EXAMINING *SPIRODELA* SMALL RNAS & THE WORLD'S SMALLEST FLOWERS

by

PAUL JAMES FOUROUNJIAN

A dissertation submitted to the

School of Graduate Studies Rutgers,

The State University of New Jersey

In partial fulfillment of the requirements

For the degree of Doctor of Philosophy

Graduate Program in

Microbiology and Molecular Genetics

Written under the direction of

Dr. JOACHIM MESSING

And approved by

_____

_____

_____

_____

New Brunswick New Jersey

May 2020

ABSTRACT OF THE DISSERTATION

EXAMINING *SPIRODELA* SMALL RNAS

& THE WORLD'S SMALLEST FLOWERS

by

PAUL JAMES FOUROUNJIAN

Dissertation Director

Dr. Joachim Messing

When I joined the duckweed community the interest was largely on sustainably providing clean water, food, and fuel as reflected in the Introduction and Conclusion. Then a fellow student of Dr. Messing, Dr. Wang, published the *Spirodela polyrhiza* 7498 genome and transcriptome. My thesis project would complement these with the *Spirodela* miRNA and target catalog, and investigate the relationship of the miRNAs 156 and 172 to the neotenous nature and rare flowering in the Lemnaceae. While characterizing the *Spirodela* miRNAs and their targets as described in Chapter 1, both Dr. Tang and Dr. Myers noticed a lack of 24nt siRNAs leading to an analysis of the RdDM pathway and transposons. This research coupled with my editing of The Duckweed Genomes textbook led to the writing of Chapters 2, 3, and 4 on small RNAs, transcriptomics, and repetitive DNA elements respectively. While studying the miRNAs and genomics I also developed flowering protocols in order to study the relationship between miR156 and 172 and floral regulation. As described in Chapter 5, I developed *Wolffia microscopica*, *Lemna minor*, *Lemna gibba*, and *Spirodela polyrhiza* flowering protocols to enable further research. I hope you find this dissertation helpful, and that these flowering protocols, the catalog of

miRNAs and their targets, the textbook chapters, and other data contribute to a deeper understanding of this fascinating family, and floral regulation.

ACKNOWLEDGEMENTS

This thesis represents the direct or indirect work of so many people, whom I'm incredibly thankful to. First and foremost is my advisor Joachim Messing. I deeply appreciate the financial support of the Waksman Fellowship and opportunity to research this project. I'm sincerely thankful for his patient guidance, passion for discovery, and the academic freedom and responsibility he provided. Like many others I look to him as a role model. He was brilliantly creative and hardworking, extremely generous in publishing instead of patenting his work, and a model of academic integrity. While we all wish he had more time, he lead a full and fulfilling life, and left the world so much better than he found it.

Dr. Messing recruited and trained many wonderful scientists, and I'm incredibly grateful for everything they've taught me and our time together. It was a pleasure working alongside my labmates Yongrui Wu, Martin Calvino, Wei Zhang, Anna Zdepski, Nelson Garcia, Jose Planta, Yin Li, Jiaqiang Dong, Zhiyong Zhang, Chenxu Liu, Guangshu Song, and Fan Feng, who taught me so much technically, culturally, and personally. Gajendran Karunanithi, Chris WAKIM, and Kaleb Friedman, it was fun and inspirational working alongside you. While I hope I taught you all I could as your mentor, I consider each of you better scientists than me and look forward to your bright futures. Jennie Ayer and Marc Probasco you were an excellent "Lab Mom and Dad team" managing our lab, greenhouse, and so much else. Thank you Jenny Tirrito and Erin Sorge and all the other staff for taking care of us and keeping the Institute running. I'd like to specially thank Professor Wenqin Wang, for taking me under her wing and mentoring me.

Professors Gerard Dismukes, I will always be grateful for the way you took me into your lab as your own student, and mentored me for the final stretch. Thank you as well professors Faith Belanger, and James White for serving on my thesis committee and providing your mentorship, foresight, and guidance through the years. I'd also like to thank my other professors and teachers notably, Janet Morrison, Matthew Cathell, John Henrikson, and Richard O'Hara, for inspiring a love of learning. Finally, thank you so much Professor Janet Slovin for everything you've taught me, and all your enthusiastic support of our research together. We couldn't have done this without your expertise and kindness.

None of this work would have been possible without the hard work of all my kind and talented collaborators. I'm incredibly grateful to Dr. Min Tu, Dr. Brian Gelfand, Dr. Yaping Feng, and Dr. Dibyendu Kumar of the Waksman Genomics core for all their work in learning and executing novel library preparations and analyses, Dr. Jie Tang, Dr. Jiong Ma, Dr. Ying Zhang, and especially Dr. Ying-Ying Zhang who sequenced *Spirodela* LT5a small RNAs, Dr. Atul Kakrana, Dr. Mayumi Nakano, Deepti Ramachandruni and Dr. Blake Myers who ran sPARTA analysis and uploaded the results to their database, all the members of the Blake Myers' and Bob Schmitz's labs who analyzed the RdDM pathway, especially Adam Bewick and Alex Harkess, and finally to Wenqin Wang and Hieu Cao my co-editors of The Duckweed Genomes. I hope you enjoy seeing our work together presented here. Thank you as well to Springer Nature, and The Plant Journal for legal permission to reproduce the Introduction, and Chapters 1-4.

I'm so happy to research duckweeds, not just because of the wonderful plants themselves, but the great people working on them in academic and commercial settings. It has been, and will be a pleasure working with all of you, even when we find out we're solving the same problem, so I'm proud to be part of this community. Professors Xu, Sree, and Appenroth it was excellent discussing our related projects at the ICDRA 2019, which I thank the Graduate School New Brunswick for sending me to. All of us, especially me, owe so much to Tamra Fakhoorian for co-founding and running the International Lemna Association.

I'm tremendously grateful for my family, Mom, Dad, Amy, my extended family, and grandparents, for all their support, patience and love along the way. I know it must be tough to raise a son who goes to school until he's 30, and thank you for everything you did. As always, I owe who I am to you and love you dearly.

Finally I'd like to thank all my friends for the great times we've had, all you've taught me, and all your encouragement. Once again I'm incredibly lucky to have so many wonderful people in my life, and can't thank you all enough.

TABLE OF CONTENTS

List of Tables

Chapter 5: Flowering and seed production across the Lemnaceae

LIST OF ILLUSTRATIONS

Introduction: Importance of Duckweeds in Basic Research and Their Industrial Applications

Chapter 1: Post-transcriptional adaptation of the aquatic plant *Spirodela polyrhiza* under stress and hormonal stimuli

Chapter 2: Small RNAs in Duckweed

Chapter 5: Flowering and seed production across the Lemnaceae

PROLOGUE

Lemnaceae Family

The 37 species and 5 genera comprising the Lemnaceae family, commonly known as duckweeds, are the smallest, fastest growing, most morphologically reduced, and widely distributed family of angiosperms (Landolt and Kandeler 1987; Ziegler et al. 2015). They have a small (0.5mm -2cm) flat leaf-like thallus structure called a frond, and the *Spirodela*, *Landoltia*, and *Lemna* genera have rhizoids while the smaller, simpler, and more recently evolved *Wolffiella* and *Wolffia* genera lack these structures (Landolt and Kandeler 1987). The duckweeds are found floating and rapidly clonally dividing on still nutrient rich waters worldwide. Fascinated by their neotenous lifestyle of rarely maturing to adulthood and flowering, we sought to study floral initiation, and the genetic mechanisms regulating flowering in this unique background, and lay the foundation for further research into this topic.

Lemnaceae as a Sustainable Crop

Lemnaceae research has a long history, where the family was a convenient model of plant biology, being heavily researched from 1950 to 1990, until most research shifted to *Arabidopsis* due to its rapid reproduction and transformation (Landolt and Kandeler 1987; Zhao et al. 2012). Duckweed research then saw a resurgence around 2009 when amidst climate change concerns, global fuel prices rose, prompting the US Department of Energy Joint Genome Initiative to fund the sequencing of the *Spirodela polyrhiza* genome to help develop a biofuel crop that didn't require arable land, clean water or

fertilizer (https://jgi.doe.gov/why-sequence-the-greater-duckweed/). Before and since that project scientists and companies have shown that duckweed can sustainably provide clean water, food, and fuel. They can either be cultivated as a water, time, and land efficient vegetable, or used to treat wastewater. In the case of agricultural wastewater the duckweed is fit to be used as animal feed, largely equivalent to soybean, and if the duckweed is treating industrial wastewater, it can sequester heavy metals and be converted to biofuels (Cao et al. 2018). While microalgae can similarly achieve these goals with a faster growth rate, high salinity tolerance, and high lipid content it is very difficult and costly to harvest, dewater, and process the algae, while the size of duckweed allows them to minimize these engineering challenges. Excited by this promising new crop, I originally joined Professor Joachim Messing's lab to work on the genetic aspects of a sustainable biofuel and water treatment project, and perhaps other commercial applications, which held great potential in sustainably providing clean water, food, and fuel. Since these applications have inspired many people and encouraged academic investigation of the Lemnaceae genomes and physiology, the text-book chapter "Importance of duckweeds in basic research and their industrial applications" serves as a great introductory chapter to both the the textbook I co-edited, The Duckweed Genomes, and the rest of this doctoral thesis.

Spirodela Genome & Starting the Thesis

In several ways the start of my research with Dr. Messing on February 4th 2013 roughly coincided with several beginnings in this field of research. Shortly after joining the lab I attended the Second International Conference on Duckweed Research and Applications

(ICDRA), with roughly 60 attendees. We mourned the loss of one of the field's founders Elias Landolt that year, heard about some of the earliest studies on beneficial bacteria and large-scale academic agronomy research, and met some of the first companies. Like the first conference, this one had talks of genome sequencing projects in progress, but no lectures or posters on flowering (Zhao et al. 2012; Lam et al. 2014).

Before any nuclear genomes were available, sequencing started with the chloroplast and mitochondrial genomes (Wang and Messing 2011; Wang et al. 2012). Then in January of 2014 now Professor Wenqin Wang, Dr. Messing, and an international team of 24 other collaborators published the *Spirodela polyrhiza* 7498 genome map enabling genomic studies (Wang et al. 2014a). This was closely followed by a transcriptomic study of ABA induced turion production in 7498 (Wang et al. 2014b). From 2014 to 2019 there were 3 other genomes published and 5 transcriptomes published (Wang and Messing 2015; An et al. 2018). As described in "The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle" Dr. Messing was interested in the juvenile, asexual lifestyle of *Spirodela* and deduced that the 24 and single copies of miRNA156 and 172 families initially annotated are likely essential genetic mechanisms in this development. I was surprised that the "Never grow up" neotenous lifestyle of these tiny plants was responsible for the fastest known plant growth rate, and also wanted to study the extremely rare, smallest flowers in the world. My thesis therefore was to investigate the role of these miRNAs on the neotenous development and floral regulation of the Lemnaceae.

Characterizing miRNAs & mRNA Targets

To learn more about these miRNAs, and further characterize *Spirodela polyrhiza* Dr. Messing provided me with small RNA sequencing data generated by Professor Bahattin Tanyolac, in 8 different conditions (Control, 0°C, 37°C, ABA, $CuCl_2$, Kinetin, KNO_3, and Sucrose) and instructed me to identify the miRNAs in the *Spirodela* genome and analyze their expression between the conditions leading to Chapter 1. Interested in biofuels during the ICDRA 2013 conference, I had forgotten about work presented by Professor Weixiong Zhang on miRNA sequencing in *Spirodela polyrhiza* strain 9509, and was unaware of now professor Jie Tang's and Professor Jiong Ma's work on strain LT5a, and believed Dr. Messing and I were the only group conducting this research. As described in Chapter 3, data on miRNA sequencing and ABA response was published as part of the 9509 genome paper (Michael et al. 2017) and Professors Tang and I combined our studies into a single publication (Fourounjian et al. 2019). Viewing these 3 sequencing projects together illustrates slight variations in miRNA family copy number based on methodology (~ 9 copies of miR156 and ~5 of miR172), variations between the most abundant families, and our commitment to apply stringent plant miRNA annotation criteria suggested by (Meyers et al. 2008; Axtell and Meyers 2018) to qualify miRNAs for miRBase's high confidence list.

While many papers identify the mRNA targets of miRNAs with prediction programs, this is challenging, and the best of them provide a roughly 50% false positive rate (Dai et al. 2011). So I suggested, and we chose to, validate our miRNA annotations and mRNA target predictions through degradome sequencing analysis of *Spirodela polyrhiza* 7498 in the same 8 conditions as the sRNA, as reviewed in Chapter 3. Our goal was to create an experimentally verified miRNA:mRNA target catalog, observe post-

transcriptional profiles in the various stimuli, and understand the function of novel miRNAs.

Spirodela RdDM & Transposable Elements

After analyzing sPARTA output data provided by Dr. Atul Kakrana in Professor Blake Myer's lab to validate 149 targets cleaved by 66 miRNAs, I worked with Mayumi Nakano and Deepti Ramachandruni, to upload the data to their online interactive database. Here, Dr. Myers noticed the lack of 24nt heterochromatic siRNAs that Dr. Tang had discussed in our manuscript. Accordingly, Dr. Alex Harkess from the Myers lab and Dr. Adam Bewick from Professor Bob Schmitz's lab, investigated the extreme lack of 24nt het-siRNAs, the severely reduced RNA-directed DNA Methylation (RdDM) pathway, and the lack of recent transposition while I provided some data and introduced them to duckweed biology (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). As discussed in Chapter 5, despite the severe lack of the RdDM "genetic immune system" to prevent the virus-like transposons from multiplying, transposons were actually very rare. We hypothesized this is due to their activation in pollen development and the extreme lack of flowering in *Spirodela* and the duckweed family as a whole.

Lemnaceae Flowering

This curiosity in transposon regulation brings us back to Lemnaceae flowering. In addition to understanding the process so we could help open the doors to genetic

inheritance studies and commercial breeding, we sought to investigate the role of miRNAs on the floral regulation and development in the Lemnaceae for a deeper understanding of neoteny. In their comprehensive 1987 monograph, Landolt and Kandeler summed up the 106 flowering studies illustrating that daylength, salycilic acid (SA) and EDDHA (ethylenediamine-N,N′-bis(2-hydroxyphenylacetic acid) are the 3 most common inducers of flowering. Then Professor Arnold Pieterse's 2013 review "Is flowering in Lemnaceae stress induced?" illustrated that EDDHA likely broke down into SA, and together with jasmonic acid (JA), these stress hormones induced flowering in duckweed. He then described the evolutionary sense of setting seeds as a way to survive stress, and reviewed the 6 papers written on Lemnaceae flowering from 1987-2013 reflecting the relative gap in the research record. These 2 publications essentially sum up the knowledge in the field when I started researching the floral regulation paper presented in Chapter 6.

Starting with *Spirodela polyrhiza* 7498 I ran experiments with daylength, SA, EDDHA, JA, EDTA (ethylenediaminetetraacetic acid), and nutrient deprivation, like those previously conducted in other strains of *Spirodela* and other medias, and didn't see any flowers. Shortly after the publication of the re-discovered *Wolffia microscopica* (strain 2005), Professor Klaus Appenroth kindly gifted the strain to our lab to study (Sree et al. 2015). We sought to compare the miRNA profile of a rarely and commonly flowering duckweed, but like Professor Sowjanya Sree's group we couldn't control flowering, and took several years to develop a protocol to reliably grow flowering and non-flowering *W. microscopica* populations. At the ICDRA 2019 we found out that we

had independently developed quite different protocols to study *Wolffia microscopica* flowering, and that Professor Sree was running mRNA sequencing of the flowers.

As a student studying Lemnaceae flowering, I was extremely fortunate to work 4 hours away from Professor Janet Slovin, the first person to publish a Lemnaceae breeding protocol, who generously agreed to join my thesis committee. She suggested experimenting with *Lemna gibba* G3, and in September 2015 I ordered strain DWC130 (Slovin parental line). I had no success with this strain, and in 2017 Professor Jiaming Zhang's group published a *Lemna gibba* G3 breeding, and cross breeding protocol using DWC114 (Sicilia Siracusa) (Fu et al. 2017). They were unaware of the breeding protocol tucked in the methods section of (Slovin and Cohen 1988), which had surprisingly never been cited for its breeding protocol. Inspired by their success with DWC114, I ordered all 4 strains of *Lemna gibba* G3, observed very different flowering phenotypes, and found that DWC114 was in fact *Lemna gibba* G3, while DWC130, DWC131, and DWC132 were *Lemna minor* strains that had been mislabeled.

Inspired by success in *Wolffia* and *Lemna*, I returned to flowering experiments with the sequenced and well-studied *Spirodela* 7498 and 9509. Using daylength, SA, and EDDHA I developed protocols with 1-6% flowering rate and a high turion formation rate, which could enable transcriptomic studies of these flowers. At ICDRA 2019 I also heard from Professor Shuqing Xu that geographic origin of the *Spirodela* strains made a large difference in flowering rate, with the European and North American strains I studied being primarily asexual.

At the start of my thesis, the global understanding of flowering was essentially that daylength and stress triggered flowering, mostly as we understood it in 1988. Then

with renewed interest and a genome sequence Dr. Messing and I sought to understand the genetic mechanisms that regulated flowering, including the role of the miRNAs 156 and 172. This meant I needed to characterize the miRNAs in the genome, and secondly, that I needed to generate flowering protocols in this highly neotenous family. Working seemingly from the least to most asexual species, and to smaller simpler genomes, I developed 3 main flowering protocols across the Lemnaceae. In *Wolffia microscopica* I developed what appears to be a rare non-daylength, stress-free flowering protocol. We hope this helps study flowering with less complicating variables. Intrigued by this, I found convincing literature from other *Arabidopsis* and other plant species demonstrating that Flowering Locus D is the link between the stress response and the promotion of flowering (Singh et al. 2014). This provided the genetic mechanism likely at work in most duckweed flowering protocols, yet possibly not in this one for *Wolffia microscopica*. In *Lemna gibba* and *minor* I barcoded the 4 strains, developed a more convenient breeding protocol, and identified 3 different flowering phenotypes. For *Spirodela polyrhiza* I developed flowering protocols useful for studying the well characterized strains 7498, and 9509. While we didn't combine our characterized miRNAs and flowering protocols to sequence and analyze the role of the miRNAs during the floral transition, we created the requisite materials and knowledge and look forward to the continued study and deeper understanding of the complex genetic mechanisms behind floral regulation.

Summary

When I joined the duckweed community the interest was largely on sustainably providing clean water, food, and fuel as reflected in the Introduction and the Conclusion. Then a

fellow student of Dr. Messing, Dr. Wang, published the *Spirodela* 7498 genome and transcriptome (Wang et al. 2014a). My thesis project was to complement these with the *Spirodela* degradome, and investigate the relationship of the miRNAs 156 and 172 to the neotenous nature and rare flowering in the Lemnaceae. While characterizing the *Spirodela* miRNAs and their targets as described in Chapter 1, both Dr. Tang and Dr. Myers noticed a lack of 24nt siRNAs leading to an analysis of the RdDM pathway and transposons. This research combined with editing The Duckweed Genomes textbook led me to the write Chapters 2, 3, and 4 on small RNAs, transcriptomics, and repetitive DNA elements respectively. While studying the miRNAs and genomics, I developed flowering protocols in order to study the relationship between miR156 and 172 and floral regulation. As described in Chapter 5, I developed *Wolffia microscopica*, *Lemna gibba*, *Lemna minor*, and *Spirodela polyrhiza* flowering protocols to enable further study. While this was a sparse field of research in 2013, my work seen here, along with that of other scientists, has advanced the field significantly in the last 6 years.

References

An D, Li C, Zhou Y, et al (2018) Genomes and Transcriptomes of Duckweeds. Front Chem 6:. doi: 10.3389/fchem.2018.00230

Axtell MJ, Meyers BC (2018) Revisiting criteria for plant miRNA annotation in the era of big data. Plant Cell 30:tpc.00851.2017. doi: 10.1105/tpc.17.00851

Cao HX, Fourounjian P, Wang W (2018) The Importance and Potential of Duckweeds as a Model and Crop Plant for Biomass-Based Applications and Beyond. In: Handbook of Environmental Materials Management. Springer International Publishing, Cham,

pp 1–16

Dai X, Zhuang Z, Zhao PX (2011) Computational analysis of miRNA targets in plants: current status and challenges. Brief Bioinform 12:115–121. doi: 10.1093/bib/bbq065

Fourounjian P, Tang J, Tanyolac B, et al (2019) Post-transcriptional adaptation of the aquatic plant *Spirodela polyrhiza* under stress and hormonal stimuli. Plant J. doi: 10.1111/tpj.14294

Fu L, Huang M, Han B, et al (2017) Flower induction, microscope-aided cross-pollination, and seed production in the duckweed Lemna gibba with discovery of a male-sterile clone. Sci Rep 7:3047. doi: 10.1038/s41598-017-03240-8

Lam E, Appenroth KJ, Michael T, et al (2014) Duckweed in bloom: The 2nd International Conference on Duckweed Research and Applications heralds the return of a plant model for plant biology. Plant Mol Biol 84:737–742. doi: 10.1007/s11103-013-0162-9

Landolt E, Kandeler R (1987) Biosystematic investigations in the family of duckweeds (Lemnaceae), Vol. 4: the family of Lemnaceae-a monographic study, Vol. 2 (phytochemistry, physiology

Meyers BC, Axtell MJ, Bartel B, et al (2008) Criteria for annotation of plant MicroRNAs. Plant Cell 20:3186–90. doi: 10.1105/tpc.108.064311

Michael TP, 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. doi: 10.1111/tpj.13400

Singh V, Roy S, Singh D, Nandi AK (2014) Arabidopsis FLOWERING LOCUS D influences systemic-acquired-resistance-induced expression and histone

modifications of WRKY genes. J Biosci 39:119–126. doi: 10.1007/s12038-013-9407-7

Slovin JP, Cohen JD (1988) Levels of Indole-3-Acetic Acid in Lemna gibba G-3 and in a Large Lemna Mutant Regenerated from Tissue Culture'

Sree KS, Maheshwari SC, Boka K, et al (2015) The duckweed Wolffia microscopica: A unique aquatic monocot. Flora - Morphol Distrib Funct Ecol Plants 210:31–39. doi: 10.1016/J.FLORA.2014.10.006

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. doi: 10.1038/ncomms4311

Wang W, Messing J (2015) Status of duckweed genomics and transcriptomics. Plant Biol 17:10–15. doi: 10.1111/plb.12201

Wang W, Messing J (2011) High-Throughput sequencing of three Lemnoideae (duckweeds) chloroplast genomes from total DNA. PLoS One 6:. doi: 10.1371/journal.pone.0024670

Wang W, Wu Y, Messing J (2012) The Mitochondrial Genome of an Aquatic Plant, Spirodela polyrhiza. PLoS One 7:e46747. doi: 10.1371/journal.pone.0046747

Wang W, Wu Y, Messing J (2014b) RNA-Seq transcriptome analysis of Spirodela dormancy without reproduction. BMC Genomics 15:60. doi: 10.1186/1471-2164-15-60

Zhao H, Appenroth K, Landesman L, et al (2012) Duckweed rising at Chengdu: Summary of the 1st International Conference on Duckweed Application and Research. Plant Mol Biol 78:627–632. doi: 10.1007/s11103-012-9889-y

Ziegler P, Adelmann K, Zimmer S, et al (2015) Relative in vitro growth rates of duckweeds (Lemnaceae) - the most rapidly growing higher plants. Plant Biol 17:33–41. doi: 10.1111/plb.12184

1 Importance of Duckweeds in Basic Research and Their Industrial Applications

Paul Fourounjian, Tamra Fakhoorian and Xuan Hieu Cao

Abstract

The Lemnaceae family, commonly called duckweeds, is 37 species of the smallest and simplest flowering plants found floating on nutrient-rich waters worldwide. Their small size and rapid clonal growth in aseptic conditions made them a stable and simple model for plant research especially from 1950 to 1990, when they were used to study plant physiology and biochemistry including auxin synthesis and sulfur metabolism. Duckweed research then saw a resurgence in 2008 when global fuel prices rose and the US Department of Energy funded the sequencing of the *Spirodela polyrhiza* genome. This launched not only the genomic investigations detailed in this book, but the regrowth of duckweed industrial applications. Thanks to their ability to quickly absorb nitrogen, phosphorous, and other nutrients while removing pathogens and growing at a rate of 13–38 dry tons/hectare year in water treatment lagoons, scientists are currently exploring ways that duckweed can convert agricultural and municipal wastewater into clean water and a high-protein animal feed. The potential of these plants for phytoremediation of heavy metals and organic compounds also allows the possibility to clean the wastewater from heavy industry while providing biofuels and even plastics. Finally, thanks to their superb nutritional profile *Wolffia* species grown in clean conditions promise to become one of the healthiest and most environmentally friendly vegetables. Given the importance of these incredible plants, it is no wonder researchers are investigating the genetic mechanisms that make it all possible.

This chapter was revised and significantly expanded upon, with the guidance of T. F., from the chapter "The Importance and Potential of Duckweeds as a Model and Crop Plant for Biomass-Based Applications and Beyond," in the Handbook on Environmental Materials Management, which X. H. C. and P. F. wrote for Springer Nature a year ago (Cao et al. 2018). We hope this chapter thoroughly explains non-genomic research and application topics, especially for those who are unfamiliar with the family.

P. Fourounjian (✉)
Waksman Institute of Microbiology,
Rutgers University, Piscataway 08854, USA
e-mail: pjf99@scarletmail.rutgers.edu

T. Fakhoorian
The International Lemna Association,
Mayfield, KY 42066, USA

X. H. Cao
Institute of Biology/Plant Physiology,
University of Halle-Wittenberg
Martin-Luther-Weinbergweg,
10, 06120 Halle, Germany

X. H. Cao et al. (eds.), *The Duckweed Genomes*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-11045-1_1

1.1 Introduction

Duckweed (known as monocotyledon family *Lemnaceae* or recently classified as subfamily *Lemnoideae* in the arum or aroid family *Araceae*) is a small group of aquatic plants with only five genera (*Spirodela*, *Landoltia*, *Lemna*, *Wolffia*, and *Wolffiella*) and 37 species (see Landolt 1986; Nauheimer et al. 2012; Sree et al. 2016). Except for *Wolffiella* (commonly named as bogmat) that is restricted to the Americas and Africa, species of other duckweed genera occur around the whole world. Although highly adaptable across a broad range of climates, most diverse species of duckweed appear in the subtropical or tropical zones. Duckweed species tend to be associated with nutrient-rich or eutrophic freshwater environments with quiet or slow-moving flow. However, they are extremely rare in deserts and are absent in the cold polar regions (Arctic and Antarctica).

