Rfam database (Rfam 11.0) (Burge et al., 2012, https://www.sanger.ac.uk/science/tools https://www.ncbi.nlm.nih.gov/gds/?term=Spirodela%20miRNA). Matched reads were eliminated from further analyses. A BLASTN search was performed on each unique sequence of remaining reads against known mature and precursor miRNAs (pre-miRNAs) from all plant species deposited in the miRBase database (Release 20) (Griffiths-Jones et al., 2007, 2006; Griffiths-Jones, 2006; Kozomara and Griffiths-Jones, 2014, 2011). Only perfectly matched reads were classified as conserved miRNAs. Finally, the remaining sRNAs were sequentially aligned to repeat-associated RNAs, exon/intron of mRNA (both in-house databases at BGI) to detect degraded fragments of either repetitive elements or mRNA. During the annotation process, each unique sRNA was annotated only once based on the following priority rules: rRNA etc. (GenBank > Rfam) > known miRNA > repeat > exon > intron. The rest of the sRNAs could not be matched to any databases and were grouped into ‘unannotated sRNAs’.

Prediction of novel miRNA. The fact that pre-miRNAs have characteristic fold-back structure was used to predict novel miRNAs in Spirodela. The unannotated sRNAs were subjected to a BLAST homology search against a S. polyrhiza EST library (accession: SRX148325) from NCBI for precursor sequences. The surrounding sequences of each matched EST sequence were extracted, and then run through structural analysis to identify novel miRNA hairpins using the Mireap program developed by BGI (http://sourceforge.net/projects/mireap/) according to published criteria (Chen et al., 2012). The secondary structures of these putative pre-miRNAs were validated by Mfold (Zuker, 2003) (Figure 3a), and only structures with the lowest folding energy were selected.

‘7498’ small RNA-seq experiment

Growth conditions. The conditions were those used to generate the EST database based on the Spirodela polyrhiza ‘7498’ genome (Wang et al., 2014a,b). Heat treatment at 37°C, cold treatment...
at 0°C, 20 mg/L CuCl₂, 300 mg/L KNO₃, 250 nm ABA, 10 nm kinetin, and 1% sucrose. To create each biological triplicate, five fronds were placed in a sterile flask in 100 mL half-strength Schenk and Hildebrandt medium adjusted by KOH to pH 5.8, and grown at 24°C under the environmental conditions of 16 h of light, until they covered the surface of the water. Cold treatment was carried out overnight, and copper treatment lasted 4 days before harvest. All other variables were present from flask inoculation until harvest. Upon harvest, all samples were immediately frozen in liquid nitrogen, and stored at −80°C until RNA extraction.

Small RNA extraction. miRNA was extracted using the Ambion mirVana™ miRNA Isolation Kit (cat no: AM1560 Life Sciences) according to the manufacturer’s instructions. Small RNA (10-40 nt) was extracted in RNase free water, and run on the Agilent® 2100 BioAnalyzer™ instrument with the small RNA Chip, and visualized using 2100 expert software BioAnalyzer.

Library construction. For the library construction, the SOLiD® Total RNA-seq Kit (Part Number 4452437) was used, to ligate adaptors, and reverse transcribe the library. cDNA purification was accomplished using the MinElute® PCR Purification Kit (Qiagen). InInvitrogen Novex® pre-cast gels (Invitrogen Novex® 10% TBE-Urea Gel 1.0 mm, 10 Well). The cDNA was amplified with SOLiD™ 5’ and 3’ PCR Primers. After amplification, the PCR product was purified using the Invitrogen PureLink® PCR Micro Kit. The yield and size distribution of the amplified DNA was assessed using the Agilent 2100 BioAnalyzer™ Instrument and the DNA 1000 Kit (Agilent®). Barcoded libraries were prepared using SOLiD™ RNA Barcoding Kit-Modules 17–32 (cat no. 4453189). Following barcoding, each library template was clonally amplified on SOLiD™ P1 DNA Beads by emulsion PCR using the Applied Biosystems SOLiD™ 4 System Templated Bead Preparation (Part. no. 4448378). Finally, sequencing was carried out on the SOLiD 5500 instrument. Although this instrument no longer in use, it was one of the most accurate machines available at the time of the sequencing and is useful for short-read sequencing.

miRPlant. miRPlant, uses a plant genome, index file, guide miRNA annotation file, and small RNA sequencing data to identify and score conserved and novel miRNA hairpins (An et al., 2014) (Figure 3b). The 32 279 643 perfectly mapped small RNA-seq reads from strain ‘7498’ were combined with the 24 881 393 reads from ‘LT5a’, and aligned against Brachypodium distachyon non-coding RNA sequences with miRNAs removed to filter out ribosomal, transfer, and other RNA sequences (ensemblegenomes.org). All miRNAs, along with the psRNA target. All miRNAs, along with the Spirodela genome sequence were uploaded into the psRNA target (Dai and Zhao, 2011). With this program, we were able to predict targets of the aligned miRNAs using the suggested confidence score of 2.0 and default target accessibility cutoffs.

‘7498’ degradome sequencing experiment

Growth conditions. Growth conditions were identical to those used in the ‘7498’ sRNA-seq experiment, with two notable changes. Each flask was inoculated with four frond clusters, grown for 2 weeks, and exposed to the specified condition. After a literature review of each stimuli, the duration of each variable was adjusted to capture a greater diversity of differently cleaved targets. The stimuli included cold treatment at 0°C as an ice water bath for 24 h, heat treatment at 37°C for 1 h, 250 nm ABA for 3 h, 20 mg/L CuCl₂ for 24 h, 10 mM kinetin for 3 h, 300 mg/L KNO₃ for 2 h, and the addition of 1% sucrose before inoculation. As before, all samples were harvested in a bacteriological hood, frozen in liquid nitrogen, and stored at −80°C until RNA extraction.

