the plant journal

SEB Society for Experimental Biology

The Plant Journal (2019) doi: 10.1111/tpj.14294

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

Paul Fourounjian[1,†], Jie Tang[2,†], Bahattin Tanyolac[1], Yaping Feng[1], Brian Gelfand[1], Atul Kakrana[3], Min Tu[1], Chris Wakim[1], Blake C. Meyers[3], Jiong Ma[4] and Joachim Messing[1,*]

[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.

*For correspondence (e-mail messing@waksman.rutgers.edu).

[†]These authors contributed equally to this work.

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 Hoeck *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)

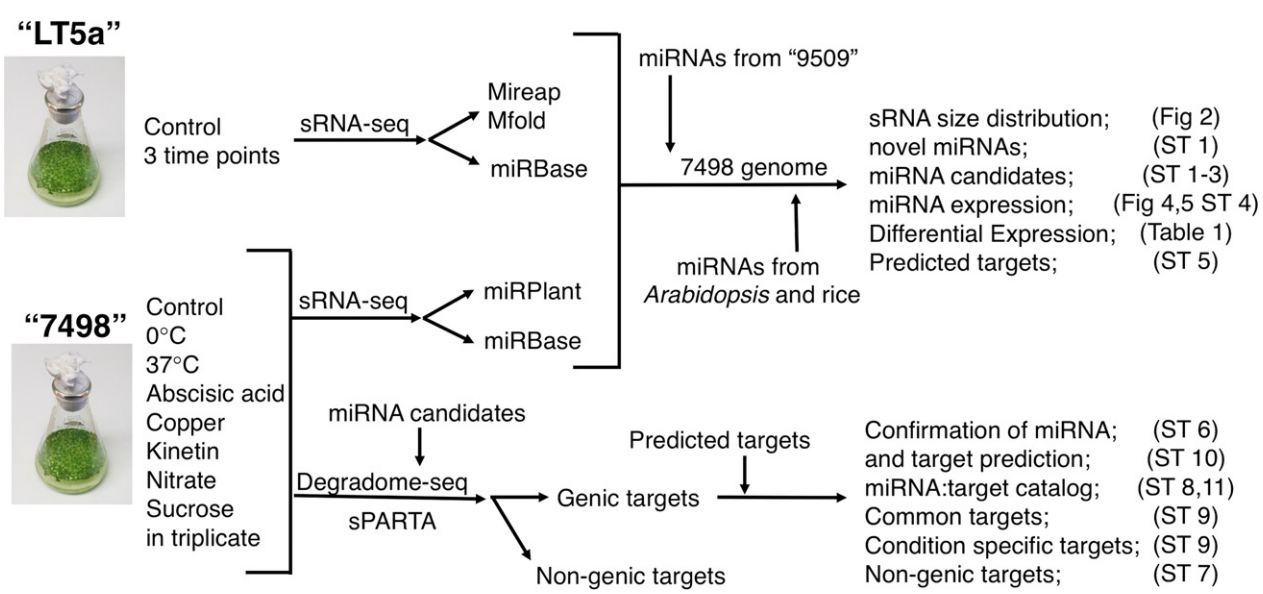

Figure 1. Study overview.
This study includes three different sequencing experiments, and eight experimental conditions. miRNA candidates were collected from both sRNA analyses, de-duplicated and then mapped to the '7498' genome. Then new miRNAs were added from strain 9509 and *Arabidopsis* and rice sequentially. These candidates were all analyzed for expression and with the degradome sequencing in the sPARTA program for post-transcriptional activity. Resulting data are described on the right, where ST stands for Supplemental Table and Fig is an abbreviation for Figure.

