Plant Communications, Volume *4*

Supplemental information

High-quality *Cymbidium mannii* genome and multifaceted regulation of crassulacean acid metabolism in epiphytes

Weishu Fan, Zheng-Shan He, Mengqing Zhe, Jing-Qiu Feng, Le Zhang, Yiwei Huang, Fang Liu, Jia-Lin Huang, Ji-Dong Ya, Shi-Bao Zhang, Jun-Bo Yang, Andan Zhu, and De-Zhu Li

Supplemental Information

High-quality *Cymbidium mannii* genome and multifaceted regulation of crassulacean acid metabolism in epiphytes

Weishu Fan[1#], Zheng-Shan He[1,2#], Mengqing Zhe[1,2], Jing-Qiu Feng[2,3], Le Zhang[1], Yiwei Huang[1,2], Fang Liu[1,2], Jia-Lin Huang[4], Ji-Dong Ya[1], Shi-Bao Zhang[3], Jun-Bo Yang[1]*, Andan Zhu[1]* and De-Zhu Li[1,2]*

[1] Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China

[2] Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan 650201, China

[3] Key Laboratory for Economic Plants and Biotechnology, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China

[4] Yuxi Normal University, Yuxi, Yunnan 653100, China

[#] These authors contribute equally

* **Corresponding authors:**

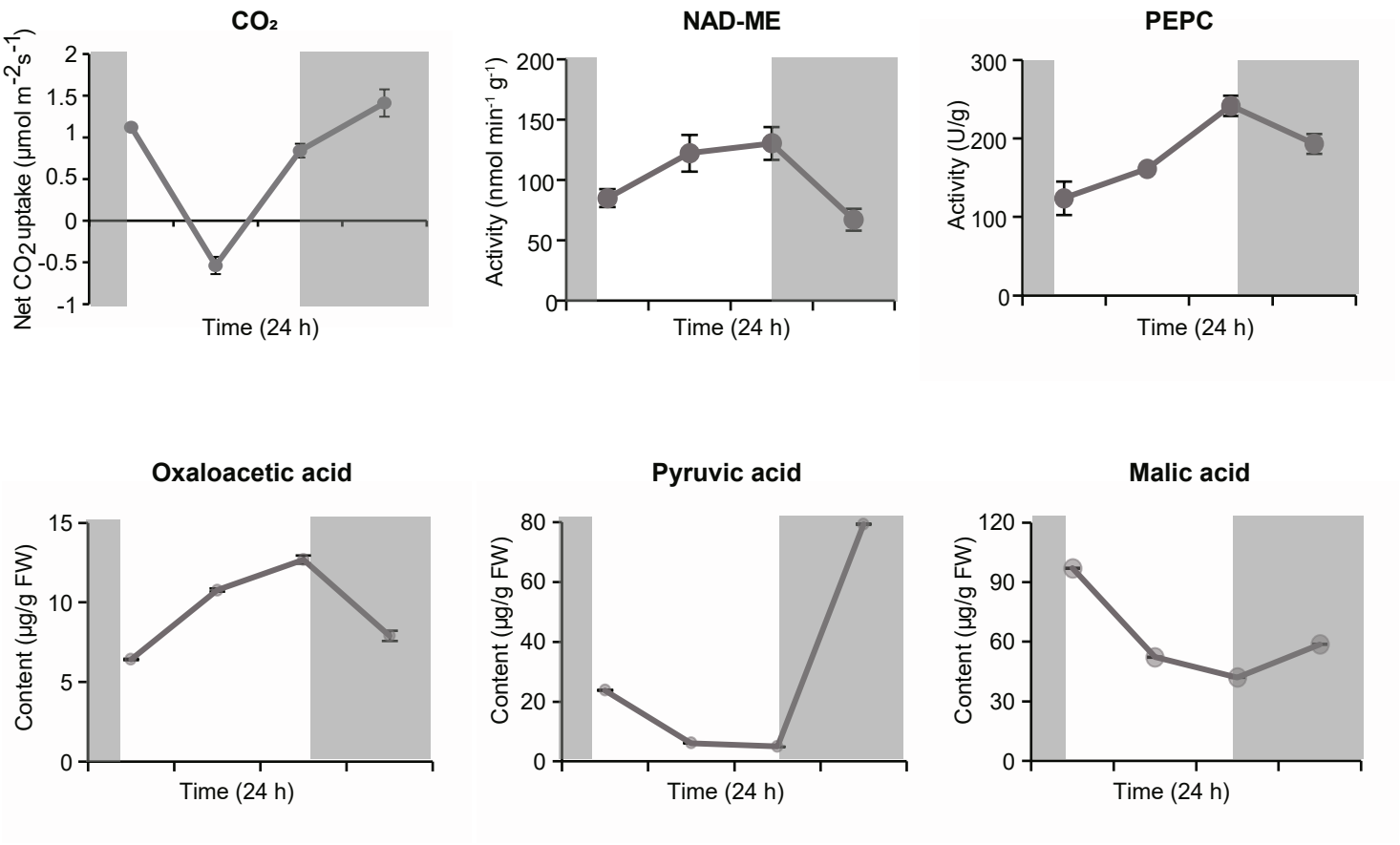
Jun-Bo Yang. Address: No. 132 Lanhei Rd, Heilongtan, Kunming, Yunnan 650201, China. Phone: +86 871 6522 3139. Email: jbyang@mail.kib.ac.cn

Andan Zhu. Address: No. 132 Lanhei Rd, Heilongtan, Kunming, Yunnan 650201, China. Phone: +86 871 6523 8370. Email: zhuandan@mail.kib.ac.cn

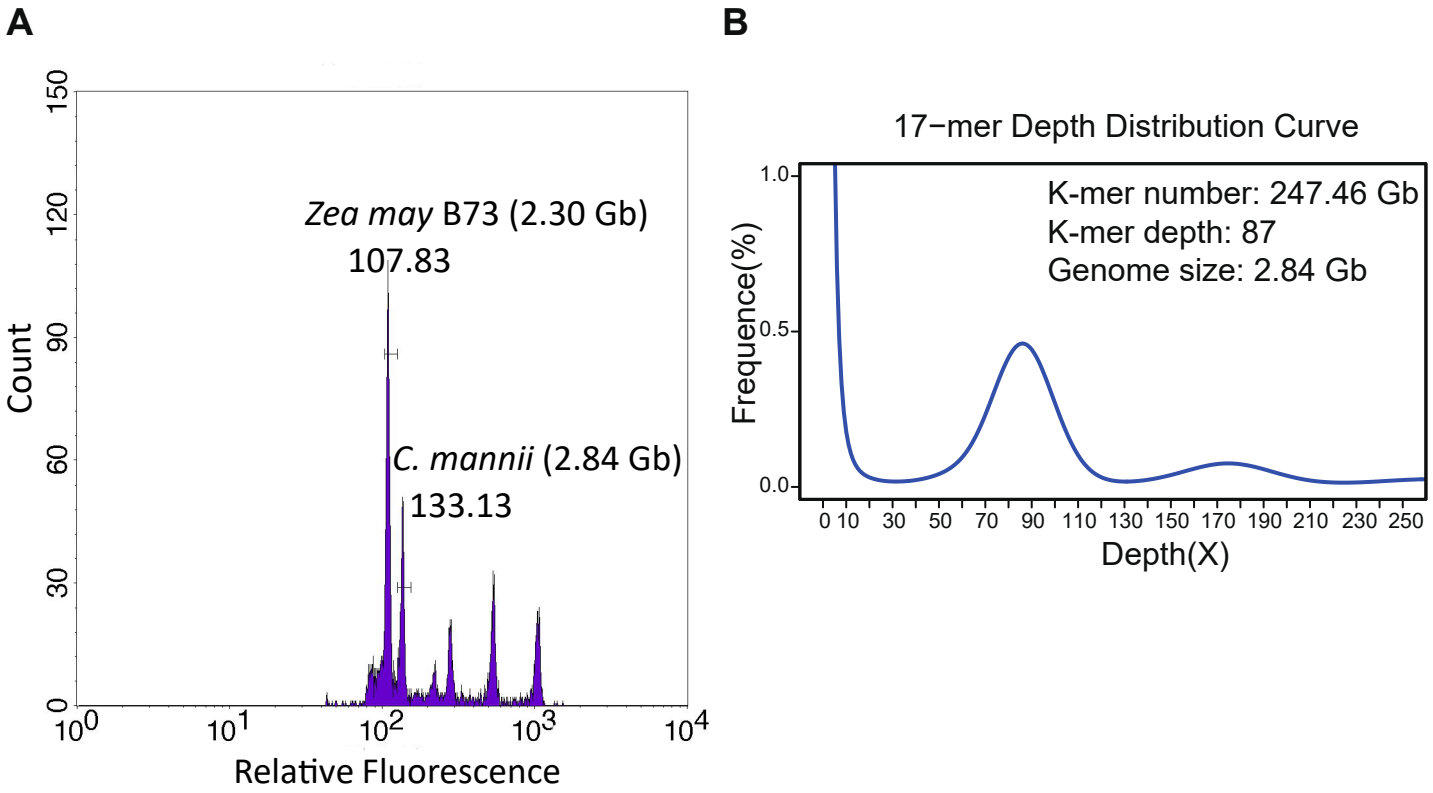
De-Zhu Li. Address: No. 132 Lanhei Rd, Heilongtan, Kunming, Yunnan 650201, China. Phone: +86 871 6522 3503. Email: dzl@mail.kib.ac.cn

