Altered Gene Regulatory Networks are Associated with the Transition from C_3 to Crassulacean Acid Metabolism in *Erycina* (Oncidiinae: Orchidaceae)

Karolina Heyduk (karohey@gmail.com)[1*], Michelle Hwang (Mxhwang@gmail.com)[1], Victor A. Albert (vaalbert@buffalo.edu)[2,3], Katia Silvera (katia.silvera@ucr.edu)[4, 5], Tianying Lan (tianying@buffalo.edu)[2], Kimberly M. Farr (kmfarr@buffalo.edu)2, Tien-Hao Chang (cardiandra@gmail.com)[2], Ming-Tsair Chan (mbmtchan@gate.sinica.edu.t)[6], Klaus Winter (winterk@si.edu) [5], and Jim Leebens-Mack (jleebensmack@uga.edu)[1]

1 – Department of Plant Biology, University of Georgia, Athens, GA, USA
2 – Department of Biological Sciences, University at Buffalo, Buffalo, NY, USA
3 – School of Biological Sciences, Nanyang Technological University, Singapore
4 – Department of Botany and Plant Sciences, University of California-Riverside, Riverside, CA, USA
5 – Smithsonian Tropical Research Institute, Panama City, Panama
6 – Agricultural Biotechnology Research Center, Academia Sinica, Nankang, Taipei, Taiwan, ROC

Correspondence:
Dr. Karolina Heyduk
karohey@gmail.com

Abstract

Crassulacean acid metabolism (CAM) photosynthesis is a modification of the core C_3 photosynthetic pathway that improves the ability of plants to assimilate carbon in water-limited environments. CAM plants fix CO_2 mostly at night, when transpiration rates are low. All of the CAM pathway genes exist in ancestral C_3 species, but the timing and magnitude of expression are greatly altered between C_3 and CAM species. Understanding these regulatory changes is key to elucidating the mechanism by which CAM evolved from C_3. Here we use two closely related species in the Orchidaceae, *Erycina pusilla* (CAM) and *Erycina crista-galli* (C_3), to conduct comparative transcriptomic analyses across multiple time points. Clustering of genes with expression variation across the diel cycle revealed some canonical CAM pathway genes similarly expressed in both species, regardless of photosynthetic pathway. However, gene network construction indicated that 149 gene families had significant differences in network connectivity and were further explored for these functional enrichments. Genes involved in light sensing and ABA signaling were some of the most differently connected genes between the C_3 and CAM *Erycina* species, in agreement with the contrasting diel patterns of stomatal conductance in C_3 and CAM plants. Our results suggest changes to transcriptional cascades are important for the transition from C_3 to CAM photosynthesis in *Erycina*.

Keywords: RNA-seq, transcriptomics, photosynthesis, time-course, gene network

Introduction

Crassulacean acid metabolism (CAM) is a carbon concentrating mechanism that evolved multiple times in response to CO_2 limitation caused by water stress. In C_3 species, stomata remain open during the day to assimilate atmospheric CO_2, but water limitation can force stomata to close, resulting in impaired CO_2 fixation at the expense of growth. When water stress is prolonged, stomatal closure in C_3 plants can become debilitating. CAM species circumvent prolonged stomatal closure by opening stomata at night and fix CO_2 nocturnally, when evapotranspiration rates are on average lower. CO_2 is temporarily stored as malic acid in the vacuoles until day time, when stomata close and malic acid is moved back into the cytosol for decarboxylation. The resulting increase of CO_2 levels near ribuslose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) results in highly efficient CO_2 reduction via C_3 photosynthesis CAM is associated with a number of anatomical, physiological and genetic change, including alterations to leaf anatomy (Nelson and Sage, 2008; Zambrano et al., 2014), stomatal opening at night, and tight regulation of metabolic genes within day/night cycles. Despite the complexity of these evolutionary novelties, CAM plants are found in a wide range of plant families, including eudicot species in the Euphorbiaceae (Horn et al., 2014) and Caryophyllales (Guralnick et al., 1984; Moore et al., 2017; Winter and Holtum, 2011) and monocot lineages in Agavoideae (Abraham et al., 2016; Heyduk et al., 2016), Orchidaceae (Silvera et al., 2009, 2010), and Bromeliaceae (Crayn et al., 2004).

The CAM pathway is well-described biochemically (Holtum et al., 2005) and contemporary genomics approaches are beginning to shed light on the genetic basis of CAM (Abraham et al., 2016; Cushman et al., 2008; Dever et al., 2015) (Fig. 1). As CO_2 enters the chloroplast-containing cells as night, it is initially converted to HCO_3^- facilitated by a carbonic anhydrase (CA). HCO_3^- is then fixed by phosphoenolpyruvate carboxylase (PEPC) using phosphoenolpyruvate (PEP) as the substrate. Carboxylation of PEP results in oxaloacetate (OAA), which is subsequently converted to malic acid by malate dehydrogenase (MDH). Malic acid is then moved into the vacuole for storage. The vacuolar transporter of malic acid is not known for certain, although previous studies have pointed to aluminum-activated malate transporters (ALMT) as a candidate (Kovermann et al., 2007; Yang et al., 2017). During the day, the malic acid is released from the vacuoles either via a passive process or through as-yet undescribed transporter. The malic acid is then decarboxylated to CO_2 and PEP using two decarboxylation pathways: NAD and/or NADP malic enzymes together with pyruvate, phosphate dikinase (PPDK), or MDH and phosphoenolpyruvate carboxykinase (PEPCK)

While the roles of canonical CAM enzymes are considered novel in CAM species, they are all present in C_3 ancestral species as well. As a result, the evolution of CAM likely involved alterations to gene copies, including changes to protein sequences and regulatory motifs. For example, PEPC in C_3 species can play several roles depending on tissue type and developmental stage, including providing carbon backbones to the citric acid cycle and providing malate for cellular pH balance (Aubry et al., 2011; Winter et al., 2015), but its core function as a carboxylating enzyme remains unchanged in C_3 and CAM species. Studies of molecular evolution of PEPC in CAM species have largely determined that there are CAM-specific copies of the enzyme that are differentially expressed in C_3 and CAM species (Gehrig et al., 2001; Lepiniec et al., 1994; Ming et al., 2015; Silvera et al., 2014). In some cases, these CAM-specific PEPC gene copies have been shown to share sequence similarity across closely related but

independently derived CAM taxa (Christin et al., 2014; Silvera et al., 2014). Additionally, genomic screens across many canonical CAM gene promoters have revealed an enrichment of circadian clock motifs (Ming et al., 2015), implicating alterations to transcription factor binding sites during CAM evolution. Because of the strong influence of the internal circadian clock on CAM (Hartwell, 2005), we might expect that genes involved in the CAM regulatory pathway should be controlled by a co-expressed circadian master regulator (Borland et al., 1999; Nimmo, 2000; Taybi et al., 2000b).

Elucidation of regulatory changes requires comparative analysis between closely related C_3 and CAM species, and this can be accomplished through RNA-Seq analyses. One of the largest plant families with multiple origins of CAM is the Orchidaceae; known for floral diversity and inhabiting a broad range of habitats, the evolution of CAM in predominantly epiphytic lineages may have also contributed to orchid diversity (Silvera et al., 2009). Epiphytic species constitute more than 70% of the Orchidaceae (Chase et al., 2015; Gravendeel et al., 2004), and many exhibit different degrees of CAM (Silvera et al., 2005). In a large proportion of these, CAM is weakly expressed relative to C_3. Weakly expressed CAM may represent an evolutionary end point, or may be an important intermediate step on the evolutionary path between constitutive C_3 and constitutive CAM. Because many genera within the Orchidaceae include both C_3 and weak/strong CAM species, the orchids are an attractive family to study the evolution of CAM photosynthesis. The subtribe Oncidiinae is one of the most diverse subtribes within Orchidaceae and it is part of a large epiphytic subfamily (Epidendroideae) in which CAM may have facilitated the expansion into the epiphytic habitat (Silvera et al., 2009). Despite the prevalence of CAM within the subtribe, the genus *Erycina* is particularly interesting because it has both CAM and C_3 species. *Erycina pusilla* is a fast-growing CAM species with transformation capability and has the potential to be a model species for studying CAM photosynthesis in monocots (Lee et al., 2015). Comparative investigations of *E. pusilla* and its C_3 relative, *E. crista-galli*, can therefore offer valuable insight into studying the evolution and regulation of CAM photosynthesis in the Orchidaceae. Through comparative, time-course RNA-Seq analysis of *E. pusilla* and *E. crista-galli*, we aim to understand 1) the changes in expression of core CAM genes between C_3 and CAM *Erycina* species and 2) which regulatory changes are required for the evolution of CAM.

Materials and Methods

Plant growth and RNA-Seq tissue collection

Erycina pusilla (L.) N.H.Williams & M.W.Chase (CAM) seedlings were cultivated on solid PSYP medium comprising 2 g/L Hyponex No. 1, 2 g/L tryptone, 20 g/L sucrose, 0.1 g/L citric acid, and 1 g/L active charcoal in flasks. The pH of the medium is adjusted to 5.4 before autoclaving and gelling with 3 g/L Phytagel. Plants were grown in 12-hour day and 12-hour night conditions over three independent dates in a growth chamber at the University at Buffalo, with temperatures set to 22-25C and lights on at 6 am for a 12 hour photoperiod. Light intensity was between 95-110 μmol m^{-2} s^{-1}. Leaf samples for RNA-sequencing were collected every 4 hours directly from plants grown on sealed flasks for the first two experiments (January and February 2015, Set 1 and Set 2 respectively) and every 2 hours from the final experiment (October 2015, Set 3), where both medium-sized and large plants were collected. Because *Erycina* species are considered miniatures and are therefore relatively small for destructive leaf

sampling, we use individual genets as biological replicates at each time point. Leaf samples were flash frozen in liquid N_2 and stored at -80°C.

Erycina crista-galli (Rchb.f.) N.H.Williams & M.W,Chase (C_3) plants were wild-collected from Peña Blanca, District of Capira, Republic of Panama at 858m above sea level, then grown and propagated in a commercial orchid greenhouse in Bajo Bonito, District of Capira, Republic of Panama. Plants were fertilized once a week week alternatively with a 20-20-20 or 16-32-16 N-P-K fertilizer. Similarly sized and aged plants were moved into an environmental growth chamber at the Smithsonian Tropical Research Institute laboratories (Panama City, Panama) in April 2016, where they were allowed to acclimate for 48 hours to the following conditions: 12 hour light/dark cycle (lights on 6 a.m.), 25°C/22°C day/night temperatures, 60% humidity, and a light intensity of 30 µmol m^{-2} s^{-1}, which is similar to the light intensity this species would experience naturally. Biological replicates (consisting of entire shoots without root tissue) were sampled every 4 hours over a 24-hour period, starting at ZT0 (lights on, 6 a.m.) with 4 replicates per time point. Tissue was flash frozen in liquid nitrogen and stored as described above for *E. pusilla*.

RNA was isolated from leaf tissue of both *Erycina* species using the RNeasy Plant Mini Kit (Qiagen). RNA samples were subsequently quantified via Nanodrop and checked for integrity with a Bioanalyzer v2100. RNA libraries were constructed using the Kapa mRNA stranded kit with a combinatorial barcoding scheme (Glenn et al., 2016). Libraries were sequenced on an Illumina NextSeq500 with PE75 reads, pooling 30-32 samples per run. A summary of the data can be found in Supplemental Table 1.