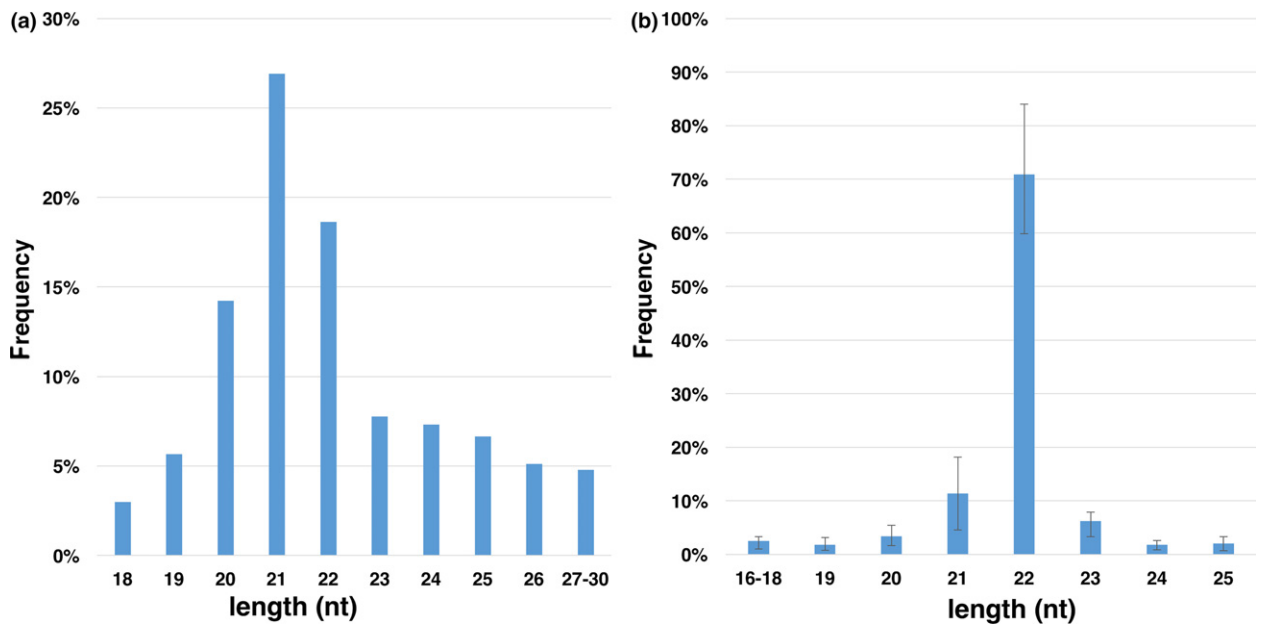

Figure 2. Length distributions of small RNAs from *Spirodela polyrhiza*.
(a) Length distribution in strain LT5a; (b) Length distributions in strain 7498. The error bars indicate the highest and lowest average percentage seen between the eight conditions sequenced.

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

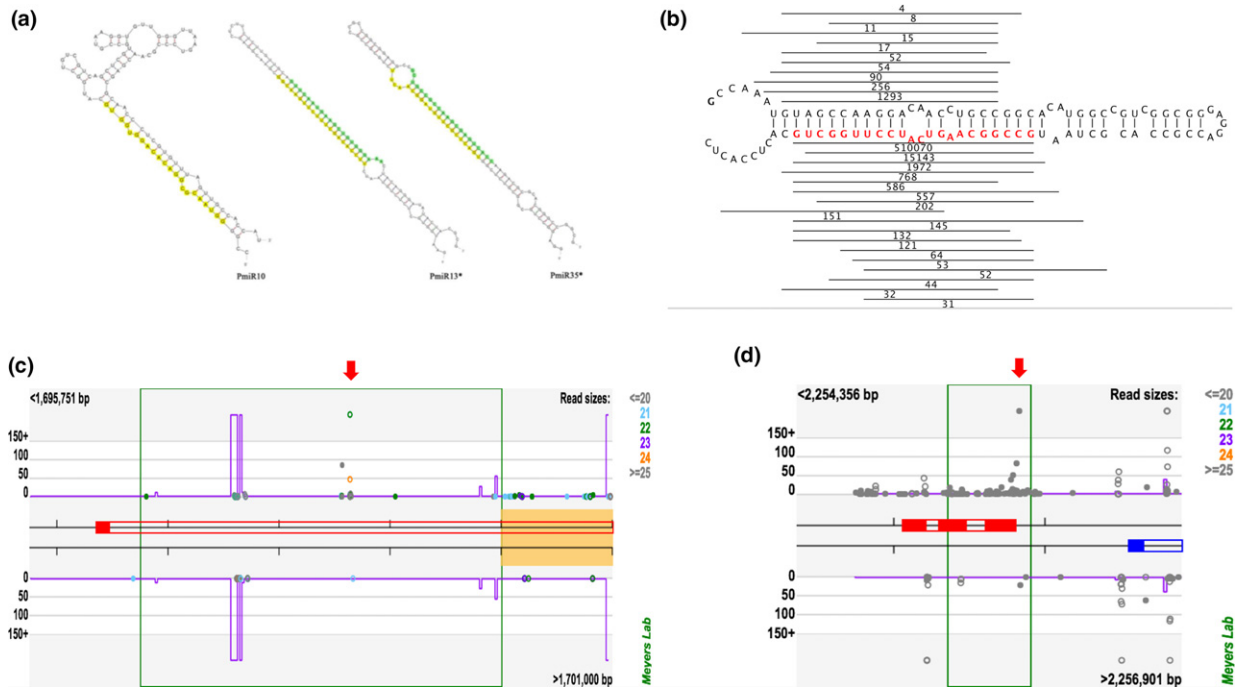

Figure 3. Visual examples of miRNA and target prediction.
(a) Mfold illustrations of predicted novel miRNAs; (b) miRPlant graph of read distribution along a predicted miR156 family hairpin; (c) Read distribution of miRNA spo-miR172d, marked with a red arrow, located along with a long terminal repeat (LTR) (Orange) in the intron of a SNF1 subunit gene Spipo1G0025900 as seen in the Myers lab web viewer; (d) GMUCT 2.0 read distribution in the same viewer showing a *sPARTA* verified cleavage of a squamosa promoter-like gene Spipo7G0024400 by miR156a at the site of the red arrow.

'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 were 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

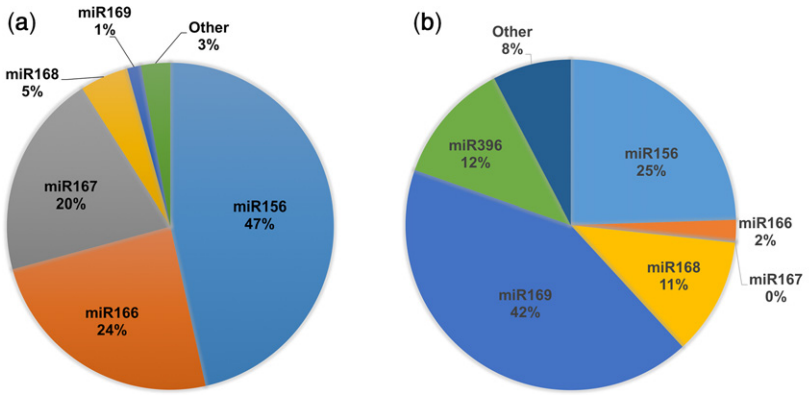

Figure 4. miRNA family expression in *Spirodela polyrhiza* under control conditions.
Expression of miRNA families in strain LT5a (a) and in strain 7498 (b) as the percentage of the total.