Short Summary

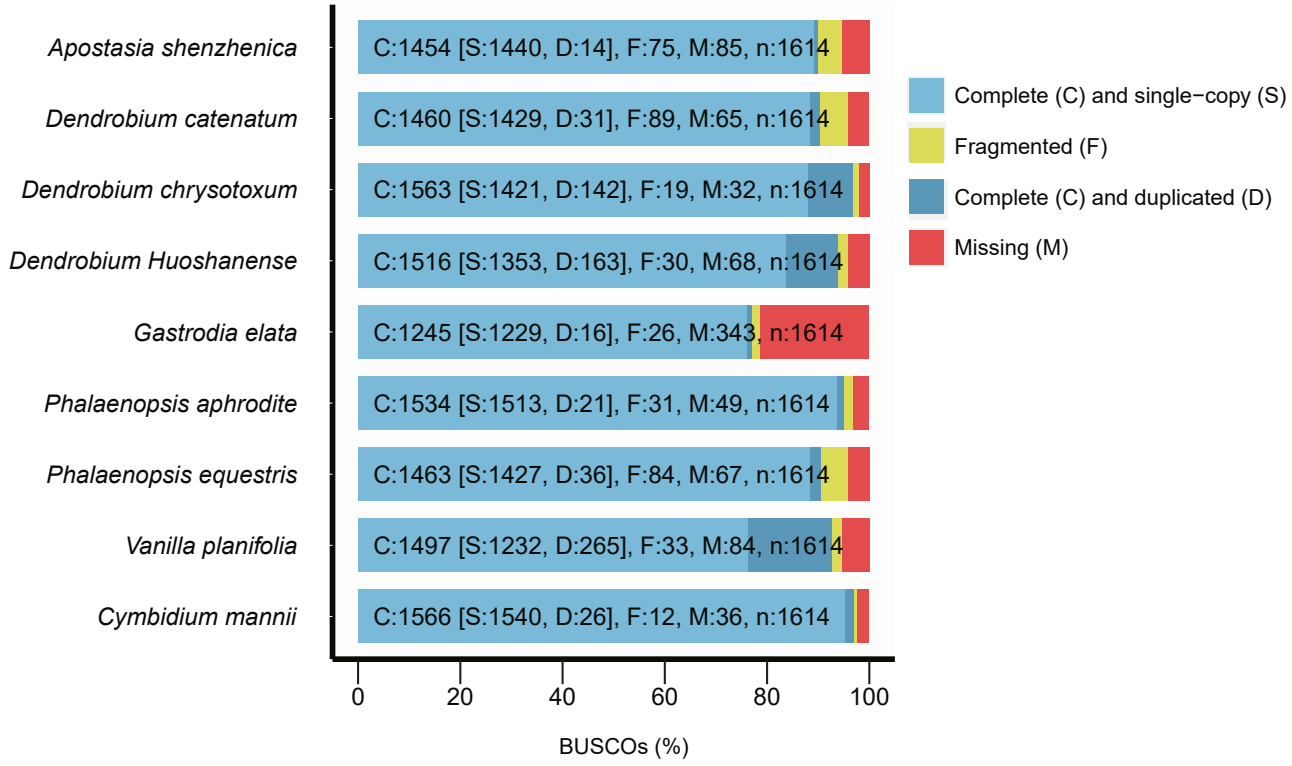
We report a high-quality chromosomal-level genome assembly of a CAM epiphyte, *Cymbidium mannii* (Orchidaceae). Our high-resolution transcriptomics, proteomics, and metabolomics data across a CAM diel cycle reflect circadian rhythmicity in metabolite accumulation in epiphytes. Genome-wide analysis of transcript and protein level regulation revealed phase shifts during the multifaceted regulation of circadian metabolism.



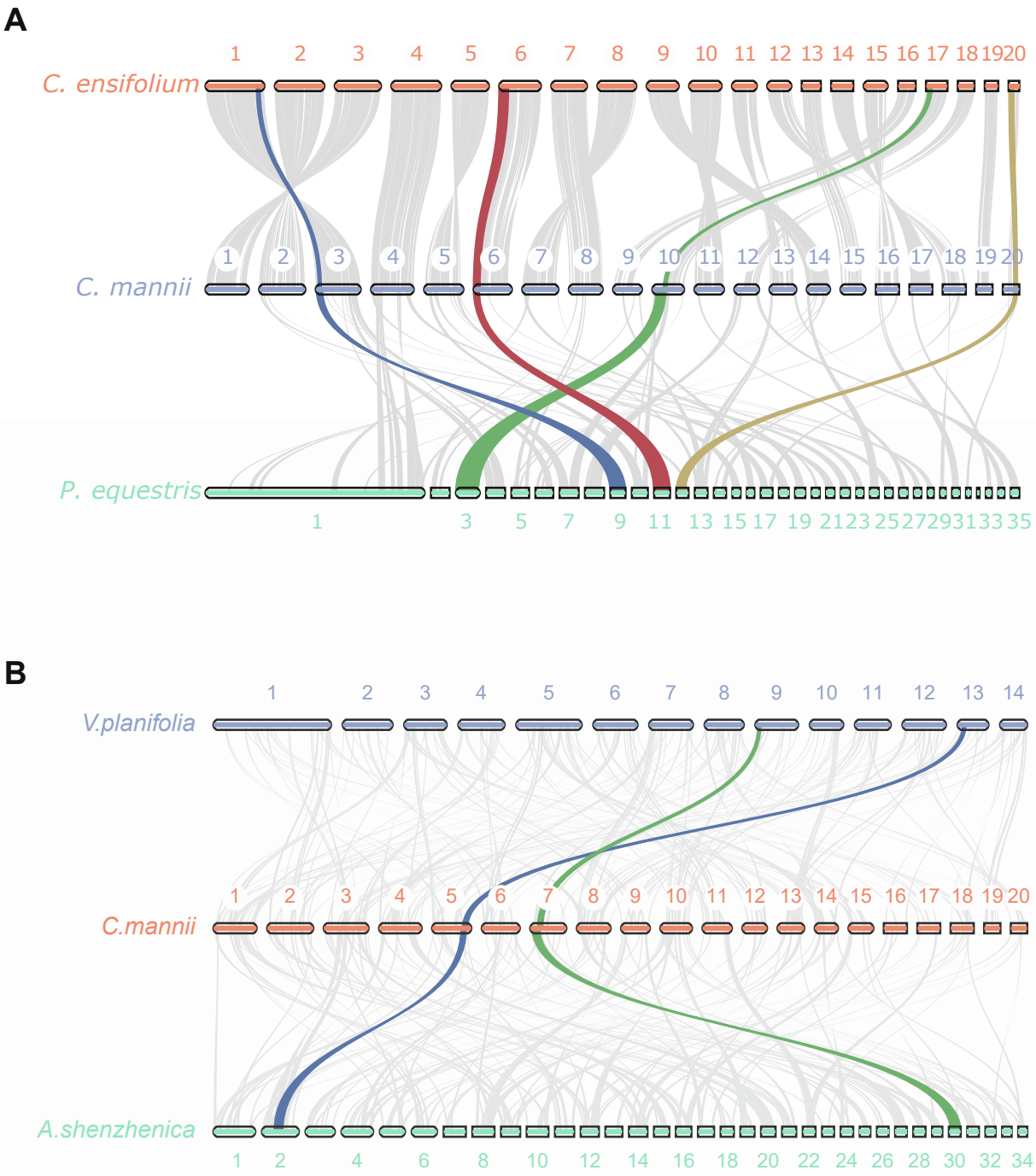
Supplementary Fig. S1. Physiological characteristics of the CAM plant *Cymbidium mannii*. The rate of net CO_2 uptake, the enzyme activities of NAD malic enzyme (NAD-ME) and PEP carboxylase (PPC), and the levels of oxaloacetic acid, pyruvic acid, and malic acid were measured over a 24-h cycle. Standard errors of replicates are indicated by error bars.



