



GENOME RESEARCH

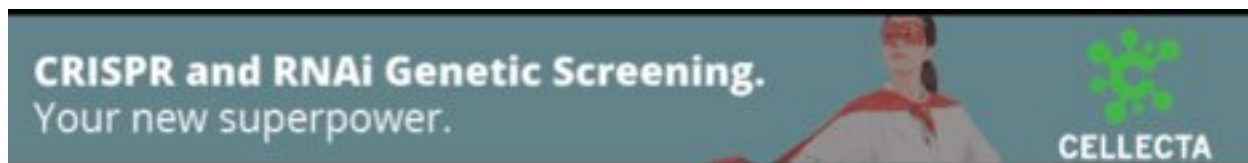
Genome and time-of-day transcriptome of *Wolffia australiana* link morphological minimization with gene loss and less growth control

Todd P. Michael, Evan Ernst, Nolan Hartwick, et al.

Genome Res. published online December 23, 2020

Access the most recent version at doi:[10.1101/gr.266429.120](https://doi.org/10.1101/gr.266429.120)

| | |
|---------------------------------|--|
| P<P | Published online December 23, 2020 in advance of the print journal. |
| Accepted Manuscript | Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version. |
| Open Access | Freely available online through the <i>Genome Research</i> Open Access option. |
| Creative Commons License | This manuscript is Open Access. This article, published in <i>Genome Research</i> , is available under a Creative Commons License (Attribution-NonCommercial 4.0 International license), as described at http://creativecommons.org/licenses/by-nc/4.0/ . |
| Email Alerting Service | Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here . |



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Published by Cold Spring Harbor Laboratory Press

TITLE

Genome and time-of-day transcriptome of *Wolffia australiana* link morphological minimization with gene loss and less growth control

AUTHORS

Todd P. Michael^{1§}, Evan Ernst^{2,10*}, Nolan Hartwick^{1*}, Philomena Chu^{3,&}, Douglas Bryant^{4,&}, Sarah Gilbert^{3&}, Stefan Ortleb⁵, Erin L. Baggs⁶, K. Sowjanya Sree⁷, Klaus J. Appenroth⁸, Joerg Fuchs⁵, Florian Jupe^{1,&}, Justin P. Sandoval¹, Ksenia V. Krasileva⁶, Ljudmylla Borisjuk⁵, Todd C. Mockler⁴, Joseph R. Ecker^{1,9}, Robert A. Martienssen^{2,10}, and Eric Lam^{3§}

AFFILIATIONS

¹Plant Molecular and Cellular Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, USA

²Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY, USA

³Department of Plant Biology, Rutgers, The State University of New Jersey, New Brunswick, NJ USA

⁴Donald Danforth Plant Science Center, St. Louis, MO USA

⁵Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

⁶Department of Plant and Microbial Biology, University of California, Berkeley, 231 Koshland Hall, Berkeley, CA USA

⁷Department of Environmental Science, Central University of Kerala, Periyar-671320,
India

⁸Friedrich Schiller University of Jena, Jena, Germany

⁹Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA,
USA

¹⁰Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, 1 Bungtown Road,
Cold Spring Harbor, NY, USA

*These authors contributed equally to the work

§Corresponding authors

CURRENT ADDRESS

&PC: Department of Energy and Process Engineering, Norwegian University of Science
and Technology, Trondheim, Norway

&FJ: Bayer Crop Science, Chesterfield, MO, USA

&DB: NewLeaf Symbiotics, BRDG Park, 1005 North Warson Rd, St. Louis, MO, USA

&SG: Department of Biology, University of North Carolina, Chapel Hill, NC 27599, USA

CORRESPONDING AUTHORS

Todd P. Michael

Research Professor

Plant Molecular and Cellular Biology Laboratory

Salk Institute for Biological Studies

10010 N Torrey Pines Rd, La Jolla, CA 92037

toddpmichael@gmail.com

Eric Lam

Distinguished Professor

Department of Plant Biology

Rutgers, the State University of New Jersey

Foran Hall, Rm. 216, 59 Dudley Road, New Brunswick, NJ 08901

ericl89@hotmail.com

RUNNING TITLE

Genomics of *Wolffia* sheds light on growth strategy

KEY WORDS

Minimal plant genome, growth optimization, time-of-day expression

ABSTRACT

Rootless plants in the genus *Wolffia* are some of the fastest growing known plant on Earth. *Wolffia* have a reduced body plan, primarily multiplying through a budding-type of asexual reproduction. Here we generated draft reference genomes for *Wolffia australiana* (Benth.) Hartog & Plas, which has the smallest genome size in the genus at 357 Mb and has a reduced set of predicted protein-coding genes at about 15,000. Comparison between multiple high-quality draft genome sequences from *W. australiana* clones confirmed loss of several hundred genes that are highly conserved amongst flowering plants, including genes involved in root developmental and light signaling pathways. *Wolffia* has also lost most of the conserved NLR genes that are known to be involved in innate immunity, as well as those involved in terpene biosynthesis, while having a significant overrepresentation of genes in the sphingolipid pathways that may signify an alternative defense system. Diurnal expression analysis revealed that only 13% of *Wolffia* genes are expressed in a time-of-day (TOD) fashion, which is less than the typical ~40% found in several model plants under the same condition. In contrast to the model plants *Arabidopsis* and rice, many of the pathways associated with multi-cellular and developmental processes are not under TOD control in *W. australiana*, where genes that cycle the condition tested predominantly have carbon processing and chloroplast-related functions. The *Wolffia* genome and TOD expression dataset thus provide insight into the interplay between a streamlined plant body plan and optimized growth.

INTRODUCTION

Wolffia has the distinction of being the duckweed genus with the smallest (*Wolffia angusta*) as well as the fastest growing (*Wolffia microscopica*) species of known flowering plants (Sree et al. 2015b). Plants belonging to this genus are highly reduced in their morphology and anatomy, lacking roots and containing only the green floating frond, which is essentially a fused leaf and stem without any vasculature (Figure 1).

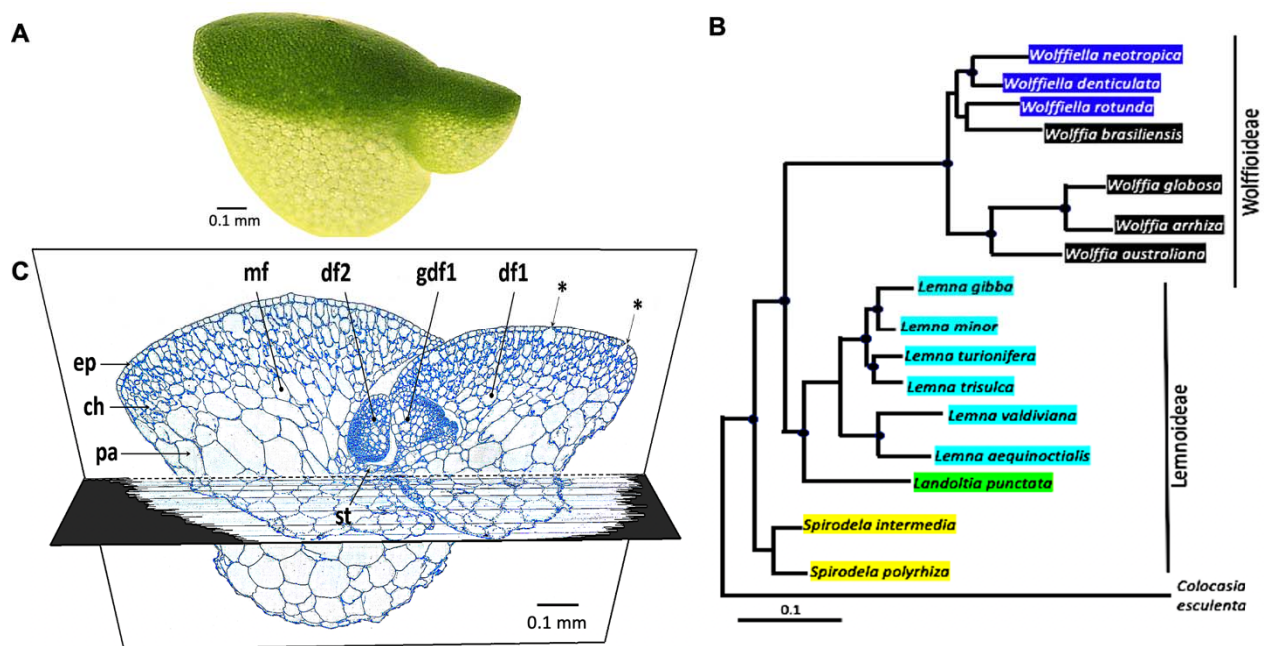


Figure 1. *Wolffia* is a simple plant with a limited number of cells and structures. A)

Brightfield image of wa8730. A video with time tracking of the asexual propagation of wa7733 in

culture can be found in the Supplemental Information. **B)** Phylogenetic relationship between representative species from the 5 genera of duckweed. Each color represents a distinct genus.

Modified from (Borisjuk et al. 2015). **C)** Cross section (1 μ m thick section) of *W. australiana*

stained with methylene blue. **mf**: mother frond; **df**: daughter frond; **gdf**: grand-daughter frond;

ep: epidermal cells; **ch**: chlorenchyma cells; **pa**: parenchyma cells; **asterisks** indicate some of the stomata; **st**: stipe tissue.

Wolffia typically measure only a few millimeters to less than a millimeter in size (Landolt 1986) and grow as colonies of two individuals, one mother frond budding and giving rise to one or more daughter fronds (Figure 1A). Anatomically, however, at least four different generations of plants (total of 10-14 individuals) can be found within one colony, highlighting their adaptive preparedness for fast vegetative multiplication (Sree et al. 2015a; Bernard et al. 1990; Lemon and Posluszny 2000) (Figure 1C; Supplemental Movie S1).

Wolffia is part of a family of aquatic non-grass monocots known as duckweeds (Lemnaceae) (Figure 1B). There are five genera of duckweed (*Spirodela*, *Landoltia*, *Lemna*, *Wolffiella* and *Wolffia*), and their genome sizes span an order of magnitude (Wang et al. 2011; Hoang et al. 2019; Bog et al. 2015). The Greater Duckweed, *Spirodela polyrhiza*, which is most basal with the largest body size and most complex organization, has the smallest genome at 158 megabase (Mb), while in *Wolffia*, the most derived genus, genome sizes range from *W. australiana* at 357 Mb to *W. arrhiza* at 1,881 Mb. *S. polyrhiza* (sp7498) was the first duckweed genome to be sequenced, which revealed a reduced set of protein coding genes at 19,623 that are conserved across flowering plants (Wang et al. 2014; An et al. 2019). A chromosome-resolved genome for a second accession sp9509 revealed that in addition to a reduced gene set, the *Spirodela* genome has low intraspecific variance, highly reduced ribosomal arrays and minimal cytosine methylation, consistent with the genome being specialized for growth (Hoang et al. 2018; Michael et al. 2017).

Green organisms from algae to higher plants partition their biology to coincide with the light-dark cycle, which enhances their ability to anticipate changing conditions (Michael et al. 2003; Dodd et al. 2005; Green et al. 2002; Ferrari et al. 2019). In the model plant *Arabidopsis thaliana*, as much as 90% of its genes are expressed, or phased, to a specific TOD to optimize growth and this global transcriptional regulation is conserved across higher plants (Michael et al. 2008a, 2008b; Filichkin et al. 2011). Since *Wolffia* is the fastest growing angiosperm known to date with a doubling time of as little as a day, we wanted to understand what special adaptations in the genome enabled this rapid growth.