Gas exchange

Individual shoots (leaves emerging from a common base) of *E. pusilla* and *E. crista-galli* species were individually sealed within a CQP 130 porometer gas-exchange cuvette (Walz, Effeltrich, Germany), located inside an environmental chamber (Environmental Growth Chambers, OH, USA) operating on 12h light (6AM to 6PM) at 28°C, and 12h dark (6PM to 6AM) cycle at 22°C. Light intensity inside the chamber was 230 µmol m^{-2} s^{-1}. Plants were watered 3-4 times daily and humidity inside the chamber was maintained at near 60%. Continuous net CO_2 exchange was measured for each plant for up to 8 day/night cycles with data points obtained every 4 minutes. For *E. pusilla*, water was withheld from one of the three plants measured (drought stress) between the fifth and the eighth day; regarding *E. crista-galli*, water was withheld from the third to the fifth day for one plant of the three measured. Data is presented for all 3 replicates of each of the two *Erycina* species in Supplemental Figure 1.

Titratable Acidity

Erycina crista-galli plants were too small for both transcriptomic and titratable acidity analysis, therefore leaf tissue from this species was collected only for RNA-Seq. Leaf samples from *E. pusilla* were collected from the same plant and at the same time such that half the plant was used for RNA-sequencing and the other half was used for titratable acidity assays. Leaves from mature plants were collected every 4 hours, flash frozen in liquid N_2, weighed, and boiled in 20% ethanol and deionized water. Titratable acidity was measured as the amount of 0.002M NaOH required to neutralize the extract to a pH of 7. Because leaf tissue was limited for *E. crista-galli* plants, we conducted titrations on three plants that were not sampled for RNAseq to

confirm their status as C_3. Samples were collected in greenhouse conditions at dawn and dusk with three replicates at each time. Titratable acidity was measured as for *E. pusilla* but using 0.001M KOH.

Transcriptome assembly

An initial *de novo* transcriptome was assembled from Sets 1 and 2 for *E. pusilla* sequences and from all samples sequenced for *E. crista-galli* using Trinity v2.0.6 (Haas et al., 2013). Reads were cleaned using Trimmomatic (Bolger et al., 2014) as implemented in Trinity, and assemblies were made on *in-silico* normalized reads. An initial evaluation of read mapping results from *E. pusilla* Sets 1, 2, and 3 showed a large degree of variation among replicates; to reduce this variation, we used only reads from Sets 1 and 2, as well as medium sized plants from Set 3. All reads for *E. crista-galli* were included in the analysis. These data were further reduced to include only 4 replicates per time points for a total of 24 samples. Replicates were chosen randomly. Read mapping and abundance estimation for transcripts was conducted separately in each species using RSEM v1.3.0 (Li and Dewey, 2011) and Kallisto (Bray et al., 2016). Transcripts with a transcripts per kilobase million mapped (TPM) < 2 were removed and the reads were re-mapped to the filtered assemblies.

Ortholog circumscription and isoform filtering

To determine gene family circumscription and annotation, all transcripts were sorted into 14 orthogroup gene families from the genomes of the following: *Amborella trichopoda, Ananas comosus, Arabidopsis thaliana, Asparagus officinalis, Brachypodium distachyon, Carica papaya, Dendrobium catenatum, Elaeis guineensis, Musa acuminata, Oryza sativa, Phalaenopsis equestris, Solanum lycopersicum , Sorghum bicolor, Spirodella polyrrhiza , Vitis vinifera* , and *Zostera marina*. Assembled transcripts were first used to query the genome database using blastx and sorted to gene families (orthogroups) based on best BLAST hit. The gene families were annotated by *Arabidopsis* members, using TAIR 10 (www.arabidopsis.org) classifications. Transcripts were retained only if they 1) had a length less than the maximum sequence of a gene family based on the sequenced genomes and 2) had a length no less than 50% of the minimum sequence length based on sequenced genome members of that gene family.

Trinity produces both gene components and subsidiary isoforms, which may represent true alternative splice isoforms or allelic or paralogous sequence variants. To mitigate dilution of read mappings to multiple isoforms, we instead used gene components (hereafter referred to as transcripts) for all further analyses (including gene level read mapping from RSEM). For gene tree estimation we took the longest isoform per component per orthogroup, using our minimum/maximum orthogroup filtered data set. Scripts for orthogroup sorting and filtering can be found at www.github.com/kheyduk/Erycina.

Time-dependent clustering

To incorporate time into our clustering analysis, we used R software package maSigPro v1.46.0 (Conesa et al., 2006; Nueda et al., 2014), which analyzes expression data for patterns across time by fitting each gene's expression pattern to a polynomial using stepwise regression. Cross-normalized read counts and a negative binomial distribution for the generalized linear models were used. For each transcript, maSigPro estimated up to a 4^{th} degree polynomial and tested the fit via ANOVA. Transcripts that had significantly time-structured expression

(Benjamini & Hochberg adjusted $p < 0.05$) were retained while all others were removed from further analysis. Additionally, any genes considered overly influential based on DFBETAS diagnostic (how much an observation affects the estimate of a regression coefficient) (Belsley et al., 1980) were also removed. In total, 1,515 transcripts from *E. pusilla* and 505 transcripts from *E. crista-galli* were removed as influential genes.

The remaining transcripts that did show time-dependent expression (n=7,066 in *E. pusilla* and n=7,127 in *E. crista-galli*) were clustered by fuzzy clustering based on similarity in expression profiles. An optimal fuzzifier *m*, a parameter that determines how much clusters can overlap, was calculated in the Mfuzz package (Kumar and E Futschik, 2007) of R for each species (*m*=1.09 for both *E. pusilla* and *E. crista-galli*). The number of groups *k* for each species was determined by choosing a value which minimizes the within-group variance (Supplemental Figure 2); a *k* of 6 was used for both *E. pusilla* and *E. crista-galli*. Z-scores of normalized counts were calculated for each gene in each cluster, as well as a median cluster expression, for each species separately.

Gene trees and expression

Gene trees were estimated for phosphoenolpyruvate carboxylase (PEPC) and its kinase (PPCK) by first aligning nucleotide sequences from *Erycina* transcripts and their associated gene family members from the sequenced genomes using PASTA (Mirarab et al., 2014), then estimating trees using RAxML (Stamatakis, 2006). Gene expression for genes of interest was plotted based on averaged transcripts per million mapped (TPM) for each replicate.

Network analysis

To identify regulatory candidates possibly involved in CAM and examine the relationships between genes within clusters, we used the ARACNe-AP v1.4 (Lachmann et al., 2016) algorithm to create networks of co-expressed transcripts from both species separately. Briefly, the algorithm randomly samples gene pairs and uses an adaptive partitioning approach to infer a pairwise mutual information (MI) statistic, or measure of statistical dependence, between them. This process is repeated iteratively for a specified number of bootstraps, while at each step removing indirect interactions. A final network is built based on the consensus of all bootstrap runs. Although ARACNe provides an option to specify transcription factors to generate a directed network by only considering interactions with a transcription factor source, we chose to generate an undirected gene co-expression network of genes that were significantly time-structured based on our maSigPro analysis, using 100 bootstrap replicates in ARACNe.

We imported network data into Cytoscape (Shannon et al., 2003) to generate visualizations and calculate network statistics. Nodes were color coded by their cluster membership and scaled to represent number of connections. Network statistics were exported and further analyzed at the orthogroup level. We calculated which orthogroups had the largest average difference in connectivity between the two species. We first calculated the average connectivity (number of directed edges, output from Cytoscape Network Analysis) for each orthogroup per species, then normalized these via Z-scores and subtracted the Z-score of the orthogroup in *E. pusilla* from that in *E. crista-galli*. Outliers – those orthogroups with the largest difference between species in connectivity – had Z-score differences above and below the upper and lower quantiles (Supplemental Figure 3).

The outlier orthogroups with large changes to connectivity between species (Supplemental Table 3) were explored for genes of interest. The largest difference in connectivity was found in a E3 ubiquitin ligase gene family shown to play a role in ABA signaling, with the *Arabidopsis* homolog known as ring finger of seed longevity1 (RSL1). To explore differences in network connections of RSL1 between the two species, we employed the diffusion algorithm (Carlin et al., 2017) in Cytoscape which finds strongly interactive nodes to a target of interest. For both species, we found the diffusion network for the RSL1 gene. Only a single gene copy of RSL1 was time-structured in *E. crista-galli*, but *E. pusilla* had two copies found in the ARACNe network. One gene copy had only a single connection to any other gene in the network and was not analyzed further. The other copy in *E. pusilla*, which had 72 directed connections, was used as the center of the diffusion network. Diffusion networks were compared for orthogroup content between species using a hypergeometric test. Gene Ontology (GO) terms were compared for the two RSL1 subnetworks and checked for enrichment using hypergeometric test (using all GO terms found in either ARACNe network as the universe), correcting for multiple testing with Bonferroni-Holm significance correction.

Results

Gas exchange patterns and titratable acidity

Gas exchange data collected continuously showed net nighttime CO_2 uptake in CAM *E. pusilla* under both well-watered conditions and while drought stressed (Fig. 2a). C_3 *Erycina crista-galli* displayed net CO_2 uptake during the light period only. There was no net uptake of CO_2 at night. Nonetheless, under drought stress, a slight decrease in respiratory loss of CO_2 at night may indicate low levels of CAM cycling. Titration data collected from the same plants and at the time of RNA-sampling in *E. pusilla* confirms CAM function in the plants used for gene expression analysis, with a significant increase in leaf titratable acids occurring towards the end of the dark period (Fig. 2A, 6AM), and a reduction in total acids during the day period. Although the C_3 *E. crista-galli* had higher overall levels of leaf acids, there was no significant diurnal fluctuation (Fig. 2b).

Clustering of genes with time-structured expression profiles

After filtering by minimum/maximum length each transcript's orthogroup, 23,596 and 26,437 genes were retained in *E. pusilla* and *E. crista-galli*, respectively. Both species had a similar number of genes that were significantly time-structured (~7,000) according to maSigPro. Each species had best fit to k=6 clusters, with three clusters showing nighttime biased expression and three with daytime bias (Fig. 3). Expression of PEPC, the initial carboxylating enzyme in the CO_2 fixation pathway at night, increased in expression just before the onset of darkness in the CAM species *E. pusilla* (Fig. 4A). In contrast, there was a low, but significant, time-structured expression pattern of PEPC in the C_3 species *E. crista-galli* (Fig. 4B). The dedicated kinase, PPCK, which phosphorylates PEPC and allows it to function in the presence of malate, likewise showed a strong nocturnal increase in expression in the CAM species, with similar levels of expression in the C_3 species (Fig. 4B).

Network Comparisons

The network for C_3 *E. crista-galli* had 119,338 directed connections between 4,828 nodes, whereas the CAM *E. pusilla* network was notably less connected, with only 76,071

connections between 4,591 nodes. Although the number of genes in each network was similar, overall connectivity of *E. pusilla* is easily seen in both the fewer number of connections as well as the mean number of connections (34.3 in *E. pusilla* vs. 50.3 in *E. crista-galli*). As expected, genes from the same co-expressed cluster (Fig. 3) were grouped within the larger ARACNe network for each species (Fig. 5A,C).