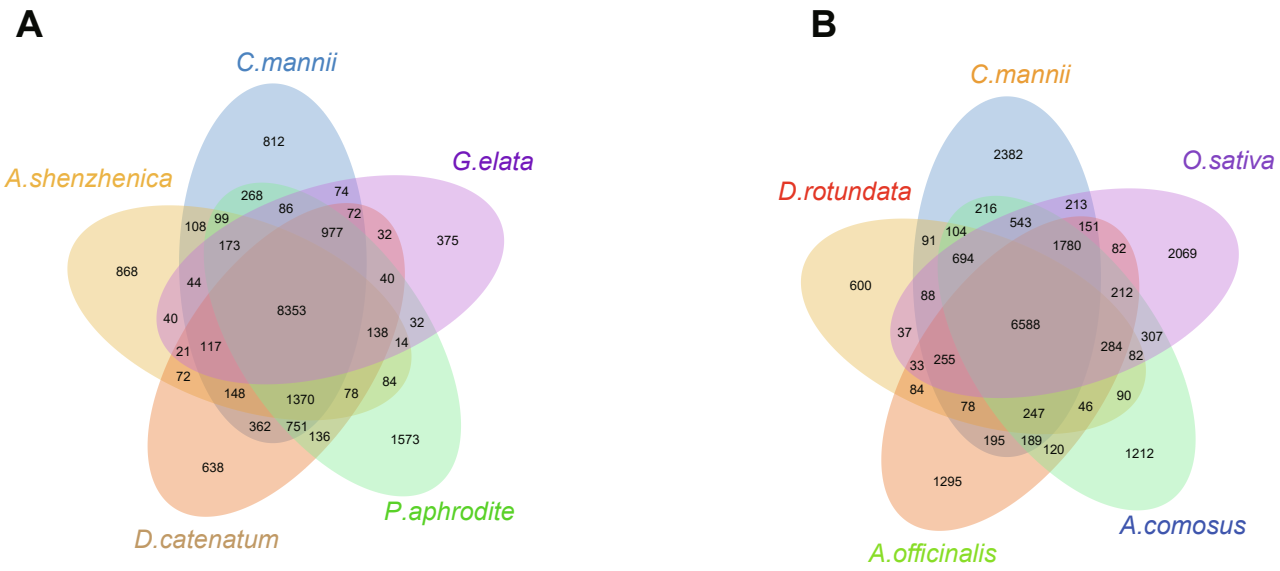
Supplementary Fig. S2. Flow cytometry (A) and *k*-mer frequency distribution analyses (B) of the genome size.



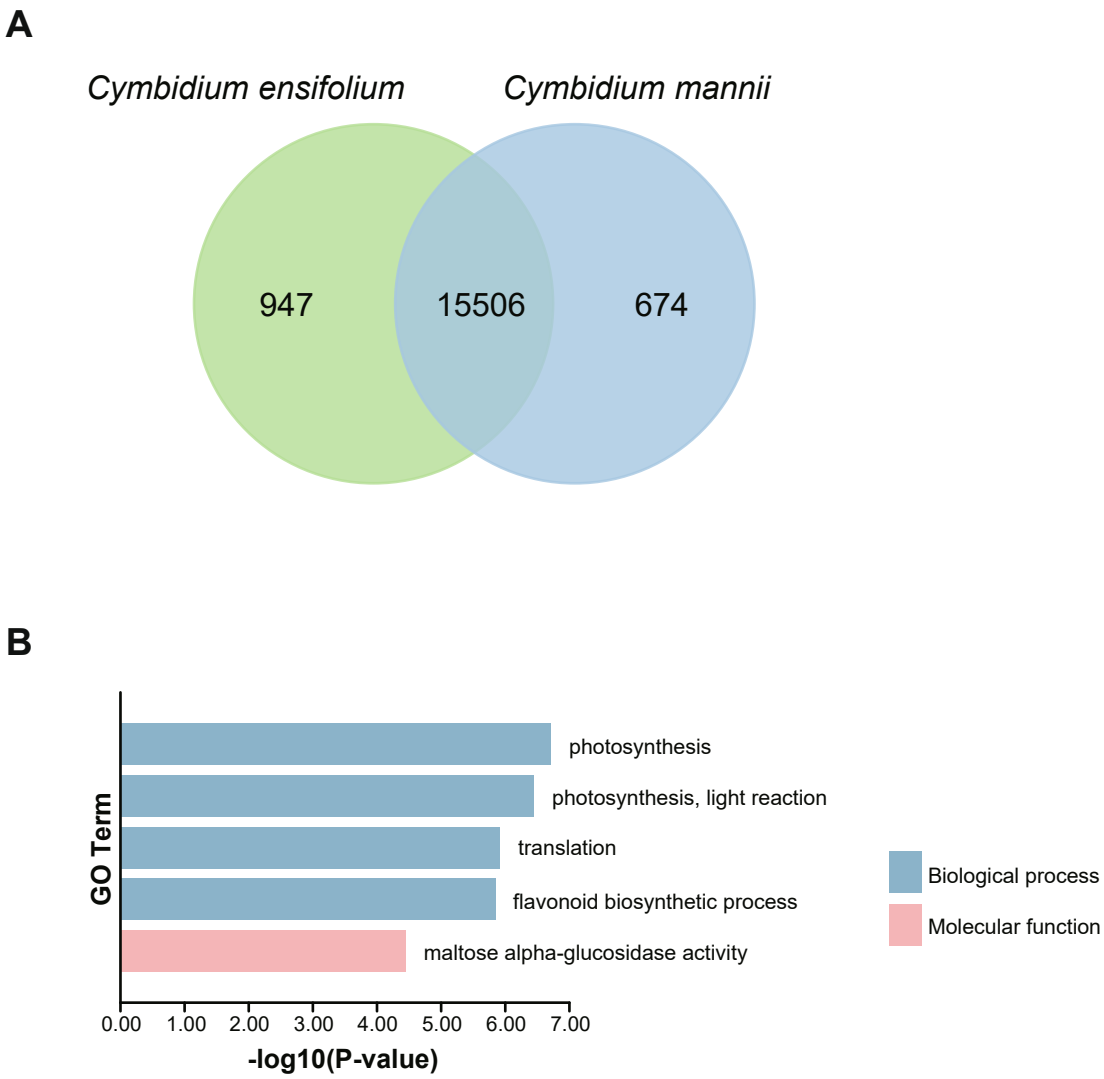
Supplementary Fig. S3. BUSCO assessment of the genome assembly with orthologous groups.



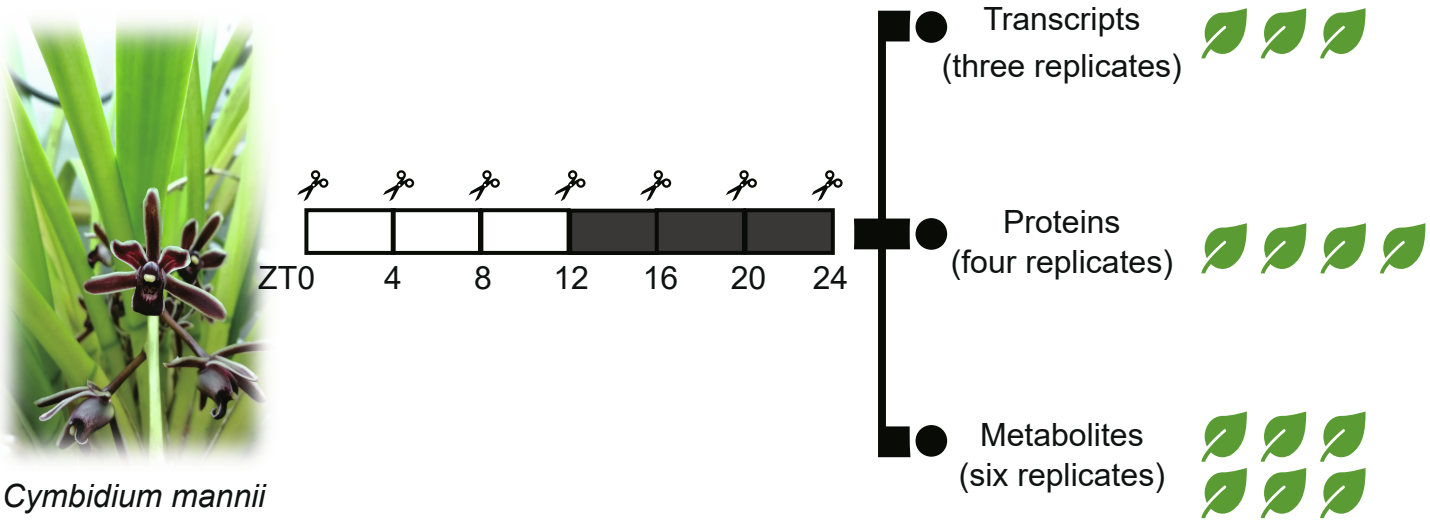
Supplementary Fig. S4. Intergenomic synteny between *C. mannii* (20 chromosomes), *C. ensifolium* (20 chromosomes), and *Phalaenopsis equestris* (35 scaffolds longer than 2 Mb) (A); and between *C. mannii*, *Vanilla planifolia* (14 chromosomes), and *Apostasia shenzhenica* (34 longest scaffolds) (B). Each syntenic line represents at least four adjacent anchor pairs. Highlighted lines represent examples of one copy of syntenic genes in all three orchid species.



