Supplemental Information for

Streamlined spatial and environmental expression signatures characterize the minimalist duckweed *Wolffia australiana*

Tom Denyer[1], Pin-Jou Wu[1], Kelly Colt[2], Bradley W. Abramson[2,3], Zhili Pang[4], Pavel Solansky[1], Allen Mamerto[2], Tatsuya Nobori[5], Joseph R. Ecker[56], Eric Lam[4*], Todd P. Michael[2*], Marja C.P. Timmermans[1*]

[1] Center for Plant Molecular Biology, University of Tübingen, Auf der Morgenstelle 32, Tübingen 72076, Germany

[2] Plant Molecular and Cellular Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

[3] Applied Sciences and Life Sciences Laboratory, Noblis, Reston, Virginia, 20191 USA

[4] Department of Plant Biology, Rutgers, The State University of New Jersey, New Brunswick, New Jersey 08901, USA

[5] Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

[6]Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

***Correspondence:**

Eric Lam, eric.lam@rutgers.edu

Todd P. Michael, tmichael@salk.edu

Marja C.P. Timmermans, marja.timmermans@uni-tuebingen.de

This document contains:

Supplemental Dataset List

Supplemental Figures S1 to S5

Supplemental Text ‘Circadian and time of day control of biological processes’

Supplemental Interactive 3D UMAPs S1 to S3

Supplemental Tables S1 to S5

SUPPLEMENTAL DATASET LIST

Supplemental Dataset S1. Differential expressed genes for dusk Superclusters with enriched GO terms

Supplemental Dataset S2. Above and below water differential expressed genes with enriched GO terms

Supplemental Dataset S3. Differential expressed genes between dusk Superclusters of common tissue types (A vs C, B vs D) with enriched GO terms

Supplemental Dataset S4. Differential expressed genes for dusk Clusters with enriched GO terms

Supplemental Dataset S5. Stomatal differential expressed genes with enriched GO terms

Supplemental Dataset S6. Daughter and mother frond differentially expressed genes with enriched GO terms

Supplemental Dataset S7. Differential expressed genes for dawn Superclusters with enriched GO terms

Supplemental Dataset S8. Differential expressed genes for dawn Clusters with enriched GO terms

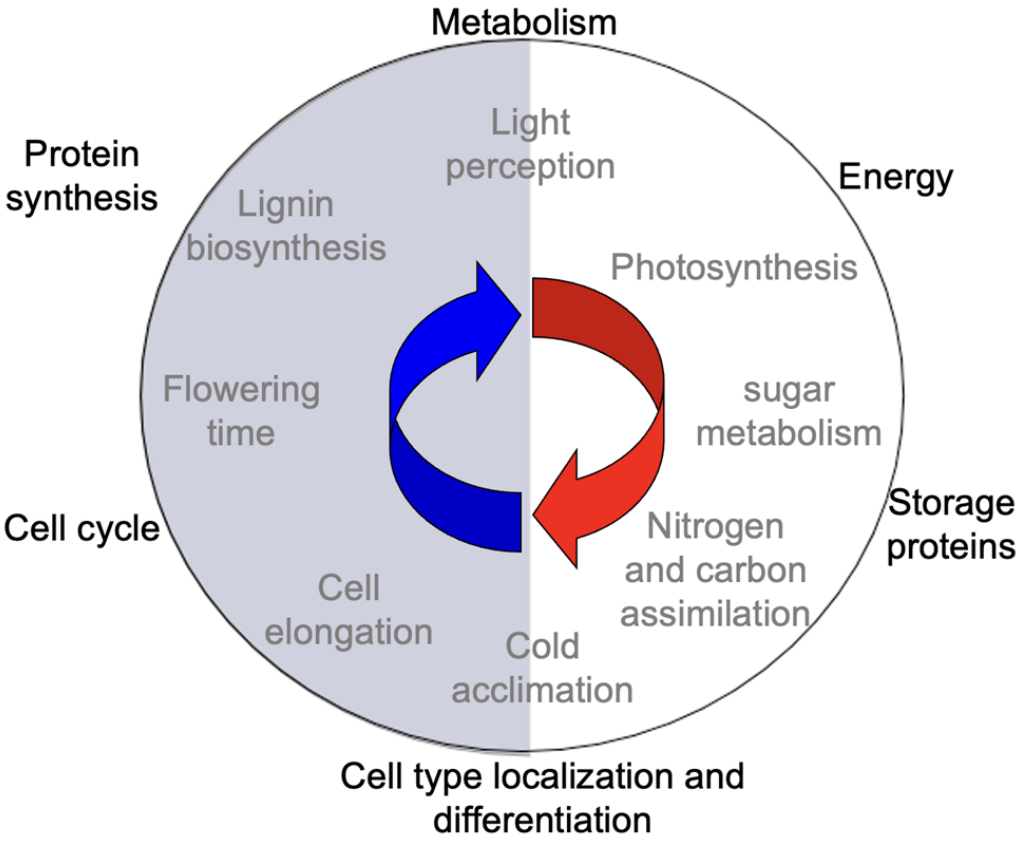
Supplemental Dataset S9. Differential expressed genes between dusk vs dawn Superclusters with enriched GO terms