Duckweed species are the smallest flowering plants with minute sizes from 0.5 mm to less than two cm (Landolt 1986). Species of duckweed can be easily distinguished morphologically from species of any other flowering plants, even closely related aquatic plants, due to their highly reduced body structure. The leaflike body of the duckweed species, sometimes called a frond or thallus, is a modified stem with only few cellular differentiations (Fig. 1.1). The growth of duckweed vegetatively occurs by budding within the pouches or cavities of the basal sections of the fronds. Each daughter frond emerging from the pouch of mother bud already contains two new generations of daughter fronds. Therefore, under optimal conditions, the growth rate of duckweed is nearly exponential. The frond number of fast-growing species (e.g., *Lemna aequinoctialis*, *Wolffiella hyalina*, and *Wolffia microscopica*) almost doubles within 24 h (Ziegler et al. 2015; Sree et al. 2015b), presenting the fastest growing flowering plants. With a miniaturized body plan and such rapid growth leading to maximum fitness, duckweed has arguably been interpreted as an example of the hypothetical Darwin–Wallace Demon for the lifetime reproductive success (Kutschera and Niklas 2015).

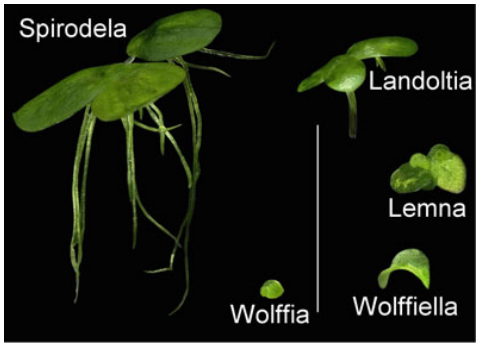


Fig. 1.1 Morphology of five representative species for duckweed genera. Spirodela: *Spirodela polyrhiza*; Landoltia: *Landoltia punctata*; Lemna: *Lemna minor*; Wolffiella: *Wolffiella lingulata*; Wolffia: *Wolffia arrhiza*. Bar: 1 cm

Only occasionally or very rarely, several species of duckweeds produce microscopic flowers in nature as well as under in vitro conditions (Fu et al. 2017; Schmitz and Kelm 2017; Sree et al. 2015a). In *Spirodela* and *Lemna* (belonging to the subfamily *Lemnoideae*), the flowering organs (1 membranous scale, 2 stamens, and 1 pistil) originate in the same pouches in which the daughter fronds are normally formed. In the subfamily *Wolffioideae* (consisting of *Wolffiella* and *Wolffia*), generative and vegetative reproductions are spatially separated occupying the floral cavity on the upper surface of the frond and the budding pouch, respectively.

Duckweed fronds are free floating on or near the surface of the water, often forming dense mats in suitable climatic and nutrient conditions. In unfavorable weather, such as drought or freezing winter seasons, in addition to flowering, several duckweed species are able to form special "resting fronds" (in the dormant phase) to persist until conditions return that can support growth. In place of a frond, the greater duckweed (*Spirodela polyrhiza*) produces a starch-rich tissue called a turion, which sinks to the bottom of the water. Turion production has been reported also for *Lemna turionifera*, *L. aequinoctialis*, *Wolffia brasiliensis*, *Wolffia borealis*, *Wolffia angusta*, *Wolffia australiana*, *Wolffia arrhiza*, *Wolffia columbiana*, and *Wolffia globosa*. These turions

do not grow any further but can germinate and start a new life cycle from the bottom of the water body or mud when the water temperature reaches about 15 °C. In addition, resting fronds of the ivy duckweed (*Lemna trisulca*) and *Wolffiella gladiata* with reduced air spaces can accumulate starch and still rather slowly grow on the bottom of the water, forming new but similar fronds. However, the common duckweed (*Lemna minor*), gibbous duckweed (*Lemna gibba*), *Lemna perpusilla*, and some strains of *Lemna japonica* produce starch-rich fronds that do not sink to the bottom of the water but are just pressed down under the ice cover during freezing temperatures. Interestingly, formation of turions as a survival and adaptive capacity of *S. polyrhiza* strains collected from a wide geographical range seems to be genetically determined and highly influenced by the mean annual temperature of habitats (Kuehdorf et al. 2013). Furthermore, the family displays significant inter- and intraspecies differences of cell physiology (e.g., starch, protein, and oil contents) together with duckweed potential for industrial applications (Alvarado et al. 2008; Appenroth et al. 2017; Hou et al. 2007; Mkandawire and Dudel 2005; Tang et al. 2015; Yan et al. 2013; Zhang et al. 2009).

Due to their small and abbreviated structures, morphological and physiological classification of the 37 duckweed species (*Spirodela*: 2 species; *Landoltia*: 1; *Lemna*: 13; *Wolffiella*: 10; *Wolffia*: 11) can be challenging. In the past decade, for species assignment as well as resolving intraspecies differences, several attempts have been carried out to employ molecular genotyping techniques, including random amplified polymorphic DNA (RAPD; Martirosyan et al. 2008), inter-simple sequence repeats (ISSR; Fu et al. 2017; Xue et al. 2012), simple sequence repeats (SSR; Feng et al. 2017), amplified fragment length polymorphism (AFLP; Bog et al. 2010, 2013), and DNA barcoding using plastid sequences (Borisjuk et al. 2015; Wang et al. 2010) or nuclear ribosomal sequences (Tippery et al. 2015). Although DNA barcoding using two plastidic barcodes aids in identifying most duckweed species (at least 30 among 37 species) in a quite simple and straight forward manner, combination of different techniques or using additional barcodes may help to unambiguously and economically assign remaining duckweed species.

The Lemnaceae family was one of the earliest model plants due to their ease of aseptic cultivation in the laboratory and simple morphology. The second volume of Landolt and Kandeler's 1987 monographic study contains 360 pages dedicated to the physiological research of the family in particular and plants as a whole (Landolt and Kandeler 1987). The professors who organized the first duckweed conference summed up the duckweed research stating that duckweeds were the main model for plant biology from 1950 to 1990, when Arabidopsis and rice were used for their sexual reproduction and applicability to terrestrial crops (Zhao et al. 2012). In that time, investigations of duckweeds revealed the tryptophan-independent synthesis of auxin (Baldi et al. 1991), translational regulation in eukaryotes (Slovin and Tobin 1982), and seven of the first stable plant mutants (Posner 1962). Today, physiological studies continue largely in the fields of circadian rhythm research, xenobiotic plant–microbe interactions, and phytoremediation and toxicology. Starting in 2011, a biannual series of international duckweed conferences in research and applications has connected and helped expand this research community and increased public awareness and recognition of duckweed economic and environmental importance (Zhao et al. 2012; Lam et al. 2014; Appenroth et al. 2015). Together with the completion of the *Spirodela* genome in the year 2014 and rapid advances in sequencing technologies, this resurgence of interest has resulted in a proliferation of genome and transcriptome sequences for duckweed species and ecotypes discussed in the remainder of this book.

One of the largest fields of duckweed research is ecotoxicology, where the widely distributed *Lemna* species *minor* and *gibba* serve as model plants to determine the effect of a compound on an ecosystem. These growth tests have been standardized in the International Organization for Standardization's protocol ISO 20079 which

handles environmental samples and the Organisation for Economic Co-operation and Development's assay OECD 221, which was developed for specific chemicals and compounds (ISO 2005; OECD 2006). Both are seven-day growth rate tests, which use different media, to measure the effective concentration of the substance, or EC_{50}, where the growth rate by frond count or frond area is half of the control. These tests date back to the 1970s and have surveyed the effects of heavy metals, pharmaceuticals, various pesticides and organic compounds, and even radioactivity on Lemna growth rate and health, helping us quickly asses and monitor environmental safety.

In order to perform major gene function studies, as well as to improve duckweed agronomic performance (Cao et al. 2016), it is required to establish an efficient system for genetic manipulation and transformation. Several stable transformation protocols for *Lemna* (Chhabra et al. 2011; Yamamoto et al. 2001), *Landoltia* (*Spirodela oligorrhiza*; Vunsh et al. 2007), and *Wolffia* (Boehm et al. 2001; Khvatkov et al. 2015) using either *Agrobacterium*-mediated or biolistic gene transfer together with a recent gene-silencing platform in *L. minor* (Cantó-Pastor et al. 2015) have been described, providing the means to further develop gene/genome-edited duckweed as a powerful biomanufacturing platform.

1.2 Current State of Duckweed-Based Applications

1.2.1 Historical

For centuries, people have seen the role duckweed can play in their food production. Perhaps by observing their livestock consume duckweed species, small-scale farmers in Southern Asia started feeding duckweed, often fresh as a portion of the diet, to their fish, ducks, chickens, pigs, and goats. In addition to animal feed, the people of Thailand, Laos, and Cambodia have consumed wild harvested and farmed *Wolffia*, mainly *globosa*, rinsed, and then incorporated into soups, salads, sauces, and a wide variety of foods (Bhanthumnavin and Mcgarry 1971). If the *Wolffia* is not cooked in with other ingredients, it is generally briefly boiled, thereby selecting a duckweed species without harmful calcium oxalate crystals and killing potential pathogens. Recently, farmer education programs in Guatemala, Indonesia, and across the globe have improved the use of duckweed to treat manure while using it as a fertilizer and expanded the practice within Asia and around the world, especially in Central America where a consortium of ~200 small-scale farmers grows duckweed and tilapia. It is estimated by the executive director of the International Lemna Association that the thousands of small-scale farmers collecting wild duckweed or growing it on site for human or animal consumption are currently a greater part of the duckweed applications by volume than the large-scale, higher tech companies.

1.2.2 Water Treatment

As global population rises, so does demand in clean water supply and wastewater treatment systems. While developed nations have often relied on a combination of aerobic bacteria degradation and chemical treatment in activated sludge systems, a variety of natural treatment systems have been growing in popularity for their often 50% lower capital and operating costs, ability to recapture nitrogen, phosphorous, and other valuable nutrients, and in some cases convert them into appropriate products. The main downsides of these natural treatment systems are their larger land requirements (up to 5 m^2/person), poorer performance at cold temperatures, and the requirement of knowledgeable and diligent staff to manage ecosystems through toxic wastewater streams, harsh weather, etc. All this indicates that natural treatment systems such as constructed wetlands are ideal in rural locations, especially of tropical countries, precisely where many of the 2.5 billion people without access to sanitary wastewater treatment live (Zhang et al. 2014b).

While a variety of plants have been used effectively in constructed wetlands, we will focus here on the 37 species of the Lemnaceae family for their global distribution, tolerance of ammonia, heavy metals, other stresses, high yield of biomass (especially at 20–30 °C), ease of harvest, high protein and starch content, and range of uses. As seen in Fig. 1.2, duckweed can treat agricultural, municipal, and even industrial wastewater streams into clean non-potable water, and a biomass that can be used for feed applications, or fuel if it was used to treat harmful industrial wastewater.

The classic example of a duckweed treatment system and feed application would be the Mirzapur Bangladesh hospital wastewater facility, which was designed by the PRISM group, monitored from 1989 to 1991, and thoroughly described in the book "Duckweed Aquaculture: A New Aquatic Farming System for Developing Countries" (Skillicorn et al. 1993). The book describes a pilot plant facility with clean effluent water of 1 mg/l BOD (biological oxygen demand) and 0.03 mg/l of both NH_3 and P, an annual duckweed dry yield of 13–38 metric tonnes/hectare year (t/ha yr), carp production of 10–15 t/ha yr, and positive economic analysis of duckweed, duckweed-fed carp, and duckweed-fed tilapia farming. As of 2015, the Mirzapur facility was still operational, profitably treating wastewater above the standards of any US city, while providing fresh, pathogen-free, sustainably farmed fish. Professor Zhao Hai's group from Chengdu Institute of Biology, Chinese Academy of Sciences, also has extensive records from their pilot plant at Dianchi Lake, in subtropical Yunnan, China (Fig. 1.2). In a year-long comparison between duckweed and water hyacinth, they found a higher total yield for water hyacinth (55 compared to 26.5 t/ha yr) and a higher nitrogen removal rate, partially due to denitrifying bacteria. However, they chose to focus on duckweed for its consistent year-round production, ~33% protein content, and biofuel potential as a low lignin, high starch ethanol feedstock (Zhao et al. 2014). In follow-up

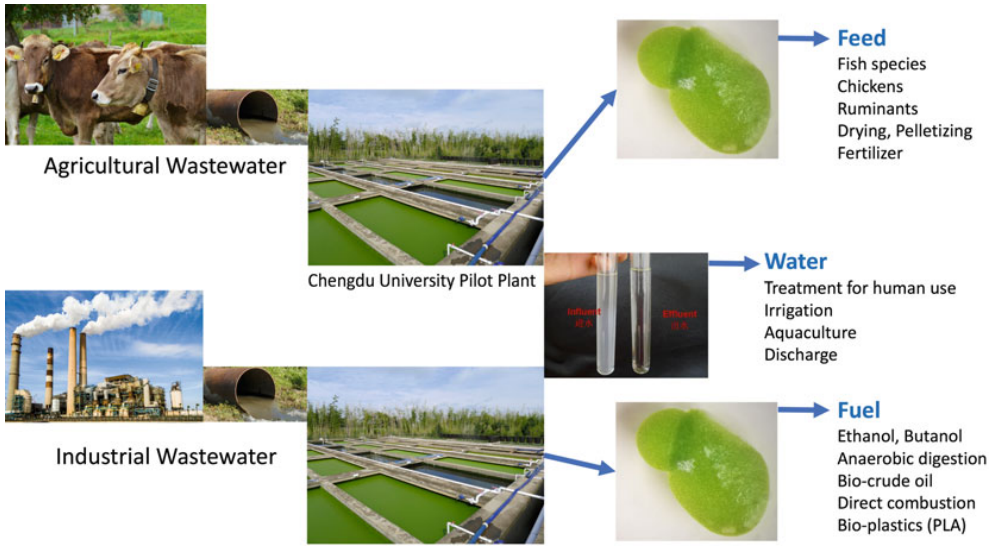


Fig. 1.2 Flowchart of duckweed wastewater treatment and biomass application. Farm and factory examples, and the pilot plant at Chengdu University. Their influent agricultural wastewater and effluent water in the two test tubes. Mother and daughter frond of *Lemna minor*. While duckweed can be grown on agricultural or industrial wastewater and used for feed or fuel, the applications of the biomass are determined by the input water source and local regulations. *Source* Hai Zhao, Chengdu University, China

experiments, they found they could increase duckweed starch content from 9.5 to 40% through 11 days of growth on clean water, and that a hydraulic residence time (HRT) of 6 days achieved their treatment standards and optimized the *Landoltia punctata* starch yield above maize and wheat to 13.9 t/ha yr. Considering that these are experimental water treatment plants, their duckweed yield is expected to rise with further optimization, or in more intensive cultivation. For their size, length of study, and abundance of publicized information, these two facilities stand as prime examples to study duckweed's water treatment capabilities, yield, and applications in practice.

If a wastewater stream comes from an industrial point source or a large municipality, it likely has persistent chemical compounds, such as textile dyes and metalworking fluids, or bioaccumulating heavy metals in it. There is a large body of academic evidence illustrating the potential of duckweed and other plants to treat wastewater from cities, tanneries, mines, metalworking shops, and textile mills by degrading compounds like pharmaceuticals and antibiotics, and accumulating phenols along with heavy metals (van der Spiegel et al. 2013). Rezania et al. reviewed the heavy metal absorption of 5 different plant species and described 19 studies evaluating *Lemna minor* and *gibba* as moderate or hyperaccumulators, often concentrating metals over 400-fold, depending on the metal and circumstance; even when used as a dried powder (Rezania et al. 2016). A table of 10 studies illustrated removal efficiencies of Cu, Cd, Pb, Zn, and 9 other metals, with the lowest being 29% and the majority being over 70%. In these cases, duckweed and its microbial communities can treat a variety of harmful wastewater streams and then be utilized outside of the food supply for biofuel applications to further concentrate the metals.

1.2.3 Bioenergy

While these applications have been researched academically, few have been practiced in large scale. The simplest bioenergy application would be direct combustion of dried duckweed, possibly as a drop-in fuel for a trash incinerator or coal-fired power plant. This would concentrate heavy metals in the smoke, which could be scrubbed, and ash for proper disposal, or even encapsulated reuse in concrete or gypsum as coal ash is in the USA. A second relatively simple option would be anaerobic digestion to produce methane. Conveniently, many municipal wastewater treatment plants already have anaerobic digesters to treat sludge, and the liquid digestate has been well studied as a fertilizer for duckweed ponds. A duckweed and pig manure mixture increased gas production 41% in comparison with pig manure, while the increased production from cow manure tapered after a 2% inclusion of duckweed (Cui and Cheng 2015).

Another possibility is pyrolysis of dried biomass or hydrothermal liquefaction (HTL) of wet biomass. Both processes are similar, yet we will focus on HTL since it conveniently avoids drying the ~90% water content duckweed biomass. In HTL, biomass and water processed at 200–400 °C and 50–200 times atmospheric pressure for 10–90 min to create aqueous solutes, H_2, CO_2, and CH_4 gasses, high molecular weight bio-char, and bio-crude oil with 95% of the energy content of diesel (Zhang et al. 2014a). A wide variety of feedstocks from algae to wood and to sewage sludge can be used, separately, or mixed, and each requires significant testing to optimize, which is likely why there are no large-scale HTL operations at the present day. The algae can be converted to crude oil with a 26–68% yield depending on the conditions, yet all the crude oil tends to have a high water content and require hydro-deoxygenation to dewater it thereby matching the stability and viscosity of petroleum crude oil. A wide range of molecules can be created and isolated so there is petrochemical potential as well. This option is interesting for its theoretical ability to match the wide variety of the crude oil applications in a carbon neutral manner and the ability to produce in hours what naturally takes ~150 million years.

The most versatile and best studied application of potentially harmful duckweed is fermentation of the starch, which can be accumulated at

rates varying from 46% after 5 days to 31% after 10 days of nutrient starvation and fermented at 95% efficiency after enzymatic pre-treatment. These fermentation processes also create protein-rich distiller's grains, which can be used as an animal feed supplement if they are not concentrating heavy metals. As the first commercially viable example of ethanol fermentation, the Andrew Young Foundation conducted a private research trial using the ecosystem technology, produced by resource recovery experts Greenbelt Resources Corporation, which was presented in a feasibility study report conducted by an independent party Agregy and submitted to the USDA in 2017. With successful feasibility determined, the foundation created a corporation called Duckweed Days LLC, which partnered with Greenbelt Resources to conduct a pilot system development project in Paso Robles, California, USA, in 2018. Leveraging its farming and agricultural expertise as well as its engineering prowess, Greenbelt has developed a species agnostic prototype cultivation, harvesting and processing system. For the biorefining of the cultivated duckweed, Greenbelt's proprietary, partially AI-operated modular machinery uses membrane filtration to produce anhydrous bioethanol that can be sold as a fuel or solvent, plus chemically safe distillers' grains that can be used as animal feed or a nutritious protein concentrate.

Ethanol is not the only fermentation product, since *Clostridium acetobutylicum* bacteria can convert the sugars of 32% starch content duckweed into a mixture of 68% butanol, with acetone and ethanol coproducts (Cui and Cheng 2015). Ethanol can of course be blended into gasoline at rates up to 10% or 85% for certain flex-fuel vehicles, while significantly more expensive butanol behaves very similarly to gasoline. Finally, the Argentinian company MamaGrande experimented with fermentation as a means to generate lactic acid for polymerization into PLA. Polylactic acid, or PLA, is a renewable and degradable plastic produced by enzymatically digesting starch to glucose, fermenting the glucose to lactic acid, and then purifying and polymerizing it. At the moment, anaerobic digestion and ethanol fermentation appear to be the best studied options, while fermentation is the only biofuel in full-scale commercial application.

1.2.4 Animal Feed

Most agricultural wastewater and certain domestic wastewater streams will have undetectable or legally permissible levels of heavy metals, enabling a design where duckweed can recycle nutrients back into the food supply, provided it is monitored for heavy metals and other hazards, and legally approved. Agricultural wastewater, which can come from greenhouses, livestock barns, anaerobic digesters, or even food processing facilities, is often heavy metal "free" and therefore diluted down to 20–50 mg/l total nitrogen for optimal duckweed growth. Considering the pilot plant examples above, and publicly posted information from Paul Skillicorn's Agriquatics Blog, we see the following steps for domestic wastewater treatment (Fig. 1.3).

First, solids will be removed by screening and then primary settling lagoons or laminar flow systems and hydrocyclones, possibly for anaerobic digestion. Secondly, there may be a buffer lagoon or lagoons, which treat soaps and other chemicals that may harm duckweed or its downstream applications. Third will be the duckweed farm portion, where a diluted influent with NH_3 concentrations of 10–30 mg/l, BOD of 15–30 mg/l, and pH from 6.0 to 7.0 will fertilize rapidly growing high-protein duckweed biomass. Fourth, ponds with slower growing, high starch content duckweed can polish wastewater as the final cleaning step. Here, once nitrogen has been depleted heavy metals will be accumulated, with the majority of municipal effluents producing duckweed passing US food and feed safety standards. HRT can vary from 6 to 15 days depending on environment, degree of effluent recirculation, and treatment standards. For example, the Mirzapur duckweed ponds reduced NH_3 from 32 to 0.03 mg/l. This high HRT increases the footprint compared to a conventional system, while providing resilience against heavy rains or community crashes that

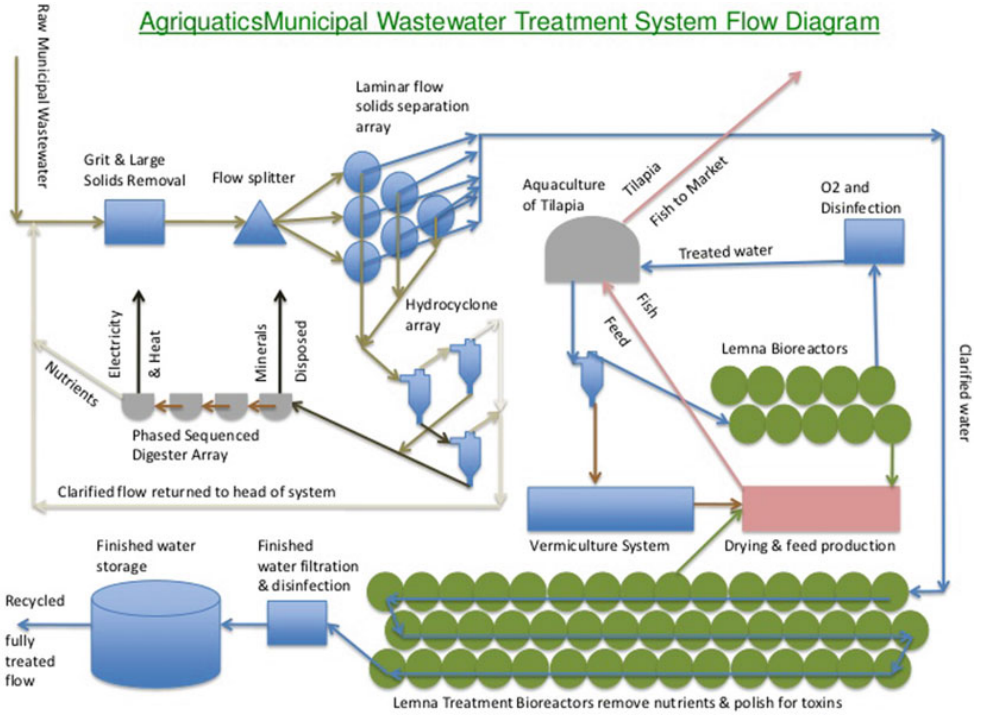


Fig. 1.3 Agriquatics wastewater treatment and aquaculture diagram for Olmito, Texas. Proposed blueprint for a municipal treatment facility designed by Agriquatics. The systems start with solids removal through laminar flow separators and hydrocyclones, and sends solids to an array of bacterial digesters, which act as an improved anaerobic digester similar to conventional methods. A series of duckweed ponds remove solutes, and their circular shape facilitates central harvesting. Water is then filtered and disinfected with conventional methods. Duckweed biomass can be tested, sterilized, and converted to Tilapia feed, while aquatic worms and duckweed purify water and provide food for the Tilapia. *Source* Paul Skillicorn, Lyndon Water Limited, UK, https://paulskillicorn.wordpress.com/about/

occasionally overwhelm smaller systems. Throughout this process, pathogens are largely killed off, evaporation is reduced 33%, mosquito populations are reduced, and odors are partially suppressed by the duckweed mat (Goopy and Murray 2003; van der Spiegel et al. 2013). Finally, polished water and duckweed biomass can be sterilized and utilized. In a budget estimate for a medium-sized treatment plant in Texas, USA, Agriquatics illustrated that their treatment system would have 52% of the capital and 66% of the total annual costs of a conventional oxidative ditch system. This budget completely excluded the proposed tilapia aquaculture system that had been proved profitable in Mirzapur. To make larger duckweed treatment systems, even more cost-effective Agriquatics has positioned them on the outskirts of cities to benefit from rising real estate value as the city grows, while providing greenspace and reducing pipe distance.

Since duckweeds have been a traditional feed for fish and poultry in South East Asia for centuries, they are now being quantitatively investigated in a variety of feed trials. In many cases, NH_3-tolerant *Lemna* and *Spirodela* species are used and harvested with dry weight protein contents of 20–30%. To minimize pathogen transfer, feed trials often use effluent from one species to grow duckweed, which is then fed to a different

species. While ozone and microwave disinfection were used in the long-term commercial operation of Mirzapur, many feed trials have simply washed with water, or just harvested duckweed, and have no report of pathogens (Goopy and Murray 2003; Skillicorn et al. 1993). Surprisingly, several studies have found duckweed, including samples from hospital wastewater to be safe as chicken and fish feed with regard to *E. coli* and *Salmonella*, with no significant differences in the quantity of five different pathogens in chickens fed on duckweed compared to control, presumably due to the severe pathogen reduction seen in wastewater treated by duckweed and its associated microbial communities (Goopy and Murray 2003; van der Spiegel et al. 2013). The feed trials often use dried duckweed as a percentage of complete commercial feed or substitute it for a percentage of the soybean or fishmeal component, with duckweed performing very similarly to soy in the case of chickens, ducks, and fish, up to a point where it is suspected that oxalates or other anti-nutritives inhibit growth (Goopy and Murray 2003; Skillicorn et al. 1993). For tilapia, inclusion rates of 30% were found equivalent to control, and 30% replacement of fishmeal component was seen as the most cost effective (Goopy and Murray 2003). An ecosystem of 5 different carp species or the grass, catla, and mirror carp and tilapia species individually can be fed on a pure duckweed diet, with a carp yield of 10–15 t/ha yr (Skillicorn et al. 1993). Duckweed was found to be beneficial in replacing ~15% of the soybean meal in the feed for chicks or broilers, and 40% in the case of laying hens (Goopy and Murray 2003; Skillicorn et al. 1993). In some cases, pig saw decreased growth in response to small inclusion rates of duckweed, while the Mong Cai piglets of Vietnam had higher growth rates than their Large White counterparts due to higher nitrogen digestibility (Goopy and Murray 2003; Gwaze and Mwale 2015). Finally, ruminants have shown promising results with high nitrogen digestibility in merino sheep, and cattle consuming and effectively digesting up to 10% of their weight in dried duckweed per day (Goopy and Murray 2003). Taken together, these results show the potential of duckweed to reduce the environmental impact of livestock by recycling nitrogen phosphorous and other nutrients that currently cause eutrophication, while partially replacing human edible soy and non-sustainable fishmeal. Furthermore, recycling wastewater to grow animal feed has been shown in several economic analyses to raise farmer income, especially in developing countries.