RESULTS

Wolffia genome

We performed whole genome sequencing and time-of-day (TOD) expression profiling of *Wolffia australiana*, the species with the smallest reported genome of the *Wolffia* genus, with an estimated size of 375 Mb for accession wa8730 and 357 Mb for accession wa7733 (Wang et al. 2011). Both accessions are from Australia with wa7733 from Mount Lofty Range, Torrens Gorge in South Australia (34° South, 138° East) and wa8730 from Singleton, Doughboy Hollow, New South Wales (32°South 151°East). These two *W. australiana* accessions wa7733 and wa8730 have doubling times of 1.56 and 1.66 days, respectively (Supplemental Table S1; Supplemental Figure S1), consistent with the previous measurement of 1.39 days for this species (Sree et al. 2015b). We sequenced and *de novo* assembled wa7733 and wa8730 using Pacific

Bioscience (PacBio) single molecule real-time sequencing (SMRT). We also generated BioNano optical maps to correct contigs and help initial scaffolding of the assemblies (Supplemental Table S2), which resulted in final assemblies of 393 and 354 Mb, with longest scaffolds of 5.3 and 1.7 Mb, and scaffold N50 lengths of 836 and 109 kb for wa7733 and wa8730, respectively (Table 1). The assembled genomes sizes are consistent with those predicted by k-mer (k=19) frequency analysis (Supplemental Figure S2), but smaller than predicted by our new flow cytometry estimates possibly reflecting missing high copy number repeat sequence (centromeres) in our assemblies (Supplemental Table S3; Supplemental Materials).

Table 1. Genome assembly statistics.

| | wa7733 | wa8730 | sp9509 | sp7498 |
|--|---------------|---------------|---------------|---------------|
| Genome size estimate flow cytometry (Mb) | 441 | 432 | 157 | 157 |
| Genome size estimate k-mer (Mb) | 343 | 342 | 164 | 185 |
| Genome final assembly (bp) | 393,842,59 | 354,558,44 | 138,592,15 | 145,274,39 |
| Contig (#) | 2,578 | 5,250 | 20 | 21 |
| Genome contig assembly size (bp) | 359,766,21 | 337,899,87 | NA | NA |
| Longest contig (bp) | 1,664,978 | 679,034 | NA | NA |
| N50 contig length (bp) | 256,298 | 102,418 | NA | NA |
| L50 contig (#) | 404 | 1,000 | NA | NA |
| Longest scaffold (bp) | 5,333,369 | 1,714,878 | 11,560,055 | 12,728,324 |
| N50 scaffold length (bp) | 836,551 | 109,493 | 7,949,387 | 8,107,549 |

| | | | | |
|------------------------------|--------|--------|--------|--------|
| L50 scaffold (#) | 122 | 753 | 8 | 8 |
| BUSCO complete scaffolds (%) | 69 | 70 | 79 | 78 |
| Full length LTRs (#) | 2,510 | 1,892 | 801 | 567 |
| Masked (%) | 50 | 50 | 26 | 25 |
| Solo:intact LTR ratio | 11 | 14 | 8 | NA |
| Protein coding genes (#) | 15,312 | 14,324 | 17,510 | 17,057 |
| cDNA mapping (%) | 100 | 99.7 | NA | NA |
| Coverage PacBio (fold) | 91 | 45 | NA | NA |
| Coverage Illumina (fold) | 80 | 68 | NA | NA |
| Mapping PacBio (%) | 98 | 96 | NA | NA |
| Mapping Illumina (%) | 98 | 95 | NA | NA |

NA, Not available

We checked the completeness of these assemblies by mapping 1,780 high confidence Sanger sequenced *W. australiana* cDNA clone sequences deposited at GenBank by the Waksman Student Scholar Program (WSSP). We found that 100% and 99.7% of the cDNAs mapped to the wa7733 and wa8730 genomes, respectively. Also, the PacBio reads used for the assembly and the Illumina reads used for polishing had a >95% mapping rate to their respective assemblies, consistent with the lack of contamination (bacterial) in the sequencing data and the completeness of the assemblies (Table 1). We were also able to identify megabases of putative centromere arrays in both assemblies with three prominent base unit sizes of 126, 167 and 250 bp (Supplemental Figure S3).

Repeat sequence is a dominant driver of genome size differences in plants (Michael and VanBuren 2020). We looked at ribosomal DNA (rDNA) and long terminal

repeat retrotransposon (LTR-RT) content in the *W. australiana* genomes to understand whether, as in *Spirodela*, the smallest genome in this genus (Wang et al. 2011; Hoang et al. 2019) *Wolffia* has also purged repetitive sequences. We found ~200 copies of the rDNA array in the *W. australiana* genomes, double that of *Spirodela* but still half of the number found in *Arabidopsis* (Supplemental Table S4). Next, we identified full-length, intact LTR-RT across both genomes and found 2,510 and 1,892 for wa7733 and wa8730 respectively, which was three times more than *Spirodela* (Table 1). Retroelements (solo + intact) made up over 50% of both of the *Wolffia* genomes, which is twice the repeat content found in *Spirodela* (Michael et al. 2017). It has been shown that the ratio between intact LTR-RTs and solo LTRs (solo:intact), which are left over after illegitimate recombination, is a proxy for how actively the genome is purging proliferating LTR-RTs (Devos et al. 2002). While *Spirodela* has a high solo:intact ratio of 8 (Michael et al. 2017), *Wolffia* has an even higher ratio at 11 to 14 in wa7733 and wa8730 respectively (Table 1), consistent with it more actively purging its TEs leading to a much smaller sized genome than other *Wolffia* species (Wang et al. 2011).

The two draft *Wolffia* genomes are highly collinear with one another at the gene level but there are examples of structural changes and gene loss and gain (Figure 2A;

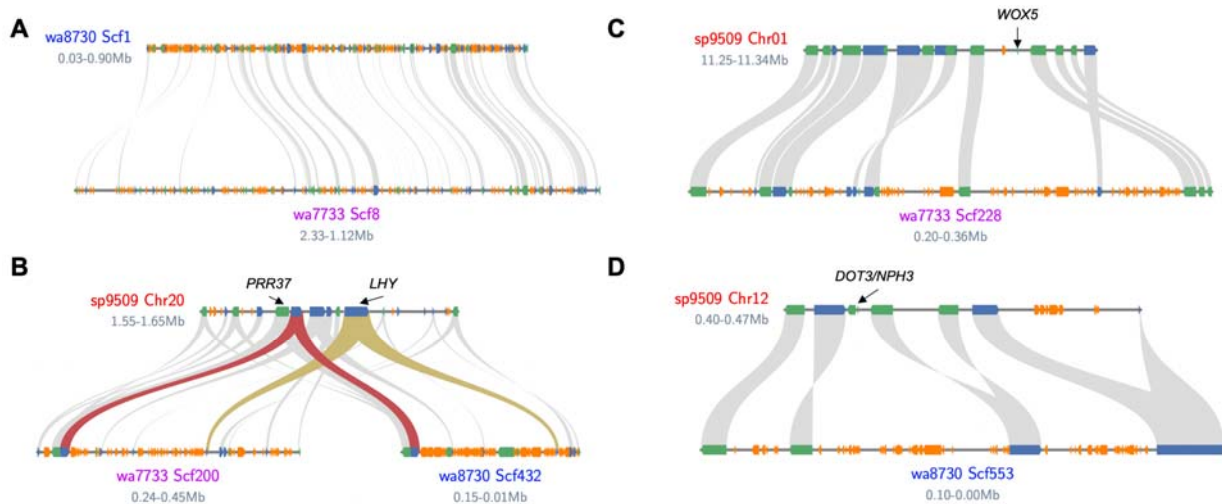


Figure 2. The *Wolffia* genome is collinear with *Spirodela* with bloating and loss of genes due to TEs.

A) wa8730 scaffold 1 (scf1) is collinear with wa7733 scf8. **B)** Conserved core circadian clock genes *PRR37* (red ribbon) and *LHY* (gold ribbon) on sp9509 Chromosome 20 (Chr20) are collinear in wa8730 scf200 and wa7733 scf432. **C)** *WOX5* is lost in *Wolffia* compared to *Spirodela* due to TE insertions (yellow boxes). **D)** *DOT3/NPH3* loci in *Spirodela* lost in *Wolffia* due to LTR-RT insertions. Grey lines represent syntenic connections. Green and blue boxes are forward and

Supplemental Figure S4). Most of the structural variations are associated with insertion/deletions (INDELs) between 50 and 500 bp (Supplemental Figure S5). In addition, both *Wolffia* genomes are largely collinear with *Spirodela* (Figure 2B, C, D; Supplemental Figure S6). One specific example is in the evolutionarily conserved linkage between core circadian clock genes *LATE ELONGATED HYPOCOTYL* (*LHY*) and *PSEUDO RESPONSE REGULATOR 37* (*PRR37*) on Chr20 of the sp9509 genome that is found as far back as bryophytes and shared across other dicots and non-grass monocots (Figure 2B). While the genic regions of the *Wolffia* genomes are mostly collinear with those of the *Spirodela* genomes, the intergenic and repeat sequences are

expanded in the *Wolffia* genomes explaining its larger genome sizes (Figure 2C, D; Supplemental Figure S6).

Comparative BUSCO analysis reveals genes missing in *Wolffia*

As an additional check of genome assembly completeness we leveraged BUSCO (**B**enchmarking **U**niversal **S**ingle-**C**opy **O**rthologs) (Simão et al. 2015), which searches the genome for near-universal single-copy orthologs to establish how much of the gene space has been properly assembled. We found that wa7733 and wa8730 contained 69% and 70% complete BUSCOs respectively (Table 1; Supplemental Table S5). The BUSCO scores for percent complete were lower than we had expected based on the contiguity (N50 length and longest contigs) as well as the cDNA and read mapping results (Table 1). Therefore, we compared the missing BUSCO genes between the two *Wolffia* assemblies and the two previously published high quality *Spirodela* genomes, sp9509 and sp7498 (Wang et al. 2014; Michael et al. 2017; Hoang et al. 2018), and found that most of these "missing genes" were significantly shared across the four duckweed assemblies (Supplemental Figure S7).

It is possible that some of the missing BUSCO genes represent important genes for land plants that are lost in the *Wolffia* genome due to its minimal body plan and lifecycle. Of the 762 and 731 missing BUSCO genes in wa7733 and wa8730 respectively, 574 (76-79%) genes were shared between them (Supplemental Figure S7; Supplemental Table S6). Two specific groups of genes emerged from this list of 574 genes; first, genes that are involved with root initiation and development, and second, genes associated with cell fate and gravitropism. Of particular interest was the loss of

WUSCHEL RELATED HOMEODOMAIN 5 (WOX5), which is a homeodomain transcription factor responsible for root stem cell maintenance in the meristem (Sarkar et al. 2007). Since *Spirodela* has roots, we indeed found a likely ortholog of *WOX5*, while the *Wolffia* version of this gene has been lost and in its expected genomic location are LTR-RT suggesting a mechanism by which this gene may have been lost (Figure 2C). Moreover, *Wolffia* is also missing *TOPLESS RELATED 2 (TPL2)*, which through chromatin-mediated repression specifies where stem cell daughters will exit stem cell fate in *Arabidopsis* (Pi et al. 2015). *Wolffia* is missing other genes associated with stem cell fate (*BLISTER, FEZ*) (Schatlowski et al. 2010; Willemsen et al. 2008), mediator complex (*MED3, MED9, MED33*) (Mathur et al. 2011), and gravitropism (*LAZY1*) (Yoshihara and Spalding 2017), which together are consistent with modified signaling and transcriptional cascades for a rootless, organless plant.

Wolffia has a reduced set of core plant genes

We predicted protein-coding gene structures and performed gene family analysis on the two *Wolffia* genomes and the two published *Spirodela* genomes using a standardized pipeline to ensure consistency (Supplemental Material). Similar to *Spirodela*, we found that *Wolffia* also had reduced gene sets in spite of their larger genome sizes. The wa7733 and wa8730 genome assemblies contain 15,312 and 14,324 predicted protein-coding genes respectively (Table 1), which is several thousand genes less than that found in the *Spirodela* genome (Wang et al. 2014; Michael et al. 2017). We compared predicted proteomes from the four duckweed genomes to 28 proteomes from complete genomes spanning algae, non-seed plants, monocots and dicots (PLAZA v4

monocots) (Van Bel et al. 2018). A multidimensional scaling (MDS) plot of the orthogroups (OGs) placed *Wolffia* and *Spirodela* next to the non-grass monocots and close to the grasses consistent with their evolutionary position (Supplemental Figure S8) (One Thousand Plant Transcriptomes Initiative 2019). Duckweeds were nested between dicot crops, basal plants and a tree, and distant from non-seed plants and algae, consistent with having a core set of higher plant proteins (Wang et al. 2014).

