Title: Gene co-expression reveals the modularity and integration of C_4 and CAM in *Portulaca*

Ian S. Gilman[1], Jose J. Moreno-Villena[1], Zachary R. Lewis[1,2], Eric W. Goolsby[3], and Erika J. Edwards[*,1]

[1] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA; [2] Present address: Nanostring Technologies, Seattle, WA, USA; [3] Department of Biology, University of Central Florida, Orlando, FL, USA

Corresponding authors: Ian S. Gilman (ian.gilman@yale.edu)

Abstract

C_4 photosynthesis and Crassulacean acid metabolism (CAM) have been considered as largely independent adaptations in spite of sharing key biochemical modules. *Portulaca* is a geographically widespread clade of over 100 annual and perennial angiosperm species that primarily use C_4, but facultatively exhibit CAM when drought stressed, a photosynthetic system known as C_4+CAM. It has been hypothesized that C_4+CAM is rare because of pleiotropic constraints, but these have not been deeply explored. We generated a chromosome-level genome assembly of *P. amilis* and sampled mRNA from *P. amilis* and *P. oleracea* during CAM induction. Gene co-expression network analyses identified C_4 and CAM gene modules shared and unique to both *Portulaca* species. A conserved CAM module linked phosphoenolpyruvate carboxylase (PEPC) to starch turnover during the day-night transition and was enriched in circadian clock regulatory motifs in the *P. amilis* genome. Preservation of this co-expression module regardless of water status suggests that *Portulaca* constitutively operate a weak CAM cycle that is transcriptionally and post-transcriptionally upregulated during drought. C_4 and CAM mostly used mutually exclusive genes for primary carbon fixation and it is likely that nocturnal CAM malate stores are shuttled into diurnal C_4 decarboxylation pathways, but we find evidence that metabolite cycling may occur at low levels. C_4 likely evolved in *Portulaca* through co-option of redundant genes and integration of the diurnal portion of CAM. Thus, the ancestral CAM system did not strongly constrain C_4 evolution because photosynthetic gene networks are not co-regulated for both daytime and nighttime functions.

Main

Modularity and evolvability have become central research areas in biology over the past three decades[1,2]. Modularization generates simple building blocks that facilitate rapid and complex adaptations through the combination of more discrete functions and the ability of simpler units to explore mutational space with minimal pleiotropic effects[1–5]. Modern "-omics" strategies have enabled detailed resolution of gene regulation in diverse study systems, and modeling results have shown that carbon metabolic networks have high potential for exaptation into evolutionary innovations in bacteria[6]. Empirical research has identified deeply conserved carbon metabolic modules from anaplerotic pathways that have been repeatedly recruited into plant photosynthetic adaptations known as carbon concentrating mechanisms (CCMs)[7]. CCMs provide classic examples of rapid, adaptive evolution through the exaptation of existing metabolic modules that are subsequently refined for new functions.

Since the early Oligocene, CCMs—typically taking the form of C_4 photosynthesis or Crassulacean acid metabolism (CAM)—have evolved many dozens, if not hundreds, of times independently[8–12]. Certain facets of CCMs, such as the main biochemical pathways, are evolutionarily accessible to all green plant lineages because they belong to deeply conserved photosynthetic and respiratory gene networks[7,13]. It is thought that by coupling these gene networks to light responses (C_4)[14] and circadian oscillators (CAM)[15], CCMs create two-stage carbon fixation pathways that address the most fundamental ecophysiological tradeoff in land plants: balancing CO_2 influx with water loss to transpiration.

Most plants assimilate carbon using C_3 photosynthesis, in which CO_2 is directly fixed by RuBisCO during the day in mesophyll cells. However, RuBisCO can also bind to O_2, triggering a set of costly reactions called photorespiration, which is exacerbated by hot, dry conditions[16].

CCMs increase the efficiency of photosynthesis by increasing the concentration of CO_2 around RuBisCO. This is achieved by first fixing carbon with phosphoenolpyruvate (PEP) carboxylase (PEPC), which is either spatially (C_4) or temporally (CAM) decoupled from final fixation of CO_2 by RuBisCO (Fig. 1a, b). In both CCMs, CO_2 is first captured by beta carbonic anhydrase (BCA) and PEPC, forming oxaloacetate (OAA) (Fig. 1a, b) in the mesophyll, but exclusively during the night in CAM (Fig. 1b). OAA is then reduced to malate by malate dehydrogenase (MDH) and stored in the vacuole overnight as malic acid. During the day, CAM plants limit gas exchange with the external environment, and instead decarboxylate stored malate with malic enzymes (MEs) to release CO_2. Most C_4 species also convert OAA to malate (NADP-type biochemistry), but others first transform it to aspartate (NAD-type biochemistry) via aspartate aminotransferase (ASP). These 4-carbon acids are transported to the bundle sheath cells and decarboxylated. In NAD-type C_4 species, ASP reversibly transforms aspartate back into OAA in bundle sheath mitochondria, where it is reduced to form malate by NAD-MDH and finally decarboxylated by NAD-ME. By separating initial CO_2 capture in the mesophyll from assimilation in the bundle sheath, C_4 increases photosynthetic rates and water use efficiency while minimizing energy and carbon lost to photorespiration[17–19]. CAM does not always reduce photorespiration[20,21], but greatly increases water use efficiency[22,23] and reduces the harmful effects of high heat and radiation caused by oxidative stress[20].

C_4 and CAM boost the efficiency of photosynthesis in different ways and are therefore hypothesized to have evolved in response to different stressors. C_4 has been primarily considered as an adaptation to high rates of photorespiration[24,25], and CAM to water limitation[8,26], though photorespiratory stress and drought stress are often experienced simultaneously[27]. Given the overlapping environmental contexts of CAM and C_4 evolution, their repeated assembly of a

similar set of biochemical modules[28] (Fig. 1a, b), and their apparent ease of evolution, it is perhaps surprising that use of both C_4 and CAM (hereafter C_4+CAM) has only been reported in three lineages: *Portulaca* (Portulacaceae)[29], *Spinifex* (Poaceae)[30], and *Trianthema* (Aizoaceae)[31]. On the other hand, it may be precisely the large overlap in CCM gene networks and metabolites that may make C_4 and CAM mutually exclusive in most lineages[32]. Plants that use both C_4 and CAM seemingly must regulate the same gene networks in contrasting patterns that may result in futile metabolite cycling and inefficient transport. For example, daytime decarboxylation activity in the mesophyll during CAM would directly oppose C_4carboxylation activity.

Portulaca, a geographically widespread clade of over 100 annual and perennial herbs and small shrubs, predominantly use C_4, but facultatively exhibit CAM in response to drought[28,29,33–41]. It has been hypothesized that their CAM cycle must be either spatially separated from the C_4 cycle, that is, relegated to a subpopulation of leaf cells such as water storage or bundle sheath[32,35]; or alternatively, that *Portulaca* operate a novel two cell CAM system, whereby CAM malate is decarboxylated in the bundle sheath[36,37]. Facultative CAM, the reversible induction of CAM in response to drought stress, is assumed to be ancestral in *Portulaca* because it has been observed in every major sub-clade assayed[38,39] as well as in *Portulaca*'s closest relatives[42–47] (Fig. 1c). Transcriptomic studies of *Portulaca* and its close relatives suggest that the ancestral CAM pathway used both NAD- and NADP-enzymes for CAM[40,41,44]. The biochemical diversity of C_4 pathways in major *Portulaca* clades and the observation of intermediate photosynthetic phenotypes (C_3-C_4) in the Cryptopetala clade[39,48,49] imply multiple origins of C_4 within a facultative CAM context (Fig. 1c). Multiple C_4 origins are further evidenced by differences in leaf ultrastructure among lineages[48–50], as Kranz anatomy is typically established early in C_4 evolution[25,51].

Photosynthetic gene activity has only been studied in detail in *P. oleracea*[40,41], an NAD-type species. Similar experiments on other *Portulaca* are needed to answer fundamental questions about the function and evolution of C_4+CAM, and the integration of new metabolic modules more generally. Facultative CAM is hypothesized to be ancestral to *Portulaca*, but expression of CAM-specific orthologs in multiple species has not yet been documented. And although multiple carboxylation and decarboxylation pathways can co-exist in some C_4 plants[52], it is not obvious why or how different C_4 biochemistries would evolve in lineages that shared an ancestral, mixed CAM biochemistry type, nor do we know the extent to which these CCMs have been integrated in various lineages.

Here, we present a chromosome-level genome assembly of *Portulaca amilis* (NADP-type C_4), the first of any C_4+CAM plant. We analyzed transcriptomic data during a CAM-induction experiment to understand the synergistic evolution of multiple metabolic modules. Concurrent transcriptomic data for *P. oleracea* allowed us to compare C_4+CAM systems across a deep split in the *Portulaca* phylogeny and discriminate between ancestral, inherited and lineage specific gene networks. We confirm previous hypotheses that *Portulaca* species share one CAM-specific PEPC ortholog *(PPC-1E1c*) and another C_4-specific ortholog *(PPC-1E1a')*[28], and that CAM evolved by linking PEPC to starch catabolism via the circadian clock[13]. We found some evidence for CAM-specific decarboxylation pathways, but the diurnal portion of CAM appears to have been largely integrated into C_4 metabolism, and we predict that nocturnal acids produced by CAM are likely shuttled into the diurnal pool of bundle sheath C_4-acids. Finally, although *P. amilis* and *P. oleracea* share some central C_4 orthologs, exclusive use of orthologs from most core gene families highlights the diversity of C_4+CAM systems and provides further evidence that C_4 evolved largely independently in multiple *Portulaca* lineages.

Results

CAM induction experiment

Significant diel fluctuations in titratable leaf acidity indicative of CAM activity were observed after seven and eight days without water in *P. amilis* and *P. oleracea*, respectively (Fig. 1d). Diel fluctuations in titratable acidity were not significantly different from zero when well-watered (one sample *t*-test; $t_{amilis}(6) = -0.522$, $p_{amilis} = 0.624$; $t_{oleracea}(6) = -0.573$, $p_{amilis} = 0.591$), but were significantly different under drought (one sample *t*-test; $t_{amilis}(6) = 7.437$, $p_{amilis} = 0.00069$; $t_{oleracea}(6) = 10.136$, $p_{amilis} = 0.00053$), and significantly greater than when well-watered (Fig. 1d). There was no significant difference between the magnitude of diel fluctuations between species (independent *t*-test, $t(12) = -1.713$, $p = 0.121$).

Genome and transcriptome assembly and annotation

The *Portulaca amilis* v1 genome assembly was 403.89 Mb in length and extremely contiguous (L50/N50 = 5 scaffolds/42.598 Mb; L90/N90 = 9 scaffolds/39.183 Mb), with nine primary scaffolds representing the nine expected chromosomes of the haploid *P. amilis* genome (Fig. 2a) based on the karyotype of a close relative (*P. grandiflora*[53]). The scaffolds were 92.9% complete and single copy as measured by the BUSCO v5 embryophyta database (C:96.9% [S:92.9%, D:4.0%], F:1.5%, M:1.6%, n:1,614), and eukaryotic BUSCOs were similarly complete (C:98.0% [S:82.7%, D:15.3%], F:1.2%, M:0.8%, n:255). Slightly less than half of the assembly was masked as repetitive elements (46.96%), which primarily consisted of LTRs (41.96%), DNA elements (29.36%), and LINEs (13.85%) (Fig. 2b). A repeat landscape of these elements did not show any sudden bursts of element activity or multiple peaks (Fig. 2b), which can signify

genomic upheaval from hybridization or polyploidy events[54,55]. The final annotation contained 53,094 gene models with 58,732 unique coding sequences (Fig. 2c).

The *P. oleracea de novo* transcriptome assembly contained 413,658 transcripts corresponding to 230,086 "unigenes" (hereafter, "genes") that were highly complete but infrequently single copy: BUSCO v5 embryophyta (C:94.7% [S:16.4%, D:78.3%], F:3.2%, M:2.1%, n:1,614), eukaryota (C:98.9% [S:16.9%, D:82.0%], F:0.0%, M:1.1%, n:255). The high number of duplicate BUSCOs was expected, as most members of the *P. oleracea* species complex are polyploid[56]. Less than half of the transcripts (198,026) were predicted to be coding sequences, which corresponded to 45,285 genes.

A total of 21,621 *P. amilis* and 44,532 *P. oleracea* genes passed filtering for differential abundance analysis[57], and over half of these genes (*P. amilis*, 62%; *P. oleracea*, 51.0%) were found to be significantly differentially abundant between well-watered and drought conditions following Benjamini Hochberg corrections[57] for false discovery ($q < 0.01$; see Methods for details).

Co-expression network analysis

The well-watered and drought *P. amilis* photosynthetic gene networks (PGNs) were similar in size and density, but had different distributions of node degree due to a small number of extremely highly connected genes belonging to modules paWW1 and paWW3 (Fig. 3a, Supplementary Fig. 1c, and Supplementary Table 1). The 104 genes of paWW1 represented all functional groups except starch synthesis, and paWW1 was characterized by C_4-like expression: very high morning abundance that rapidly tapered off by late afternoon. Indeed, paWW1 contained at least one homolog of all core elements of the C_4 pathway; including *paPPC-1E1a'*, the C_4-specific *PPC* ortholog used by *P. oleracea*[28,40,41] (Fig. 3c). The most similar drought

module to paWW1 in terms of constituent genes, paD6, was 31% smaller (72 genes), while the largest drought module (paD1) was characterized by dark period peak expression and consisted primarily of light response and circadian transcription factors, and starch catabolism-related genes (Fig. 3b). Similar to *P. amilis*, the largest well-watered *P. oleracea* module (poWW2) contained *poPPC-1E1a'*; however, the *P. oleracea* transcriptome contained five *poPPC-1E1a'* transcripts distributed between two modules with C_4-like expression (poWW2 and poWW1) (Fig. 3e). There was a greater size discrepancy between the *P. oleracea* PGNs, and the drought PGN was much larger and denser than the well-watered PGN (Fig. 3e, f, and Supplementary Table 1).

