Supplemental Material

Title

Genome and time-of-day transcriptome of *Wolffia australiana* link morphological extreme minimization with un-gated plant growth

Authors

Todd P. Michael[1$], Evan Ernst[2,12*], Nolan Hartwick[1*], Philomena Chu[3,&], Douglas Bryant[4,&], Sarah Gilbert[3&], Stefan Ortleb[5], Erin L. Baggs[6], K. Sowjanya Sree[7], Klaus J. Appenroth[8], Joerg Fuchs[5], Florian Jupe[1,&], Justin P. Sandoval[1], Ksenia V. Krasileva[6], Ljudmylla Borisjuk[5], Todd C. Mockler[4], Joseph R. Ecker[1,9], Robert A. Martienssen[2,10], and Eric Lam[3$]

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

Table of Contents

1. *Wolffia* biology
    1.1. The potential of *Wolffia* as a future high protein crop
    1.2. *Wolffia* growth rate measurements
    1.3. Ultrastructure of *W. australiana* and Imaging
    1.4. *Wolffia australiana* genome size estimates
    1.5. *Wolffia* growth for genome sequencing and time of day (TOD) RNA-seq
2. *Wolffia* Genomics
    2.1. Genome and RNA sequencing
    2.2. Genome assembly
    2.3. BioNano Genomics maps
    2.4. BUSCO (Benchmarking Universal Single-Copy Orthologs) analysis
    2.5. Ribosomal (rDNA) prediction and copy number
    2.6. Centromere and telomere identification
    2.7. Repeat annotation
    2.8. Solo:intact (S:I) ratio analysis
    2.9. Gene prediction and annotation
    2.10. Synteny and variation analysis
    2.11. Gene family analysis
    2.12. Circadian clock, flowering time and temperature gene orthogroups
3. *Wolffia* Time of day (TOD) analysis
    3.1. Time of day (TOD) expression analysis
    3.2. HAYSTACK versus JTKcycle cycling predictions
    3.3. TOD cis-element identification with ELEMENT
    3.4. TOD cis-element evolutionary analysis
    3.5. Transcription factors annotation
    3.6. Gene Ontology (GO) analysis of cycling genes
    3.7. Orthogroup cycling across *Wolffia*, *Arabidopsis* and rice

1 *Wolffia* biology

1.1 The potential of *Wolffia* as a future high protein crop

Wolffia is the most derived of all the duckweed with highly reduced morphology and anatomy, lacking roots and containing only the green floating frond, which is essentially a fused leaf and stem without any vasculature (Figure 1A, B). *Wolffia microscopica* is the fastest known growing angiosperm with a doubling time close to a day. *Wolffia globosa* and *Wolffia arrhiza* have been traditionally eaten in Southeast Asian countries like Thailand, Laos and Cambodia. It is referred to as Khai Nam, which means “water eggs.” Not only is the protein content high (~30% dry weight), the level for different amino acids is well above the recommendation of the World Health Organization (WHO) (Klaus-J Appenroth et al. 2018). Although the fatty acid content is low, its composition is suitable for the human diet. Interestingly, the mineral content can be modulated as per the desired content (Klaus-J Appenroth et al. 2018). Additionally, it was found that the plant extracts did not possess any cytotoxic or anti-proliferative effects on tested human cell lines (Sree et al. 2019). Taken together with the high growth rate, *Wolffia* has the added potential to be biofortified and developed as a more nutritious food crop for the future. Currently there are several companies bringing Wolffia to the market as a high protein crop.

1.2 *Wolffia* growth rate measurements

Four *W. australiana* (Benth.) Hartog & Plas accessions 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 (Supplemental Figure S1). All four *W. australiana* accessions were pre-cultivated for four weeks with a subculture interval of 1 week at 25°C and continuous white light of an intensity of 100 µmol m^{-2} s^{-1}. For growth rate measurement, 10 ± 1 fronds from the pre-cultures were inoculated into 100 ml nutrient medium N contained in 150 ml beakers (K-J Appenroth, Teller, and Horn 1996). The plants were grown for 1 week at 25°C and continuous white light of an intensity of 100 µmol m^{-2} s^{-1}. The final

frond count was taken on day 7. The experiment was carried out with 6 replicates and was repeated at an interval of 1 month (Supplemental Table S1). All values are average ±SE. Explanation of Relative Growth Rate (RGR), Doubling Time (DT) and Relative (weekly) Yield (RY) can be found in Ziegler et al., (Ziegler et al. 2015). For the two draft genomes reported in this study, we find that wa7733 and wa8730 have doubling times of 1.56 and 1.66 days respectively (Supplemental Table S1), consistent with the previous measurement of 1.39 days (Sree, Sudakaran, and Appenroth 2015).

1.3 Ultrastructure of *W. australiana* and Imaging

W. australiana was grown on 2% agar (Agar-Agar, danish, Carl Roth GmbH, Karlsruhe, Germany) with half-strength MS salts (Murashige and Skoog 1962) under sterile conditions with 16 h day (white light of 85 µmol $m^{-2}s^{-1}$) and 8 h night at 25 ± 1°C. Frond proliferation was monitored using a digital microscope Keyence VHX-5000 (Keyence Deutschland GmbH, Neu-Isenburg) by capturing images at 30 min intervals during 45 hours. In order to demonstrate growth dynamic, the video file was converted to time-lapse movie MP4 format (Supplemental Movie S1). For histological examinations, intact plants were fixed in FAA (Weigel and Glazebrook 2008), infiltrated with Spurr Resin (Plano, Wetzlar, Germany) and polymerized at 70°C for 24 h. The serial block sectioning was performed using Semimikrotom Leica Ultracut UTC (Leica Mikrosysteme GmbH, Vienna, Austria); the semi-thin 1 µm tissue sections were stained with Methylenblau-Borax (Waldeck GmbH & Co. KG, Muenster) and the stack of 311 micro images was obtained using digital microscope Keyence VHX-5000 with 600× lens magnification (Keyence Deutschland GmbH, Neu-Isenburg) and ImageJ software (Schindelin et al. 2012). The creation of three-dimensional models from a set of images (3D-reconstruction) was performed using Amira-Avizo software (ThermoFisherScientific) (Figure 1B).

1.4 *Wolffia australiana* genome size estimates

The genome size of *W. australiana* has been reported to be 375 ± 25 Mb and 357 ± 8 Mb for accessions wa8730 and wa7733 respectively by flow cytometry (Wenqin Wang, Kerstetter, and Michael 2011). More recently the genome size of wa7540 was

estimated by flow cytometry to be 432 ± 6 Mb (P. T. N. Hoang et al. 2019). These estimates were carried out with different internal standards, which could reflect the differences in size, or they could represent real variation between *W. australiana* accessions. We estimated the genome sizes of all three accessions by k-mer frequency using Illumina sequencing reads. Using an estimated 68, 80 and 62 fold Illumina sequencing coverage (with a 375 Mb genome size) for wa8730, wa7733 and wa7540 respectively, the k-mer (k=19) count was generated using jellyfish (v2.2.4) using no k-mer cut off (no -U), which enables counting of up to 10,000 k-mers to capture the high copy number repeats (centromere, rDNA and young TEs). The k-mer frequency was estimated by analyzing the Jellyfish histo file with GenomeScope (Vurture et al. 2017) and verified with an in-house pipeline (Supplemental Figure S2). The k-mer profiles are consistent with diploid homozygous plants with one predominant k-mer frequency peak and heterozygosity ranging from 0.29-0.60%. The k-mer frequency genome size estimates were very similar at 342-345 Mb with 31-33% repeat sequence (Supplemental Figure S2).

The k-mer genome size estimates results are consistent with the 357-375 Mb flow cytometry genome size estimate (Wenqin Wang, Kerstetter, and Michael 2011), but the consistency (within 3 Mb) in k-mer genome size estimates suggested the past flow cytometry estimates may not reflect the true genome size variation between accessions. Therefore, we measured all three accessions using the same method as Hoang et al., (P. T. N. Hoang et al. 2019), which uses radish (*Raphanus sativus*) as the reference, and found that the genome size estimates were also similar to one another but higher at 432, 441 and 432 Mb for wa7733, wa8730 and wa7540 respectively (Supplemental Figure S3). Since the genome assembly sizes were 395 and 355 Mb for wa7733 and wa8730 respectively, and both Illumina and PacBio reads covered greater than 95% of the genome assemblies, which confirms that 100 Mb is not missing from the assemblies, the flow cytometry results most likely are a high estimate of genome size. However, it does confirm that the genome size variation between the accessions is small (1-3 Mb). The differences in assembled genome size (395 vs. 355 Mb) could just reflect the amount of high copy number repeats (centromere and rDNA) (Supplemental Figure S2,3; Supplemental Table S4) assembled based on their coverage (91x vs. 41x),

which would be consistent with the k-mer genome size estimate of 342-345 Mb (Supplemental Figure S1).

1.5 *Wolffia australiana* growth for genome sequencing and time of day (TOD) RNA-seq

Fresh sterile clones for wa8730 and wa7733 from the Rutgers Duckweed Stock Cooperative (http://www.ruduckweed.org/) were inoculated into 100 ml half-strength MS salts (Murashige and Skoog 1962) in 150 ml beakers. Clones were grown under 12 h light (100 µmol m^{-2} s^{-1}) and 12 h dark at constant 20°C for three weeks to ensure enough biomass for genome sequencing and TOD RNA-seq. This growing condition is known as light/dark/hot/hot (LDHH) (Michael, Mockler, et al. 2008). For the 2-day time course experiment, roughly 600 mg of tissue was sampled every 4 hr over 2 days for a total of 13 time points for wa8730 and at 0 hr and 12 hr for a total of 2 time points for wa7733 (Supplemental Figure S10). Samples were collected in the dark under a dim green light, flash frozen in liquid nitrogen and stored at -80 ^{o}C. RNA was extracted using the total RNA isolation protocol from mirVana miRNA Isolation Kit (Thermo Scientific, Cat. No. AM1560). The RNA quality was validated using gel electrophoresis and the Qubit RNA IQ Assay (ThermoFisher, USA). High molecular weight (HMW) genomic DNA was isolated 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). The DNA sample was run overnight on a low concentration agarose gel (<0.5% W/V) to ensure the DNA has HMW (>40 kb) with Lambda DNA used as a control.

2. *Wolffia* genomics

2.1 Genome and RNA sequencing

PacBio (Pacific Biosciences, USA) libraries were constructed with HMW DNA using the manufacturer's protocol and were size selected for 30 kb fragments on the BluePippin system (Sage Science, USA) followed by subsequent purification using AMPure XP beads (Beckman Coulter, USA). The PacBio libraries were sequenced on a

PacBio RSII system with P6C4 chemistry. In total, 16 and 36 Gb of filtered PacBio reads were generated, which represents 41x and 91x coverage for wa8730 and wa7733 respectively. The same batch of HMW genomic DNA was used to construct Illumina (Illumina, USA) DNA-seq libraries for correcting residual errors in the PacBio assembly. Libraries were constructed using the KAPA HyperPrep Kit (Kapa Biosystems, Switzerland) followed by sequencing on an Illumina HiSeq2500 with 2x150 bp paired-ends (wa7733 and wa7540). Wa8730 was sequenced on an Illumina MiSeq with 2x311 bp paired-ends. Stranded RNA-seq libraries were constructed using 2ug of total RNA quantified using the Qubit RNA HS assay kit (Invitrogen, USA) with the Illumina TruSeq Stranded Total RNA with Ribo-Zero Plant (Illumina, USA). Multiplexed libraries were pooled and sequenced on an Illumina HiSeq2500 under paired-end 150 bp mode. All raw sequencing data has been submitted to NCBI (Supplemental Figure S21).

2.2 Genome assembly

PacBio data was error corrected and assembled using Canu (v1.5) (Koren et al. 2017), which has produced accurate and contiguous assembly for homozygous plant genomes (VanBuren et al. 2020). The following parameters were modified: minReadLength=2000, GenomeSize=375Mb, minOverlapLength=1000. Assembly graphs were visualized after each iteration of Canu in Bandage (Wick et al. 2015) to assess complexities related to repetitive elements and homoeologous regions. 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 greater than 30x coverage of Illumina paired-end data for wa8730 and wa7733 respectively. 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. Parameters for Pilon were modified as follows: --flank 7, --K 49, and --mindepth 15. Pilon was run recursively three times using the modified corrected assembly after each round. The statistics for the final Canu based PacBio assembly are summarized (Table 1). Completeness of the genome was checked by both looking at the mapping of the Illumina reads to the genome assemblies, as well as mapping high quality Sanger sequenced cDNA clones. The average Illumina read mapping to the genome assembly at each pilon step was 95

and 97% for wa8730 and wa7733 respectively (Table 1), consistent with the assemblies covering almost all of the genome space. In addition, 1,987 high quality Sanger sequenced cDNA clones from the Waksman Student Scholars Program (WSSP) (https://www.ncbi.nlm.nih.gov/nuccore/?term=Waksman+Student+Scholars+program+wolffia+australiana) were mapped to the genome assemblies with 100 and 99.7% of the cDNAs mapping to the wa7733 and wa8730 genomes, respectively (Table 1). The WSSP cDNA sequences were mapped to the wa8730 and wa7733 genome assemblies using minimap2 (v2.17-r941) with default settings (Li 2018).