Almost all of the *Spirodela* and *Wolffia* proteins were found in OGs (93-98%; Supplemental Table S7), with the majority (33-41%) having only one protein per OG

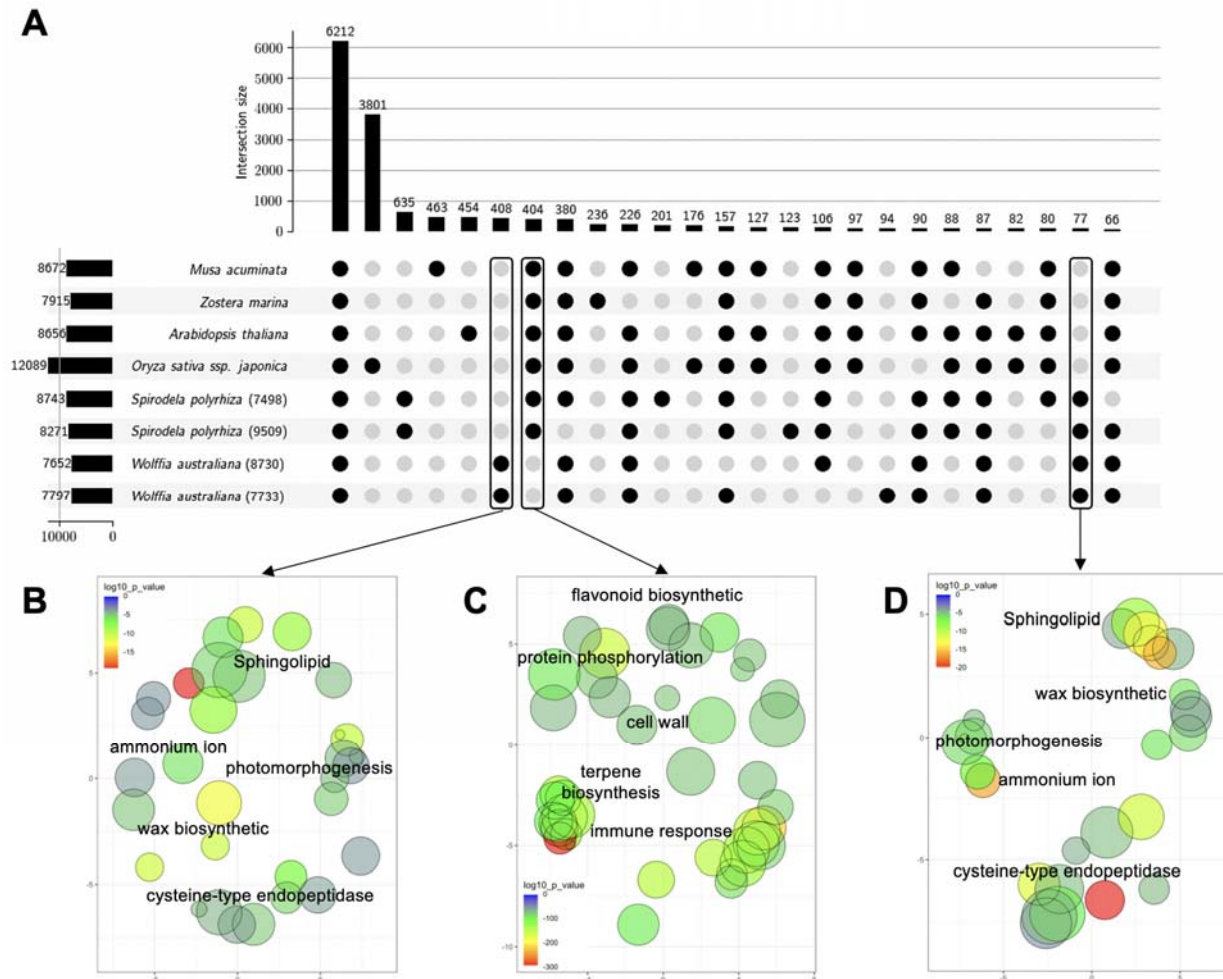


Figure 3. *Wolffia* orthogroups (OGs) analysis reveals significant Gene Ontology (GO) terms relating to growth and defense. A) An upset plot showing the orthogroups (OGs) present (black circles) or missing (grey circle) across the four duckweed genomes (sp9509, sp7498, wa7733 and wa8730), model species (*Arabidopsis thaliana*, *Oryza sativa ssp. Japonica*), and two non-grass monocots (*Musa acuminata*, *Zostera marina*). Boxes indicate *Wolffia*-specific, *Wolffia*-missing and duckweed-specific genes. B) Significant (FDR < 0.05) Gene Ontology (GO) terms derived from *Wolffia*-specific OGs (A) were plotted in two-dimensional semantic space by multidimensional scaling (MDS) with the color reflecting significance (P-value) and circle size reflecting GO frequency. C) GO terms from *Wolffia* missing OGs. D) GO terms from duckweed-specific OGs. All GO terms were summarized using semantic similarity (SimRel), and P-values from overrepresentation with REVIGO.

(Supplemental Figure S9), which means that duckweed has a core set of proteins with few retained paralogs (expanded families). In contrast, species like *Arabidopsis*, rice, *Brachypodium* and Maize have almost 20% of their proteins in OGs with more than 10 paralogs (Supplemental Figure S9). There were 408 and 635 *Wolffia* and *Spirodela* specific OGs respectively (as compared to rice, *Arabidopsis*, *Zostera* and banana), and 77 OGs exclusive to both (Figure 3A). The OGs unique to *Wolffia* can be summarized into the significant Gene Ontology (GO) categories (FDR <0.05) of sphingolipid biosynthesis, photomorphogenesis, wax biosynthetic and cysteine-type endopeptidase (Figure 3B; Supplemental Table S8); we found an overlapping set duckweed-specific significant GO terms (Figure 3D). We took a look at the genes that made up these *Wolffia* unique OGs using our annotation, and found that they are associated with cell wall architecture (Fasciclin-like arabinogalactan proteins) (Johnson et al. 2011), environment-specific expression orchestration (Nuclear transcription factor Y) (Zhao et al. 2016), flowering time (Casein kinase 1-like HD16) (Hori et al. 2013), and sphingolipids (Huby et al. 2020) (Figure 3B; Supplemental Table S9).

In contrast, the significant OGs (FDR >0.05) missing in *Wolffia* included cell wall, flavonoid biosynthesis, protein phosphorylation, immune response and terpene biosynthesis (Figure 3C; Supplemental Table S10). Looking at specific genes with known function suggests there is a loss of OGs involved in meristem development (*FAF*, *BON1*, *SCL3/11*, *POLAR*, *PSD*, *SZ1*), chromatin (*MED4*, *POLD3*), and light signaling (*BBX12*, *DOT3/NPH3*, *GBF4*) (Supplemental Table S10). Another OG that is completely missing in *Wolffia* but has a large family in *Arabidopsis* (n=11), rice (n=9) and most other land

plants, is the *CASPARIAN STRIP MEMBRANE DOMAIN PROTEIN (CASPs)* family (Roppolo et al. 2014).

Some of the most significant missing GO terms were in the terpene biosynthesis pathway, while the most significant *Wolffia*-specific as well as duckweed-specific GOs were in the sphingolipid related pathways (Figure 3B, D). Since both terpenes and sphingolipids may play a predominant role in plant defense (Singh and Sharma 2015; Huby et al. 2020), it is possible that *Wolffia* has traded the terpene pathway for sphingolipids or that the aquatic environment favors the latter. Related to genes involved in defense, one of the most conserved gene families in plants is the NLR (Nucleotide-binding Leucine-rich Repeat) superfamily encoding hundreds of disease resistance (R) genes that provide innate immunity to pathogen associated molecular patterns or effector triggered resistance responses that are associated with systemic activation of broad spectrum immunity (Jones and Dangl 2006). Members of this family of proteins are known to be the most rapidly evolving genes in plants and they are likely under strong selection pressure (van Wersch and Li 2019). In the sp9509 genome, we have previously annotated only 58 NLR genes in contrast to the 178 NLRs that are known in *Arabidopsis* and 387 in *Brachypodium* (Michael et al. 2017). In the wa7733 and wa8730 genomes, we found only a single canonical NLR gene. While it has conserved homologs in the *Spirodela* genomes, it is very divergent from NLR genes of other species. Additionally, two NLR-like genes that contained incomplete NB-ARC domains were identified in the *Wolffia* genome assemblies and these genes are highly conserved across duckweeds, *Arabidopsis* and rice.

Another gene loss was associated with light signaling and the circadian clock in *Wolffia*. We reasoned that since *Wolffia* has apparently optimized for fast growth through rapid multiplication that it may have expanded light signaling and circadian gene families in order to better fine-tune its physiological response to the environment. However, we observed the opposite with *Wolffia*, which only had one third of light signaling and circadian clock genes as compared to that in other land plants (30 vs. ~90), and half of those found in *Spirodela* and *Zostera* (Supplemental Table S11). Of the conserved single copy BUSCO genes that are also light and circadian-related, *Wolffia* is missing *WITH NO LYSINE (K) KINASE 1 (WNK1)* (Murakami-Kojima et al. 2002), *TANDEM ZINC KNUCKLE (TZP)* (Loudet et al. 2008), *FAR1-RELATED FACTOR1 (FRF1)* (Ma and Li 2018) and several of the light harvesting complex genes. In addition, *Wolffia* is missing the core clock component *TIMING OF CAB EXPRESSION 1 (TOC1/PRR1)*, which is also missing in *Spirodela* and *Zostera*, yet present in all other non-grass monocots. *TOC1/PRR1* specifically binds the promoter of *CELL DIVISION CONTROL 6 (CDC6)* to regulate the time of DNA replication licensing and growth in *Arabidopsis* (Fung-Uceda et al. 2018). In contrast, *Wolffia* had similar numbers of proteins for flowering time (*FLOWERING LOCUS T [FT]* and *CONSTANS [CO]*) and temperature response (*HEAT SHOCK PROTEINS [HSP]* and *C-repeat/DRE-Binding Factor [CBF]*) (Supplemental Table S11).

Time-of-day (TOD) expression networks

W. australiana doubles in just over a day (Supplemental Table S1), and since growth is controlled through TOD expression networks, we set out to produce a

temporally resolved transcriptome dataset for one of the two *Wolffia* accessions.

Wa8730 was grown under standard diurnal conditions of photocycles and constant temperature, which is 12 hours (hrs) of light (L) at 20°C, and 12 hrs of dark (D) at 20°C (referred to as LDHH, light/dark/hot/hot) for three weeks after transfer. Fronds were then sampled every 4 hrs over 2 days for a total of 13 time points (Supplemental Figure S10). We found that 83% (11,870) of the predicted genes were expressed significantly under the LDHH condition (Supplemental Figure S11; Supplemental Table S12). We estimated the number of genes showing cycling behavior using HAYSTACK ($R > 0.8$) and found that 13% (1,638) of the expressed genes displayed a TOD expression pattern under the LDHH condition (Supplemental Figure S11; Supplemental Table S13). We also predicted cycling genes using another popular tool called JTKcycle and found slightly fewer cycling genes (11.5%) with significant overlap with the HAYSTACK results (Supplemental Figure S12; Supplemental Table S14); all subsequent analyses were conducted with the HAYSTACK results. Wa8730 genes displayed peak expression, or phase, every hour over the day, similar to TOD time courses under the LDHH condition in other plants, with more genes peaking at morning or evening specific phases (Figure 4E; Supplemental Figure S12). A possible contributing factor for the reduced number of TOD controlled genes in *Wolffia* under the LDHH condition is the lower percentage of transcription factors (TFs) that cycle (11%) compared to other species like *Arabidopsis* and rice (Supplemental Table S17).