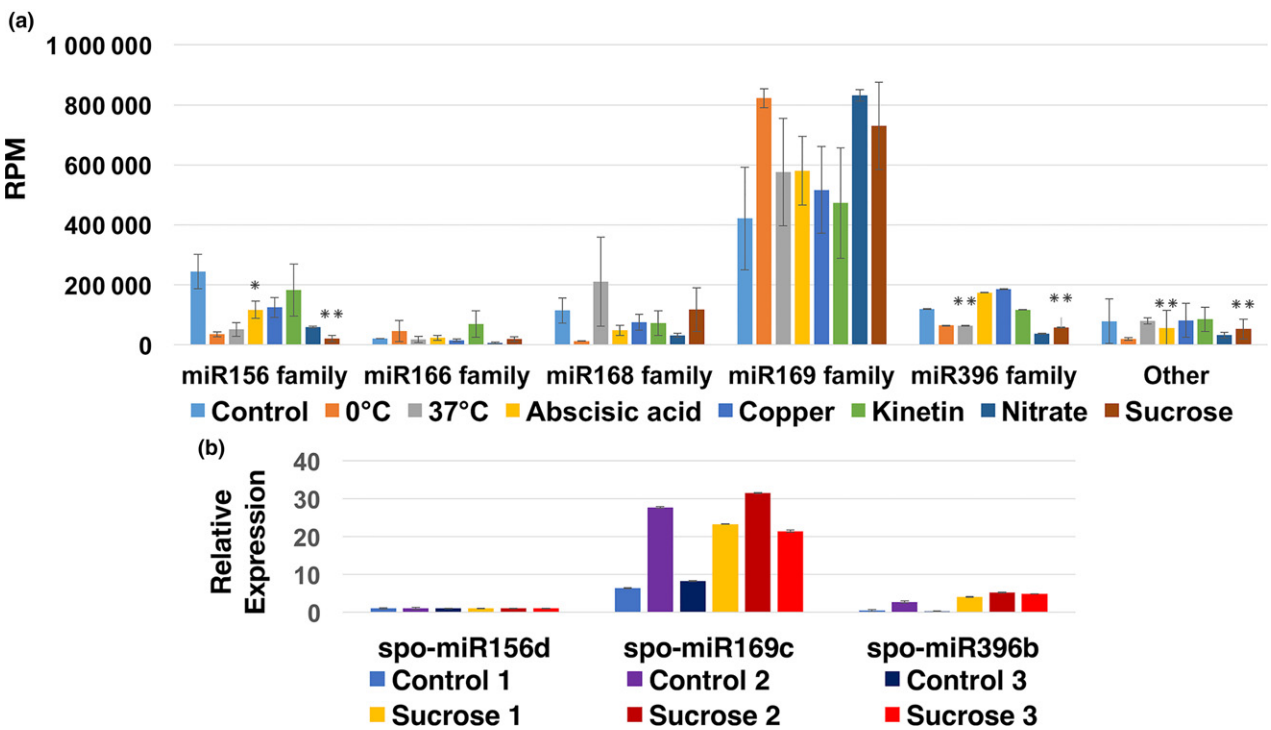

Figure 5. miRNA family expression under eight different conditions.
(a) Sequencing data for the five most abundant miRNA families and all other miRNAs, from biological triplicates of *S. polyrhiza* strain 7498 under eight conditions. The data are expressed as reads per million (RPM) of all miRNA reads ± the standard error of the mean (SEM). *Represents that a single miRNA sequence in that family or group was differentially expressed compared with control conditions with an adjusted $P < 0.05$ in Deseq2, while ** indicates $P < 0.01$. These differentially expressed miRNAs are listed in Table 1. Complete expression data for individual miRNAs are presented in Table S4; (b) Expression of spo-miR169c and 396b as fold change of miR156d ± the standard difference of error as measured by qPCR.

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

Table 1 Differentially expressed miRNAs

Condition	miRNA name	*P*-value adjusted	RPM in control	Relative expression
ABA	spo-miR156d-3p	0.03886	5900.0	15%
ABA	spo-miR171	0.04690	4919.6	16%
ABA	spo-miRnovel14-5p	0.03886	173.1	5%
ABA	spo-miRnovel21-3p	0.03886	68.7	480%
ABA	spo-miRnovel48-3p	0.00022	2451.4	7%
37°C	spo-miR396a	0.00925	30.3	698%
Sucrose	spo-miR156c	0.00146	2819.7	4%
Sucrose	spo-miR156c-3p	0.00002	198.6	1%
Sucrose	spo-miR156d-3p	0.00012	5900.0	6%
Sucrose	spo-miR171	0.00013	4919.6	8%
Sucrose	spo-miR224	0.00803	37.0	0%
Sucrose	spo-miR396b-3p	0.04752	254.9	5%
Sucrose	spo-miR396c	0.00658	383.0	14%
Sucrose	spo-miR8005	0.04052	104.5	0%
Sucrose	spo-miRnovel3-5p	0.01754	72.2	2%
Sucrose	spo-miRnovel14-5p	0.00146	173.1	5%
Sucrose	spo-miRnovel37	0.01232	554.0	11%
Sucrose	spo-miRnovel40	0.00080	85.4	0%
Sucrose	spo-miRnovel48-3p	0.00004	2451.4	5%

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.