Comparison of network connectivity of the time-structured genes found 149 outlier transcripts that had large species differences in the number of connected directed edges (Additional file 4); these were largely skewed toward increased connectivity in *E. crista-galli* (n=90). Annotations of these outliers revealed a number of genes involved in stomatal opening/closing and ABA signaling. GO term enrichment indicates that outliers that skew toward more connectivity in C_3 *E. crista-galli* were enriched for vacuolar and tonoplast membrane proteins and potassium and calcium transport. Genes that were more connected in CAM *E. pusilla* were enriched for genes involved in aldehyde dehydrogenase activity, among other functions (Supplemental Table 4).

A E3 ubiquitin ligase also known as ring finger of seed longevity1 (RSL1) had the greatest difference in connectivity between the two species. RSL1 has been shown to be a negative regulator of ABA signaling (Bueso et al., 2014) and was chosen as a center node for comparison between the two species. The diffusion algorithm used to create subnetworks defaults to producing a subnetwork that is 10% of the total nodes in the larger network; as a result, both species subnetworks were roughly the same size, containing about 400 genes. However, the connectivity of those subnetworks differed greatly (Fig. 5B,D); the C_3 *E. crista-galli* RSL1 subnetwork contained 30,392 connections, whereas the CAM *E. pusilla* network had only 9,244. The subnetworks differed in their gene content as well. There contained 429 and 427 orthogroups in *E. pusilla* and *E. crista-galli*, respectively, but only 57 orthogroups were shared between the two (this was not significantly under-enriched via a hypergeometric test (p=1)). While all of the genes in the *E. crista-galli* RSL1 subnetwork were in night-biased expression clusters (485), *E. pusilla* had more subnetwork genes in day-biased clusters (298) than in night biased ones (162). GO term enrichment indicates both subnetworks are enriched for chloroplast, chloroplast stroma, and photosynthesis (Supplemental Table 5).

While both subnetworks were centered on RSL1 and generally were enriched for similar types of genes involved in chloroplast functions and photosynthesis, there were substantial differences in gene content between the networks. The focal gene RSL1 acts as a master negative regulator of ABA signaling pathway by targeting pyrabactin resistance 1 (PYR1) and PYR-like (PYL) ABA receptors for degradation. More generally its role in protein ubiquitination is relatively unknown. RSL1 was the third most connected node in the C_3 *E. crista-galli* subnetwork (419 directed connections; most connected node had 423 connections). In CAM *E. pusilla*, RSL1 was 173rd out of 460 genes in connectivity. *E. pusilla* had a number of ABA responsive genes in its RSL1 subnetwork that *E. crista-galli* did not, including protein phosphatase 2C (PP2C), a gene encoding a member of the Snf1-related kinase family, plus a homolog of *ABA Overly Sensitive 5*. Recent work has shown that ABA responses in stomata including *PYR/PYL* and downstream genes are responsible for constitutive stomatal aperture, as well as stomatal responses to drought stress (Gonzalez-Guzman et al., 2012).

Erycina crista-galli, on the other hand, had a number of light sensing and circadian clock genes that CAM *E. pusilla* did not have in the RSL1 subnetwork, including *lov kelch protein 2* (LKP2/ZTL gene family), *time for coffee* (*TIC*), *phytochrome B/D* (*PHYB*), and *protein phosphatase 2A* (*PP2A*). LKP2/ZTL and PHYB are both involved in light sensing (blue and red/far-red, respectively). Specifically, *LKP2/ZTL* genes are thought to regulate light induced protein degradation via their function as E3 ligases (Demarsy and Fankhauser, 2009; Ito et al., 2012), and *ztl* mutants in *Arabidopsis* show a prolonged clock period under constant light due to the lack of degradation of clock components via ZTL (Más et al., 2003; Somers et al., 2000). TIC has been found to be responsible for the amplitude of the circadian clock but is not thought to be directly involved in light signaling (Hall et al., 2003). TIC is also implicated in daytime transcriptional induction via its association with the central circadian oscillator late elongated hypocotyl (LHY) (Ding et al., 2007). Finally, PP2A is a member of a large family of plant phosphoprotein phosphatases (PPP) with several cellular roles (Uhrig et al., 2013). PP2A specifically has been implicated in brassinosteroid signaling, light signaling via dephosphorylation of phototropin2 (Tseng and Briggs, 2010), flowering time control (Kim et al., 2002), as well as the induction of CAM under certain abiotic stresses (Cushman and Bohnert, 1999).

Although the C_3 *E. crista-galli* RSL1 subnetwork contained circadian and light sensing transcripts that were absent in the *E. pusilla* RSL1 subnetwork, these transcripts largely did not have different expression patterns between the two species with the exception of *PHYB* (Fig. 6B). The contrasting gene content of the RSL1 subnetwork suggests light sensing and circadian regulatory cascades comprise large differences between C_3 and CAM species, rather than levels of gene expression, which were quite similar.

Discussion

Shared gene expression patterns

While traditionally it was thought that canonical CAM genes should have large differences in timing and magnitude of expression between C_3 and CAM species, recent work has highlighted that between closely related C_3 and CAM species, that may not always be the case (Heyduk et al., 2018b). In *Erycina*, it appears that a similar pattern holds. Both the CAM and C_3 species have time-structured expression of PEPC albeit at very different expression levels. While this alone says little about PEPC's function in both species, the high nocturnal expression of PEPC's dedicated kinase, PPCK, in both *E. pusilla* and *E. crista-galli* suggests that PEPC is being phosphorylated in both species and therefore has the need to function in the presence of malate.

It is worth noting that all C_3 species have the genes involved in the CAM cycle. Many, including PEPC, function in anaplerotic reactions of the TCA cycle. PEPC has also been shown to have a role in malate production for osmotic regulation of stomatal aperture and in CO_2 fixation in guard cells of tobacco (Asai et al., 2000; Daloso et al., 2015). Because most RNA-seq or gene expression studies to date in C_3 species sample during the day, understanding how common nocturnal *PEPC* expression is across flowering plants will require more nighttime gene expression studies in C_3 species, especially in lineages closely related to CAM species. While nearly all the genes in the CAM CO_2 fixation pathway have known functions in C_3 species, *PPCK* is a notable exception. Phosphorylation of PEPC by PPCK in CAM and C_4 species is well-described (Jiao and Chollet, 1989; Nimmo et al., 1986; Taybi et al., 2000a); malate and

other organic acids act as negative regulators of PEPC, but phosphorylation of PEPC renders it immune to these negative effects. Thus, phosphorylated PEPC via PPCK is required for high levels of CAM and C_4 malic acid production. In C_3 species, however, there is no clear need for heavily phosphorylated PEPC, especially at night, making the expression of *PEPC* and *PPCK* in C_3 *E. crista-galli* intriguing (though see Sullivan et al., 2004 and Fukayama et al., 2006 for work on nocturnal *PPCK* expression in soybean and rice, respectively).

Alterations in regulatory pathways between C_3 and CAM orchids

Network analysis of co-expressed genes and subsequent comparisons between species can give insights not only into the changes in expression, but also the degree to which a given gene changes in connectivity between species. Extensive connectivity for a gene has long thought to be a signal of a "hub" or master regulatory gene – one that cannot experience large changes in timing or magnitude of expression without significant perturbations to the entire network. A major assumption of co-expression networks is that they rely on mRNA as an accurate predictor of downstream processes; while this is not always the case, recent work showed that although mRNA and protein networks differed in their gene content, they overlapped in gene ontologies and were predictive of pathway regulation in maize (Walley et al., 2016). Additionally, it has been shown that a correlation exists between connectivity of a gene within a network and the evolutionary conservation of the gene's sequence across a number of flowering plant species (Masalia et al., 2017). It is therefore somewhat surprising to observe that as many as 149 gene families have large changes in connectivity between the two closely related *Erycina* species.

The 149 outlier gene families in *Erycina* were enriched for functions in protein degradation via phosphorylation and ubiquitination. While typically differences in phenotype are considered the result of changes in the abundance of gene products, our data highlight the importance of considering protein degradation as well. The differences in connectivity of protein degradation pathway genes were unbiased between species – in other words, genes involved in protein degradation have increased connectivity in both species. More interestingly, genes that had increased connectivity in C_3 *E. crista-galli* were enriched for GO terms involved in potassium and chloride channels and membrane proteins associated with chloroplasts and vacuoles. Greater connectivity of such genes in the C_3 species indicates an increased reliance on ion and metabolite fluxes. In stomatal guard cells, which make up a smaller portion of the whole-leaf transcriptome, these fluxes directly affect stomatal aperture and may play a role in alternative regulation of stomatal opening in C_3 and CAM species.

Regulatory changes in ABA, light, and clock perception

Stomatal opening in CAM species has been vastly understudied, despite the opportunities it presents for understanding a fundamental biological process. In C_3 species, stomatal opening is thought to be regulated by blue and red light inputs, whereas stomatal closing is driven by efflux of potassium cations. It remains largely unknown how stomata sense darkness, but experimental data have suggested a large role of CO_2 concentrations on stomatal aperture (Cockburn et al., 1979). Draw down of CO_2 concentrations at night in the intercellular airspace would result in stomatal opening, whereas high concentrations of CO_2 from decarboxylation during the day may promote stomatal closure. CO_2 concentrations undoubtedly play some role in the inverted stomatal aperture of CAM species, but more contemporary genomic work has implicated

additional levels of regulation in CAM stomata (Abraham et al., 2016; Cushman and Bohnert, 1997; Wai et al., 2017). For example, ABA may be one signaling molecule for nocturnal stomatal closure (Desikan et al., 2004; Gonzalez-Guzman et al., 2012; Merilo et al., 2013). Gene expression results coupled with network analysis in *Erycina* indicate that both ABA signaling and light sensing are likely to be altered. Indeed, the gene with the largest difference in connectivity between species is *RSL1*, which encodes a E3 ubiquitin ligase known to function in stomatal response to ABA. Expression of RSL1 is higher in CAM *E. pusilla* and is clustered with day-biased genes (Fig. 6A), whereas in C_3 *E. crista-galli* expression is about half that in *E. pusilla* and slightly increases in expression during the dark period. These expression patterns are consistent with stomatal regulation between C_3 and CAM species. In *E. crista-galli* nighttime ABA may play a larger role in drought-induced or nighttime stomatal closure than in the CAM *E. pusilla,* where the nocturnal stomatal closure driven by ABA signaling must be repressed to allow for nighttime CO_2 acquisition. It is also unknown what causes daytime stomatal closure in CAM species; while high intracellular CO_2 concentrations may play a role, so might ABA signaling. *RSL1*'s high connectivity to other genes in the subnetwork of *E. crista-galli* suggests that alterations to regulatory networks are also important for fully functional CAM.

Stomata, while highly responsive to ABA, are also strongly regulated by light inputs. The gene encoding phytochrome B (phyB), a photoreceptor that both regulates transcriptional responses to red and far-red light as well as entrains the circadian clock (Goosey et al., 1997; Más et al., 2000; Ni et al., 1999), was differentially regulated and expressed in the two *Erycina* species (Fig. 6B). While both species had copies of *phyB* that showed time-structured expression, only *E. crista-galli* had a copy that was in the same network as *RSL1* and had many connections to other genes in the same network. A single copy of *phyB* was time-structured in CAM *E. pusilla*, and the expression level relative to copies in C_3 *E. crista-galli* was quite low. Instead, the constitutively expressed copy of *phyB* in *E. pusilla* had the highest expression, but what it's role might be given constant expression across time is unclear. The stark difference in both connectivity and expression levels of *phyB* in *E. pusilla* and *E. crista-galli* suggests that light-induced transcriptional regulation has a greater role in the C_3 species, and that phyB mediated transcriptional regulation in CAM species may be significantly reduced. Additionally, the presence of *PP2A* in the C_3 *E. crista-galli* RSL1 subnetwork, but not in *E. pusilla,* further indicates that light induced responses are differentially regulated (although expression was similar between the two species, Fig. 6D). PP2A is, among many other tasks, responsible for dephosphorylation of photoropin 2, which subsequently promotes stomatal opening (Tseng and Briggs, 2010; Uhrig et al., 2013).