The phase of expression of core clock genes is conserved across species (Filichkin et al. 2011). Therefore, we looked at the expression of the core clock-related proteins to both validate the time-course and establish if their expression is also

conserved in *Wolffia*. At the core of the clock is a family of single MYB (sMYB) domain transcription factors *LHY* and *CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)*, along with a related family called *REVEILLE (RVE)* (McClung 2019). While *Arabidopsis* has 12 sMYB genes similar to *CCA1/LHY/RVE*, *Wolffia* has 4 and only 2 cycle with a dawn

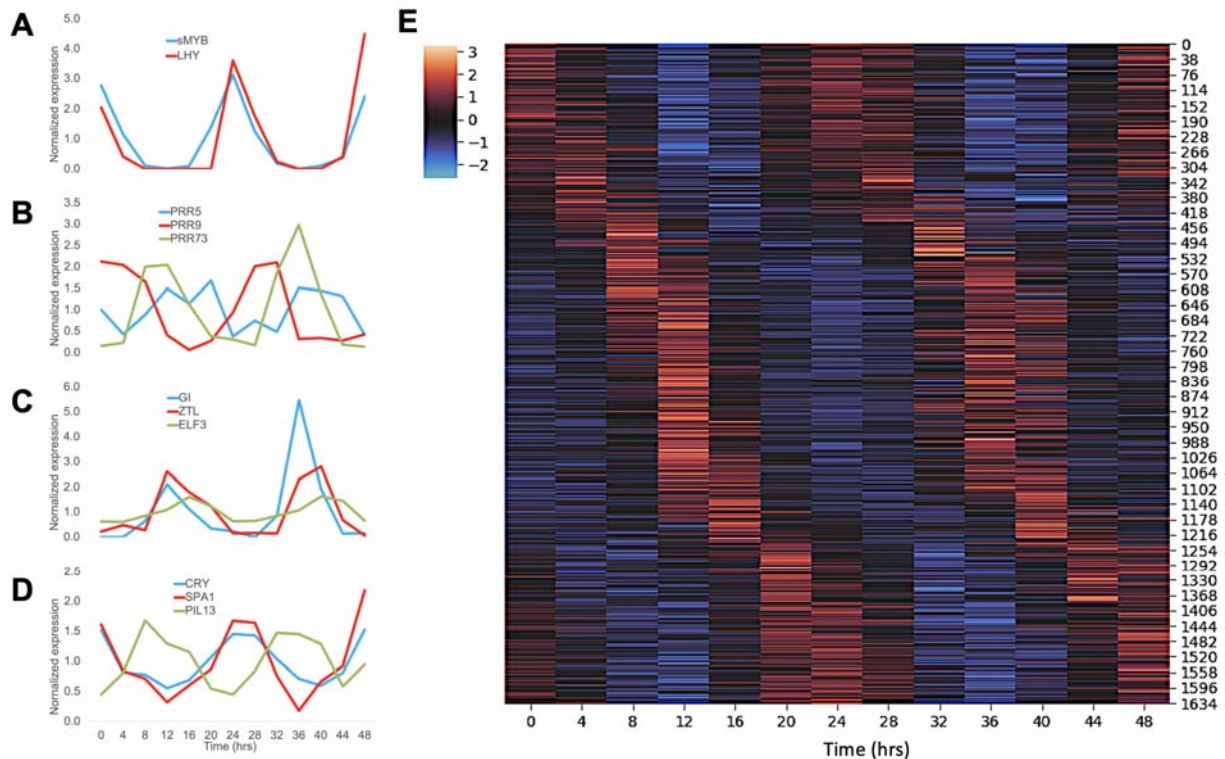


Figure 4. Core circadian and light genes cycle in *Wolffia* with global expression over the day. A-D) Circadian clock and light signaling TOD gene expression in wa8730 is similar to that found in other plants. **A)** sMYB (blue) and LHY (red) cycle with dawn specific expression. **B)** PRR5 (blue), PRR9 (red), and PRR73 (green) show waves of expression similar to other species. **C)** GI (blue), ZTL (red) and ELF3 (green) cycle with evening specific expression. **D)** Light signaling genes CRY (blue), SPA1 (red) and PIL3 (green) cycle over the day. **E)** Heatmap of the cycling genes with red indicating high expression and blue indicating low expression. Genes are on the y-axis and time is on the x-axis.

phase; one is the likely *CCA1/LHY* ortholog and the other is orthologous to *RVE7* (Figure 4A). The other half of the core circadian clock negative-feedback loop is the evening expressed *TOC1*, which is missing in *Wolffia*. *TOC1* is part of a five-member gene family

of PRRs, which display “waves” of expression across the day (*PRR9*, dawn; *PRR7*, midday; *PRR5*, dusk; *PRR3*, evening; *TOC1/PRR1*, evening) (Michael and McClung 2003). *Wolffia* only had three PRRs (plus 3 and 4 ARR for wa7733 and wa8730 respectively): *WaPRR9*, *WaPRR7/3* and *WaPRR5*, which are phased to dawn, dusk and evening respectively (Figure 4B). Other core circadian proteins such as *GIGANTEA* (*G1*), *EARLY FLOWERING 3* (*ELF3*) and *FLAVIN-BINDING, KELCH REPEAT, F-BOX* (*FKF1*) have evening expression as in other plants (Figure 4C), as well as circadian-regulated light signaling genes (Figure 4D). Despite having a reduced set of circadian and light signaling genes (Supplemental Table S11), TOD expression is conserved across the core circadian clock proteins.

As another check of the time course, we looked to see if TOD cis-elements that we have found to be conserved in other plant species are also found in *Wolffia* (Michael et al. 2008b). We searched promoters (500 bp upstream) of genes predicted to be

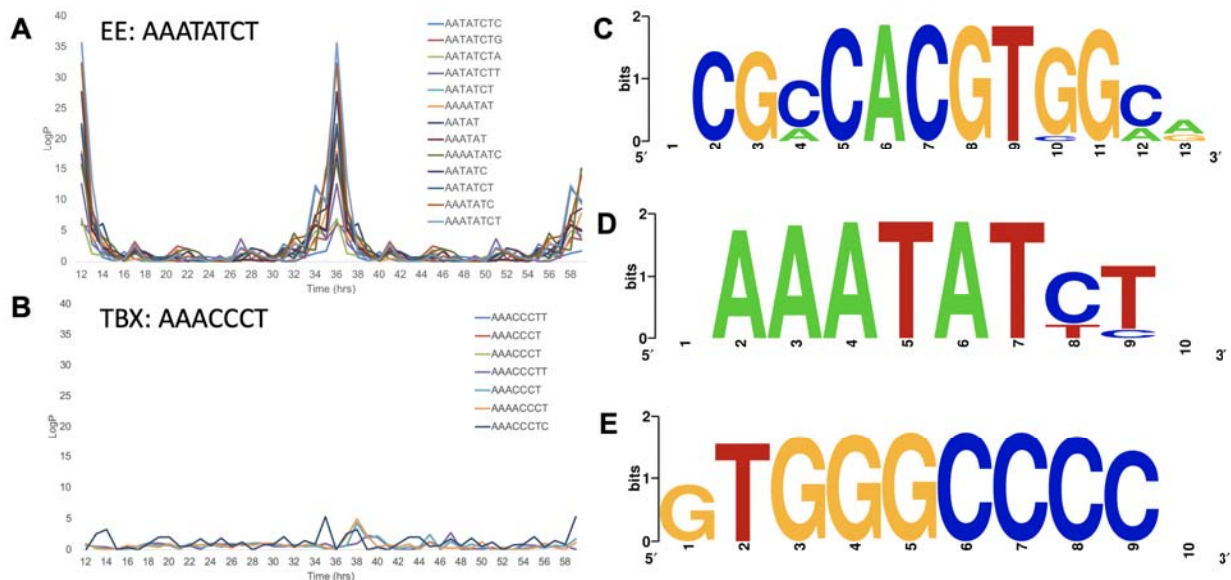


Figure 5. *Wolffia* has conserved TOD cis-elements but lacks others. **A)** The Evening Element (EE:AAATATCT) is overrepresented in genes with evening-specific expression. **B)** The Telobox (TBX:AAACCCT) is not significantly overrepresented in *Wolffia* promoters of cycling genes. **C)** Sequence logo of the significantly overrepresented Gbox (CACGTG) in *Wolffia*. **D)** Sequence logo of the EE overrepresented cis-elements. **E)** Sequence logo of the Protein box (PBX: TGGGCC) overrepresented cis-elements.

expressed at the same time of day, and found the same cis-elements that we have identified across all other plants tested to date (Michael et al. 2008b; Zdepski et al. 2008; Filichkin et al. 2011). For instance, the Evening Element (EE: AAATATCT), which was identified in early microarray experiments and promoter bashing (Harmer et al. 2000; Michael and McClung 2002), is highly overrepresented in the promoters of evening expressed genes (Figure 5A,D), while the Gbox (CACGTG) and its derivatives are highly overrepresented at dawn (Figure 5C), and the Protein Box (PBX: GTGGGCCCC) is overrepresented late in the night (Figure 5E, Supplemental Table S15,S16). In contrast, the TeloBox (TBX: AAACCCT), which is usually overrepresented in genes expressed around midnight, was not significant in *Wolffia* (Figure 5B). The lack of the TBX could partly explain the decreased number of cycling genes or it could mean that *Wolffia* genes with the TBX do not cycle like they do in *Arabidopsis*, rice and poplar (Filichkin et al. 2011). To test these options, we leveraged our empirically proven informatic method to assign cycling and phase information from *Arabidopsis* to its *Wolffia* reciprocal best blast (RBB) ortholog (Filichkin et al. 2011; Michael et al. 2008b; Zdepski et al. 2008). We found that the TBX was overrepresented in the promoters of *Wolffia* orthologs assigned the phase from *Arabidopsis*, consistent with *Wolffia* TBX-containing genes not cycling under the conditions tested (Supplemental Figure S14). In *Arabidopsis*, genes associated with protein synthesis and other activities that occur in the middle of the night contain the TBX cis-element, suggesting *Wolffia* may have lost the TOD coordination for these pathways.

***Wolffia* cycling genes are focused on core energy acquisition pathways**

Since *Wolffia* has fewer TOD controlled genes under LDHH, we wanted to know if this was due to fewer pathways cycling or just a result of *Wolffia* having fewer genes per family (orthogroups). If the later were true, then we would expect *Wolffia* to have a similar number of cycling OGs compared to other plants that have expanded gene families. Therefore, we compared the *Wolffia* cycling genes against two high quality time courses from *Arabidopsis* and rice generated under the same LDHH conditions (Michael et al. 2008b; Filichkin et al. 2011). Of the OGs that had at least one gene cycling in *Arabidopsis*, rice and *Wolffia*, we found that 49% (4,293/8,724), 74% (6,025/8,063) and 18% (1,442/7,844) were cycling respectively, which are similar numbers to overall cycling for each species, and consistent with larger gene families not playing a significant role in percent cycling in any of the species (Figure 6B; Supplemental Figure 15). 81% of cycling *Wolffia* OGs were shared with *Arabidopsis* and rice, suggesting the cycling pathways in *Wolffia* are a subset of those found in other plants. In addition, the OGs that did cycle in *Wolffia* were more likely to share the same mean phase of expression with *Arabidopsis* and rice, consistent with the conservation of TOD expression that has been found in other studies (Filichkin et al. 2011; Reyes et al. 2017) (Supplemental Figure S16).

We next looked to see which *Wolffia* pathways were cycling, and by proxy through comparison with data from *Arabidopsis* and rice for those pathways that are not cycling in *Wolffia*, by looking at the GO terms that were significantly overrepresented at specific times over the day. *Wolffia* had 92 significant GO terms that are TOD specific, whereas *Arabidopsis* and rice had 238 and 253 respectively (Figure 6A; Supplemental Figure S17; Supplemental Table S18-20). The 20 significant GO terms shared across all three

species were summarized into sulfate assimilation, response to abiotic stimulus, ribosomal small subunit biogenesis, photosynthesis, photosystem I, starch catabolism and chlorophyll biosynthesis (Figure 6C, E). The remaining 72 *Wolffia*-specific TOD GO terms also are focused on carbon metabolism and summarized into monocarboxylic acid catabolism, cellular response to sulfate starvation, fructose 1,6-bisphosphate metabolism, and UDP-galactose transmembrane transport (Figure 6F). Both the shared and species-specific significant TOD GO terms are focused on energy acquisition in *Wolffia*. In contrast, the 421 significant GO TOD terms in *Arabidopsis* or rice are summarized into 15 biological processes consistent with plants that have more complex structure (Figure 6D) (Michael et al. 2008b; Filichkin et al. 2011). In addition, several of the summarized GO terms that are missing in *Wolffia*, but found in *Arabidopsis* and rice, are related to protein biosynthesis pathways, which is consistent with the loss of the TBX cis-element in *Wolffia* (Figure 5B; Supplemental Figure S14). These results show that *Wolffia* has retained a core set of cycling genes focused on energy acquisition and utilization, while other pathways common in model plants have been released from TOD control. Since the primary role of the circadian clock is to gate processes to coincide with specific times of day (Michael et al. 2003), and the circadian clock is dispensable in plants, the reduced TOD control in *Wolffia* could reflect its less orchestrated continuous growth pattern.