We expected the CAM cycle to be broken into two co-expression networks: one that builds PEP from transitory starches and carboxylates CO_2 into malate via BCA, PEPC, and MDH from dusk until dawn, and a second that decarboxylates malate into CO_2 for fixation by RuBisCO during the day (Fig. 1b). As reported in *P. oleracea*[28,40,41], *P. amilis* used the *PPC* paralog *paPPC-1E1c* for CAM (Fig. 3d), and this gene belonged to smaller modules with abundance profiles that peaked across the light-dark transition, and were dominated by starch degradation and circadian clock genes (paWW5 and paD4; Fig. 3a, b, d). Extracting the genes that remained co-expressed (preserved) across well-watered and drought conditions resulted in 28 genes of the *paPPC-1E1c* modules (42%) (Fig. 4a). With the exception of one photorespiratory gene (*paCAT1-2*, FUN_025647), all had functions related to carboxylation (*paPPC-1E1c* and *paBCA1-2*), starch metabolism and catabolism, or circadian rhythm and light response. At least one paralog of all members of the starch degradation pathway were preserved, except amylase (*AMY*) and beta-amylase (*BAM*), but one *BAM* homolog (*paBAM3-2*, FUN_041354) was recovered exclusively in the drought module, along with one paralog of

191 aluminum activated malate transporter 4 (*paALMT4-2*, FUN_033599), which imports malate into
192 the vacuole. However, no *MDH* homologs were recovered in this nocturnal CAM network, and
193 most were recovered in drought modules paD6, paD2, and paD5 with C_4-like expression.
194 The *P. oleracea* transcriptome contained multiple transcripts of *poPPC-1E1c* that were
195 recovered in two well-watered modules (poWW4 and poWW8) and three drought modules
196 (poD2, poD5, and poD9) that were consistent with use in CAM (Fig. 3e, f). The preserved genes
197 were functionally similar to those preserved in the *P. amilis* nocturnal CAM network, and were
198 involved with starch metabolism and catabolism, and the circadian clock (Fig. 4b). Among the
199 genes preserved with *poPPC-1E1c* was cytoplasmic *poNAD-MDH1*
200 (TRINITY_DN7462_c3_g1), which was the most highly expressed *NAD-MDH* under drought.
201 Furthermore, many of the genes that link phosphorylative starch degradation to PEP generation
202 were preserved in the nocturnal CAM network, including *AMY*, fructose-bisphosphate aldolase
203 (*FBA*), glyceraldehyde-3-phosphate dehydrogenase (*GAP*), and phosphofructokinase (*PFK*).
204 Although not preserved, the *poPPC-1E1c* drought modules included copies of
205 phosphoglucomutase (*PGM*) and enolase (*ENO*), the final two steps in PEP generation from
206 starch or soluble sugars.
207 We used the *paPPC-1E1c* and *poPPC-1E1c* drought modules to identify the set of genes
208 common to both nocturnal CAM networks (Fig. 4c, d). By taking the intersection of the two sets
209 of *PPC-1E1c* drought module genes (i.e., preserved and drought exclusive genes), we uncovered
210 21 *P. amilis* and 24 *P. oleracea* orthologs (some genes had multiple paralogs in the *P. oleracea*
211 network). These orthologous nocturnal CAM networks contained one gene involved in starch
212 synthesis (starch branching enzyme 3, *SBE3*), five phosphorylative starch degradation genes, and
213 alpha-glucan phosphorylase 2 (*PHS2*), which is also part of phosphorylative starch metabolism

and catabolism (Fig. 4e). These orthologous nocturnal CAM networks covered much of the phosphorylative starch degradation pathway that forms the basis of PEP generation in most CAM plants[58,59], but the interactions among many genes were different between species (Fig. 4c, d). The largest group of shared genes were core circadian clock elements: *Adagio* (*ADO*), also known as *ZEITLUPE* (*ZTL*), *1* and *3*, two *Arabidopsis pseudo-response regulator 5* (*APRR5*) paralogs, *EARLY FLOWERING* (*ELF*) *1* and *4*, *GIGANTEA* (*GI*), one *LUX ARRHYTHMO* (*LUX*) paralog, and one *REVEILLE 6* (*RVE6*) paralog (Fig 4c, d, e). Although the two species' nocturnal CAM networks shared many key orthologs, there were notable differences. The *P. oleracea* network did not contain an *ALMT* homolog, but did contain three paralogs of *PEPC kinase 1* (*PPCK1*) that were all highly abundant and significantly upregulated under drought ($q < 1.0 \times 10^{-7}$) (Supplementary Fig. 4). PPCK increases the activity of PEPC, which is inhibited by malate, forming a negative feedback loop[60]. While all three *poPPCK* transcripts were upregulated under drought with peak nighttime expression indicative of use in CAM, all had higher daytime levels of abundance under drought as well—which points towards a dual role in both C_4 and CAM. The single *PPCK* ortholog in *P. amilis* (FUN_002684) was also significantly more abundant under drought ($q = 0.0011$) and had peak nocturnal expression (Supplementary Fig. 3), but it was recovered in drought module paD1, which increased in abundance throughout the night (Fig. 3b). PPCK activity is typically essential for highly active CCMs, and a CAM-like expression pattern in *paPPCK1* would suggest low PEPC activity during C_4. However, we found a derived amino acid residue, glutamic acid (E), in *paPPC-1E1a'* at the same location where an aspartic acid (D) residue (D509) was demonstrated to reduce malate inhibition in *Kalanchoë* and *Phalaenopsis* orchids[61]. In *Kalanchoë* and *Phalaenopsis*, the D509 residue was shown to be derived from either arginine (R), lysine (K), or histidine (H), the latter of which is present at

position 509 in *poPPC-1E1a'*. Glutamic acid is functionally similar to aspartic acid, and may similarly reduce malate inhibition; thus, allowing for high PEPC activity during C_4 without the need for PPCK.

We considered two modules each in *P. amilis* (paWW1 + paWW3 and paD5 + paD6) and *P. oleracea* (poWW2 + poWW1 and poD3 + poD4) when calculating the preservation of the C_4 pathway because these pairs of modules had C_4-like expression patterns and were strongly connected (Fig. 3a, e). In the *P. amilis* C_4 PGNs, 86 genes (38%) were preserved (Supplementary Fig. 4), which represented of all core C_4 (*ALAAT*, *ASP*, *BCA*, *ME*, *MDH*, *PPC-1E1a'*) and PEP regeneration gene families (adenylate kinase (*ADK*), sodium/pyruvate cotransporter (*BASS*), sodium-hydrogen antiporter (*NDH*), soluble inorganic pyrophosphatase (*PPA*), pyruvate, phosphate dikinase (*PPDK*), and PEP/phosphate translocator (*PPT*)). It also contained members of the entire transitory starch pathway that builds and transports various sugars and carbohydrates (e.g., *ENO*, *GAP*, phosphoglycerate kinase (*PGK*), *PGM*, and triose phosphate/phosphate translocator (*TPT*)). In addition to *paBASS2-1* (FUN_026309), which imports pyruvate into the chloroplast as a substrate for PEP generation, we recovered two paralogs of methyl erythritol phosphate (*MEP*), also known as *RETICULATA-RELATED* (*RER*), *4* (*A. thaliana* ortholog AT5G12470). MEP transporters have been implicated as mesophyll pyruvate importers and bundle sheath pyruvate exporters in maize[62,63], and thus may serve a similar function to BASS. The preserved modules also contained key enzymatic regulators (e.g., serine/threonine-protein phosphatase (*PP2A*) subunits and RuBisCO activase (*RCA*)) and transcription factors and clock elements (e.g., *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*)).

Portulaca amilis and *P. oleracea* span a fairly deep node in the *Portulaca* phylogeny (Fig. 1c) and it was unclear which facets of C_4 they might share. A similar fraction of C_4 module

genes were preserved between well-watered and drought C_4 PGNs in *P. oleracea* (117; 41.2%), and these largely consisted of the same gene families (58.3% overlap with preserved *P. amilis* C_4 gene families). To identify the orthologous components of the *P. amilis* and *P. oleracea* C_4 gene networks, we took the intersection of their well-watered C_4 PGNs, which contained 186 and 165 homologs, respectively, that represented 258 unique ortholog groups. Roughly one quarter (69/258; 26.7%) of these ortholog groups were recovered in the common C_4 networks, while 117 (45.3%) and 72 (27.9%) were exclusive to *P. amilis* and *P. oleracea*, respectively. The common networks included many of the most highly expressed copies of key C_4 genes, including the PEP generation pathway, and most of the photorespiratory pathway (Supplementary Fig. 5).

Indicative of the unique evolutionary histories of *P. amilis* and *P. oleracea*, each species' C_4 PGN contained exclusive orthologs of most C_4 gene families, some of which represent species-specific duplication events. Both species exhibited exclusive use of orthologs of *ADK*, *ALAAT*, *ASP*, *BCA*, *MDH*, *ME*, *NDH*, *PP2A*, *PPA*, mitochondrial uncoupling protein (*PUMP*), RuBisCO, and *RCA*, among others. Some exclusive homologs fit biochemical expectations. Exclusive to *P. amilis* (NADP-type C_4) were three chloroplastic *NADP-ME4* paralogs (FUN_006401, FUN_006404, and FUN_006411) which did not have orthologs in the *P. oleracea* transcriptome. These three copies were on the same strand and physically close in the genome (within ~85kb); they possibly represent tandem duplication events of *paNADP-ME4-4* (FUN_006423), which is ~50kb downstream of FUN_006411 (Fig. 2a). In *P. oleracea*, two *ASP* (TRINITY_DN71325_c0_g1 and TRINITY_DN346_c1_g3) and three *ALAAT* (TRINITY_DN43865_c0_g1, TRINITY_DN43865_c0_g2, and TRINITY_DN81553_c0_g1) homologs were found in orthogroups that had no *P. amilis* members. Many of the genes that were exclusive to either C_4 PGN were circadian clock and light response transcription factors, or

regulatory enzymes. For example, unique orthologs of the *FAR-RED IMPAIRED RESPONSE* family were part of each C_4 PGN, while only the *P. amilis* C_4 PGN contained an *APRR* homolog, and only the *P. oleracea* C_4 PGN contained a *LATE ELONGATED HYPOCOTYL* (*LHY*) transcription factor.

We used the drought C_4 PGNs to identify orthologous ancestral elements of the daytime portion of the CAM cycle. We expected this portion of the CAM cycle to be nearly indistinguishable from the C_4 cycle (Fig. 1a, b), but to only exhibit C_4-like expression under drought stress. To be considered part of the ancestral CAM cycle, elements needed to both be shared between the two drought C_4 PGNs and be significantly upregulated under drought ($q <$ 0.05). Eleven orthologs met these potentially CAM-induced criteria, but only one had a function directly related to daytime CO_2 release from nocturnal malate stores: chloroplastic *NADP-ME4-5* (FUN_015913; TRINITY_DN81025_c0_g1). *PUMP5-1*—also known as dicarboxylate transporter 1 (*DIC1*)—of both species also appeared CAM-induced, but was not significantly differentially abundant when considering the entire time series in *P. oleracea* ($q = 0.24$). The *P. amilis* C_4 PGNs contained a drought exclusive mitochondrial *NAD-ME2-1* (FUN_037145) that was significantly upregulated ($q < 0.01$), and a second, preserved ortholog (*NAD-ME1-1*; FUN_021996) that was significantly upregulated under drought ($q < 0.05$). Coupled with upregulated *PUMP5-1* expression, this suggests that the NAD decarboxylation pathway, which occurs mainly in the mitochondria, may still be elicited by CAM in *P. amilis* despite NADP-type C_4. In contrast, constitutive expression of the same mitochondrial orthologs in *P. oleracea* (*poNAD-ME1-1* and *poNAD-ME2-1*) is consistent with use in C_4, suggesting that *P. oleracea* may have recruited its ancestral CAM NAD-decarboxylating pathway into its C_4 system. Finally, and unexpectedly, one *P. amilis* paralog of the RuBisCO small subunit B (*paRBCS-B-1*;

306 FUN_016561) fit a CAM induced abundance pattern ($q < 0.001$) and was expressed on the same
307 order of magnitude as other *RBCS* homologs.
308
309 *Motif identification and enrichment*
310 Enriched motifs were detected in at least one region (i.e., 3'UTR, 5'UTR, introns, or upstream
311 promoter) for gene models in many *P. amilis* modules. However, no enriched motifs were
312 identified for paWW1 or paWW3—which contained the bulk of C_4 genes when well-watered. In
313 contrast, the 5'UTRs and upstream promoter regions of the preserved *paPPC-1E1c* module were
314 enriched in 10 unique motifs ($p < 0.01$), all of which had circadian or light response functions
315 (Table 1). These motifs were all annotated to Myb-related families, except bHLH74, which
316 contains a G-box motif (CACGTG), and governs flowering time in response to blue light[64]. All
317 enriched motifs contained either an Evening Element (EE; AAATATCT) or GATA (HGATAR)
318 motif. EE and GATA motifs often serve as binding sites for core circadian oscillator
319 transcription factors such as RVEs, CCA1, LHY, and APRR1, also known as TIMING OF CAB
320 1 (TOC1)[65–67].
321 There were no enriched motifs in or near genes preserved with *paPPC-1E1a'*, and the *P.*
322 *amilis* C_4 PGN genes were only enriched in four motifs: one Myb (MA1028.1), two bZIP
323 (MA0968.1, MA1033.1), and one AP2/ERF (MA1261.1) factor (Supplementary Table 2). While
324 these motifs were significantly enriched ($p < 0.01$), they were much less widespread than the
325 enriched motifs found in the preserved CAM genes. At the low end, the KAN4 motif
326 (MA1033.1) was present in just five 5'UTRs (8.5% of sequences), and the most widespread
327 (bZIP68, MA0968.1) was present in 12 sequences (20.3%). Motifs found in the preserved
328 *paPPC-1E1c* genes were present in 20.7-58.6% of the sequences, and all 10 enriched motifs
329 were found at least once in the 5'UTR and promoter region of *paPPC-1E1c*.