Differences in light input sensing and signaling between C_3 and CAM species is not surprising, but relatively little work has focused on this aspect of CAM biology. Previous work assessed light responses in the facultative CAM species *Portulacaria afra* (Lee and Assmann, 1992) and *Mesembryanthemum crystallinum* (Tallman et al., 1997) and showed that both species had reduced stomatal response to blue light signals when relying on the CAM cycle for carbon fixation (though see Ceusters et al., 2014). Additionally, *M. crystallinum* had reduced guard cell zeaxanthin production during the day in the CAM state compared to the C_3 state. The reduction in zeaxanthin in the CAM state was shown to be a result of the downregulation of the pathway, rather than an aberration in guard cell chloroplasts. Changes to regulation and expression of

phyB in *Erycina* further support significantly altered light-induced pathways in CAM species relative to C_3.

TIME FOR COFFEE (*TIC*) also stood out in its altered connectivity between the C_3 and CAM species. In C_3 *E. crista-galli*, TIC was present in the RSL1 diffusion network whereas it was not for *E. pusilla*. TIC has been shown to be involved in maintaining the amplitude of the circadian clock in *Arabidopsis* (Ding et al., 2007; Hall et al., 2003), as well as regulating metabolic homeostasis and response to environmental cues (Sanchez-Villarreal et al., 2013). *tic* mutants in *Arabidopsis* showed large phenotypic effects ranging from late flowering to anatomical abnormalities. The *tic* mutants also showed extreme tolerance to drought, likely due to increased amounts of osmolytes such as proline and myo-inositol, as well as an accumulation of starch. As a result of the pleiotropic effects of TIC, altered network status of TIC in CAM *E. pusilla* compared to C3 *E. crista-galli* likely results a complex alteration in phenotype. In general, the mechanism that link the circadian clock and CAM photosynthesis are unknown (Boxall et al., 2005), and research to uncover circadian regulation of CAM is limited to transcriptomic studies. Gene expression comparisons between CAM *Agave* and C_3 *Arabidopsis* revealed changes to the timing of expression of *REVEILLE 1*, a clock output gene that integrates the circadian network to metabolic activity (Yin et al., 2018). While it's possible that changes to the timing of *REVEILLE 1* are required for CAM evolution, expression differences may also be a result of lineage-specific changes to expression unrelated to CAM. Comparisons of closely related C3 and CAM species of *Erycina* suggest that alteration of transcriptional cascades from circadian oscillators may play a role in the evolution of CAM, rather than large scale changes to the timing or abundance of expression, but additional work to link clock outputs to the CAM phenotype are necessary.

Implications for the evolution of CAM in Oncidiinae

Both CAM and C_4 photosynthesis have evolved multiple times across the flowering plant phylogeny, suggesting that the evolution of these complex traits cannot be insurmountably difficult. Recent physiological work demonstrates that the evolution of anatomical traits required for CAM or C_4 often predates the emergence of strong, constitutive carbon concentrating mechanisms (Christin et al., 2013; Heyduk et al., 2016). Other transcriptomic work has shown that closely related C_3 and CAM species share the expression of canonical CAM genes, especially *PEPC* and *PPCK* as seen here in *Erycina* (Heyduk et al., 2018b). It has even been suggested that many C_3 plants already have the nocturnal CAM cycle in place for fixation of respired CO_2 and generation of amino acids (Bräutigam et al., 2017), but this alone is unlikely to entirely explain the repeated and relatively frequent emergence of CAM on the angiosperm phylogeny.

Indeed it appears that gene expression alone would not facilitate the large-scale transition from C_3 to CAM. Recent comparative work across multiple C_3 and C_4 transcriptomes highlighted the recurrent co-option of highly expressed genes from C_3 species into C_4 (Moreno-Villena et al., 2018). The initial co-option of highly expressed gene copies happened early in the evolutionary trajectory between C_3 and C_4, and later steps included the refinement of C_3 enzymes, including kinetic and tissue specificity. It is quite likely that a similar model of evolution holds for CAM, in that C_3 relatives of CAM lineages have been shown to have similar expression patterns of canonical CAM genes (though, it is worth noting, they are not typically highly expressed in the

C_3 species). In *Erycina*, both C_3 and CAM species share similar expression profiles for *PEPC* and *PPCK*, with the latter having nearly identical levels of expression in both taxa. It is possible that low levels of nocturnal CO_2 fixation via PEPC evolved in the ancestor of both species for non-photosynthetic reasons; for example, tobacco leaves (Kunitake et al., 1959) and cotton ovules (Dhindsa et al., 1975) show various levels of carbon concentration, and fixation of cytosolic CO_2 via PEPC is required to replenish the citric acid cycle (Aubry et al., 2011). Because these pathways already exist, slight upregulation of some components in a shared ancestor may have enabled the origins of CAM in certain lineages (Bräutigam et al., 2017).

However, gene expression alone clearly does not make a species CAM, and further refinement is necessary. Refinement of CAM may take the form of improving secondary metabolic processes or genomic characteristics that allow for strong and constitutive CAM to exist. For example, fine tuning of carbohydrate turnover is necessary for CAM function (Borland et al., 2016; Ceusters et al., 2014). Experiments placing *Mesembryanthemum crystallinum* in CO_2-free air at night resulted in a dampened CAM cycle (Dodd et al., 2003) and comparative RNA-seq in facultative CAM and C_3/CAM comparisons has shown increased reliance on carbohydrate breakdown as CAM function increases (Brilhaus et al., 2016; Heyduk et al., 2018a). Circadian regulation of CAM genes in pineapple was shown to be the result of promoters that induce evening expression (Ming et al., 2015), and genes that link the clock to metabolic outputs had shifts in expression phasing between *Agave* (CAM) and *Arabidopsis* (C_3) (Yin et al., 2018). Variation in timing and magnitude of expression of various light sensing and clock genes has been described not only here in *Erycina*, but also in *Agave* (Abraham et al., 2016). In all, the growing evidence suggests that the secondary processes that make a CAM plant, including stomatal regulation, light sensing and downstream transcriptional responses, and carbohydrate metabolism feedbacks, undergo fine-tuning along the evolutionary trajectory between C_3 and CAM. Further work that characterizes closely related C_3 and CAM species has the potential to greatly advance our understanding of the integration of nighttime CO_2 acquisition with more complex regulatory pathways.

Data Availability

Raw sequence reads are available on NCBI Short Read Archive, under the BioProject PRJNA483943. All scripts used that are not part of existing programs are available at www.github.com/kheyduk.

Acknowledgements

The authors would like to sincerely thank Orquideas Tropicales in Panama for generously donating *Erycina crista-galli* plants. Additional gratitude goes to Saravanaraj Ayyampalayam for assistance with data processing, Jeremy Ray for assistance with RNA preparation, and Rishi Masalia for giving feedback on figures. This work was supported by NSF grants to J.L.M, V.A, and K.S (DEB #1442199 and 1442190) and the by Smithsonian Tropical Research Institute.

Author Contributions Statement

JLM and VAA conceived and led the project; KS and KW collected, phenotyped plants, and conducted continuous gas exchange measurements, and KS additionally sampled *E. crista-galli* for RNA-seq; VAA and TL grew and sampled *Erycina pusilla* individuals for RNA and physiology; KH and MH sequenced and analyzed the data and prepared the manuscript; JLM

oversaw general experimental design, and all authors contributed to the final version of this manuscript.

Conflict of Interest

The authors declare no personal, professional, or financial conflicts of interest.

References

Abraham, P. E., Yin, H., Borland, A. M., Weighill, D., Lim, S. D., De Paoli, H. C., et al. (2016). Transcript, protein and metabolite temporal dynamics in the CAM plant Agave. *Nat. Plants* 2, 16178. doi:10.1038/nplants.2016.178.

Asai, N., Nakajima, N., Tamaoki, M., Kamada, H., and Kondo, N. (2000). Role of Malate Synthesis Mediated by Phosphoenolpyruvate Carboxylase in Guard Cells in the Regulation of Stomatal Movement. *Plant Cell Physiol.* 41, 10–15. doi:10.1093/pcp/41.1.10.

Aubry, S., Brown, N. J., and Hibberd, J. M. (2011). The role of proteins in C3 plants prior to their recruitment into the C4 pathway. *J. Exp. Bot.* 62, 3049–3059. doi:10.1093/jxb/err012.

Belsley, D. A., Kuuh, E., and E., W. R. (1980). *Wiley Series in Probability and Statistics*. doi:10.1002/9780470316559.oth1.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–20. doi:10.1093/bioinformatics/btu170.

Borland, A. M., Guo, H.-B., Yang, X., and Cushman, J. C. (2016). Orchestration of carbohydrate processing for crassulacean acid metabolism. *Curr. Opin. Plant Biol.* 31, 118–124. doi:10.1016/j.pbi.2016.04.001.

Borland, A. M., Hartwell, J., Jenkins, G. I., Wilkins, M. B., Nimmo, H. G., Borland, A. M., et al. (1999). Metabolite Control Overrides Circadian Regulation of Phosphoenolpyruvate Carboxylase Kinase and CO□ Fixation in Crassulacean Acid Metabolism. *Plant Physiol.* 121, 889–896.

Boxall, S. F., Foster, J. M., Bohnert, H. J., Cushman, J. C., Nimmo, H. G., and Hartwell, J. (2005). Conservation and Divergence of Circadian Clock Operation in a Stress-Inducible Crassulacean Acid Metabolism Species Reveals Clock Compensation against Stress. *PLANT Physiol.* 137, 969–982. doi:10.1104/pp.104.054577.

Bräutigam, A., Schlüter, U., Eisenhut, M., and Gowik, U. (2017). On the Evolutionary Origin of CAM Photosynthesis. *Plant Physiol.* 174, 473–477. doi:10.1104/pp.17.00195.

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi:10.1038/nbt.3519.

Brilhaus, D., Bräutigam, A., Mettler-Altmann, T., Winter, K., and Weber, A. P. M. (2016). Reversible Burst of Transcriptional Changes during Induction of Crassulacean Acid Metabolism in Talinum triangulare. *Plant Physiol.* 170, 102–22. doi:10.1104/pp.15.01076.

Bueso, E., Rodriguez, L., Lorenzo-Orts, L., Gonzalez-Guzman, M., Sayas, E., Muñoz-Bertomeu, J., et al. (2014). The single-subunit RING-type E3 ubiquitin ligase RSL1 targets PYL4 and PYR1 ABA receptors in plasma membrane to modulate abscisic acid signaling. *Plant J.* 80, 1057–1071. doi:10.1111/tpj.12708.

Carlin, D. E., Demchak, B., Pratt, D., Sage, E., and Ideker, T. (2017). Network propagation in the cytoscape cyberinfrastructure. *PLOS Comput. Biol.* 13, e1005598. doi:10.1371/journal.pcbi.1005598.