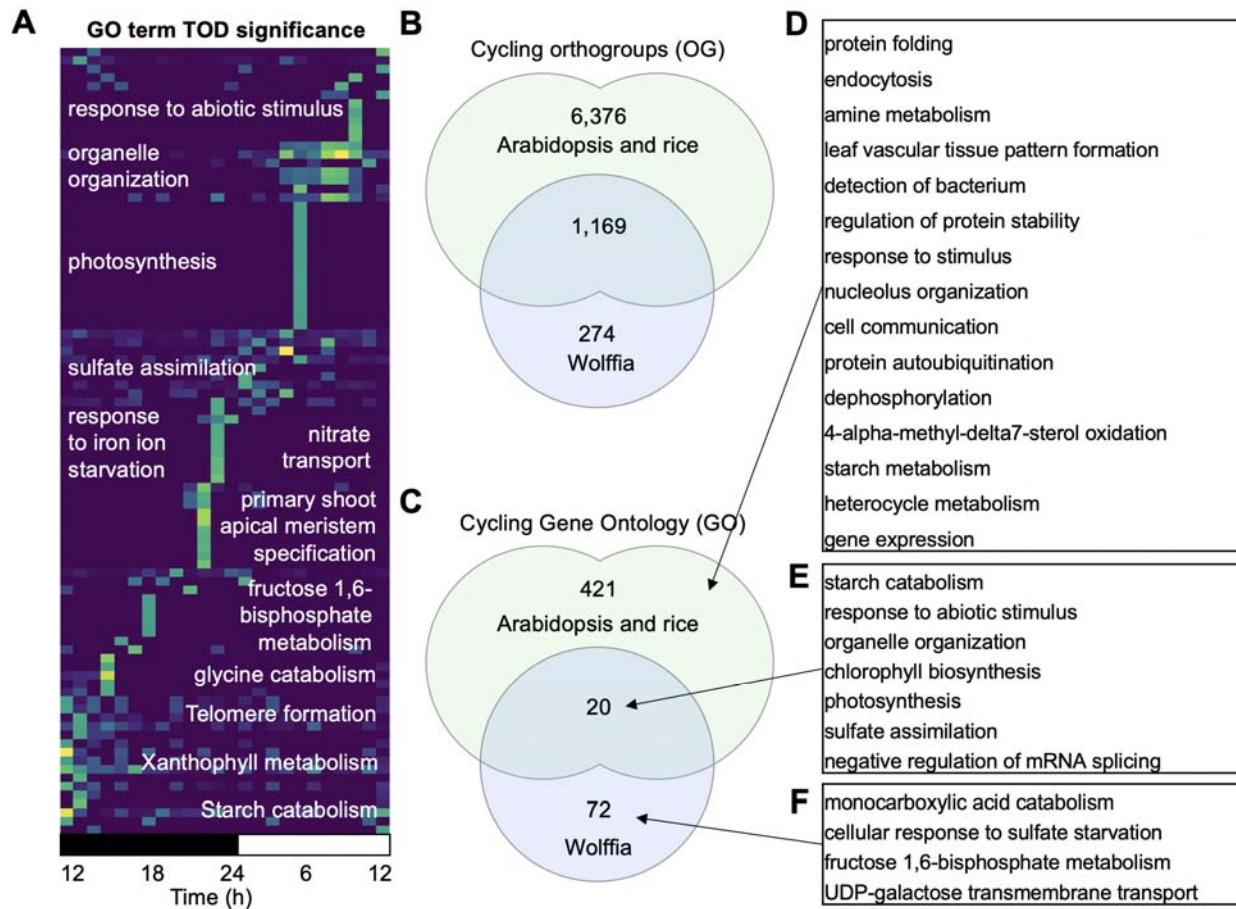


Figure 6. *Wolffia* cycling genes are focused on energy acquisition. A) Gene Ontology (GO) term overrepresentation by TOD for *Wolffia* with GO term summaries for 6 hr bins. B) Venn diagram of cycling orthogroups (OGs) with at least one gene from *Wolffia*, *Arabidopsis* and rice. C) Venn diagram of cycling GO terms from *Wolffia*, *Arabidopsis* and rice. D) GO term summary for the 421 *Arabidopsis* and rice cycling GO terms. E) GO term summary for the 20 cycling GO terms shared across all three species. F) GO term summary for the 72 *Wolffia*-specific cycling GO terms. All GO terms were summarized using semantic similarity (SimRel), and P-values from overrepresentation with REVIGO.

DISCUSSION

Here we present two draft genomes for different accessions of *W. australiana*, which has a relatively small genome and contains a minimal set of about 15,000 genes. A key finding of our work is that *Wolffia* has a reduced number of TOD regulated genes (13%) compared to other plants under the same assay conditions, and the genes that

remain regulated are specific to photosynthesis and carbon metabolism. Since *Wolffia* is small (around 1 mm in size), fast growing (DT ~1d), has a minimal set of core plant genes, and grows in direct contact with the media, it offers advantages analogous to the yeast system, which opens up new research opportunities in plants.

W. australiana has the smallest genome across the 11 species in the genus tested, which have an average size of 1,136 Mb, and is half the size of the next smallest species of *W. brasiliensis* at 776 Mb (Wang et al. 2011). Despite the range in genome sizes, all *Wolffia* species have a reduced body plan of a frond with just several thousand cells and no roots (Figure 1). They also have a similar fast growth rate of around a day (Supplemental Table S1) with no obvious relationship to genome size (Sree et al. 2015b). *Wolffia* contains the most derived of all of the duckweed species (Figure 1C), with body plan reductions compared to the most basal *Spirodela* genus. The results that several key genes associated with root and light-specific development were disrupted by LTR-RT in *W. australiana* versus *Spirodela* suggests that at some level the changes in morphology are the result of TE activity. Like *Spirodela*, which also has a small genome, the LTR-RT solo::intact ratio is high in *W. australiana* consistent with LTR-RT being purged through illegitimate recombination (Devos et al. 2002). Therefore, it is probable that *W. australiana* has a relatively small genome compared to other *Wolffia* species due to the active purging of LTR-RT that in the past have helped to shape its unique gene repertoire and body plan.

Wolffia, and the Lemnaceae family in general, represent extreme examples of plant morphology minimization, and these adaptations are reflected in their reduced yet representative gene sets. Multiple independently assembled genomes of both *Spirodela*

and *Wolffia* share a common set of missing BUSCO genes (Supplemental Figure S7), and many of these genes in *Wolffia* represent genes associated with its morphological innovations (Figure 3; Supplemental Table S6). Low BUSCO scores have also been noted in other plants with morphological innovations like the parasitic dodder plant (*Cuscuta*) and carnivorous bladderwort (*Utricularia*) that both also lack roots and leaf structures like *Wolffia* (Ibarra-Laclette et al. 2013; Sun et al. 2018; Vogel et al. 2018). *Cuscuta australis* shares many of the same gene losses as *Wolffia* such as *WOX5* (root apical stem cell maintenance), *LOP1* (leaf patterning and root development) and the entire *CASP* family (casparian strip) (Sun et al. 2018). *Wolffia* is also missing several families of the small signaling peptide *CLAVATA3/ESR-RELATED* (*CLE*) that regulate various aspects of cell fate and meristem size (Jun et al. 2010). Loss of *CLE3* peptide known to play key roles in orchestrating meristem size in plants is correlated with the unconventional mode of organogenesis in duckweed in which the stipe tissue functions as the site where new meristems are continuously initiated and develop sequentially (Figure 1C) (Lemon and Posluszny 2000). Despite the extensive gene loss, *Wolffia* still maintains a core set of gene families (orthogroups, OGs) common to flowering plants that have minimal expansions (Supplemental Table S7; Supplemental Figure S9).

Wolffia has lost a host of genes associated with the intersection between light signaling, phytohormones, circadian clock, growth, immunity and development that may provide insight into its floating ball-shaped morphology. It is critical for a multi-organ plant to position itself relative to the Earth (gravitropism) and the sun (phototropism) for proper development, as well as the means to communicate between the different parts of the organism through systemic signaling (Vandenbrink et al. 2014). *Wolffia* has lost many

genes of the *NPH3/RPT2-Like* family (*NRL*), which are required for several auxin-mediated growth processes, including phototropism (root and shoot), petiole positioning, leaf expansion, chloroplast accumulation, stomatal opening and circadian control of PSII photosynthetic efficiency (Christie et al. 2018). *NPH3* is the founding member of the *NRL* family with a close paralog *DOT3*, mutants of which fail to exhibit phototropism and have aberrant venation patterning in *Arabidopsis* respectively (Petricka et al. 2008; Motchoulski and Liscum 1999; Liscum and Briggs 1995). Compared to *Arabidopsis*, *Spirodela* has only one *NPH3/DOT3*, which is lost in *Wolffia* through an LTR-RT disruption (Figure 2D). In addition, *Wolffia* is missing the family of *LAZY* proteins that act as central integrators of gravity sensing with the formation of auxin gradients to control plant architecture (Yoshihara and Spalding 2017). The loss of these two key phototropic and gravitropic pathways provide clues as to how *Wolffia* has streamlined its body plan.

Lastly, the attrition of the highly conserved NLR family that are known to contain many R genes in plants required to mediate defense signaling and systemic resistance indicates that these genes are largely dispensable for a fast-growing and structurally simple plant. Elucidating how the stipe tissue continually functions to generate new meristematic centers that are orchestrated to produce new daughters and the mechanism that provides *Wolffia* with robust defense to pathogens in an R-gene independent manner will reveal much new plant biology. For the latter, the emphasis of the sphingolipid-related pathways observed in our work here may suggest their importance for defense signaling in *Wolffia* while there is also some evidence for an amplification of the antimicrobial peptide pathways in *Spirodela* (An et al. 2019). These suggestions remain to be examined in greater details and tested in the future.

One of the best described plant growth processes at the molecular level is that of hypocotyl elongation in *Arabidopsis* (Creux and Harmer 2019). The circadian clock restricts or “gates” the growth during the night hours (dark) through the core clock protein TOC1 binding to the PHYTOCHROME INTERACTING FACTORS (PIF) transcription factors and interaction with one of the core feedback loops mediated by the Evening Complex (EC: *ELF3/ELF4/LUX*) (Seluzicki et al. 2017). *Wolffia*, *Spirodela* and *Zostera* all are missing *TOC1* while having several other PRRs, suggesting that they may replace *TOC1* function in aquatic non-grass monocots. In addition, *Wolffia* is missing the orthologs for *PIF3/4* while having the other factors of the EC. Also, light regulated growth is controlled at some level by two distinct types of nuclear photobodies, one of which is defined by *TZP* (Huang et al. 2016). Loss of *TZP* in *Arabidopsis* results in smaller plants while overexpression results in large plants that don't stop growing, consistent with it being a central integrator of light (*PHYA*) signaling into plant growth (Loudet et al. 2008; Huang et al. 2016; Zhang et al. 2018). Overall, the loss of growth gating pathways in *Wolffia* is consistent with the significant decrease in circadian, light and flowering time genes (Supplemental Table S11), which is in contrast to some CAM (Crassulacean acid metabolism) and crop plants that display expansion of circadian genes (Wai et al. 2019; Lou et al. 2012). These results suggest that the genome innovations responsible for the change in body plan in *Wolffia* may be closely linked to the loss of specific light-gated growth.