Discussion

Plant photosynthetic adaptations can serve as models for exploring central themes of biology, such as the evolution of modularity and evolvability, gene and genome duplication, convergent evolution, and the evolution of novel phenotypes. Network analyses provide powerful tools for identifying and comparing modules of genes related to functions of interest against a background of tens of thousands of genes with complicated temporal expression patterns. C_4+CAM photosynthesis has only been reported in three plant lineages, *Portulaca*[29], *Spinifex*[30], and *Trianthema*[31], but CCMs have evolved many dozens of times[8–12] and the use of multiple CCMs has been documented in aquatic plant lineages[68,69]. CCM evolution is a classic example of the repeated assembly of a new function via the integration of multiple, existing gene networks[7]. When coupled with patterns of transcript abundance and placed in an evolutionary context, our network results provide an initial interpretation of how these metabolic pathways interact in *Portulaca,* and how they were evolutionarily assembled.

The evolution of C_4+CAM in Portulaca

Facultative CAM is ancestral in *Portulaca*, and likely evolved by coupling elements of the TCA cycle (i.e., PEPC and BCA) to the phosphorylative starch degradation pathway via the circadian clock. CAM in both species was restricted to a minimal set of tightly regulated nocturnal pathways that linked *PPC-1E1c* to the transitory starch pathway, with only parts of the decarboxylation pathways distinguishable from C_4. This aligns with previous hypotheses that the initial steps in CAM evolution do not require "rewiring" of gene networks, but rather occur by upregulating nocturnal organic acid production that occurs in all plants[13]. Subsequent canalization of these fluxes occurs by bringing these networks under circadian control. If C_4

354 evolution began in *Portulaca* while the nighttime portion of CAM included derived gene
355 networks, but the daytime portion did not, CAM would have likely imposed few pleiotropic
356 constraints given redundant paralogs, because recruitment of genes into C_4 would not have
357 required modification of existing CAM decarboxylation pathways in the mesophyll. Sufficient
358 paralogs of NAD- and NADP-type enzymes were present in the most recent common ancestor of
359 *P. amilis* and *P. oleracea* to allow C_4 and CAM to use separate genes, and for differential
360 enzyme co-option during their independent C_4 origins. However, transcriptomic data presented
361 here and elsewhere[40,44] imply that CAM-specific decarboxylation pathways did exist in the
362 common ancestor to *Portulaca*. It therefore seems likely that the initial steps towards C_4 evolved
363 within a particular window of CAM evolution before other aspects of CAM were fixed, such as
364 diurnal stomatal closure or changes to leaf ultrastructure that greatly reduced intercellular
365 diffusion of metabolites.

366 Shared use of some enzymes and pathways suggests that the most recent common
367 ancestor of *Portulaca* may have had some C_4 characters. Shared use of photorespiratory
368 modules, found here, and proto-C_4 anatomy in C_3-C_4 intermediate taxa[49,50] suggest that C_2
369 metabolism—the enhanced refixation of photorespired CO_2 by preferential expression of various
370 photorespiratory enzymes[25,70]—may also be ancestral to *Portulaca*. Like other CCMs, C_2
371 requires the recruitment of existing gene networks, and our comparison of well-watered C_4 gene
372 networks revealed many shared photorespiratory orthologs that may point to a single origin.
373 However, it is difficult to distinguish between a scenario in which these orthologs were co-opted
374 into C_2 metabolism once near the stem of extant *Portulaca*, or more recently along multiple
375 branches in parallel. Shared use of PEPC-1E1a' in C_4 has been hypothesized to have occurred
376 through parallel recruitment[28], stemming from "preconditioning", such as increased expression[71]

or adaptive amino acid substitutions[29]. C_2 metabolism is hypothesized to typically be followed by rapid evolution of a C_4 system[25,70], and therefore represents a limit of what we expect to be shared among *Portulaca* lineages with distinct morphologies and biochemistries. Broader sampling across the *Portulaca* phylogeny is needed to precisely delimit the shared initial steps towards C_4 and subsequent parallel refinement of C_4+CAM.

Integration of C_4 and CAM in Portulaca

Portulaca amilis and *P. oleracea* have canonical C_4 pathways, but constitutively operate very weak CAM cycles (at least at the transcriptional level) that are upregulated under drought. The core elements of C_4 and CAM were moderately preserved in both species, and where parts of the C_4 and CAM pathways could theoretically overlap, we often found alternative homologs. Increased CAM activity was associated with a reduction in size of C_4 PGNs, but there was no evidence that this was due to dual roles of genes in nocturnal CAM physiology under drought. If transcript abundances of C_4 and CAM related genes can serve as a proxy for enzymatic and metabolic activity, conflicting intracellular physiological cues from CAM likely played a minor role in altering C_4 activity. Given the strong diffusion gradients between mesophyll and bundle sheath of C_4 species[72], we suspect that the evolution of C_4 in *Portulaca* has resulted in two-cell CAM, as has been previously hypothesized[37], where malate produced by CAM in the mesophyll is largely decarboxylated in the bundle sheath, but it is unclear if overnight storage occurs in the mesophyll, bundle sheath, or both (Fig. 5). Drought-induced upregulation of decarboxylating enzymes was consistent with expectations given C_4 biochemistries. Upregulation of *paNAD-ME1-1* and *paNAD-ME2-1* during CAM induction in *P. amilis* indicated that C_4 did not co-opt the entirety of the diurnal portion of the CAM cycle (Fig. 5). In contrast, constitutive use of *poNAD-ME1-1* and *poNAD-ME2-1* in *P. oleracea* was consistent with recruitment and

integration of the ancestral CAM NAD-decarboxylating pathway into C_4, while the NADP-pathway maintained its role in CAM (Supplementary Fig. 6).

Our network results confirmed that the nocturnal portion of CAM is a fixed gene network—rather than the result of drought-induced rewiring—that is both transcriptionally and post-transcriptionally regulated under drought to increase CAM activity. Gas exchange measurements in *Portulaca* species show substantial net CO_2 loss during the night when well-watered[39–42]. Therefore, well-watered expression levels of the preserved CAM-related genes are either not high enough to re-fix all nocturnally respired CO_2 or must have their products regulated at a post-transcriptional level. Evidence for the latter is given by expression patterns of PPCK, which increases the activity of PEPC[60]. *PPCK* was upregulated in both species under drought, and during the dark period in particular, which implies a role in increasing nocturnal PEPC activity for CAM under drought. In *P. oleracea*, the maintenance of high daytime *poPPCK* expression with a nocturnal peak in abundance suggests that PPCK may play a dual role in both C_4 and CAM. In contrast, we were surprised to find that *paPPCK* had expression only consistent with use in CAM, as we are not aware of other C_4 species without high morning expression of *PPCK*; but a derived amino acid residue in PEPC-1E1a' has the potential to increase its activity by reducing malate inhibition, as has been show in *Kalanchoë* and *Phalaenopsis*[61]. Therefore, the avoidance of pleiotropic effects has occurred through redundant homologs (e.g., decarboxylating enzymes) and adaptive mutations, but *P. oleracea* provides evidence that pleiotropy may not constrain some genes.

Portulaca as a model for metabolic evolution

The natural history of *Portulaca* provides a compelling model system for studying many evolutionary and ecophysiological questions, and its unique metabolism and experimental

practicality make it a powerful tool for genetic engineering. Forward engineering of both C_4[73] and CAM[74] have become priorities for agriculture as arid lands expand and populations grow. *Portulaca* comprises over 100 species that are found in nearly every country and have invaded some continents multiple times[56]. Despite the small genome size of *P. amilis*, duplication events deeper in the history of the Portulacineae[75] have played large roles in the evolution of C_4+CAM, with implications for neofunctionalization and the evolution of modularity. With the generation of the first *Portulaca* genome, we provide a resource for studying these past events and for the functional genetic work to understand and engineer CCMs.

CAM is found in every major clade of the Portulacineae to varying degrees[34,43,45,46,76,77] (Fig. 1c), and stronger and constitutive CAM have evolved in *Portulaca*'s closest relatives, Anacampserotaceae and Cactaceae[43,78,79]. These three clades overlap in climate space[77] and in many parts of their geographic ranges[80], exhibit succulent morphologies capable of photosynthesis in their stems[43,81], and likely share the same ancestral CAM cycle. It remains an open question why only *Portulaca* evolved C_4+CAM, while stronger CAM evolved in close relatives. C_4 and CAM consume more energy than C_3 photosynthesis, and we lack models to predict environmental conditions that would select for C_4 in a CAM plant (or vice versa). Selective pressures to evolve C_2 in the facultative CAM ancestor of *Portulaca* may have been no different than in C_3 plants, as CAM does not necessarily reduce photorespiration, and in some cases can exacerbate it[20]. Future progress on these topics will be made through continued study of the C_3-C_4 Cryptopetala clade that uses C_2 metabolism.

Despite an increased general interest in CCMs over recent years, and *Portulaca*'s C_4+CAM system in particular, we have only a limited understanding of the benefits of combining C_4 and CAM. The advantages of facultative CAM vary between lineages, as does

carbon gain, which is typically <10% of the daytime photosynthetic activity[82,83]. Facultative CAM increases water use efficiency and can prevent reduced growth and reproductive output under drought while also reducing photoinhibition (radiation induced damage to photosystems)[82,83]. Measures of chlorophyll *a* in *P. oleracea* over a water limitation experiment provide evidence that facultative CAM reduces photoinhibition[40]. A photoprotective role of CAM is further corroborated by rapid recovery of photosynthetic rate upon rewatering[38,39], indicating little or no damage is done to photosystems. Gas exchange measurements also demonstrate that *Portulaca* continue to grow despite several days of drought stress[38,39]. Furthermore, drought stress may increase CO_2 leakiness of bundle sheath in C_4 plants, a problem augmented when carboxylation activity decreases relative to decarboxylation[84]. Facultative CAM may provide some carbon and photoprotection during drought that accelerates the transition to full photosynthetic capacity upon rewatering and allows for growth and reproduction despite large water deficits.

The annotated *P. amilis* genome and metabolic modules identified here, combined with recent resources for transgenic research in *Portulaca*[85] signal a significant shift toward the development of an experimental, functional genomics-based research program in C_4+CAM. On one level, we suggest that *P. amilis* is a better candidate than *P. oleracea* for further model system development, in part because *P. oleracea* exists within the context of a larger species complex of over a dozen described species and subspecies with uncertain relationships and multiple whole genome duplication events[56]. *P. amilis* has an indistinguishable life history from *P. oleracea* (i.e., 3-5 week life cycle, thousands of seeds produced, self-fertilization), but does not suffer from the same taxonomic challenges and variations in ploidy. On the other hand, we view the *Portulaca* clade as providing not one, but at least three, evolutionary origins of

C_4+CAM systems, and there is much to be learned in understanding their similarities and differences.

Conclusions

The evolution of modularity is a common theme in studies of evolutionary development and the origination of novel phenotypes. Network-based analyses can distinguish complex and highly overlapping metabolic modules, and establish their orthology between species. C_4+CAM provides a unique opportunity to understand how a single module that is typically recruited into two distinct functions is regulated when both of the new functions operate in a single organism. Remarkably, their integration appears possible in part via additional modularization, with little overlap in C_4 and CAM gene networks. Assignment of homologs to functions and generation of a high-quality genome for *P. amilis* were necessary first steps for future functional genetic work on C_4+CAM. Perhaps the most pertinent questions surround the utility of C_4+CAM, which could be addressed in part by knock-down or knock-out *PPC-1E1c* lines to study the fitness of a purely C_4 *Portulaca*. In addition to functional genetic experiments, similar CAM-induction experiments are needed throughout *Portulaca* to establish homology between metabolic pathways across the multiple origins; especially in the C_3-C_4 Cryptopetala clade, which will provide greater clarity on which elements of C_4+CAM are ancestrally shared across multiple origins in *Portulaca*.

Materials and Methods

Plant materials and CAM-induction experiment

Specimens of *Portulaca amilis* and *P. oleracea* were collected in Casselberry, Florida, U.S.A. and propagated in the Plant Environmental Center at Brown University. All studied *Portulaca* are selfing, and many are primarily cleistogamous[86,87]; cleistogamous flowers self-pollinate in

the developing flower bud without the need of a vector, and most buds progress directly to fruit set without opening. S2 seeds were germinated at Yale's Marsh Botanical Garden. One S2 *P. amilis* was removed from soil, gently washed to remove soil, flash frozen whole in liquid N_2, and shipped overnight to Dovetail Genomics (Scotts Valley, CA, U.S.A.) for whole genome sequencing and assembly. An S4 individual was collected and vouchered at the Yale University Herbarium in the Peabody Museum of Natural History.

S3 *P. amilis* and S2 *P. oleracea* were germinated in a growth chamber with the following environmental conditions: 00:00-06:00, dark, 22°C; 06:00-20:00, lights on, 25°C; 20:00-23:59, dark, 22°C. Plants were regularly bottom watered and the mean photosynthetically active radiation (PAR) at plant level was ~385 mol $\cdot$ m^{-2} $\cdot$ s^{-1} during the 14 hour photoperiod. After plants had reached maturity and branched sufficiently (~4 weeks), eight plants of each species were selected for CAM-induction: six biological replicates and two indicator plants. Indicator plants were monitored daily for fluctuations in titratable acidity to determine CAM induction, and were not included in any further analyses. To induce CAM, water was withheld from plants, and significant (one sample *t*-test, $p < 0.01$) increases in nocturnal titratable acidity were observed in *P. amilis* and *P. oleracea* indicator plants after seven and eight days, respectively. Leaf tissue was collected from all six biological replicates for titratable acidity assays at 16:00 and 04:00 (the following day) and bulk RNA at 10:00, 16:00, 22:00, and 04:00 (the following day) on day 0 (referred to as "well-watered") and days 8 and 9 (referred to "drought").