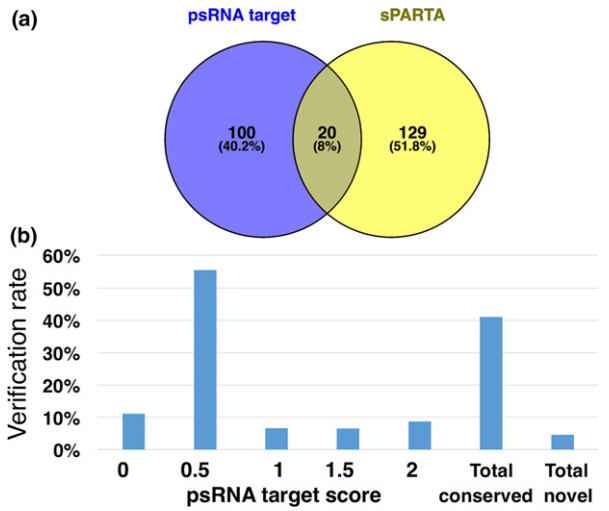

Figure 6. Overlap of target predictions and sequenced validations.
(a) Overlap of the mRNA targets predicted by psRNA target in blue and cleaved targets detected by sPARTA in yellow. Overlap indicates cleavage by the same miRNA family and predictions were for those miRNAs validated by the sPARTA analysis; (b) Verification rate of target predictions based on psRNA target scores with 0 being the best and two being acceptable. Verification rate for all predicted targets is also shown according to conservation status of the miRNA.

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 (Table S9). Therefore, it appears that the value of additional libraries was largely to identify condition-specific cleavage events.

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 trifoliate*, *Medicago truncatula* and *Oryza sativa* (Fahlgren *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). It is intriguing to note that this group of species with lower levels of 24-nt siRNAs includes not only angiosperms, such as duckweed and mustard, but also multiple ancient gymnosperms such as Norway spruce, *Ginkgo biloba*, moss,

Table 2 Condition-specific adaptations

Condition	DE miRNAs	Condition-specific targets	Notes
0°C	0	7	Osa-miR3979-5p cleaved two genes, one being a transcription factor
37°C	1	15	Mostly energy generating enzymes No miR156 activity detected
ABA	5	3	Three novel miRNAs were differentially expressed
Control	0	15	Eight novel miRNAs cleaved condition-specific targets
Copper	0	18	Polyphenol oxidases and ROS target genes
Kinetin	0	12	Five transcription factors uniquely cleaved. Highest miR156:172 ratio
Nitrate	0	14	Highest miR169 family expression
Sucrose	13	19	miR156 and miR396 family members along with five novel miRNAs downregulated. Many metabolically related mRNA targets

Summary table indicating the number of differentially expressed miRNAs and number of condition-specific mRNA targets. The notes describe some interesting features of the post-transcriptional profile. 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 mRNAs (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 mRNAs 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 mRNAs, 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)