The finding that *Wolffia* has fewer cycling genes under LDHH conditions presents an unexpected paradox when compared with other minimalist organisms that have been studied. *Ostreococcus tauri* is a single-cell alga that is the smallest eukaryote

(picoeukaryote) with a 13 Mb genome, but a functional circadian clock made up of the core negative feedback loop of *CCA1-PPR1* and almost all of the genes cycling in a TOD fashion (Monnier et al. 2010). Similarly, the model microalga *Chlamydomonas reinhardtii* has a core circadian clock and 80% of its genes cycle in a TOD fashion (Zones et al. 2015). In contrast, multi-cellular, multi-organ plants tested under the LDHH condition have a higher number of cycling genes, such as *Arabidopsis* (45%), rice (41%), poplar (30%), douglas-fir (29%) and *Brachypodium* (27%) (Michael et al. 2008b; Filichkin et al. 2011; MacKinnon et al. 2019; Cronn et al. 2017). Therefore, it would seem that a simplified (i.e. fewer celled) plant like *Wolffia* would have almost all of its transcriptome TOD regulated while multi-organ plants would have some processes that would not require TOD expression. The fact that only the core photosynthetic and carbon pathways remain under TOD control under LDHH suggests that most processes in *Wolffia* are uncoupled from the environment.

However, it has been noted that a reduced number of genes (~5.2%) cycle in Norway spruce seedlings (*Picea abies*), which could be species-specific or reflect the age of the plants sampled, since other gymnosperms, under normal spring-summer conditions, have a similar number of cycling genes as other model plants (Ferrari et al. 2019; Cronn et al. 2017). In gymnosperms the hypothesis is that in overwintering needles that are not growing, only core photosynthesis is required for coupling to the environment. In both cases, perhaps the common driving force to minimize TOD control of genes is to economize on energy expenditure to survive harsh winters for gymnosperms and rapid, continuous growth for *Wolffia*. This would be consistent with the retention of TOD control for genes involved in energy acquisition and storage for both

cases. In a similar way, it is formally possible that *Wolffia* just has fewer genes cycling under the condition tested (LDHH), although it has been shown that there is a high level of overlap of cycling genes across environmental conditions (Michael et al. 2008b; Filichkin et al. 2011).

Since *Wolffia* is in direct contact with the environment (water) where temperature and nutrients are most likely in a relatively constant state over the course of the day, it is possible that *Wolffia* has uncoupled these processes from TOD expression required in an environment on land. While most conserved TOD cis-elements are found (Figure 5) (Michael et al. 2008b), the loss of TOD overrepresentation of the TBX in *Wolffia*, but yet the identification of it through cycling-orthologs with *Arabidopsis* (Supplemental Figure S14), suggest that TOD regulation of the highly conserved TBX-controlled pathways (such as protein synthesis) have been lost in *Wolffia* (Filichkin et al. 2011). The absence of TBX regulation could reflect a general loss of key regulatory switches (fewer TFs) associated with the circadian clock and light signaling (Supplemental Table S11,16,17). Since many of the gene losses in *Wolffia* link development with light signaling, it is possible that the evolutionary path to a highly reduced plant with simple architecture and continuous growth also resulted in the loss of light-specific gated growth.

Wolffia is like the yeast of flowering plants with a core set of angiosperm genes, small size, rapid unrestricted growth, and growing in direct contact with its environment. Before *Arabidopsis*, duckweed was widely used as a model plant in Plant Biology (Lam et al. 2014). In fact, duckweeds were central in elucidating photoperiodic flowering (*Lemna perpusilla*) and auxin biosynthetic pathways by radioisotope labeling (*Lemna gibba*) (Hillman 1976; Rapparini et al. 1999). *Wolffia* still has the genes for flowering and

could be developed as a genetic system that is distinct from that of *Arabidopsis*, where crosses may be made by mixing flowering strains and collecting the seeds that sink to the bottom of the growth medium. The limited number of cells and cell types in *Wolffia* could provide a simplified model to dissect cell specific regulation and how plant cells directly respond to specific chemicals at the organismal level.

METHODS

Growth

W. australiana (Benth.) Hartog & Plas clones 7211 (Australia, victoria), 7540 (New Zealand), 7733 (Australia, South Australia), and 8730 (Australia, New South Wales) were maintained at the Rutgers Duckweed Stock Cooperative (<http://www.ruduckweed.org/>) or at the stock collection at the University of Jena, Germany. The specific growth conditions for growth assays, imaging and the TOD time course are detailed in the Supplemental Materials (Appenroth et al. 1996). Explanation of Relative Growth Rate (RGR), Doubling Time (DT) and Relative (weekly) Yield (RY) have been previously described (Ziegler et al. 2015). The stack of 311 *Wolffia* micro images was obtained using digital microscope Keyence VHX-5000 with 600x lens magnification (Keyence Deutschland GmbH, Neu-Isenburg) and ImageJ software (Schindelin et al. 2012).

Genome sequencing and assembly

HMW genomic DNA was isolated from young teff leaf tissue for both PacBio and Illumina sequencing. A modified nuclei preparation was used to extract HMW gDNA and residual

contaminants were removed using phenol chloroform purification (Lutz et al. 2011). Genome size was estimated by k-mer frequency with Jellyfish and GenomeScope (Vurture et al. 2017), as well as by flow cytometry as previously described (Hoang et al. 2019). PacBio data was error corrected and assembled using Canu (v1.5) (Koren et al. 2017), and assembly graphs were visualized after each iteration of Canu in Bandage (Wick et al. 2015). A consensus was first generated using the PacBio reads and three rounds of racon (v1.3.1) (Vaser et al. 2017). The raw PacBio contigs were polished to remove residual errors with Pilon (v1.22) (Walker et al. 2014) using Illumina paired-end sequence. Illumina reads were quality-trimmed using Trimmomatic followed by aligning to the assembly with bowtie 2 (v2.3.0) (Langmead and Salzberg 2013) under default parameters. To test completeness, the Waksman Student Scholars Program (WSSP) (<https://www.ncbi.nlm.nih.gov/nucore/?term=Waksman+Student+Scholars+program+wolffia+australiana>) cDNA sequences were mapped to the wa8730 and wa7733 genome assemblies using minimap2 (v2.17-r941) (Li 2018). Bionano optical maps were prepared as previously described (Kawakatsu et al. 2016) with minor modifications (Supplemental Materials). Assemblies (wa8730, wa7733, sp9509, sp7498) were benchmarked using the BUSCO (v3) liliopsida odb10 database (Simão et al. 2015). Identification of high copy number repeats such as rDNA and centromere arrays in the two *Wolffia* accessions were performed as previously described (Hoang et al. 2018).

Genome annotation

Custom repeat libraries were constructed for each species following the MAKER-P basic protocol (Campbell et al. 2014). RepeatModeler (v1.0.8) was run against the genome

assembly to produce an initial *de novo* library (Smit and Hubley 2008-2015). Sequences with blastx hits (e-value 1×10^{-10}) to a UniProt database of plant protein coding genes were removed along with 50 bp flanking sequences. The resulting custom library was used with RepeatMasker (v4.0.7) with default settings (Smit et al. 2015). Protein coding genes were annotated for all four duckweed genomes with the MAKER 3.01.02 pipeline (Holt and Yandell 2011). For protein coding gene predictions, a comprehensive transcriptome assembly for each species was developed using PASA (Haas et al. 2003). Illumina RNA-seq reads for both sp9509 and wa8730 were trimmed with skewer (v0.2.2) (Jiang et al. 2014), aligned to the genome assembly with HISAT2 (v2.1.0) (Kim et al. 2015), and assembled with Trinity (v2.6.6) (Grabherr et al. 2011; Haas et al. 2013). EST sequences from sp7498 (Wang et al. 2014) were downloaded from SRA (SRR497624) and assembled with Newbler (v3.0) (Margulies et al. 2005). All available assembled duckweed transcripts were passed to MAKER as evidence. Protein homology evidence consisted of all UniProtKB/Swiss-Prot (Schneider et al. 2009; UniProt Consortium 2019) plant proteins and the following proteomes: *Arabidopsis thaliana*, *Elaeis guineensis*, *Musa acuminata*, *Oryza sativa*, *Spirodela polyrhiza*, and *Zostera marina* (Goodstein et al. 2012; Singh et al. 2013). Three different approaches were used for *ab initio* gene prediction. RNA-seq alignments along with the soft-masked assembly were passed to BRAKER 2.1.0 to train species-specific parameters for AUGUSTUS (v3.3.3) and produce the first set of coding gene predictions (Stanke et al. 2008, 2006; Hoff et al. 2016, 2019). A second set of predictions was produced by generating a whole-genome multiple sequence alignment of all duckweed species with Cactus (Paten et al. 2011) and running AUGUSTUS in CGP mode (König et al. 2016). The final set of predictions

was produced by running AUGUSTUS within the MAKER pipeline, using the BRAKER-generated species parameters along with protein and assembled RNA-seq alignments passed as evidence. MAKER was run and allowed to select the gene models most concordant with the evidence among these three prediction datasets. MAKER-P (Campbell et al. 2014) standard gene builds were generated by running InterProScan (v5.30-69) (Jones et al. 2014) and retaining only those predictions with a Pfam domain or having evidence support (AED score < 1.0). Orthogroups and orthologues were identified across 29 proteomes from the Plaza 4.0 Monocots database (Van Bel et al. 2018) and the duckweed MAKER-P standard build proteomes with alternate transcripts removed. OrthoFinder (v2.2.7) (Emms and Kelly 2019, 2015) was run against all-vs-all proteome alignments computed with DIAMOND (Buchfink et al. 2015) using the standard workflow. To create a consistent set of GO term classifications for every protein in each species in the dataset the 28 Plaza monocot (v4) proteomes and the four duckweed proteomes were processed by eggNOG-mapper (Huerta-Cepas et al. 2017). The significant GO terms (FDR < 0.05) were summarized and visualized using REVIGO (Supek et al. 2011). The NLRs (Nucleotide-binding Leucine-rich Repeat proteins) were predicted using NLR-Annotator (Steuernagel et al. 2020) that scans genomic sequences for MEME based sequence motifs (Bailey et al. 2009). In addition the proteomes were queried for the presence of the NB-ARC domain (PF00931) (Sarris et al. 2016). Genomic and proteomic NLR predictions were combined to create a non-redundant list of putative NLRs, each putative NLR with an available gene model was then run through interpro-scan (Jones et al. 2014) for further domain prediction. The set of NLRs from

either accession and species were then aligned by MAFFT (Kato and Standley 2016) to identify orthologs.

Time of day (TOD) time course analysis

TOD time course data was analyzed with similar methods that have been described with some modifications described below (Michael et al. 2008b; Wai et al. 2019; MacKinnon et al. 2019; Filichkin et al. 2011). HISAT2 (Kim et al. 2015) was used to align RNA-seq reads to the wa8730 assembly. The resultant alignments were processed by Cuffquant and CuffNorm (Trapnell et al. 2012) to generate normalized expression counts for each gene for each time point. Genes with mean expression across the 13 timepoints below 1 FPKM were filtered prior to being processed by HAYSTACK (Michael et al. 2008b). As a check cycling genes were also predicted cycling genes using JTKcycle (Hughes et al. 2010). Once cycling genes in wa8730 were identified, we were able to find putative cis-acting elements associated with TOD expression. Promoters, defined as 500 bp upstream of genes, were extracted for each gene in wa8730 and processed by ELEMENT (Michael et al. 2008a, 2008b; Mockler et al. 2007). Our threshold for identifying a k-mer as being associated with cycling was an FDR less than 0.05 in at least one of the comparisons.

DATA ACCESS

The raw PacBio data, Illumina resequencing, and RNA-seq reads generated in this study were deposited to the NCBI BioProject database

(<https://www.ncbi.nlm.nih.gov/bioproject/>) under accession number PRJNA615235

(individual SRA accession numbers in Table S21). The genome assembly wa8730 and

wa7733 are available from CoGe (<https://genomeevolution.org/>) under genome id56605 and id56606 respectively.