Titratable acidity was assessed by boiling 0.2-0.3g fresh leaf tissue in 60mL 20% EtOH until the volume was reduced by half (~30mL). Distilled water was then added to return the volume to 60mL. This process was repeated, and the supernatant was then covered and allowed to cool to room temperature. Samples were then titrated to a pH of 7.0 using 0.002M NaOH, and

recorded in units of μEq H^+ per gram fresh mass—calculated as volume titrant (μL) × titrant molarity (M) / tissue mass (g).

Total leaf RNA was extracted using Option 2 (CTAB/TRIzol/sarkosyl) from Jordon-Thaden et al.[88]. Sample purity was measured using a NanoDrop (Invitrogen by Thermo Fisher Scientific, Carlsbad, CA), and a representative set of 12 samples were further assayed on using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) for fragment length distribution and RNA integrity (RIN); RIN scores ≥ 7 were considered high quality. Library preparation included poly(A) tailing to pull down mRNA, and sequencing was done on an Illumina HiSeq (2x100bp paired-end reads), with an expected 25 million reads per sample. Four of the 96 total libraries failed during sequencing; all were *P. oleracea* samples (one 10:00 drought, one 16:00 drought, and two 04:00 drought).

Genome and transcriptome assembly and annotation

The *Portulaca amilis* v1 genome was assembled *de novo* by Dovetail Genomics, using one Chicago[89] and two Dovetail Hi-C[90] libraries. The Chicago library was sequenced on an Illumina HiSeqX, resulting in 178 million 2x150bp paired-end reads that yielded 169.1x physical coverage of the genome (1-100kb pairs). The two Dovetail Hi-C libraries were also sequenced on an Illumina HiSeqX; libraries one and two contained 202 million and 342 million 2x150bp paired-end reads, respectively, which together provided 52,626.57x physical coverage of the genome (10-10000kb pairs). Meraculous 2[91] was used to generate an assembly from 791 million paired-end shotgun read pairs (mean insert size ~550bp and ~350bp) with kmer size 91. The Meraculous assembly and shotgun, Chicago, and Hi-C reads were then used for iterative

scaffolding using HiRise v2.1.6-072ca03871cc[89,90] to create the final assembly consisting of 4,053 scaffolds.

The *P. amilis* v1 genome was annotated using funannotate v1.6.0 (http://doi.org/10.5281/zenodo.3354704), a modular software package for genomic scaffold cleaning, genomic element prediction (i.e., genes, UTRs, tRNAs, etc.), functional annotation, and comparative genomics. All annotation steps after genome assembly were run within funannotate unless otherwise noted, and a step-by-step walkthrough of the entire annotation process can be found on GitHub (https://github.com/isgilman/Portulaca-amilis-genome).

Using funannotate's 'clean' function, scaffolds were checked for duplicates with a "leave one out" strategy where, starting with the shortest scaffold, one scaffold was left out and mapped back to the remaining scaffolds with minimap2[92]. Scaffolds were sorted by size from largest to smallest using funannotate's 'sort' function. Outside of funannotate, repetitive element analysis was conducted with RepeatModeler v1.0.11 and RepeatMasker v4.0.9-p2 (http://www.repeatmasker.org). RepeatModeler was used to create a custom repeat database for *P. amilis* and further classification of unknown repeats was done with TEclass v2.1.3[93]. The custom repeats were then passed to RepeatMasker, where complex repeats were soft masked. An additional round of masking was done using the Viridiplantae repeat database from RepBase (RepBaseRepeatMaskerEdition - 26 October 2018)[94] and transposable element data from Dfam v3.0[95]. All repetitive elements were then cleaned using the RepeatMasker utility script 'ProcessRepeats'.

Prior to *ab initio* gene prediction, we used funannotate's 'train' function to create a preliminary set of high-quality gene models for training prediction software using the soft masked scaffolds and RNAseq reads from the CAM-induction experiment. Paired-end raw

563 RNAseq reads were filtered by quality (phred < 25) and trimmed to remove adapters, poly-A/T
564 tails longer than 12 bp, and low-quality tails (phred < 20) using PRINSEQ v0.20.4[96]. To capture
565 the greatest diversity of transcripts and provide sufficient depth to call rare isoforms, we pooled
566 all RNAseq reads from the six biological replicates across four time points and two conditions
567 (well-watered and drought), totaling ~1.58 billion reads from mature leaf tissue. RNAseq reads
568 were then normalized using the 'insilico_read_normalization.pl' utility script within Trinity
569 v2.8.5[97]. Funannotate's 'train' function first builds a coordinate-sorted BAM file and associated
570 index file using HISAT2 v2.1.0[98] for an initial genome-guided transcriptome assembly with
571 Trinity. The resulting initial transcripts then were fed through the PASA pipeline v2.3.3[99] to
572 generate gene models and predicted coding sequences using TransDecoder v5.5.0
573 (https://transdecoder.github.io). Kallisto v0.45.1[100] was then used to quantify abundance data,
574 and PASA gene models with abundance at least 10% of the most abundant gene model at each
575 respective locus were retained for training the *ab initio* predictors Augustus v3.2.3[101], SNAP
576 v2013_11_29[102], glimmerHMM v3.0.4[103], and GeneMark-ET[104].
577 Funannotate's 'predict' function was then used to produce an initial set of gene models
578 from the scaffolds using high abundance PASA models, *ab initio* prediction software, and
579 evidence from transcripts and predicted proteins. First, transcript alignment information from the
580 coordinate-sorted BAM file and protein alignments from Diamond v0.9.24[105] and Exonerate
581 v2.4.0[106] were used to generate a 'hints' file for Augustus and GeneMark-ET. Augustus models
582 with at least 90% exon overlap with protein alignments were selected as high-quality models
583 (HiQ). Gene models from the above predictors, HiQ models, transcripts, and predicted proteins
584 were then used as sources of evidence to create consensus models in EVidenceModeler v1.1.1[107]
585 with weights Augustus:1, HiQ:1, GeneMark-ET:1, PASA:10, SNAP:1, glimmerHMM:1,

proteins:1, transcripts:1.The resulting consensus gene models were filtered to remove any proteins less than 50 amino acids, models that began or ended in assembly gaps, and transposable elements. Next, tRNA gene models were identified with tRNAscan-SE v2.0[108]. Final consensus gene models were constructed by funannotate's 'update' function which added and updated UTRs and refined exon boundaries of alternatively spliced models.

To assemble the *de novo* transcriptome for *P. oleracea*, we again filtered all paired-end raw RNAseq reads by quality (phred < 25) and trimmed to remove adapters, poly-A/T tails longer than 12 bp, and low-quality tails (phred < 20) using PRINSEQ v0.20.4. As with *P. amilis*, we pooled all *P. oleracea* RNAseq reads across time points, conditions, and replicates. Pooled reads were input into Trinity v2.11.0 and assembled *de novo*. We reduced the redundancy of the resulting contigs using CD-HIT v4.8.1[109] by collapsing contigs with at least 98% similarity (-c 98) and open reading frames were identified with TransDecoder. Orthology between *P. amilis* and *P. oleracea* peptide sequences was established using OrthoFinder[110] v2.5.2 and the primary peptide sequences from genomes of *Ananas comosus*, *Amaranthus hypochondriacus*, *Arabidopsis thaliana*, *Brachypodium distachyon*, *Helianthus annuus*, *Kalanchoë fedtschenkoi*, *Phalaenopsis equistris*, *Sedum album*, *Sorghum bicolor*, *Vitis vinifera*, and *Zea mays*.

RNAseq differential abundance analysis

We used sleuth[111] v0.30.0 in R v3.6.1 to assess differential abundance of RNAseq reads quantified by Kallisto for each species independently. For *P. amilis*, we mapped reads to the predicted mRNA transcripts from the genome annotation, and for *P. oleracea* we used the redundancy reduced *de novo* transcriptome. All analyses were conducted at the gene level by providing a transcript-to-gene map to sleuth generated from the final set of gene models and setting 'gene_mode' = TRUE. Abundance data were quantified in transcripts per million (TPM)

and differential abundance was assessed using a likelihood ratio test between the reduced design matrix, where abundance was purely a function of sampling time ('y ~ time_point'), and the full design matrix that included time and water treatment ('y ~ time_point + treatment'). *P*-values for unique transcripts were aggregated for each gene model and transformed into *q*-values by adjusting for false discovery rate (Benjamini Hochberg correction).

Co-expression network analysis

We used Weighted Gene Correlation Network Analysis (WGCNA)[112] v1.69 in R v3.6.1 to construct co-expression networks and identify modules, and NetworkX[113] v2.5 in Python v3.7.3 to visualize and measure network features. WGCNA was run separately for each species and condition, resulting in four networks. Using the normalized abundance data for all biological replicates from sleuth, we first divided samples into well-watered and drought conditions for separate network construction (hereafter referred to as the well-watered and drought datasets, networks, modules, etc). We then removed genes from each dataset with maximum expression less than 5 TPM. Remaining genes with excessive missing data or too low variance were removed with the WGCNA function 'goodSampleGenes'. Outlier samples were detected using hierarchical clustering with 'hclust' (stats v3.6.2; R-core team), and removed from each dataset. We determined the soft threshold powers for each network separately using 'pickSoftThreshold'; the soft threshold power was 8 for both well-watered and drought *P. amilis* networks, and 14 and 7 for *P. oleracea* well-watered and drought networks, respectively. We constructed networks and identified modules using 'blockwiseModules' with corType="bicor", maxPOutliers=0.05, TOMType="signed", deepSplit=2, and mergeCutHeight=0.25 for all networks. With corType="bicor", we allowed modules to contain both correlated and anti-correlated expression

time profiles. We set minModuleSize=30 for *P. amilis* networks and minModuleSize=100 for *P. oleracea* networks.

After filtering and outlier detection within WGCNA, the final *P. amilis* and *P. oleracea* co-expression datasets contained 26,760 genes for 24 samples and 26,337 genes for 22 samples, respectively. Simultaneous network construction and module detection partitioned genes into 41 well-watered and 21 drought *P. amilis* modules and 29 well-watered and 20 drought *P. oleracea* modules (Supplementary Fig. 1a, b), where modules are mutually exclusive clusters of highly correlated genes. In order to extract photosynthetic gene networks (PGNs), we subsetted the resulting adjacency matrices from WGCNA using over 500 (*P. amilis*) and over 600 (*P. oleracea*) manually curated genes with functions broadly related to photosynthesis and removed edges with adjacency < 0.25. This resulted in networks containing 368-466 genes (Supplementary Table 1). We binned these genes into functional categories reflecting major metabolic pathways (e.g., Calvin-Benson cycle, photorespiration, and starch degradation) and enzymatic reactions or functions (e.g., aminotransferase and PEPC regulation). As many enzymes catalyze reversible reactions or function in multiple pathways, these categorizations are intended to reflect the primary role each protein plays and we recognize that alternative categorizations are possible. For example, RuBisCO is the central protein in both the Calvin-Benson cycle and photorespiration, and MDH functions in both the initial carboxylation pathway that generates malate from CO_2 and in the decarboxylation pathway that liberates CO_2 from malate in NAD-type C_4 biochemistry, albeit to a lesser extent.

Motif identification and enrichment

Putative *cis*-regulatory motif identification and enrichment was done using the MEME Suite[114] v5.0.2. First, 5'- and 3'UTRs, introns, and 2kb upstream regions were extracted from all *P.*

amilis gene models. The MEME tool[115] was used to identify *de novo* motifs for each co-expression module and region using an e-value threshold of 0.05 and a minimum motif size of 6bp. We downloaded the JASPAR CORE 2018 motif database[116], supplemented it with the *de novo* motifs identified in Wai et al.[117] related to CAM, and finally added *de novo P. amilis* motifs after checking for redundancy with Tomtom[118]. We then used AME (--scoring avg --method fisher --hit-lo-fraction 0.25 --evalue-report-threshold 10.0)[119] with random sets of control sequences to measure motif enrichment in every module-region combination, as well as the preserved genes of the *PPC-1E1c* and C_4 modules.

Acknowledgements

We thank C. M. Mason (University of Central Florida) for identification and collection of *P. amilis*, Karolina Heyduk (University of Hawaii) for insightful comments on early drafts of this work, and Christopher Bolick at Yale's Marsh Botanic Gardens, who's support made this work possible. This research was funded by the National Science Foundation (IOS-1754662 to EJE and IOS-1708941 to EWG).

Data availability

Genomic scaffolds, peptide and coding sequences, assembled transcriptome, and annotations (including repeats) for *Portulaca amilis* are available through the Phytozome web portal (https://phytozome-next.jgi.doe.gov). Raw RNAseq data for *P. amilis* and *P. oleracea* are available through the NCBI's SRA within BioProject PRJNA732408.

Code availability

680 Walkthroughs of genome annotation and RNAseq analyses (including differential abundance and
681 co-expression network analyses) and all custom code are available on GitHub at
682 https://github.com/isgilman/Portulaca-amilis-genome and https://github.com/isgilman/Portulaca-
683 coexpression, respectively.