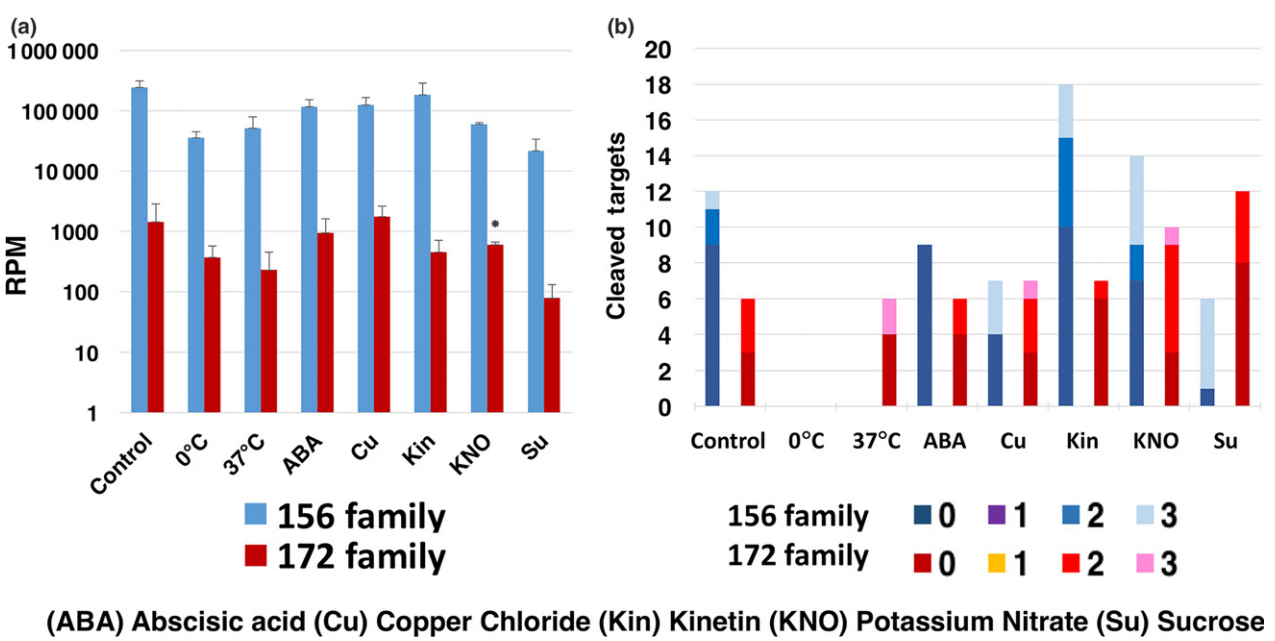

Figure 7. Abundance and target cleavage of miR156 and miR172 families.
(a) Read count of miR156 and miR172 family mature sequences as reads/million miRNA reads. Error bars are SEM of three biological replicates. *Represents that the miR156 family was significantly more abundant than the 172 family under potassium nitrate conditions with a $P < 0.05$. (b) Cleavage activity of miR156 and miR172 families expressed as number of cleaved targets in the three degradome libraries. Degree of cleavage is indicated by class. Class 0 reads indicate the miRNA cleavage site was the most abundant degradome read on the transcript, while Class 1 indicates a tie. Classes 2 and 3 are above and below median degradome reads, respectively, indicating less miRNA regulation.

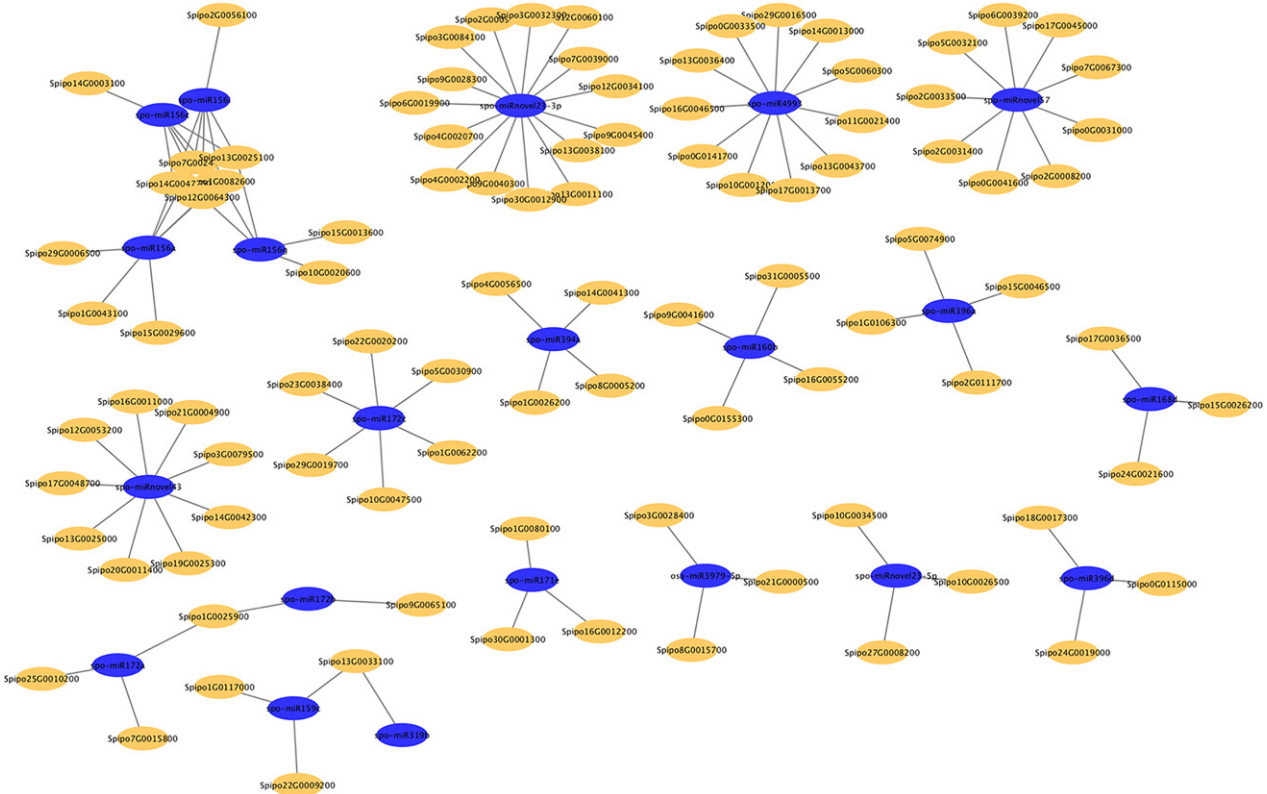

Figure 8. miRNA and target network from Table S8 represented in Cytoscape. Blue nodes are the 21 most connected miRNAs with the most targets, whereas yellow represents mRNA targets from filtered genic *sPARTA* results. The miR156 family is seen in the top left sharing most of their targets, and similarly the miR172 and miR159 and 319 families share their targets in the bottom left. Apart from those shared targets every mRNA was found to be regulated by one miRNA with a maximum of 15 targets.

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 $CuCl_2$, 300 mg/L KNO_3, 250 nM ABA, 10 mM 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). Following cDNA purification, 60–80 bp size selection was performed in Invitrogen 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* noncoding RNA sequences with miRNAs removed to filter out ribosomal, transfer, and other RNA sequences using the Galaxy wrapper (Afgan *et al.*, 2016), and then used as the sRNA library of miRPlant. All miRNA hairpin predictions with a score above 3 had their mature sequences aligned against miRBase mature miRNAs and previously predicted novel miRNAs using BLAST+ to predict hairpin structures (Camacho *et al.*, 2009; Cock *et al.*, 2015). Then novel miRNAs with at least 20 reads for the mature strand, and one for the passenger were included, with several being filtered at the authors' discretion due to repetitive sequences that matched hundreds of mRNA targets, or equal numbers of active and passenger reads. Finally, these structures with their read alignment were analyzed by the revised criteria for plant miRNA annotation to ensure only high quality annotations enter miRBase (Axtell and Meyers, 2018) (Table S2).