Ceusters, J., Borland, A. M., Taybi, T., Frans, M., Godts, C., and De Proft, M. P. (2014). Light

642 quality modulates metabolic synchronization over the diel phases of crassulacean acid
643 metabolism. *J. Exp. Bot.* 65, 3705–3714. doi:10.1093/jxb/eru185.
644 Chase, M. W., Cameron, K. M., Freudenstein, J. V., Pridgeon, A. M., Salazar, G., van den Berg,
645 C., et al. (2015). An updated classification of Orchidaceae. *Bot. J. Linn. Soc.* 177, 151–174.
646 doi:10.1111/boj.12234.
647 Christin, P.-A., Arakaki, M., Osborne, C. P., Bräutigam, A., Sage, R. F., Hibberd, J. M., et al.
648 (2014). Shared origins of a key enzyme during the evolution of C4 and CAM metabolism. *J.*
649 *Exp. Bot.* 65, 3609–3621. doi:10.1093/jxb/eru087.
650 Christin, P.-A., Osborne, C. P., Chatelet, D. S., Columbus, J. T., Besnard, G., Hodkinson, T. R.,
651 et al. (2013). Anatomical enablers and the evolution of C4 photosynthesis in grasses. *Proc.*
652 *Natl. Acad. Sci. U. S. A.* 110, 1381–6. doi:10.1073/pnas.1216777110.
653 Cockburn, W., Ting, I. P., and Sternberg, L. O. (1979). Relationships between Stomatal Behavior
654 and Internal Carbon Dioxide Concentration in Crassulacean Acid Metabolism Plants. *Plant*
655 *Physiol.* 63, 1029–32. doi:10.1104/PP.63.6.1029.
656 Conesa, A., Nueda, M. J., Ferrer, A., and Talón, M. (2006). maSigPro: a method to identify
657 significantly differential expression profiles in time-course microarray experiments.
658 *Bioinformatics* 22, 1096–1102. doi:10.1093/bioinformatics/btl056.
659 Crayn, D. M., Winter, K., and Smith, J. A. C. (2004). Multiple origins of crassulacean acid
660 metabolism and the epiphytic habit in the Neotropical family Bromeliaceae. *Proc. Natl.*
661 *Acad. Sci. U. S. A.* 101, 3703–8. doi:10.1073/pnas.0400366101.
662 Cushman, J. C., and Bohnert, H. J. (1997). Molecular Genetics of Crassulacean Acid
663 Metabolism. *Plant Physiol.* 113, 667–676. Available at:
664 http://www.ncbi.nlm.nih.gov/pubmed/15012212.
665 Cushman, J. C., and Bohnert, H. J. (1999). CRASSULACEAN ACID METABOLISM:
666 Molecular Genetics. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 305–332.
667 doi:10.1146/annurev.arplant.50.1.305.
668 Cushman, J. C., Tillett, R. L., Wood, J. A., Branco, J. M., and Schlauch, K. A. (2008). Large-
669 scale mRNA expression profiling in the common ice plant, Mesembryanthemum
670 crystallinum, performing C3 photosynthesis and Crassulacean acid metabolism (CAM). *J.*
671 *Exp. Bot.* 59, 1875–94. doi:10.1093/jxb/ern008.
672 Daloso, D. M., Antunes, W. C., Pinheiro, D. P., Waquim, J. P., AraÚJo, W. L., Loureiro, M. E.,
673 et al. (2015). Tobacco guard cells fix CO_2 by both Rubisco and PEPcase while sucrose acts
674 as a substrate during light-induced stomatal opening. *Plant. Cell Environ.* 38, 2353–2371.
675 doi:10.1111/pce.12555.
676 Demarsy, E., and Fankhauser, C. (2009). Higher plants use LOV to perceive blue light. *Curr.*
677 *Opin. Plant Biol.* 12, 69–74. doi:10.1016/J.PBI.2008.09.002.
678 Desikan, R., Cheung, M.-K., Clarke, A., Golding, S., Sagi, M., Fluhr, R., et al. (2004). Hydrogen
679 peroxide is a common signal for darkness- and ABA-induced stomatal closure in *Pisum*
680 *sativum*. *Funct. Plant Biol.* 31, 913. doi:10.1071/FP04035.
681 Dever, L. V, Boxall, S. F., Kneřová, J., and Hartwell, J. (2015). Transgenic perturbation of the
682 decarboxylation phase of Crassulacean acid metabolism alters physiology and metabolism
683 but has only a small effect on growth. *Plant Physiol.* 167, 44–59.
684 doi:10.1104/pp.114.251827.
685 Dhindsa, R. S., Beasley, C. A., and Ting, I. P. (1975). Osmoregulation in Cotton Fiber:
686 Accumulation of Potassium and Malate during Growth. *Plant Physiol.* 56, 394–8. Available
687 at: http://www.ncbi.nlm.nih.gov/pubmed/16659311 [Accessed September 21, 2018].

688 Ding, Z., Millar, A. J., Davis, A. M., and Davis, S. J. (2007). TIME FOR COFFEE encodes a
689 nuclear regulator in the Arabidopsis thaliana circadian clock. *Plant Cell* 19, 1522–36.
690 doi:10.1105/tpc.106.047241.
691 Dodd, A. N., Griffiths, H., Taybi, T., Cushman, J. C., and Borland, A. M. (2003). Integrating diel
692 starch metabolism with the circadian and environmental regulation of Crassulacean acid
693 metabolism in Mesembryanthemum crystallinum. *Planta* 216, 789–797.
694 doi:10.1007/s00425-002-0930-2.
695 Fukayama, H., Tamai, T., Taniguchi, Y., Sullivan, S., Miyao, M., and Nimmo, H. G. (2006).
696 Characterization and functional analysis of phospho *enol* pyruvate carboxylase kinase genes
697 in rice. *Plant J.* 47, 258–268. doi:10.1111/j.1365-313X.2006.02779.x.
698 Gehrig, H., Heute, V., and Kluge, M. (2001). New partial sequences of phosphoenolpyruvate
699 carboxylase as molecular phylogenetic markers. *Mol. Phylogenet. Evol.* 20, 262–274.
700 doi:10.1006/mpev.2001.0973.
701 Glenn, T. C., Nilsen, R., Kieran, T. J., Finger, J. W., Pierson, T. W., Bentley, K. E., et al. (2016).
702 Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries
703 (iTru & iNext). *bioRxiv*. Available at:
704 http://biorxiv.org/content/early/2016/06/15/049114 [Accessed May 31, 2017].
705 Gonzalez-Guzman, M., Pizzio, G. A., Antoni, R., Vera-Sirera, F., Merilo, E., Bassel, G. W., et
706 al. (2012). Arabidopsis PYR/PYL/RCAR receptors play a major role in quantitative
707 regulation of stomatal aperture and transcriptional response to abscisic acid. *Plant Cell* 24,
708 2483–96. doi:10.1105/tpc.112.098574.
709 Goosey, L., Palecanda, L., and Sharrock, R. A. (1997). Differential patterns of expression of the
710 Arabidopsis PHYB, PHYD, and PHYE phytochrome genes. *Plant Physiol.* 115, 959–69.
711 doi:10.1104/PP.115.3.959.
712 Gravendeel, B., Smithson, A., Slik, F. J. W., and Schuiteman, A. (2004). Epiphytism and
713 pollinator specialization: drivers for orchid diversity? *Philos. Trans. R. Soc. Lond. B. Biol.*
714 *Sci.* 359, 1523–1535. doi:10.1098/rstb.2004.1529.
715 Guralnick, L. J., Rorabaugh, P. A., and Hanscom, Z. (1984). Seasonal Shifts of Photosynthesis in
716 Portulacaria afra (L.) Jacq. *Plant Physiol.* 76, 643–646. doi:10.1104/pp.76.3.643.
717 Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Philip, D., Bowden, J., et al. (2013).
718 *De novo transcript sequence reconstruction from RNA-Seq: reference generation and*
719 *analysis with Trinity*. Nature Protocols doi:10.1038/nprot.2013.084.De.
720 Hall, A., Bastow, R. M., Davis, S. J., Hanano, S., McWatters, H. G., Hibberd, V., et al. (2003).
721 The TIME FOR COFFEE gene maintains the amplitude and timing of Arabidopsis
722 circadian clocks. *Plant Cell* 15, 2719–29. doi:10.1105/tpc.013730.
723 Hartwell, J. (2005). The co-ordination of central plant metabolism by the circadian clock.
724 *Biochem. Soc. Trans.* 33, 945–8. doi:10.1042/BST20050945.
725 Heyduk, K., McKain, M. R., Lalani, F., and Leebens-Mack, J. (2016). Evolution of CAM
726 anatomy predates the origins of Crassulacean acid metabolism in the Agavoideae
727 (Asparagaceae). *Mol. Phylogenet. Evol.* 105, 102–113.
728 Heyduk, K., Ray, J. N., Ayyampalayam, S., and Leebens-Mack, J. (2018a). Shifts in gene
729 expression profiles are associated with weak and strong Crassulacean acid metabolism. *Am.*
730 *J. Bot.* doi:10.1002/ajb2.1017.
731 Heyduk, K., Ray, J. N., Ayyampalayam, S., Moledina, N., Borland, A., Harding, S., et al.
732 (2018b). Nocturnal expression of Crassulacean acid metabolism (CAM) genes predates the
733 origin of CAM in the genus Yucca. *bioRxiv*, 371344. doi:10.1101/371344.