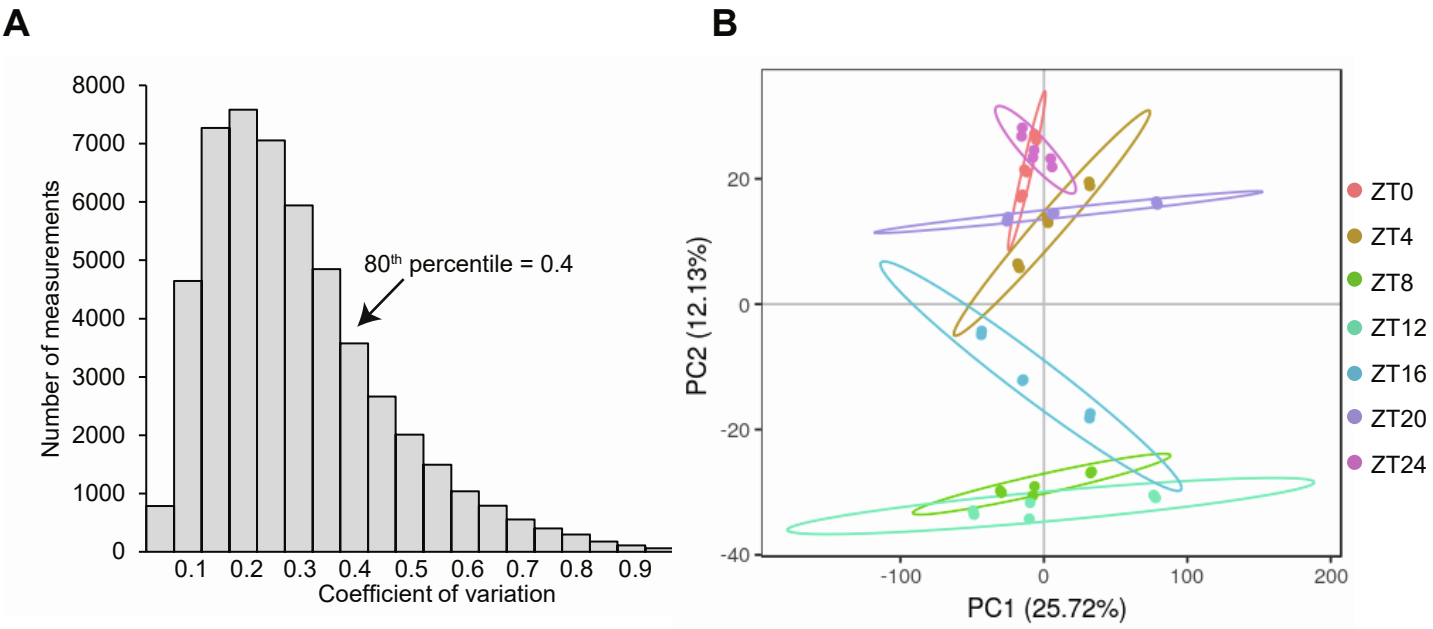
Supplementary Fig. S5. OrthoMCL clustering analysis of orthologous groups from Orchidaceae species **(A)** and the shared and unique gene families among *C. mannii* and four representative monocotyledons (*Ananas comosus*, *Asparagus officinalis*, *Dioscorea rotundata*, and *Oryza sativa*) **(B)**. Each number within the Venn diagram represents the number of orthologous groups.



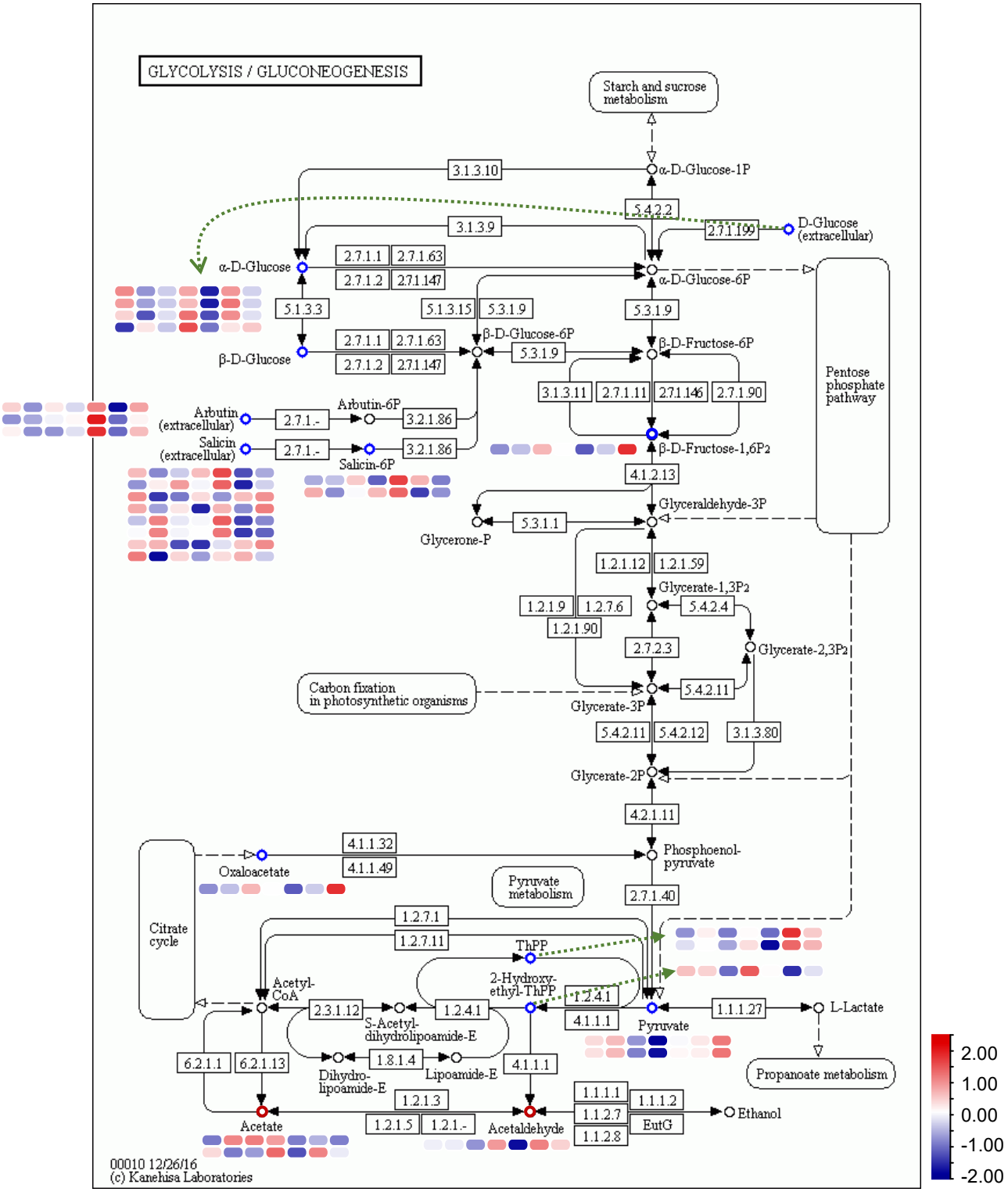
Supplementary Fig. S6. Orthologous groups comparison between *C. mannii* and *C. ensifolium*. (A) Shared and unique gene families between *C. mannii* and *C. ensifolium*. **(B)** GO enrichment of the shared gene families.