Supplemental Dataset S10. *Wolffia* gene orthologues in model plants

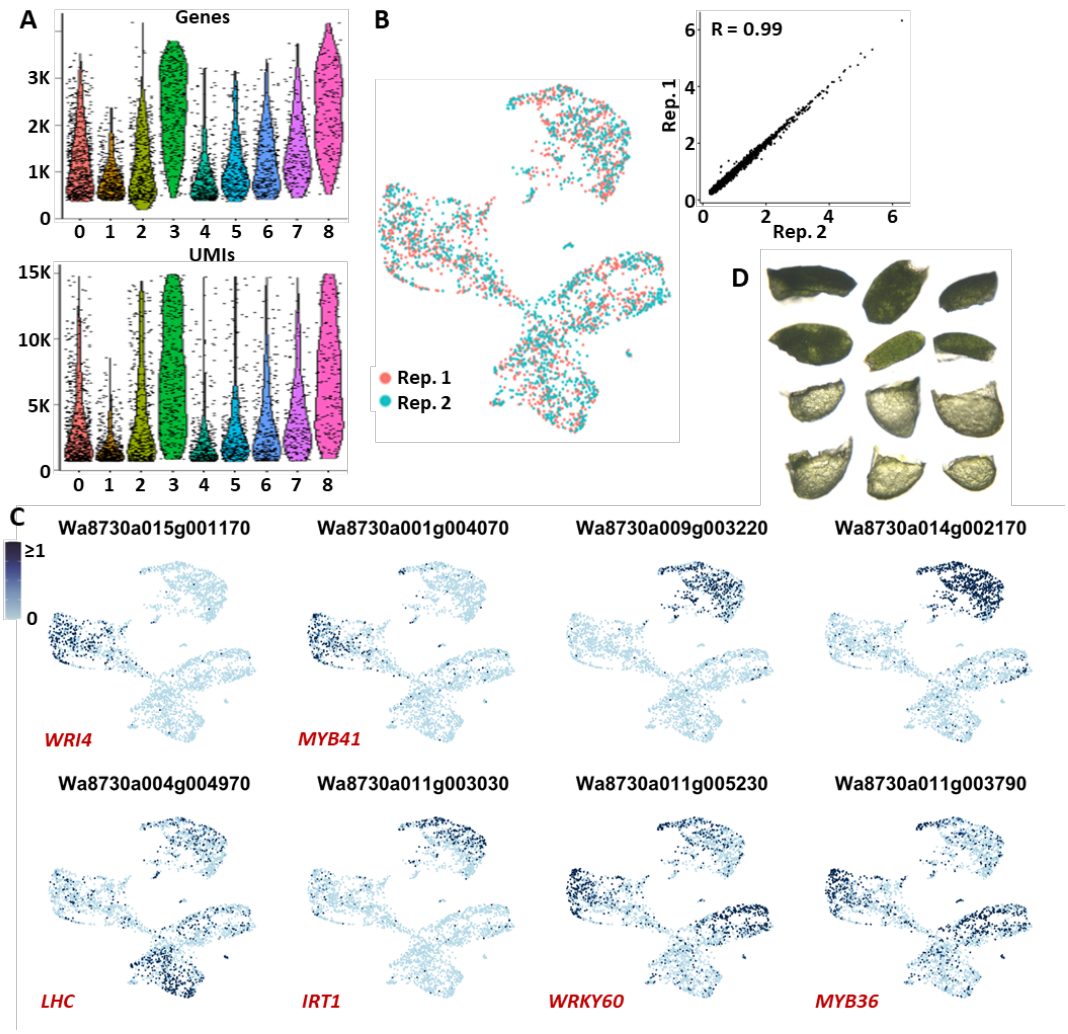
Supplemental Dataset S11. Genes linked to the cell cycle or circadian clock and their *Wolffia* orthologues