ACKNOWLEDGEMENTS

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research program under Award Number DE-SC0018244. Duckweed research at the Lam Lab is also supported in part by a grant from Hatch project (#12116) from the New Jersey Agricultural Experiment Station at Rutgers University. RM and JRE are Investigators of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTION

TPM and EL conceived the study, oversaw the experiments and wrote the manuscript; DB and TCM assembled the genome; FJ, JPS and JRE generated Bionano optical maps and scaffolded the genome; KSS and KJA conducted growth experiments; KJA and JF performed genome size estimations with flow cytometry; SO and LB performed sectioning, microscopy and developed images and time-lapse video of *Wolffia*; PC and SG carried out tissue sampling and RNA isolation for the circadian transcriptome study; EE and RM predicted protein coding genes and protein family (orthogroup) analysis; EB and KK performed NB-ARC gene curation and analysis; TPM and NH conducted orthogroups, Gene Ontology terms and TOD analyses. All authors have read and approved the manuscript.

DISCLOSURE DECLARATION

No competing interests are declared by the authors.

REFERENCES

- An D, Zhou Y, Li C, Xiao Q, Wang T, Zhang Y, Wu Y, Li Y, Chao D-Y, Messing J, 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.
- Appenroth K-J, Sree KS, Bog M, Ecker J, Seeliger C, Böhm V, Lorkowski S, Sommer K, Vetter W, Tolzin-Banasch K, et al. 2018. Nutritional Value of the Duckweed Species of the Genus *Wolffia* (Lemnaceae) as Human Food. *Front Chem* **6**: 483.
- Appenroth K-J, Teller S, Horn M. 1996. Photophysiology of turion formation and germination in *Spirodela polyrhiza*. *Biol Plant* **38**: 95.
- Bernard FA, Bernard JM, Denny P. 1990. Flower Structure, Anatomy and Life History of *Wolffia australiana* (Benth.) den Hartog & van der Plas. *Bull Torrey Bot Club* **117**: 18–26.
- Bog M, Lautenschlager U, Landrock MF, Landolt E, Fuchs J, Sowjanya Sree K, Oberprieler C, Appenroth K-J. 2015. Genetic characterization and barcoding of taxa in the genera *Landoltia* and *Spirodela* (Lemnaceae) by three plastidic markers and amplified fragment length polymorphism (AFLP). *Hydrobiologia* **749**: 169–182. <http://dx.doi.org/10.1007/s10750-014-2163-3>.

- 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* **17**: 42–49.
- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**: 59–60.
- Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, Lei J, Achawanantakun R, Jiao D, Lawrence CJ, et al. 2014. MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol* **164**: 513–524.
- Christie JM, Suetsugu N, Sullivan S, Wada M. 2018. Shining Light on the Function of NPH3/RPT2-Like Proteins in Phototropin Signaling. *Plant Physiol* **176**: 1015–1024.
- Creux N, Harmer S. 2019. Circadian Rhythms in Plants. *Cold Spring Harb Perspect Biol* **11**. <http://dx.doi.org/10.1101/cshperspect.a034611>.
- Cronn R, Dolan PC, Jogdeo S, Wegrzyn JL, Neale DB, St Clair JB, Denver DR. 2017. Transcription through the eye of a needle: daily and annual cyclic gene expression variation in Douglas-fir needles. *BMC Genomics* **18**: 558.
- Devos KM, Brown JKM, Bennetzen JL. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res* **12**: 1075–1079.
- Dodd AN, Salathia N, Hall A, Kévei E, Tóth R, Nagy F, Hibberd JM, Millar AJ, Webb

- AAR. 2005. Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science* **309**: 630–633.
- Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**: 18.
- Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**: 238.
- Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**: 157.
- Ferrari C, Proost S, Janowski M, Becker J, Nikoloski Z, Bhattacharya D, Price D, Tohge T, Bar-Even A, Fernie A, et al. 2019. Kingdom-wide comparison reveals the evolution of diurnal gene expression in Archaeplastida. *Nat Commun* **10**: 737.
- Filichkin SA, Breton G, Priest HD, Dharmawardhana P, Jaiswal P, Fox SE, Michael TP, Chory J, Kay SA, Mockler TC. 2011. Global profiling of rice and poplar transcriptomes highlights key conserved circadian-controlled pathways and cis-regulatory modules. *PLoS One* **6**: e16907.
- Fung-Uceda J, Lee K, Seo PJ, Polyn S, De Veylder L, Mas P. 2018. The Circadian Clock Sets the Time of DNA Replication Licensing to Regulate Growth in Arabidopsis. *Dev Cell* **45**: 101–113.e4.
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W,

- Hellsten U, Putnam N, et al. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* **40**: D1178–86.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**: 644–652.
- Green RM, Tingay S, Wang Z-Y, Tobin EM. 2002. Circadian rhythms confer a higher level of fitness to *Arabidopsis* plants. *Plant Physiol* **129**: 576–584.
- Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, et al. 2003. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* **31**: 5654–5666.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**: 1494–1512.
- Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, Wang X, Kreps JA, Kay SA. 2000. Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* **290**: 2110–2113.
- Hillman WS. 1976. Calibrating duckweeds: light, clocks, metabolism, flowering. *Science* **193**: 453–458.
- Hoang PNT, Michael TP, Gilbert S, Chu P, Motley ST, Appenroth KJ, Schubert I, Lam E.

2018. Generating a high-confidence reference genome map of the Greater Duckweed by integration of cytogenomic, optical mapping, and Oxford Nanopore technologies. *Plant J* **96**: 670–684.
- Hoang PTN, Schubert V, Meister A, Fuchs J, Schubert I. 2019. Variation in genome size, cell and nucleus volume, chromosome number and rDNA loci among duckweeds. *Sci Rep* **9**: 3234.
- Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**: 767–769.
- Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-Genome Annotation with BRAKER. *Methods Mol Biol* **1962**: 65–95.
- Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**: 491.
- Hori K, Ogiso-Tanaka E, Matsubara K, Yamanouchi U, Ebana K, Yano M. 2013. Hd16, a gene for casein kinase I, is involved in the control of rice flowering time by modulating the day-length response. *Plant J* **76**: 36–46.
- Huang H, Alvarez S, Bindbeutel R, Shen Z, Naldrett MJ, Evans BS, Briggs SP, Hicks LM, Kay SA, Nusinow DA. 2016. Identification of Evening Complex Associated Proteins in Arabidopsis by Affinity Purification and Mass Spectrometry. *Mol Cell Proteomics* **15**: 201–217.

- Huby E, Napier JA, Baillieul F, Michaelson LV, Dhondt-Cordelier S. 2020. Sphingolipids: towards an integrated view of metabolism during the plant stress response. *New Phytol* **225**: 659–670.
- Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol Biol Evol* **34**: 2115–2122.
- Hughes ME, Hogenesch JB, Kornacker K. 2010. JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *J Biol Rhythms* **25**: 372–380.
- Ibarra-Laclette E, Lyons E, Hernández-Guzmán G, Pérez-Torres CA, Carretero-Paulet L, Chang T-H, Lan T, Welch AJ, Juárez MJA, Simpson J, et al. 2013. Architecture and evolution of a minute plant genome. *Nature* **498**: 94–98.
- Jiang H, Lei R, Ding S-W, Zhu S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**: 182.
- Johnson KL, Kibble NAJ, Bacic A, Schultz CJ. 2011. A fasciclin-like arabinogalactan-protein (FLA) mutant of *Arabidopsis thaliana*, *fla1*, shows defects in shoot regeneration. *PLoS One* **6**: e25154.
- Jones JDG, Dangl JL. 2006. The plant immune system. *Nature* **444**: 323–329.
<http://dx.doi.org/10.1038/nature05286>.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function

classification. *Bioinformatics* **30**: 1236–1240.

Jun J, Fiume E, Roeder AHK, Meng L, Sharma VK, Osmont KS, Baker C, Ha CM, Meyerowitz EM, Feldman LJ, et al. 2010. Comprehensive analysis of CLE polypeptide signaling gene expression and overexpression activity in *Arabidopsis*. *Plant Physiol* **154**: 1721–1736.

Katoh K, Standley DM. 2016. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* **32**: 1933–1942.

Kawakatsu T, Huang S-SC, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y, et al. 2016. Epigenomic Diversity in a Global Collection of *Arabidopsis thaliana* Accessions. *Cell* **166**: 492–505.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**: 357–360.

König S, Romoth LW, Gerischer L, Stanke M. 2016. Simultaneous gene finding in multiple genomes. *Bioinformatics* **32**: 3388–3395.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**: 722–736.

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**: 737–742.

- Landolt E. 1986. Biosystematic investigations in the family of duckweeds (Lemnaceae)(Vol. 2.) The family of Lemnaceae-a monographic study. vol. 1. *Veroff Geobot Inst ETH* **71**: 1–563.
- Langmead B, Salzberg SL. 2013. Langmead. 2013. Bowtie2. *Nat Methods* **9**: 357–359.
- Lemon GD, Posluszny U. 2000. Comparative Shoot Development and Evolution in the Lemnaceae. *Int J Plant Sci* **161**: 733–748.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.
- Liscum E, Briggs WR. 1995. Mutations in the NPH1 locus of *Arabidopsis* disrupt the perception of phototropic stimuli. *Plant Cell* **7**: 473–485.
- Loudet O, Michael TP, Burger BT, Le Metté C, Mockler TC, Weigel D, Chory J. 2008. A zinc knuckle protein that negatively controls morning-specific growth in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **105**: 17193–17198.
- Lou P, Wu J, Cheng F, Cressman LG, Wang X, McClung CR. 2012. Preferential retention of circadian clock genes during diploidization following whole genome triplication in *Brassica rapa*. *Plant Cell* **24**: 2415–2426.
- Lutz KA, Wang W, Zdepski A, Michael TP. 2011. Isolation and analysis of high quality nuclear DNA with reduced organellar DNA for plant genome sequencing and resequencing. *BMC Biotechnol* **11**: 54.

- MacKinnon KJ-M, Cole BJ, Yu C, Coomey JH, Hartwick NT, Remigereau M-S, Duffy T, Michael TP, Kay SA, Hazen SP. 2020. Changes in ambient temperature are the prevailing cue in determining *Brachypodium distachyon* diurnal gene regulation. *New Phytol.* <http://dx.doi.org/10.1111/nph.16507>.
- Ma L, Li G. 2018. FAR1-RELATED SEQUENCE (FRS) and FRS-RELATED FACTOR (FRF) Family Proteins in Arabidopsis Growth and Development. *Frontiers in Plant Science* **9**. <http://dx.doi.org/10.3389/fpls.2018.00692>.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376–380.
- Mathur S, Vyas S, Kapoor S, Tyagi AK. 2011. The Mediator Complex in Plants: Structure, Phylogeny, and Expression Profiling of Representative Genes in a Dicot (*Arabidopsis*) and a Monocot (*Rice*) during Reproduction and Abiotic Stress. *Plant Physiology* **157**: 1609–1627. <http://dx.doi.org/10.1104/pp.111.188300>.
- McClung CR. 2019. The Plant Circadian Oscillator. *Biology* **8**. <http://dx.doi.org/10.3390/biology8010014>.
- Michael TP, Breton G, Hazen SP, Priest H, Mockler TC, Kay SA, Chory J. 2008a. A morning-specific phytohormone gene expression program underlying rhythmic plant growth. *PLoS Biol* **6**: e225.
- Michael TP, Bryant D, Gutierrez R, Borisjuk N, Chu P, Zhang H, Xia J, Zhou J, Peng H, El Baidouri M, 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.

Michael TP, McClung CR. 2003. Enhancer trapping reveals widespread circadian clock transcriptional control in *Arabidopsis*. *Plant Physiol* **132**: 629–639.

Michael TP, McClung CR. 2002. Phase-specific circadian clock regulatory elements in *Arabidopsis*. *Plant Physiol* **130**: 627–638.

Michael TP, Mockler TC, Breton G, McEntee C, Byer A, Trout JD, Hazen SP, Shen R, Priest HD, Sullivan CM, et al. 2008b. Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS Genet* **4**: e14.

Michael TP, Salomé PA, Yu HJ, Spencer TR, Sharp EL, McPeck MA, Alonso JM, Ecker JR, McClung CR. 2003. Enhanced fitness conferred by naturally occurring variation in the circadian clock. *Science* **302**: 1049–1053.

Michael TP, VanBuren R. 2020. Building near-complete plant genomes. *Curr Opin Plant Biol* **54**: 26–33.