2.3 BioNano Genomics maps

Bionano optical maps were prepared as previously described (Kawakatsu et al. 2016) with minor modifications; HMW DNA was extracted from up to 10 g whole plant tissue. Briefly, extracted HMW DNA was nicked with the enzyme Nt.BspQI (NEB, Beverly, MA, USA), fluorescently labeled, repaired, and stained overnight according to the Bionano Genomics nick-labeling protocol. Nick-labeled DNA was run on a single flow cell on the Irys platform (Bionano Genomics, San Diego, CA, USA). The IrysView software (Bionano Genomics; version 2.5.1) was used to quality filter the raw data (>100 kb length, >2.5 signal/noise ratio) and molecules were assembled into contigs using the “small optArguments” parameters. Resulting Bionano cmaps were compared against the different assemblies using Bionano RefAlign and collapsed regions or artificial expansions were detected as structural variations using the structomeIndel.py script (https://github.com/RyanONeil/structome) (Supplemental Table S4).

2.4 BUSCO (Benchmarking Universal Single-Copy Orthologs) analysis

BUSCO is typically used to assess the quality of a draft genome assembly by determining the number of highly conserved single-copy orthologs found in the assembly (Simão et al. 2015). We benchmarked each *Wolffia* draft assembly (wa8730 and wa7733), as well as the complete chromosome resolved *Spirodela* genomes (sp9509 and sp7498) using the BUSCO (v3) liliopsida odb10 database (Simão et al. 2015). This produced a list of BUSCO genes found to be complete, missing, duplicated or partially found in each genome and forms the basis of a contingency table

represented by an upset plot (Table 1; Supplemental Table S5; Supplemental Figure S7). Despite the high-quality draft genomes for *Wolffia* and *Spirodela*, the complete BUSCO scores were between 66-74%, which is low compared to other high quality plant genome assemblies. Therefore, we hypothesized that the BUSCO genes (liliopsida odb10) were not conserved in duckweed due to its reduced body plan and basal position in the plant kingdom. A chi-square test found that the missing genes in the duckweeds were not independent. Subsequent *post hoc* tests showed that genes found to be missing in any duckweed were at increased chance of being found missing in the other duckweeds. Further, genes found to be missing in either *Wolffia* were at increased chance of being found missing in the other. (Supplemental Figure S7). In order to get more information about the missing BUSCO genes, the representative ancestral proteins were BLAST against *Arabidopsis* protein models to identify the *Arabidopsis* locus and annotation (Supplemental Table S6).

2.5 Ribosomal (rDNA) prediction and copy number

We found that *Spirodela* has a surprisingly small number of ribosomal DNA (rDNA) arrays compared to other plants; sp9509 has ~80 arrays whereas *Arabidopsis* has ~400 with the same size genome (Michael et al. 2017). We hypothesized that *Spirodela* may not require as many rDNA arrays due to its rapid growth and short life cycle. Similarly, we would expect the *Wolffia* genome to have fewer rDNA arrays since it also grows rapidly and has a short life cycle. *Wolffia* rDNA was identified with the *Spirodela* 18S, 5.8S, 26S and 5S sequence (Michael et al. 2017) using BLAST against the genome wa8730 and wa7733 genome assemblies (Supplemental Table S4). The *Wolffia* consensus rDNA sequence was used to search the genomes to determine how many repeats were in each assembly, and consistent with the contiguity statistics for the two genomes wa7733 has about double the number of assembled rDNA as compared to wa8730 (Supplemental Table S4). To estimate the number of rDNA repeats in the two *Wolffia* accessions we performed a coverage analysis with the Illumina short reads (P. N. T. Hoang et al. 2018). We created a library with one full length rDNA array, one 5S array, and the single copy gene *GIGANTEA* (*GI*), and mapped the Illumina reads using minimap2 v2.17-r941) (Li 2018). The number of rDNA repeats was estimated by dividing the coverage across the rDNA arrays by the coverage over the single copy

gene (*GI*). We found that both *Wolffia* genomes had a reduced repeat content with ~200 copies, double that of *Spirodela* but still half of the number found in *Arabidopsis* (Supplemental Table S4).

2.6 Centromere and telomere identification

We have found that with PacBio long read-based assemblies another measure of completeness is the identification of high copy repeats such as centromeres and telomeres sequences (VanBuren et al. 2015). We employed a searching strategy to identify the centromeres that leverages the idea that the highest copy number tandem repeat (TR) will be the centromere in most genomes (Melters et al. 2013). We searched the genomes using tandem repeat finder (TRF; v4.09) using modified settings (1 1 2 80 5 200 2000 -d -h) (Benson 1999). TRF results were reformatted, summed and plotted to find the highest copy number TR per our previous methods (VanBuren et al. 2015). In contrast to *Spirodela* where there is either a dispersed (holocentric) centromere or no centromere array (Michael et al. 2017; Harkess et al. 2020), we identified that both *Wolffia* genomes contain the same three distinct putative centromeric repeats with base repeat units of 126, 167 and 250 bp (Supplemental Figure S2). All three putative centromere repeats had higher order repeats (HOR) consistent with them being part of centromere arrays that are currently active or active in the past. The fact that there are at least three putative centromere arrays could reflect a history in *Wolffia* of a changing centromere, or that *W. australiana* is a hybrid of several species; more *Wolffia* genomes should resolve this evolutionary history. Also, from the TRF output we identified telomeric arrays (AAACCCT). Like most plants both Wolffia genomes have the archetypal sequence, yet they are longer than other long read plant genomes. The *Arabidopsis* and *Spirodela* genome assemblies based on Oxford Nanopore Technologies (ONT) long reads have centromeres that are ~5 kb (Harkess et al. 2020; Michael et al. 2018; P. N. T. Hoang et al. 2018). In contrast, the *Wolffia* telomere sequences are as long as 20 kb, which considering its rapid growth is an interesting finding to be followed up.

2.7 Repeat annotation

Several different types of repeats were identified in the *Wolffia* genomes including long terminal repeat retrotransposons (LTR-RTs) and ribosomal DNA (rDNA) arrays. Custom repeat libraries were constructed for each species following the MAKER-P basic protocol (Campbell et al. 2014). RepeatModeler 1.0.8 was run against the genome assembly to produce an initial *de novo* library (Arian F. A. Smit and Hubley 2008-2015). Specifically, LTRharvest (GenomeTools v1.5.10) with the specific settings (-xdrop 37 -motif tgca -motifmis 1 -minlenltr 100 -maxlenltr 3000 -mintsd 2) were used to find full length LTR-RTs (Ellinghaus, Kurtz, and Willhoeft 2008). We identified 2,510 and 1,892 full length LTR-RTs in wa7733 and wa8730 respectively (Table 2). Called repeats with blastx hits (e-value 1 x 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 (A. F. A. Smit, Hubley, and Green 2015) to produce hard- and soft-masked versions of the genome assemblies.

2.8 Solo:intact (S:I) ratio analysis

When LTR-RTs proliferate via copy and paste, they either cause the genome to "bloat," which has been seen in most large plant genomes, or they are "purged" through mechanisms like illegitimate recombination (Michael 2014). Illegitimate LTR-RTs recombination results in solo LTRs, one terminal repeat without its pair or internal sequence (Ma, Devos, and Bennetzen 2004; Devos, Brown, and Bennetzen 2002). In general, genomes that are bloated have Solo:Intact (S:I) LTR ratios <1, while genomes that are purging LTR-RTs have S:I ratios >1. *Spirodela* has an S:I ratio of 8.52, the highest reported in plants to date, with some families of LTRs having more than 200 solo LTR per intact LTR-RT (Michael et al. 2017). As in *Spirodela*, we only called solo LTRs for the families that had at least one full length LTR-RT. We found that *Wolffia* has an even higher ratio of 11 to 14 in wa7733 and wa8730 respectively, consistent with *W. australiana* more actively purging its TEs leading to a much smaller sized genome compared to other *Wolffia* species that span an order of magnitude (Wenqin Wang, Kerstetter, and Michael 2011).

2.9 Gene prediction and annotation

Protein coding genes were annotated for all four duckweed genomes (sp7498, sp9509, wa7733 and wa8730) with the MAKER-P pipeline (v3.01.02) (Holt and Yandell 2011). For both sp9509 and wa8730, Illumina RNA-seq reads were trimmed with skewer (v0.2.2) (Jiang et al. 2014), aligned to the genome assembly with HISAT2 (v2.1.0) (Kim, Langmead, and Salzberg 2015), and assembled with Trinity (v2.6.6) (Grabherr et al. 2011; Haas et al. 2013) in genome-guided mode. Trinity was run independently in *de novo* mode, and all transcripts from the Trinity runs were aligned to the genome and reassembled with PASA (Haas et al. 2003) to form a comprehensive transcriptome assembly for each species. EST sequences from sp7498 (W. 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-P 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 (v2.1.0) 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-P pipeline, using the BREAKER-generated species parameters along with protein and assembled RNA-seq alignments passed as evidence. Finally, 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).

2.10 Synteny and variation analysis

The draft *Wolffia* genomes were compared at both the DNA and protein level to determine collinearity between wa8730 and wa7733, as well as with the chromosome resolved *Spirodela* genomes. First the genomes were compared at the DNA level by them against themselves or the *Spirodela* genomes using both minimap2 (minimap2 -x asm5) or mummer (nucmer -maxmatch -l 100 -c 500). The minimap2 alignments were compared using dotplotly (https://github.com/tpoorten/dotPlotly) to make dotplots, which was consistent with collinearity of the assemblies and the genera (data not shown). The DNA mummer alignments were analyzed for variation using Assemblytics (http://assemblytics.com/), which enabled the identification of insertions, deletions, repeat expansions/contractions and tandem expansions/contractions (Supplemental Figure S5). Protein alignments were generated with both SynMap on the CoGe platform (https://genomevolution.org/coge/) and MCscan (https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version)). Both wa8730 and wa7733 were highly collinear with the chromosome resolved sp9509 genome (Supplemental Figure S4). The MCscan jcvi.graphics.synteny module was used to visualize the syntenic blocks between the genomes (Figure 2; Supplemental Figure S6).

2 .11 Gene family analysis

Orthogroups (OGs) were identified across 29 proteomes from the PLAZA (v4) Monocots database (Van Bel et al. 2018) and the duckweed MAKER-P standard build proteomes with alternate transcripts removed. The 29 species included are: cre,*Chlamydomonas reinhardtii*; mco,*Micromonas commoda*; smo,*Selaginella moellendorffii*; atr,*Amborella trichopoda*; sly,*Solanum lycopersicum*; vvi,*Vitis vinifera*; ptr,*Populus trichocarpa*; ath,*Arabidopsis thaliana*; pab,*Picea abies*; mpo,*Marchantia polymorpha*; sp9509,*Spirodela polyrhiza*; sp7498,*Spirodela polyrhiza*; wa7733,*Wolffia australiana*; wa8730,*Wolffia australiana*; bdi,*Brachypodium distachyon*; hvu,*Hordeum vulgare*; obr,*Oryza brachyantha*; osaj,*Oryza sativa japonica*; osai,*Oryza sativa indica*; oth,*Oropetium thomaeum*; sbi,*Sorghum bicolor*; sit,*Setaria italica*; tae,*Triticum aestivum*; zma,*Zea mays*; zjn,*Zoysia japonica* ssp. nagirizaki; aco,*Ananas comosus*; egu,*Elaeis guineensis*; mac,*Musa acuminata*; ped,*Phyllostachys edulis*; peq,*Phalaenopsis equestris*; zoma,*Zostera marina*; ppa,*Physcomitrella patens*. OrthoFinder (v2.2.7)

(Emms and Kelly 2019, 2015) was run against all-vs-all proteome alignments computed with DIAMOND blastp using the standard workflow (Buchfink, Xie, and Huson 2015). The OG summary table for the 32 species is presented (Supplemental Table 7).

Leveraging the OGs, we clustered the proteomes using multidimensional scaling (MDS) and found that the Duckweed genomes clustered with the non-grass monocots (Supplemental Figure S8). 33-41% of *Wolffia* wa7733 and wa8730) and *Spirodela* (sp9509 and sp7498) genes are found in OGs with only one gene, whereas *Arabidopsis*, rice, brachy (*Brachypodium distachyon*) and maize have between 12-18% and over 20% in OGs with more than 10 genes. In contrast, less than 10% of duckweed genes are found in OGs with more than 10 genes (Supplemental Figure S9). We also identified OGs that are specific to *Wolffia* (must have wa8730 and wa7733) and no other species in that OG, as well as duckweed specific, and missing in *Wolffia* (Figure 3). Then we associated these species specific OGs with their Gene Ontology (GO) terms using EggNOG-mapper (Huerta-Cepas et al. 2017) (Supplemental Table S8). The GO categories unique to *Wolffia* (versus all species) revealed sphingolipid biosynthesis, photomorphogenesis, wax metabolism and cofactor catabolism (Figure 3; Supplemental Table S9). The GO categories missing in Wolffia (versus all species) revealed immobilization stress, lithium-ion transport, fatty alcohol metabolism, thermal perception, cell wall thickening, and phloem development (Figure 3; Supplemental Table S9). Since *Arabidopsis* has the highest quality annotation and number of functionally defined genes, we looked at the *Arabidopsis* genes in the OG missing in *Wolffia* (Supplemental Table S10).