734 Holtum, J. A. M., Smith, J. A. C., and Neuhaus, H. E. (2005). Intracellular transport and
735 pathways of carbon flow in plants with crassulacean acid metabolism. *Funct. Plant Biol.* 32,
736 429. doi:10.1071/FP04189.
737 Horn, J. W., Xi, Z., Riina, R., Peirson, J. A., Yang, Y., Dorsey, B. L., et al. (2014). Evolutionary
738 bursts in Euphorbia (Euphorbiaceae) are linked with photosynthetic pathway. *Evolution (N.*
739 *Y).* 68, 3485–3504. doi:10.1111/evo.12534.
740 Ito, S., Song, Y. H., and Imaizumi, T. (2012). LOV Domain-Containing F-Box Proteins: Light-
741 Dependent Protein Degradation Modules in Arabidopsis. *Mol. Plant* 5, 573–582.
742 doi:10.1093/MP/SSS013.
743 Jiao, J.-A., and Chollet, R. (1989). Regulatory seryl-phosphorylation of C4 phosphoenolpyruvate
744 carboxylase by a soluble protein kinase from maize leaves. *Arch. Biochem. Biophys.* 269,
745 526–535. doi:10.1016/0003-9861(89)90136-7.
746 Kim, D.-H., Kang, J.-G., Yang, S.-S., Chung, K.-S., Song, P.-S., and Park, C.-M. (2002). A
747 phytochrome-associated protein phosphatase 2A modulates light signals in flowering time
748 control in Arabidopsis. *Plant Cell* 14, 3043–56. doi:10.1105/TPC.005306.
749 Kovermann, P., Meyer, S., Hörtensteiner, S., Picco, C., Scholz-Starke, J., Ravera, S., et al.
750 (2007). The Arabidopsis vacuolar malate channel is a member of the ALMT family. *Plant*
751 *J.* 52, 1169–1180. doi:10.1111/j.1365-313X.2007.03367.x.
752 Kumar, L., and E Futschik, M. (2007). Mfuzz: a software package for soft clustering of
753 microarray data. *Bioinformation* 2, 5–7. Available at:
754 http://www.ncbi.nlm.nih.gov/pubmed/18084642 [Accessed September 20, 2016].
755 Kunitake, G., Stitt, C., and Saltman, P. (1959). Dark Fixation of CO(2) by Tobacco Leaves.
756 *Plant Physiol.* 34, 123–7. doi:10.1104/PP.34.2.123.
757 Lachmann, A., Giorgi, F. M., Lopez, G., and Califano, A. (2016). ARACNe-AP: Gene network
758 reverse engineering through adaptive partitioning inference of mutual information.
759 *Bioinformatics* 32, 2233–2235. doi:10.1093/bioinformatics/btw216.
760 Lee, D. M., and Assmann, S. M. (1992). Stomatal responses to light in the facultative
761 Crassulacean acid metabolism species, Portulacaria afra. *Physiol. Plant.* 85, 35–42.
762 doi:10.1111/j.1399-3054.1992.tb05260.x.
763 Lee, S. H., Li, C. W., Liau, C. H., Chang, P. Y., Liao, L. J., Lin, C. S., et al. (2015).
764 Establishment of an Agrobacterium-mediated genetic transformation procedure for the
765 experimental model orchid Erycina pusilla. *Plant Cell. Tissue Organ Cult.* 120, 211–220.
766 doi:10.1007/s11240-014-0596-z.
767 Lepiniec, L., Vidal, J., Chollet, R., Gadal, P., and Cretin, C. (1994). Phosphoenolpyruvate
768 carboxylase: structure, regulation and evolution. *Plant Sci.* 99, 111–124. doi:10.1016/0168-
769 9452(94)90168-6.
770 Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data
771 with or without a reference genome. *BMC Bioinformatics* 12, 323. doi:10.1186/1471-2105-
772 12-323.
773 Más, P., Devlin, P. F., Panda, S., and Kay, S. A. (2000). Functional interaction of phytochrome B
774 and cryptochrome 2. *Nature* 408, 207–211. doi:10.1038/35041583.
775 Más, P., Kim, W.-Y., Somers, D. E., and Kay, S. A. (2003). Targeted degradation of TOC1 by
776 ZTL modulates circadian function in Arabidopsis thaliana. *Nature* 426, 567–570.
777 doi:10.1038/nature02163.
778 Masalia, R. R., Bewick, A. J., Burke, J. M., Irizarry, R., Bertranpetit, J., and Laayouni, H.
779 (2017). Connectivity in gene coexpression networks negatively correlates with rates of

molecular evolution in flowering plants. *PLoS One* 12, e0182289. doi:10.1371/journal.pone.0182289.

Merilo, E., Laanemets, K., Hu, H., Xue, S., Jakobson, L., Tulva, I., et al. (2013). PYR/RCAR receptors contribute to ozone-, reduced air humidity-, darkness-, and CO2-induced stomatal regulation. *Plant Physiol.* 162, 1652–68. doi:10.1104/pp.113.220608.

Ming, R., VanBuren, R., Wai, C. M., Tang, H., Schatz, M. C., Bowers, J. E., et al. (2015). The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* 47, 1435–42. doi:10.1038/ng.3435.

Mirarab, S., Nguyen, N., and Warnow, T. (2014). "PASTA: Ultra-Large Multiple Sequence Alignment," in *Research in Computational Molecular Biology: 18th Annual International Conference, RECOMB 2014, Pittsburgh, PA, USA, April 2-5, 2014, Proceedings*, ed. R. Sharan (Cham: Springer International Publishing), 177–191. doi:10.1007/978-3-319-05269-4_15.

Moore, A., de Vos, J. M., Hancock, L. P., Goolsby, E., and Edwards, E. J. (2017). Targeted Enrichment of Large Gene Families for Phylogenetic Inference: Phylogeny and Molecular Evolution of Photosynthesis Genes in the Portullugo (Caryophyllales). *bioRxiv*. Available at: http://biorxiv.org/content/early/2017/06/04/145995 [Accessed June 19, 2017].

Moreno-Villena, J. J., Dunning, L. T., Osborne, C. P., and Christin, P.-A. (2018). Highly Expressed Genes Are Preferentially Co-Opted for C4 Photosynthesis. *Mol. Biol. Evol.* 35, 94–106. doi:10.1093/molbev/msx269.

Nelson, E. A., and Sage, R. F. (2008). Functional constraints of CAM leaf anatomy: tight cell packing is associated with increased CAM function across a gradient of CAM expression. *J. Exp. Bot.* 59, 1841–50. doi:10.1093/jxb/erm346.

Ni, M., Tepperman, J. M., and Quail, P. H. (1999). Binding of phytochrome B to its nuclear signalling partner PIF3 is reversibly induced by light. *Nature* 400, 781–784. doi:10.1038/23500.

Nimmo, G. A., Nimmo, H. G., Hamilton, I. D., Fewson, C. A., and Wilkins, M. B. (1986). Purification of the phosphorylated night form and dephosphorylated day form of phosphoenolpyruvate carboxylase from Bryophyllum fedtschenkoi. *Biochem. J.* 239, 213–20. doi:10.1042/BJ2390213.

Nimmo, H. G. (2000). The regulation of phosphoenolpyruvate carboxylase in CAM plants. *Trends Plant Sci.* 5, 75–80. doi:10.1016/S1360-1385(99)01543-5.

Nueda, M. J., Tarazona, S., and Conesa, A. (2014). Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics* 30, 2598–2602. doi:10.1093/bioinformatics/btu333.

Sanchez-Villarreal, A., Shin, J., Bujdoso, N., Obata, T., Neumann, U., Du, S.-X., et al. (2013). *TIME FOR COFFEE* is an essential component in the maintenance of metabolic homeostasis in *Arabidopsis thaliana*. *Plant J.* 76, n/a-n/a. doi:10.1111/tpj.12292.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–504. doi:10.1101/gr.1239303.

Silvera, K., Santiago, L. S., Cushman, J. C., and Winter, K. (2009). Crassulacean Acid Metabolism and Epiphytism Linked to Adaptive Radiations in the Orchidaceae. *Plant Physiol.* 149. Available at: http://www.plantphysiol.org/content/149/4/1838.short [Accessed June 21, 2017].

Silvera, K., Santiago, L. S., Cushman, J. C., and Winter, K. (2010). The incidence of

826 crassulacean acid metabolism in Orchidaceae derived from carbon isotope ratios: a checklist
827 of the flora of Panama and Costa Rica. *Bot. J. Linn. Soc.* 163, 194–222. doi:10.1111/j.1095-
828 8339.2010.01058.x.
829 Silvera, K., Santiago, L. S., and Winter, K. (2005). Distribution of crassulacean acid metabolism
830 in orchids of Panama: evidence of selection for weak and strong modes. *Funct. Plant Biol.*
831 32, 397. doi:10.1071/FP04179.
832 Silvera, K., Winter, K., Rodriguez, B. L., Albion, R. L., and Cushman, J. C. (2014). Multiple
833 isoforms of phosphoenolpyruvate carboxylase in the Orchidaceae (subtribe Oncidiinae):
834 implications for the evolution of crassulacean acid metabolism. *J. Exp. Bot.* 65, 3623–36.
835 doi:10.1093/jxb/eru234.
836 Somers, D. E., Schultz, T. F., Milnamow, M., and Kay, S. A. (2000). ZEITLUPE Encodes a
837 Novel Clock-Associated PAS Protein from Arabidopsis. *Cell* 101, 319–329.
838 doi:10.1016/S0092-8674(00)80841-7.
839 Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with
840 thousands of taxa and mixed models. *Bioinformatics* 22, 2688–90.
841 doi:10.1093/bioinformatics/btl446.
842 Sullivan, S., Jenkins, G. I., and Nimmo, H. G. (2004). Roots, Cycles and Leaves. Expression of
843 the Phosphoenolpyruvate Carboxylase Kinase Gene Family in Soybean. *Plant Physiol.* 135.
844 Available at: http://www.plantphysiol.org/content/135/4/2078.full [Accessed September 14,
845 2017].
846 Tallman, G., Zhu, J., Mawson, B. T., Amodeo, G., Nouhi, Z., Levy, K., et al. (1997). Induction
847 of CAM in Mesembryanthemum crystallinum Abolishes the Stomatal Response to Blue
848 Light and Light-Dependent Zeaxanthin Formation in Guard Cell Chloroplasts. *Plant Cell*
849 *Physiol.* 38, 236–242. doi:10.1093/oxfordjournals.pcp.a029158.
850 Taybi, T., Patil, S., Chollet, R., and Cushman, J. C. (2000a). A minimal serine/threonine protein
851 kinase circadianly regulates phosphoenolpyruvate carboxylase activity in crassulacean acid
852 metabolism-induced leaves of the common ice plant. *Plant Physiol.* 123, 1471–82.
853 Available at:
854 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=59103&tool=pmcentrez&render
855 type=abstract [Accessed October 10, 2015].
856 Taybi, T., Patil, S., Chollet, R., Cushman, J. C., Taybi, T., Patil, S., et al. (2000b). A Minimal
857 Serine / Threonine Protein Kinase Circadianly Regulates Phosphoenolpyruvate Carboxylase
858 Activity in Crassulacean Acid Metabolism-Induced Leaves of the Common Plant. *Plant*
859 *Physiol.* 123, 1471–1481.
860 Tseng, T.-S., and Briggs, W. R. (2010). The Arabidopsis rcn1-1 mutation impairs
861 dephosphorylation of Phot2, resulting in enhanced blue light responses. *Plant Cell* 22, 392–
862 402. doi:10.1105/tpc.109.066423.
863 Uhrig, R. G., Labandera, A.-M., and Moorhead, G. B. (2013). Arabidopsis PPP family of
864 serine/threonine protein phosphatases: many targets but few engines. *Trends Plant Sci.* 18,
865 505–513. doi:10.1016/J.TPLANTS.2013.05.004.
866 Wai, C. M., VanBuren, R., Zhang, J., Huang, L., Miao, W., Edger, P. P., et al. (2017). Temporal
867 and spatial transcriptomic and microRNA dynamics of CAM photosynthesis in pineapple.
868 *Plant J.* doi:10.1111/tpj.13630.
869 Walley, J. W., Sartor, R. C., Shen, Z., Schmitz, R. J., Wu, K. J., Urich, M. A., et al. (2016).
870 Integration of omic networks in a developmental atlas of maize. *Science* 353, 814–8.
871 doi:10.1126/science.aag1125.

Winter, K., and Holtum, J. A. M. (2011). Induction and reversal of crassulacean acid metabolism in *Calandrinia polyandra*: effects of soil moisture and nutrients. *Funct. Plant Biol.* 38, 576–582. Available at: https://doi.org/10.1071/FP11028.

Winter, K., Holtum, J. A. M., and Smith, J. A. C. (2015). Crassulacean acid metabolism: a continuous or discrete trait? *New Phytol.* doi:10.1111/nph.13446.

Yang, X., Hu, R., Yin, H., Jenkins, J., Shu, S., Tang, H., et al. (2017). The Kalanchoë genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. *Nat. Commun.* 8. doi:10.1038/s41467-017-01491-7.

Yin, H., Guo, H.-B., Weston, D. J., Borland, A. M., Ranjan, P., Abraham, P. E., et al. (2018). Diel rewiring and positive selection of ancient plant proteins enabled evolution of CAM photosynthesis in Agave. *BMC Genomics* 19, 588. doi:10.1186/s12864-018-4964-7.