Considering the economic and environmental benefits, and the success of duckweed as feed for a variety of livestock species, there will likely be a rapid expansion of the duckweed agricultural sector and its use as a sustainable animal feed. In the FAO's 2012 estimates, global demand for non-fish animal protein is expected to increase at 1.3% per year till 2050, with the largest growth of 4.2% in South Asia, with similar numbers in the 2030 projection (Alexandratos and Bruinsma 2012). Roughly, half of this increase is expected to be as poultry (OECD/FAO 2017). Additionally, the largest increase in animal protein supply will be aquaculture, which was ~17% of global fish supply in 1990, grew largely in Asia between 4 and 10% per year, and is forecasted to exceed the global catch in 2020 (OECD/FAO 2017). The livestock sector is, however, very feed, land, and water intensive, and all reports stress the need to reduce the environmental impact particularly through improving the feed supply. With their ability to treat agricultural wastewater on non-arable land and provide an affordable protein-rich feed, a greater number of farmers are turning to duckweeds as a cheap sustainable feed source. There are currently several commercial ventures and hundreds of thousands of small-scale farmers growing duckweed primarily in Asia and Central America feeding tilapia, ducks, chicken, and pigs. Since they are sustainably feeding the livestock species in the regions where the FAO expects the largest growth in the world, it is natural to expect this industry to grow. While working with farmer education programs in Guatemala and Indonesia, the ILA, International Lemna Association, has seen an increase in educational activities for small-scale farmers and 20% more businesses seeking to enter the industry for the past 7 years (Table 1.1, Director of the ILA).

Table 1.1 Summary of the duckweed applications in use or development and the major companies working on them

Application	Company (if blank academic)	Genera
Human food	Hinoman, Green Onyx, Parabel	Wolffia, Lemna
Protein isolate	Plantible, Parabel, CAIS	Lemna
Livestock	Many small-scale farmers	Lemna, Spirodela, others
Conversion chemicals	MamaGrande	Lemna
Wastewater treatment	MamaGrande, CAIS	Mixture
Space life support	Space Lab Technologies	Lemna, Wolffia
Isolation chemicals	CAIS	Mixture
Transformation		
Specialty (cosmetics, pets, tea)		
Biofuels or energy	Greenbelt Resources	

1.2.5 Human Nutrition

The high growth rate, protein content, and success in a variety of animal feed trials naturally beg the question of whether duckweeds could be a healthy and environmentally friendly food for humans? As previously stated, the *Wolffia* genus of the duckweed family has been traditional cuisine in Thailand, Burma, and Laos for centuries, since they lack the kidney stone forming calcium oxalate crystals found in the other genera. At the time of writing, there are three large companies producing *Wolffia* or *Lemna* for human consumption, namely Hinoman with greenhouse precision agriculture cultivation, Parabel with open pond *Lemna* cultivation and protein extraction, and Green Onyx, which has developed robotic farming systems that can dispense *Wolffia* on demand. Due to their successful scale-up since their founding in 2010, and abundant public information, we will focus on the Israeli company Hinoman here. They currently grow *Wolffia* (aka Mankai™) on formulated, clean water media in greenhouses with automated energy-efficient climate control and harvesting systems operated by their cultivation algorithm. Through this system, they are able to grow a pesticide- and herbicide-free vegetable year-round, with a fraction of the water used in cultivation of soy, spinach, or kale, (http://www.hinoman.biz/what-we-do/). Their product is stable with approximately 25% carbohydrate content, 45% protein content, and a complete and bioavailable amino acid profile such as egg or soy, with a higher PDCAAS than soy. They have currently conducted three publicly visible clinical trials demonstrating the protein and iron bioavailability, as well as the mitigating effect on Glycemic Index of their *Wolffia*, and posted multiple recipes for their product, which will soon be made available to consumers.

Furthermore, compared to kale *Wolffia* is more abundant in most minerals and vitamins A, B2, B12, and E, which survive the gentle drying process. An extensive academic investigation of the species *Wolffia microscopica* confirmed the high mineral content and that the protein (~25% of dry weight) exceeded WHO recommendations, while finding abundant antioxidants and a high omega-3 content (Ω6/Ω3 ratio is 0.61) for the relatively scarce lipids (Appenroth et al. 2017, 2018). Fresh, or dry powdered *Wolffia*, with a neutral taste, can be juiced, consumed fresh, or incorporated into breads, pastas, and sports nutrition products (Fig. 1.4). With supporting data from academic laboratories, records of historical consumption, and thorough testing of their product for harmful metals and oxalates, Hinoman and Green Onyx were able to achieve the generally recognized as safe (GRAS) status for the *Wolffia* species *arrhiza*

Fig. 1.4 *Wolffia* fortified breads. Hinoman has tested the addition of *Wolffia* to multiple food and beverage products. Note the retention of the chlorophyll pigments throughout the baking process, and unchanged texture and leavening of the bread. *Source* http://www.hinoman.biz

and *globosa* in the USA in 2015 and 2016, respectively. Now, with South East Asia, Israel, and the USA recognizing select *Wolffia* species and *Lemna minor* as human food the crop and its producers have significant potential to grow and provide abundant plant protein for minimal land, water, and energy inputs.

With their small size, growth rate, aquatic lifestyle, and high protein content, the duckweeds provide a promising new crop to grow and an assortment of cultivation and preparation processes for human consumption. Given the growing consumer demand for novel vegetables and healthy leafy greens, companies like Hinoman and Green Onyx grow these tiny nutritious vegetables in clean environments with robotic systems and plan to bring them into our grocery stores and homes both frozen and fresh. The global market for plant-based protein (57% of total global protein supply; Henchion et al. 2017) has been growing at 12.3% per year from 2013 to 2016, and is anticipated to grow 6.7% annually from 2018 to 2021, when it is anticipated to exceed 1 billion USD. Seeing this demand for protein, Plantible Foods is developing a gentle protein isolation process using *Lemna* in order to create a colorless, tasteless protein isolate with the physical properties of egg whites to create a vegan product that can finally match the textures of many beloved foods. Additionally, Parabel has chosen to sell its duckweed product as a high protein powder. Given the expansion of the plant protein market in both whole and extract formats, and their current progress, we expect these and other companies to increase in size, dramatically, providing a healthy and environmentally friendly alternative to less efficient protein sources.

As seen above, duckweed wastewater treatment performs well in tropical and subtropical environments, requires more land, yet less funding to operate, and even has the potential to generate revenue if duckweed biomass and clean effluent are well utilized. Agricultural wastewater can be converted into animal feed supplements, while industrial effluents can be treated to degrade or accumulate harmful chemicals and heavy metals while producing bioenergy, according to the laws of the land. The duckweed has proven to be a suitable food source for both humans and livestock, and will likely play an expanding role in meeting future food demands. There is plenty more to learn at the International Lemna Association and The Charms of Duckweed Web sites, and in the Duckweed Forum newsletter. Given the tremendous diversity of species, strains, environments, and applications, along with the relatively recent commercial interest, duckweed researchers are continuously rediscovering what is possible and practical.

1.3 Future Prospects in Duckweed-Based Applications

The field of duckweed applications has made tremendous progress recently. For centuries, it was harvested from wild ponds and used as a vegetable or animal feed in certain parts of the world, and largely in the twenty-first century humans have recognized the potential of these tiny overlooked plants and started applying them to wastewater treatment, and larger-scale animal feed and human nutrition operations. While certain applications are mature enough for large-scale deployment, those discussed below include important developing technologies. In terms of scale and possibility of duckweed

applications, we believe in 2019 we are still looking at the tip of the iceberg.

Due to the success and low prices of other crops, many companies growing duckweed are focused on high-tech, high-value applications to avoid commodity markets. Similar to protein extracts, several high-value products, like sugars, antioxidants, and oils, are being extracted from duckweed biomass in academic and commercial research laboratories. Appenroth et al. conducted a thorough investigation of *W. microscopica* and found a complete plant protein, roughly 150 mg carotenoids and 22 mg of tocopherols/gram dry weight, and an oil profile of 61% poly-unsaturated fatty acids with a high content omega-3 s and a phytosterol content minimum fivefold higher than common plant oils, presenting several healthy, high-value compounds that may be extracted (Appenroth et al. 2017). After or without extraction of certain compounds or protein, biomass can be converted to other products, for example MamaGrande's research in converting starch to sugar, and then polylactic acid valued at ~$2000 USD/ton. After enzymatically converting starch to sugar, the sugars can be fractionated and sold, or converted to levulinic, formic, or succinic acid (Liu et al. 2018). Pyrolysis and HTL discussed above can be used to create bio-char, gases, and a bio-crude oil. A subset of a single sample of duckweed derived bio-crude oil contained over 100 distinct compounds, mainly ketones, alcohols, fatty acids, and cyclic compounds (Duan et al. 2013). When considering the variables of biomass, solvents, temperature, pressure, and time HTL, pyrolysis can be adjusted to offer countless compounds that can be created and fractionated. Finally, there are a variety of other high-value application niches that duckweed can be used for including tea, cosmetics, pet food, and aquarium plants, which have been tested on small scale and may develop further. Major crops such as corn and soy have been used as feedstocks for hundreds of uses including food-thickening agents, cosmetics, construction adhesives, and ink. It is therefore reasonable to expect that as duckweed abundance grows there will be a greater number and variety of applications.

Another sector where duckweed species will likely play an expanding role is water reclamation and supply. In 2018, the Duckweed forum issue 22 described 23 companies in 9 countries, with 4 each working in water quality testing and water treatment (Shoham 2018). Provided the perpetual rise of water pollution and increased testing, and the roughly 50% lower capital and operating costs of duckweed (Skillicorn 2013) and constructed wetland (Zhang et al. 2014b) treatment systems compared to their bacterial counterparts, these industries are expected to grow, likely more so in developing countries. Sadly, 14 years of satellite observations reveal decreasing clean water availability across the world and in heavily populated areas like California, the Middle East, Northern India, and Northern China where groundwater is being depleted (Rodell et al. 2018). Many regions suffer clean water scarcity for at least 1 month of the year resulting in inadequate supply for people as well as agricultural losses. Duckweed treatment systems to reclaim water, as well as water efficient duckweed crops, with many other measures, might be utilized in these and other regions to increase supply. Similar to water reclamation, there is a lesser known need for phosphorous reclamation, since our current practice is to mine and refine phosphorous deposits, fertilize our crops, and then let the phosphorous run directly off of fields and into the ocean, or through our wastewater treatment systems into the ocean where it causes eutrophication damage like the Gulf of Mexico hypoxic zone. Economically mineable, organically available phosphorous is expected to be scarce by 2050 or 2100, and production might decline by 2030 raising its price possibly beyond the reach of poorer farmers (Childers et al. 2011). Fortunately, phosphorous can be recycled by better farming practices or by using more aquatic plants and other methods to recapture more than the current rate of 50% from human wastes. While phosphorous is a critical macronutrient and prime example, many other fertilizers have similar life cycles and would follow the phosphorous in any reduce, recapture, and reuse applications. Given the water and fertilizer scarcities this century will

likely pose to billions of people, we sincerely hope that duckweed-based water treatment systems and many other water and nutrient reclamation technologies will be applied at larger scale to "close the loop" and avoid scarcity.

One of the earliest companies to work with duckweed, Biolex Therapeutics, saw the rapidly growing high protein biomass of *Lemna* as a great expression platform for transgenic proteins. They produced several complex antibodies, including one to target Leukemia, and trademarked the term Plantibodies™, yet sadly went out of business. Since their closure, there have been improvements in the transgenic expression within several duckweed species. There have even been academic papers reporting over 20 transgenic therapeutic proteins in duckweed reaching as high as 7% of total soluble protein (Balaji et al. 2016). Given the lower cost of production and lower risk of transmissible pathogens compared to mammalian cell lines, duckweed may once again provide genetically engineered proteins for medical or other applications.

Catalyst Agri-Innovations Society (CAIS) works with a number of diverse companies in several locations including an on-farm anaerobic digester with nutrient extraction and at a land-based fish farm. All of their work is on efficiency and resource recovery at the food/energy/water nexus in the overall agriculture domain. They currently work with several wastewater treatments like the Trident Processes system for separating manure solids, anaerobic digestion to extract energy, and duckweed or algae to remove solutes. Wastewater from multiple species is anaerobically digested to generate methane and energy, and the digestate moves on to enclosed stacked shelf growth chambers filled with duckweed. After doubling in under 48 h and cleaning the water, duckweed is fermented to separate protein from high-value simple sugars. Christopher Bush, Co-founder of CAIS, has worked with the XPRIZE Foundation, designing competitions including "Feeding the Next Billion." The team also works with the HeroX platform where a sponsor can publicly host a problem and cash prize for the solution, greatly increasing the number of scientists who can see and solve the problem and learn from the winning solution. This type of modern interdisciplinary research center, consulting firm, or incubator that relies on datasets from large sensor arrays and crowd sourcing looks to be increasing in popularity, and we look forward to the variety of applications that will be developed where duckweed will play a role as one of several options to reclaim resources or feed people and livestock more effectively.

Perhaps given their ability to clean wastewater while providing food and fresh air, duckweeds can be seen as not only a crop species, but a life support system. The current water recovery system on the International Space Station relies on complex chemical treatments and reagents while generating wastes, which has NASA interested in developing closed-loop life support systems for long-term missions. Many plants develop poorly in microgravity and produce inedible biomass, so non-gravitropic aquatic plants and specifically duckweeds have been studied for space flight in closed-loop systems, microgravity simulations, and space flights since 1966 (Landolt and Kandeler 1987; Gale et al. 1989; Bluem and Paris 2003). *Lemna aequinoctialis* was even found to have a 32% increase in growth rate in simulated microgravity (Yuan and Xu 2017). Therefore, Space Lab Technologies, LLC is currently collaborating with the University of Colorado at Boulder on a Phase 2 grant from NASA to develop the μG-LilyPond™ growth chamber as part of a life support system (Escobar and Escobar 2017). Part of their project is studying how bursts of high intensity light can stimulate production of carotenoids, vitamin E, and other nutritious secondary metabolites (Demmig-Adams and Adams 2002), and how these bursts within the light regimen can be optimized for energy use, plant yield, and nutritional content. Thanks to their high growth rate, ability to grow in shallow trays, preference for ammonia, and entirely edible nutritious biomass duckweed are currently the prime candidates for the system. Presently, it is designed to provide fresh food and oxygen, with the eventual goal of converting urine to clean water. Based on the previous

literature, the goal is to create a 1 m^3 system capable of treating the wastewater and CO_2 of 4 crew members and providing an edible vegetable yield up to 250 g of dry weight per day (Gale et al. 1989; Landolt and Kandeler 1987). The μG-LilyPond™ system will need to overcome the unique challenges of space missions including size and weight restrictions, controlled growth and harvest in microgravity, water delivery via capillary action, sterility, minimal human maintenance, and rapid recovery from failures (Escobar and Escobar 2017). This intimate reliance on duckweed in a closed-loop system provides both a technical and a symbolic example of how humans and duckweed complement each other, and how we can use the smallest plants to solve the largest challenges.

Acknowledgments We'd like to thank Paul Skillicorn from Agriquatics, Professor Zhao Hai from Chengdu University, Darren Eng from Greenbelt Resources, Eduardo Mercovich from MamaGrande, Ron Salpeter from Hinoman, Tsipi and Ben Shoham from Green Onyx, Maurits van de Ven from Plantible Foods, Christopher Bush from CAIS, and Christine Escobar from Space Lab Technologies for sharing data on their companies and research and reviewing this document to ensure its accuracy.

References

Alexandratos N, Bruinsma J (2012) World agriculture towards 2030/2050: the 2012 revision. Food and Agriculture Organization of the United Nations (FAO), Rome

Alvarado S, Guédez M, Lué-Merú MP, Nelson G, Alvaro A, Jesús AC, Gyula Z (2008) Arsenic removal from waters by bioremediation with the aquatic plants water hyacinth (*Eichhornia crassipes*) and lesser duckweed (*Lemna minor*). Bioresour Technol 99 (17):8436–8440

Appenroth KJ, Crawford DJ, Les DH (2015) After the genome sequencing of duckweed—how to proceed with research on the fastest growing angiosperm? Plant Biol (Stuttg) 17:1–4. https://doi.org/10.1111/Plb.12248

Appenroth K-J, Sree KS, Böhm V, Hammann S, Vetter W, Leiterer M, Jahreis G (2017) Nutritional value of duckweeds (*Lemnaceae*) as human food. Food Chem 217:266–273

Appenroth KJ, Sree KS, Bog M, Ecker J, Boehm V, Lorkowski S, Sommer K, Vetter W, Tolzin-Banasch K, Kirmse R (2018) Nutritional value of the duckweed species of the genus *Wolffia* (Lemnaceae) as human food. Frontiers in chemistry 6:483

Balaji P, Satheeshkumar PK, Venkataraman K, Vijayalakshmi MA (2016) Expression of anti-tumor necrosis factor alpha (TNFalpha) single-chain variable fragment (scFv) in *Spirodela punctata* plants transformed with Agrobacterium tumefaciens. Biotechnol Appl Biochem 63(3):354–361. https://doi.org/10.1002/bab.1373

Baldi BG, Maher BR, Slovin JP, Cohen JD (1991) Stable Isotope Labeling, in Vivo, of d- and l-Tryptophan Pools in *Lemna gibba* and the Low Incorporation of Label into Indole-3-Acetic Acid. Plant Physiol 95 (4):1203–1208. https://doi.org/10.1104/pp.95.4.1203

Bhanthumnavin K, Mcgarry MG (1971) *Wolffia* arrhiza as a possible source of inexpensive protein. Nature 232 (5311):495

Bluem V, Paris F (2003) Possible applications of aquatic bioregenerative life support modules for food production in a Martian base. Adv Space Res 31(1):77–86

Boehm R, Kruse C, Voeste D, Barth S, Schnabl H (2001) A transient transformation system for duckweed (*Wolffia columbiana*) using Agrobacterium-mediated gene transfer. J Appl Bot 75(3–4):107–111

Bog M, Baumbach H, Schween U, Hellwig F, Landolt E, Appenroth KJ (2010) Genetic structure of the genus *Lemna* L. (*Lemnaceae*) as revealed by amplified fragment length polymorphism. Planta 232 (3):609–619. https://doi.org/10.1007/s00425-010-1201-2

Bog M, Schneider P, Hellwig F, Sachse S, Kochieva EZ, Martyrosian E, Landolt E, Appenroth KJ (2013) Genetic characterization and barcoding of taxa in the genus *Wolffia* Horkel ex Schleid. (Lemnaceae) as revealed by two plastidic markers and amplified fragment length polymorphism (AFLP). Planta 237 (1):1–13. https://doi.org/10.1007/s00425-012-1777-9

Borisjuk N, Chu P, Gutierrez R, Zhang H, Acosta K, Friesen N, Sree KS, Garcia C, Appenroth KJ, Lam E (2015) Assessment, validation and deployment strategy of a two-barcode protocol for facile genotyping of duckweed species. Plant Biol (Stuttg) 17:42–49. https://doi.org/10.1111/Plb.12229

Cantó-Pastor A, Mollá-Morales A, Ernst E, Dahl W, Zhai J, Yan Y, Meyers BC, Shanklin J, Martienssen R (2015) Efficient transformation and artificial miRNA gene silencing in *Lemna minor*. Plant Biol (Stuttg) 17:59–65. https://doi.org/10.1111/plb.12215

Cao HX, Wang W, Le HTT, Vu GTH (2016) The power of CRISPR-Cas9-induced genome editing to speed up plant breeding. Int J Genomics 2016:10. https://doi.org/10.1155/2016/5078796

Cao HX, Fourounjian P, Wang W (2018) The importance and potential of duckweeds as a model and crop plant for biomass-based applications and beyond. In:

Hussain CM (ed) Handbook of environmental materials management. Springer International Publishing, Cham, pp 1–16. https://doi.org/10.1007/978-3-319-58538-3_67-1

Chhabra G, Chaudhary D, Sainger M, Jaiwal PK (2011) Genetic transformation of Indian isolate of *Lemna minor* mediated by *Agrobacterium tumefaciens* and recovery of transgenic plants. Physiol Mol Biol Plants 17(2):129–136. https://doi.org/10.1007/s12298-011-0059-5

Childers DL, Corman J, Edwards M, Elser JJ (2011) Sustainability challenges of phosphorus and food: solutions from closing the human phosphorus cycle. Bioscience 61(2):117–124. https://doi.org/10.1525/bio.2011.61.2.6

Cui W, Cheng JJ (2015) Growing duckweed for biofuel production: a review. Plant Biol (Stuttg) 17:16–23. https://doi.org/10.1111/plb.12216

Demmig-Adams B, Adams WW 3rd (2002) Antioxidants in photosynthesis and human nutrition. Science 298 (5601):2149–2153. https://doi.org/10.1126/science.1078002

Duan P, Chang Z, Xu Y, Bai X, Wang F, Zhang L (2013) Hydrothermal processing of duckweed: effect of reaction conditions on product distribution and composition. Bioresour Technol 135:710–719. https://doi.org/10.1016/j.biortech.2012.08.106

Escobar CM, Escobar AC (2017, July) Duckweed: a tiny aquatic plant with enormous potential for bioregenerative life support systems. In: 47th International conference on environmental systems, Charleston, South Carolina, 16–20 July 2017. ICES, p 281

Feng B, Fang Y, Xu Z, Xiang C, Zhou C, Jiang F, Wang T, Zhao H (2017) Development of a new marker system for identification of *Spirodela polyrhiza* and *Landoltia punctata*. International Journal of Genomics 2017:8. https://doi.org/10.1155/2017/5196763

Fu L, Huang M, Han B, Sun X, Sree KS, Appenroth KJ, Zhang J (2017) Flower induction, microscope-aided cross-pollination, and seed production in the duckweed *Lemna gibba* with discovery of a male-sterile clone. Sci Rep 7(1):3047. https://doi.org/10.1038/s41598-017-03240-8

Gale J, Smernoff DT, Macler BA, MacElroy RD (1989) Carbon balance and productivity of *Lemna gibba*, a candidate plant for CELSS. Adv Space Res 9(8):43–52

Goopy J, Murray P (2003) A review on the role of duckweed in nutrient reclamation and as a source of animal feed. Asian-Australas J Anim Sci 16(2):297–305

Gwaze FR, Mwale M (2015) The prospect of Duckweed in pig nutrition: a review. J Agric Sci 7(11):189

Henchion M, Hayes M, Mullen AM, Fenelon M, Tiwari B (2017) Future protein supply and demand: strategies and factors influencing a sustainable equilibrium. Foods 6(7). https://doi.org/10.3390/foods6070053

Hou W, Chen X, Song G, Wang Q, Chang CC (2007) Effects of copper and cadmium on heavy metal polluted waterbody restoration by duckweed (*Lemna minor)*. Plant Physiol Biochem 45(1):62–69

International Organization of Standardization (2005) Water quality—Determination of the toxic effect of water constituents and waste water on duckweed (*Lemna minor*)—Duckweed growth inhibition test. ISO Geneva (CH), ISO guideline 20079

Khvatkov P, Chernobrovkina M, Okuneva A, Pushin A, Dolgov S (2015) Transformation of *Wolffia arrhiza* (L.) Horkel ex Wimm. Plant Cell Tissue Organ Cult (PCTOC) 123(2):299–307

Kuehdorf K, Jetschke G, Ballani L, Appenroth KJ (2013) The clonal dependence of turion formation in the duckweed *Spirodela polyrhiza*—an ecogeographical approach. Physiol Plant. https://doi.org/10.1111/ppl.12065

Kutschera U, Niklas KJ (2015) Darwin-Wallace Demons: survival of the fastest in populations of duckweeds and the evolutionary history of an enigmatic group of angiosperms. Plant Biol (Stuttg) 17:24–32. https://doi.org/10.1111/Plb.12171

Lam E, Appenroth KJ, Michael T, Mori K, Fakhoorian T (2014) Duckweed in bloom: the 2nd international conference on duckweed research and applications heralds the return of a plant model for plant biology. Plant Mol Biol 84 (6):737–742. https://doi.org/10.1007/s11103-013-0162-9

Landolt E (1986) The family of Lemnaceae—a monographic study, vol 1. Veroffentlichungen des Geobotanischen Institutes der Eidgenossischen Technischen Hochschule, Stiftung Rubel, Zurich

Landolt E, Kandeler R (1987) Biosystematic investigations in the family of duckweeds (*Lemnaceae*), Vol. 4: The family of Lemnaceae-a monographic study, Vol. 2 (phytochemistry, physiology, application, bibliography). Veroeffentlichungen des Geobotanischen Instituts der ETH, Stiftung Ruebel (Switzerland)

Liu C, Feng Q, Yang J, Qi X (2018) Catalytic production of levulinic acid and ethyl levulinate from uniconazole-induced duckweed (*Lemna minor*). Bioresour Technol 255:50–57. https://doi.org/10.1016/j.biortech.2018.01.087

Martirosyan EV, Ryzhova NN, Skryabin KG, Kochieva EZ (2008) RAPD analysis of genome polymorphism in the family Lemnaceae. Russ J Genet 44(3):360–364. https://doi.org/10.1134/s1022795408030198

Mkandawire M, Dudel EG (2005) Accumulation of arsenic in *Lemna gibba* L. (duckweed) in tailing waters of two abandoned uranium mining sites in Saxony, Germany. Sci Total Environ 336 (1):81–89

Nauheimer L, Metzler D, Renner SS (2012) Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils. New Phytol 195 (4):938–950. https://doi.org/10.1111/j.1469-8137.2012.04220.x

OECD/FAO (2017) OECD-FAO agricultural outlook 2017–2026. OECD Publishing, Paris. https://doi.org/10.1787/19991142

OECD (2006) Test No. 221: *Lemna* sp. growth inhibition test. OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, Paris. https://doi.org/10.1787/9789264016194-en

Posner HB (1962) Characteristics of X-ray-induced aberrants of *Lemna perpusilla* 6746. Plant Cell Physiol 3(3):275–284