2.12 Circadian clock, flowering time and temperature gene orthogroups

We constructed a list of 55 *Arabidopsis* genes with functional roles in the circadian clock, flowering time and temperature responses (Lou et al. 2012). We searched OGs for these genes and then summarized the number of genes per OG, manually checking when numbers didn't match expectations or were different between accessions of the same species (Supplemental Table S11). Overall *Wolffia* and *Spirodela* have fewer clock, flowering temperature genes consistent with fewer total proteins. Specifically, the number of core circadian clock genes is about a third of that

found in *Arabidopsis*, with *TOC1/PRR1*, *SIC/WARP* and *TZP* missing. In contrast, flowering time genes such as *FT* and *CO/COL* have similar numbers to other monocots, consistent with the flowering pathway being conserved in *Wolffia*. *Wolffia* has the same number of *FT* paralogs as rice, and double that of *Spirodela*. While flowering has not been recorded for *Spirodela*, *Wolffia* readily flowers, suggesting that the flowering time pathway is intact in *Wolffia*.

Wolffia is continuously in the water, and therefore could experience fewer daily environmental changes associated with temperature. For instance, one of the GO overrepresented terms for genes missing in *Wolffia* was thermal adaptation. Therefore, we looked at several classes of proteins thought to play a role in temperature regulation such as the *C-repeat/DRE-Binding Factor* (*CBF*) and *Heat Shock Proteins* (*HSP*). CBFs have been shown to play a cold sensing and *PHYB*-dependent temperature-regulated growth (Dong et al. 2020; Kidokoro et al. 2017), while *HSPs* play a role in thermotolerance during heat stress (Sharma et al. 2019). Both *CBFs* and *HSPs* have similar numbers of genes as compared to other monocots consistent with the conservation of these pathways in *Wolffia*.

3 *Wolffia* Time of day (TOD) analysis

3.1 Time of day (TOD) expression analysis

TOD time course data was analyzed with similar methods that have been described with some modifications (Michael, Mockler, et al. 2008; Wai et al. 2019; MacKinnon et al. 2019; Filichkin et al. 2011). HISAT2 (v2.1.0) (Kim, Langmead, and Salzberg 2015) was used to align RNA-seq reads to the wa8730 assembly. The resultant alignments were processed by Cuffquant and CuffNorm (v2.2.1) (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 as “not expressed” prior to being processed by HAYSTACK (Michael, Mockler, et al. 2008) to identify genes that show cycling expression behavior (Supplemental Figure S11; Supplemental Table S12). HAYSTACK operates by correlating the observed expression levels of each gene with a variety of user specified models that represent archetypal cycling behavior. We used a model file containing sinusoid, spiking traces, and various

rough linear interpolations of sinusoids with periods ranging from 20 hours to 28 hours in one-hour increments and phases ranging from 0-23 hours in one-hour increments. Genes that correlated with their best fit model (R > 0.8) were classified as cyclers with phase and period defined by the best fit model (Figure 4; Supplemental Figure S13). The threshold for calling cycling genes is based on previous validated observations (Michael, Mockler, et al. 2008; Wai et al. 2019; MacKinnon et al. 2019; Filichkin et al. 2011), although in some cases it is too stringent as evidenced by the exclusion of one of the core circadian clock genes *LATE ELONGATED HYPOCOTYL* (*LHY*) that clearly cycles (Figure 4A) but has an R (0.78) just below the threshold.

3.2 HAYSTACK versus JTKcycle cycling predictions

There are several methods for identification of cycling genes, and we have leveraged HAYSTACK for the majority of our studies due to its ability to identify cycling genes with non-sinusoidal patterns (Michael, Mockler, et al. 2008; Wai et al. 2019; MacKinnon et al. 2019; Filichkin et al. 2011). However, we usually check the HAYSTACK results against another popular method for predicting cycling genes, JTKcycle (Hughes, Hogenesch, and Kornacker 2010). Allowing periods 20, 24 and 28 hrs, and using a significance cutoff of an adjusted P-value (<0.05) we found 11.5% (1,370/11,870) of expressed genes were predicted to cycle (Supplemental Table 14). There was significant overlap of genes (964) that were predicted as cycling between JTKcycle and HAYSTACK, yet some differences (Supplemental Figure 12). For instance, *LHY* was predicted as cycling by JTKcycle (adjusted P < 0.01) with a period of 28 h. Therefore, the difference may reflect period estimations used by the two programs. We also restricted JTKcycle to 24 h and 1,124 genes were predicted to cycle (P<0.05); *LHY* was not predicted to cycle using these settings. Since the HAYSTACK cycling classifications were robust in our previous studies, they formed the basis of our promoter analysis of wa8730 and our GO term analysis of wa8730.

3.3 TOD cis-element identification with ELEMENT

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, Breton, et al. 2008; Michael, Mockler, et al. 2008; Mockler et al. 2007). ELEMENT generates an exhaustive background model of all 3-7 k-mer using all of the promoters in the genome, and then identifies overrepresented k-mers (3-7 bp) from the promoters for a specified gene list. Promoters for cycling genes were split according to their predicted peak time of expression into “phase” gene lists and k-mers that were overrepresented in any of these 24 promoter sets were identified by ELEMENT. By splitting up cycling genes according to their associated phase, we gained the power to identify k-mers associated with TOD-specific cycling behavior at every hour over the day (Supplemental Table S15). Our threshold for identifying a k-mer as being associated with cycling was an FDR < 0.05 in at least one of the comparisons. The significant k-mers were clustered according to sequence similarity identifying the Morning Element (ME), Gbox, and Evening Element (EE) motifs (Figure 5; Supplemental Table S16). Interestingly, the TeloBox (TBX) motif was not identified as significant (Figure 5). We also asked how often these four cis-elements are found in the promoters of genes that are not called TOD expressed. These four conserved TOD cis-elements are found in non-TOD regulated genes at a similar rate as the background model (5-10%), whereas they are found in cycling genes >15% of the time and >20% at specific times over the day. Since a limited number of non-TOD expressed genes contain these cis-elements in their promoters, this could mean that up to 5-10% of genes could cycle under different conditions then tested here. In fact, we have shown in *Arabidopsis*, rice, and poplar that despite having significant overlap between different TOD conditions (LDHH vs. LDHC vs. LLHC), there can be a 10-20% difference between conditions (Michael, Mockler, et al. 2008; Filichkin et al. 2011).

In order to identify the potential transcription factors (TFs) that target the identified cis-elements, we added an additional tool to the ELEMENT pipeline. Recently, the *in vitro* genomic binding sites for many TFs (~500) have been elucidated in *Arabidopsis* using the DNA affinity purification sequencing (DAP-seq) assay (O’Malley et al. 2016). Leveraging the position weight matrices (PWM) of TF binding sites from DAP-seq, we searched for the TFs that target our TOD cis-elements using TOMTOM (Gupta et al. 2007). This resulted in 168 matches to TFs that potentially bind our TOD

cis-elements, representing potential cycling regulators (Supplemental Table 16). Of course, these are *Arabidopsis* binding sites and TFs, so they only serve as a guide to look at the orthologs in *Wolffia*.

3.4 TOD cis-element evolutionary analysis

Due to the conserved nature of TOD expression networks across plants, the orthologues across species have similar expression patterns and underlying cis-elemnts in their promoters. We have leveraged the conserved nature of TOD expression networks to predict cycling behavior and cis-elements across phylogenetically diverse plants (Michael, Mockler, et al. 2008; Zdepski et al. 2008), and validated these predictions with empirical TOD expression data (Filichkin et al. 2011). These predictive tools also enable us to evaluate why genes are not cycling or cis-elements are not significant, such as the case in *Wolffia* with the conserved TBX cis-element. Therefore, we leveraged a high-quality *Arabidopsis* LDHH dataset (Michael, Mockler, et al. 2008) to assign the phase of expression to reciprocal best blast (RBB) matches of wa8730 and the green alga *Micromonas commoda* (Mco). We used the green alga to generate as much diversity in this comparison as possible and also since it is a single-celled organism we reasoned it might have similar cycling pathways as *Wolffia*. The phase of TOD gene expression from *Arabidopsis* was assigned to the RBB gene matches of both wa8730 and Mco, and the wa8730 and Mco promoters were analyzed using ELEMENT as above. The significant k-mers were evaluated for known elements and TBX was significant across all three species as well as other known elements such as the EE and GATA (Supplemental Figure 14). These results demonstrate that TOD expression networks are conserved as far back as single cell algae. However, the fact that the conserved TBX is not overrepresented in the *Wolffia* cycling dataset, yet overrepresented when the RBB genes are tested with *Arabidopsis* phases, suggests that the control of the TBX-related pathways are no longer TOD regulated in *Wolffia*. In *Arabidopsis* and rice, the TBX controls pathways related to protein regulation, which could mean that Wolffia has released protein-related activities from TOD control. Note, from this analysis we do not know whether the TBX-related pathways are also regulated in the green algae lineage; this would need to be tested directly. What we do know is

that like all plants tested to date, the TBX is overrepresented in RBB matches of cycling genes.

3.5 Transcription factors annotation

Transcription factors (TFs) are the key regulators of gene expression, often the regulatory hubs associated with key pathways by binding cis-elements. We predicted TFs and organized them by family for the *Wolffia* genomes, as well as the *Spirodela* (sp9509), *Arabidopsis* and rice genomes, using PlantTFcat (http://plantgrn.noble.org/PlantTFcat/) (Dai et al. 2013). In 103 different TF families, *Arabidopsis* and rice had two times the number of TFs (~3,500) compared to *Wolffia* and *Spirodela* (~1,500), which is consistent with the higher gene counts in *Arabidopsis* and rice; they had a similar percentage of TFs per genome (10-13%) (Supplemental Table S17). 53 of the TF families represented a higher percentage in *Wolffia* (wa8730) versus either *Arabidopsis* and rice; WD40 TFs were almost double the overall percentage in duckweeds compared to rice and *Arabidopsis*. Next, we looked at what percentage of the predicted TFs cycle to see if we can better understand the number of cycling genes. 172 of the wa8730 TFs are predicted to cycle, which is 1.3% of the expressed genes and 11.6% of the TFs. Overall, the percentage of cycling TFs for *Wolffia*, *Arabidopsis* and rice reflected the global cycling percentages at 11.6%, 26.7% and 68.6% respectively under the LDHH condition. Of the cycling *Wolffia* TFs, there were 47 families represented with WD40, C2H2, AP2-EREBP and MYB having 23, 16, 12 and 12 members respectively. The TFs we identified using our identified TOD cis-elements with the DAP-seq database resulted in 168 potential TOD TFs (Supplemental Table S16), which is close to the number of TFs we found to cycle. The DAP-seq TFs spanned 30 families with AP2-EREBP, bZIP and bHLH having 68, 21 and 11 respectively.

3.6 Gene Ontology (GO) analysis of cycling genes

We used GO annotations to identify cellular functions associated with cycling genes. EggNOG-mapper (Huerta-Cepas et al. 2017) was used to provide annotations for each gene establishing a mapping between GO terms and sets of genes associated

with those terms. In total, 6,874 genes in wa8730 could be associated with one or more GO terms from a population of 6,945 GO terms. Cycling calls from HAYSTACK were used to separate genes into 25 categories, one for each phase and non-rhythmic. The elim (eliminating genes) method was used to identify GO terms that were overrepresented among cycling categories. This allows us to associate GO terms with specific cycling behavior (Figure 6; Supplemental Figure S17; Supplemental Table S18-20).

3.7 Orthogroup cycling across *Wolffia*, *Arabidopsis* and rice

The fact that *W. australiana* has fewer genes compared to other species, in addition to smaller gene families, led us to wonder if the reduced number of cycling genes could just reflect fewer paralogs. For instance, maybe *Arabidopsis* and rice have more cycling genes because they have more paralogs that cycle; in other words, the fact that they have larger gene families leads to more cycling genes. We reasoned that if the reduced number of cycling genes in *Wolffia* was due to its smaller gene families (paralogs), then the number of cycling orthogroups should be similar to other species. We chose to look at *Arabidopsis* and rice since they both have high quality genomes and parallel cycling datasets in the same LDHH condition, and larger gene families than *Wolffia* (Supplemental Figure S9, 15) (Filichkin et al. 2011; Michael, Mockler, et al. 2008).

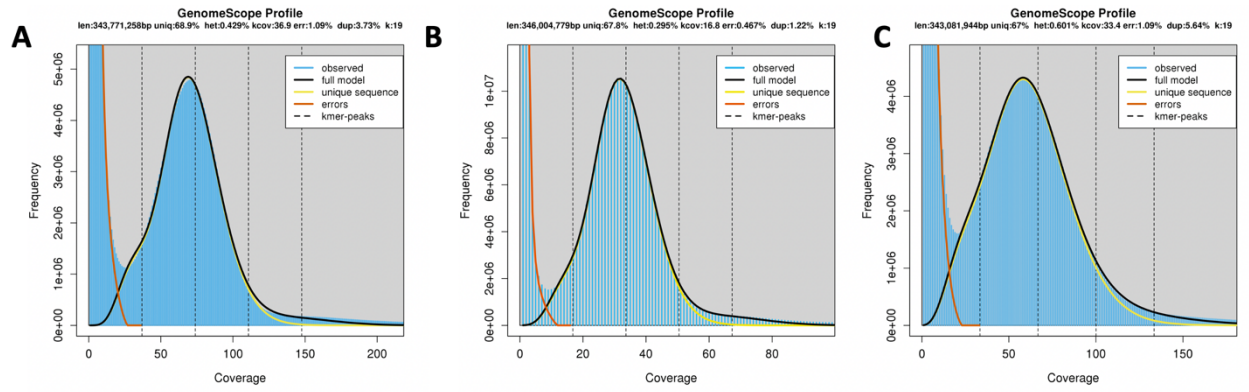
First, we confirmed that the *Wolffia* time course and resulting cycling predictions were similar to *Arabidopsis* and rice. The *Wolffia*, *Arabidopsis* and rice time courses had the expected morning and evening abundance of predicted cycling genes (Supplemental Figure S13). For each orthogroup (OG), for each species, we computed a circular mean of the phases of the cycling genes belonging to that OG in that species. This allows us to use the OGs to map from some cycling gene in wa8730 to a 'cycling behavior' in orthologous genes in *Arabidopsis* and rice. We used this mapping to create a KDE histogram plot representing the density of these pairs: phase of cycling gene, orthologous 'cycling behavior.' This plot represents how cycling behavior in wa8730 genes related to cycling behavior in *Arabidopsis* and rice orthologues. For example,

genes that cycle and peak at dusk (11-14h) in wa8730 are associated more with dusk genes in *Arabidopsis* and rice than with morning genes (Supplemental Figure S16).