Supplemental Dataset S12. Genes referred to by common names

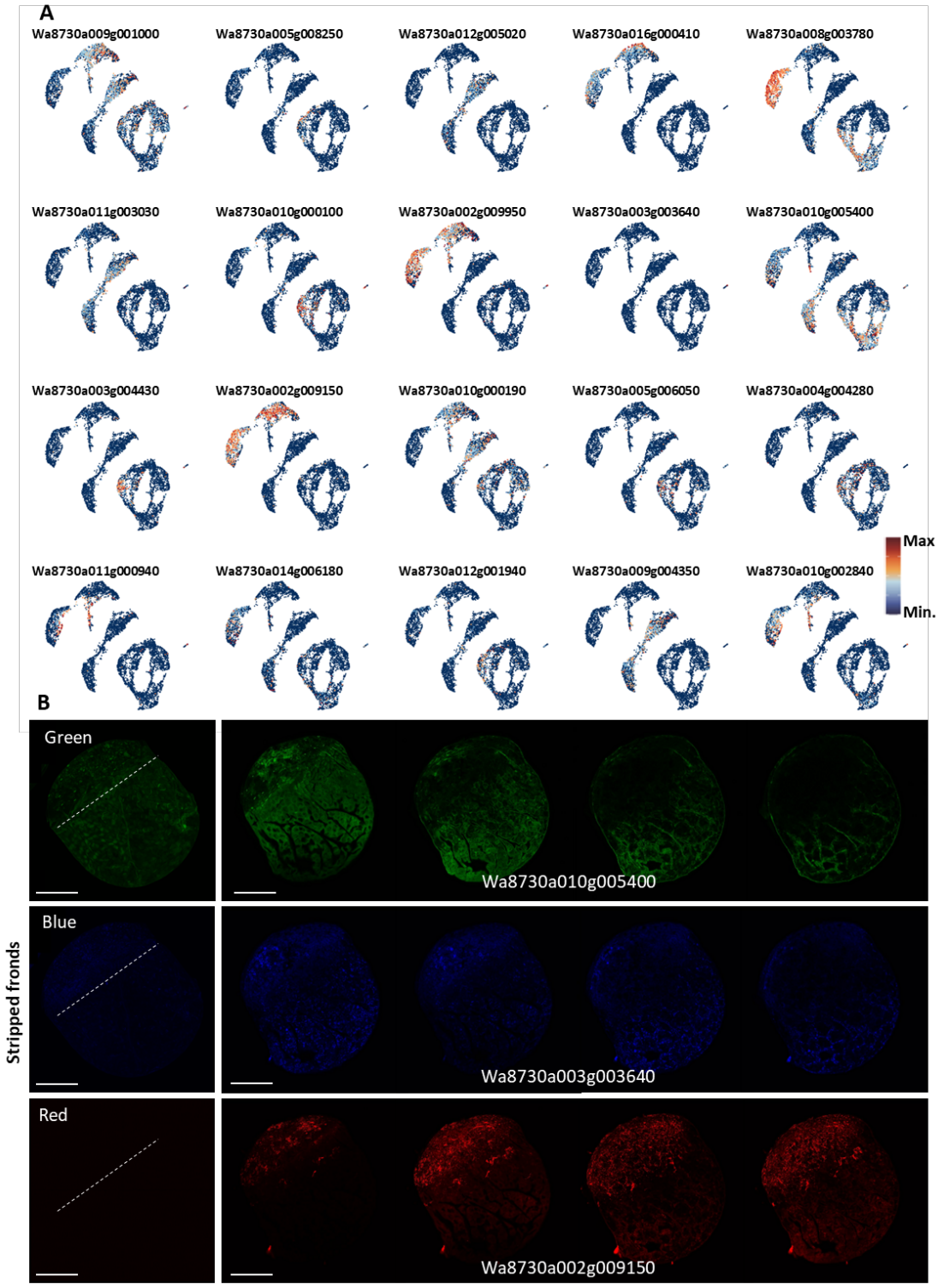
SUPPLEMENTAL FIGURES



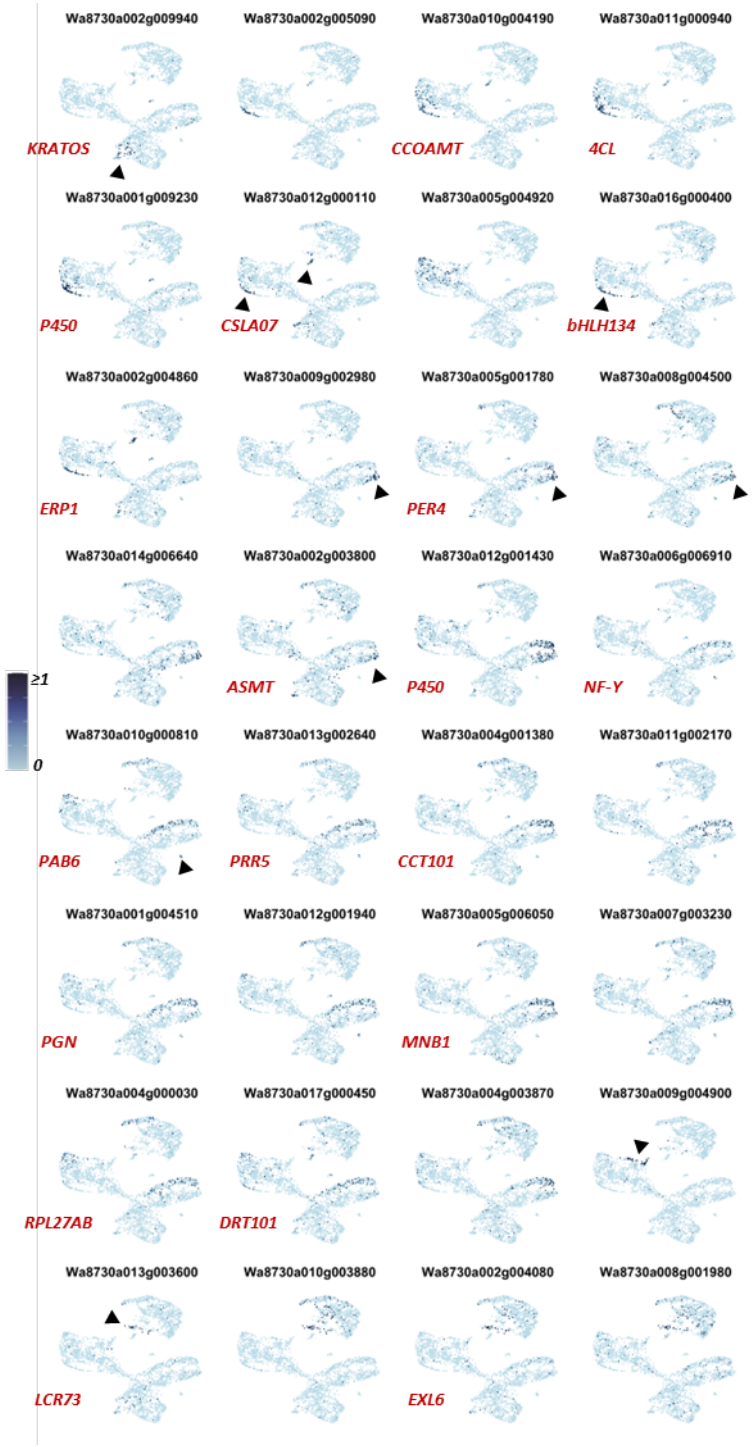
Supplemental Figure S1. Summary of TOD gene networks (related to Supplementary Text, below)



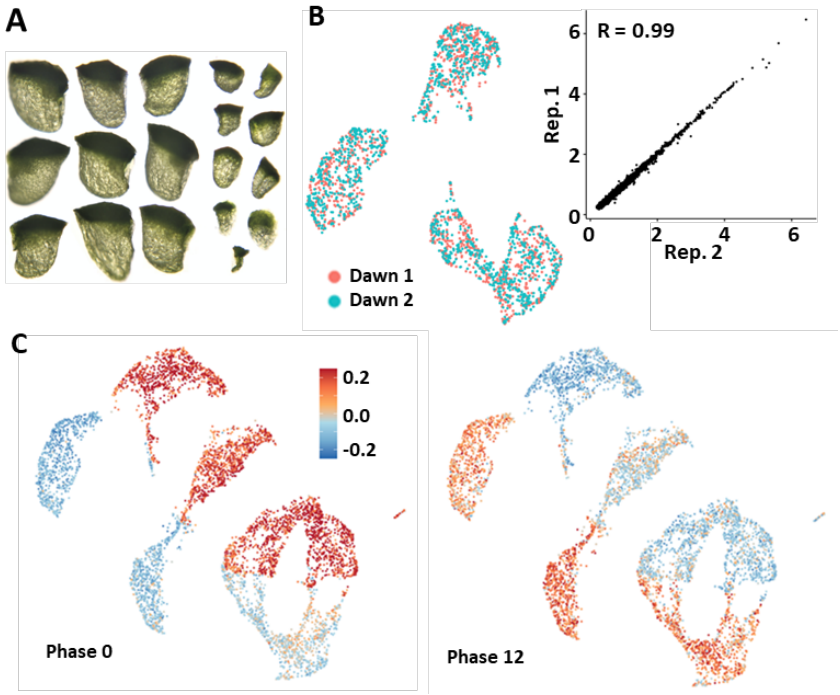
Supplemental Figure S2. A high-quality single cell atlas of whole *Wolffia* plants (related to Figure 1). **(A)** Distribution of the number of genes and UMIs per cell across clusters. Data for high quality cells (UMI > 650) within the dusk atlas are shown. **(B)** UMAP visualization highlighting cells from individual replicates (left) and Pearson's correlation of expression values between replicates (right) show that the two individual replicates within the dusk dataset are highly congruent. Only genes with expression in more than 10% of cells in each replicate were considered in each analysis. **(C)** UMAP projections of normalized expression for select differentially expressed genes within the atlas showing degrees of Supercluster specificity. Names of *Arabidopsis* orthologues are given in red. (Dataset S10). **(D)** Microscopic images of *Wolffia* plants bisected into above- (top two rows) and below-water (bottom two rows) segments.