Rezania S, Taib SM, Din MFM, Dahalan FA, Kamyab H (2016) Comprehensive review on phytotechnology: heavy metals removal by diverse aquatic plants species from wastewater. J Hazard Mater 318:587–599

Rodell M, Famiglietti JS, Wiese DN, Reager JT, Beaudoing HK, Landerer FW, Lo MH (2018) Emerging trends in global freshwater availability. Nature 557 (7707):651–659. https://doi.org/10.1038/s41586-018-0123-1

Schmitz U, Kelm H (2017) First discovery of flowering *Wolffia arrhiza* in Central Europe. Aquat Bot 143:33–35

Shoham T (2018) On a mission to introduce duckweed as the 21st century cash crop: Current active industrial efforts. Duckweed Forum 6(3):92–97

Skillicorn P (2013) The agriquatics system vs activated sludge: a comparative analysis. https://paulskillicorn.wordpress.com/2013/04/17/the-agriquatics-system-vs-activated-sludge-a-comparative-analysis/. Accessed 9 Sept 2019

Skillicorn P, Spira W, Journey W (1993) Duckweed aquaculture: a new aquatic farming system for developing countries. World Bank

Slovin JP, Tobin EM (1982) Synthesis and turnover of the light-harvesting chlorophylla/b-protein in *Lemna gibba* grown with intermittent red light: possible translational control. Planta 154(5):465–472. https://doi.org/10.1007/BF01267815

van der Spiegel M, Noordam MY, van der Fels-Klerx HJ (2013) Safety of novel protein sources (Insects, microalgae, seaweed, duckweed, and rapeseed) and legislative aspects for their application in food and feed production. Compr Rev Food Sci F 12(6):662–678. https://doi.org/10.1111/1541-4337.12032

Sree KS, Bog M, Appenroth KJ (2016) Taxonomy of duckweeds (*Lemnaceae*), potential new crop plants. Emir J Food Agr 28(5):291–302. https://doi.org/10.9755/ejfa.2016-01-038

Sree KS, Maheshwari SC, Boka K, Khurana JP, Keresztes A, Appenroth KJ (2015a) The duckweed *Wolffia* microscopica: a unique aquatic monocot. Flora 210:31–39. https://doi.org/10.1016/j.flora.2014.10.006

Sree KS, Sudakaran S, Appenroth K-J (2015b) How fast can angiosperms grow? Species and clonal diversity of growth rates in the genus *Wolffia* (Lemnaceae). Acta Physiol Plant 37(10):204

Tang J, Li Y, Ma J, Cheng J (2015) Survey of duckweed diversity in Lake Chao and total fatty acid, triacylglycerol, profiles of representative strains. Plant Biology 17(5):1066–1072

Tippery N, Les D, Crawford D (2015) Evaluation of phylogenetic relationships in Lemnaceae using nuclear ribosomal data. Plant Biology 17(s1):50–58

Vunsh R, Li J, Hanania U, Edelman M, Flaishman M, Perl A, Wisniewski JP, Freyssinet G (2007) High expression of transgene protein in Spirodela. Plant Cell Rep 26(9):1511–1519. https://doi.org/10.1007/s00299-007-0361-4

Wang W, Wu Y, Yan Y, Ermakova M, Kerstetter R, Messing J (2010) DNA barcoding of the Lemnaceae, a family of aquatic monocots. BMC Plant Biol 10:205. https://doi.org/10.1186/1471-2229-10-205

Xue HL, Xiao Y, Jin YL, Li XB, Fang Y, Zhao H, Zhao Y, Guan JF (2012) Genetic diversity and geographic differentiation analysis of duckweed using inter-simple sequence repeat markers. Mol Biol Rep 39(1):547–554. https://doi.org/10.1007/s11033-011-0769-3

Yamamoto Y, Rajbhandari N, Lin X, Bergmann B, Nishimura Y, Stomp A-M (2001) Genetic transformation of duckweed *Lemna gibba* and *Lemna minor*. Vitro Cell Develop Biol-Plant 37(3):349–353. https://doi.org/10.1007/s11627-001-0062-6

Yan Y, Candreva J, Shi H, Ernst E, Martienssen R, Schwender J, Shanklin J (2013) Survey of the total fatty acid and triacylglycerol composition and content of 30 duckweed species and cloning of a Δ6-desaturase responsible for the production of γ-linolenic and stearidonic acids in Lemna gibba. BMC Plant Biol 13(1):201

Yuan J, Xu K (2017) Effects of simulated microgravity on the performance of the duckweeds *Lemna* aequinoctialis and *Wolffia* globosa. Aquat Bot 137:65–71. https://doi.org/10.1016/j.aquabot.2016.11.010

Zhang C, Duan P, Xu Y, Wang B, Wang F, Zhang L (2014a) Catalytic upgrading of duckweed biocrude in subcritical water. Bioresour Technol 166:37–44. https://doi.org/10.1016/j.biortech.2014.05.022

Zhang DQ, Jinadasa K, Gersberg RM, Liu Y, Ng WJ, Tan SK (2014b) Application of constructed wetlands for wastewater treatment in developing countries–a review of recent developments (2000–2013). J Environ Manag 141:116–131

Zhang X, Zhao FJ, Huang Q, Williams PN, Sun GX, Zhu YG (2009) Arsenic uptake and speciation in the rootless duckweed *Wolffia* globosa. New Phytol 182 (2):421–428

Zhao Y, Fang Y, Jin Y, Huang J, Bao S, Fu T, He Z, Wang F, Zhao H (2014) Potential of duckweed in the conversion of wastewater nutrients to valuable

biomass: a pilot-scale comparison with water hyacinth. Bioresour Technol 163:82–91. https://doi.org/10.1016/j.biortech.2014.04.018

Zhao H, Appenroth K, Landesman L, Salmean AA, Lam E (2012) Duckweed rising at Chengdu: summary of the 1st international conference on Duckweed application and research. Plant Mol Biol 78 (6):627–632. https://doi.org/10.1007/s11103-012-9889-y

Ziegler P, Adelmann K, Zimmer S, Schmidt C, Appenroth KJ (2015) Relative in vitro growth rates of duckweeds (*Lemnaceae*)—the most rapidly growing higher plants. Plant Biol (Stuttg) 17:33–41. https://doi.org/10.1111/plb.12184

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)

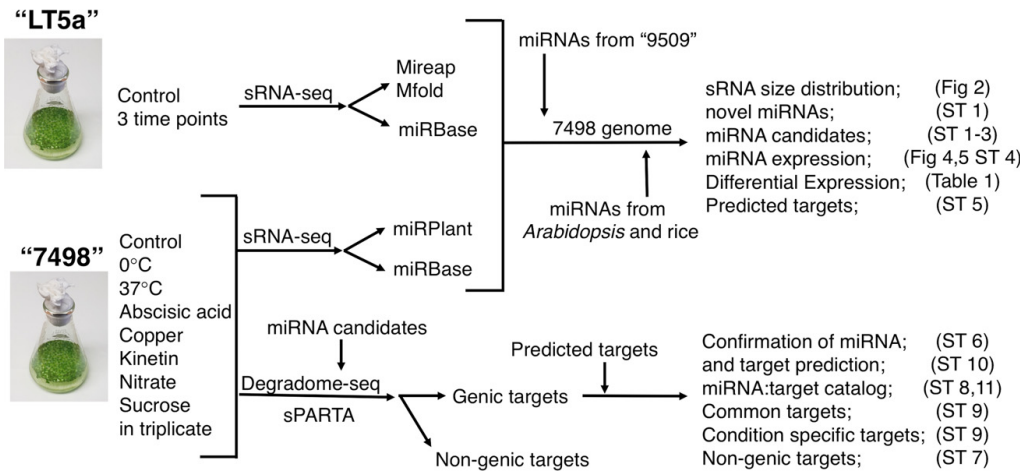


Figure 1. Study overview.
This study includes three different sequencing experiments, and eight experimental conditions. miRNA candidates were collected from both sRNA analyses, de-duplicated and then mapped to the '7498' genome. Then new miRNAs were added from strain 9509 and *Arabidopsis* and rice 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.

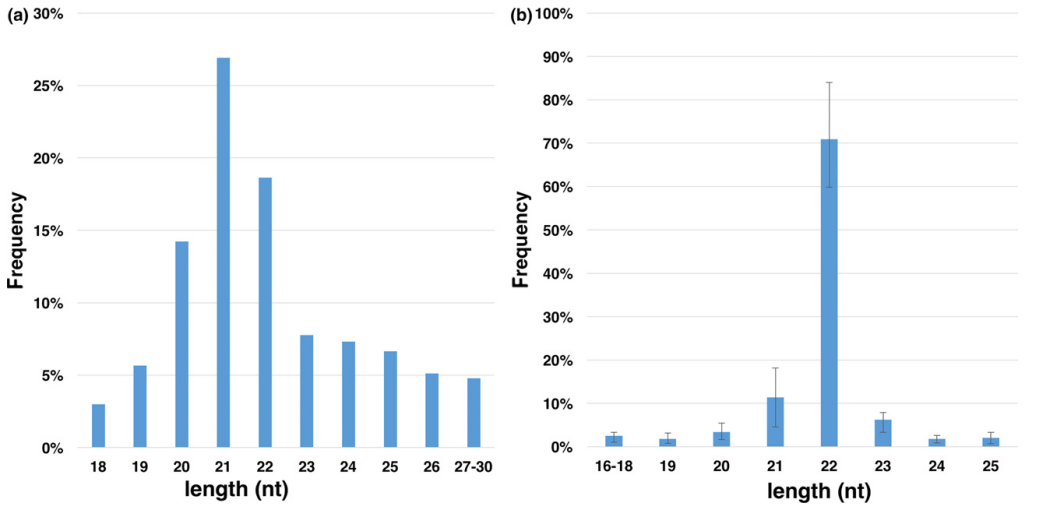


Figure 2. Length distributions of small RNAs from *Spirodela polyrhiza*.
(a) Length distribution in strain LT5a; (b) Length distributions in strain 7498. The error bars indicate the highest and lowest average percentage seen between the eight conditions sequenced.

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

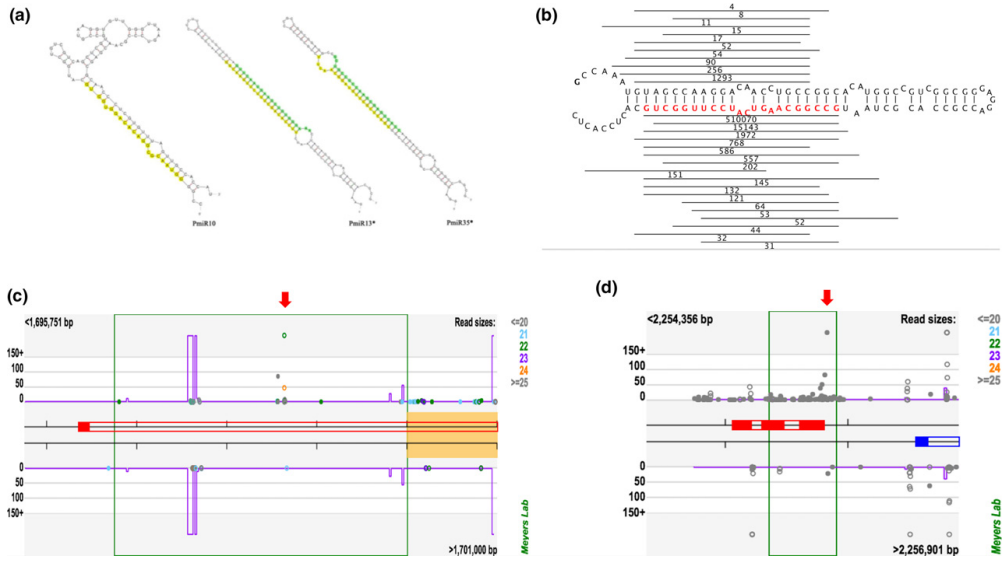


Figure 3. Visual examples of miRNA and target prediction.
(a) Mfold illustrations of predicted novel miRNAs; (b) miRPlant graph of read distribution along a predicted miR156 family hairpin; (c) Read distribution of miRNA spo-miR172d, marked with a red arrow, located along with a long terminal repeat (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

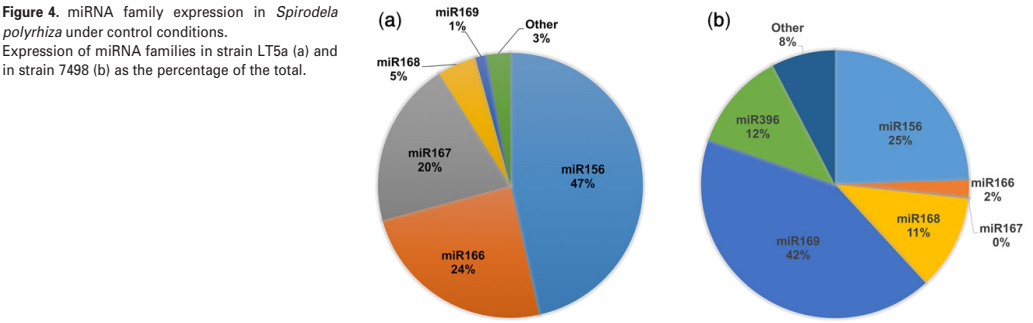


Figure 4. miRNA family expression in *Spirodela polyrhiza* under control conditions.
Expression of miRNA families in strain LT5a (a) and in strain 7498 (b) as the percentage of the total.

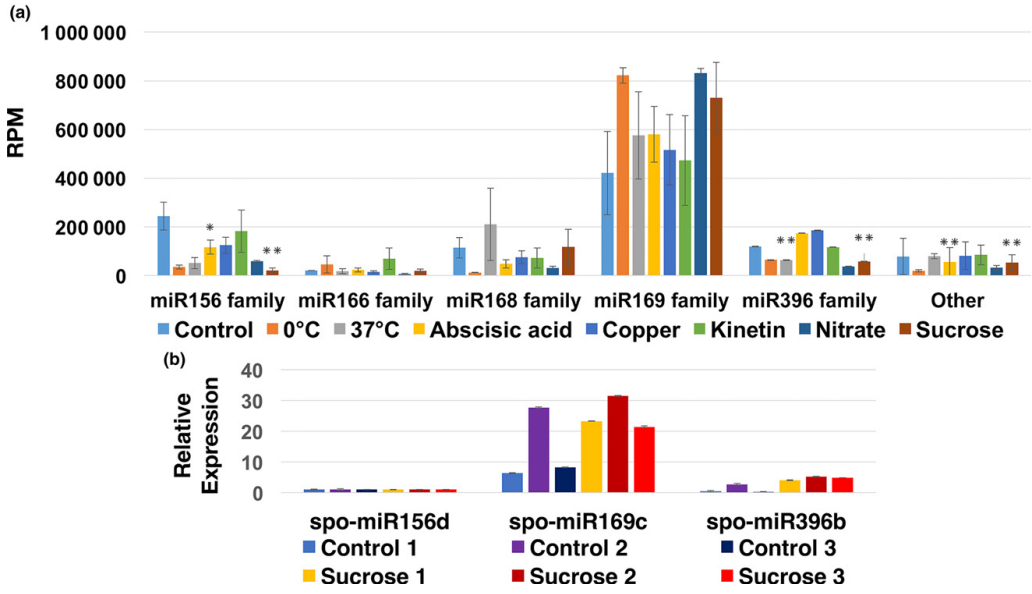


Figure 5. miRNA family expression under eight different conditions.
(a) Sequencing data for the five most abundant miRNA families and all other miRNAs, from biological triplicates of *S. polyrhiza* strain 7498 under eight conditions. The data are expressed as reads per million (RPM) of all miRNA reads ± 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.

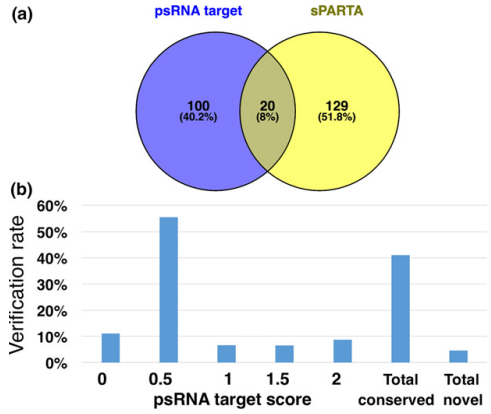


Figure 6. Overlap of target predictions and sequenced validations.
(a) Overlap of the mRNA targets predicted by psRNA target in blue and cleaved targets detected by sPARTA in yellow. Overlap indicates cleavage by the same miRNA family and predictions were for those miRNAs validated by the sPARTA analysis; (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)

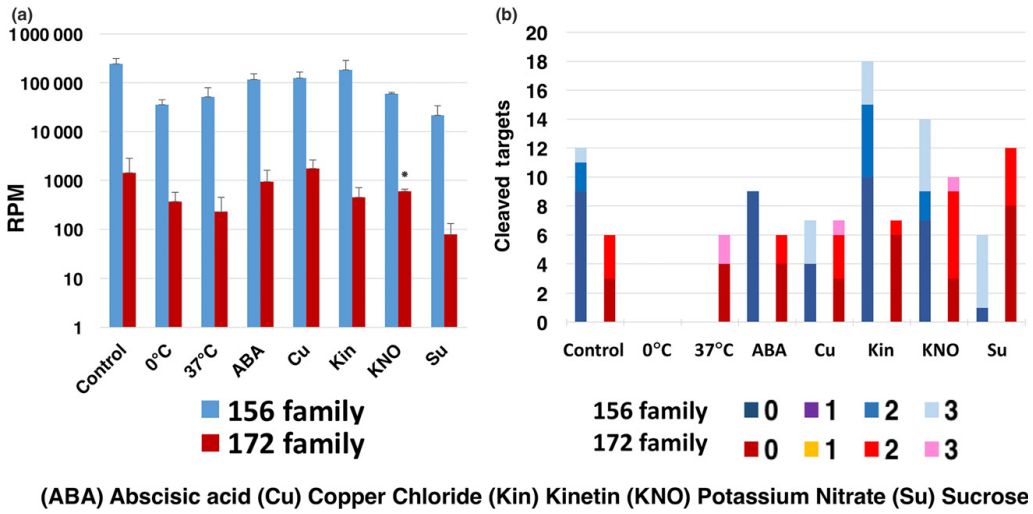


Figure 7. Abundance and target cleavage of miR156 and miR172 families.
(a) Read count of miR156 and miR172 family mature sequences as reads/million miRNA reads. Error bars are SEM of three biological replicates. *Represents that the miR156 family was significantly more abundant than the 172 family under 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.

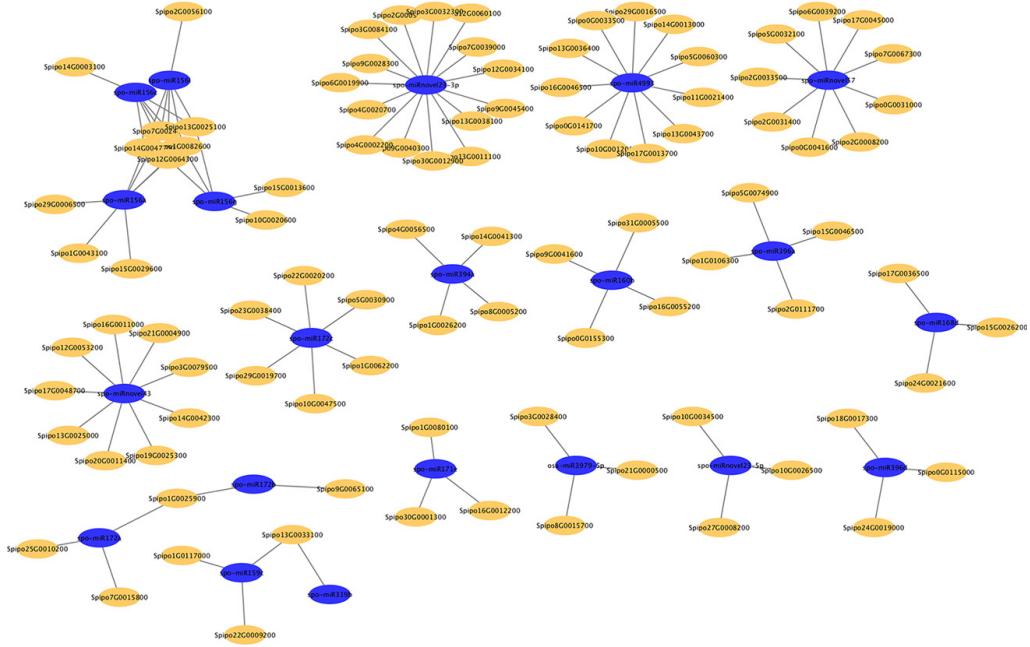


Figure 8. miRNA and target network from Table S8 represented in Cytoscape. Blue nodes are the 21 most connected miRNAs with the most targets, whereas yellow represents mRNA targets from filtered genic *sPARTA* results. The miR156 family is seen in the top left sharing most of their targets, and similarly 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.

Small RNAs in Duckweeds

16

Paul Fourounjian

Abstract

Within eukaryotic genomes, there are several types of small RNAs including sn, sno, si, and miRNAs. With respect to the Lemnaceae family, the vast majority of the research to date has been conducted in *Spirodela polyrhiza*, focused mainly on the miRNAs. This research consists of three small RNA-sequencing experiments in strains from China, Germany, and the USA, with each experiment identifying conserved miRNAs and predicting novel miRNAs and targets. While the novel miRNAs and recently discovered miRNAs fluctuated, the family size and expression of well-known miRNA families was consistent between the three experiments. While miRBase likely contains many incorrect annotations, these miRNAs were annotated according to strict criteria and analyzed for the miRBase high confidence list. They were further characterized through degradome sequencing, which confirmed half of the conserved miRNAs and a third of the novel. Finally, *Spirodela polyrhiza* has a surprisingly low abundance of 24nt sRNAs, which are required to suppress transposon proliferation.

P. Fourounjian (✉)
Waksman Institute of Microbiology, Rutgers University, Piscataway 08854, USA
e-mail: pjf99@scarletmail.rutgers.edu

As scientists moved from sequencing the ϕX174 virus in 1977 to prokaryote genomes, simple eukaryotes, and then the first plant, *Arabidopsis thaliana* in the year 2000, they saw that these larger complex genomes were made of so much more than genes. We now know that eukaryotic genomes contain a host of structural repeats such as the centromere and telomere regions. There are also large stretches of tandem repeats, also called satellite DNA. Then, there are the virus-like transposable elements that are often copied and spread across the genome. Many of the transcribed RNA sequences are small RNAs like small interfering, micro, and small nucleolar RNAs (si, mi, and snoRNAs) that bind to protein complexes to regulate gene expression and assemble ribosomes. Larger RNA transcripts include long non-coding RNAs and the high copy number ribosomal and transfer RNAs (lnc, r, and tRNAs) that translate mRNAs to proteins. Each genome also contains plenty of pseudogenes, which are non-functional due to mutations. Finally, the genome contains the protein-coding genes themselves, with all their introns, exons, cis- and trans-regulatory elements and terminators, which are 2% of the human genome and roughly 20% of a typical angiosperm genome, with wide variation due to genome size differences.

Within this genome, there are several types of transcribed RNAs, with the longer varieties including m, r, t, and lncRNA. While the first three types are well characterized, long

X. H. Cao et al. (eds.), *The Duckweed Genomes*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-11045-1_16

non-coding RNA (lncRNA) wasn't discovered until 1990 (Brannan et al. 1990). These spliced and polyadenylated RNAs function in epigenetic regulation, the generation and sequestration of miRNAs, and various other functions. While most of the studies have been run in animals, thousands of lncRNAs have been annotated in plant genomes, including IPS1, which sequesters miR399 with a non-cleavable target bulge in response to phosphate starvation across many plant species (Franco-Zorrilla et al. 2007).

The small RNAs in plants include sno, si, and miRNAs, with the snoRNAs evolutionarily conserved back to Archaea. They are produced from their own RNA precursors, or introns, which are cleaved by endonucleases and trimmed by exonucleases, until only the protein bound 60–250 bp snoRNA remains; they then guide the protein complex's methylation and pseudouridylation of rRNAs in the nucleolus. It is even hypothesized that snoRNAs gave rise to miRNAs based on their similarity in processing including some overlap of enzymes, their similar hairpin structure, and combination of function (Scott and Ono 2011). There have been reports of snoRNAs with miRNA-like characteristics, and vice versa, and even small RNAs with complete sno and miRNA function in animals, plants, and yeast. In plants, both miRNAs and siRNAs are cut to 22 and 21nt lengths by dicer-like proteins 1 and 4, respectively, and loaded onto Ago1 in the RISC, with the main difference being that an RNA hairpin is processed into a miRNA for mRNA gene suppression, while a dsRNA is diced into many siRNAs for pathogen gene silencing.

When the *Spirodela polyrhiza* genome was published in 2014, prediction programs were able to detect miRNA precursors through homologous sequences and RNA folding software (Wang et al. 2014a). In strain 7498, all miRBase plant mature sequences were mapped back to the genome, and flanking sequences analyzed by RNAfold and miRCheck (Denman 1993; Jones-Rhoades and Bartel 2004). The search predicted 413 miRNAs belonging to 93 families. This survey based on DNA sequencing aimed to provide all possible miRNA genes, for comparison to other plant genomes, with the eventual aim of detecting their activity in later RNA-seq experiments.

The earliest attempt at sequencing and analyzing *S. polyrhiza* miRNAs predated the published genome. This experiment, run at Peking University Shenzhen Graduate School, was run on strain LT5a, isolated from Lake Tai, using three populations grown in SH media for 1, 3, and 5 days under control conditions. Using 18-31nt sRNA on a HiSeq 2000 Illumina platform, they sequenced 24 million reads, 3.5 of which matched conserved miRNAs in miRBase, and 7.6 million that were not annotated in GenBank or Rfam. These 7.6 million reads were analyzed by the MIREAP program and validated by Mfold to identify 41 predicted novel miRNAs (Zuker 2003). A summary of this and the other small RNA-seq experiments is available in Table 16.1.

In strain 9509, conserved and novel miRNAs were identified through small RNA-sequencing and an analysis of read count and distribution (Michael et al. 2017). The study used 10 sRNA libraries from a SOLiD5500 sequencer, aligned to the genome allowing 1 mismatch, and then annotated if the candidate has a stable hairpin structure, sufficient miR reads, more than 1 miR* read, and a 2 or 3 nt 3′ overhang (Table 16.1). They identified conserved miRNAs by checking for a strong BLAST homology to not only the mature, but also hairpin structures in miRBase. Next they used the program TargetFinder with a cutoff score of 4 to identify the predicted targets (Fahlgren and Carrington 2010). These transcription and structural requirements lead to the prediction of 59 conserved miRNAs in 22 families, and 29 novel miRNAs, with 29 of the conserved and 25 of the novel miRNAs being predicted to regulate 991 mRNA targets.