Next, we asked if *Wolffia* had more or less cycling OGs to determine if the reduced number of cycling genes was a result of gene family size. There are 11,506 OGs that have at least one expressed gene for *Arabidopsis*, rice or *Wolffia* with 8724, 8063 and 7844 for each respectively (Supplemental Figure S15A-C). Taking the orthogroups that are shared across *Arabidopsis*, rice and *Wolffia* (must have at least one gene in an orthogroup), we found that 49% (4,293/8,724), 74% (6,025/8,063) and 18% (1,442/7,844) of shared orthogroups cycle (Figure 6B; Supplemental Figure 15D). Not only do more genes cycle in *Arabidopsis* and rice, more shared orthogroups cycle, consistent with fewer gene families cycling in *Wolffia*. There are 658 *Wolffia*-specific (not in *Arabidopsis* and rice) orthogroups with one or more genes, and only 12 of them cycle (1.8%). Therefore, we conclude that the fewer cycling genes in *Wolffia* is not due to smaller gene families. Further analysis of GO terms associated with the cycling OGs is consistent with *Wolffia* having a subset of pathways controlled in a TOD fashion compared to *Arabidopsis* and rice where processes associated with more complex plant architecture (cell types, tissues and life history) cycle (Figure 6; Supplemental Figure S16).

Supplemental Figures

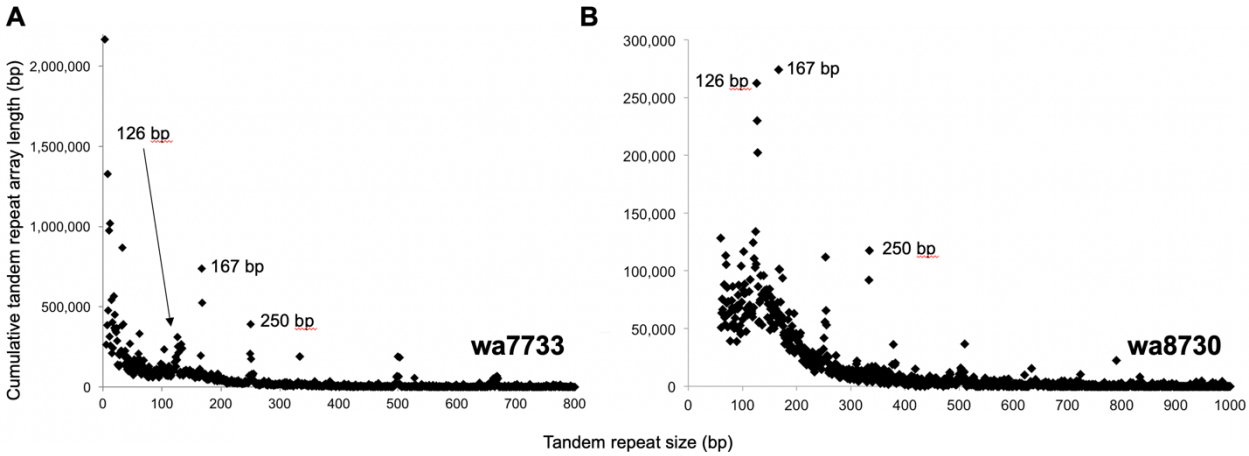
Supplemental Figure S1. *Wolffia australiana* under growth assay. *W. australiana was* pre-cultivated for four weeks with a subculture interval of 1 week at 25°C and continuous white light of an intensity of 100 µmol m^{-2} s^{-1}. For growth rate measurement, 10 ± 1 fronds from the pre-cultures were inoculated into 100 ml nutrient medium N contained in 150 ml beakers. The plants were grown for 1 week at 25°C and continuous white light of an intensity of 100 µmol m^{-2} s^{-1}.



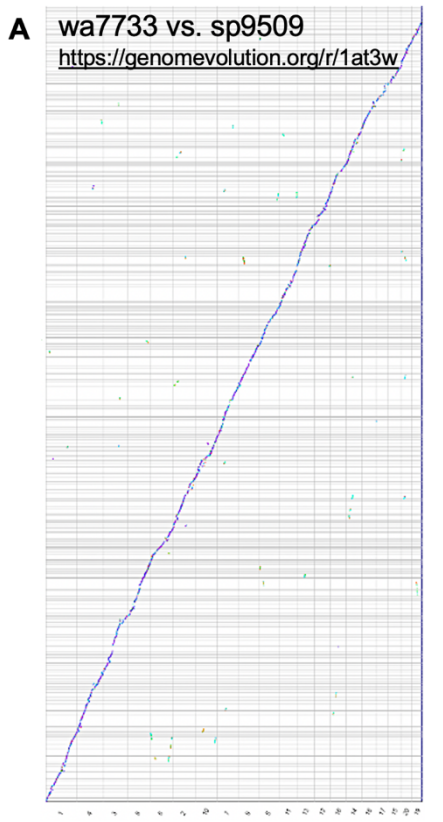
D

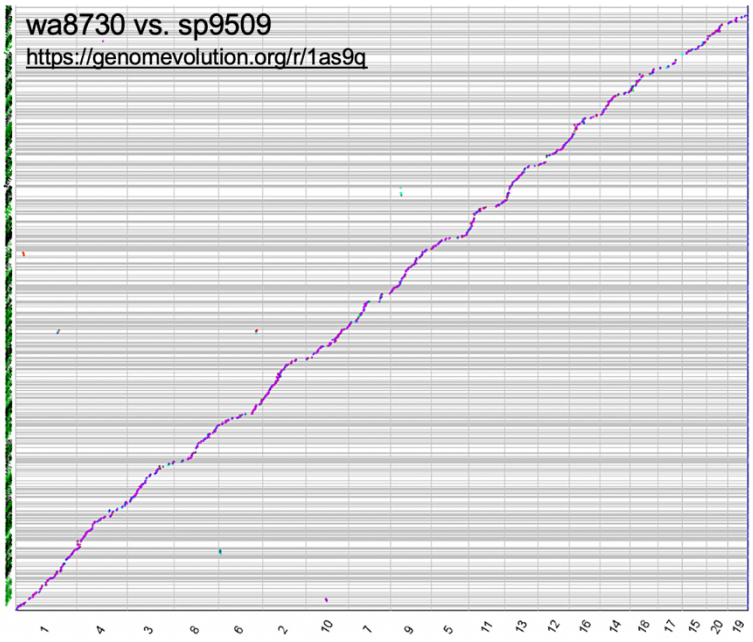
	wa7733		wa8730		wa7540	
	min	max	min	max	min	max
Heterozygosity	0.43%	0.43%	0.29%	0.30%	0.60%	0.60%
Genome length (bp)	343,520,172	343,771,258	345,863,147	346,004,779	342,809,947	343,081,944
Genome repeat (bp)	106,901,092	106,979,228	111,346,478	111,392,074	112,983,649	113,073,294
Genome unique (bp)	236,619,080	236,792,030	234,516,670	234,612,705	229,826,298	230,008,650
Repeat (%)	31.1	31.1	32.2	32.2	33.0	33.0
Model fit	92.21%	96.13%	94.73%	98.29%	94.74%	98.34%
Read error (%)	1.09%	1.09%	0.47%	0.47%	1.09%	1.09%
K-mer	19		19		19	

Supplemental Figure S2. Genome size estimated by K-mer (k=19) frequency. A) wa7733, B) wa8740, and C) wa7540. K-mers (19 bp) were counted with jellyfish and the resulting ”histo” file was plotted using GenomeScope. D) Table of GenomeScope results. All three genomes are estimated to be ~340 Mb with 0.3-0.4% heterozygosity and 100 Mb (30%) repeat content.



Supplemental Figure S3. *Wolffia* putative centromere base and higher order repeats. A) wa7733 and B) wa8730 assemblies were analyzed with Tandem Repeat Finder (TRF) to identify abundant repeats. The identified repeats were summed by repeat length and plotted as a function of repeat number to reveal the potential centromere arrays that are usually the highest repeats in genome assemblies. *W. australiana* has three distinct centromere repeats with the base array size of 126, 167 and 250 bp.

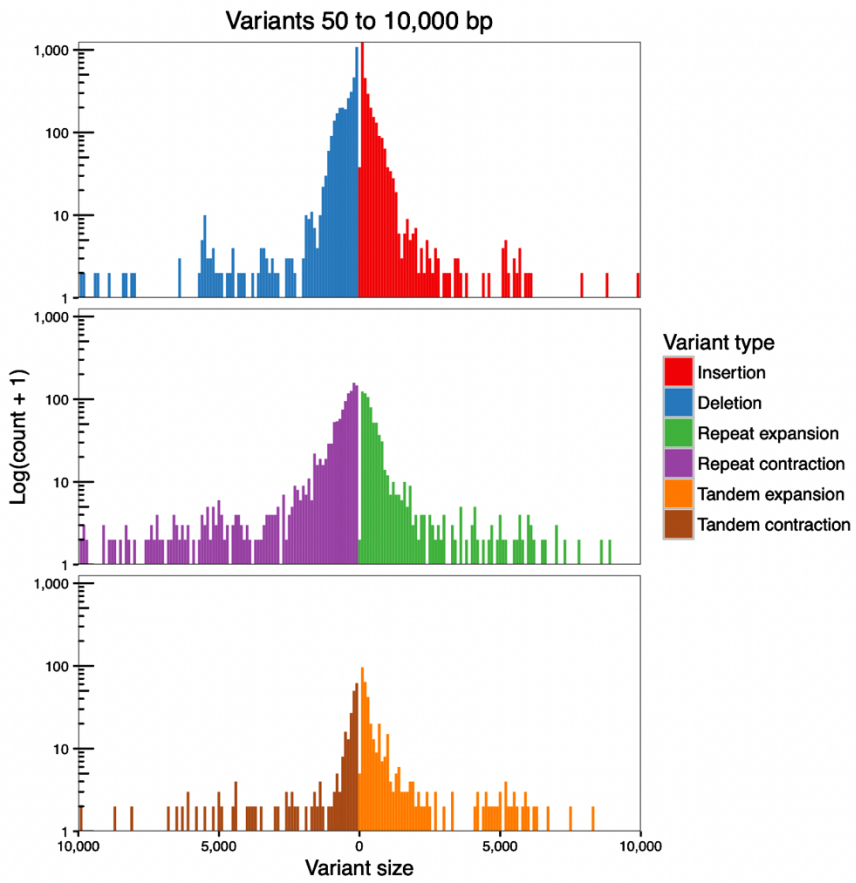




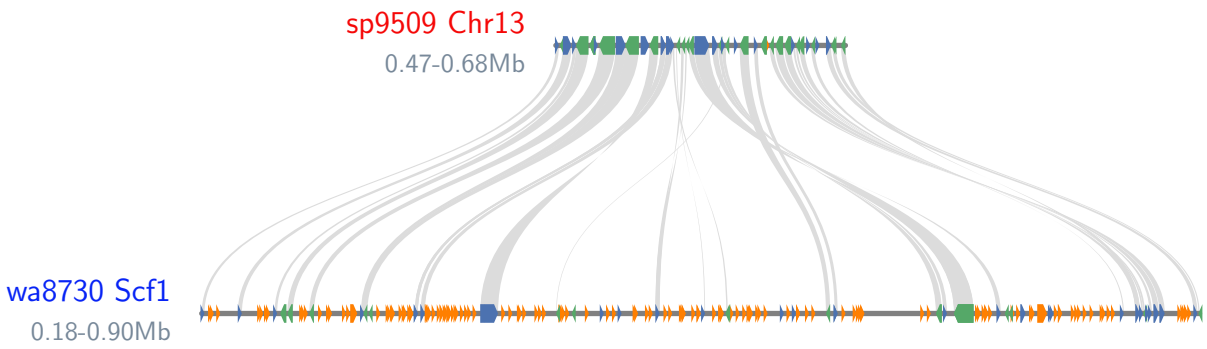
C

	wa7733	wa8730
Total gene level syntenic blocks (>5)	814	518
Average number of genes per block	10	7
Minimum genes per block	5	5
Maximum genes per block	92	21
Total genes in syntenic blocks	7,936	3,622

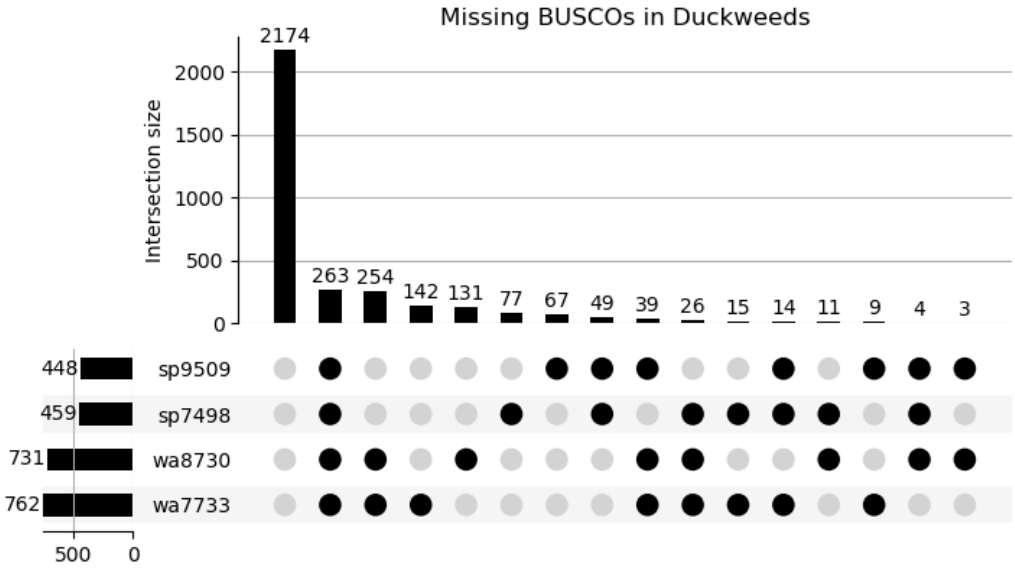
Supplemental Figure S4. *Wolffia* is collinear with sp9509. A) wa7733 vs. sp9509 and B) wa8730 vs. sp9509. C) Summary of the gene level syntenic blocks. Genomes were aligned using the last and syntenic regions based on proteins were identified and plotted using SynMap on CoGe (https://genomevolution.org/coge/).