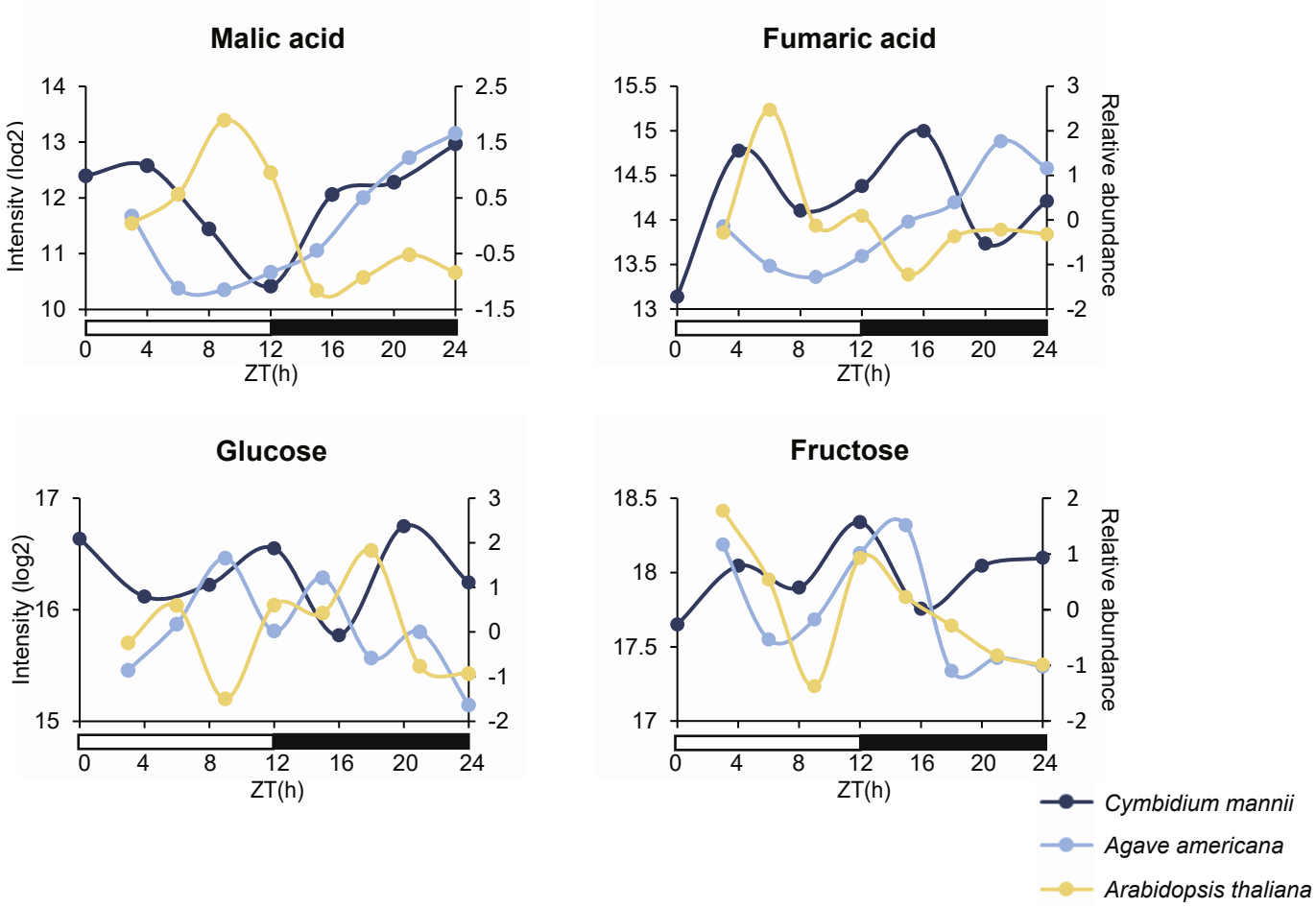
Supplementary Fig. S7. Sampling details and experimental design. Illustration of sample collection time points and replicates in a diel cycle.



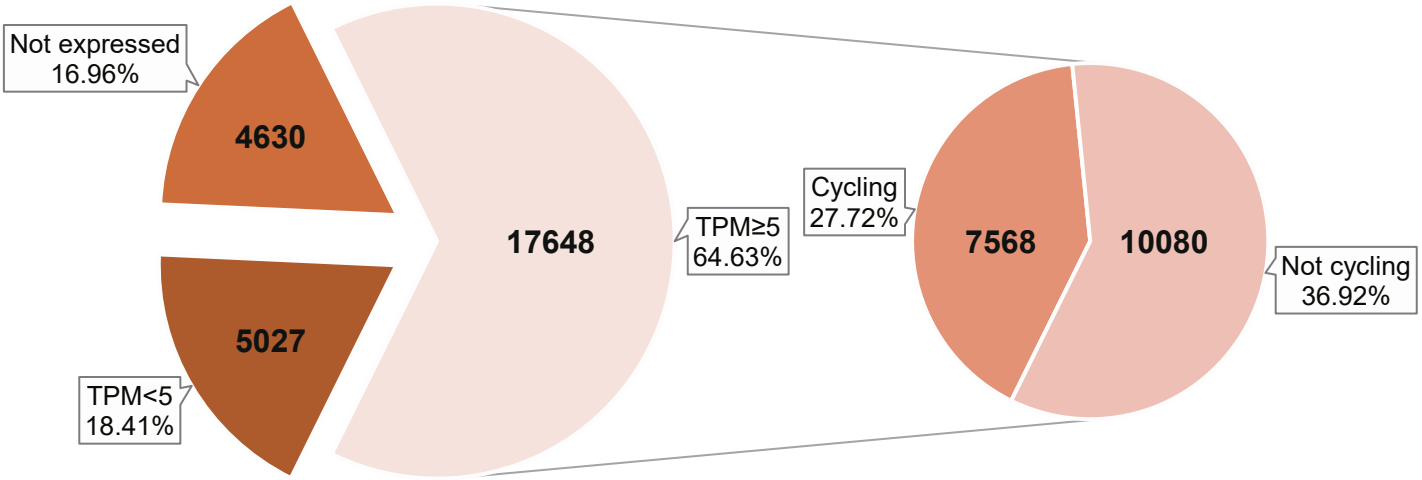
Supplementary Fig. S8. Quality assessment of the *C. mannii* metabolome. (A) Distribution summary of coefficient of variation values from each metabolite among replicates. **(B)** Principal component analysis of metabolomes at different time points.



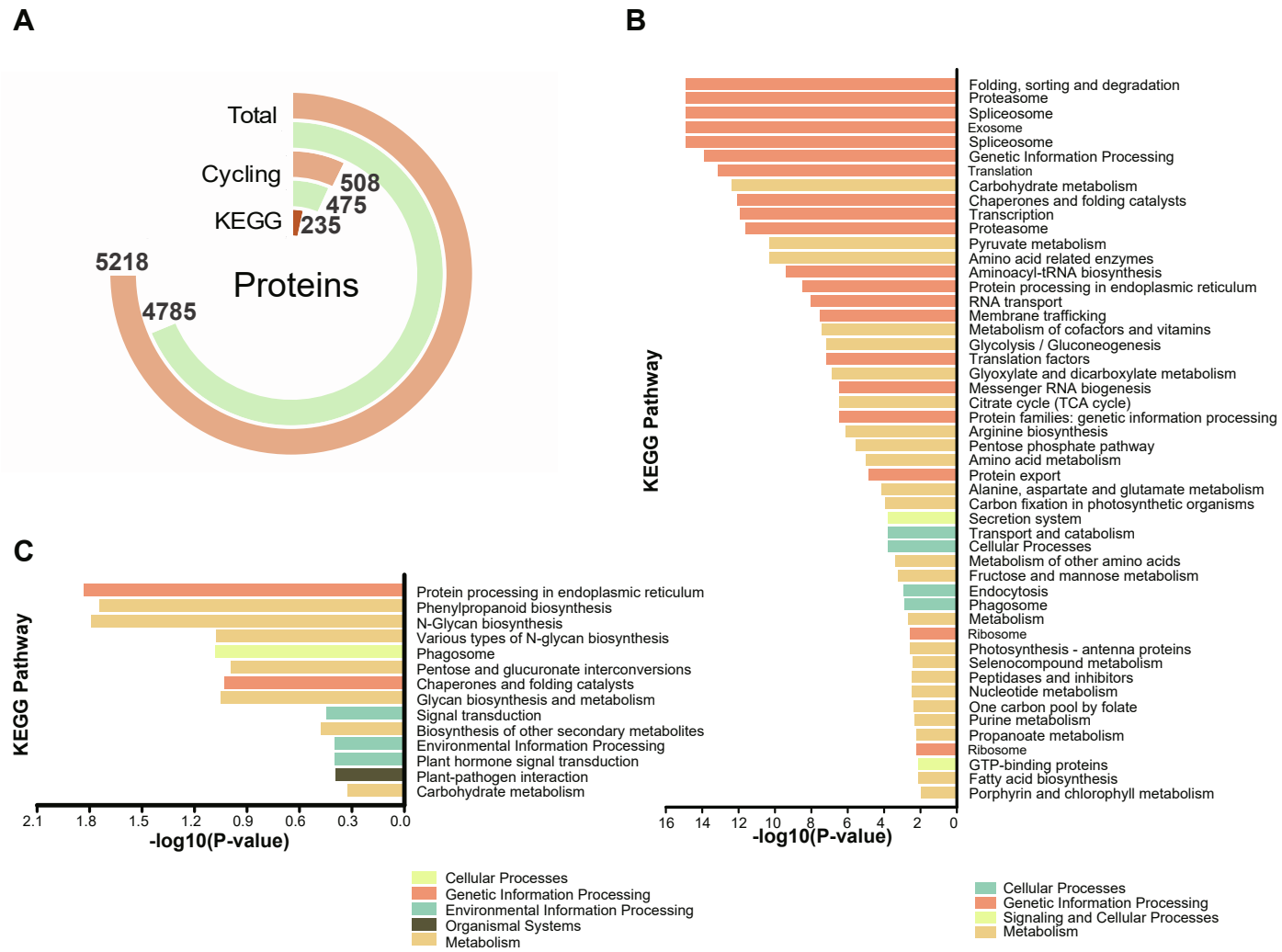
Supplementary Fig. S9. Relative intensity of the identified metabolites in the glycolysis/gluconeogenesis pathway. The KEGG map of the glycolysis/gluconeogenesis (map00010) pathway was downloaded from the KEGG website (https://www.genome.jp/kegg-bin/show_pathway?map00010). Blue circles represent the identified primary metabolites, and red circles represent secondary metabolites. The relative intensity heatmap is shown beside the corresponding genes, and multiple heatmaps show all possible primary metabolites. Each heatmap shows the normalized intensity at ZT0, ZT4, ZT8, ZT12, ZT16, ZT20, and ZT24 (from left to right; also see supplementary Table S10).