Zambrano, V. A. B., Lawson, T., Olmos, E., Fernández-García, N., and Borland, A. M. (2014). Leaf anatomical traits which accommodate the facultative engagement of crassulacean acid metabolism in tropical trees of the genus Clusia. *J. Exp. Bot.* 65, 3513–23. doi:10.1093/jxb/eru022.

Figure 1 - A simplified diagram of the Crassalucean acid metabolism (CAM) pathway under day and night conditions. ALMT9 – aluminum activated malate transporter; CA – carbonic anhydrase; MDH – malate dehydrogenase; OAA – oxaloacetate; ME – malic enzyme (NAD or NADP); PEPC – phosphoenolypyruvate; PEPCK – PEP carboxykinase; PPCK – PEPC kinase; PPDK – pyruvate, phosphate dikinase.

Figure 2 – Gas exchange and titratable acidity for A) *Erycina pusilla* and B) *Erycina crista-galli*. Gas exchange is shown for a single plant; replicate plant gas exchange plots can be found in Supplemental Figure 1. Drought induction is indicated with a red arrow. Titrations are shown for dawn and dusk; full titratable acidity values are in Supplemental Table 2.

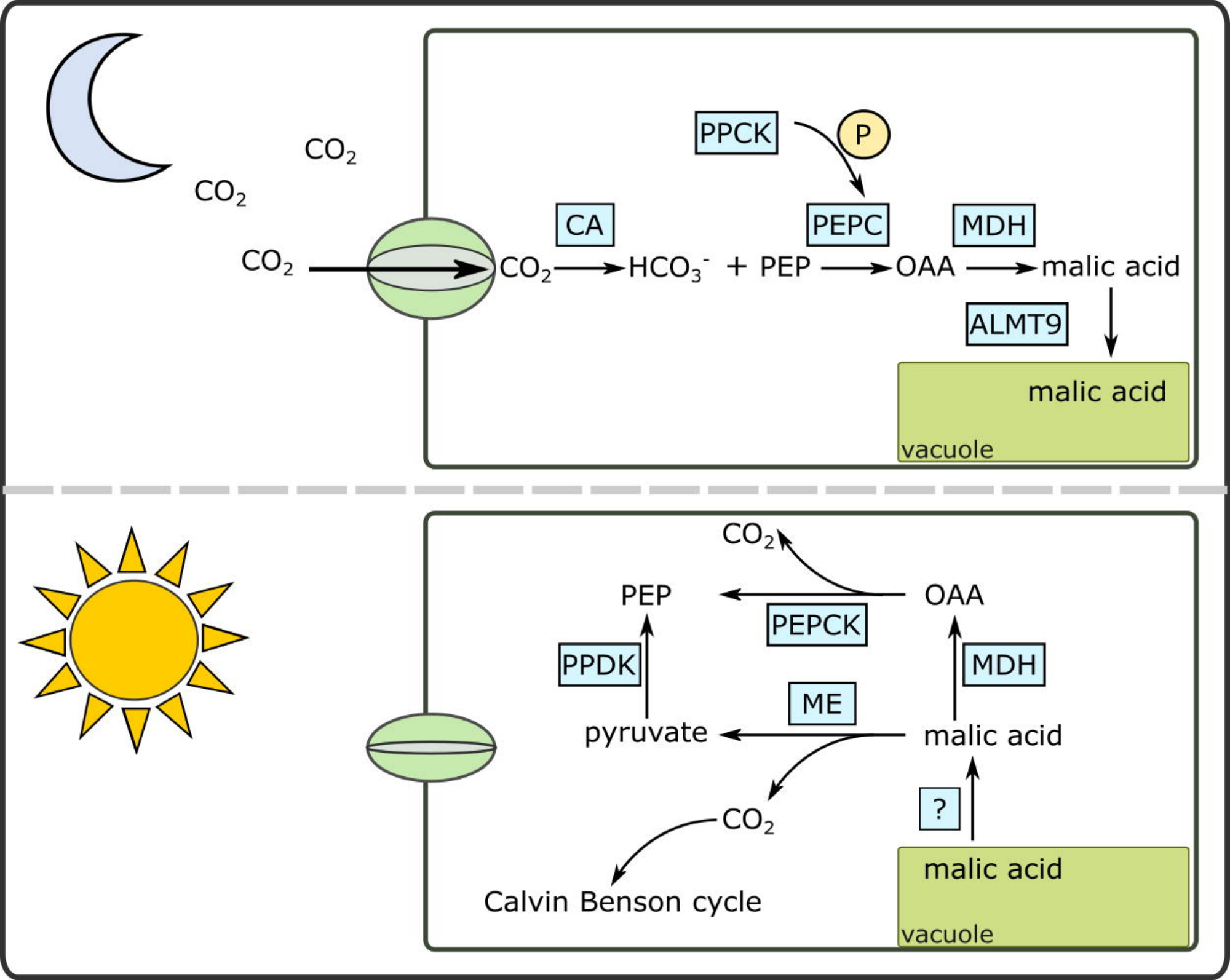
Figure 3 – Expression z-scores for each gene in each cluster for both A) *E. pusilla* and B) *E. crista-galli*, with median expression shown in the black line. Clusters represent co-expressed genes within each species' time-structured transcripts. Cool colors (blue and purple) are clusters with nighttime biased expression, whereas warmer colors (orange and yellow) are clusters whose transcripts increase in expression during the day.

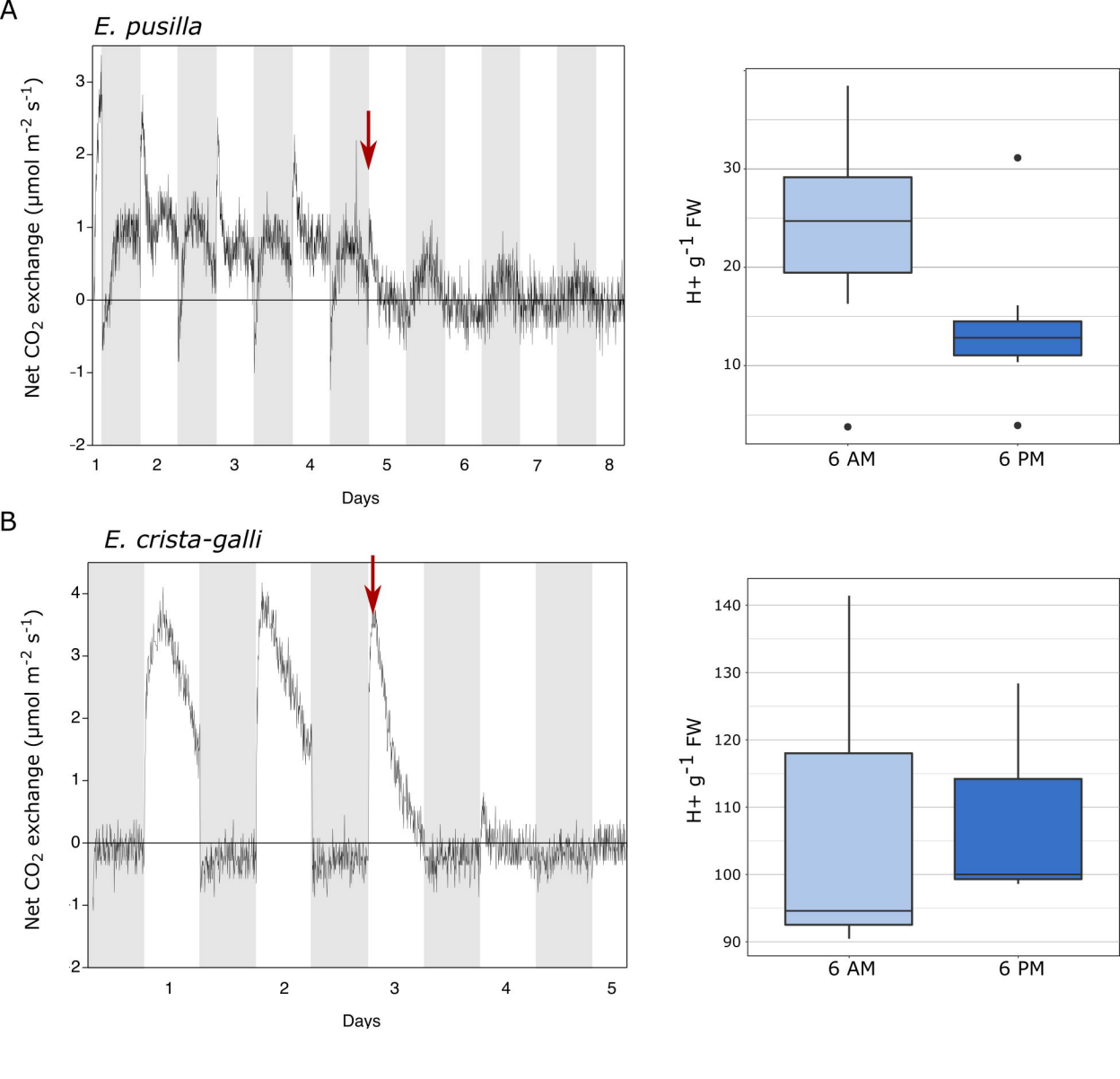
Figure 4 – Expression of A) phosphoenolpyruvate carboxylase (PEPC) and B) PEPC kinase in *E. pusilla* (squares, blue tones) and *E. crista-galli* (circles, purple tones). Different shades of color represent different assembled gene copies, and only transcripts that were found to be significantly time-structured are shown. Average TPM and standard deviation, scaled to the maximum mean TPM value across all copies and species, is plotted for all 6 time points, with the grey background indicating nighttime samples.