684

685 **Tables**

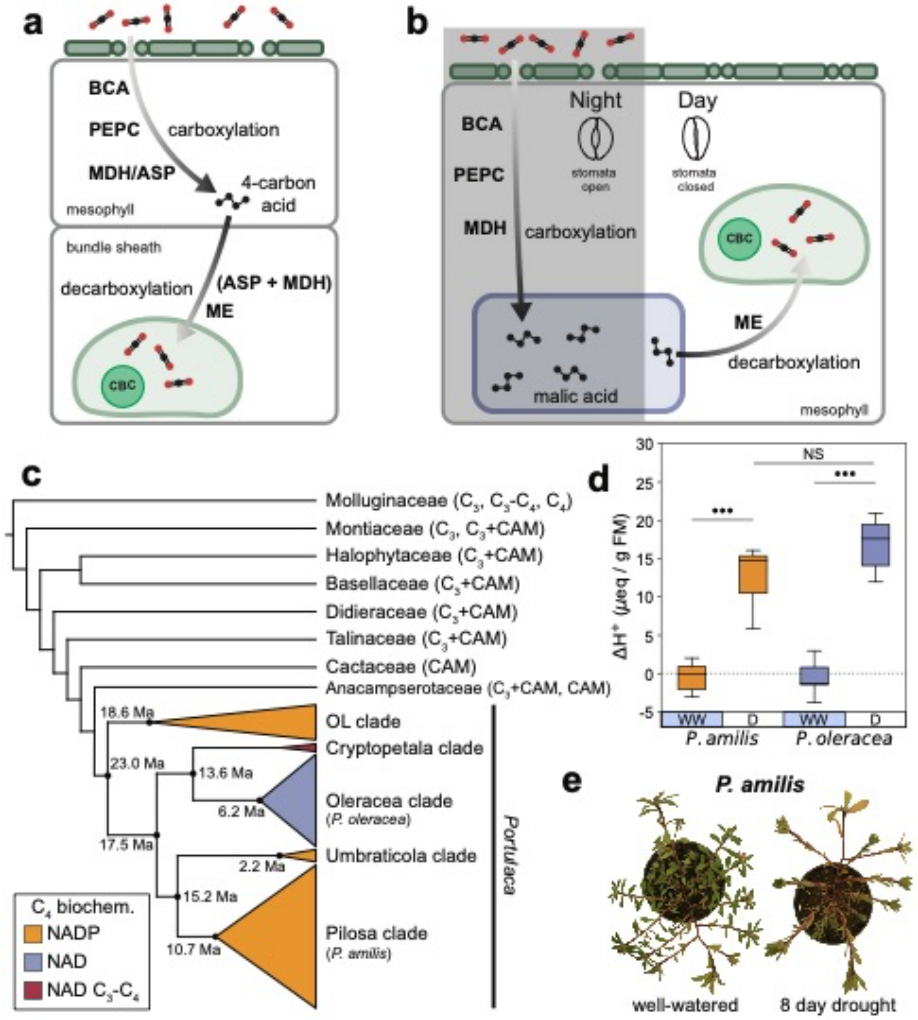
686 Table 1. Motifs enriched in the 5'UTR of genes preserved in the *paPPC-1E1c* module

Motif name	JASPAR matrix ID	Consensus sequence*	Associated TF	*Arabidopsis* ortholog
RVE7L	MA1191.1	HHHVA**AAATATCT**WA	REVEILLE 7-like (RVE7L)	AT3G10113
RVE1	MA1184.1	HHVW**AAATATCT**WH	REVEILLE 1 (RVE1)	AT5G17300
RVE8	MA1182.1	**AGATAT**TTTWDD	LHY-CCA1-like 5 (LCL5)/REVEILLE 8 (RVE8)	AT3G09600
RVE4	MA1187.1	**AGATAT**TTTT	LHY-CCA-like 1 (LCL1)/REVEILLE 4 (RVE4)	AT5G02840
RVE6	MA1183.1	**AGATAT**TTTTD	REVEILLE 6 (RVE6)	AT5G52660
RVE5	MA1190.1	**AGATAT**TTTTD	REVEILLE 5 (RVE5)	AT4G01280
LHY1	MA1185.1	**AGATAT**TTTN	LATE ELONGATED HYPOCOTYL 1 (LHY1)	AT1G01060
EPR1	MA1401.1	RA**AAATATCT**WA	Early-phytochrome-responsive (EPR1)/REVEILLE 7 (RVE7)	AT1G18330

CCA1	MA0972.1	**AAATATCT**	CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)	AT2G46830
bHLH74	MA1360.1	**AGATAT**	ACTIVATOR FOR CELL ELONGATION 2 (ACE2)/Basi helix-loop-helix protein 74 (bHLH74)	AT1G10120

687 *Bold text indicates EE (AAATATCT) or GATA (HGATAR) motifs

688 Figure 1. Simplified C_4 (a) and CAM (b) pathways showing shared carboxylation and
689 decarboxylation pathways. ASP, aspartate aminotransferase; BCA, beta carbonic anhydrase;
690 CBC, Calvin-Benson cycle; MDH, malate dehydrogenase; ME, malic enzyme; PEPC, PEP
691 carboxylase. Cartoon dendrogram of the Portullugo (Portulacineae + Molluginaceae),
692 highlighting major *Portulaca* clades (c), which are colored by primary C_4 biochemical pathway.
693 Mean age estimates for focal *Portulaca* clades are from Ocampo and Columbus[56]. Night-day
694 differences in titratable acidity of experimental *P. amilis* (orange) and *P. oleracea* plants (purple)
695 (d). WW, well-watered; D, drought. Asterisks connecting lines indicate significant differences
696 between well-watered and drought treatments (independent *t*-test; "***", $p < 0.001$; "NS", $p =$
697 0.121). Images of *P. amilis* when well-watered and after 8 days without water (e).



698

699 Figure 2. Idiogram of the nine primary scaffolds of the *P. amilis* genome with key C_4 and CAM
700 genes highlighted (a). Breakdown of repetitive elements of the *P. amilis* genome (b); the
701 horizontal bar chart shows the fractions of major classes of repetitive elements and the repeat
702 landscape (histogram) shows the relative abundances of repetitive elements versus the Kimura
703 substitution level from the consensus. Scatterplot of number of gene models versus the log10-
704 transformed genome size for 107 angiosperm genomes (c) with notable benchmarking and CCM
705 genomes highlighted; data is from Zhao and Schranz[120]. Line in (c) shows results of ordinary
706 least squares regression; one sample $t_{slope}(106) > 0$, $p_{slope} = 0.0017$.

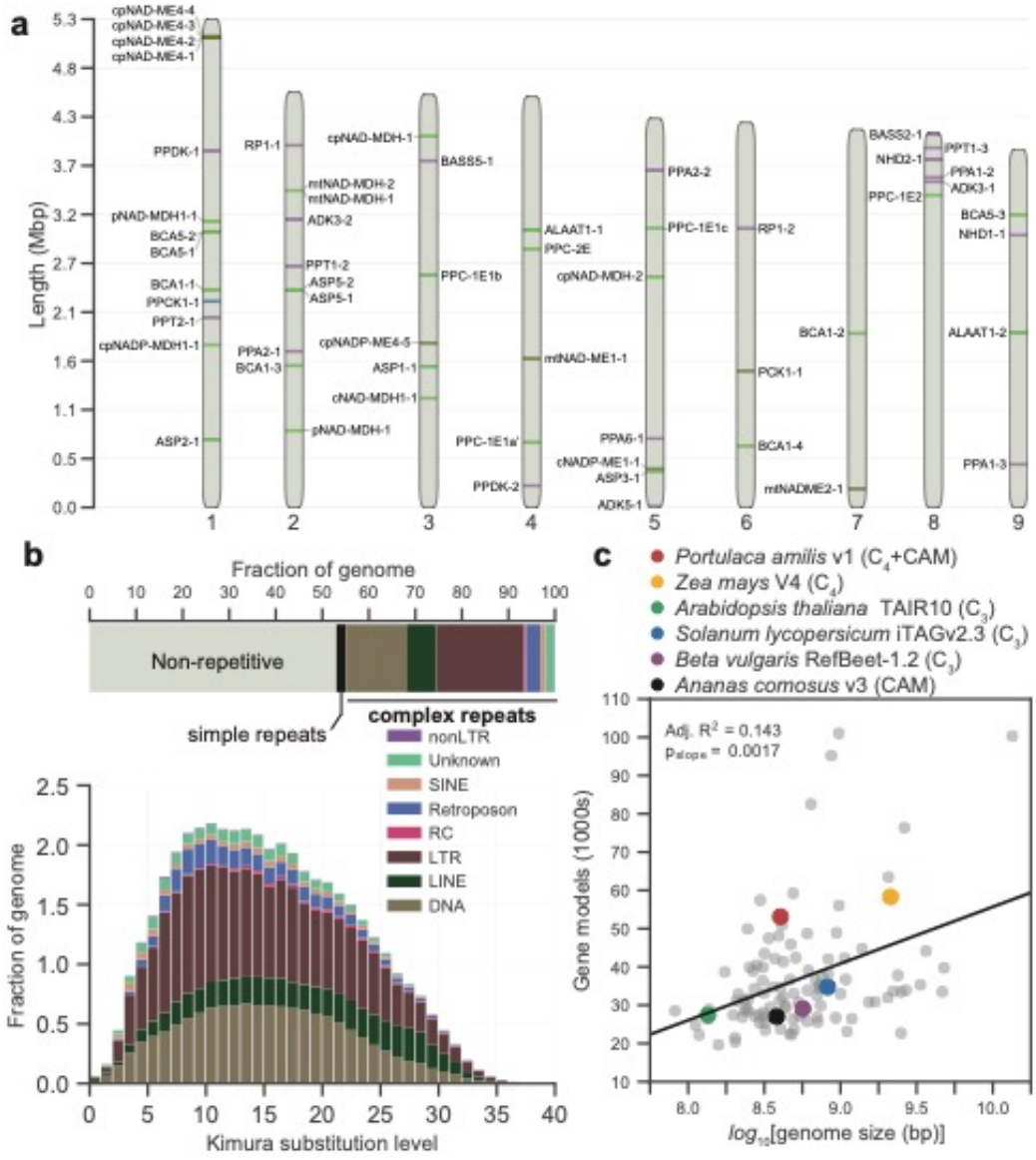


Figure 3. Photosynthetic gene networks (PGNs). Well-watered and drought PGNs of *P. amilis* (a-b), and *P. oleracea* (e-f) colored by WGCNA co-expression module identity. Each node in the graph represents one gene and node sizes represent their degrees. Each module's size and functional composition is shown in the histogram, and the *z*-score normalized expression of each module's constituent genes is shown along the horizontal axis. An example of the *z*-score normalized expression for module paWW1 is enlarged in (a) to show individual gene expression with the median highlighted in bold. Transcripts per million (TPM) normalized expression profiles of two focal *PPC* transcripts (*paPPC-1E1a'* and *paPPC-1E1c*) are shown in (c) and (d), respectively; points show median of size biological replicates and whiskers show interquartile range; black and red lines represent well-watered and drought samples, respectively.