Supplementary Fig. S10. Comparison of abundance profiles of the selected metabolites of malic acid, fumaric acid, glucose, and fructose between CAM (terrestrial *Agave americana* and epiphytic *C. mannii*) and C_3 (*Arabidopsis thaliana*) representatives.

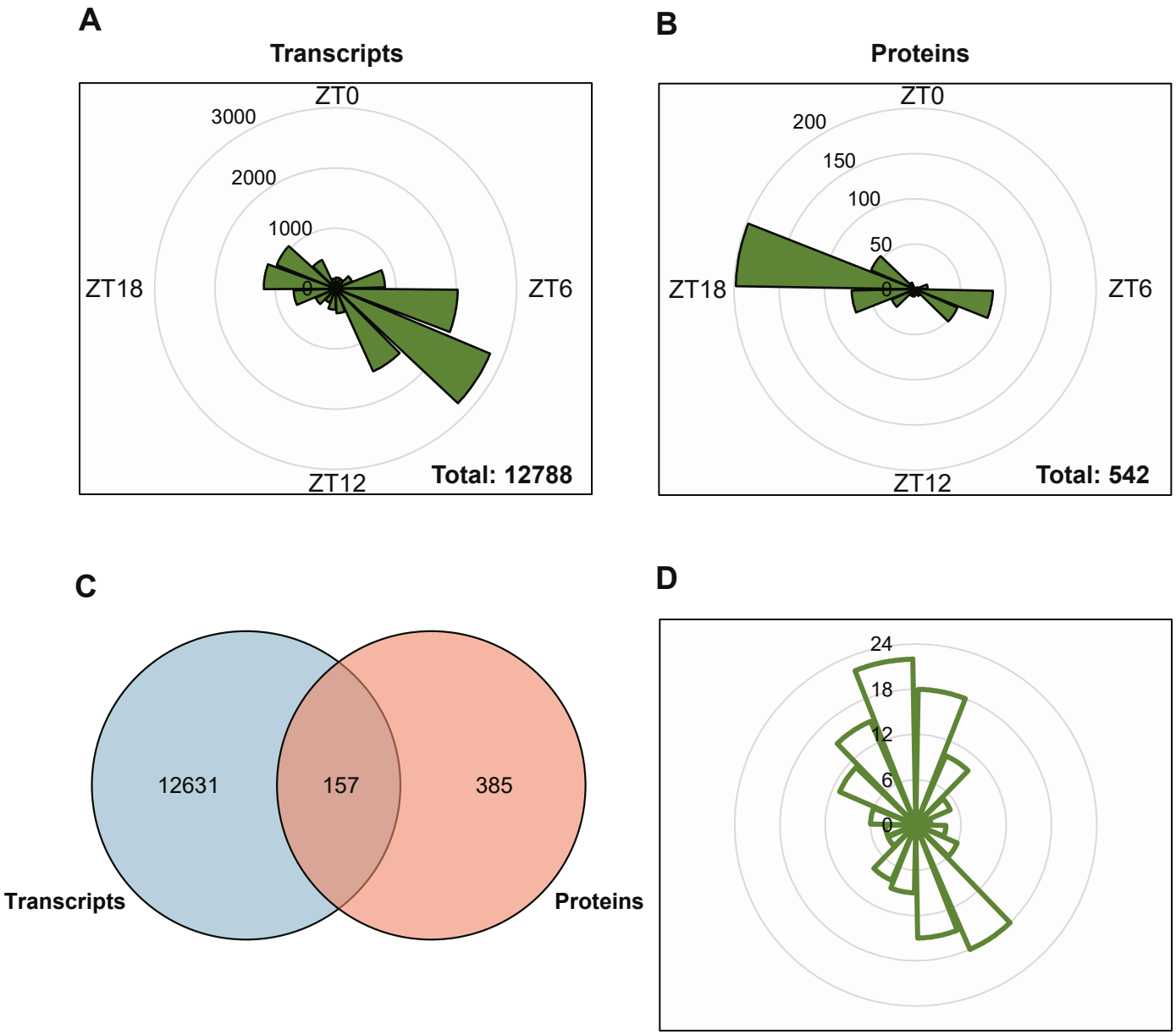


Supplementary Fig. S11. Expressed and cycling genes in *C. mannii*. The total number of genes is equal to the total annotated genes in *C. mannii*. Genes with a zero TPM value in all seven time points were considered “not expressed”. Genes with the total TPM of all time points lower than 5 were considered “low expressed”, and these genes were also excluded from the cycling analysis. The cycling genes were predicted by JTK_CYCLE with an FDR <0.05 and a period of 24 h.

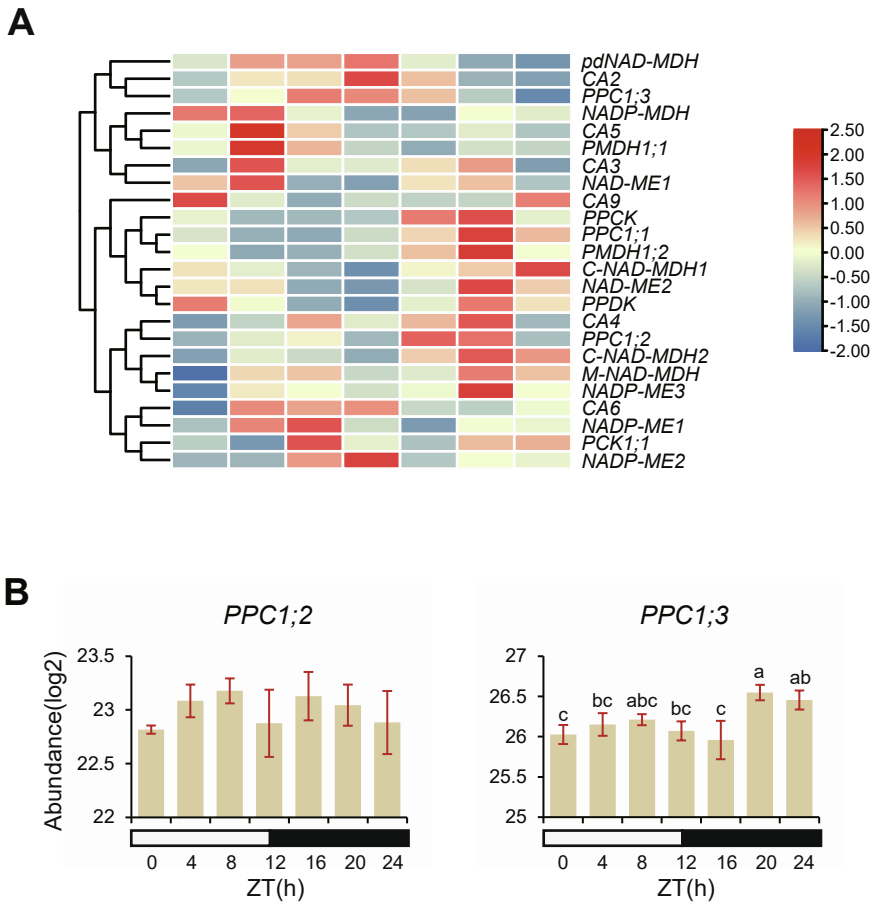


Supplementary Fig. S12. DIA identified proteins and their KEGG annotations.