Mockler TC, Michael TP, Priest HD, Shen R, Sullivan CM, Givan SA, McEntee C, Kay SA, Chory J. 2007. The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis. *Cold Spring Harb Symp Quant Biol* **72**: 353–363.

Monnier A, Liverani S, Bouvet R, Jesson B, Smith JQ, Mosser J, Corellou F, Bouget F-Y. 2010. Orchestrated transcription of biological processes in the marine picoeukaryote *Ostreococcus* exposed to light/dark cycles. *BMC Genomics* **11**: 192.

Motchoulski A, Liscum E. 1999. Arabidopsis NPH3: A NPH1 photoreceptor-interacting protein essential for phototropism. *Science* **286**: 961–964.

Murakami-Kojima M, Nakamichi N, Yamashino T, Mizuno T. 2002. The APRR3 component of the clock-associated APRR1/TOC1 quintet is phosphorylated by a novel protein kinase belonging to the WNK family, the gene for which is also transcribed rhythmically in *Arabidopsis thaliana*. *Plant Cell Physiol* **43**: 675–683.

Murashige T, Skoog F. 1962. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum* **15**: 473–497.

<http://dx.doi.org/10.1111/j.1399-3054.1962.tb08052.x>.

One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**: 679–685.

Paten B, Earl D, Nguyen N, Diekhans M, Zerbino D, Haussler D. 2011. Cactus: Algorithms for genome multiple sequence alignment. *Genome Res* **21**: 1512–1528.

Petricka JJ, Clay NK, Nelson TM. 2008. Vein patterning screens and the defectively organized tributaries mutants in *Arabidopsis thaliana*. *Plant J* **56**: 251–263.

Pi L, Aichinger E, van der Graaff E, Llavata-Peris CI, Weijers D, Hennig L, Groot E, Laux T. 2015. Organizer-Derived WOX5 Signal Maintains Root Columella Stem Cells through Chromatin-Mediated Repression of CDF4 Expression. *Dev Cell* **33**: 576–588.

- Rapparini F, Cohen JD, Slovin JP. 1999. Indole-3-acetic acid biosynthesis in *Lemna gibba* studied using stable isotope labeled anthranilate and tryptophan. *Plant Growth Regul* 27: 139–144.
- Reyes P de L, de los Reyes P, Romero-Campero FJ, Teresa Ruiz M, Romero JM, Valverde F. 2017. Evolution of Daily Gene Co-expression Patterns from Algae to Plants. *Frontiers in Plant Science* 8. <http://dx.doi.org/10.3389/fpls.2017.01217>.
- Roppolo D, Boeckmann B, Pfister A, Boutet E, Rubio MC, Dénervaud-Tendon V, Vermeer JEM, Gheyselinck J, Xenarios I, Geldner N. 2014. Functional and Evolutionary Analysis of the CASPARIAN STRIP MEMBRANE DOMAIN PROTEIN Family. *Plant Physiol* 165: 1709–1722.
- Sarkar AK, Luijten M, Miyashima S, Lenhard M, Hashimoto T, Nakajima K, Scheres B, Heidstra R, Laux T. 2007. Conserved factors regulate signalling in *Arabidopsis thaliana* shoot and root stem cell organizers. *Nature* 446: 811–814.
- Sarris PF, Cevik V, Dagdas G, Jones JDG, Krasileva KV. 2016. Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biol* 14: 8.
- Schatlowski N, Stahl Y, Hohenstatt ML, Goodrich J, Schubert D. 2010. The CURLY LEAF interacting protein BLISTER controls expression of polycomb-group target genes and cellular differentiation of *Arabidopsis thaliana*. *Plant Cell* 22: 2291–2305.
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. 2012. Fiji: an open-source platform for

biological-image analysis. *Nat Methods* **9**: 676–682.

Schneider M, Lane L, Boutet E, Lieberherr D, Tognolli M, Bougueleret L, Bairoch A.

2009. The UniProtKB/Swiss-Prot knowledgebase and its Plant Proteome Annotation Program. *J Proteomics* **72**: 567–573.

Seluzicki A, Burko Y, Chory J. 2017. Dancing in the dark: darkness as a signal in plants.

Plant Cell Environ **40**: 2487–2501.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO:

assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.

Singh B, Sharma RA. 2015. Plant terpenes: defense responses, phylogenetic analysis,

regulation and clinical applications. *3 Biotech* **5**: 129–151.

Singh R, Ong-Abdullah M, Low E-TL, Manaf MAA, Rosli R, Nookiah R, Ooi LC-L, Ooi S-

E, Chan K-L, Halim MA, et al. 2013. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* **500**: 335–339.

Smit AFA, Hubley R. 2008-2015. *RepeatModeler Open-1.0*.

<http://www.repeatmasker.org>.

Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-4.0. 2013--2015.

Sree KS, Dahse H-M, Chandran JN, Schneider B, Jahreis G, Appenroth KJ. 2019.

Duckweed for Human Nutrition: No Cytotoxic and No Anti-Proliferative Effects on Human Cell Lines. *Plant Foods Hum Nutr* **74**: 223–224.

- Sree KS, Sowjanya Sree K, Maheshwari SC, Boka K, Khurana JP, Keresztes Á, Appenroth K-J. 2015a. The duckweed *Wolffia microscopica*: A unique aquatic monocot. *Flora - Morphology, Distribution, Functional Ecology of Plants* **210**: 31–39. <http://dx.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**: 204.
- Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**: 637–644.
- Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**: 62.
- Steuernagel B, Witek K, Krattinger SG, Ramirez-Gonzalez RH, Schoonbeek H-J, Yu G, Baggs E, Witek A, Yadav I, Krasileva KV, et al. 2020. The NLR-Annotator tool enables annotation of the intracellular immune receptor repertoire. *Plant Physiol.* <http://dx.doi.org/10.1104/pp.19.01273>.
- Sun G, Xu Y, Liu H, Sun T, Zhang J, Hettenhausen C, Shen G, Qi J, Qin Y, Li J, et al. 2018. Large-scale gene losses underlie the genome evolution of parasitic plant *Cuscuta australis*. *Nat Commun* **9**: 2683.
- Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes

long lists of gene ontology terms. *PLoS One* **6**: e21800.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**: 562–578.

UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**: D506–D515.

Van Bel M, Diels T, Vancaester E, Kreft L, Botzki A, Van de Peer Y, Coppens F, Vandepoele K. 2018. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res* **46**: D1190–D1196.

Vandenbrink JP, Kiss JZ, Herranz R, Medina FJ. 2014. Light and gravity signals synergize in modulating plant development. *Front Plant Sci* **5**: 563.

van Wersch S, Li X. 2019. Stronger When Together: Clustering of Plant NLR Disease resistance Genes. *Trends Plant Sci* **24**: 688–699.

Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* **27**: 737–746.

Vogel A, Schwacke R, Denton AK, Usadel B, Hollmann J, Fischer K, Bolger A, Schmidt MH-W, Bolger ME, Gundlach H, et al. 2018. Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris*. *Nat Commun* **9**: 2515.

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. 2017. GenomeScope: fast reference-free genome profiling from short reads.

Bioinformatics **33**: 2202–2204.

Wai CM, Weise SE, Ozersky P, Mockler TC, Michael TP, VanBuren R. 2019. Time of day and network reprogramming during drought induced CAM photosynthesis in *Sedum album*. *PLoS Genet* **15**: e1008209.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**: e112963.

Wang W, Haberer G, Gundlach H, Gläßer C, Nussbaumer T, Luo MC, Lomsadze A, Borodovsky M, Kerstetter RA, Shanklin J, et al. 2014. The *Spirodela polyrhiza* genome reveals insights into its neotenuous reduction fast growth and aquatic lifestyle. *Nat Commun* **5**: 3311.

Wang W, Kerstetter RA, Michael TP. 2011. Evolution of Genome Size in Duckweeds (Lemnaceae). *J Bot* **2011**. <https://www.hindawi.com/journals/jb/2011/570319/abs/> (Accessed April 11, 2019).

Weigel D, Glazebrook J. 2008. Fixation, embedding, and sectioning of plant tissues. *CSH Protoc* **2008**: db.prot4941.

Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**: 3350–3352.

Willemsen V, Bauch M, Bennett T, Campilho A, Wolkenfelt H, Xu J, Haseloff J, Scheres B. 2008. The NAC domain transcription factors FEZ and SOMBRERO control the

orientation of cell division plane in *Arabidopsis* root stem cells. *Dev Cell* **15**: 913–922.

Yoshihara T, Spalding EP. 2017. LAZY Genes Mediate the Effects of Gravity on Auxin Gradients and Plant Architecture. *Plant Physiol* **175**: 959–969.

Zdepski A, Wang W, Priest HD, Ali F, Alam M, Mockler TC, Michael TP. 2008. Conserved Daily Transcriptional Programs in *Carica papaya*. *Trop Plant Biol* **1**: 236–245.

Zhang S, Li C, Zhou Y, Wang X, Li H, Feng Z, Chen H, Qin G, Jin D, Terzaghi W, et al. 2018. TANDEM ZINC-FINGER/PLUS3 Is a Key Component of Phytochrome A Signaling. *Plant Cell* **30**: 835–852.

Zhao H, Wu D, Kong F, Lin K, Zhang H, Li G. 2016. The *Arabidopsis thaliana* Nuclear Factor Y Transcription Factors. *Front Plant Sci* **7**: 2045.

Ziegler P, Adelman K, Zimmer S, Schmidt C, Appenroth K-J. 2015. Relative in vitro growth rates of duckweeds (Lemnaceae) - the most rapidly growing higher plants. *Plant Biol* **17 Suppl 1**: 33–41.

Zones JM, Blaby IK, Merchant SS, Umen JG. 2015. High-Resolution Profiling of a Synchronized Diurnal Transcriptome from *Chlamydomonas reinhardtii* Reveals Continuous Cell and Metabolic Differentiation. *Plant Cell* **27**: 2743–2769.

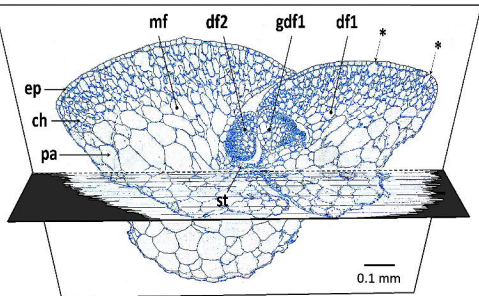
Table 1. Genome assembly statistics.

| | wa7733 | wa8730 | sp9509 | sp7498 |
|--|---------------|---------------|---------------|---------------|
| Genome size estimate flow cytometry (Mb) | 441 | 432 | 157 | 157 |
| Genome size estimate k-mer k=19 (Mb) | 343 | 342 | 164 | 185 |
| Genome final assembly (bp) | 393,842,592 | 354,558,446 | 138,592,155 | 145,274,398 |
| Contig (#) | 2,578 | 5,250 | 20 | 21 |
| Genome contig assembly size (bp) | 359,766,217 | 337,899,876 | NA | NA |
| Longest contig (bp) | 1,664,978 | 679,034 | NA | NA |
| N50 contig length (bp) | 256,298 | 102,418 | NA | NA |
| L50 contig (#) | 404 | 1,000 | NA | NA |
| Longest scaffold (bp) | 5,333,369 | 1,714,878 | 11,560,055 | 12,728,324 |
| N50 scaffold length (bp) | 836,551 | 109,493 | 7,949,387 | 8,107,549 |
| L50 scaffold (#) | 122 | 753 | 8 | 8 |
| BUSCO complete scaffolds (%) | 69 | 70 | 79 | 78 |
| Full length LTRs (#) | 2,510 | 1,892 | 801 | 567 |
| Masked (%) | 50 | 50 | 26 | 25 |
| Solo:intact LTR ratio | 11 | 14 | 8 | NA |
| Protein coding genes (#) | 15,312 | 14,324 | 17,510 | 17,057 |
| cDNA mapping (%) | 100 | 99.7 | NA | NA |
| Coverage PacBio (fold) | 91 | 45 | NA | NA |
| Coverage Illumina (fold) | 80 | 68 | NA | NA |
| Mapping PacBio (%) | 98 | 96 | NA | NA |
| Mapping Illumina (%) | 98 | 95 | NA | NA |

A



C



B