Supplemental Figure S3. Genes examined by PHYTOMap multiplex *in situ* RNA detection (related to Figure 2). **(A)** UMAP projections of normalized expression for the tissue or condition specific genes examined by PHYTOMap (Table 2). **(B)** Left - Stacked confocal images of a *Wolffia* frond 'stripped' of fluorescent probes displaying a degree of autofluorescence, primarily in the green channel. Right - Four sequential optical sections in individual channels of the identical pre-stripped frond (set 3) showing distinct expression profiles as expected (Table 2). See Figure 2D for the composite image. Scale bar, 250 µm.



Supplemental Figure S4. Functional specialization within epidermal and parenchymal cells (related to Figure 3). UMAP projections of normalized gene expression for select genes differentially expressed across clusters showing intra-clustal localization (PCT1 <0.25, PCT2 <0.05, Dataset S4). Names of select *Arabidopsis* orthologues are given in red (Dataset S10). Arrows highlight localized expression.



Supplemental Figure S5. *Wolffia* cells show mother/daughter and time of day distinctions (related to Figures 3-5). **(A)** Microscopic images of *Wolffia* plants bisected into mother (left) and daughter (right) fronds. **(B)** UMAP visualization highlighting cells from individual replicates (left) and Pearson's correlation of expression values between replicates (right) show that the two replicates for the dawn atlas are highly congruent. Only genes with expression in more than 10% of cells in each replicate were considered in each analysis. **(C)** UMAP projections of 'gene set activity' (Seurat) profiles of time-of-day cycling genes as determined by bulk RNA-Seq (Michael et al., 2020).

SUPPLEMENTAL TEXT

Circadian and time of day control of biological processes

Under natural conditions, plants experience changing day lengths and thermocycles that they must integrate via the circadian clock to synchronize with the environment over time (Oravec and Greenham, 2022; Michael et al., 2008). A growing number of studies have assayed time of day (TOD) regulated genes expression across plant species under standard lab conditions (for example, Bläsing et al., 2005, Michael et al., 2008; Hazen et al. 2009; Khan et el., 2010; Filichkin et al. 2011; Sato et al., 2013; Ferrari et al., 2019; Wai et al., 2019; Greenham et al., 2020; Lai et al., 2020; Michael et al., 2020). In general, across plant species, over 50% of genes cycle under specific diurnal conditions of light and/or temperature cycles, while between 5-20% genes cycle under circadian conditions of continuous light/temperature. There are exceptions, such as *Wolffia australiana*, for which only 13% of its genes cycling under diurnal light cycles (Michael et al., 2020), and spruce that only has 5.2% of genes cycling under diurnal long day conditions (Ferrari et al., 2019). In *Arabidopsis thaliana*, when eleven different conditions of light, temperature and daylength were tested in whole seedlings (shoot and root), more than 90% of genes display TOD expression under at least one condition, consistent with different environmental conditions playing specific roles in phasing biological processes (Michael et al., 2008). In the only field based developmental time course in rice, between 5-9% and 57-88% of genes cycle in the roots and above ground leaf tissue, respectively (Michael, 2022b), while only 29.9-40.8% of genes cycle under lab conditions of light, temperature or light and temperature (Filichkin et al., 2011). These results in *Arabidopsis* and rice suggest there are additional TOD cues beyond light and temperature that together result in almost global TOD expression in some plants, perhaps under some conditions cycling information is obscured due to cell-specific expression as well (Swift et al., 2021).

The fact that almost all genes have peak and trough expression at specific TOD in plants is consistent with plants being sessile organisms that need to be able to partition their biological processes to the correct TOD to anticipate the daily changing environment (Green et al., 2002; Michael et al., 2003; Dodd et al., 2005). One can infer which biological process, cellular compartment and Molecular Function are controlled at what TOD by looking at the Gene Ontology (GO) enrichment for genes with the same phase of expression (usually TOD peak expression). The pioneering study to look at gene function as a function of TOD used microarrays to identify genes controlled by the circadian clock under continuous conditions and highlighted several pathways with TOD expression. The study found that phenylpropanoid biosynthesis genes peak before dawn, photosynthesis genes peak at midday, starch-mobilizing genes peak during the evening, and cell wall synthesis genes peak after midnight (Harmer et al., 2000). Several studies followed that showed under diurnal conditions of light and temperature, many major processes and gene families are controlled in a TOD specific manner (Figure S1, Bläsing et al., 2005; Covington and Harmer 2007; Michael et al., 2008; Pan et al., 2009; Ferrari et al., 2019).