Alongside the miRNA prediction, they were able to predict trans-acting siRNAs (tasiRNAs), from the sRNA library using previously established criteria (Howell et al. 2007; Johnson et al. 2009). Reads matching cDNA and the corresponding genomic regions had miRNA results filtered out, and then, 50nt candidate transcripts were required to have over 100 reads, with over

Table 16.1 Summary of sRNA-sequencing experiments

Strain	LT5a	7498	9509
Conditions	Control (SH media, 16 h days, 23 °C)	Control, heat, cold, abscisic acid, copper, kinetin, nitrate, sucrose	Control, abscisic acid
# reads	25 million	32 million	N/A
# conserved miRNAs	158	58	59
# novel miRNAs	41	14	29
# targets	N/A	162	991
# DE miRNAs	N/A	15	12

70% being 21nt in length. These are sufficient to distinguish randomly degraded transcripts from mRNAs that had been transcribed into dsRNA and then diced into 21nt tasiRNAs. TargetFinder was then used with a cutoff of 6 and a requirement of two miRNA bind sites to identify the targeted genes. This search yielded two cleaved TAS3 genes, and the miR393 targeting another putative TAS gene that was also found in oil palm and banana.

The most recent miRNA survey started with strain 7498 grown in three replicate flasks of eight growth conditions: control, cold, heat, abscisic acid, copper, kinetin, nitrate, and sucrose stimuli. After harvest, RNA extraction, and size selection, 32 million reads of the 24 libraries were sequenced on the SOLiD5500 platform and mapped to the genome (Table 16.1). These results were filtered against *Brachypodium distachyon* non-coding RNAs, with miRNAs removed, and analyzed in miRPlant (An et al. 2014). Criteria required a miRPlant score greater than 3.0, over 20 miR reads and at least 1 miR* read. This yielded 58 conserved miRNAs and 14 novel miRNAs after the removal of those that had already been found in strain 9509. When consolidated with the results from strain LT5a and mapped back to the strain 7498 genome, these two showed a strong degree of overlap resulting in 63 conserved and 45 novel miRNAs. These miRNAs were then further judged by the stringent criteria for plant miRNA annotation by sRNA-seq indicating that only 30 were highly confident based on structure and read count (Axtell and Meyers 2018). These miRNAs were then used to predict 163 targets with a psRNATarget score better than 2 (Dai and Zhao 2011), with roughly half corresponding to novel miRNAs.

The first prediction of miRNAs based on genome sequence and hairpin structure saw 413 possible miRNAs, and this number dropped to 58 and 59 once the miRNAs were being predicted based on sequencing results (Table 16.1). Of the 413 miRNAs, many were from recently discovered families, with only 121 that corresponding to those 58 families sequenced in 7498 at 119 genomic loci. While numbers of miRNA loci within families mostly agree, the copy number of a few families based on expression data differs from the 7498 genome survey as shown in Table 16.2. Perhaps the 24 copies of miR156 include a number of unexpressed pseudogenes from duplication events. When the strain 7498 and 9509 conserved miRNA families were compared 20 overlapped, while two were only found in the 9509 genome, and the 7498 study included 11 less commonly conserved one-member miRNA families not observed in strain 9509. This overlap of family and sequence number of highly conserved families suggests we have robust identification of the expressed, heavily conserved miRNA families, while lower confidence previously reported and novel miRNAs require further investigation to characterize.

While much attention is always paid to proper identification and mapping of miRNAs in the first sequencing experiments of a genome, measuring miRNA abundance is also essential. Since miRNA families have high sequence homology and target the same family of gene targets, these results are grouped by expression of certain

families. The three experiments studied strains LT5a, 9509, and 7498 which originated in China, USA, and Germany, providing a global perspective of the species. The control conditions were largely similar using Schenk & Hildebrandt medium at a pH of 5.8, with the known variations mainly being the 15 °C night time temperature, and relatively young cultures for LT5a and harvesting based on water surface coverage in 7498. While these expression results from strains across the world grown in control conditions vary in rank and abundance of miRNA families (Table 16.3), the same six families are within the top 5 in two of the three experiments demonstrating their prominent roles. As seen in Table 16.4, these miRNA families and their target gene families regulate growth, meristem development, and stress responses.

Strain 9509 was also exposed to 1uM ABA, which was shown to induce turion production irreversibly after 3 days (Wang et al. 2014b; Kuehdorf et al. 2014). At the 10 h time point, this hormonal stimuli changed the expression of 12 conserved miRNAs (over 100 reads in control, over twofold expression change in ABA), with the 169 and 396 families being underexpressed and the 159 and 168 families doubling in abundance (Michael et al. 2017). Then, at the 5-day time point, there were 28 miRNAs and targets with significant overexpression of the miRNA and underexpression of the mRNA compared to control and vice versa. Twelve of the miRNAs were novel miRNAs with relatively low expression, large fold change differences, and a wide variety of targets. Similar to the transcriptomic study at day 3, this experiment saw a decline in chloroplast proteins and an increase in polyphenol producing enzymes (Wang et al. 2014b; Michael et al. 2017).

The survey of miRNAs in strain 7498 in the control, cold, heat, abscisic acid, copper, kinetin, nitrate, and sucrose stimuli yielded a striking

Table 16.2 Copy number variation of miRNA families between three publications

miRNA family	7498 genome survey	9509 sRNA-seq	7498 sRNA-seq
156	24	6	9
159	1	3	4
169	9	5	7
396	11	5	9

Table 16.3 miRNA expression of control conditions of three strains of *Spirodela polyrhiza*

LT5a	7498	9509
156 (47%)	156 (41%)	160 (68%)
166 (24%)	168 (18%)	169 (7%)
167 (20%)	396 (16%)	166 (6%)
168 (5%)	169 (6%)	528 (5%)
169 (1%)	166 (4%)	159 (3%)

Table 16.4 Biological roles of prominent miRNA families

miRNA family	mRNA target family	Biological role
miR156	SPLs	Maintains juvenile tissues
miR166	HDZipIIIs	Regulates meristems
miR167	IARs	Reduced by drought
miR168	Ago1	Viral defense
miR169	NFYs	Drought and stress response
miR396	GRFs	Regulates meristems

result in that miR169c was between 33 and 82% of the reads in each condition, with large variability between the three biological replicates. This result was believed by the authors of the study to be an experimental artifact due to the lack of this expression in the other experiments, the only partial replication of the expression in the qPCR follow-up, and the current reputation of the SOLiD5500 sequencer. With this one sequence ignored and the dataset renormalized, we can accurately see the responses of other miRNAs to the various conditions. There were large increases in miR166 expression under the influence of cold and kinetin and miR168 in the heat and sucrose conditions. The meristem regulating 396 familiesy doubled expression in response to the heat, ABA, and copper stimuli. Finally, miR156, which maintains the juvenile, neotenous life cycle of the duckweed family, decreased over fourfold in response to sucrose, which was the condition responsible for 13 of the 19 instances of differential miRNA expression indicating that the mixotrophic lifestyle often used in laboratory experiments is quite different from duckweed grown in an outdoor setting.

Accurate miRNA annotation is quite difficult, since miRNAs are vastly outnumbered by similarly sized siRNAs in the genome, and even the more stringent miRNA prediction programs supply tens or hundreds of false predictions. An analysis in 2014 suggested that 75% of the land plant miRNA families in miRBase are questionable, especially those with only a single member (Taylor et al. 2014). In an attempt to manage the large number of submissions and false positives coming in, miRBase has established criteria for its high confidence miRNAs that analyze the structure of the hairpin, the read distribution along it, and the miR, and miR* read count. For plants in miRBase release 21, there are currently 6942 hairpins in 2408 distinct miRNA families, with only 587 from 227 families (9.7%) making the high confidence cutoff (Griffiths-Jones 2006). As an attempt to preserve miRNA annotation confidence, 21 of the leading minds of the field wrote the plant miRNA annotation criteria in 2008 that has since been updated by two of them thanks to new information and sequencing capabilities (Meyers et al. 2008; Axtell and Meyers 2018). The plant miRNA annotation criteria are generally more stringent than the high confidence criteria, except for the latter's requirement of 10 miR* reads, since plant miRNA biogenesis is quite specific. Both miRNA studies in *Spirodela* annotated miRNAs based on homology according to the 2008 criteria, with most of these being well-conserved, high-confidence miRNA families. The conserved miRNAs with family names above 535 are relatively likely to be based off of lower confidence annotations in previous reports. The novel miRNAs from strain 9509 were predicted in 2017 using cutoffs very similar to the 2018 criteria, demonstrating a high degree of confidence, while those predicted in the LT5a and 7498 study had a lower degree of confidence. These authors reviewed all their data, with the revised criteria finding that 30 of the 47 hairpin structures met the current standards.

In addition to applying the stringent structural and read distribution filters above, the authors of the 2018 study verified miRNAs through a method called degradome sequencing where uncapped mRNAs are sequenced and aligned to miRNA target sites to measure evidence of precise miRNA cleavage above random mRNA degradation. There were several methods available at the time, and the authors chose the GMUCT2.0 library for its read length and minimal PCR amplification and the sPARTA program for its accuracy in analysis of the degradome data (Kakrana et al. 2014; Willmann et al. 2014). Biological triplicate libraries of the same eight conditions observed in the miRNA-sequencing study were sequenced on the Illumina NextSeq 500, yielding 911 million total reads. When running the sPARTA program, the *Spirodela* 7498 gene models were extended 150nt upstream and 250nt downstream, since many of the degradome reads were from the UTRs of the mRNAs. The sequencing verified activity of 66 miRNAs on 149 targets. For the 42 conserved miRNAs, the targets were mainly the transcription factor families reported in other plant species. While these essential developmental transcription factors mostly made up the

targets sequenced in over half of the conditions, 71% of the cleaved targets were sequence specific underscoring the importance of sampling a variety of post-transcriptional responses. Notably sucrose had the largest number of condition-specific results including metabolic and signaling proteins indicating a large shift in the mixotrophic lifestyle. This included a complete reversal where miR172 went from cleaving half as many targets as miR156 to twice as many despite being 0.4% of its expression. This suggested that sucrose may be inducing a less neotenous phenotype, and that highly expressed miRNAs are not necessarily highly active. Of the 81 novel miRNAs predicted within the three separate experiments of *Spirodela,* 24 were validated with 66 targets. This 30% validation rate, evenly spread between the three experiments, is consistent with similar surveys in other plant genomes thanks to the low expression and number of targets compared to conserved miRNAs, and the likelihood that novel miRNAs may be false predictions (Song et al. 2010; Li et al. 2010; Yang et al. 2013). While degradome evidence is a great way to confirm miRNAs, it does require co-expression and mRNA cleavage meaning that non-supported miRNAs may be found as active in later experiments with the right conditions and sequencing depth.

In order to provide other scientists easy access for further analysis, the raw data is available for LT5a results at GSE55208, 9509 at PRJNA308109, and 7498 at PRJNA473779 (SRP149336). As a second approach to increase transparency, ease replication, and enable further research, the data from the 2018 study and some of its analysis can be viewed in the Galaxy server as a history of the analysis, which includes the option of extracting the workflow and adapting it to analyze similar data Spirodela7498Galaxy-history (Afgan et al. 2016). Then, as a third method to make the data quick to review and useful to the community, the 7498 results are now displayed on an interactive viewer hosted by the Myers lab at the Danforth center https://mpss.danforthcenter.org/tools/mirna_apps/comPARE.php. Here the user can search for miRNAs, targets and sequences, see the expression across the 24 libraries, and download expression data (Fig. 16.1) (Nakano et al. 2006). The goal of this data accessibility was to enable other scientists to explore beyond the miRNAs, to the phased small interfering RNAs, the possible lncRNA intergenic targets in the degradome sequencing, or any other striking discovery within the datasets.

While the primary focus of both sRNA-seq experiments was to analyze miRNAs, Professor Jie Tang working with strain LT5a noted a surprising lack of 24nt RNAs typically found in plant genomes. These are often comparable in expression to the 22 and 21nt miRNAs, but they were rare as 7.3% of the small RNAs in strain

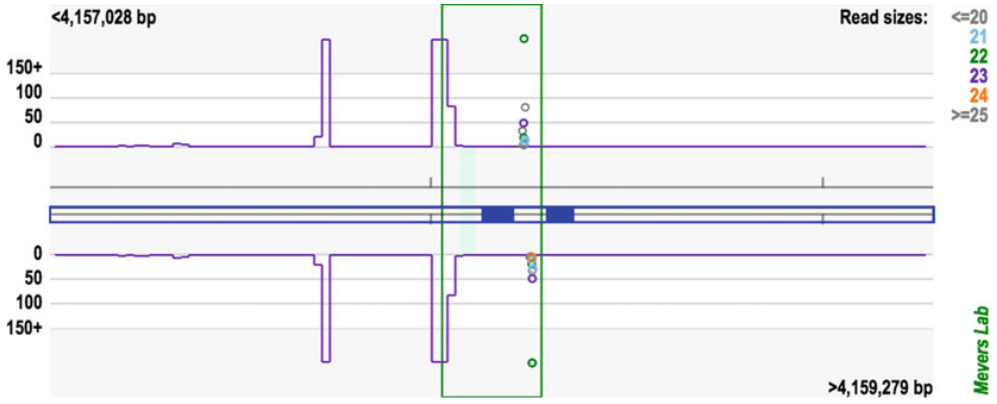


Fig. 16.1 View of small RNA browser showing high expression of the 22nt miR396d in the intron of the unknown protein Spipo10G0052600 in the control 1 library

LT5a, and 1% in 7498. In other plant species, 24nt RNAs are a part of the RNA-directed DNA methylation pathway where transposons are transcribed into single-stranded and then double-stranded RNA, diced into 24nt heterochromatic small RNAs, and then used to guide a protein network that methylates matching sequences and then silences them as heterochromatin. Accordingly, Michael et al. also studied DNA methylation in the *Spirodela* 9509 genome and found it to be the least methylated plant sequenced! This DNA methylation pathway in duckweeds is a new and exciting field of study summarized in Chap. 5 that appears to be the cutting edge of small RNA research in the Lemnaceae.

Acknowledgements I would like to thank Professor Jie Tang of Chengdu University for providing data on his submitted and published work respectively and reviewing this manuscript to ensure its accuracy.

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. https://doi.org/10.1093/nar/gkw343

An J, Lai J, Sajjanhar A et al (2014) miRPlant: an integrated tool for identification of plant miRNA from RNA sequencing data. BMC Bioinform 15:275. https://doi.org/10.1186/1471-2105-15-275

Axtell MJ, Meyers BC (2018) Revisiting criteria for plant miRNA annotation in the era of big data. Plant Cell 30:tpc.00851.2017. https://doi.org/10.1105/tpc.17.00851

Brannan CI, Dees EC, Ingram RS, Tilghman SM (1990) The product of the H19 gene may function as an RNA. Mol Cell Biol 10:28–36. https://doi.org/10.1128/MCB.10.1.28

Dai X, Zhao PX (2011) PsRNATarget: A plant small RNA target analysis server. Nucleic Acids Res 39: W155–W159. https://doi.org/10.1093/nar/gkr319

Denman RB (1993) Using RNAFOLD to predict the activity of small catalytic RNAs. Biotechniques 15:1090–1095

Fahlgren N, Carrington JC (2010) miRNA target prediction in plants. Humana Press, pp 51–57

Franco-Zorrilla JM, Valli A, Todesco M et al (2007) Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet 398 (39):1033. https://doi.org/10.1038/ng2079

Griffiths-Jones SM (2006) The microRNA sequence database. Methods Mol Biol 342:129–138

Howell MD, Fahlgren N, Chapman EJ et al (2007) Genome-wide analysis of the RNA-dependent RNA POLYMERASE6/DICER-LIKE4 pathway in Arabidopsis reveals dependency on miRNA- and tasiRNA-directed targeting. Plant Cell 19:926–942. https://doi.org/10.1105/tpc.107.050062

Johnson C, Kasprzewska A, Tennessen K et al (2009) Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. Genome Res 19:1429–1440. https://doi.org/10.1101/gr.089854.108

Jones-Rhoades MW, Bartel DP (2004) Computational identification of plant MicroRNAs and their targets, including a stress-induced miRNA. Mol Cell 14:787–799. https://doi.org/10.1016/J.MOLCEL.2004.05.027

Kakrana A, Hammond R, Patel P et al (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. https://doi.org/10.1093/nar/gku693

Kuehdorf K, Jetschke G, Ballani L, Appenroth K-J (2014) The clonal dependence of turion formation in the duckweed *Spirodela polyrhiza*—an ecogeographical approach. Physiol Plant 150:46–54. https://doi.org/10.1111/ppl.12065

Li Y-F, Zheng Y, Addo-Quaye C et al (2010) Transcriptome-wide identification of microRNA targets in rice. Plant J 62:742–759. https://doi.org/10.1111/j.1365-313X.2010.04187.x

Meyers BC, Axtell MJ, Bartel B et al (2008) Criteria for annotation of plant MicroRNAs. Plant Cell 20:3186–3190. https://doi.org/10.1105/tpc.108.064311

Michael TP, 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. https://doi.org/10.1111/tpj.13400

Nakano M, Nobuta K, Vemaraju K et al (2006) Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. Nucleic Acids Res 34:D731–D735. https://doi.org/10.1093/nar/gkj077

Scott MS, Ono M (2011) From snoRNA to miRNA: dual function regulatory non-coding RNAs. Biochimie 93:1987–1992. https://doi.org/10.1016/J.BIOCHI.2011.05.026

Song C, Wang C, Zhang C et al (2010) Deep sequencing discovery of novel and conserved microRNAs in trifoliate orange (citrus trifoliata). BMC Genom 11:431. https://doi.org/10.1186/1471-2164-11-431

Taylor RS, Tarver JE, Hiscock SJ, Donoghue PCJ (2014) Evolutionary history of plant microRNAs. Trends Plant Sci 19:175–182. https://doi.org/10.1016/J.TPLANTS.2013.11.008

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. https://doi.org/10.1038/ncomms4311

Wang W, Wu Y, Messing J (2014b) RNA-Seq transcriptome analysis of Spirodela dormancy without reproduction. BMC Genom 15:60. https://doi.org/10.1186/1471-2164-15-60

Willmann MR, Berkowitz ND, Gregory BD (2014) Improved genome-wide mapping of uncapped and cleaved transcripts in eukaryotes-GMUCT 2.0. Methods 67:64–73. https://doi.org/10.1016/j.ymeth.2013.07.003

Yang J, Liu X, Xu B et al (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. https://doi.org/10.1186/1471-2164-14-9

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415. https://doi.org/10.1093/nar/gkg595

Transcriptome Responses of *Spirodela polyrhiza*

13

Paul Fourounjian

Abstract

In order to analyze the transcriptome of any species, RNA-seq has become the gold standard and evolved into a variety of library preparations and sequencing platforms to study more than mRNA abundance. This chapter reviews the transcriptional studies of *Spirodela polyrhiza*, the best-characterized member of the Lemnaceae family in a genomic sense. To date, there have been three studies of its transcriptome. The first two analyzed ribosomal RNA depleted total RNA of fronds and fronds developing into turions after exposure to abscisic acid. The first study analyzed 154 down-regulated genes involved in growth and 208 upregulated genes involved in starch, anthocyanin production, and seed development. The second study found 66 sites where chloroplast mRNAs were edited to create a functional protein, supporting the hypothesis that mRNA editing was evolved once, and the conservation of editing sites was phylogenetically correlated. The third study, also performed in the 7498 ecotypes, was sequencing of the uncapped polyadenylated transcripts. While the main aim was to observe miRNA induced cleavage, differences in the post-transcriptional regulation or abundance of degraded transcripts across the eight sequencing conditions can be observed. Taken together, these studies cover mRNA expression, post-transcriptional editing, and finally degradation.

Scientists have been interested in gene expression ever since discovering the central dogma of biology and have developed a number of methods over the decades to measure RNA quantity. After reverse transcription was discovered in 1970, Northern blot and Sanger sequencing followed in 1977, qPCR came out in the late 1980s, and then in 2005, the Roche 454 sequencing platform applied shotgun genome sequencing technology to massively parallel RNA-sequencing and quantification (Cieślik and Chinnaiyan 2017). The data from these Roche and Illumina sequencers were typically thousands to millions of 50–200 nt reads that need to be mapped to the genome and reassembled to determine splicing patterns and gene expression as fragments per kilobase per million (FPKM). Joining the next-generation sequencers are the high throughput, long-read sequencers like PacBio and Oxford Nanopore systems available in 2011 which often produce 20–200 kB reads that can easily span entire mRNA and long non-coding RNA transcripts (1–2 kb), thereby eliminating the reassembly steps to more precisely map the transcriptome with its splicing patterns and alternative polyadenylation sites. Unfortunately, the present challenge with these

P. Fourounjian (✉)
Waksman Institute of Microbiology, Rutgers University, Piscataway 08854, USA
e-mail: pjf99@scarletmail.rutgers.edu

X. H. Cao et al. (eds.), *The Duckweed Genomes*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-11045-1_13

reads is their indels and mismatches which can be corrected via deeper PacBio sequencing of the same read, alignment with the reference genome, and alignment with short-read sequencing, all of which can be combined (An et al. 2018a). With these technologies, it is now possible to accurately characterize and measure the transcriptome of virtually any species with a reference genome or through *de novo* assembly.

When looking at the RNA-seq studies of the duckweed family, we see that most of the gene expression analyses have been done in the recently sequenced *Lemna minor*, or in *Landoltia punctata* and *Lemna aequinoctialis* with *de novo* transcriptomes assembled from small reads (An et al. 2018b). These species have genome sizes ranging from 379 to 650 Mb and de novo transcriptomes of 74,797 and 72,105 unique contigs, while *Lemna minor* has 22,382 annotated genes, and *Spirodela* has 19,623 and 18,507 in strains 7498 and 9509, respectively, suggesting a wide variety of possible mRNAs from the roughly 20,000 genes found in Lemnaceae genomes (Tao et al. 2013; Wang et al. 2014a, 2016; Van Hoeck et al. 2015; Michael et al. 2017; Yu et al. 2017). In addition to the transcriptomic analysis of ABA-induced turion formation, there have been a couple of other RNA-seq experiments within *Spirodela polyrhiza* that measure aspects other than gene expression to understand the full complexity of these plant transcriptomes. A reinvestigation of turion development RNA-seq identified the chloroplast genes that undergo mRNA editing and how this relates to the rest of the monocots. Another study was the sequencing of the degraded RNAs in *Spirodela* within eight conditions as a measure of miRNA induced cleavage, which could also show a rough measure of expression. These three studies provide researchers a chance to witness mRNA expression, editing, and degradation.

One unique aspect of the duckweed lifecycle is the formation of turions. Their role as an asexual organ of perennation makes them analogous to both seeds, which sexually give rise to an entire organism, and tubers or buds, which asexually survive winter and other unfavorable conditions. In 2014, an RNA-seq study was performed on *Spirodela* during the development of turions, making it the first genome wide-study of gene expression in a plant tuber, which was followed by a study of potatoes in 2015 (Wang et al. 2014b). It was previously discovered that 3 days of exposure to 10 μM abscisic acid, ABA, induced irreversible turion development and an increase of two enzymes involved in starch and cell wall production. This study, therefore, compared four biological replicates of *Spirodela* fronds with and without 3 days of exposure to ABA by sequencing 15–41 million 75 bp reads of ribosomal, rRNA, depleted total RNA on a SOLiD 5500 instrument. They were able to map reads to the nuclear, mitochondrial, and chloroplast genomes, with 28–39% of reads deriving from the organelles. Reads were aligned with bowtie and tophat, normalized and compared in cufflinks, and annotated for GO term enrichment through blast2go and GOseq. The results showed 154 genes down-regulated during turion development, meaning that they were minimally four-fold less abundant, with roughly half being 0.2–0.1 the expression compared to control. These results had a false discovery rate, FDR, less than 0.01 thanks to the eight biological replicates. The 154 down-regulated genes were largely involved in carbon fixation, protein synthesis, DNA replication, and growth in general since turions no longer grow. For the 208 upregulated genes, the GO term enrichment showed that many of these upregulated genes functioned in starch and anthocyanin production, hormone response and signal transduction, cell wall synthesis, and seed dehydration. There were 13 genes in cell wall and anthocyanin production that were specific to turion induction. Similar to a desiccating seed of a terrestrial plant developing turions upregulated five and expressed two previously silent genes of the late embryogenesis abundant protein family. These LEA family proteins protect other proteins and confer resistance to dehydration, salinity, and cold stress. This transcriptomic study was properly timed to observe not only the structural changes of turion development, but the signaling pathway. They noticed upregulation of seven ABA-responsive, three ethylene-responsive, and two heat shock

responsive transcription factors. There were also ABA transcription factor binding sites in 30 of the upregulated genes, while 119 had a bind site for ethylene-responsive transcription factors. This pathway matches the ABA or environment triggered, calcium-dependent signal pathway observed in maturing seeds, reinforcing the similarity of turions and seeds on a molecular, invisible level.

The same authors performed a second investigation of the rRNA depleted RNA-seq experiment in fronds and developing turions (Wang et al. 2015). Since 26% of the total RNA sequenced mapped to the chloroplast genome, they had 1000-fold coverage of most genes after stringent filtering. The PPR proteins are a massive family, characterized by the 35 amino acid pentatricopeptide repeat motif that specifically binds the 4th and 34th residues in the pfam model to an RNA base, creating a pattern of these motifs that bind to a specific RNA sequence (Barkan et al. 2012; Manna 2015). While these PPR proteins are found in prokaryotes and eukaryotes acting in splicing, processing, editing, stability, and translation of RNAs, this study focused on the DYW-type PPRs that correct certain missense mutations in the plastid genome by editing the mRNA from a cytosine to uracil residue, thereby creating a functional mRNA and protein product. Mapping the RNA-seq reads and detecting C to U SNPs with SAMtools revealed 66 sites of RNA editing with an average efficiency of 76% and a range of 6–100%. Comparison to developing turions showed very similar gene expression with no differentially expressed genes. There were, however, six over and five under edited sites (>two-fold difference, p value <0.05) in seven genes during turion development compared to fronds. So while expression was constant, 1/6th of the sites were differentially edited, thereby altering the functional protein abundance of seven genes. These differences in editing efficiency even varied as much as 8–100% at multiple sites within the same gene due to the sequence-specific nature of PPR protein editing. A phylogenetic analysis with the Mega6 program revealed the 66 editing sites in *Spirodela* had an 81% overlap with the 75 in coconut. There was a 42 and 38% overlap with the 35 and 26 sites observed in the more evolutionarily distant rice and maize. This correlation confirms the hypothesis of a single origin of RNA editing PPR proteins in the early land plants like ferns, that have hundreds of edited sites that were gradually reduced and differentiated over time to ~80 in the basal monocots and 25–40 in the more recent angiosperm species.