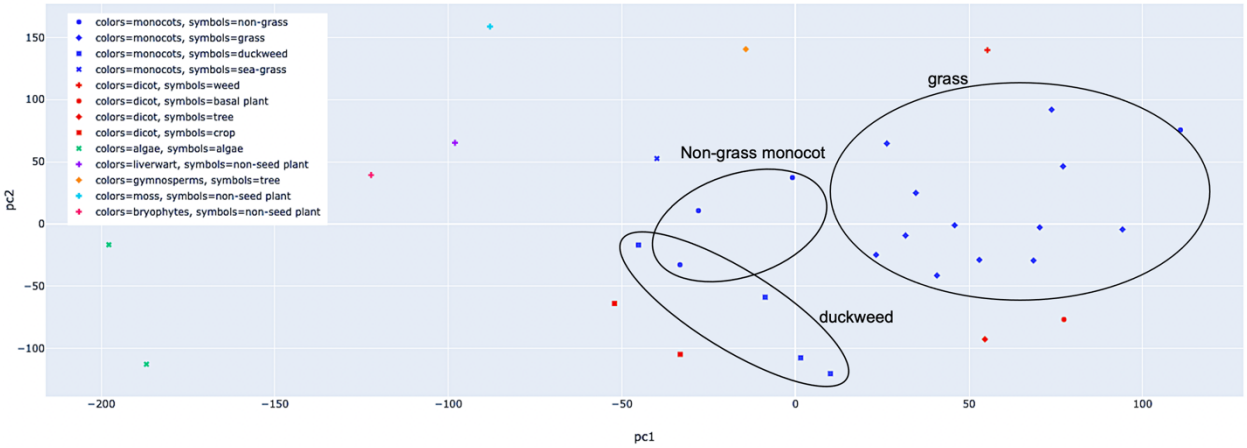
Supplemental Figure S5. *Wolffia* genomes differ mostly by small INDELs. Wa8730 and wa7733 were compared using mummer and alignments were analyzed for variation using Assemblytics (http://assemblytics.com/).



Supplemental Figure S6. *Wolffia* genome is collinear with sp9509. Proteins were aligned between sp9509 and wa8730 to identify syntenic regions. Grey bars are syntenic regions, green boxes are forward genes and blue boxes are reverse genes. The regions are to scale, highlighting the expansion in size of the wa8730 genome due to repeats (orange boxes).



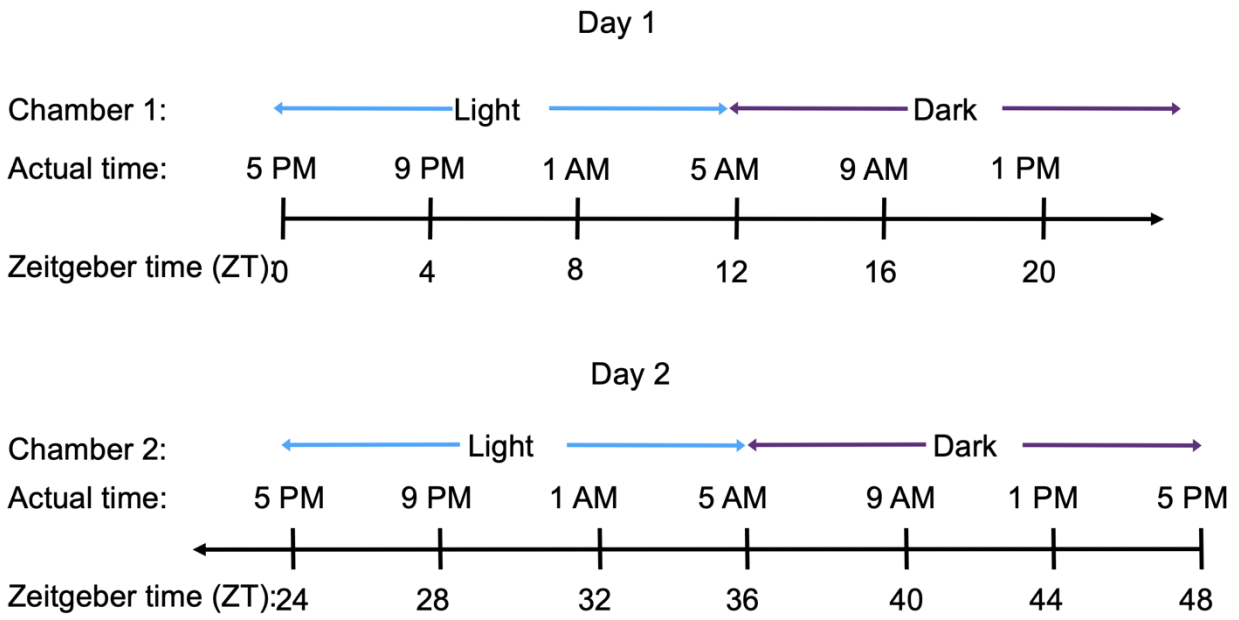
Supplemental Figure S7. *Wolffia* and *Spirodela* missing a similar set of BUSCO genes. An upset plot showing the number of BUSCO (Benchmarking Universal Single-Copy Orthologs) genes missing across the four duckweed genome assemblies: sp9509, sp7498, wa8730, and wa7733. There are 2174 BUSCO genes that are found in all four duckweed genomes. There are missing 263 BUSCO genes shared across the duckweed.



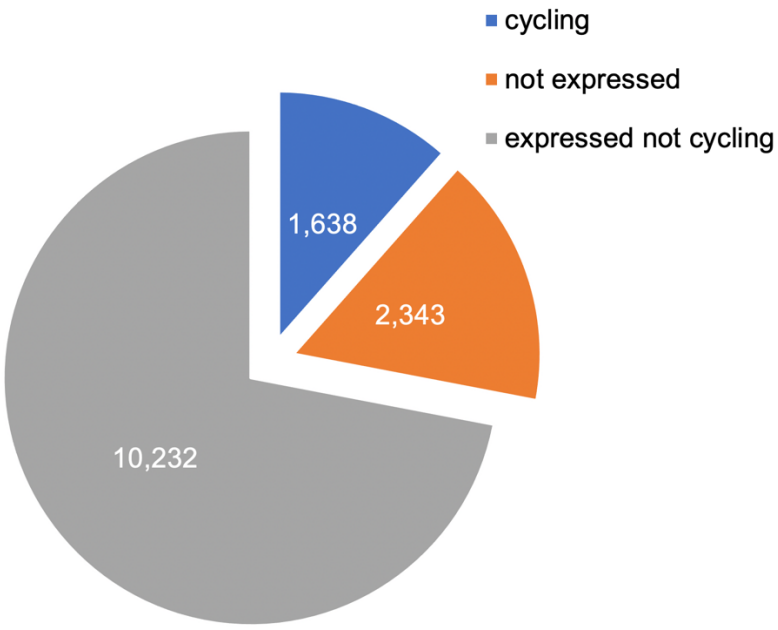
Supplemental Figure S8. Multi-dimensional scaling (MDS) based on orthogroups (OGs) reveals relationships between the grass monocots, non-grass monocots and duckweeds. Circles highlight the grass, non-grass and duckweed groupings. Monocots (blue): non-grass (circle), *Musa acuminate*, *Phalaenopsis equestris*, *Ananas comosus*, *Elaeis guineensis*; grass (diamond), *Phyllostachys edulis*, *Brachypodium distachyon, Hordeum vulgare, Oryza brachyantha, Oryza sativa japonica, Oryza sativa indica, Oropetium thomaeum, Sorghum bicolor, Setaria italica, Triticum aestivum, Zea mays, Zoysia japonica ssp. nagirizakizoma*; Duckweed (square), wa8730, wa7733, sp9509, sp7498; seagrass (x) *Zostera marina*. Dicot (red): weed (plus), *Arabidopsis thaliana*; basal plant (circle) *Amborella trichopoda*; tree (diamond), *Populus trichocarpa*; crop (square) *Solanum lycopersicum, Vitis vinifera*. Algae (green, x), *Micromonas commode*, *Chlamydomonas reinhardtii*. Non-seed plants, Liverwart (purple, plus), *Marchantia polymorpha*; moss (blue, cross), *Physcomitrella patens*; bryophytes (red, plus), *Selaginella moellendorffii*. Gymnosperms (orange, diamond), *Picea abies*.

Number of genes per-species in orthogroup	Wa7733	Wa8730	Sp9509	Sp7498	*Zostera*	*Arabidopsis*	rice	brachy	maize
1	41	45	37	33	26	18	18	19	12
2	20	20	19	18	17	14	12	13	13
3	9	9	9	10	10	9	8	9	9
4	6	5	6	6	7	7	5	6	8
5	3	3	4	4	5	5	4	4	6
6	3	3	4	4	4	4	3	3	5
7	2	2	2	3	3	3	2	2	5
8	2	2	2	2	2	3	2	2	3
9	1	1	2	2	2	2	2	2	3
10	1	1	1	1	2	2	2	2	2
11-15'	3	3	4	4	6	7	6	6	7
16-20	2	2	2	2	2	4	3	4	5
21-50	3	2	3	4	4	8	8	8	9
51-100	1	1	1	1	1	2	3	3	2
101-150	1	1	0	0	1	1	2	1	1
151-200	0	0	1	1	0	0	0	0	1
201-500	0	0	0	0	0	0	0	0	1
501-1000	0	0	0	0	0	0	0	0	0
1001+	0	0	0	0	0	0	0	0	0

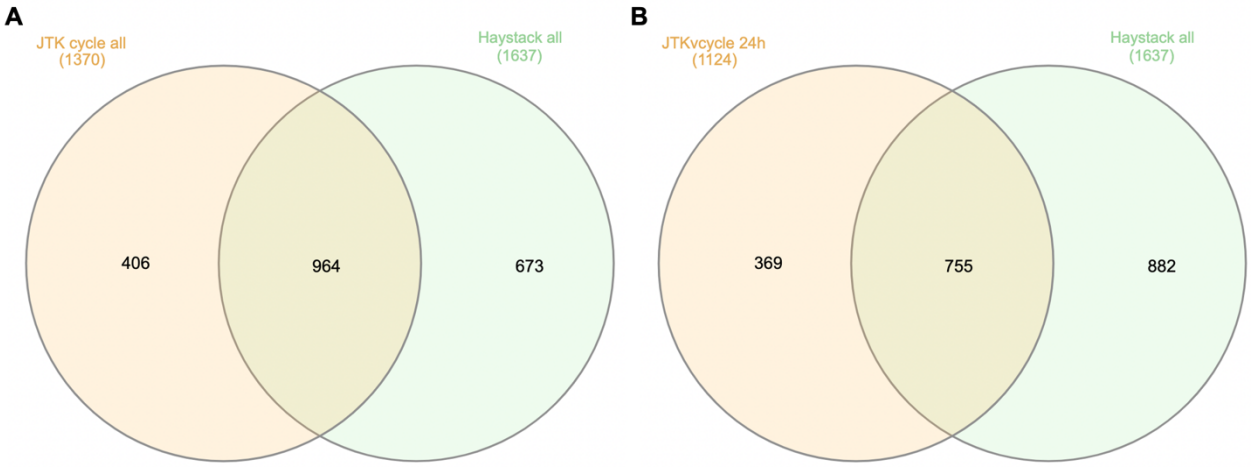
Supplemental Figure S9. *Wolffia* and *Spirodela* genes found in small gene families (orthogroups). 33-41% of *Wolffia* (wa7733 and wa8730) and *Spirodela* (sp9509 and sp7498) genes are found in orthogroups (OGs) with only one gene, whereas *Arabidopsis*, rice, brachy (*Brachypodium distachyon*) and maize have between 12-18% and over 20% in OGs with more than 10 genes. In contrast, less than 10% of duckweed genes are found in OGs with more than 10 genes.