In *Wolffia* there are 92 significant GO terms that are TOD specific (*Arabidopsis* and rice have 238 and 253 respectively). While the shared genes are associated with growth and photosynthesis, the GO terms specific to *Arabidopsis* and rice are associated with terms related to their more complex body plans and lifestyles. When

the *Wolffia* TOD GO terms (Michael et al., 2020) are summarized into dawn, midday, evening and midnight, we find:

- Dawn (0-6 hrs) - Response to abiotic stimulus, pigment metabolism, negative regulation of mRNA splicing via spliceosome
- Midday (6-12 hrs) - Photosynthesis, seed dormancy process, protein localization to endoplasmic reticulum, organelle organization
- Evening (12-18 hrs) - Telomere formation via telomerase, xanthophyll metabolism, starch catabolism, cytoplasmic translation
- Midnight (18-24 hrs) - Sulfate assimilation, nitrate transport, response to iron ion starvation, plastoglobule organization, primary shoot apical meristem specification, fructose 1,6-bisphosphate metabolism, glycine catabolism (full lists can be found in Michael et al., 2020).

In our scRNA-seq dataset we identify differentially expressed genes (DEGs) between the dawn (Zeitgeber (ZT) 0) and dusk (ZT12) Superclusters (Dataset S9). Using a strict cut-off (Log_2 FC >1 and adjusted p value <0.05), we find 364, 164, 157 and 157 TOD-specific DEGs for Superclusters CA, CB, CC and CD, respectively (Dataset S9). Consistent with the *Wolffia* TOD bulk RNA-seq, photosynthesis, response to light stimulus and chloroplast organization were enriched GO terms in the dawn DEGs, while translation and plastid were enriched GO terms in the dusk DEGs (Dataset S9). In the below water Superclusters, there were fewer significantly enriched GO terms in the dusk DEGs compared to the above water, suggesting either that the biological activities are more dispersed in these cell types or that our end of day (ZT12) timepoint didn't capture the TOD processes that occur in the middle of the night (midnight) such as cell cycle and protein synthesis (Dataset S9). Taken together, the scRNA-seq not only validates the TOD expression and predicted biological activities, it also suggests that there may be nuance at the single cell level that will require additional studies in whole plants.

References not otherwise in main text

Green RM., Tingay S, Wang Z-Y, Tobin EM. 2002. Circadian Rhythms Confer a Higher Level of Fitness to Arabidopsis Plants. *Plant Physiology* **129**, 576–84.

Greenham K, Sartor RC, Zorich S, Lou P, Mockler TC, McClung CR. 2020. Expansion of the Circadian Transcriptome in *Brassica Rapa* and Genome-Wide Diversification of Paralog Expression Patterns. *eLife* **9**.

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**, 2110–13.

Hazen SP, Naef F, Quisel T, Gendron JM, Chen H, Ecker JR, Borevitz JO, Kay SA. 2009. Exploring the Transcriptional Landscape of Plant Circadian Rhythms Using Genome Tiling Arrays. *Genome Biology* **10**, R17.

Khan S, Rowe SC, Harmon FG. 2010. Coordination of the Maize Transcriptome by a Conserved Circadian Clock. *BMC Plant Biology* **10**, 126.

Lai X, Bendix C, Yan L, Zhang Y, Schnable JC, Harmon FG. 2020. Interspecific Analysis of Diurnal Gene Regulation in Panicoid Grasses Identifies Known and Novel Regulatory Motifs. *BMC Genomics* **21**, 428.

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

Pan Y, Michael TP, Hudson ME, Kay SA, Chory J, Schuler MA. 2009. Cytochrome P450 Monooxygenases as Reporters for Circadian-Regulated Pathways. *Plant Physiology* **150**, 858–78.

Sato Y, Takehisa H, Kamatsuki K, Minami H, Namiki N, Ikawa H, Ohyanagi H, Sugimoto K, Antonio BA, Nagamura Y. 2013. RiceXPro Version 3.0: Expanding the Informatics Resource for Rice Transcriptome. *Nucleic Acids Research* **41**, D1206–13.

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 Genetics* **15,** e1008209.