Another indirect study of *Spirodela* mRNA expression was the degradome experiment found in Fourounjian et al. (2019). The primary purpose of this experiment was to confirm the cleavage activity of miRNAs on target mRNAs and observe regulatory differences between the biological triplicates of the control, 0 °C, 37 °C, ABA, kinetin, copper, nitrate, and sucrose conditions by sequencing 28–63 million uncapped mRNAs per library. While this degradome sequencing is not a perfect correlate to mRNA expression, the normalized read count of each gene (not kilobase normalized), its expression pattern can be viewed in this program hosted by the Myers laboratory of the Danforth Center https://mpss.danforthcenter.org/~private/dbs/index.php?SITE=messing_SPIRODELA_PARE. These patterns can even reveal unannotated exons, since all reads were polyadenylated. Finally, the miRNA cleavage study revealed that 15 genes, mainly well-conserved transcription factors were expressed and cleaved in four or more conditions, while 71% of the results were condition-specific targets, many of which had more structural and metabolic functions. Of these conditional specific changes, sucrose created the largest difference, followed by copper and heat exposure. This large transcriptomic and metabolic change of sucrose addition suggests that laboratory experiments modeling duckweeds in outdoor applications should avoid this often added media component.

The assembly of the *Spirodela* genome for strains 7498 and 9509 provided not only a scaffold for easy and accurate mapping of RNA-seq data, but a context for the gene expression. This is both in a physical sense for the chromatin

modeling and DNA methylation studies (Cao et al. 2016; Michael et al. 2017), and in a physiological sense where the studies of turion development, for example, can be linked to the results (Kuehdorf et al. 2014). As it stands the research community can observe the Spirodela transcriptome in two or eight conditions as mRNA expression, editing, and degradation. It is expected that the transcriptional research will expand to include more stimuli exposure, tissue specific, life cycle, and microbe interaction experiments. These *Spirodela* genomes and transcriptome studies facilitate research across the family by providing a reference for the other genomes or transcriptomes. Even in cases of de novo assemblies and isoform sequencing of any other related species the *Spirodela* genome will provide annotated and characterized gene models.

References

An D, Cao H, Li C et al (2018a) Isoform sequencing and state-of-art applications for unravelling complexity of plant transcriptomes. Genes (Basel) 9:43. https://doi.org/10.3390/genes9010043

An D, Li C, Zhou Y et al (2018b) Genomes and transcriptomes of duckweeds. Front Chem 6. https://doi.org/10.3389/fchem.2018.00230

Barkan A, Rojas M, Fujii S et al (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. PLoS Genet 8:e1002910. https://doi.org/10.1371/journal.pgen.1002910

Cao HX, Vu GTH, Wang W et al (2016) The map-based genome sequence of *Spirodela polyrhiza* aligned with its chromosomes, a reference for karyotype evolution. New Phytol 209:354–363. https://doi.org/10.1111/nph.13592

Cieślik M, Chinnaiyan AM (2017) Cancer transcriptome profiling at the juncture of clinical translation. Nat Rev Genet 19:93–109. https://doi.org/10.1038/nrg.2017.96

Fourounjian P, Tang J, Tanyolac B, Feng Y, Gelfand B, Kakrana A, Tu M, Wakim C, Meyers BC, Ma J, Messing J (2019) Post-transcriptional adaptation of the aquatic plant under stress and hormonal stimuli. Plant J

Kuehdorf K, Jetschke G, Ballani L, Appenroth K-J (2014) The clonal dependence of turion formation in the duckweed *Spirodela polyrhiza*—an ecogeographical approach. Physiol Plant 150:46–54. https://doi.org/10.1111/ppl.12065

Manna S (2015) An overview of pentatricopeptide repeat proteins and their applications. Biochimie 113:93–99. https://doi.org/10.1016/J.BIOCHI.2015.04.004

Michael TP, 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. https://doi.org/10.1111/tpj.13400

Tao X, Fang Y, Xiao Y et al (2013) Comparative transcriptome analysis to investigate the high starch accumulation of duckweed (*Landoltia punctata*) under nutrient starvation. Biotechnol Biofuels 6:72. https://doi.org/10.1186/1754-6834-6-72

Van Hoeck A, Horemans N, Monsieurs P et al (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 8:1–13. https://doi.org/10.1186/s13068-015-0381-1

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. https://doi.org/10.1038/ncomms4311

Wang W, Wu Y, Messing J (2014b) RNA-Seq transcriptome analysis of Spirodela dormancy without reproduction. BMC Genom 15:60. https://doi.org/10.1186/1471-2164-15-60

Wang W, Zhang W, Wu Y et al (2015) RNA editing in chloroplasts of *Spirodela polyrhiza*, an aquatic monocotelydonous species. PLoS ONE 10:e0140285. https://doi.org/10.1371/journal.pone.0140285

Wang W, Li R, Zhu Q et al (2016) Transcriptomic and physiological analysis of common duckweed Lemna minor responses to NH4+ toxicity. BMC Plant Biol 16:92. https://doi.org/10.1186/s12870-016-0774-8

Yu C, Zhao X, Qi G et al (2017) Integrated analysis of transcriptome and metabolites reveals an essential role of metabolic flux in starch accumulation under nitrogen starvation in duckweed. Biotechnol Biofuels 10:167. https://doi.org/10.1186/s13068-017-0851-8

Repetitive Sequences: Impacts and Uses in the Spirodela Genome

8

Paul Fourounjian

Abstract

Repetitive DNA, consisting of small and large satellite repeats and transposable elements, comprises over 50% of most plant genomes. The Lemnaceae family demonstrates a ~12-fold difference in genome size and relatively similar number of genes, indicating a wide variability in repeat content. The best studied genome of the family *Spirodela polyrhiza* had a normal total satellite DNA content, yet a surprisingly high 50% of those were dinucleotide microsatellite repeats. The telomeres and 119 bp centromere repeats were typical, although ribosomal repeats appear scarce. Genomic studies showed a small number of 24nt heterochromatic siRNAs accompanied by the lowest rate of DNA methylation seen in any plant sequenced at 9% and low rates of heterochromatin formation. Despite this low level of regulation, the transposable elements are unexpectedly rare and old. In fact, they even show high rates of DNA methylation and high rates of inactivation through illegitimate recombination. This suggests that the scarce 24nt siRNAs are surprisingly effective and an intriguing topic of further research.

P. Fourounjian (✉)
Waksman Institute of Microbiology, Rutgers University, Piscataway 08854, USA
e-mail: pjf99@scarletmail.rutgers.edu

In the early years of DNA and chromosome research, structural components of chromosomes were noticed as patterns in DNA and protein stains, often in the centromeric or telomeric regions. Once DNA sequencing began it was uncovered that virtually all eukaryotic genomes contain significant portions of repetitive DNA, previously thought of as "junk DNA" (Biscotti et al. 2015). In plants, repetitive elements comprise the majority of most genomes sequenced, ranging from a mere 14% in the grain teff to 85% in maize (Wendel et al. 2016). These repetitive elements can be categorized into tandem repeats which aid in chromosome structure, and longer interspersed repeats derived from transposable elements (TEs). As of 2018 there are two published sequences for *Spirodela polyrhiza* clones 7498 and 9509, and the *Lemna minor* 5500, along with draft genomes of two *Lemna* species *minor* and *gibba* and the *Wolffia* species *australiana* (Unpublished), (Wang et al. 2014; Van Hoeck et al. 2015; Ernst and Martienssen 2016; Michael et al. 2017). Similar to other angiosperms as a whole, these genomes vary considerably in size, but not significantly in gene number (Table 8.1). The Lemnaceae family displays a 12-fold difference in genome size from the smallest sequenced monocot *Spirodela polyrhiza* to the 1881 megabase *Wolffia arrhiza* (Wang et al. 2011). A recent review on plant genome architecture summarized that these size variations between genomes are due to common whole genome duplication, followed by

X. H. Cao et al. (eds.), *The Duckweed Genomes*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-11045-1_8

Table 8.1 Lemnaceae genome size and gene content

Species, clone	Genome size (Megabases)	Gene copy #
S. polyrhiza, 7498	158	19,623
S. polyrhiza, 9509	158	18,507
L. minor, 5500	481	22,382
L. minor, 8627	800	NA
L. gibba, 7742	450	21,830
W. australiana	~380	NA

reduction of coding genes, and proliferation of transposable elements (Wendel et al. 2016). Taken in summary, these repeats play a large role in genomic size and composition and chromosomal structure, in the duckweeds and eukaryotes as a whole.

When DNA was separated by density gradient centrifugation tandem repeats with differential AT/GC content created satellite bands above and below the majority of DNA eventually leading to the name satellite DNA. These tandem repeats range in size from the 180 bp corresponding to a nucleosome to tiny 2 nucleotide microsatellite repeats. They were found to have structural implications in centromeres and telomeres where they maintain heterochromatic structure, and disruptions of their expression have been shown to lead to genomic instability and cancer (Biscotti et al. 2015). The strain 7498 genome study showed that the small *Spirodela polyrhiza* genome had a normal number of satellite DNA repeats, at 1.3% of the genome. Yet while most plants have 10–100 bp minisatellites making up roughly half of the total satellite DNA, strain 7498 satellite DNA was 50% microsatellite repeats, largely comprised of GA repeats, which may have been mutated from methylated CG heterochromatin sequences (Wang et al. 2014; Michael et al. 2017). For the *Lemna minor* 5500 genome, we know that satellite and microsatellite repeats made up 0.6 and 3% of the genome, indicating a similar enrichment of microsatellite repeats (Van Hoeck et al. 2015). In a follow-up study assembling the 32 pseudo-molecules into 20 chromosomes relied on the telomeric repeats of TTTAGGG and the suspected centromeric repeats to help support the confidence of the pseudomolecule assembly (Cao et al. 2016). Another analysis of the 7498 and 9509 strains of *Spirodela* was run using longer reads for better resolution of repeat regions and found a high homology with few indels and less than 0.06% heterozygosity in SNPs. They found that a previously reported 138 bp centromeric repeat was found at 1 centromere and that 19 of 20 chromosomes contained large numbers of a 119 bp centromeric repeat (Melters et al. 2013; Michael et al. 2017). Additionally, they found an extremely low ribosomal DNA copy number of 81 compared to 570 in the similarly sized *Arabidopsis thaliana* genome. In summary, while the centromeres and telomeres of *Spirodela polyrhiza* are consistent with other plant genomes, the microsatellite repeats are very abundant and the ribosomal repeats are very rare.

Probably, the most interesting repeat elements are the transposable elements (TEs), which include DNA copying transposons, RNA copying retrotransposons with autonomous versions capable of replicating themselves and non-autonomous versions of each. Thanks to this replication potential, these selfish genes are always attempting to proliferate, while the plant host genome is perpetually suppressing them and removing them through illegitimate recombination. This push and pull occurring in countless plant species shows that of our crop plants TEs can comprise as little as 14% of the genome in teff and as much as 85% in maize (Wendel et al. 2016). In the annotation of the 7498 genome, LTR retrotransposons were annotated based on homology and found to be 15.5% of the genome, which agreed with its size, while the transposons were too distant from their homologs in their genomes an unable to be annotated (Wang

et al. 2014). This lack of homology is due to the age of the transposons, which mutate over time. In *Spirodela,* the relatively few LTRs (264) had an average age of 4.3 million years, while the average in Brachypodium and rice was found to be 1.8 and 0.7 million years, respectively. In the later analysis of the 9509 genome, TEs were annotated by homology to other known TEs, and by mapping 22–24nt siRNAs known to regulate them through methylation. This showed that the genome is 25% TEs, with a Gypsy/Copia ratio of 1.5. In accordance with the age of the LTRs, the *Spirodela* genome was found to be purging them through illegitimate recombination resulting in the highest ratio of deactivated solo to intact LTRs seen in any plant genome.

After the *Spirodela* 7498 genome was published, the draft genome of *Lemna minor* 5500 was published due to its importance in ecotoxicological studies (Van Hoeck et al. 2015). While *Lemna minor* strains vary in genome size from 323 to 760 Mb strain 5500 is 481 Mb in size and only has 14% more annotated genes than *Spirodela polyrhiza* 7498 (Table 8.1). Compared to *Spirodela* 94.5% of the difference in genome size is due to repeats. These repeats make up 61% of the genome and 36% of the genome is TEs, mainly retrotransposons, which is slightly higher than *Spirodela*. The count of LTRs increased ~10-fold to 210,531. There was a final category of unclassified repeats that made up 21% of the genome. In strain, 7498 DNA-based transposons were difficult to annotate based on their old age and low homology, and in strain, 9509 the annotation relied on siRNAs. Therefore, the unclassified repeats may include many ancient unannotated transposons.

The relative lack of TEs in *Spirodela* brought attention to the RNA directed DNA methylation (RdDM) pathway. This is a mechanism of silencing transposons through siRNAs where Pol IV creates a ssRNA transcript and RDR2 makes it a dsRNA (Matzke et al. 2015). Then, DCL3 cleaves it into 24nt het-siRNAs (heterochromatic) that are loaded onto AGO4, which binds to DRM2 and RDM1 proteins that methylate the 5' end of cytosine in GC, CHG, and CHH sequences. To finish the process a collection of proteins in a histone-modifying complex converts the methylated TE sequence to silenced heterochromatin. This pathway is highly conserved across all land plants, with the notable outlier of the Norway Spruce, which has relatively few 24nt het-siRNAs, mainly localized to reproductive organs (Matzke et al. 2015).

In *Spirodela polyrhiza,* it was noticed that 24nt sRNAs were rare, comprising 7.3% of the small RNAs in strain LT5a and 1% in strain 7498 (Fourounjian et al. 2019). While the 9509 genome had the lowest DNA methylation rate of any plant sequenced at 9%, the TEs had an average methylation rate of 20% (Michael et al. 2017). Furthermore, older TEs were annotated based on the mapping of 22–24nt siRNAs, suggesting that they were expressed and active. The *Spirodela* genome also revealed a low number of old TEs suggesting that it has been very successful at halting their proliferation (Wang et al. 2014; Michael et al. 2017). Taken together it looks like the RdDM pathway is working with little to no 24nt het-siRNAs. This could be similar to the results seen in Norway spruce where 24nt het-siRNAs are localized to flowers, which are very rare in *Spirodela*, or perhaps other mechanisms may be at play. The mystery of how the Lemnaceae, particularly *Spirodela*, regulate their TEs is an exciting field of research that is still currently unfolding.

Acknowledgements Thank you to Dr. Alex Harkess for reviewing this chapter to confirm its accurate, but not complete description of the RdDM pathway.

References

Biscotti MA, Olmo E, Heslop-Harrison JS (2015) Repetitive DNA in eukaryotic genomes. Chromosom Res 23:415–420. https://doi.org/10.1007/s10577-015-9499-z

Cao HX, Vu GTH, Wang W et al (2016) The map-based genome sequence of Spirodela polyrhiza aligned with its chromosomes, a reference for karyotype evolution. New Phytol 209:354–363. https://doi.org/10.1111/nph.13592

Ernst E, Martienssen R (2016) Status of the Lemna gibba 7742a and Lemna minor 8627 genomes. Duckweed Forum, 12

Fourounjian P, Tang J, Tanyolac B, Feng Y, Gelfand B, Kakrana A, Tu M, Wakim C, Meyers BC, Ma J, Messing J (2019) Post-transcriptional adaptation of the aquatic plant under stress and hormonal stimuli. Plant J

Matzke MA, Kanno T, Matzke AJM (2015) RNA-directed DNA methylation: the evolution of a complex epigenetic pathway in flowering plants. Annu Rev Plant Biol 66:243–267. https://doi.org/10.1146/annurev-arplant-043014-114633

Melters DP, Bradnam KR, Young HA et al (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol 14:R10. https://doi.org/10.1186/gb-2013-14-1-r10

Michael TP, 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. https://doi.org/10.1111/tpj.13400

Van Hoeck A, Horemans N, Monsieurs P et al (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. https://doi.org/10.1186/s13068-015-0381-1

Wang W, Haberer G, Gundlach H et al (2014) The Spirodela polyrhiza genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. Nat Commun 5:1–13. https://doi.org/10.1038/ncomms4311

Wang W, Kerstetter RA, Michael TP (2011) Evolution of genome size in Duckweeds (Lemnaceae). J Bot 2011:1–9. https://doi.org/10.1155/2011/570319

Wendel JF, Jackson SA, Meyers BC, Wing RA (2016) Evolution of plant genome architecture. Genome Biol 17:1–14. https://doi.org/10.1186/s13059-016-0908-1

Flowering and seed production across the Lemnaceae

Paul Fourounjian[1*], Janet Slovin[2], Joachim Messing[1,3]
[1] Waksman Institute of Microbiology, Rutgers University, New Jersey, USA
[2] Genetic Improvement of Fruits & Vegetables Laboratory, USDA, Beltsville, MD
[3]Deceased 9/13/2019
[*]Corresponding author

Paul Fourounjian Orchid ID 0000-0002-5714-3892
Janet Slovin Orchid ID 0000-0001-6609-7034
Joachim Messing Orchid ID 0000-0001-8260-2643

Abstract

The Lemnaceae family represents the smallest, simplest, and fastest growing angiosperms. This growth rate is partially due to the family's neotenous lifestyle, where instead of maturing and producing flowers, the plants asexually bud in a juvenile state, with extremely rare maturation and flowering. Here we provide flowering protocols for 3 of the 5 genera to promote further research. While almost all Lemnaceae flowering protocols are circadian or stress induced, we believe this *Wolffia microscopica* flowering protocol is neither. The protocol for flowering *Spirodela polyrhiza* 7498 and 9509 strains enables sequencing studies in the best characterized strains in the family. Then the different phenotypes of *Lemna gibba* and *minor* strains provide a chance to study essential genes of floral development, while *Lemna gibba* provides a rare breeding protocol. Hopefully these 3 flowering protocols below facilitate further research of this neotenous family, and these rare flowers.

i. W. *microscopica* in 16 hour daylength in E media in either 6 well plates or flasks
ii. *Lemna gibba* and *minor* in continuous light in E media ±20μM salicylic acid
iii. *Spirodela polyrhiza* in continuous light in 50ml petri dishes of Hoagland's media ±1.5μM salicylic acid

Keywords: Lemnaceae, flowering protocols, miRNAs, ODNs

Flowering and seed production across the Lemnaceae

Paul Fourounjian[1*], Janet Slovin[2], Joachim Messing[1,3]
[1] Waksman Institute of Microbiology, Rutgers University, New Jersey, USA
[2] Genetic Improvement of Fruits & Vegetables Laboratory, USDA, Beltsville, MD
[3]Deceased 9/13/2019
[*]Corresponding author

Intro

The 5 genera and 37 species comprising the Lemnaceae family, commonly known as duckweeds, are the smallest, fastest growing, most morphologically reduced, and widely distributed family of angiosperms (Landolt and Kandeler 1987; Ziegler et al. 2015). They have a small (0.5mm - 2cm), flat leaf-like frond. The *Spirodela*, *Landoltia*, and *Lemna* genera have rhizoid structures while the smaller, simpler, and more recently evolved *Wolffiella* and *Wolffia* genera lack these structures (Landolt and Kandeler 1987). The duckweeds are found floating and rapidly clonally dividing on still, nutrient rich, waters worldwide. They're also a promising crop that can grow at rates of 13-38 dry tons/hectare year while cleaning agricultural or industrial wastewater and producing a biomass that can be used as animal feed or biofuel depending on whether it is concentrating heavy metal or other hazardous materials out of the water (Cao et al. 2018; Fourounjian et al. 2020, Skillicorn et al., 1993). Alternatively, they can be a resource-efficient food source with *Wolffia* either wild harvested or grown in greenhouses, and *Lemna* species used to extract protein concentrates (Cao et al., 2018, Fourounjian et al. 2020). Deeper academic understanding of this family's unique biology, will improve the ways we use them to sustainably provide clean water, food, and energy in the future.

Due to their small size, asexual, clonal growth, ease of aseptic cultivation, and simple morphology, the Lemnaceae were used as model plants from 1950 to 1990 for studying plant biology topics ranging from hormone and amino acid synthesis to chemical and temperature responses (Landolt and Kandeler 1987; Zhao et al. 2012). Dr. Landolt collected hundreds of asexual duckweed strains cataloging them with a 4 digit code. These are now stored at the Rutgers Duckweed Stock Cooperative (RDSC) with newer strains assigned a 3 digit code eg. DWC130. Now Lemnaceae research includes genomics and transcriptomics studies thanks to the publications of the *Spirodela polyrhiza* 7498 and 9509 and *Lemna minor* 5500 and other upcoming genome drafts (Wang et al. 2014; Van Hoeck et al. 2015; Michael et al. 2017; An et al. 2018).

Members of the Lemnaceae grow in a juvenile and asexual, aka neotenous, manner indefinitely, and then produce flowers in a matter of days if exposed to the correct stimuli. Therefore their floral regulation can be seen as a simple presence or absence of flowers, instead of the earlier or later flowering, complicated by multiple pathways, studied in other plants (Spanudakis and Jackson 2014). The Lemnaceae were used to search for the mythical "florigen" first hypothesized in 1937 (Chaïlakhyan 1937), and The Family of Lemnaceae Volume 1,2, (1987) and (Pieterse 2013) describe 112

studies that investigated factors affecting floral regulation. After day-length, which was paramount in every study, salicylic acid (SA) was the most studied inducer. The chelating agent ethylenediamine-N,N′-bis(2-hydroxyphenylacetic acid) (EDDHA), was also tested in many studies.

While many experiments and studies of the duckweeds have found ways to induce flowering in the laboratory, only 2 accessible studies have resulted in seeds and are therefore suitable for breeding protocols. While studying IAA accumulation in a mutant line of *Lemna gibba* G3, Slovin and Cohen (1988) found they could induce flowering by growing on E media with continuous light and shaking for 1 hour twice a day. Seeds that fell to the bottom of the culture flask then germinated on fresh media. Recently two strains of *L. gibba*, were found to optimally flower on 20μM SA. In modified Hoagland's medium solidified with 0.6% agar (Fu et al. 2017). Under these conditions, They saw that *Lemna gibba* G3, strain DWC114 (from the RDSC) produced flowers and viable pollen on E+20μM SA, while the other strain (5504) had non-viable pollen. Cross-pollination of strain 5504 with strain DWC114 pollen produced a hybrid *Lemna* clone (Fu et al. 2017).

With this knowledge of flowering and the availability of new strains and genomes we sought to create flowering and breeding protocols in 3 of the 5 genera, and expand this re-emerging field of research.

Results

Wolffia microscopica

With *Wolffia microscopica* 2005 recently rediscovered (Sree et al. 2015) we attempted to develop a consistent flowering protocol for *W. microscopica* . We screened day length and EDDHA content in flasks of Hoagland's (Hg) media finding that sucrose greatly increased flowering rate, that 16 hours of light improved flowering rate, and that 25 or 75μM of EDDHA greatly increased flowering rate and survivability (Online Resource 1). Further tests in 50ml petri dishes demonstrated great variability in flowering rate and no influence of EDDHA on flowering rate (Online Resource 2). For non-flowering populations flasks of Hg media or E media with sucrose in either 16 or 24 hours of light work very well, although rare (<1%) instances of flasks with a few flowers have been observed (Online Resource 1). The container that *Wolffia* is grown in was shown to cause a great difference in flowering rate from 0% in 100ml flasks to 38±2% in 6 well plates with 10ml of E media, providing a convenient flowering protocol (Fig. 1). This protocol produced plentiful fertile pistils, with a drop of stigmatic fluid almost the size of the frond, and dehiscent anthers (Fig. 2) Pollen from these anthers had a 56% (94 of 169) viability rate, and 28% pollen tube formation rate (65 of 235) (Fig. 2). Despite gentle shaking pollination of fertile flowering samples fruits and seeds were never observed.

Spirodela polyrhiza

In order to create a flowering protocol for *Spirodela* we screened the sequenced strains 7498 and 9509 in 3 medias E media, Hg, and Schenk Hildebrandt (SH) with 0-3μM SA and 25, 75, and 150 μM of EDDHA (Online Resource 3). E media, E media at pH 5.8, and EDDHA were unsuccessful and lethal. The flowering optimums of 4% were

SH and Hg media with 1.5 or 2 μM SA in 24 hr light. Since SH media was lethal to strain 9509 by day 14 experiments were continued in Hg media. Next we screened different containers, which showed significantly higher flowering rates for 7498, that 50ml petri dishes and 100ml flasks outperformed 50ml flasks, and that optimal SA was 1 and 1.5 μM (Online Resource 4). One month old instead of 1-2 week old input biomass had similar results of 6% flowering rate in 7498 and 1% in 9509 (Online Resource 5). Since Far-red/red light ratio is known to affect phytochrome B and flowering time in many species (Halliday et al. 1994), both strains were grown with and without 1.5 μM SA in continuous blue and red light with 38 μmoles photons /m^2/s of Far red light and 24 hours and 16 hours of light from the typical fluorescent "daylight bulbs". Flowering was only seen in 7498 grown with SA under continuous light from "daylight"bulbs" (Online Resource 6). Transferring strain 7498 from 16 hour days to 50ml petri dishes of Hg+1.5μMSA in 12 hour days could induce flowering rates of 6±3% at 10 days (Online Resource 7), yet replication experiments resulted in little to no floral induction (Online Resource 6), discouraging this approach to floral induction.

Despite the persistently low flowering rate and fertile flower rate, *Spirodela* flowers were available to study. Both strains produced fertile pistils and dehiscent anthers, although 7498 did so at a higher rate (Fig. 3). Growth in petri dishes occasionally effected the morphology of the flowers, and possibly their abundance, with 7498 flowers growing in the middle of the frond instead of the usual meristem locations, and 9509 flowers occasionally never leaving the spathe, aborting, and detaching. One example even showed the purple underside covering the anthers, and a rhizoid sticking vertically up (Fig. 4). This odd morphology may be responsible for further reducing the flowering rate of strain 9509. Analyzing the dehiscent anthers showed that 69% (161 of 234) pollen grains were viable and 48% (47 of 97) germinated in 7498, while 9509 had 26% (207 of 803) and 9.7% (10 of 103) viability and germination rate (Fig. 5). All flasks and plates were examined for settled fruits and seeds, yet only turions were found at an average rate of 518 ±98 and 309 ±66 per 100ml flask for 7498 and 9509 respectively. There was no statistically significant relationship between turion count and SA content or daylength. Gentle rotary shaking of fertile and developing flowers of both strains, and manual self-pollination of fertile pistils of 7498 with dehiscent anthers did not result in fruit formation.