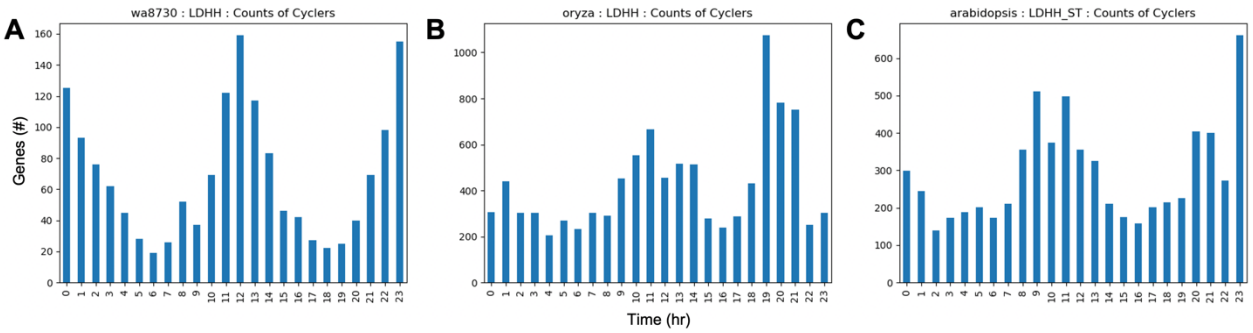
Supplemental Figure S10. *Wolffia* time course design. Two chambers were set to 12 hrs or light and 12 hrs of dark and 20^{o}C but 12 hrs apart opposite phases so plants could be collected over a 12 hr period without collecting during “actual night.” Samples were collected every four (4) hours over two days for a total of 13 time points. Zeitgeber time (ZT) is the time relative to the light dark cycle where ZT0 equals lights on.



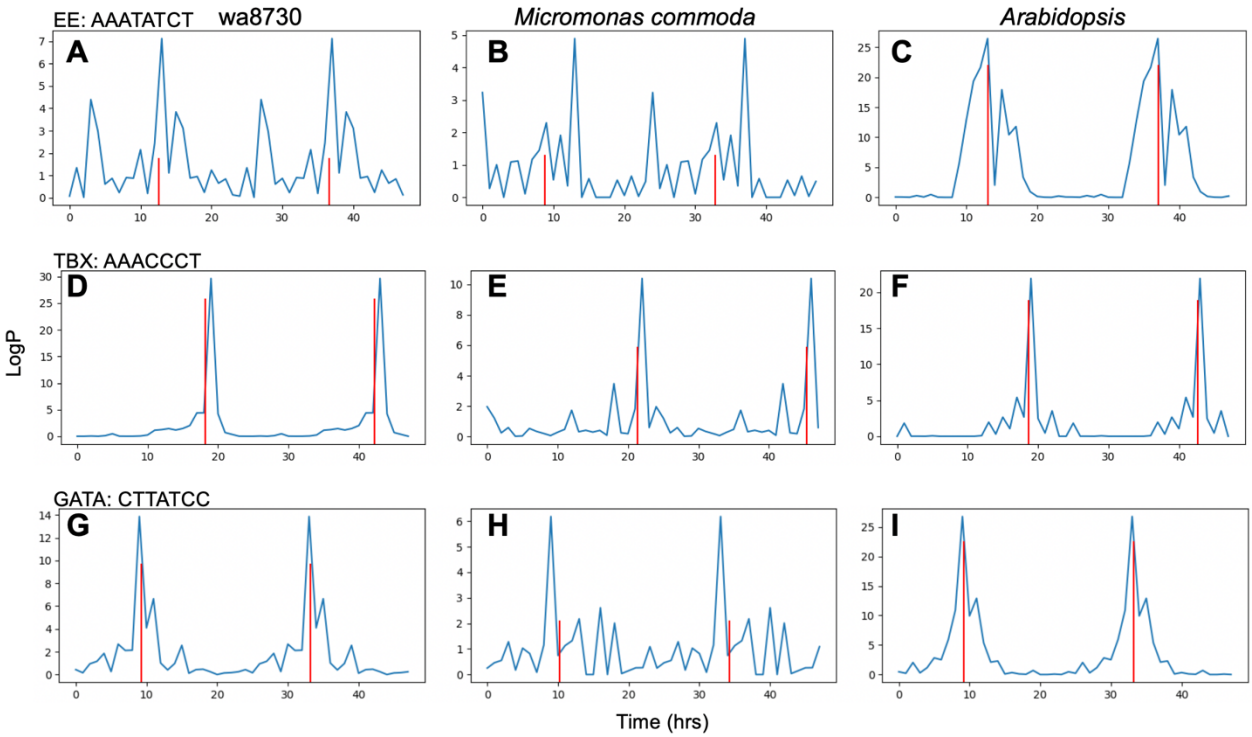
Supplemental Figure S11. Break down of expressed and cycling genes in wa8730. Total number of genes reflects the predicted protein coding genes in wa8730. Genes with mean expression across the 13 timepoints below 1 FPKM were considered "not expressed" (2,343, orange). Genes that correlated with their best fit model at a threshold of R > 0.8 were classified as "cycling" with phase and period defined by the best fit model (1,638; 13%). The rest were classified as "expressed not cycling" (10,232).



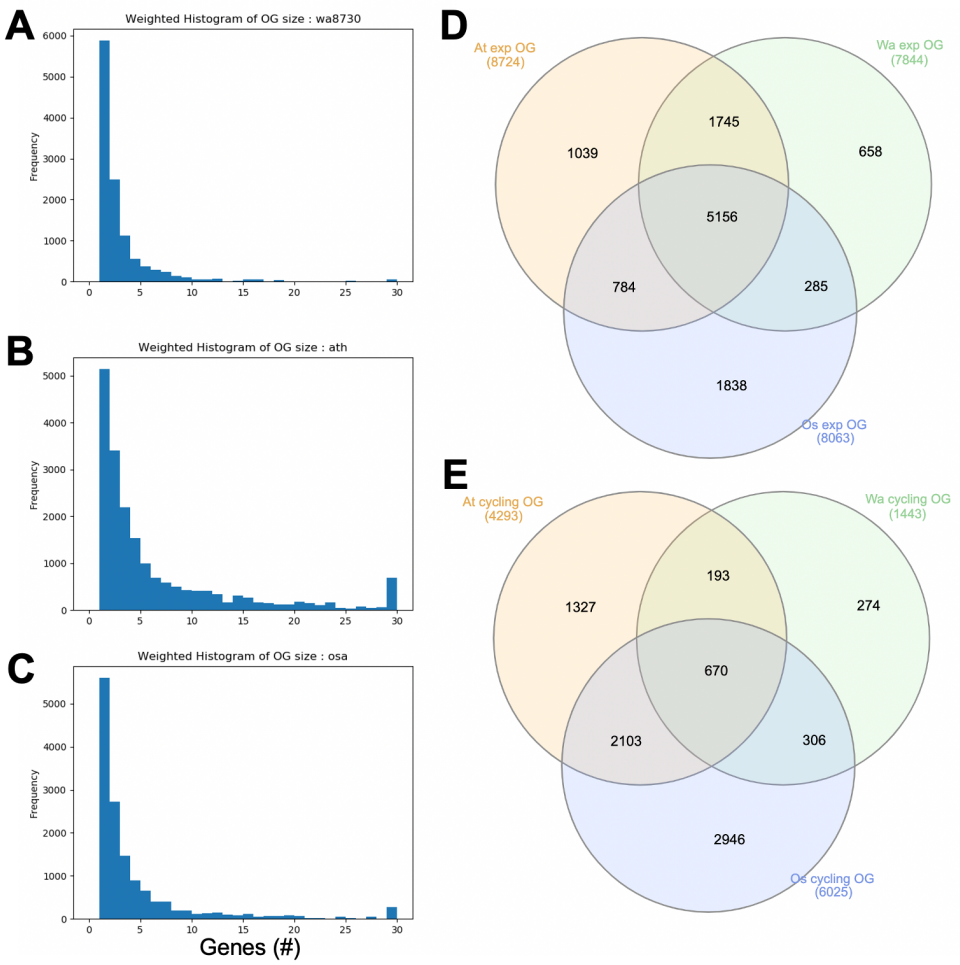
Supplemental Figure S12. Comparison of *Wolffia* cycling genes predicted using two different software tools. Wolffia genes that were not expressed were removed from the cycling prediction and JTKcycle was run using two different settings: all periods (20, 24 and 28 hrs; “JTK cycle all”) and only 24 hrs (“JTK cycle 24h”) A) Venn diagram of overlapping predicted genes between “JTKcycle all” (P<0.05) versus HAYSTACK (R>0.8). B) Venn diagram of overlapping predicted genes between “JTKcycle 24h” (P<0.05) versus HAYSTACK (R>0.8). Both methods produced similar number of cycling genes while JTKcycle consistently called fewer cycling genes.



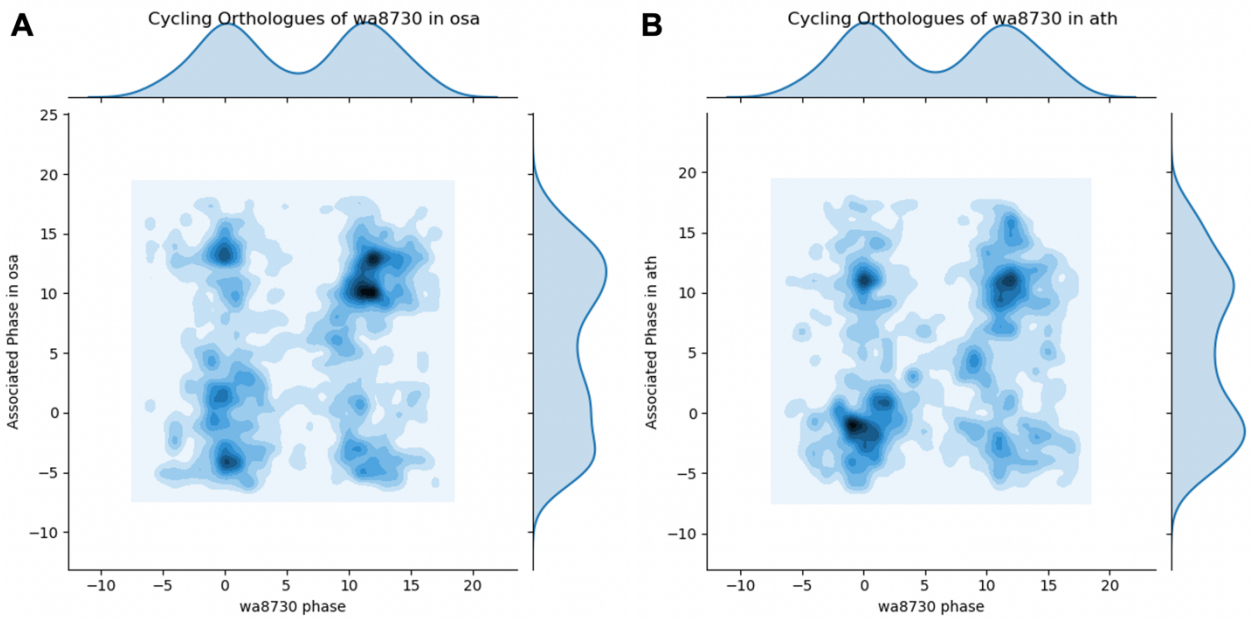
Supplemental Figure S13. Distribution of cycling genes over the day. A) wa8730, B) *Oryza* (rice; *Oryza sativa*), C) *Arabidopsis* (*Arabidopsis thaliana*). Genes that correlated with their best fit model at a threshold of R > 0.8 were classified as “cycling” with phase and period defined by the best fit model.