SUPPLEMENTAL INTERACTIVE 3D UMAPS

The Supplemental Information folder contains three .html files that can be opened on a browser. These show 3D UMAP projections for the dusk-, dawn- and combined atlases. Dragging the image allows the viewer to explore the projection while a panel on the right allows the viewer to isolate select clusters, rotate the image and zoom.

Supplemental Interactive 3D UMAP S1 – Dusk dataset

Supplemental Interactive 3D UMAP S2 – Dawn dataset

Supplemental Interactive 3D UMAP S1 – Combined dusk/dawn dataset

SUPPLEMENTAL TABLES

Supplemental Table S1. Organization of clusters within annotated Superclusters of dusk and dawn datasets.

Condition	**Dusk**									**Dawn**							
Supercluster	CA			CB		CC		CD		CA		CB	CC		CD		
Supercluster Annotation	Above Water Epidermis			Above Water Parenchyma		Below Water Epidermis		Below Water Parenchyma		Above Water Epidermis		Above Water Parenchyma	Below Water Epidermis		Below Water Parenchyma		
Clusters	C1	C3	C7	C2	C5	C0	C8	C4	C6	C1	C3	C2	C0	C6	C4	C5	C7

Supplemental Table S2. Genes analyzed using PHYTOMap with average expression values in the integrated dusk/dawn scRNA-Seq dataset.

Gene	Fluorescence Channel[a]	Set	Notes	Expression Dusk[b]				Expression Dawn[b]			
				CA	CB	CC	CD	CA	CB	CC	CD
Wa8730a009g001000	Blue	1		1.14	0.77	0.91	0.78	0.23	0.28	0.26	0.33
Wa8730a005g008250	Green	1		0.09	0.04	0.06	0.71	0.06	0.02	0.07	0.40
Wa8730a012g005020	Magenta	1		0.18	0.05	1.08	0.17	0.02	0.02	0.81	0.14
Wa8730a016g000410	Red	1		1.35	0.00	0.00	0.05	1.24	0.00	0.03	0.03
Wa8730a008g003780	Blue	2		0.29	0.09	0.00	0.17	3.00	1.51	0.65	1.94
Wa8730a011g003030	Green	2		0.27	0.08	1.00	0.26	0.10	0.04	0.69	0.35
Wa8730a010g000100	Magenta	2		0.15	0.15	0.12	1.00	0.08	0.06	0.09	1.01
Wa8730a002g009950	Red	2		1.14	0.05	0.04	0.07	1.55	0.03	0.01	0.29
Wa8730a003g003640	Blue	3	No Data	0.00	0.01	0.00	0.04	0.00	0.00	0.00	0.21
Wa8730a010g005400	Green	3		0.12	0.36	0.21	0.19	0.80	1.28	1.25	1.20
Wa8730a003g004430	Magenta	3	No Data	0.10	0.06	0.07	0.99	0.03	0.07	0.02	0.92
Wa8730a002g009150	Red	3		1.91	0.05	0.03	0.10	1.94	0.04	0.04	0.02
Wa8730a010g000190	Blue	4		0.86	0.57	0.91	0.57	0.04	0.12	0.09	0.27
Wa8730a005g006050	Green	4		0.06	0.11	0.10	0.66	0.01	0.12	0.06	0.60
Wa8730a004g004280	Magenta	4	No Data	0.11	0.31	0.06	0.88	0.03	0.11	0.01	0.62
Wa8730a011g000940	Red	4		0.56	0.06	0.04	0.08	0.80	0.05	0.12	0.23
Wa8730a014g006180	Blue	5		0.04	0.02	0.03	0.11	0.55	0.31	0.26	0.23
Wa8730a012g001940	Green	5		0.08	0.05	0.13	0.58	0.13	0.10	0.10	0.45
Wa8730a009g004350	Magenta	5	No Data	0.09	0.02	0.82	0.02	0.02	0.01	0.62	0.06
Wa8730a010g002840	Red	5		0.45	0.04	0.02	0.04	1.00	0.23	0.16	0.25

[a]Genes were analyzed in sets of four with individual fluorescent probe.

[b]Average normalized expression across cells of a Supercluster. Expression values are natural logarithm transformed (log(1+value)).

Supplemental Table S3. Genes differentially expressed between dusk and dawn in various Superclusters, and crossover with those identified in a bulk tissue study (Michael et al., 2020).

Supercluster (SC)[a]	CA	CB	CC	CD	In any SC	In all SCs	In all SC and Bulk RNA-seq[b]
Up in dusk	179	72	73	68			
Up in dawn	185	92	84	89			
Total	364	164	157	157	456	76	58

[a] Supercluster DEGs - Log_2 FC>1 and adjusted p value <0.05

[b] Significance cut off for Bulk RNA-seq, R >0.8

Supplemental Table S4. scRNA-seq Metadata (as per Grones et al., 2024). Necessary reported information to allow evaluation and repetition of a plant single cell/nucleus experiment.