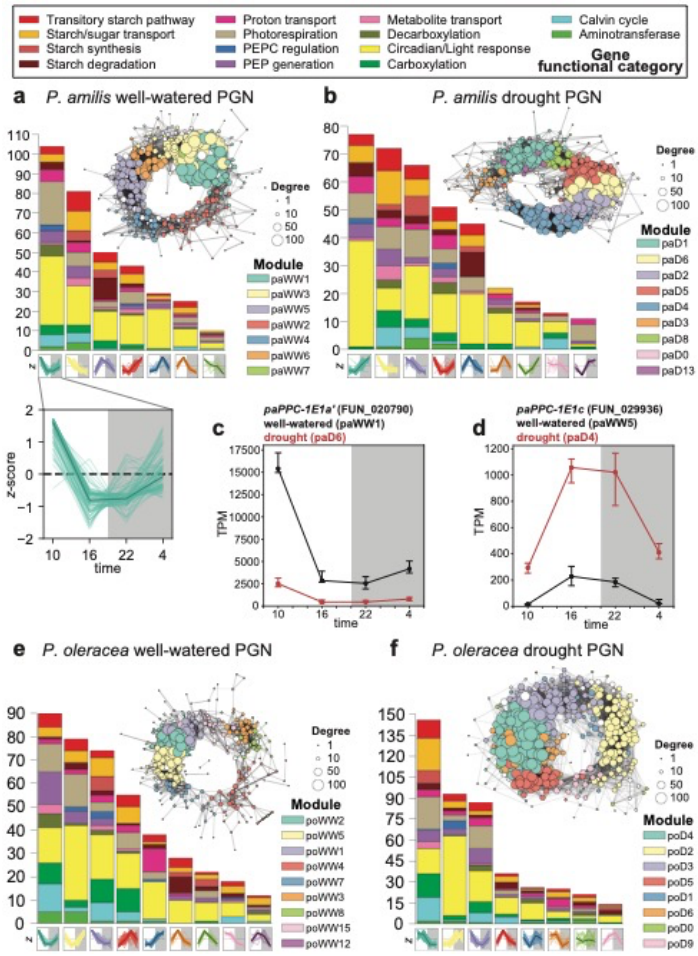
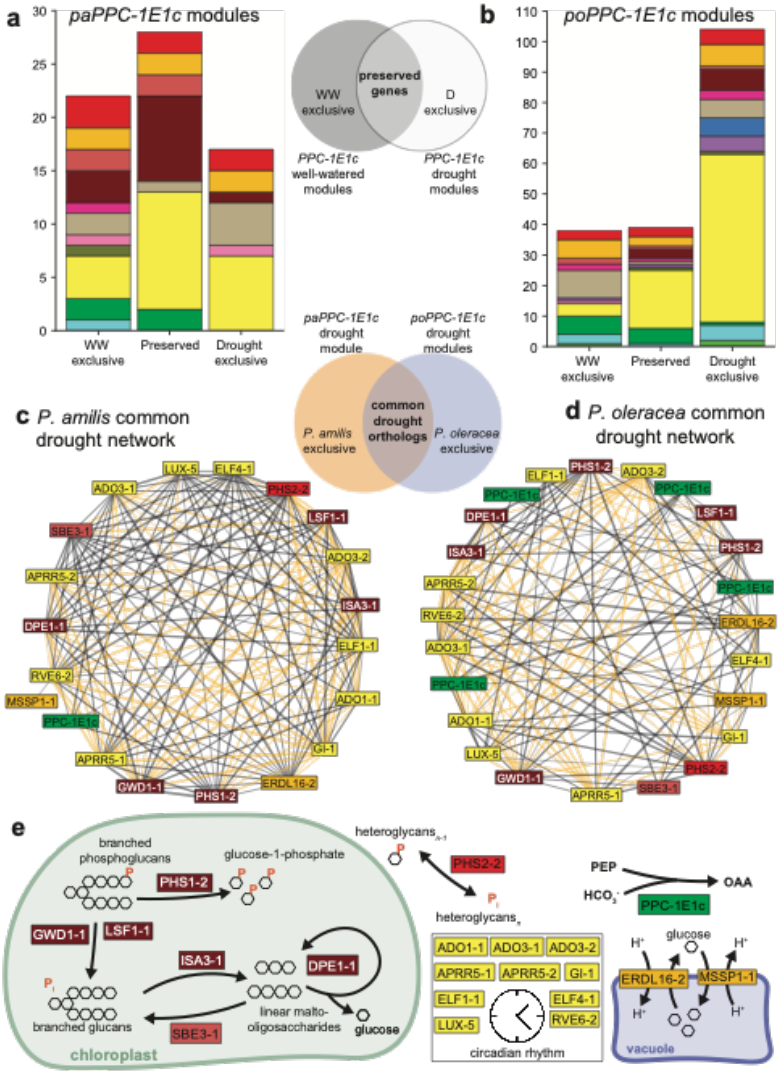
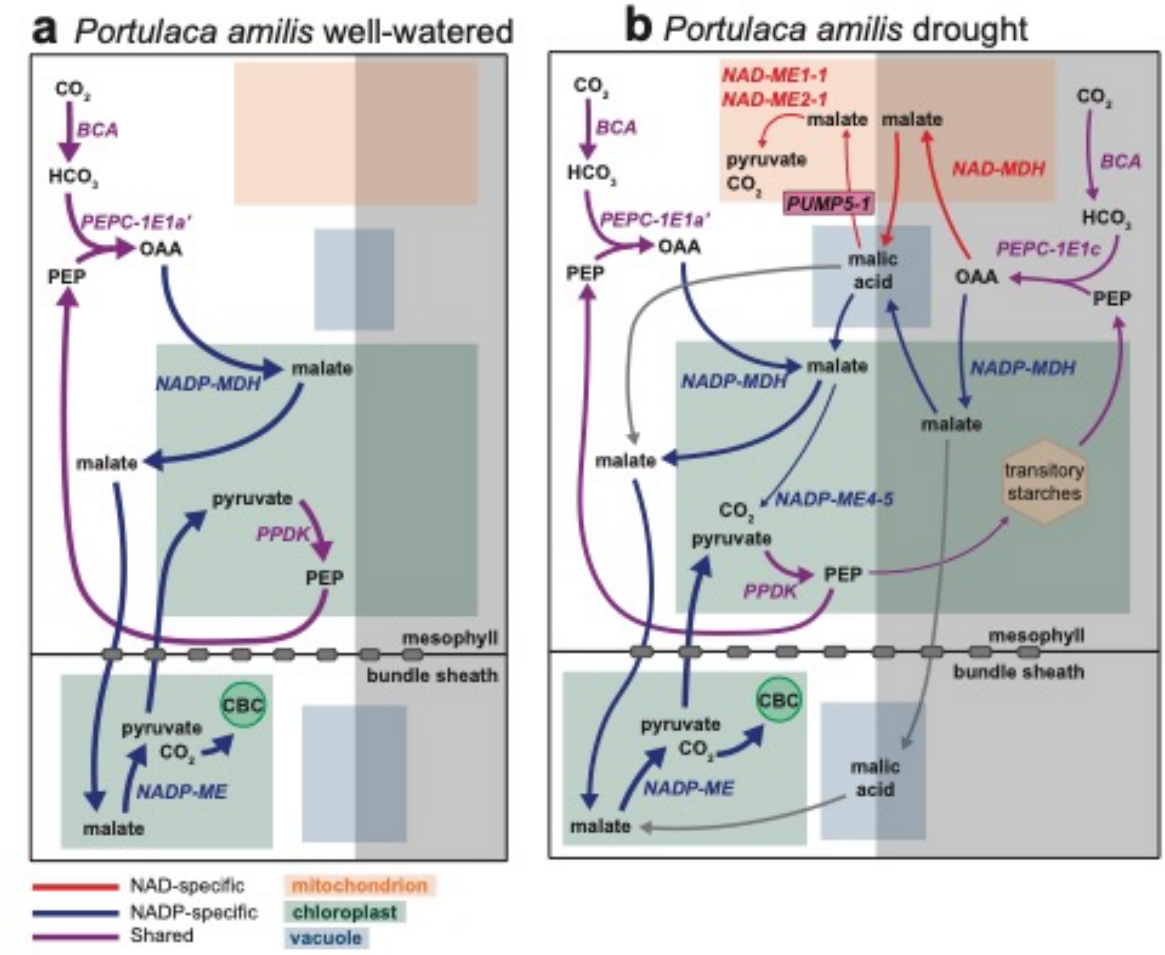


Figure 4. *PPC-1E1c* PGN preservation and subnetworks of common genes. Stacked bar charts showing the number and functional composition of genes that were exclusive to the well-watered or drought *PPC-1E1c* PGNs, or preserved across treatments in *P. amilis* (a) and *P. oleracea* (b). Subnetworks induced by genes common to both *P. amilis* (c) and *P. oleracea* (d) drought *PPC-1E1c* PGNs; drought PGNs include preserved and drought exclusive genes. Orange edges represent correlations among genes shared between species, while black edges are unique to each respective species. Metabolic pathways of the shared *PPC-1E1c* subnetworks (e), showing PEPC, sugar transport, and starch metabolism. Colors of bar plot units (a and b) and gene boxes (c-e) indicate functional categories as in Fig. 3.



726 Figure 5. Hypothesized biochemical cycles of C_4+CAM in *P. amilis* when well-watered (a) and
727 droughted (b). Red, blue, and purple lines indicate NAD-specific, NADP-specific, and shared
728 reactions, respectively, and grey lines show novel, possible pathways for malate that are unique
729 to C_4+CAM photosynthesis. Line thicknesses are indicative of relative fluxes through pathways.
730 ALAAT, alanine aminotransferase; ASP, aspartate aminotransferase; BCA, beta carbonic
731 anhydrase; CBC, Calvin-Benson cycle; NAD-MDH, NAD-dependent malate dehydrogenase;
732 NADP-MDH, NADP-dependent malate dehydrogenase; NAD-ME, NAD-dependent malic
733 enzyme; NADP-ME, NADP-dependent malic enzyme; PEP, phosphoenolpyruvate; PEPC, PEP
734 carboxylase; PPDK, pyruvate, phosphate dikinase; OAA, oxaloacetate.
735



736 1. Wagner, G. P. & Altenberg, L. Perspective: Complex adaptations and the evolution of
737 evolvability. *Evolution* **50**, 967–976 (1996).
738 2. Clune, J., Mouret, J.-B. & Lipson, H. The evolutionary origins of modularity. *Proc. R. Soc. B*
739 *Biol. Sci.* **280**, 20122863 (2013).
740 3. Kashtan, N. & Alon, U. Spontaneous evolution of modularity and network motifs. *Proc. Nat.*
741 *Acad. Sci. USA* **102**, 13773–13778 (2005).
742 4. Kashtan, N., Noor, E. & Alon, U. Varying environments can speed up evolution. *Proc. Nat.*
743 *Acad. Sci. USA* **104**, 13711–13716 (2007).
744 5. Lenski, R. E., Ofria, C., Pennock, R. T. & Adami, C. The evolutionary origin of complex
745 features. *Nature* **423**, 139–144 (2003).
746 6. Barve, A. & Wagner, A. A latent capacity for evolutionary innovation through exaptation in
747 metabolic systems. *Nature* **500**, 203–206 (2013).
748 7. Heyduk, K., Moreno-Villena, J. J., Gilman, I. S., Christin, P.-A. & Edwards, E. J. The genetics
749 of convergent evolution: insights from plant photosynthesis. *Nat. Rev. Gen.* **313**, 1–9 (2019).
750 8. Keeley, J. E. & Rundel, P. W. Evolution of CAM and C_4 Carbon-Concentrating Mechanisms.
751 *Int. J. Plant Sci.* **164**, S55–S77 (2003).
752 9. Edwards, E. J. *et al.* The origins of C_4 grasslands: integrating evolutionary and ecosystem
753 science. *Science* **328**, 587–591 (2010).
754 10. Christin, P.-A., Osborne, C. P., Sage, R. F., Arakaki, M. & Edwards, E. J. C_4 eudicots are not
755 younger than C_4 monocots. *J. Exp. Bot.* **62**, 3171–3181 (2011).
756 11. Sage, R. F. A portrait of the C_4 photosynthetic family on the 50th anniversary of its
757 discovery: species number, evolutionary lineages, and Hall of Fame. *J. Exp. Bot.* **67**, 4039–4056
758 (2016).

759 12. Edwards, E. J. Evolutionary trajectories, accessibility and other metaphors: the case of C_4
760 and CAM photosynthesis. *New Phyt.* **223**, 1742–1755 (2019).
761 13. Bräutigam, A., Schlüter, U., Eisenhut, M. & Gowik, U. On the Evolutionary Origin of CAM
762 Photosynthesis. *Plant Physiol.* **174**, 473–477 (2017).
763 14. Hibberd, J. M. & Covshoff, S. The regulation of gene expression required for C_4
764 photosynthesis. *Ann. Rev. Plant Biol.* **61**, 181–207 (2010).
765 15. Hartwell, J. Annual Plant Reviews: Endogenous Plant Rhythms. in (eds. Hall, A. J. W. &
766 McWatters, H.) vol. 21 211–236 (Blackwell Publishing Ltd, 2006).
767 16. Peterhansel, C. *et al.* Photorespiration. *The Arabidopsis Book* **8**, e0130 (2010).
768 17. Ogren, W. L. Photorespiration: Pathways, Regulation, and Modification. *Ann. Rev. Plant*
769 *Physiol.* **35**, 415–442 (1984).
770 18. Sage, R. F. The evolution of C_4 photosynthesis. *New Phyt.* **161**, 341–370 (2004).
771 19. Taylor, S. H. *et al.* Physiological advantages of C_4 grasses in the field: a comparative
772 experiment demonstrating the importance of drought. *Global Change Biol.* **20**, 1992–2003
773 (2014).
774 20. Niewiadomska, E. & Borland, A. M. Crassulacean acid metabolism: A cause or consequence
775 of oxidative stress in planta? in *Progress in Botany* (eds. Lüttge, U., Beyschlag, W. B. & Murata,
776 J.) vol. 69 247–266 (Springer-Verlag Berlin Heidelberg, 2008).
777 21. Lüttge, U. Photorespiration in Phase III of Crassulacean Acid Metabolism: Evolutionary and
778 Ecophysiological Implications. in *Progress in Botany* (eds. Lüttge, U., Beyschlag, W., Büdel, B.
779 & Francis, D.) vol. 72 371–384 (Springer-Verlag Berlin Heidelberg, 2011).

780 22. Szarek, S. R. & Ting, I. P. Photosynthetic efficiency of CAM plants in relation to C_3 and C_4
781 plants. in *Environmental and Biological Control of Photosynthesis* (ed. Marcelle, R.) 289–299
782 (Dr. W. Junk b.v., 1975).
783 23. Winter, K., Aranda, J. & Holtum, J. A. M. Carbon ssotope composition and water-use
784 efficiency in plants with Crassulacean acid metabolism. *Funct. Plant Biol.* **32**, 381–388 (2005).
785 24. Ehleringer, J. R., Cerling, T. E. & Helliker, B. R. C_4 photosynthesis, atmospheric CO_2, and
786 climate. *Oecologia* **112**, 285–299 (1997).
787 25. Sage, R. F., Sage, T. L. & Kocacinar, F. Photorespiration and the evolution of C_4
788 photosynthesis. *Ann. Rev. Plant Biol.* **63**, 19–47 (2012).
789 26. Raven, J. A. & Spicer, R. A. The evolution of Crassulacean acid metabolism. in
790 *Crassulacean Acid Metabolism Biochemistry, Ecophysiology and Evolution* (eds. Winter, K. &
791 Smith, J. A. C.) 360–388 (Springer, 1996).
792 27. Arakaki, M. *et al.* Contemporaneous and recent radiations of the world's major succulent
793 plant lineages. *Proc. Nat. Acad. Sci. USA* **108**, 8379–8384 (2011).
794 28. Christin, P.-A. *et al.* Shared origins of a key enzyme during the evolution of C_4 and CAM
795 metabolism. *J. Exp. Bot.* **65**, 3609–3621 (2014).
796 29. Koch, K. & Kennedy, R. A. Characteristics of Crassulacean acid metabolism in the succulent
797 C_4 dicot, *Portulaca oleracea* L. *Plant Physiol.* **65**, 193–197 (1980).
798 30. Ho, C.-L., Chiang, J.-M., Lin, T.-C. & Martin, C. E. First report of C_4 /CAM-cycling
799 photosynthetic pathway in a succulent grass, *Spinifex littoreus* (Brum. f.) Merr., in coastal
800 regions of Taiwan. *Flora* **254**, 194–202 (2019).