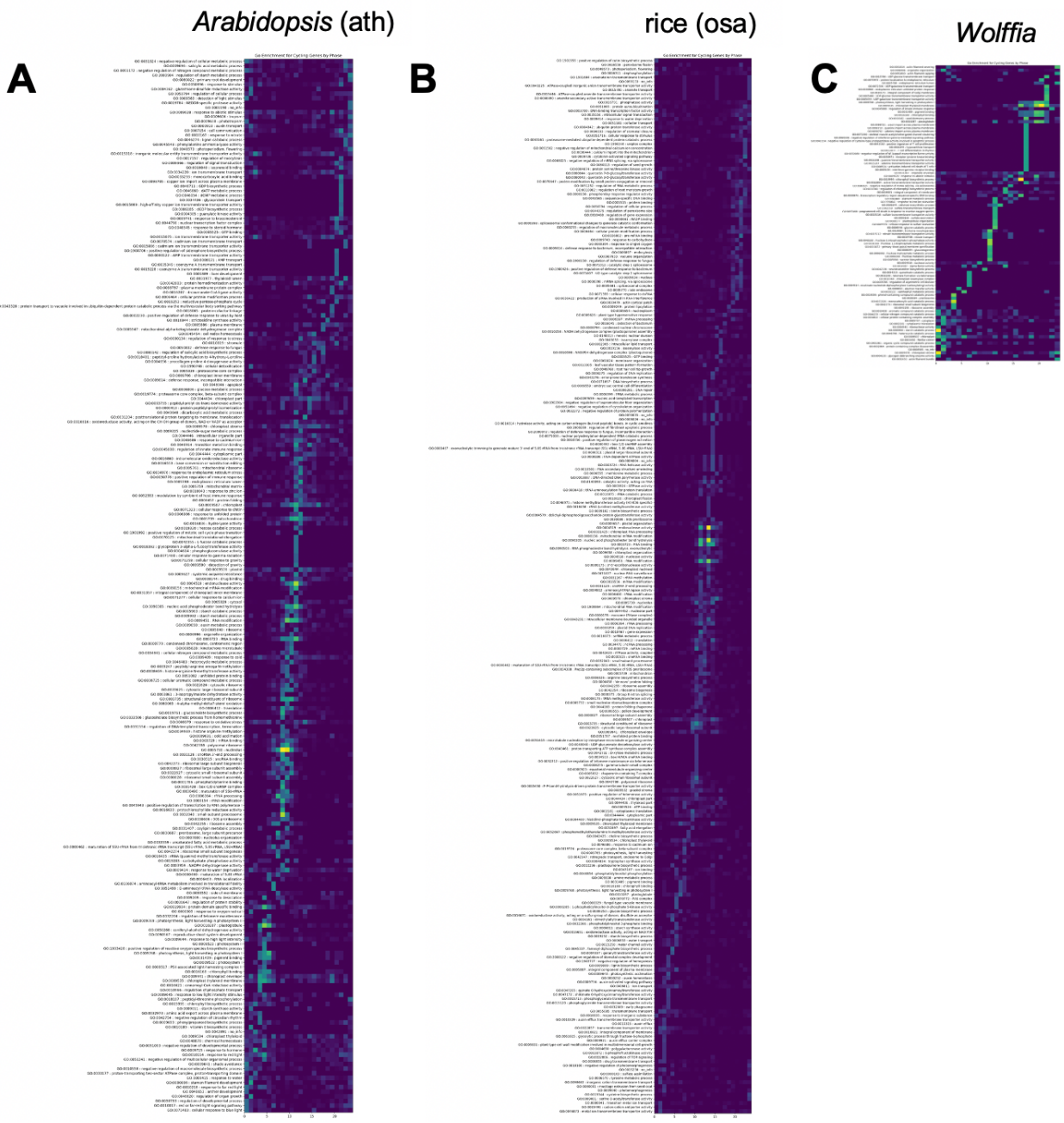
Supplemental Figure S14. Conserved TOD cis-elements from Algae to *Arabidopsis*. A-C) The EE is overrepresented across all three species. D-F) The TBX is overrepresented across all three species. G-I) The GATA cis-element is overrepresented across all three species. *Arabidopsis* cycling genes under LDHH were used to find Wolffia and *Micromonas commode* (Mco) orthologs, and the phase was assigned to those genes. Promoters were then searched for overrepresented cis-elements.



Supplemental Figure S15. *Wolffia* cycling orthogroups (OGs). Weighted histogram of the number of genes per OGs that have at least one expressed gene for A) Wolffia (wa8730), B) *Arabidopsis* (ath) and C) rice (osa). As seen in Supplemental Figure 9 for all genes regardless of expression, *Wolffia* has fewer genes per OG compared to *Arabidopsis* and rice that have some families with many genes per OG. D) Venn diagram of OGs with at least one expressed gene for *Arabidopsis* (At), *Wolffia* (Wa) and rice (Os). E) Venn diagram of cycling OGs with at least one expressed gene for *Arabidopsis* (At), *Wolffia* (Wa) and rice (Os); separated version of Figure 6B.



Supplemental Figure S16. Comparison of cycling genes between *Wolffia* (wa8730), *Arabidopsis* (ath) and rice (osa). Predicted phases were compared between wa8730 and osa (A), as well as wa8730 and ath (B). Density (darker blue) represents the number of genes with an associated phase. Using the wa8730 cycling genes, the ath or osa orthologues were identified and then the predicted phase was compared. While most genes are on the diagonal, which is consistent with similar phases across the species, there are some genes with different phases which could reflect that ath and osa have multiple wa8730 orthologues.



Supplemental Figure S17. Time of day (TOD) overrepresentation of Gene Ontology (GO) terms in *Wolffia*, rice and *Arabidopsis*. A) *Arabidopsis* (ath), B) rice (osa) and C) *Wolffia* significantly overrepresented TOD GO terms (y-axis) by TOD (x-axis). Dark color is less significant and light color is more significant.

Supplemental Tables

Supplemental Table S1. *Wolffia australiana* growth rates.
Supplemental Table S2. BioNano Genomics optical map summary.
Supplemental Table S3. *Wolffia australiana* genome size estimates by flow cytometry.
Supplemental Table S4. Ribosomal DNA content of the *Wolffia* genomes.
Supplemental Table S5. BUSCO scores across *Wolffia* and *Spirodela* genomes.
Supplemental Table S6. Missing BUSCO genes in wa8730 with *Arabidopsis* representative and annotation.
Supplemental Table S7. Orthogroup summary (OG) comparing *Wolffia*, *Spirodela* and PLAZA monocot proteomes.
Supplemental Table S8. *Wolffia* unique orthogroup (OG) Gene Ontology (GO) terms.
Supplemental Table S9. Gene in *Wolffia* unique orthogroups (OGs).
Supplemental Table S10. *Arabidopsis* gene IDs, annotation and orthogroups (OGs) for genes missing in *Wolffia*.
Supplemental Table S11. Circadian, light flowering time and temperature related gene numbers across species.
Supplemental Table S12. Expression across all *Wolffia* genes.
Supplemental Table S13. Significant cycling wa8730 genes as estimated by HAYSTACK.
Supplemental Table S14. Significant cycling wa8730 genes as estimated by JTKcycle.
Supplemental Table S15. All potential cis-elements in cycling promoters in wa8730 using ELEMENT.
Supplemental Table S16. Summary of significant TOD cis-elements and their DAP-seq transcription factor (TF) family matches.
Supplemental Table S17. Numbers of transcription factors in *Wolffia*, *Spirodela*, rice and *Arabidopsis*.
Supplemental Table S18. *Wolffia* TOD significant Gene Ontology (GO) terms.
Supplemental Table S19. *Arabidopsis* (ath) TOD significant Gene Ontology (GO) terms.
Supplemental Table S20. Rice (osa) TOD significant Gene Ontology (GO) terms.
Supplemental Table S21. NCBI SRA datasets for *Wolffia*.

References

Appenroth, K-J, S. Teller, and M. Horn. 1996. “Photophysiology of Turion Formation and Germination inSpirodela Polyrhiza.” *Biologia Plantarum* 38 (1): 95.

Appenroth, Klaus-J, K. Sowjanya Sree, Manuela Bog, Josef Ecker, Claudine Seeliger, Volker Böhm, Stefan Lorkowski, et al. 2018. “Nutritional Value of the Duckweed Species of the Genus Wolffia (Lemnaceae) as Human Food.” *Frontiers in Chemistry* 6 (October): 483.

Benson, G. 1999. “Tandem Repeats Finder: A Program to Analyze DNA Sequences.” *Nucleic Acids Research* 27 (2): 573–80.

Buchfink, Benjamin, Chao Xie, and Daniel H. Huson. 2015. “Fast and Sensitive Protein Alignment Using DIAMOND.” *Nature Methods* 12 (1): 59–60.

Campbell, Michael S., Meiyee Law, Carson Holt, Joshua C. Stein, Gaurav D. Moghe, David E. Hufnagel, Jikai Lei, et al. 2014. “MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations.” *Plant Physiology* 164 (2): 513–24.

Dai, Xinbin, Senjuti Sinharoy, Michael Udvardi, and Patrick Xuechun Zhao. 2013. “PlantTFcat: An Online Plant Transcription Factor and Transcriptional Regulator Categorization and Analysis Tool.” *BMC Bioinformatics* 14 (November): 321.

Devos, Katrien M., James K. M. Brown, and Jeffrey L. Bennetzen. 2002. “Genome Size Reduction through Illegitimate Recombination Counteracts Genome Expansion in Arabidopsis.” *Genome Research* 12 (7): 1075–79.

Dong, Xiaojing, Yan Yan, Bochen Jiang, Yiting Shi, Yuxin Jia, Jinkui Cheng, Yihao Shi, et al. 2020. “The Cold Response Regulator CBF1 Promotes Arabidopsis Hypocotyl Growth at Ambient Temperatures.” *The EMBO Journal* 39 (13): e103630.

Ellinghaus, David, Stefan Kurtz, and Ute Willhoeft. 2008. “LTRharvest, an Efficient and Flexible Software for de Novo Detection of LTR Retrotransposons.” *BMC Bioinformatics* 9 (January): 18.

Emms, David M., and Steven Kelly. 2015. “OrthoFinder: Solving Fundamental Biases in Whole Genome Comparisons Dramatically Improves Orthogroup Inference Accuracy.” *Genome Biology* 16 (August): 157.

———. 2019. “OrthoFinder: Phylogenetic Orthology Inference for Comparative

Genomics." *Genome Biology* 20 (1): 238.

Filichkin, Sergei A., Ghislain Breton, Henry D. Priest, Palitha Dharmawardhana, Pankaj Jaiswal, Samuel E. Fox, Todd P. Michael, Joanne Chory, Steve A. Kay, and Todd C. Mockler. 2011. "Global Profiling of Rice and Poplar Transcriptomes Highlights Key Conserved Circadian-Controlled Pathways and Cis-Regulatory Modules." *PloS One* 6 (6): e16907.

Goodstein, David M., Shengqiang Shu, Russell Howson, Rochak Neupane, Richard D. Hayes, Joni Fazo, Therese Mitros, et al. 2012. "Phytozome: A Comparative Platform for Green Plant Genomics." *Nucleic Acids Research* 40 (Database issue): D1178–86.

Grabherr, Manfred G., Brian J. Haas, Moran Yassour, Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, et al. 2011. "Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome." *Nature Biotechnology* 29 (7): 644–52.

Gupta, Shobhit, John A. Stamatoyannopoulos, Timothy L. Bailey, and William Stafford Noble. 2007. "Quantifying Similarity between Motifs." *Genome Biology* 8 (2): R24.

Haas, Brian J., Arthur L. Delcher, Stephen M. Mount, Jennifer R. Wortman, Roger K. Smith Jr, Linda I. Hannick, Rama Maiti, et al. 2003. "Improving the Arabidopsis Genome Annotation Using Maximal Transcript Alignment Assemblies." *Nucleic Acids Research* 31 (19): 5654–66.

Haas, Brian J., Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D. Blood, Joshua Bowden, Matthew Brian Couger, et al. 2013. "De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis." *Nature Protocols* 8 (8): 1494–1512.

Harkess, Alex, Fionn McLoughlin, Natasha Bilkey, Kiona Elliott, Ryan Emenecker, Erin Mattoon, Kari Miller, et al. 2020. "A New Spirodela Polyrhiza Genome and Proteome Reveal a Conserved Chromosomal Structure with High Abundances of Proteins Favoring Energy Production." https://doi.org/10.1101/2020.01.23.909457.

Hoang, Phuong N. T., Todd P. Michael, Sarah Gilbert, Philomena Chu, S. Timothy Motley, Klaus J. Appenroth, Ingo Schubert, and Eric Lam. 2018. "Generating a High-Confidence Reference Genome Map of the Greater Duckweed by Integration

of Cytogenomic, Optical Mapping, and Oxford Nanopore Technologies." *The Plant Journal: For Cell and Molecular Biology* 96 (3): 670–84.

Hoang, Phuong T. N., Veit Schubert, Armin Meister, Jörg Fuchs, and Ingo Schubert. 2019. "Variation in Genome Size, Cell and Nucleus Volume, Chromosome Number and rDNA Loci among Duckweeds." *Scientific Reports* 9 (1): 3234.

Hoff, Katharina J., Simone Lange, Alexandre Lomsadze, Mark Borodovsky, and Mario Stanke. 2016. "BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS." *Bioinformatics* 32 (5): 767–69.

Hoff, Katharina J., Alexandre Lomsadze, Mark Borodovsky, and Mario Stanke. 2019. "Whole-Genome Annotation with BRAKER." *Methods in Molecular Biology* 1962: 65–95.

Holt, Carson, and Mark Yandell. 2011. "MAKER2: An Annotation Pipeline and Genome-Database Management Tool for Second-Generation Genome Projects." *BMC Bioinformatics* 12 (December): 491.

Huerta-Cepas, Jaime, Kristoffer Forslund, Luis Pedro Coelho, Damian Szklarczyk, Lars Juhl Jensen, Christian von Mering, and Peer Bork. 2017. "Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper." *Molecular Biology and Evolution* 34 (8): 2115–22.

Hughes, Michael E., John B. Hogenesch, and Karl Kornacker. 2010. "JTK_CYCLE: An Efficient Nonparametric Algorithm for Detecting Rhythmic Components in Genome-Scale Data Sets." *Journal of Biological Rhythms* 25 (5): 372–80.

Jiang, Hongshan, Rong Lei, Shou-Wei Ding, and Shuifang Zhu. 2014. "Skewer: A Fast and Accurate Adapter Trimmer for next-Generation Sequencing Paired-End Reads." *BMC Bioinformatics* 15 (June): 182.