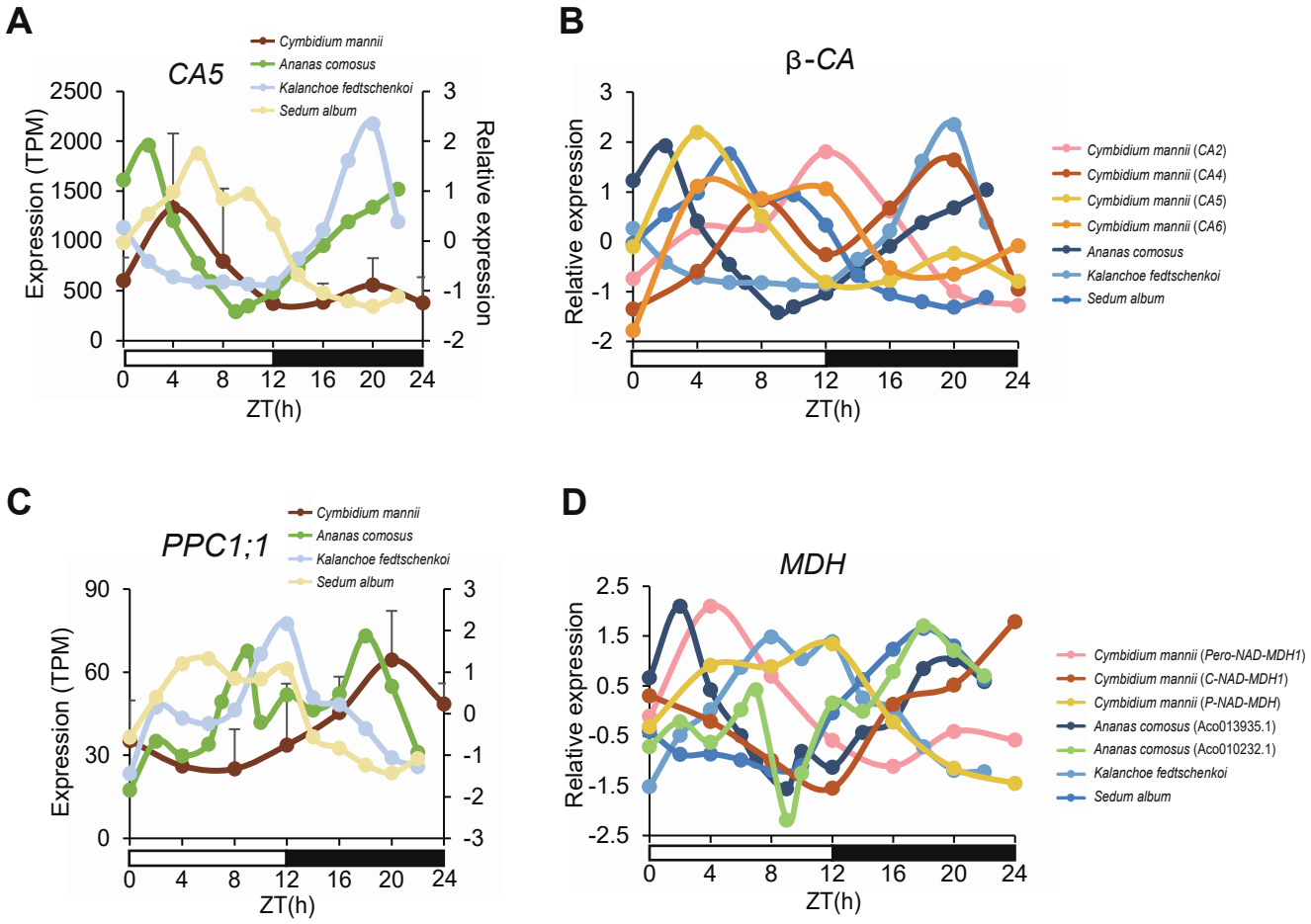
(A) Identified total and cycling proteins in *C.mannii*. Identified protein numbers are in orange and annotated gene IDs are in green. **(B)** KEGG enrichment analysis (corrected $p < 0.05$) of all proteins. The top 50 categories are shown in the bar chart. **(C)** KEGG enrichment analysis ($p < 0.05$) of the cycling proteins.



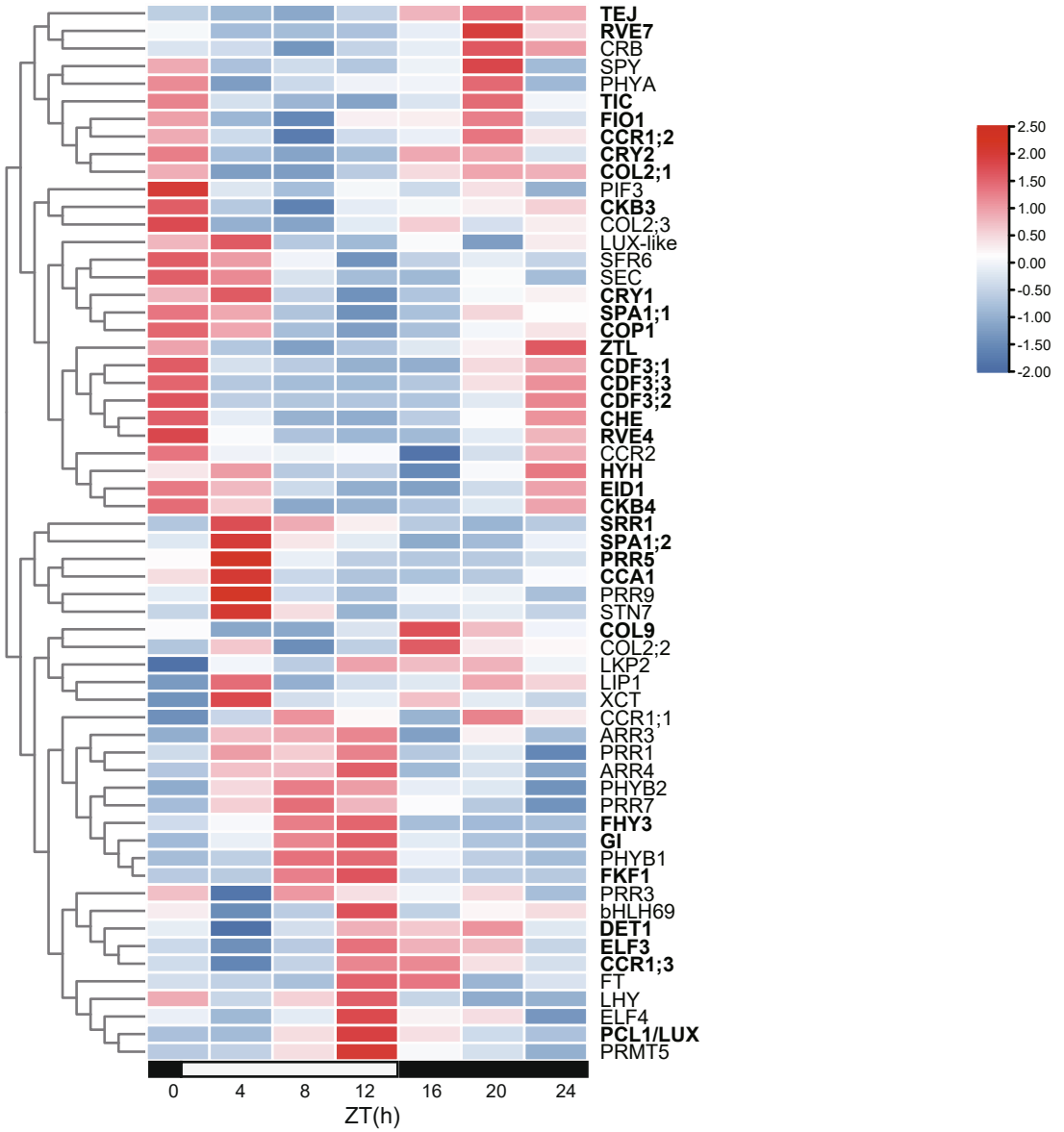
Supplementary Fig. S13. Phase shift between proteomic and expression profiles in *Agave*. Distribution of peak time in rhythmic transcripts **(A)** and proteins **(B)** across the diel cycle. **(C)** Unique and shared genes between cycling transcripts and proteins. **(D)** Phase shift time between shared transcripts and proteins. The circular scales show the shifted hours.