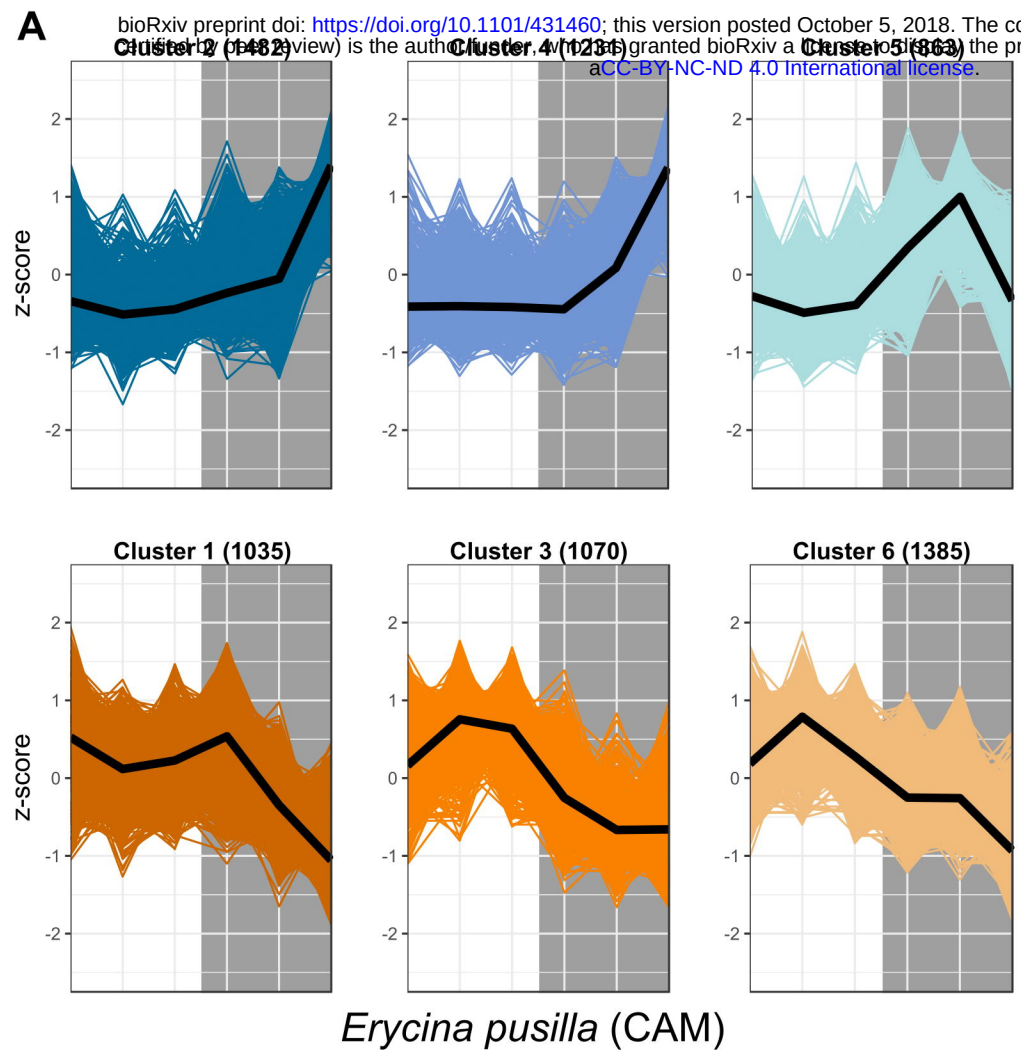
Figure 5 – Network constructions of A) *E. pusilla* and C) *E. crista-galli* based on ARACNe network analysis, and the ring finger seed longevity1 (RSL1) subnetworks for B) *E. pusilla* and D) *E. crista-galli* calculated via the diffusion algorithm in Cytoscape. Colors correspond to

915 Figure 3 cluster colors – cooler colors (blues in A and B and purples in C and D) are clusters that
916 have an increase in expression at night and warmer colors (oranges in A and B and yellows in C
917 and D) are clusters that have increases in expression during the day. Dots represent genes, scaled
918 by the number of edges (maximum number of directed edges=435).

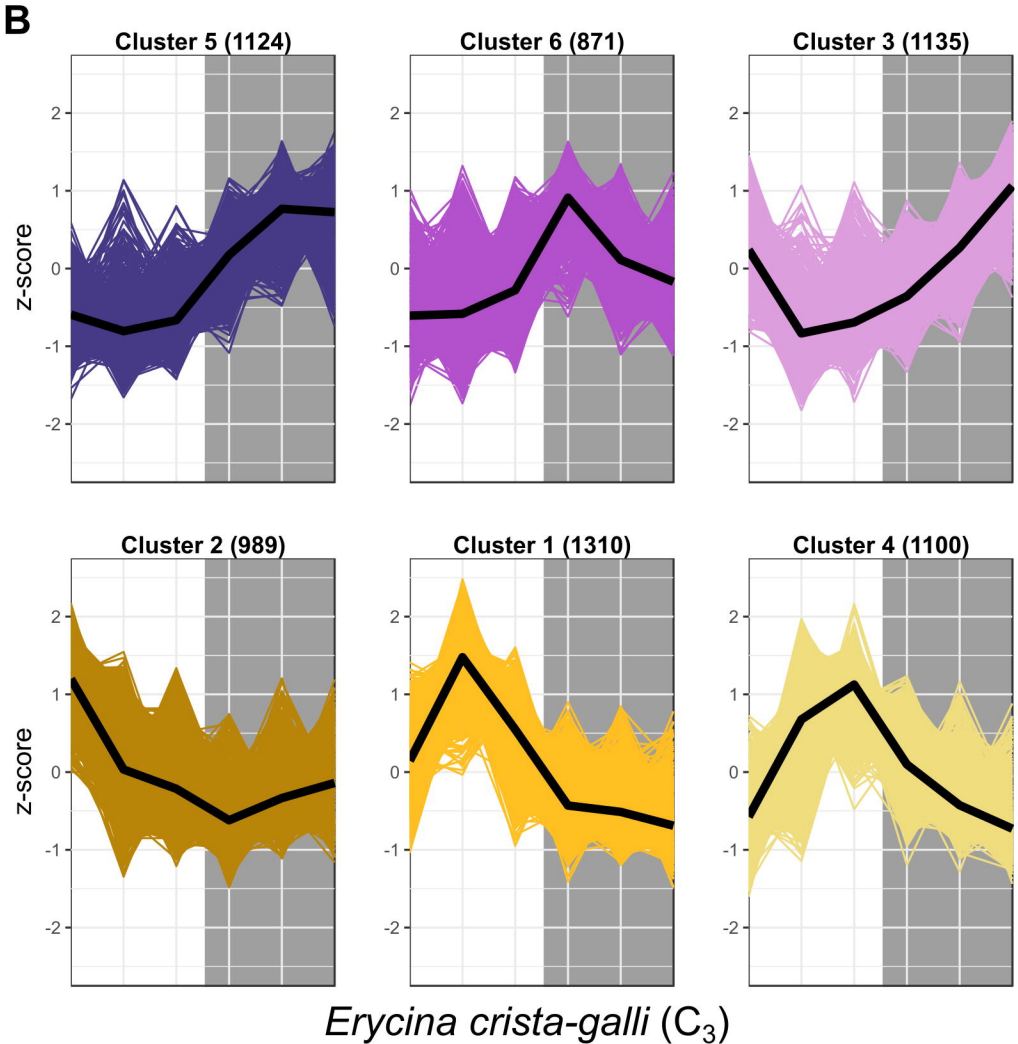
919 **Figure 6** - Expression of A) ring finger of seed longevity 1 (RSL1), B) phytochrome B/D
920 (PHYB) C) time for coffee (TIC), and D) protein phosphatase 2A in *E. pusilla* (squares) and *E.*
921 *crista-galli* (circles). Points are colored by which category of cluster they belong to: blues and
922 purples are genes with increases in expression in the dark in *E. pusilla* and *E. crista-galli*,
923 respectively, while oranges and yellows are genes with increases in daytime expression in *E.*
924 *pusilla* and *E. crista-galli*, respectively (see Fig. 3). Time-structured genes are marked with an
925 asterisk, while those that were not significantly time-structured are shown in greys. Number of
926 directed edges per gene is shown if they belonged to the RSL1-subnetwork. Average TPM and
927 standard deviation, scaled to the maximum mean TPM value across all copies and species, is
928 plotted for all 6 time points, with the grey background indicating nighttime samples.


CO2
CO2
CO2
CA
CO2
HCO3- + PEP
PPCK
P
PEPC
OAA
MDH
malic acid
ALMT9
malic acid
vacuole
CO2
PEP
PEPCK
OAA
PPDK
MDH
ME
pyruvate
malic acid
CO2
?
Calvin Benson cycle
malic acid
vacuole

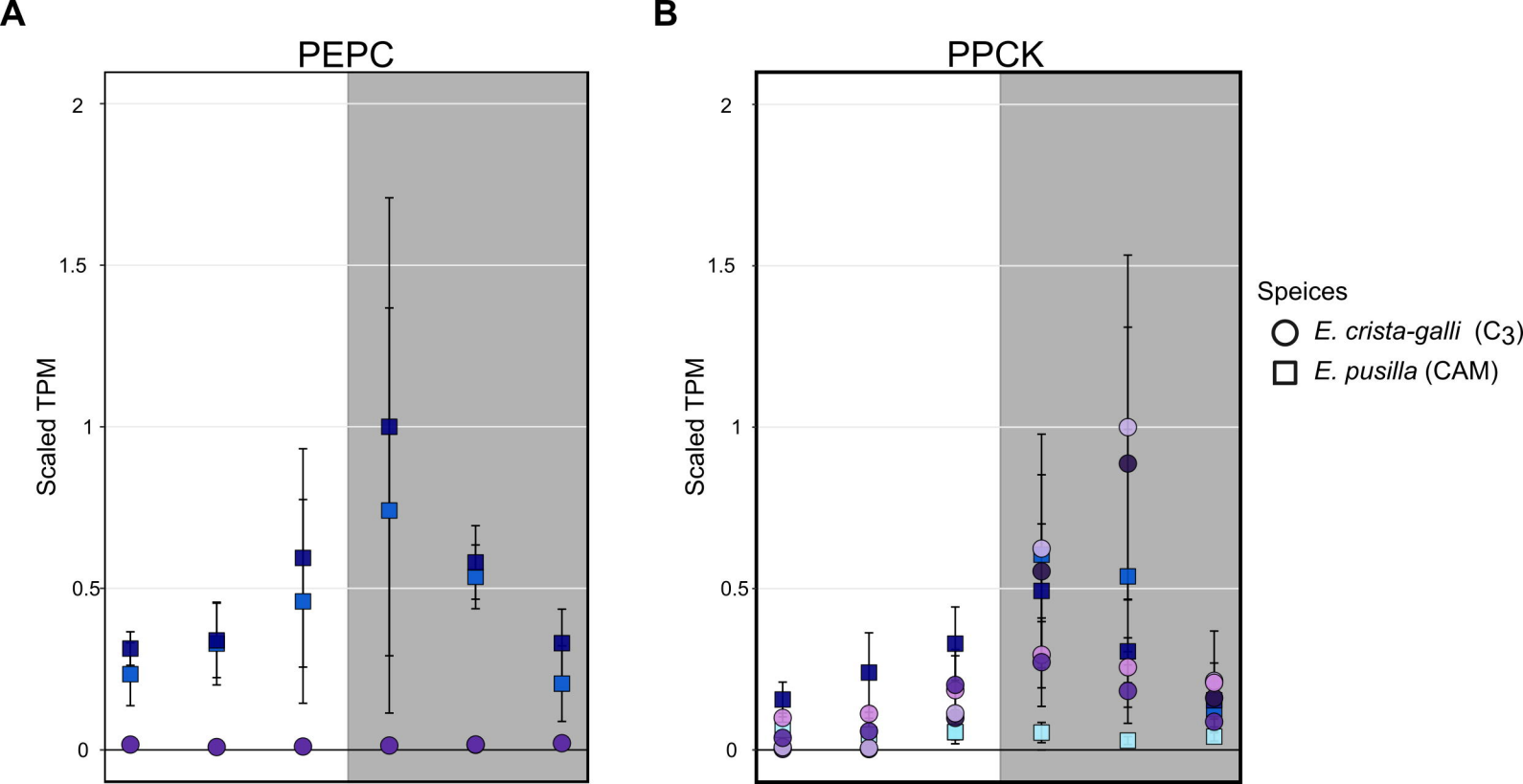

A
E. pusilla
Net CO2 exchange (µmol m-2 s-1)
Days
H+ g-1 FW
6 AM
6 PM
B
E. crista-galli
Net CO2 exchange (µmol m-2 s-1)
Days
H+ g-1 FW
6 AM
6 PM



Erycina pusilla (CAM)



Erycina crista-galli (C_3)


A
PEPC
B
PPCK
Scaled TPM
2
1.5
1
0.5
0
Speices
E. crista-galli (C3)
E. pusilla (CAM)