RNA extraction. Total RNA was extracted using the Plant RNAeasy Qiagen kit, using a modified protocol. RNA was kept in a chilled block. DNase digestion was skipped, as DNA would not be amplified in library construction, and the protocol shortened to about 25 min to minimize the degradation at the 5’ end of cleaved transcripts.

GMUCT 2.0 library construction. The following protocol was modified from Willmann et al. (2014) RNA quality/quantity were evaluated using the RNA Nano Plant Assay on BioAnalyzer.
Libraries were only constructed with 30 μg of total RNA, with a RNA Integrity Number (RIN) of 7.5 or greater and clearly defined 28s rRNA peak greater in height than 18s rRNA. miRNA was selected using Dynabeads oligo-oligo (Ambion). At least 300 ng of mRNA with less than 10% rRNA was used according to the miRNA assay on Bioanalyzer (Ambilent). 5' ligation with the small RNA 5' adapter (RA5) was performed as previously described (Willmann et al., 2014). A second poly-A selection was performed using 100 μL Dynabeads oligo(oligo) (Ambion). The GMUCT 2.0 cDNA was synthesized using a primer with a sequence CTGGAGTTCCTTGACCCGAGAATTCCANNNNNN and SuperScript III (IDT DNA, Thermo Fisher). After synthesis, excess primer and dimers were removed using 1.8× AMPure XP (Beckman Coulter). Libraries were amplified using Small RNA Index Primers RPI and RPIIX (Illumina) for no more than 15 cycles amplification. Final libraries were size selected using double sided SPRI selection with AMPure XP, with 0.3×/0.7× volumes of AMPure to remove fragments over 700 bp and under 300 bp. Library quantity and quality were evaluated using High Sensitivity DNA assay for Bioanalyzer, and Qubit Fluorimeter (Agilent, Thermo Fisher). Libraries were made, diluted to 2 μl, and used for denaturation, loading, and running on the Illumina NextSeq 500, using High Output 150 Cycle kits (Illumina). Read length was set to 160 bp. All libraries were uploaded to the Danforth mpss database and Myers lab viewer, in which results were searchable by miRNA or target (Nakano et al., 2006).

**Prediction of GMUCT-supported targets of miRNAs.** To find miRNA targets supported by experimental data such as from degradome, GMUCT and PARE sequencing, sPARTA software was applied. It integrates genome annotations, small RNA sequencing and GMUCT sequencing data to identify miRNA-mediated cleavage of targets at whole genome level (Kakrana et al., 2014). For this analysis, we first expanded the *Spirodela polyrhiza* ‘7498’ gene annotations – 150 nt upstream of start codons and 250 nt downstream of stop codons, to compensate for unannotated UTR regions using the ‘StopBed’ program in the Galaxy wrapper (Quinlan and Hall, 2010). Next, sPARTA analysis was run as described (Arikitt et al., 2014) with a few exceptions: (a) stringent complementarity-based score cutoff (score ≈ 3.0) was used; (b) for gene targets a minimum of 10 degradome reads were required along with a window ratio of 0.75 or more; (c) for targets from intergenic or unannotated regions a minimum of 20 degradome reads supporting the cleavage were required along with a window ratio of 0.80 or more. Finally, these filtered targets were manually checked in Integrative Genomics Viewer (IGV) (Thorvaldsdottir et al., 2013) along with the genome annotations and GMUCT read abundances (Figure 3c).

**ACKNOWLEDGEMENTS**

We would like to thank Ying Zhang, Ying-Ying Zhang, and Dibyendu Kumar for their assistance with data analysis, and Mayumi Nakano and Deepthi Ramachandruni for database construction.

**DECLARATIONS**

Ethics approval and consent to participate.

Not applicable.

**CONSENT FOR PUBLICATION**

Not applicable.

**DATA ACCESSIBILITY**


**FUNDING INFORMATION**

Funding was provided by the Selman Waksman Chair in Molecular Genetics.

**AUTHOR CONTRIBUTIONS**

BT, JT and PF prepared biomass and extracted RNA, BT and BG assembled and sequenced libraries, JT, YF, BG, AK, MT, CW and PF analyzed the data, JT and PF wrote the manuscript as co-first authors, BM, JM and JM guided the project.

**CONFLICT OF INTEREST**

The authors declare no competing interests.

**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article.

Table S1. 220 conserved and novel miRNAs.
Table S2. miRNA hairpin structures.
Table S3. Genomic loci of 134 of the 220 predicted miRNAs.
Table S4. miRNA expression as RPM throughout eight conditions and three replicate libraries.
Table S5. mRNA targets as predicted by psRNA target.
Table S6. Filtered genic sPARTA results.
Table S7. Filtered intergenic sPARTA results.
Table S8. Unique miRNA:target pairs in genic, filtered sPARTA results.
Table S9. Target cleavage distribution throughout eight conditions.
Table S10. psRNA and sPARTA target mRNA overlap.
Table S11. A final catalogue of 575 miRNA target pairs of varying confidence.

**REFERENCES**


ate orange (Citrus trifoliata). BMC Genom., 11, 431.

ing. BMC Plant Biol. 8, 25.


Willmann, M.R., Berkowitz, N.D. and Gregory, B.D. (2014) Improved genome-wide mapping of uncapped and cleaved transcripts in eukaryotes- 
GMUCT 2.0. Methods, 87, 84–73.


Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilber- 
652.


Zhao, C.-Z., Xia, H., Frazier, T., Yao, Y.-Y., Bi, Y.-P., Li, A.-Q., Li, M.-J., Li, C.- 

prediction. Nucleic Acids Res. 31, 3406-3415.