Jones, Philip, David Binns, Hsin-Yu Chang, Matthew Fraser, Weizhong Li, Craig McAnulla, Hamish McWilliam, et al. 2014. "InterProScan 5: Genome-Scale Protein Function Classification." *Bioinformatics* 30 (9): 1236–40.

Kawakatsu, Taiji, Shao-Shan Carol Huang, Florian Jupe, Eriko Sasaki, Robert J. Schmitz, Mark A. Urich, Rosa Castanon, et al. 2016. "Epigenomic Diversity in a Global Collection of Arabidopsis Thaliana Accessions." *Cell* 166 (2): 492–505.

Kidokoro, Satoshi, Koshi Yoneda, Hironori Takasaki, Fuminori Takahashi, Kazuo

Shinozaki, and Kazuko Yamaguchi-Shinozaki. 2017. “Different Cold-Signaling Pathways Function in the Responses to Rapid and Gradual Decreases in Temperature.” *The Plant Cell* 29 (4): 760–74.

Kim, Daehwan, Ben Langmead, and Steven L. Salzberg. 2015. “HISAT: A Fast Spliced Aligner with Low Memory Requirements.” *Nature Methods* 12 (4): 357–60.

König, Stefanie, Lars W. Romoth, Lizzy Gerischer, and Mario Stanke. 2016. “Simultaneous Gene Finding in Multiple Genomes.” *Bioinformatics* 32 (22): 3388–95.

Koren, Sergey, Brian P. Walenz, Konstantin Berlin, Jason R. Miller, Nicholas H. Bergman, and Adam M. Phillippy. 2017. “Canu: Scalable and Accurate Long-Read Assembly via Adaptive K-Mer Weighting and Repeat Separation.” *Genome Research* 27 (5): 722–36.

Langmead, B., and S. L. Salzberg. 2013. “Langmead. 2013. Bowtie2.” *Nature Methods* 9: 357–59.

Li, Heng. 2018. “Minimap2: Pairwise Alignment for Nucleotide Sequences.” *Bioinformatics* 34 (18): 3094–3100.

Lou, Ping, Jian Wu, Feng Cheng, Laura G. Cressman, Xiaowu Wang, and C. Robertson McClung. 2012. “Preferential Retention of Circadian Clock Genes during Diploidization Following Whole Genome Triplication in Brassica Rapa.” *The Plant Cell* 24 (6): 2415–26.

Lutz, Kerry A., Wenqin Wang, Anna Zdepski, and Todd P. Michael. 2011. “Isolation and Analysis of High Quality Nuclear DNA with Reduced Organellar DNA for Plant Genome Sequencing and Resequencing.” *BMC Biotechnology* 11 (May): 54.

MacKinnon, K. J. M., B. J. Cole, C. Yu, and J. H. Coomey. 2019. “Changes in Ambient Temperature Are the Prevailing Cue in Determining Brachypodium Distachyon Diurnal Gene Regulation.” *bioRxiv*. https://www.biorxiv.org/content/10.1101/762021v2.abstract.

Ma, Jianxin, Katrien M. Devos, and Jeffrey L. Bennetzen. 2004. “Analyses of LTR-Retrotransposon Structures Reveal Recent and Rapid Genomic DNA Loss in Rice.” *Genome Research* 14 (5): 860–69.

Margulies, Marcel, Michael Egholm, William E. Altman, Said Attiya, Joel S. Bader, Lisa

A. Bemben, Jan Berka, et al. 2005. “Genome Sequencing in Microfabricated High-Density Picolitre Reactors.” *Nature* 437 (7057): 376–80.

Melters, Daniël P., Keith R. Bradnam, Hugh A. Young, Natalie Telis, Michael R. May, J. Graham Ruby, Robert Sebra, et al. 2013. “Comparative Analysis of Tandem Repeats from Hundreds of Species Reveals Unique Insights into Centromere Evolution.” *Genome Biology* 14 (1): R10.

Michael, Todd P. 2014. “Plant Genome Size Variation: Bloating and Purging DNA.” *Briefings in Functional Genomics* 13 (4): 308–17.

Michael, Todd P., Ghislain Breton, Samuel P. Hazen, Henry Priest, Todd C. Mockler, Steve A. Kay, and Joanne Chory. 2008. “A Morning-Specific Phytohormone Gene Expression Program Underlying Rhythmic Plant Growth.” *PLoS Biology* 6 (9): e225.

Michael, Todd P., Douglas Bryant, Ryan Gutierrez, Nikolai Borisjuk, Philomena Chu, Hanzhong Zhang, Jing Xia, et al. 2017. “Comprehensive Definition of Genome Features in Spirodela Polyrhiza by High-Depth Physical Mapping and Short-Read DNA Sequencing Strategies.” *The Plant Journal: For Cell and Molecular Biology* 89 (3): 617–35.

Michael, Todd P., Florian Jupe, Felix Bemm, S. Timothy Motley, Justin P. Sandoval, Christa Lanz, Olivier Loudet, Detlef Weigel, and Joseph R. Ecker. 2018. “High Contiguity Arabidopsis Thaliana Genome Assembly with a Single Nanopore Flow Cell.” *Nature Communications* 9 (1): 541.

Michael, Todd P., Todd C. Mockler, Ghislain Breton, Connor McEntee, Amanda Byer, Jonathan D. Trout, Samuel P. Hazen, et al. 2008. “Network Discovery Pipeline Elucidates Conserved Time-of-Day–Specific Cis-Regulatory Modules.” *PLoS Genetics* 4 (2): e14.

Mockler, T. C., T. P. Michael, H. D. Priest, R. Shen, C. M. Sullivan, S. A. Givan, C. McEntee, S. A. Kay, and J. Chory. 2007. “The DIURNAL Project: DIURNAL and Circadian Expression Profiling, Model-Based Pattern Matching, and Promoter Analysis.” *Cold Spring Harbor Symposia on Quantitative Biology* 72: 353–63.

Murashige, Toshio, and Folke Skoog. 1962. “A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures.” *Physiologia Plantarum*. https://doi.org/10.1111/j.1399-3054.1962.tb08052.x.

O'Malley, Ronan C., Shao-Shan Carol Huang, Liang Song, Mathew G. Lewsey, Anna Bartlett, Joseph R. Nery, Mary Galli, Andrea Gallavotti, and Joseph R. Ecker. 2016. "Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape." *Cell* 166 (6): 1598.

Paten, Benedict, Dent Earl, Ngan Nguyen, Mark Diekhans, Daniel Zerbino, and David Haussler. 2011. "Cactus: Algorithms for Genome Multiple Sequence Alignment." *Genome Research* 21 (9): 1512–28.

Schindelin, Johannes, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, et al. 2012. "Fiji: An Open-Source Platform for Biological-Image Analysis." *Nature Methods* 9 (7): 676–82.

Schneider, Michel, Lydie Lane, Emmanuel Boutet, Damien Lieberherr, Michael Tognolli, Lydie Bougueleret, and Amos Bairoch. 2009. "The UniProtKB/Swiss-Prot Knowledgebase and Its Plant Proteome Annotation Program." *Journal of Proteomics* 72 (3): 567–73.

Sharma, Mohan, Zeeshan Zahoor Banday, Brihaspati N. Shukla, and Ashverya Laxmi. 2019. "Glucose-Regulated HLP1 Acts as a Key Molecule in Governing Thermomemory." *Plant Physiology* 180 (2): 1081–1100.

Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs." *Bioinformatics* 31 (19): 3210–12.

Singh, Rajinder, Meilina Ong-Abdullah, Eng-Ti Leslie Low, Mohamad Arif Abdul Manaf, Rozana Rosli, Rajanaidu Nookiah, Leslie Cheng-Li Ooi, et al. 2013. "Oil Palm Genome Sequence Reveals Divergence of Interfertile Species in Old and New Worlds." *Nature* 500 (7462): 335–39.

Smit, A. F. A., R. Hubley, and P. Green. 2015. "RepeatMasker Open-4.0. 2013--2015."

Smit, Arian F. A., and Robert Hubley. 2008-2015. *RepeatModeler Open-1.0*. http://www.repeatmasker.org.

Sree, K. Sowjanya, Hans-Martin Dahse, Jima N. Chandran, Bernd Schneider, Gerhard Jahreis, and Klaus J. Appenroth. 2019. "Duckweed for Human Nutrition: No Cytotoxic and No Anti-Proliferative Effects on Human Cell Lines." *Plant Foods for*

Human Nutrition 74 (2): 223–24.

Sree, K. Sowjanya, Sailendharan Sudakaran, and Klaus-J Appenroth. 2015. “How Fast Can Angiosperms Grow? Species and Clonal Diversity of Growth Rates in the Genus Wolffia (Lemnaceae).” *Acta Physiologiae Plantarum / Polish Academy of Sciences, Committee of Plant Physiology Genetics and Breeding* 37 (10): 204.

Stanke, Mario, Mark Diekhans, Robert Baertsch, and David Haussler. 2008. “Using Native and Syntenically Mapped cDNA Alignments to Improve de Novo Gene Finding.” *Bioinformatics* 24 (5): 637–44.

Stanke, Mario, Oliver Schöffmann, Burkhard Morgenstern, and Stephan Waack. 2006. “Gene Prediction in Eukaryotes with a Generalized Hidden Markov Model That Uses Hints from External Sources.” *BMC Bioinformatics* 7 (February): 62.

Trapnell, Cole, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David R. Kelley, Harold Pimentel, Steven L. Salzberg, John L. Rinn, and Lior Pachter. 2012. “Differential Gene and Transcript Expression Analysis of RNA-Seq Experiments with TopHat and Cufflinks.” *Nature Protocols* 7 (3): 562–78.

UniProt Consortium. 2019. “UniProt: A Worldwide Hub of Protein Knowledge.” *Nucleic Acids Research* 47 (D1): D506–15.

Van Bel, Michiel, Tim Diels, Emmelien Vancaester, Lukasz Kreft, Alexander Botzki, Yves Van de Peer, Frederik Coppens, and Klaas Vandepoele. 2018. “PLAZA 4.0: An Integrative Resource for Functional, Evolutionary and Comparative Plant Genomics.” *Nucleic Acids Research* 46 (D1): D1190–96.

VanBuren, Robert, Doug Bryant, Patrick P. Edger, Haibao Tang, Diane Burgess, Dinakar Challabathula, Kristi Spittle, et al. 2015. “Single-Molecule Sequencing of the Desiccation-Tolerant Grass Oropetium Thomaeum.” *Nature* 527 (7579): 508–11.

VanBuren, Robert, Ching Man Wai, Xuewen Wang, Jeremy Pardo, Alan E. Yocca, Hao Wang, Srinivasa R. Chaluvadi, et al. 2020. “Exceptional Subgenome Stability and Functional Divergence in the Allotetraploid Ethiopian Cereal Teff.” *Nature Communications* 11 (1): 884.

Vaser, Robert, Ivan Sović, Niranjan Nagarajan, and Mile Šikić. 2017. “Fast and Accurate de Novo Genome Assembly from Long Uncorrected Reads.” *Genome*

Research 27 (5): 737–46.

Vurture, Gregory W., Fritz J. Sedlazeck, Maria Nattestad, Charles J. Underwood, Han Fang, James Gurtowski, and Michael C. Schatz. 2017. “GenomeScope: Fast Reference-Free Genome Profiling from Short Reads.” *Bioinformatics* 33 (14): 2202–4.

Wai, Ching Man, Sean E. Weise, Philip Ozersky, Todd C. Mockler, Todd P. Michael, and Robert VanBuren. 2019. “Time of Day and Network Reprogramming during Drought Induced CAM Photosynthesis in Sedum Album.” *PLoS Genetics* 15 (6): e1008209.

Walker, Bruce J., Thomas Abeel, Terrance Shea, Margaret Priest, Amr Abouelliel, Sharadha Sakthikumar, Christina A. Cuomo, et al. 2014. “Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement.” *PloS One* 9 (11): e112963.

Wang, Wenqin, Randall A. Kerstetter, and Todd P. Michael. 2011. “Evolution of Genome Size in Duckweeds (Lemnaceae).” *Journal of Botany* 2011 (July). https://doi.org/10.1155/2011/570319.

Wang, W., G. Haberer, H. Gundlach, C. Gläßer, T. Nussbaumer, M. C. Luo, A. Lomsadze, et al. 2014. “The Spirodela Polyrhiza Genome Reveals Insights into Its Neotenous Reduction Fast Growth and Aquatic Lifestyle.” *Nature Communications* 5: 3311.

Weigel, Detlef, and Jane Glazebrook. 2008. “Fixation, Embedding, and Sectioning of Plant Tissues.” *CSH Protocols* 2008 (February): db.prot4941.

Wick, Ryan R., Mark B. Schultz, Justin Zobel, and Kathryn E. Holt. 2015. “Bandage: Interactive Visualization of de Novo Genome Assemblies.” *Bioinformatics* 31 (20): 3350–52.

Zdepski, Anna, Wenqin Wang, Henry D. Priest, Faraz Ali, Maqsudul Alam, Todd C. Mockler, and Todd P. Michael. 2008. “Conserved Daily Transcriptional Programs in Carica Papaya.” *Tropical Plant Biology* 1 (3-4): 236–45.

Ziegler, P., K. Adelmann, S. Zimmer, C. Schmidt, and K-J Appenroth. 2015. “Relative in Vitro Growth Rates of Duckweeds (Lemnaceae) - the Most Rapidly Growing Higher Plants.” *Plant Biology* 17 Suppl 1 (January): 33–41.