Genome alignment of conserved and predicted miRNAs. To conduct a complete survey of *Spirodela* miRNAs we included all the novel miRNAs from *Spirodela polyrhiza* strain '9509,' and the two conserved miRNA families not found in 'LT5a' and '7498' (Michael *et al.*, 2017). Additionally, we included members of *Arabidopsis thaliana*, or *Oryza sativa* conserved miRNA families, not observed in *Spirodela*, to detect potential miRNA genes that had not been expressed. Within Galaxy software (Afgan *et al.*, 2016) all miRNAs were mapped to the *Spirodela* '7498' genome to inspect their alignments, and determine their location using Bowtie 2 (Langmead *et al.*, 2009; Langmead and Salzberg, 2012), with no mismatches, then again allowing one mismatch. Results were converted from BAM to BED using BEDTools (Quinlan and Hall, 2010). The total filtered sRNA read library used in miRPlant was perfectly aligned with the genome and compared with the miRNA loci for slight adjustments, if needed, to reflect the most common sequence length. Genomic loci were then used to designate letters to conserved miRNAs, starting with chromosome 1.

miRNA expression quantification. Perfectly aligned sRNA reads were aligned against the *Brachypodium distachyon* noncoding RNA sequences with miRNAs removed to filter out ribosomal, transfer, and other RNA sequences (ensemblegenomes.org). Next, predicted miRNAs were quantified within libraries using the Salmon program (Patro *et al.*, 2015) in Galaxy wrapper (default settings except kmer length of 19, single strand, forward strand) to generate data shown in Figure 5(a). DeSeq2 (Love *et al.*, 2014) was used to identify differentially expressed miRNAs for the qPCR, biological triplicate flasks of *Spirodela* strain 7498 were grown under the same conditions. RNA was again extracted using Ambion mirVana™ miRNA Isolation Kit, DNase treated, reverse transcribed using the miScript Plant RT kit, and amplified using the miScript SYBR green kit in an Applied Biosystems Step-one plus instrument. All genes produced consistent products without expression of the negative controls.

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 differentially cleaved targets. The stimuli included cold treatment at 0°C as an ice water bath for 24 h, heat treatment at 37° for 1 h, 250 nM ABA for 3 h, 20 mg/L $CuCl_2$ for 24 h, 10 mM kinetin for 3 h, 300 mg/L KNO_3 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 RNeasy 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

(Agilent). 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. mRNA was selected using Dynabeads oligo-dt (Ambion). At least 300 ng of mRNA with less than 10% rRNA was used according to the mRNA Assay on BioAnalyzer (Agilent); 5′ ligation with the small RNA 5′ adaptor (RA5) was performed as previously described (Willmann *et al.*, 2014). A second poly-A selection was performed using 100 μL Dynabeads oligo(dt) (Ambion). The GMUCT 2.0 cDNA was synthesized using a primer with sequence CTGGAGTTCCTTGGCACCCGAGAATTCCANNNNNN 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 RPIX (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). Equimolar pools of libraries were made, diluted to 2 nM, 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 'SlopBed' program in the Galaxy wrapper (Quinlan and Hall, 2010). Next, sPARTA analysis was run as described (Arikit *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 Deepti Ramachandruni for database construction.

DECLARATIONS

Ethics approval and consent to participate.

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

DATA ACCESSIBILITY

NCBI accession numbers: 'LT5a' small RNA sequencing data GSE55208, '7498' small RNA libraries and '7498' degradome libraries PRJNA473779 (SRP149336). The analyses run in Galaxy can be found at https://usegalaxy.org/u/paul-fouroun jian/h/mirna-loci-and-expression. The data can also be interactively viewed in the Myers lab genome browser in the MPSS database at https://mpss.danforthcenter.org/dbs/inde x.php?SITE=spirodela_PARE and https://mpss.danforthcente r.org/dbs/index.php?SITE=spirodela_sRNA.

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

Afgan, E., Baker, D., van den Beek, M. *et al.* (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44**, W3–W10.

Allen, E., Xie, Z., Gustafson, A.M., Sung, G.-H., Spatafora, J.W. and Carrington, J.C. (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat. Genet.* **36**, 1282–1290.

An, J., Lai, J., Sajjanhar, A., Lehman, M.L. and Nelson, C.C. (2014) miRPlant: an integrated tool for identification of plant miRNA from RNA sequencing data. *BMC Bioinformatics*, **15**, 275.

Arikit, S., Xia, R., Kakrana, A. *et al.* (2014) An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes. *Plant Cell Online*, **26**, 4584–4601.

Aukerman, M.J. (2003) Regulation of flowering time and floral organ identity by a MicroRNA and Its APETALA2-like target genes. *Plant Cell Online*, **15**, 2730–2741.

Axtell, M.J. and Meyers, B.C. (2018) Revisiting criteria for plant miRNA annotation in the era of big data. *Plant Cell*, **30**, tpc.00851.2017.