	Details	**Experimental information**
Biological material	Species	*Wolffia australiana*
	Accession	8730
	Tissue type	Whole plant
	Detailed growth conditions	Plants grown at 12 h light (100 µE)/12 h dark cycles at a constant 24 °C in 100 ml nutrient medium (0.5 × Schenk & Hildebrandt, 0.1% sucrose, pH 6.7).
	Harvest conditions	Dusk (lights off) or dawn (lights on)
Sample preparation	Isolation protocol	Digestion, 1.5% Cellulase R-10, 1% Maceroenzyme, 0.5% Pectolyase
	Total sample preparation time	Digestion to loading 2 H
	Estimated cell/nuclei number loaded	~5,000/sample
	Instrument/Method/Kit	10x Genomics Single Cell 3’ Reagent Kit v 3.1
	Cell viability test	Microscopic analysis
Libraries	Library construction	10x Genomics Single Cell 3’ Reagent Kit v 3.1
	Amplification method	As per protocol (11x PCR cycles for cDNA amplification, 12x cycles for final amplification)
Sequence results	Instrument/method	NovoSeq PE150
	N° sequenced reads	20,000 reads per cell
Raw data	Reference genome	*W. australiana* line 8730 (Wa8730.asm201904v2.fasta)
	Annotation version	V2
	Mapping method (incl. software, customized settings)	STAR (version 2.7.10a)
	Sequencing saturation	48.6%

	Fraction of reads in cells (10x Genomics)	~49.6 %
Processed data	N° captured cells/nuclei	5,604 cells (total)
	N° high quality cells/nuclei	5,604 cells (total)
	Filter criteria: % mitochondrial reads/cell or nucleus	Not filtered by this criterion
	Filter criteria: % chloroplast reads/cell or nucleus	Not filtered by this criterion
	Filter criteria: Minimum N° UMI/cell or nucleus	650 < Number of UMI per cell < 15,000
	N° total detected transcripts	Dusk: 12,825 Dawn: 12,565
	Doublet rate	Estimated at ~4 % based on cell capture (10x Genomics)
	Batch correction method for merging (incl. reasoning for batch correction)	Unnecessary due to replicates overlapping very well
Validation	Method of automatic annotation of clusters	n/a
	Method of manual annotation (markers, gene function info)	DEG analysis, GO enrichment analysis, PHYTOMap analysis of Marker Genes
	Verification in planta (e.g. Number of markers used for validation)	PHYTOMap analysis of ~24 Marker Genes across tissue types
Data availability	Analysis scripts & codes (GitHub)	n/a
	Excel Tables DEG for each cluster	See Datasets S1, S4, S7 & S8
	Objects/count matrix in repository (which one, where?)	NCBI project number PRJNA615235
	On-line tool/browser URL	https://www.zmbp-resources.uni-tuebingen.de/timmermans/plant-single-cell-browser/

Supplemental Table S5. NCBI SRA accessions for the *Wolffia* bulk and single cell RNA-seq datasets generated in this study.

Accession	**Title**
SRR28238825	RNA-seq of *Wolffia:* whole frond - Rep3
SRR28238826	RNA-seq of *Wolffia:* whole frond - Rep2
SRR28238827	RNA-seq of *Wolffia:* whole frond - Rep1
SRR28238828	RNA-seq of *Wolffia:* daughter frond - Rep3
SRR28238829	RNA-seq of *Wolffia:* daughter frond - Rep2
SRR28238817	RNA-seq of *Wolffia:* daughter frond - Rep1
SRR28238818	RNA-seq of *Wolffia:* mother frond - Rep3
SRR28238819	RNA-seq of *Wolffia:* mother frond - Rep2
SRR28238820	RNA-seq of *Wolffia:* mother frond - Rep1
SRR28238821	RNA-seq of *Wolffia:* below water protion - Rep3
SRR28238822	RNA-seq of *Wolffia:* below water protion - Rep2
SRR28238823	RNA-seq of *Wolffia:* below water protion - Rep1
SRR28238824	RNA-seq of *Wolffia:* above water protion - Rep3
SRR28238830	RNA-seq of *Wolffia:* above water protion - Rep2
SRR28238831	RNA-seq of *Wolffia:* above water protion - Rep1
SRR29417744	scRNA-seq of *Wolffia* at dusk - Rep1
SRR29417743	scRNA-seq of *Wolffia* at dusk - Rep2
SRR29417746	scRNA-seq of *Wolffia* at dawn - Rep1
SRR29417745	scRNA-seq of *Wolffia* at dawn - Rep2