801 31. Winter, K., Garcia, M., Virgo, A., Ceballos, J. & Holtum, J. A. M. Does the C_4 plant
802 *Trianthema portulacastrum* (Aizoaceae) exhibit weakly expressed crassulacean acid metabolism
803 (CAM)? *Funct. Plant Biol.* **48**, 655–665 (2020).
804 32. Sage, R. F. Are Crassulacean acid metabolism and C_4 photosynthesis incompatible? *Funct.*
805 *Plant Biol.* **29**, 775–785 (2002).
806 33. Koch, K. E. & Kennedy, R. A. Crassulacean acid metabolism in the succulent C_4 dicot,
807 *Portulaca oleracea* L. under natural environmental conditions. *Plant Physiol.* **69**, 757–761
808 (1982).
809 34. Guralnick, L. J. & Jackson, M. D. The occurrence and phylogenetics of Crassulacean acid
810 metabolism in the Portulacaceae. *Int. J. Plant Sci.* **162**, 257–262 (2001).
811 35. Guralnick, L. J., Edwards, G., Ku, M. S. B., Hockema, B. & Franceschi, V. R. Photosynthetic
812 and anatomical characteristics in the C_4-crassulacean acid metabolism-cycling plant, *Portulaca*
813 *grandiflora*. *Funct. Plant Biol.* **29**, 763–773 (2002).
814 36. Lara, M. V., Disante, K. B., Podestá, F. E., Andreo, C. S. & Drincovich, M. F. Induction of a
815 Crassulacean acid like metabolism in the C_4 succulent plant, *Portulaca oleracea* L.:
816 Physiological and morphological changes are accompanied by specific modifications in
817 phosphoenolpyruvate carboxylase. *Photosynth. Res.* **77**, 241 (2003).
818 37. Lara, M. V., Drincovich, M. F. & Andreo, C. S. Induction of a Crassulacean acid-like
819 metabolism in the C_4 Succulent Plant, *Portulaca oleracea* L.: Study of enzymes involved in
820 carbon fixation and carbohydrate metabolism. *Plant Cell Physiol.* **45**, 618–626 (2004).
821 38. Holtum, J. A. M., Hancock, L. P., Edwards, E. J. & Winter, K. Optional use of CAM
822 photosynthesis in two C_4 species, *Portulaca cyclophylla* and *Portulaca digyna*. *J. Plant. Physiol.*
823 **214**, 91–96 (2017).

824 39. Winter, K., Sage, R. F., Edwards, E. J., Virgo, A. & Holtum, J. A. M. Facultative
825 Crassulacean acid metabolism in a C_3–C_4 intermediate. *J. Exp. Bot.* **108**, 8379–9 (2019).
826 40. Ferrari, R. C. *et al.* C_4 and Crassulacean acid metabolism within a single leaf: Deciphering
827 key components behind a rare photosynthetic adaptation. *New Phyt.* **225**, 1699–1714 (2020).
828 41. Ferrari, R. C. *et al.* Exploring C_4 -CAM plasticity within the *Portulaca oleracea* complex.
829 *Scientific Reports* **10**, 14237–14 (2020).
830 42. Edwards, E. J. & Diaz, M. Ecological physiology of *Pereskia guamacho*, a cactus with
831 leaves. *Plant Cell Environ.* **29**, 247–256 (2006).
832 43. Guralnick, L. J., Cline, A., Smith, M. & Sage, R. F. Evolutionary physiology: The extent of
833 C_4 and CAM photosynthesis in the genera *Anacampseros* and *Grahamia* of the Portulacaceae. *J.*
834 *Exp. Bot.* **59**, 1735–1742 (2007).
835 44. Brilhaus, D., Bräutigam, A., Mettler-Altmann, T., Winter, K. & Weber, A. P. M. Reversible
836 burst of transcriptional changes during induction of Crassulacean acid metabolism in *Talinum*
837 *triangulare*. *Plant Physiol.* **170**, 102–122 (2016).
838 45. Holtum, J. A. M., Hancock, L. P., Edwards, E. J. & Winter, K. Crassulacean acid metabolism
839 in the Basellaceae (Caryophyllales). *Plant Biol.* **20**, 409–414 (2018).
840 46. Hancock, L. P., Holtum, J. A. M. & Edwards, E. J. The evolution of CAM photosynthesis in
841 Australian *Calandrinia* reveals lability in C_3+CAM phenotypes and a possible constraint to the
842 evolution of strong CAM. *Integr. Comp. Biol.* **59**, 517–534 (2019).
843 47. Holtum, J. A. M., Hancock, L. P., Edwards, E. J. & Winter, K. CAM photosynthesis in desert
844 blooming *Cistanthe* of the Atacama, Chile. *Funct. Plant Biol.* **48**, 691–702 (2021).
845 48. Ocampo, G. *et al.* Evolution of leaf anatomy and photosynthetic pathways in Portulacaceae.
846 *Am. J. Bot.* **100**, 2388–2402 (2013).

847 49. Voznesenskaya, E. V., Koteyeva, N. K., Edwards, G. E. & Ocampo, G. Unique
848 photosynthetic phenotypes in *Portulaca* (Portulacaceae): C_3-C_4 intermediates and NAD-ME C_4
849 species with Pilosoid-type Kranz anatomy. *J. Exp. Bot.* **68**, 225–239 (2017).
850 50. Voznesenskaya, E. V., Koteyeva, N. K., Edwards, G. E. & Ocampo, G. Revealing diversity
851 in structural and biochemical forms of C_4 photosynthesis and a C_3-C_4 intermediate in genus
852 *Portulaca* L. (Portulacaceae). *J. Exp. Bot.* **61**, 3647–3662 (2010).
853 51. Williams, B. P., Johnston, I. G., Covshoff, S. & Hibberd, J. M. Phenotypic landscape
854 inference reveals multiple evolutionary paths to C_4 photosynthesis. *eLife* **2**, e00961 (2013).
855 52. Furbank, R. T. Evolution of the C_4 photosynthetic mechanism: Are there really three C_4 acid
856 decarboxylation types? *J. Exp. Bot.* **62**, 3103–3108 (2011).
857 53. Sharma, A. K. & Bhattacharyya, N. K. Cytogenetics of some members of Portulacaceae and
858 related families. *Caryologia* **8**, 257–274 (1956).
859 54. McClintock, B. The significance of responses of the genome to challenge. *Science* **226**, 792–
860 801 (1984).
861 55. Dion-Côté, A.-M., Renaut, S., Normandeau, E. & Bernatchez, L. RNA-seq reveals
862 transcriptomic shock involving transposable elements reactivation in hybrids of young lake
863 whitefish species. *Mol. Biol. Evol.* **31**, 1188–1199 (2014).
864 56. Ocampo, G. & Columbus, J. T. Molecular phylogenetics, historical biogeography, and
865 chromosome number evolution of *Portulaca* (Portulacaceae). *Mol. Phy. Evol.* **63**, 97–112 (2012).
866 57. Benjamini, Y. & Hochberg, Y. Controlling the false discovery date: A practical and powerful
867 approach to multiple testing. *J. Royal Stat. Soc. Ser. B Methodology* **57**, 289–300 (1995).

868 58. Weise, S. E., Wijk, K. J. van & Sharkey, T. D. The role of transitory starch in C_3, CAM, and
869 C_4 metabolism and opportunities for engineering leaf starch accumulation. *J. Exp. Bot.* **62**, 3109–
870 3118 (2011).
871 59. Borland, A. M., Guo, H.-B., Yang, X. & Cushman, J. C. Orchestration of carbohydrate
872 processing for Crassulacean acid metabolism. *Curr. Opin. Plant Biol.* **31**, 118–124 (2016).
873 60. Nimmo, H. G. The regulation of phosphoenolpyruvate carboxylase in CAM plants. *Trends*
874 *Plant Sci.* **5**, 75–80 (2000).
875 61. Yang, X. *et al.* The *Kalanchoë* genome provides insights into convergent evolution and
876 building blocks of Crassulacean acid metabolism. *Nat. Comm.* **8**, 1–15 (2017).
877 62. Friso, G., Majeran, W., Huang, M., Sun, Q. & Wijk, K. J. van. Reconstruction of metabolic
878 pathways, protein expression, and homeostasis machineries across maize bundle sheath and
879 mesophyll chloroplasts: Large-scale quantitative proteomics using the first maize genome
880 assembly. *Plant Physiol.* **152**, 1219–1250 (2010).
881 63. Chang, Y.-M. *et al.* Characterizing regulatory and functional differentiation between maize
882 mesophyll and bundle sheath cells by transcriptomic analysis. *Plant Physiol.* **160**, 165–177
883 (2012).
884 64. Liu, Y., Li, X., Li, K., Liu, H. & Lin, C. Multiple bHLH proteins form heterodimers to
885 mediate CRY2-dependent regulation of flowering-time in *Arabidopsis*. *PLoS Genetics* **9**,
886 e1003861 (2013).
887 65. Reyes, J. C., Muro-Pastor, M. I. & Florencio, F. J. The GATA family of transcription factors
888 in *Arabidopsis* and rice. *Plant Physiol.* **134**, 1718–1732 (2004).
889 66. Gendron, J. M. *et al. Arabidopsis* circadian clock protein, TOC1, is a DNA-binding
890 transcription factor. *Proc. Nat. Acad. Sci. USA* **109**, 3167–3172 (2012).

891 67. McClung, C. The plant circadian oscillator. *Biology* **8**, 14–17 (2019).

892 68. Zhang, Y. *et al.* Biochemical and biophysical CO_2 concentrating mechanisms in two species

893 of freshwater macrophyte within the genus *Ottelia* (Hydrocharitaceae). *Photosynth. Res.* **121**,

894 285–297 (2013).

895 69. Huang, W. M. *et al.* Different CO_2 acclimation strategies in juvenile and mature leaves of

896 *Ottelia alismoides*. *Photosynth. Res.* **138**, 219–232 (2018).

897 70. Mallmann, J. *et al.* The role of photorespiration during the evolution of C_4 photosynthesis in

898 the genus Flaveria. *eLife* **3**, e02478 (2014).

899 71. Moreno-Villena, J. J., Dunning, L. T., Osborne, C. P. & Christin, P.-A. Highly expressed

900 genes are preferentially co-opted for C_4 photosynthesis. *Mol. Biol. Evol.* **35**, 94–106 (2018).

901 72. Leegood, R. C. The intercellular compartmentation of metabolites in leaves of *Zea mays* L.

902 *Planta* **164**, 163–171 (1985).

903 73. Schuler, M. L., Mantegazza, O. & Weber, A. P. M. Engineering C_4 photosynthesis into C_3

904 chassis in the synthetic biology age. *Plant Cell* **87**, 51–65 (2016).

905 74. Borland, A. M. *et al.* Engineering Crassulacean acid metabolism to improve water-use

906 efficiency. *Trends Plant Sci.* **19**, 327–338 (2014).

907 75. Wang, N. *et al.* Evolution of Portulacineae marked by gene tree conflict and gene family

908 expansion associated with adaptation to harsh environments. *Mol. Biol. Evol.* **36**, 112–126

909 (2019).

910 76. Ting, I. P. & Hanscom, Z. Induction of acid metabolism in *Portulacaria afra*. *Plant Physiol.*

911 **59**, 511–514 (1977).

912 77. Edwards, E. J. & Ogburn, R. M. Angiosperm responses to a low- CO_2 world: CAM and C_4

913 photosynthesis as parallel evolutionary trajectories. *Int. J. Plant Sci.* **173**, 724–733 (2012).

914 78. Nobel, P. S. *Environmental Biology of Agaves and Cacti*. (Cambridge University Press,
915 2003).
916 79. Edwards, E. J. & Donoghue, M. J. *Pereskia* and the origin of the cactus life-form. *Am. Nat.*
917 **167**, 777–793 (2006).
918 80. Eggli, U. *Illustrated Handbook of Succulent Plants: Dicotyledons*. (Springer, 2004).
919 81. Ogburn, R. M. & Edwards, E. J. Anatomical variation in Cactaceae and relatives: Trait
920 lability and evolutionary innovation. *Am. J. Bot.* **96**, 391–408 (2009).
921 82. Herrera, A. Crassulacean acid metabolism and fitness under water deficit stress: If not for
922 carbon gain, what is facultative CAM good for? *Annals of Bot.* **103**, 645–653 (2008).
923 83. Winter, K. Ecophysiology of constitutive and facultative CAM photosynthesis. *J. Exp. Bot.*
924 **2**, 16178–14 (2019).
925 84. Ghannoum, O. C_4 photosynthesis and water stress. *Annals of Bot.* **103**, 635–644 (2009).
926 85. Ferrari, R. C. *et al.* Developing *Portulaca oleracea* as a model system for functional
927 genomics analysis of C_4/CAM photosynthesis. *Funct. Plant Biol.* **7**, 666–682 (2020).
928 86. Geesink, R. An account of the genus *Portulaca* in Indo-Australia and the Pacific
929 (Portulacaceae). *Blumea* **17**, (1969).
930 87. Kim, I. & Carr, G. D. Reproductive biology and uniform culture of *Portulaca* in Hawaii.
931 *Pacific Sci.* **44**, 123–129 (1990).
932 88. Jordon-Thaden, I. E., Chanderbali, A. S., Gitzendanner, M. A. & Soltis, D. E. Modified
933 CTAB and TRIzol protocols improve RNA extraction from chemically complex embryophyta.
934 *App. Plant Sci.* **3**, 1400105–22 (2015).
935 89. Putnam, N. H. *et al.* Chromosome-scale shotgun assembly using an in vitro method for long-
936 range linkage. *Genome Res.* **26**, 342–350 (2016).