Berezikov, E., Cuppen, E. and Plasterk, R.H.A. (2006) Approaches to microRNA discovery. *Nat. Genet.* **38**, S2–S7.

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.

Brenner, W.G., Romanov, G.A., Köllmer, I., Bürkle, L. and Schmülling, T. (2005) Immediate-early and delayed cytokinin response genes of *Arabidopsis thaliana* identified by genome-wide expression profiling reveal novel cytokinin-sensitive processes and suggest cytokinin action through transcriptional cascades. *Plant J.* **44**, 314–333.

Burge, S.W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E.P., Eddy, S.R., Gardner, P.P. and Bateman, A. (2012) Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* **41**, 1–7.

Calviño, M., Bruggmann, R. and Messing, J. (2011) Characterization of the small RNA component of the transcriptome from grain and sweet sorghum stems. *BMC Genom.*, **12**, 356.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

Cao, H.X., Vu, G.T.H., Wang, W., Appenroth, K.J., Messing, J. and Schubert, I. (2016) The map-based genome sequence of *Spirodela polyrhiza* aligned with its chromosomes, a reference for karyotype evolution. *New Phytol.* **209**, 354–363.

Carrington, J.C. and Ambros, V. (2003) Role of microRNAs in plant and animal development. *Science (80-.)* **301**, 336–338

Chen, L., Wang, T., Zhao, M., Tian, Q. and Zhang, W.H. (2012) Identification of aluminum-responsive microRNAs in *Medicago truncatula* by genome-wide high-throughput sequencing. *Planta*, **235**, 375–386.

Chou, C.-H., Chang, N.-W., Shrestha, S. *et al.* (2016) miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res.* **44**, D239–D247.

Chuck, G., Cigan, A.M., Saeteurn, K. and Hake, S. (2007) The heterochronic maize mutant Corngrass1 results from overexpression of a tandem microRNA. *Nat. Genet.* **39**, 544–549.

Cock, P.J.A., Chilton, J.M., Grüning, B., Johnson, J.E. and Soranzo, N. (2015) NCBI BLAST+ integrated into Galaxy. *Gigascience*, **4**, 0–6.

Dai, X. and Zhao, P.X. (2011) PsRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* **39**, W155–W159.

Dai, X., Zhuang, Z. and Zhao, P.X. (2011) Computational analysis of miRNA targets in plants: current status and challenges. *Brief. Bioinform.* **12**, 115–121.

Dolgosheina, E.V., Morin, R.D., Aksay, G., Sahinalp, S.C., Magrini, V., Mardis, E.R., Mattsson, J. and Unrau, P.J. (2008) Conifers have a unique small RNA silencing signature. *RNA*, **14**, 1508–1515.

Fahlgren, N., Howell, M.D., Kasschau, K.D. *et al.* (2007) High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS ONE*, **2**, 219–234.

Fattash, I., Voss, B., Reski, R., Hess, W.R. and Frank, W. (2007) Evidence for the rapid expansion of microRNA-mediated regulation in early land plant evolution. *BMC Plant Biol.* **7**, 13.

Griffiths-Jones, S.M. (2006) The microRNA sequence database. *Methods Mol. Biol.* **342**, 129–38.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. and Enright, A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34**, D140–D144.

Griffiths-Jones, S., Saini, H. and van Dongen, S. (2007) miRBase_tools for microRNA genomics. *Nucleic Acids Res.* **36**, D154–D158.

Hammond, S.M., Bernstein, E., Beach, D. and Hannon, G.J. (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature*, **404**, 293–296.

Hego, E., Vilain, S., Barré, A., Claverol, S., Dupuy, J.-W., Lalanne, C., Bonneu, M., Plomion, C. and Mench, M. (2016) Copper stress-induced changes in leaf soluble proteome of Cu-sensitive and tolerant *Agrostis capillaris* L. populations. *Proteomics*, **16**, 1386–1397.

Heyl, A., Ramireddy, E., Brenner, W.G., Riefler, M., Allemeersch, J. and Schmulling, T. (2008) The transcriptional repressor ARR1-SRDX suppresses pleiotropic cytokinin activities in Arabidopsis. *Plant Physiol.* **147**, 1380–1395.

Hsu, S.Da, Lin, F.M., Wu, W.Y. *et al.* (2011) MiRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **39**, D163–D169.

Hsu, S.Da, Tseng, Y.T., Shrestha, S. *et al.* (2014) MiRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res.* **42**, D78–D85.

Hutvagner, G. and Simard, M.J. (2008) Argonaute proteins: key players in RNA silencing. *Nat. Rev. Mol. Cell Biol.* **9**, 22–32.

Kakrana, A., Hammond, R., Patel, P., Nakano, M. and Meyers, B.C. (2014) SPARTA: a parallelized pipeline for integrated analysis of plant miRNA and cleaved mRNA data sets, including new miRNA target-identification software. *Nucleic Acids Res.* **42**, 1–13.

Khraiwesh, B., Pugalenthi, G. and Fedoroff, N.V. (2013) Identification and analysis of red sea mangrove (*Avicennia marina*) microRNAs by high-throughput sequencing and their association with stress responses. *PLoS ONE*, **8**, e60774.

Kidner, C.A. and Martienssen, R.A. (2004) Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. *Nature*, **428**, 81–84.

Kozomara, A. and Griffiths-Jones, S. (2011) MiRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39**, D152–D157.

Kozomara, A. and Griffiths-Jones, S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68–D73.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–9.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H. and Kim, V.N. (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051–4060.