Lemna

Lemna gibba G3 was first described by Kandeler in 1955, and over time it had 5 named strains in the RDSC inventory in 2017. These are: DWC114, which was used by Fu et al. (2017); DWC130 and DWC131 the parental and auxin mutant line respectively described by Slovin and Cohen (1988); DWC132, contributed by Biolex Corporation, and DWC128 from the Waksman Collection, which we didn't order. After seeing no flowers from DWC130, and the results of (Fu et al. 2017) we ordered 4 of the 5 *Lemna gibba* G3 strains. While DWC131was listed as the larger auxin mutant line, DWC114 was in fact the largest of the 4 strains indicating likely mislabeling. After seeing marked differences in floral development between the 4 strains we DNA barcoded them with the chloroplast atpH-F primers (Wang et al. 2010). Barcoding revealed that DWC114 is *L. gibba*, while DWC130, DWC131, and probably DWC132 are *L. minor* (Table 1). Since DWC114 is the normal size for *Lemna gibba*, the jsR1 mutant appears to have been lost

in one of the collections multiple transfers. Our colleagues at the RDSC have updated their database accordingly, to ensure the accuracy of the global collection.

As preliminary screening for flowering rate all 4 strains of previously labelled *L. gibba* G3 were grown in flasks of E media under continuous light with 0-100μM SA, and in 16 or 24 hour daylength with either 20μM SA or 20μM SA and 75μM EDDHA (Online Resource 8). This showed that DWC114 had the highest flowering rate, that 24 hours of light promoted flowering, that the optimum SA concentration was 20μM, and that EDDHA had no significant additive effect on floral induction. To confirm the high flowering rate of DWC114 at lower SA concentrations E + 0, 10, 20, and 30 μM SA flasks were tested in triplicate, in 16 and 24 hour day length, indicating similar flower rates and seed yields for 10-30μM SA (Online Resource 9).

These experiments revealed differences in flowering rate as well as morphology in the conditions we tested. Fig. 6 shows the different flowers of all 4 strains. Strain DWC114 produced perfect flowers, with a small pistil and stigmatic fluid drop and 2 dehiscent anthers. Strains DWC131 and DWC132 both produced large pistils and large drops of stigmatic fluid, with two white, non-dehiscent anthers, prone to detaching and sinking to the bottom of the flask. Strain DWC130 appeared to produce no flowers when viewed through 3.5X magnification glasses, but under 16X magnification on the stereoscope it was seen to produce small pistils with no stigmatic fluid, and no anthers at all. Thanks to the work of the Zhang laboratory we analyzed the pollen of the 3 strains of *Lemna* testing viability with the Alexander's stain and germination on plates of E media (Peterson et al. 2010; Fu et al. 2017) (Fig. 7). As expected strain DWC114's dehisced anthers' pollen was largely viable and fertile (83% viability), with a 43% pollen tube formation rate (59 of 131). Meanwhile the non-dehiscent anthers of DWC131 and DWC132 were almost entirely non-viable with 3 of 227 and 11 of 454 pollen grains appearing viable respectively and no pollen producing tubes in the hundreds of pollen grains tested.

In the interest of developing breeding protocols we tested DWC114 in the E media with yeast extract and bactopeptone (Eye media) originally described in (Slovin and Cohen 1988), and 75 μM of EDDHA with and without 20μM of SA, and found the results largely similar, possibly due to the utilization of the same pathway, and no improvement from Eye media (Fig. 8). A primary test of 0, 10, 20, 30 μM SA in 16 and 24 hours of light showed flowering rates close to 20% in the samples with SA, and abundant seed production. A second test with 5 and 6 replicate flasks in 16 and 24 hour days respectively showed 20% flowering rates, and 16 ±9 seeds, and 32 ±5 seeds per flask. While this was noticeably lower than the previous experiment, it suggests that continuous light should be used for seed production. The seeds produced have a 65% or higher germination rate by day 3 in a 50ml petri dish of liquid E media, in both 16 or 24 hour day length (Fig. 9). The seeds can be harvested via transfer pipette, left in a petri dish to dry overnight, wrapped in parafilm, and stored at 4°C for long-term storage. If stored in diluted E media, light must be blocked to prevent germination. Wild collected *L. gibba* seeds have been reported germinating at 70% efficiency when stored in water for 2 years at room temperature, with the number dropping to 5.6% for dry storage, and 1% after 3 years of wet storage (Rejmánková 1976). To eliminate fungal contamination seeds were surface sterilized with 5% bleach for 3 or 10 minutes and washed 3 times, or in 70% ethanol, and air dried. Since all 3 methods had germination rates over 97%, and no

contamination from a total of 145 seeds month old seeds, ethanol sterilization is recommended for its convenience.

Tables

Table 1 DNA Barcoding Results of DWC114, DWC130, DWC131, and DWC132
Lemna strains were DNA barcoded using the Choloplast atpH-atpF intergenic region primers according to the protocol (Wang et al. 2010). NCBI BLAST results are provided with atpF-atpH intergenic spacer indicating the Chloroplast ATPase subunit I gene atpF-atpH intergenic spacer. With the matK spacer primers DWC132 also had 4 matches (100% cover, >99% identity) to the *Lemna minor* matK gene or plastid genome before *Lemna trisulca* was the 5th best match.

Match	Description	Query Cover %	E value	Percentage identity	Accession
DWC114 1st	***Lemna gibba*** strain RDSC 5504 atpF-atpH intergenic spacer	83	0.00	99.49	KX212889.1
DWC114 2nd	***Lemna gibba*** strain RDSC 7741 atpF-atpH intergenic spacer	83	0.00	99.49	KX212887.1
DWC130 1st	***Lemna minor*** strain RDSC 7210 atpF-atpH intergenic spacer	100	0.00	99.03	KX212888.1
DWC130 2nd	***Lemna minor*** chloroplast, complete genome	100	0.00	99.03	DQ400350.1
DWC131 1st	***Lemna minor*** strain RDSC 7210 atpF-atpH intergenic spacer	100	0.00	99.38	KX212888.1
DWC131 2nd	***Lemna minor*** chloroplast, complete genome	100	0.00	99.38	DQ400350.1
DWC132 1st	***Lemna japonica*** strain 0216 atpH-atpF intergenic spacer	99	0.00	100	KJ921747.1
DWC132 2nd	***Lemna minor*** strain RDSC 7210 atpF-atpH intergenic spacer	99	0.00	100	KX212888.1

Table 3 Materials used
All medias were pH adjusted with KOH, supplemented with 1% sucrose, and autoclaved. All flasks had cotton stoppers and loose foil covers unless stated otherwise.

Medias	Full Name	concentration	pH	Supplier
E	E	Online Resource 11	4.6	
Hg	Hoagland's	1.63g/L	5.8	(Cassion Labs Smithfield, UT)

SH	Shcenk Hildebrandt	1.6g/L	5.8	(Sigma-Aldrich,St. Louis, MO)
Growth containers	**Media Volume**	**Flask Volume**		**Supplier**
flask	100ml	250ml		(VWR, PA)
50ml flask	50ml	125ml		(VWR, PA)
petri dish	50ml			(Kord-Valmark [2905], USA)
6 well plate	10ml			(Corning, NY)

Figure Legends

Fig. 1 *W. microscopica* flowering rate
Flowering rates of W. *microscopica* in E media in100ml flasks, 50ml petri dishes with and without parafilm, and 10 ml 6 well plates with and without parafilm.

Fig. 2 *W. microscopica* flowers and pollen analysis
A Vertical view of mature pistil and anther. B Horizontal view of mature pistil. C Horizontal view of a dehiscing anther. D,E Alexander pollen stains of the anthers, and pollen grains respectively. Live pollen is pink or red, clear or blue is dead. F Pollen tube formation assay.

Fig. 3 *Spirodela polyrhiza* flowers
A,B Fertile pistil and anther respectively of strain 7498. C,D Developing pistil and anther of strain 9509

Fig. 4 Anomalous *Spirodela polyrhiza* flowers seen in plates
A Strain 7498 often had flowers in the center of the frond. B Developing pistil and anthers of strain 9509 being overgrown by the underside of the frond. The vertical portion appears to be a rhizoid. C Developing pistil and anthers enveloped in a spathe and detached from the frond

Fig. 5 *Spirodela polyrhiza* pollen analysis
A,B, and C are Alexander pollen stains of the anthers and pollen grains, and the pollen tube formation assay respectively of dehiscent anthers of strain 7498.
D,E, and F are Alexander pollen stains of the anthers and pollen grains, and the pollen tube formation assay respectively of dehiscent anthers of strain 9509

Fig. 6 Flowers of 4 *Lemna* strains growing on E medium
A Strain DWC114 Perfect flower with 1 stigma, and 2 dehiscent anthers. B Strain DWC130 pistil with no stigmatic fluid C Strain DWC131 pistil and mature non-dehiscent anther. D Strain DWC132 pistil with a drop of stigmatic fluid.

Fig. 7 Pollen viability and germination assays for *Lemna* DWC114, DWC131, and DWC132

A is strain DWC114 pollen, B is strain DWC131 pollen, C is strain DWC132 pollen
Panels 1 and 2 are Alexander stains of the anthers and pollen grains. The scale bars are 50μm.
Panel 3 is the pollen tube formation assay after 1 hour of germination on E media. The scale bars are 100μm.

Fig. 8 Flowering rate and seed production in strain DWC114
A Flowering rate in Eye media. The * indicates a significant difference ($p<0.05$) between the two conditions.
B Total seeds at the bottom of the flask.

Fig. 9 Seed development and germination in DWC114
A Fruit growing in the meristematic pouch
B Fruit and a seed. The fruit contains 2 seeds
C Ungerminated seeds and germinating seedlings.
D Lateral view of a seedling capable of floating
E Graph of germinated, floating seeds

Online Resource 1 *W. microscopica* flowering rates in Hg media screening EDDHA, sucrose, and daylength and containers

Online Resource 2 *W. microscopica* flowering rates in flask and petri dishes, with and without sucrose and EDDHA

Online Resource 3 *Spirodela polyrhiza* flowering screen in E+, Hg+, and SH+ media 0-3uM SA, 0-150uM EDDHA

Online Resource 4 *Spirodela* Hg+ 0-3uM SA petri dishes, 50ml, and 100ml flasks

Online Resource 5 Month old *Spirodela* 24 hour daylength 1, 1.5, 2 μM SA experiment

Online Resource 6 Summary of *Spirodela* screens and experiments

Online Resource 7 *Spirodela* 16 or 24 hour daylength to 12 hour daylength flowering rate graph

Online Resource 8 Flowering rate screen for 3 strains of *Lemna* graph
A, B, C Flowering rate in 24 hr daylength in flask of E media with 0-100μM SA for strains DWC114, DWC131, and DWC132 respectively. D,E,F Flowering rate in 16 or 24 hour days with E media, E+ 20μM SA, and E+20μM SA + 75μM EDDHA for strains DWC114, DWC131, and DWC132 respectively.

Online Resource 9 DWC114 flower rate and seed count in 16, 24 hour daylength, 0, 10, 20, 30 μM SA

Online Resource 10 E media recipe

Discussion

Wolffia microscopica

While previous publications on flowers almost always focus on a single genera of the Lemnaceae this work was an overview of 3 different genera, with the goal of establishing simple and robust flowering protocols to enable further study of floral regulation and flower physiology. Thanks to the rediscovery of the prolifically flowering *Wolffia microscopica* we started there (Sree et al. 2015). We quickly saw both a high flowering rate, and a high degree of variability between replicates within, and replication of experiments. One possible explanation was epigenetic encoding of the *Wolffia* culture's previous nutrient availability, light regimen, and age of culture. While it can take several generations for methylation changes to be inherited, it is also important to recall from the microscopy that a typical *Wolffia* frond contains 4 or 5 generations of daughter-frond within daughter-frond meristems (Sree et al. 2015), which may further contribute and prolong the epigenetic "memory" of *Wolffia microscopica*. Accordingly, experimental replication was improved once the previous month's history of the input biomass was replicated as Appenroth suggested in Useful Methods 1, issue 8 of the Duckweed Forum. This valuable practice of 2 or 4 week acclimation to any experimental control conditions was carried out throughout the study.

Since the container the *Wolffia* was grown in affected flowering, we wondered what the mechanism was. While culture input, and age, chelators like EDDHA, hormones like SA, gas exchange, and even media surface area to volume ratio likely play roles in regulation of flowering, we chose to investigate the role of evaporation. Despite the importance of flowering and seed formation to avoid dessication, presence or absence of water-retaining parafilm was only a moderate factor, with the volume, or surface area to volume ratio, of the container playing a larger role (Fig. 1). While the mechanism for increased flower induction remains ambiguous, growth of *Wolffia microscopica* 2005 in either a flask, or a parafilm sealed plate provides a convenient protocol to study floral regulation and development with minimal responses to other stimuli.

While we tested the flower inducing effect of EDDHA like previous researchers, it had no effect on flowering rate (Online Resource 2). In addition to its possible chelating effects, EDDHA is hypothesized to break down into an SA-like molecule (Tanaka et al. 1979). So while daylength and stress seem to be the key ingredients of duckweed flowering protocols (Landolt and Kandeler 1987; Pieterse 2013), it appears that W. *microscopica* may be responding to some other stimuli. A literature review of other plant species showed that when biotic or abiotic stress triggers Systemic Acquired Resistance (SAR), Flowering Locus D (FLD) is expressed and indirectly inhibits specific WRKY transcription factors (Jaskiewicz et al. 2011; Singh et al. 2013, 2014a, b; Banday and Nandi 2015). In addition to its role in promoting the SAR response, FLD is well known for its effect in demethylating the histones of Flowering Locus C, which inhibits transcription of the FT and SOC1 floral promoters (He et al. 2003; He and Amasino 2005; Liu et al. 2007). Therefore, it appears the strong overlap between SA response and flowering observed in the Lemnaceae and other plants is largely caused by FLD

demethylating and transcriptionally suppressing FLC, which suppress the key floral promoters. While the orthologs of these proteins are likely involved in most stress-induced Lemnaceae flowering protocols, it would be interesting to see if W. *microscopica* relies on this pathway, or some other non-circadian pathway capable of triggering flowering.

Spirodela polyrhiza

Considering the great genetic and transcriptomic studies of *Spirodela polyrhiza* strains 7498 and 9509, we chose to investigate the molecular aspects of floral regulation. The vast body of floral regulation research in Lemnaceae is largely without evidence of genetic mechanisms since the majority of Lemnaceae flowering research was conducted before the sequencing of the human and Arabidopsis genomes (Initiative 2000; Venter et al. 2001; Consortium 2001), before the first miRNA was discovered in Caenorhabditis (Lee et al. 1993), or miRNAs were found in Arabidopsis (Reinhart et al. 2002). Nearly every overview of gene networks involved in flowering responses to day length, stress, and accumulated sugar states that they are largely regulated by two highly conserved microRNAs, miR156 and miR172 (Aukerman 2003; Wu et al. 2009; Srikanth and Schmid 2011; Spanudakis and Jackson 2014). Between the *Spirodela polyrhiza* strain 7498 and 9509 genomes and the miRNA sequencing in strains LT5a, 7498 and 9509 we see there are ~9 miR156 and ~5 miR172 loci in the genome, that miR156 was the most expressed miRNA family in "Lt5a", and that the miR156:172 ratio ranged from 71-408 in "7498" in 8 different stress and hormonal stimuli conditions (Wang et al. 2014; Michael et al. 2017; Fourounjian et al. 2019). These ratios of the juvenile marker miR156, and the adult marker miR172, of at least one of the members of the family, provide molecular evidence of the neotenous nature that must be overcome to induce flowering.

Since birth order is known to influence growth rate and fitness in *Lemna minor*, and turion induction rate in *Spirodela polyrhiza* we used 4 frond clusters and 3-5 frond clusters of *Lemna* and *Spirodela* respectively, with first daughter fronds still attached to the grandmother frond, to help standardize growth and flowering rate (Barks and Laird 2015, 2016; Mejbel and Simons 2018). Considering the flowering rates of *Wolffia* and *Lemna* in 24 hours of light, and that continuous light disrupts the circadian rhythm after a few days (Miwa et al. 2006), we developed flowering protocols using continuous light. We also tested a 12 hour daylength since most strains of *Spirodela polyrhiza* were classified as day neutral plants (Landolt and Kandeler 1987 pg. 261), but found it had lower flowering rates than 24 hour light (Online Resource 6,7). Therefore growth in continuous light with transfers from 100ml Hg flasks to 50ml peri dishes of Hg+1.5μM SA produced the highest flowering rates of 6% and 1% in strains 7498 and 9509 respectively (Online Resource 6). While both 7498 and 9509 produced fertile pistils and pollen, gentle shaking, and manual self-pollination did not induce fruit or seed development, suggesting the need for crosspollination as seen in (Fu et al. 2017).

Curiously, while a 100ml flask in our best flowering conditions would produce up to 39 mature and developing flowers, the flasks at day 28 would have an average of 240 and 134 turions for 7498 and 9509 respectively, leading us to suspect that turion dormancy may be a more common stress response than flowering for these *Spirodela polyrhiza* strains. Finally, we hope these experiments will be helpful for those interested

in studying flowering in these strains, and insightful for researchers looking into Lemnaceae flowering as a whole.

Lemna

While researching *Lemna* flowers we were surprised to find that in the same conditions DWC130 produced pistils without stigmatic fluid, DWC131 and DWC132 produced stigmatic fluid, and non-dehiscent anthers, and that DWC114 produced pistils with stigmatic fluid and both dehiscent, and non-dehiscent anthers (Fig. 6). Given the marked differences in floral morphology and fertility of these genetically related strains of *Lemna minor* and *gibba* these strains might provide a suitable biomass for transcriptomics sequencing to identify critical genes in floral development that are either present as different alleles, or differentially transcribed between them.

For anyone interested in studying Lemnaceae breeding *Lemna gibba* G3 (strain DWC114) is currently the ideal strain. It flowers readily, provides abundant dehiscent anthers, and self-pollinates to provide the first published duckweed breeding protocols. The first breeding protocol called for *Lemna gibba* G3 to be grown in E media under 24 hours of light, with gentle shaking on a rotary shaker 1 hour twice a day (Slovin and Cohen 1988), while the second suggested growing *Lemna gibba* G3 on plates of Modified Hoagland's media, with 20μM SA, and manually self or cross pollinating the flowers under the stereoscope (Fu et al. 2017). To make future floral and breeding experiments easier we found that G3 will flower and set seed in both 16 and 24 hours of light, with E +20uMSA, and a gentle stir whenever counting flowers. Obviously for cross-pollination with known parents manual pollination under the stereoscope is ideal.

Although DWC114 can self-pollinate, *Spirodela polyrhiza* and *Wolffia microscopica*, seem to be employing some sort of self-incompatibility (SI) mechanism. While many species rely on differential timing of pistil and anther development to avoid self-pollination, this strategy wouldn't benefit a clonal population of Lemnaceae, as stated in Landolt Volume 1, which would increase the importance of SI genes. While the mechanisms involving the determinants at the S locus are diverse, and have been evolved at least 35 times in the angiosperms, very little is known about them in monocots (Fujii et al. 2016), and certain monocot orchids appear to have a novel SI mechanism (Niu et al. 2017). Most understanding of SI in Lemnaceae comes from the flowering to seed setting ratio of populations observed in the wild, which led to the hypothesis that 22 species are likely self-compatible (Landolt Volume 1 pg. 169-179). We know that *Wolffia welwitschii* was able to self-pollinate (Witte 1985), and that 2 *L. minor* populations were self-incompatible, but capable of cross fertilization (Caldwell 1899), as described in (Landolt and Kandeler 1987). These results, along with our own illustrate that SI is variable across the family. The presence of SI may be related to the ability of a species to form turions or sunken fronds, as they all represent dormancy strategies.

Summary

After looking back to previous research for guidance we confirmed that daylength, SA, and EDDHA, which likely breaks down into SA are key inducers flowering in the Lemnaceae. While the mechanisms used to be quite mysterious, we are fortunate today to have a well-studied connection between SA, SAR, FLD, FLC, FT and

ultimately flowering studied mainly in the model plant *Arabidopsis*. While the Lemnaceae rarely flowers in nature we hope the following protocols.

iv. W. *microscopica* in 16 hour daylength in E media in either 6 well plates or flasks
v. *Lemna gibba* and *minor* in continuous light in E ±20μM SA
vi. *Spirodela polyrhiza* in continuous light in 50ml petri dishes of Hg ±1.5μM SA

Will provide convenient methods for inducing and studying these flowers. Now with new protocols and molecular tools a new slew of research questions to study the genetic mechanisms come to mind.
1. Might the flowering *Wolffia microscopica* utilize a stress-free flowering pathway?
2. Will flowering *Spirodela polyrhiza* reveal new insights in transposon mobility and DNA methylation?
3. Will the *Lemna minor* and *gibba* strains reveal the critical genes for pistil fertility and anther formation and dehiscence?
With these and of course many other questions and possibilities in mind it's clear that the study of floral induction and development of other breeding protocols within the Lemnaceae family is exciting in its own right and that these species also provide a unique perspective as a basal monocot model to study the genetic mechanisms of floral regulation by a presence or absence of flowers instead of early or delayed flowering phenotypes.

Methods

Plant material and culture media

All strains of *Spirodela polyrhiza*, *Lemna gibba*, *Lemna minor* and *Wolffia microscopica* described are available from the RDSC. *Wolffia microscopica* 2005 was kindly provided by Professor Appenroth. DWC114, DWC130, DWC131, and DWC132 were barcoded at the chloroplast ATPase subunit I gene atpF-atpH intergenic spacer, according to the PCR protocol described in (Wang et al. 2010) followed by the recommended BLAST search at NCBI.

Medias and containers are described in Table 3. E media was prepared essentially according to Slovin and Cohen (1988), supplemented with 1% sucrose, and adjusted to pH 4.6 (Online Resource 11). Additions of yeast extract and bactopeptone were made before autoclaving. SA and EDDHA were prepared as stock solutions of 100mM in 100% ethanol and water respectively, filter sterilized, and added to cooled media. Flasks were fitted with cotton stoppers and loose aluminum foil covers and petri dishes and 6 well plates were wrapped with parafilm to maintain sterility and reduce evaporation.

New cultures of *Wolffia microscopica* were inoculated with a single 1.5x1.5cm mesh loopful of floating fronds from a 100ml flask with a 3-5 week old healthy, mature culture with sunken fronds. *Spirodela* experiments were started with three 3-5 frond clusters from a 7 or 14 day old 100ml Hg flask. For *L. gibba* and *L. minor* all experiments started with 3 four-frond clusters from a 7 or 14 day old population growing in E media.

Unless specified all experiments were carried out at 23°C under continuous light (32-47 µmol m^{-2} s^{-1}) provided by General Electric Daylight 6500K fluorescent bulbs.

Flower counts, and breeding protocols

Flowers were counted if they were developing or mature, with both a flower having only a pistil, or a perfect flower (pistil with 2 anthers) being counted as one. Flowering % is expressed as developing and mature flowers/ frond number, with the count either being every frond in the flask, or over 100 fronds. Seed counts were the total number per 100ml flask. Seeds were stored dry, or in a flask of depleted media diluted 1:2 with sterile water at 4°C.

Unless otherwise specified, flowers were counted on days 7, 10, 14, 17, 21, 24, 28 after subculture.

Pollen viability and fertility

Pollen from all strains that produced anthers was tested for both viability and fertility. A modified Alexander's stain protocol (Peterson et al. 2010) was used to test viability with live pollen staining red or pink, while inviable pollen appeared blue or clear. Pollen fertility was tested by gently spreading anthers over an agar plate of E media, and looking for pollen tube formation under the stereoscope 1 hour later. Fertility was measured as pollen with tubes/ total pollen.

Author Contributions

PF Designed and conducted experiments. JS and JM conceived, guided, and supervised the research. PF and JS wrote the manuscript.

Acknowledgements

PF and JS dedicate this manuscript to Professor Joachim Messing who conceived and guided the study "to investigate the neotenous nature of the *Spirodela* genome", yet passed away before its completion. PF thanks Kaleb Friedman for helping conduct some of the experiments. PF Thanks Professors Xu, Sree and Appenroth for their insightful discussion during the ICDRA 2019 conference.

References

An D, Li C, Zhou Y, et al (2018) Genomes and Transcriptomes of Duckweeds. Front Chem 6:. doi: 10.3389/fchem.2018.00230

Banday ZZ, Nandi AK (2015) Interconnection between flowering time control and activation of systemic acquired resistance. Front Plant Sci 6:174. doi: 10.3389/FPLS.2015.00174

Barks PM, Laird RA (2016) A multigenerational effect of parental age on offspring size but not fitness in common duckweed (Lemna minor). J Evol Biol 29:748–756. doi: 10.1111/jeb.12823

Barks PM, Laird RA (2015) Senescence in duckweed: Age-related declines in survival, reproduction and offspring quality. Funct Ecol 29:540–548. doi: 10.1111/1365-2435.12359

Caldwell OW (1899) On the Life-History of Lemna minor. Bot Gaz. doi: 10.1086/327786

Cao HX, Fourounjian P, Wang W (2018) The Importance and Potential of Duckweeds as a Model and Crop Plant for Biomass-Based Applications and Beyond. In: Handbook of Environmental Materials Management. Springer International Publishing, Cham, pp 1–16

Chaïlakhyan MK (1937) Hormonal theory of plant development.