Supplementary Fig. S14. Core CAM gene expression and proteome abundance of *PPC* in *C. mannii*. (A) Transcript expression of all core CAM genes. Low expressed (TPM <5) genes are not shown. **(B)** Protein abundance of the other *PPC* copies in *C. mannii*. Significant differences in protein abundance were compared and marked with letters on the top. CA, carbonic anhydrase; NAD-MDH, NAD-malate dehydrogenase; PPC, phosphoenolpyruvate carboxylase; PPCK, phosphoenolpyruvate carboxylase kinase; PPDK, pyruvate orthophosphate dikinase.

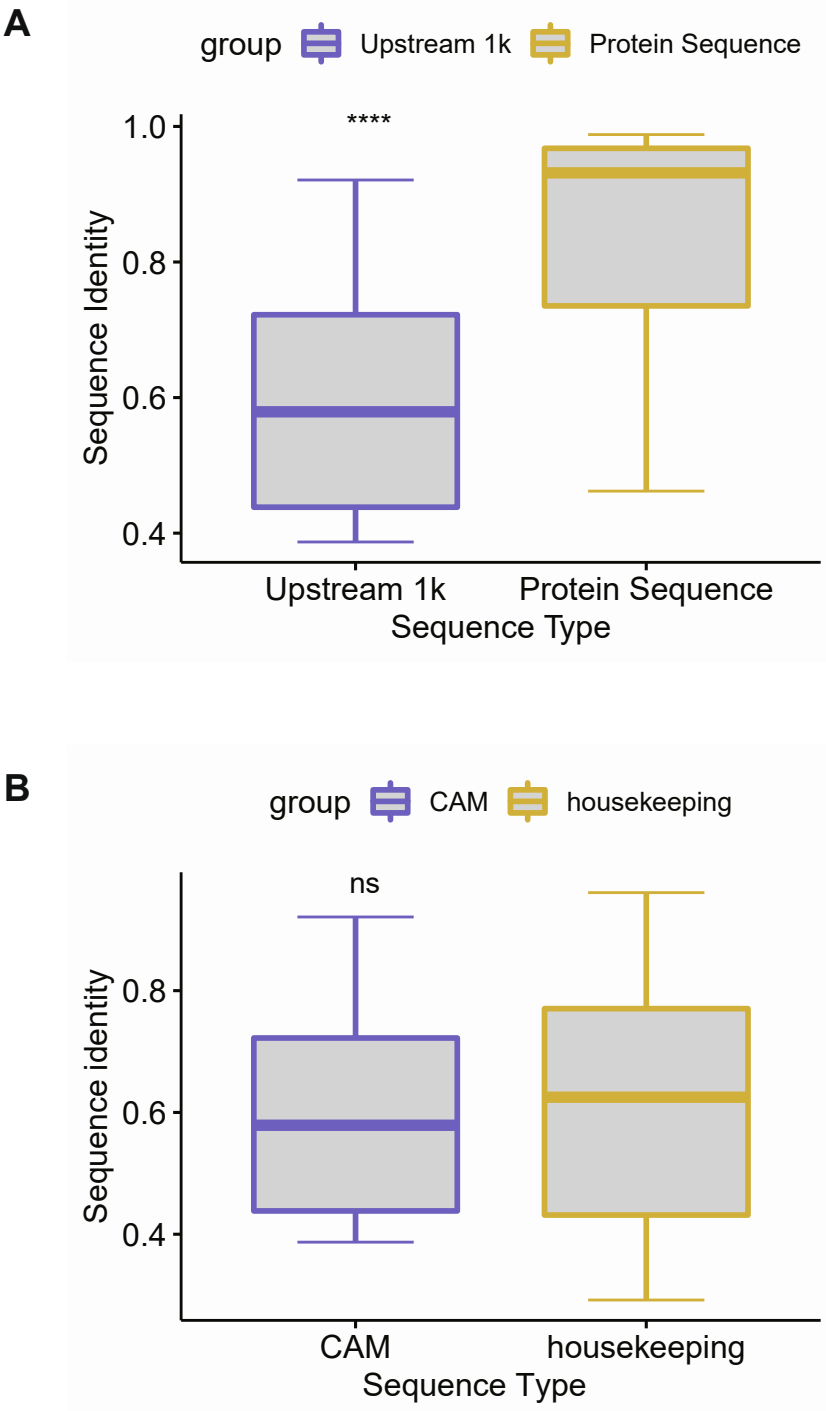


Supplementary Fig. S15. Comparison of the other copies of the selected core CAM genes.

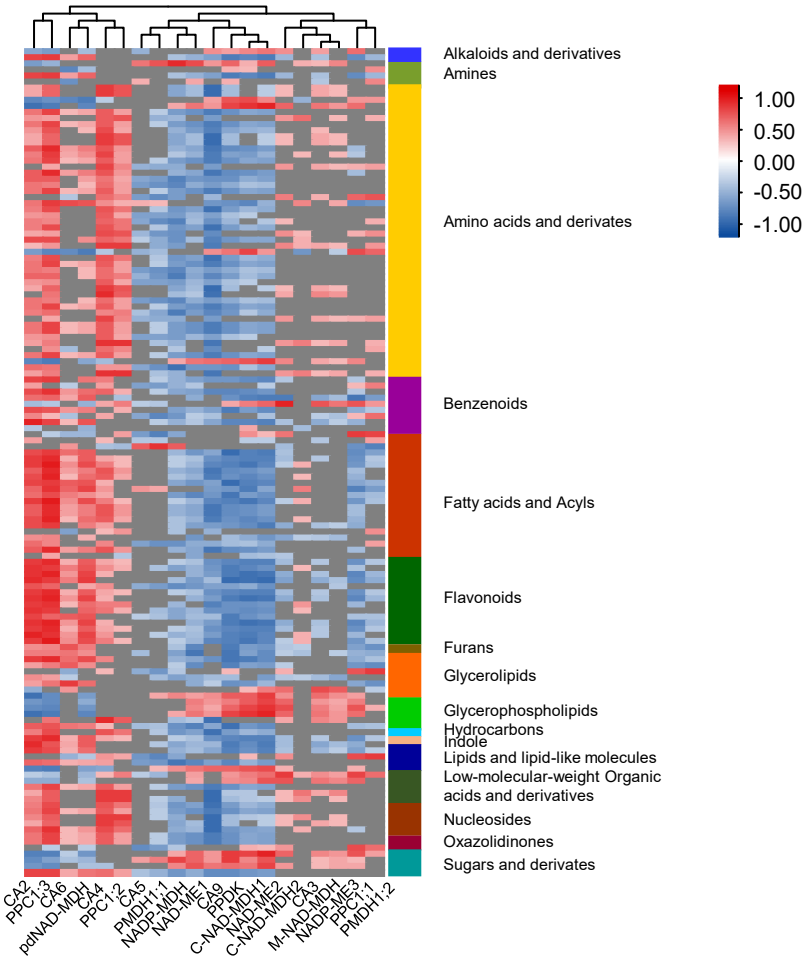


Supplementary Fig. S16. Circadian clock gene expression in *C. mannii*.

Heatmap showing the average TPM expression from the circadian clock-associated genes across the diel cycle. Bold gene names represent cycling genes.



Supplementary Fig. S17. Sequence identity comparisons between *C. mannii* and *C. ensifolium*. (A) Core CAM genes protein sequences and their upstream 1 kb sequence identity comparisons. **(B)** 1 kb upstream sequence identities between core CAM genes and selected housekeeping genes (*actin*, *beta-tubulin*, *ubiquitin*, *glucose-6-phosphate 1-dehydrogenase* and *phosphoglycerate kinase 1*).



Supplementary Fig. S18. Correlation analysis between selected core CAM genes and cycling metabolites. Heatmap of Spearman correlation between mRNA levels (TPM) and metabolite intensities. Significant correlation values ($p < 0.05$) are shown in the heatmap, and nonsignificant values are shown in gray.