Li, Y.-F., Zheng, Y., Addo-Quaye, C., Zhang, L., Saini, A., Jagadeeswaran, G., Axtell, M.J., Zhang, W. and Sunkar, R. (2010) Transcriptome-wide identification of microRNA targets in rice. *Plant J.* **62**, 742–759.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.

Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A. and Matzke, A.J. (2000) Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* **19**, 5194–5201.

Michael, T.P., Bryant, D., Gutierrez, R. *et al.* (2017) Comprehensive definition of genome features in *Spirodela polyrhiza* by high-depth physical mapping and short-read DNA sequencing strategies. *Plant J.* **89**, 617–635.

Nakano, M., Nobuta, K., Vemaraju, K., Tej, S.S., Skogen, J.W. and Meyers, B.C. (2006) Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. *Nucleic Acids Res.* **34**, D731–D735.

Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D. and Grewal, S.I.S. (2004) RITS acts in cis to promote RNA interference–mediated transcriptional and post-transcriptional silencing. *Nat. Genet.* **36**, 1174–1180.

Olsen, J.L., Rouzé, P., Verhelst, B. *et al.* (2016) The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature*, **530**, 331–335.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C. and Weigel, D. (2003) Control of leaf morphogenesis by microRNAs. *Nature*, **425**, 257–263.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C. (2015) Salmon provides accurate, fast, and bias-aware transcript expression estimates using dual-phase inference. *bioRxiv* **021592**.

Qiu, D., Pan, X., Wilson, I.W., Li, F., Liu, M., Teng, W. and Zhang, B. (2009) High throughput sequencing technology reveals that the taxoid elicitor methyl jasmonate regulates microRNA expression in Chinese yew (*Taxus chinensis*). *Gene*, **436**, 37–44.

Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

Reis, R.S., Hart-Smith, G., Eamens, A.L., Wilkins, M.R. and Waterhouse, P.M. (2015) Gene regulation by translational inhibition is determined by Dicer partnering proteins. *Nat. Plants*, **1**, 14027.

Schwab, R., Palatnik, J.F., Riester, M., Schommer, C., Schmid, M. and Weigel, D. (2005) Specific effects of microRNAs on the plant transcriptome. *Dev. Cell*, **8**, 517–527.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software

environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–504.

Song, C., Wang, C., Zhang, C., Korir, N., Yu, H., Ma, Z. and Fang, J. (2010) Deep sequencing discovery of novel and conserved microRNAs in trifoliate orange (*Citrus trifoliata*). *BMC Genom.*, **11**, 431.

Sunkar, R., Zhou, X., Zheng, Y., Zhang, W. and Zhu, J.-K. (2008) Identification of novel and candidate miRNAs in rice by high throughput sequencing. *BMC Plant Biol.* **8**, 25.

Szittya, G., Moxon, S., Santos, D.M., Jing, R., Fevereiro, M.P., Moulton, V. and Dalmay, T. (2008) High-throughput sequencing of *Medicago truncatula* short RNAs identifies eight new miRNA families. *BMC Genom.*, **9**, 593.

Tang, J., Zhang, F., Cui, W. and Ma, J. (2014) Genetic structure of duckweed population of *Spirodela*, *Landoltia* and *Lemna* from Lake Tai, China. *Planta*, **239**, 1299–1307.

Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192.

Van Hoeck, A., Horemans, N., Monsieurs, P., Cao, H.X., Vandenhove, H. and Blust, R. (2015) The first draft genome of the aquatic model plant Lemna minor opens the route for future stress physiology research and biotechnological applications. *Biotechnol. Biofuels*, **8**, 188.

Vazquez, F. (2006) Arabidopsis endogenous small RNAs: highways and byways. *Trends Plant Sci.* **11**, 460–468.

Voinnet, O. (2009) Origin, biogenesis, and activity of plant MicroRNAs. *Cell*, **136**, 669–687.

Wang, W., Haberer, G., Gundlach, H. *et al.* (2014a) The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. *Nat. Commun.* **5**, 1–13.

Wang, W., Wu, Y. and Messing, J. (2014b) RNA-seq transcriptome analysis of *Spirodela* dormancy without reproduction. *BMC Genom.*, **15**, 60.

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*, **67**, 64–73.

Wu, G., Park, M.Y., Conway, S.R., Wang, J., Weigel, D. and Scott, R. (2009) The sequential actions of miR156 and miR172 regulates developmental timing in Arabidopsis. *Cell*, **138**, 750–759.

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**, 642–652.

Xu, J., Cheng, J.J. and Stomp, A.-M. (2012) Growing *Spirodela polyrrhiza* in Swine wastewater for the production of animal feed and fuel ethanol: a pilot study. *CLEAN - Soil, Air, Water*, **40**, 760–765.

Yang, J., Liu, X., Xu, B., Zhao, N., Yang, X. and Zhang, M. (2013) Identification of miRNAs and their targets using high-throughput sequencing and degradome analysis in cytoplasmic male-sterile and its maintainer fertile lines of *Brassica juncea*. *BMC Genom.*, **14**, 9.

Zhao, C.-Z., Xia, H., Frazier, T., Yao, Y.-Y., Bi, Y.-P., Li, A.-Q., Li, M.-J., Li, C.-S., Zhang, B.-H. and Wang, X.-J. (2010) Deep sequencing identifies novel and conserved microRNAs in peanuts (*Arachis hypogaea* L.). *BMC Plant Biol.* **10**, 3.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415.