Fu L, Huang M, Han B, et al (2017) Flower induction, microscope-aided cross-pollination, and seed production in the duckweed Lemna gibba with discovery of a male-sterile clone. Sci Rep 7:3047. doi: 10.1038/s41598-017-03240-8

Fujii S, Kubo KI, Takayama S (2016) Non-self- and self-recognition models in plant self-incompatibility. Nat. Plants 2

Halliday KJ, Koornneef M, Whitelam GC (1994) Phytochrome B and at Least One Other Phytochrome Mediate the Accelerated Flowering Response of Arabidopsis thaliana L. to Low Red/Far-Red Ratio. Plant Physiol 104:1311–1315. doi: 10.1104/pp.104.4.1311

He Y, Amasino RM (2005) Role of chromatin modification in flowering-time control. Trends Plant Sci 10:30–35. doi: 10.1016/J.TPLANTS.2004.11.003

He Y, Michaels SD, Amasino RM (2003) Regulation of flowering time by histone acetylation in Arabidopsis. Science 302:1751–4. doi: 10.1126/science.1091109

Jaskiewicz M, Conrath U, Peterhälnsel C (2011) Chromatin modification acts as a memory for systemic acquired resistance in the plant stress response. EMBO Rep 12:50–55. doi: 10.1038/embor.2010.186

Landolt E, Kandeler R (1987) Biosystematic investigations in the family of duckweeds (Lemnaceae), Vol. 4: the family of Lemnaceae-a monographic study, Vol. 2 (phytochemistry, physiology

Liu F, Quesada V, Crevillén P, et al (2007) The Arabidopsis RNA-Binding Protein FCA Requires a Lysine-Specific Demethylase 1 Homolog to Downregulate FLC. Mol Cell 28:398–407. doi: 10.1016/J.MOLCEL.2007.10.018

Mejbel HS, Simons AM (2018) Aberrant clones: Birth order generates life history diversity in Greater Duckweed, Spirodela polyrhiza. Ecol Evol 8:2021–2031. doi: 10.1002/ece3.3822

Michael TP, 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. doi: 10.1111/tpj.13400

Miwa K, Serikawa M, Suzuki S, et al (2006) Conserved expression profiles of circadian clock-related genes in two Lemna species showing long-day and short-day photoperiodic flowering responses. Plant Cell Physiol 47:601–612. doi: 10.1093/pcp/pcj027

Niu SC, Huang J, Zhang YQ, et al (2017) Lack of S-RNase-based gametophytic self-incompatibility in orchids suggests that this system evolved after the monocot-eudicot split. Front Plant Sci 8:. doi: 10.3389/fpls.2017.01106

Peterson R, Slovin JP, Chen C (2010) A simplified method for differential staining of aborted and non-aborted pollen grains. Int J Plant Biol 1:13. doi: 10.4081/pb.2010.e13

Pieterse AH (2013) Is flowering in Lemnaceae stress-induced? A review. Aquat Bot

104:1–4. doi: 10.1016/J.AQUABOT.2012.08.002
Rejmánková E (1976) Germination of seeds of Lemna gibba. Folia Geobot Phytotaxon 11:261–267. doi: 10.1007/BF02909473
Singh V, Banday ZZ, Nandi AK (2014a) Exogenous application of histone demethylase inhibitor trans-2-phenylcyclopropylamine mimics FLD loss-of-function phenotype in terms of systemic acquired resistance in Arabidopsis thaliana. Plant Signal Behav 9:. doi: 10.4161/psb.29658
Singh V, Roy S, Giri MK, et al (2013) *Arabidopsis thaliana FLOWERING LOCUS D* Is Required for Systemic Acquired Resistance. Mol Plant-Microbe Interact 26:1079–1088. doi: 10.1094/MPMI-04-13-0096-R
Singh V, Roy S, Singh D, Nandi AK (2014b) Arabidopsis FLOWERING LOCUS D influences systemic-acquired-resistance-induced expression and histone modifications of WRKY genes. J Biosci 39:119–126. doi: 10.1007/s12038-013-9407-7
Skillicorn P, Spira W, Journey W (1993) Duckweed aquaculture: a new aquatic farming system for developing countries.
Slovin JP, Cohen JD (1988) Levels of Indole-3-Acetic Acid in Lemna gibba G-3 and in a Large Lemna Mutant Regenerated from Tissue Culture'
Spanudakis E, Jackson S (2014) The role of microRNAs in the control of flowering time. J Exp Bot 65:365–380. doi: 10.1093/jxb/ert453
Sree KS, Maheshwari SC, Boka K, et al (2015) The duckweed Wolffia microscopica: A unique aquatic monocot. Flora - Morphol Distrib Funct Ecol Plants 210:31–39. doi: 10.1016/J.FLORA.2014.10.006
Tanaka O, Cleland CF, Hillman3 WS (1979) Inhibition of flowering in the long-day plant Lemna gibba G3 by Hutner's medium and its reversal by salicylic acid
Van Hoeck A, Horemans N, Monsieurs P, et al (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. doi: 10.1186/s13068-015-0381-1
Wang W, Haberer G, Gundlach H, et al (2014) The Spirodela polyrhiza genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. Nat Commun 5:1–13. doi: 10.1038/ncomms4311
Wang W, Wu Y, Yan Y, et al (2010) DNA barcoding of the Lemnaceae, a family of aquatic monocots
Witte KE (1985) Zur Kultur von Wolfiella welwitschii Monod. Aqua-Planta 10:7. doi: 10.1007/s00497-019-00375-8
Zhao H, Appenroth K, Landesman L, et al (2012) Duckweed rising at Chengdu: Summary of the 1st International Conference on Duckweed Application and Research. Plant Mol Biol 78:627–632. doi: 10.1007/s11103-012-9889-y
Ziegler P, Adelmann K, Zimmer S, et al (2015) Relative in vitro growth rates of duckweeds (Lemnaceae) - the most rapidly growing higher plants. Plant Biol 17:33–41. doi: 10.1111/plb.12184

Figure Legends

Fig. 1 *W. microscopica* flowering rate
Flowering rates of W. *microscopica* in E media in100ml flasks, 50ml petri dishes with and without parafilm, and 10 ml 6 well plates with and without parafilm.

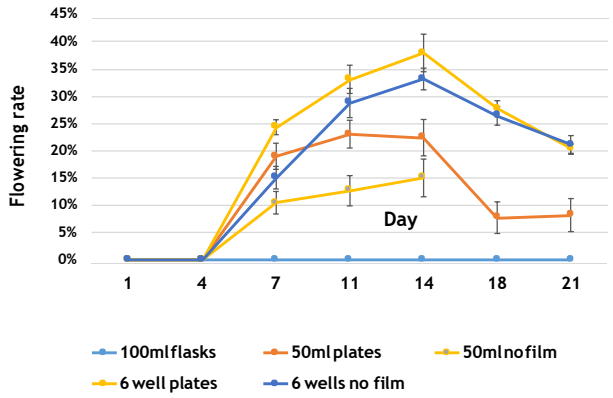


Fig. 2 *W. microscopica* flowers and pollen analysis
A Vertical view of mature pistil and anther. B Horizontal view of mature pistil. C Horizontal view of a dehiscing anther. D, E Alexander pollen stains of the anthers, and pollen grains respectively. Live pollen is pink or red, clear or blue is dead. F Pollen tube formation assay.

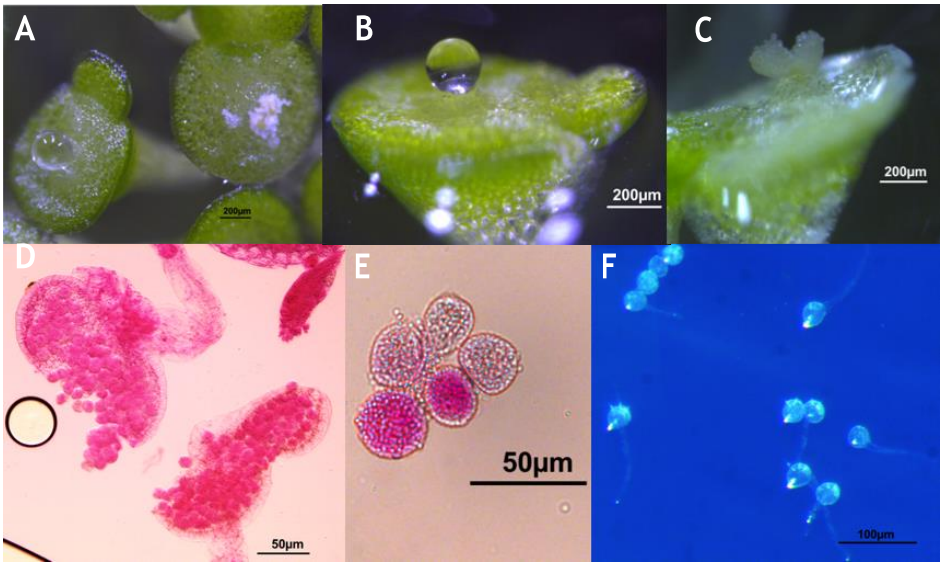


Fig. 3 *Spirodela polyrhiza* flowers
A,B Fertile pistil and anther respectively of strain 7498. C,D Developing pistil and anther of strain 9509

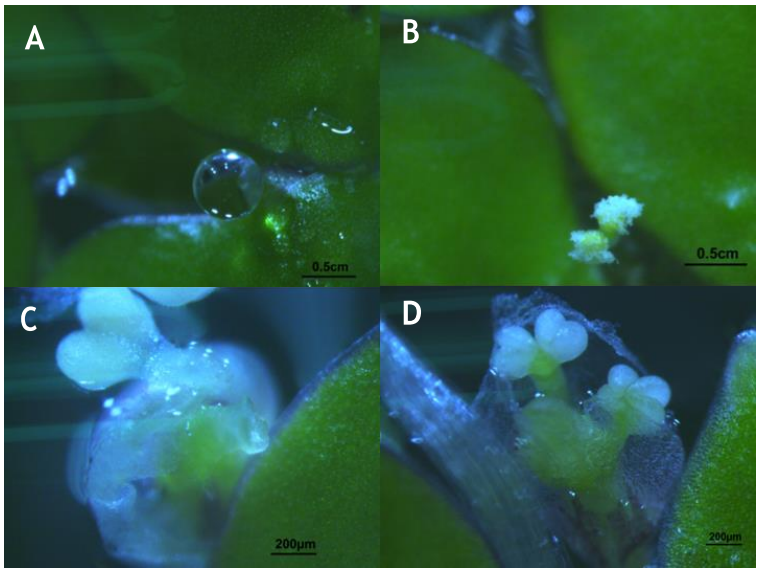


Fig. 4 Anomalous *Spirodela polyrhiza* flowers seen in plates
A Strain 7498 often had flowers in the center of the frond. B Developing pistil and anthers of strain 9509 being overgrown by the underside of the frond. The vertical portion appears to be a rhizoid. C Developing pistil and anthers enveloped in a spathe and detached from the frond

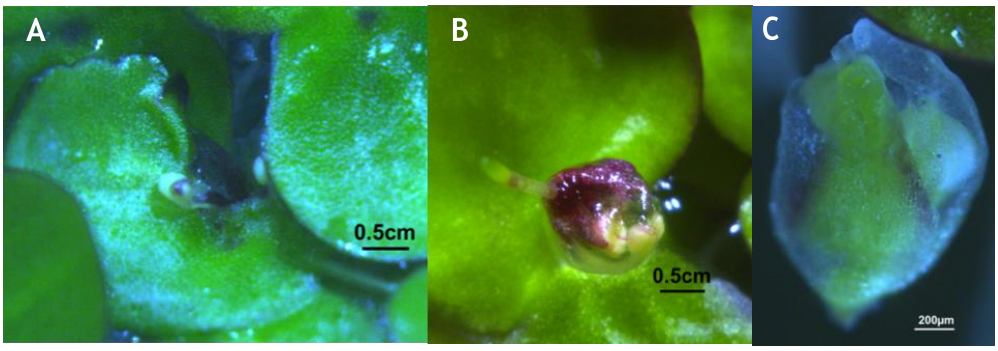


Fig. 5 *Spirodela polyrhiza* pollen analysis
A,B, and C are Alexander pollen stains of the anthers and pollen grains, and the pollen tube formation assay respectively of dehiscent anthers of strain 7498.
D,E, and F are Alexander pollen stains of the anthers and pollen grains, and the pollen tube formation assay respectively of dehiscent anthers of strain 9509

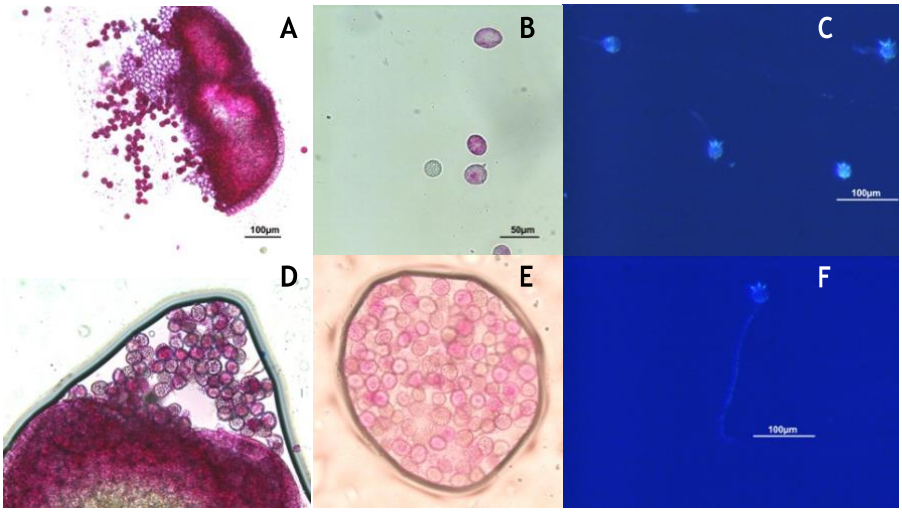


Fig. 6 Flowers of 4 *Lemna* strains growing on E medium
A Strain DWC114 Perfect flower with 1 stigma, and 2 dehiscent anthers. B Strain DWC130 pistil with no stigmatic fluid C Strain DWC131 pistil and mature non-dehiscent anther. D Strain DWC132 pistil with a drop of stigmatic fluid.

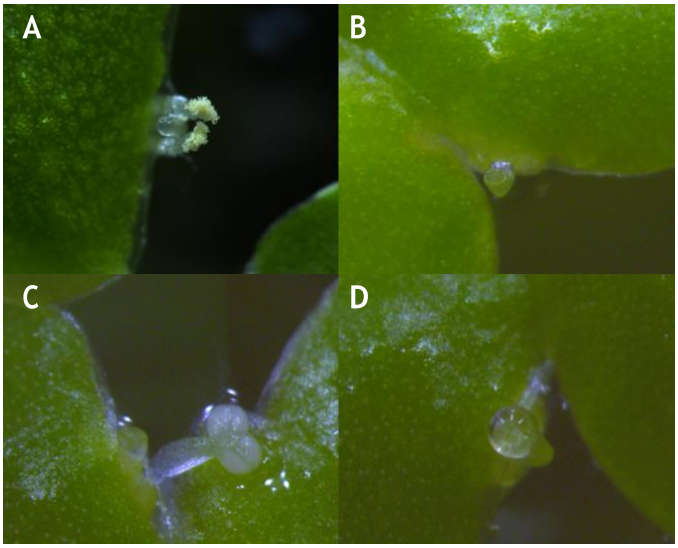


Fig. 7 Pollen viability and germination assays for *Lemna* DWC114, DWC131, and DWC132
A is strain DWC114 pollen, B is strain DWC131 pollen, C is strain DWC132 pollen
Panels 1 and 2 are Alexander stains of the anthers and pollen grains. The scale bars are 50μm.
Panel 3 is the pollen tube formation assay after 1 hour of germination on E media. The scale bars are 100μm.

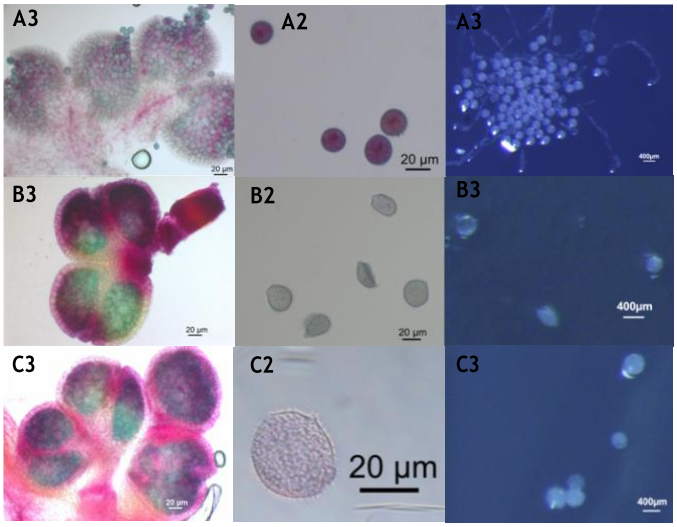


Fig. 8 Flowering rate and seed production in strain DWC114
A Flowering rate in Eye media. The * indicates a significant difference ($p<0.05$) between the two conditions.
B Total seeds at the bottom of the flask.

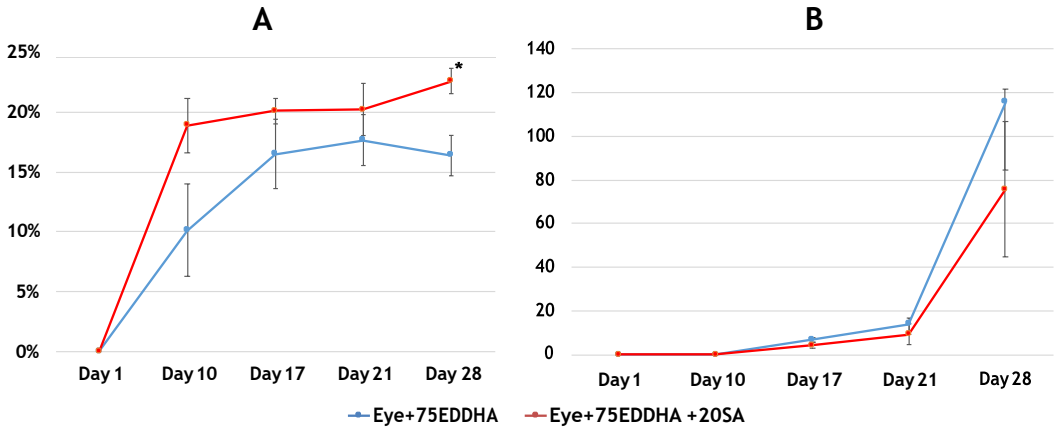
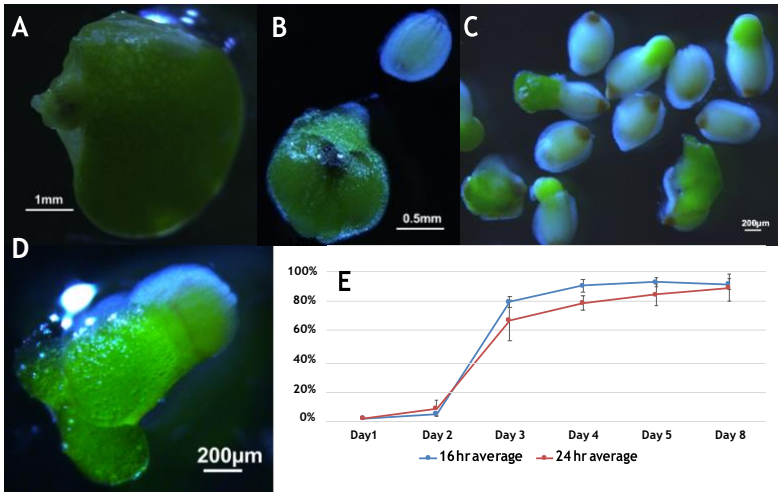


Fig. 9 Seed development and germination in DWC114
A Fruit growing in the meristematic pouch
B Fruit and a seed. The fruit contains 2 seeds
C Ungerminated seeds and germinating seedlings.
D Lateral view of a seedling capable of floating
E Graph of germinated, floating seeds



CONCLUSION

Impact of the Research

After these years of study at a public university I hope my work will contribute to the creation of public goods, both academic and practical. Chapter 1 expanded the *Spirodela* genome map with the miRNA catalog, which we made available to the community before its release in miRBase release 22. I'm proud we also verified 66 miRNAs and 149 targets with degradome sequencing to ensure quality predictions and posted these results in an interactive database. This provided an open, accessible genomic resource while opening further investigation of transposons and the RNA directed DNA methylation (RdDM) pathway in this family described in Chapters 2, 3, and 4.

For Chapter 5 I developed multiple flowering protocols across the family to investigate the genetic mechanisms of floral regulation. Sadly, while Dr. Messing and I created a miRNA catalog and the flowering protocols to study, we ran out of time to investigate these mechanisms through RNA sequencing. We hope other scientists may be intrigued and inspired to pick up these tools and answer these research questions.

While almost all duckweed flowering protocols are either circadian or stress driven (Pieterse 2013), we found a *Wolffia microscopica* protocol that may not rely on either pathway. Therefore sequencing both the flowering and non-flowering *W. microscopica* growing in E media in either flasks or plates may reveal new insights on floral regulation pathways.

Building on the work described in Slovin and Cohen 1988 and Fu et al. 2017, I produced a more convenient DWC114 breeding protocol, which may expedite breeding

research. Additionally the reclassification of DWC130, DWC131, and DWC132 as *Lemna minor* improved the accuracy of the global collection, and provided *Lemna* strains with no anthers, sterile anthers, and dehiscent anthers that may be compared to investigate essential genes of floral development.

Spirodela polyrhiza 7498 offers the simplest and best studied genome in the Lemnaceae making it ideal for genomic and transcriptomic studies (Wang et al. 2014b, a; Fourounjian et al. 2019; An et al. 2019). For those interested in continuing research of this strain, our flowering protocol provides a chance to observe floral regulation with and without circadian influences, using continuous light, which disrupts circadian synchronicity (Miwa et al. 2006). We also propose here a simple method to dissect the genes involved in flowering from the hundreds of genes involved in the salicylic acid (SA) induced systemic acquired resistance (SAR). Comparing the daughter fronds in the control and SA conditions allows identification of SAR genes, while comparing the daughter frond to the flower in the SA condition highlights floral development genes. Finally, while the RdDM pathway is severely compromised in *Spirodela polyrhiza* 7498 and 9509 (Michael et al. 2017; An et al. 2019, Harkess, Bewick et al. unpublished), it would be interesting to see if this pathway is expressed and active during pollen development.

Applications Research and Work

As seen throughout this thesis, Dr. Messing and I shared a passion not only for expanding the knowledge of the academic community, but applying our research to help the world. We both had a passion for applying duckweed to sustainably provide clean water, food,

and fuel. His other research projects like cysteine and methionine rich maize, celiac safe glutenins in maize, sequencing the rice genome, and shotgun genome sequencing will and have helped many people. While understanding neoteny and floral regulation may expedite duckweed breeding to produce better strains for food production or water treatment, we also pursued non-thesis side projects to improve duckweed applications, such as my involvement with the International Lemna Association (ILA) and submerged growth research.

Inspired by a question from professor Matthew Cathell and the Space Lab Technologies system designed to grow duckweed in stacked trays on the International Space Station, I looked into the literature and found that no one had grown duckweed underwater for more than 24 hours. So when Kaleb Friedman joined our lab I worked with him on this project as his independent study. Compared to stacked systems, submerged growth systems hold promise as a way to reduce hardware and fully utilize vertical space, so we shouldn't have been too surprised when Professor Sergey Dolgov presented an image of a similar system at the conference in 2019.

We found the depth that *Wolffia* can survive at with and without sucrose as a carbon source, and then tested the growth rate of aerated vertical cylinders with and without 2% CO_2 supplementation. Despite using filtered pipettes for aeration, we found that fungal and bacterial contaminants thrived in the oxygenated environment. We looked to the literature of fungal pathogen control, and found only 2 papers describing the susceptibility of certain species, and infection rates at different temperatures (Rejmankova et al. 1986; Flaishman et al. 1997). So we tested the effectiveness of different concentrations, and combinations of the antibiotic cefotaxime, and the

commercial aquarium fungicide Pimafix (1% *Pimenta racemose* [Bay Leaf] oil) in preventing bacterial and fungal growth. We're preparing these results as a manuscript that may start and restart research in the exciting and understudied fields of submerged growth and fungal pest prevention.

Once I finished my course curriculum, I joined the ILA's industry conference calls to broaden my knowledge of duckweed applications. Over time I progressed from listening, to being the secretary, to hosting the calls, and even helping develop the Knowing to Growing award where the ILA publicly thanked the authors of the academic paper that advanced the industry the most. Dr. Messing was supportive of this side project, allowing me to present both our floral research, and an ILA poster highlighting industrial production challenges that are promising, understudied areas of academic research at the conference in 2019. This ILA involvement also enabled me to write the applications portion of Cao et al. 2018, and its expanded update the introduction of this thesis; which are the only reviews of the duckweed industry since Landolt and Kandeler 1987. There I collaborated with 10 company leaders to describe their work in order to create a document that could be cited by academic researchers to demonstrate the economic and ecological value of duckweed applications to funding agencies. This chapter was also meant to quickly introduce and educate newcomers to help grow this duckweed research community.

Closing Remarks

Looking at this body of work I see a high confidence miRNA and target catalog, post-transcriptional profiles of *Spirodela* in 8 different conditions and flowering protocols for 4 different species. Still the question of "How do the miRNAs 156 and 172 regulate

flowering in the Lemnaceae?" remains. While I was able to assemble the tools to investigate it, I now see a variety of different approaches using these flowering protocols, the miRNA catalog, and many other methods to continue researching this broad question across the family. My hope is that the scientific community utilizes these resources to gain a deeper understanding of this wonderful plant family with it's unique flowers, and that we might translate that understanding to build a more sustainable future.

References

An D, Zhou Y, Li C, et al (2019) Plant evolution and environmental adaptation unveiled by long-read whole-genome sequencing of Spirodela. Proc Natl Acad Sci U S A 116:18893–18899. doi: 10.1073/pnas.1910401116

Cao HX, Fourounjian P, Wang W (2018) The Importance and Potential of Duckweeds as a Model and Crop Plant for Biomass-Based Applications and Beyond. In: Handbook of Environmental Materials Management. Springer International Publishing, Cham, pp 1–16

Flaishman MA, Hadar E, Hallak-Herr E (1997) First Report of Pythium myriotylum on Lemna gibba in Israel. Plant Dis. 81:550

Fourounjian P, Tang J, Tanyolac B, et al (2019) Post-transcriptional adaptation of the aquatic plant *Spirodela polyrhiza* under stress and hormonal stimuli. Plant J. doi: 10.1111/tpj.14294

Fu L, Huang M, Han B, et al (2017) Flower induction, microscope-aided cross-pollination, and seed production in the duckweed Lemna gibba with discovery of a male-sterile clone. Sci Rep 7:3047. doi: 10.1038/s41598-017-03240-8

Landolt E, Kandeler R (1987) Biosystematic investigations in the family of duckweeds (Lemnaceae), Vol. 4: the family of Lemnaceae-a monographic study, Vol. 2 (phytochemistry, physiology

Michael TP, 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. doi: 10.1111/tpj.13400

Miwa K, Serikawa M, Suzuki S, et al (2006) Conserved expression profiles of circadian clock-related genes in two Lemna species showing long-day and short-day photoperiodic flowering responses. Plant Cell Physiol 47:601–612. doi: 10.1093/pcp/pcj027

Pieterse AH (2013) Is flowering in Lemnaceae stress-induced? A review. Aquat Bot 104:1–4. doi: 10.1016/J.AQUABOT.2012.08.002

Rejmankova E, Blackwell M, Culley DD (1986) Dynamics of fungal infection in duckweeds (Lemnaceae). Veroff Geobot Inst ETH, Stift Rubel, Zurich 87:178–189

Slovin JP, Cohen JD (1988) Levels of Indole-3-Acetic Acid in Lemna gibba G-3 and in a Large Lemna Mutant Regenerated from Tissue Culture'

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. doi: 10.1038/ncomms4311

Wang W, Wu Y, Messing J (2014b) RNA-Seq transcriptome analysis of Spirodela dormancy without reproduction. BMC Genomics 15:60. doi: 10.1186/1471-2164-15-60