937 90. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals
938 folding principles of the human genome. *Science* **326**, 289–293 (2009).
939 91. Chapman, J. A. *et al.* Meraculous: De novo genome assembly with short paired-end reads.
940 *PLoS One* **6**, e23501 (2011).
941 92. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–
942 3100 (2018).
943 93. Abrusán, G., Grundmann, N., DeMester, L. & Makalowski, W. TEclass—a tool for
944 automated classification of unknown eukaryotic transposable elements. *Bioinformatics* **25**, 1329–
945 1330 (2009).
946 94. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in
947 eukaryotic genomes. *Mobile DNA* **6**, 11 (2015).
948 95. Hubley, R. *et al.* The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **44**, D81-
949 9 (2016).
950 96. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets.
951 *Bioinformatics* **27**, 863–864 (2011).
952 97. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a
953 reference genome. *Nat. Biotech.* **29**, 644–652 (2011).
954 98. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome
955 alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotech.* **37**, 907–915
956 (2019).
957 99. Haas, B. J. *et al.* Improving the *Arabidopsis* genome annotation using maximal transcript
958 alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).

959 100. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq
960 quantification. *Nat. Biotech.* **34**, 525–527 (2016).
961 101. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically
962 mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
963 102. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
964 103. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open
965 source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
966 104. Besemer, J. & Borodovsky, M. GeneMark: web software for gene finding in prokaryotes,
967 eukaryotes and viruses. *Nucleic Acids Res.* **33**, W451-4 (2005).
968 105. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using
969 DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
970 106. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence
971 comparison. *BMC Bioinformatics* **6**, 31 (2005).
972 107. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler
973 and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
974 108. Chan, P. P. & Lowe, T. M. tRNAscan-SE: Searching for tRNA genes in genomic
975 sequences. *Methods Mol. Biol.* **1962**, 1–14 (2019).
976 109. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-
977 generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
978 110. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome
979 comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157
980 (2015).

111. Pimentel, H., Bray, N. L., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods* **14**, 687–690 (2017).

112. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

113. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure, dynamics, and function using NetworkX. in *Proceedings of the 7th Python in Science Conference (SciPy 2008)* (eds. Varopuax, G., Vaught, T. & Millman, J.) 11–16 (2008).

114. Bailey, T. L. *et al.* MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).

115. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. in *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology* 28–36 (AAAI Press, 1994).

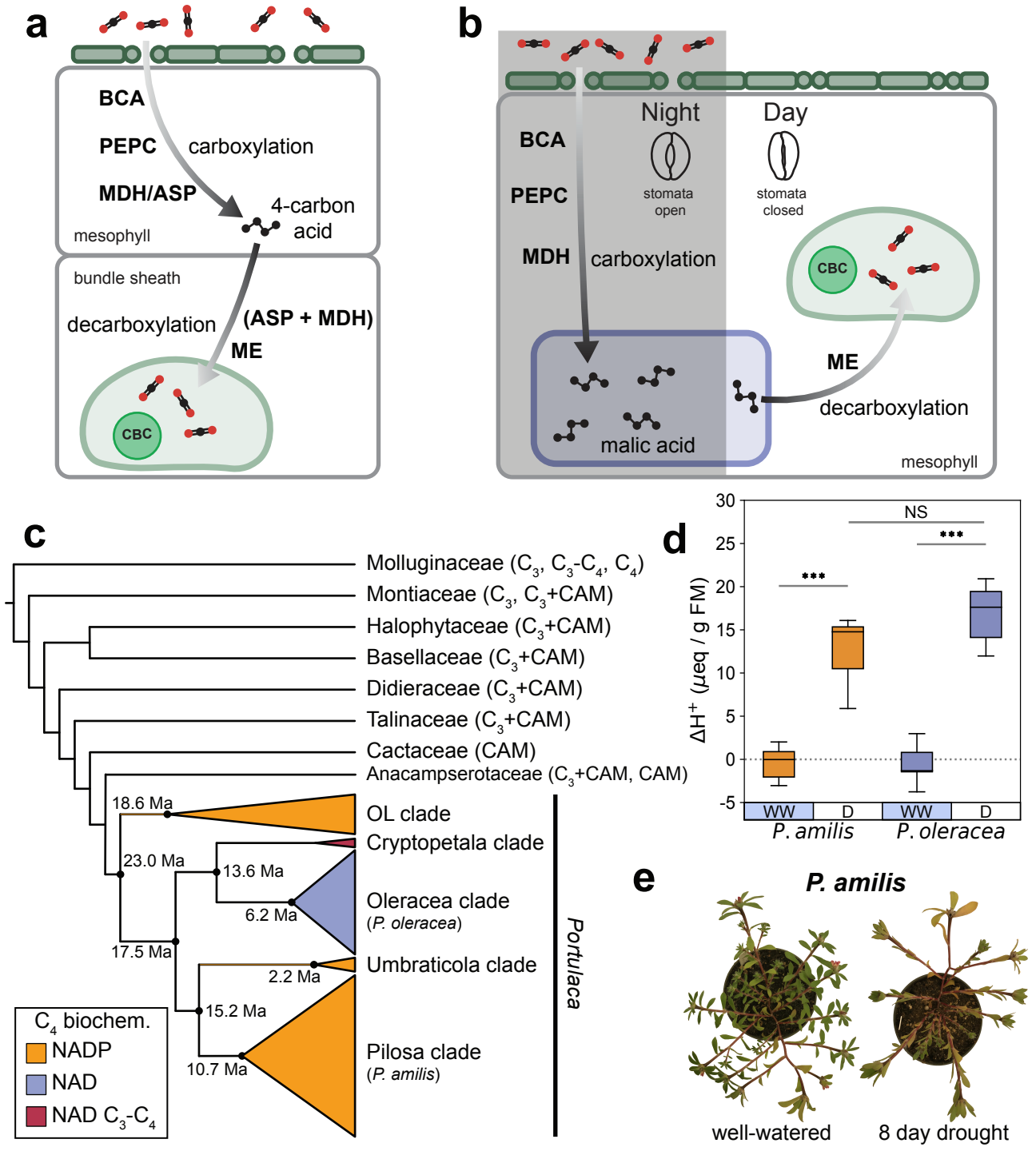
116. Khan, A. *et al.* JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260-266 (2018).

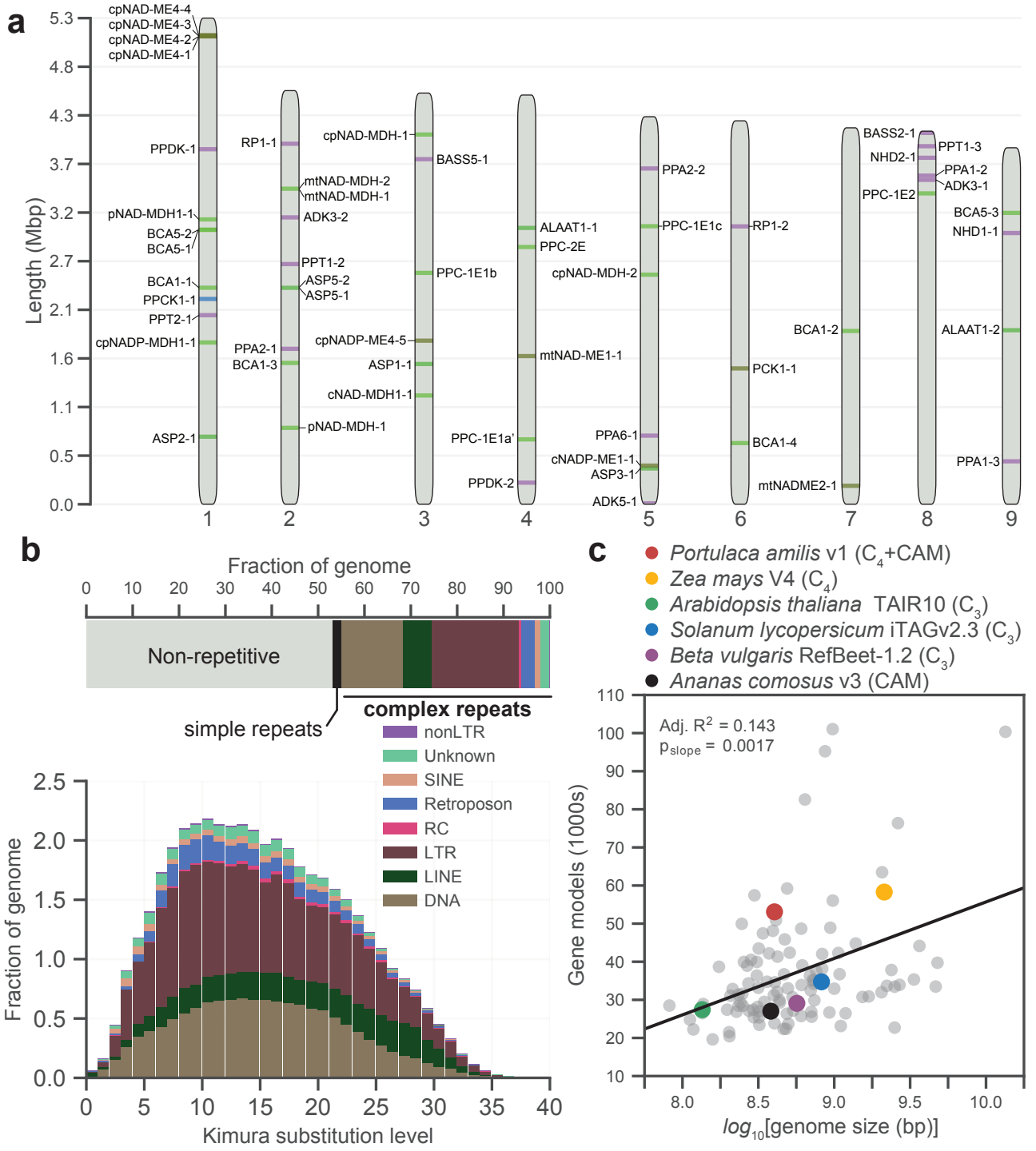
117. Wai, C. M. *et al.* Time of day and network reprogramming during drought induced CAM photosynthesis in *Sedum album*. *PLoS Genetics* **15**, e1008209 (2019).

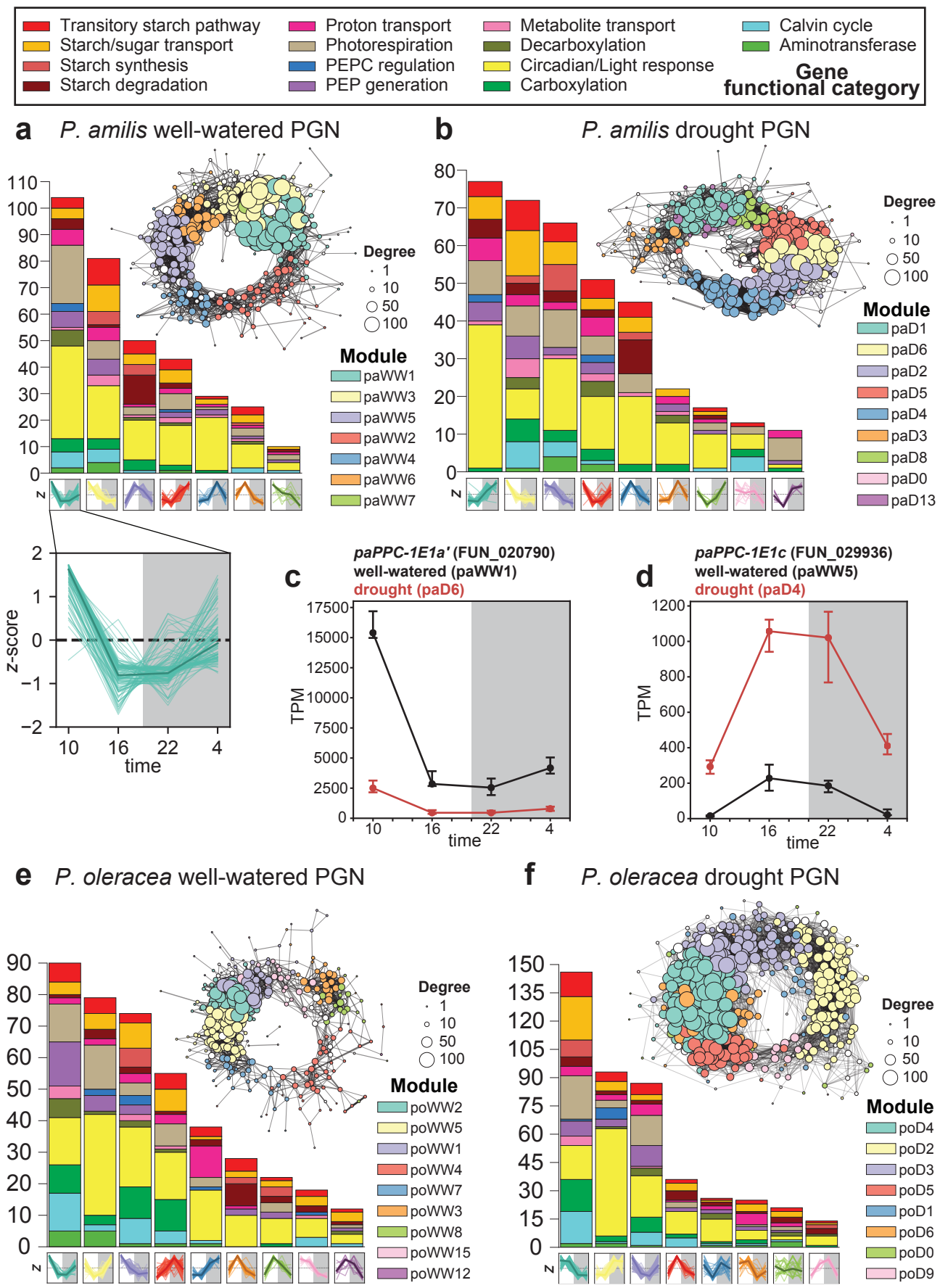
118. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S. Quantifying similarity between motifs. *Genome Biol.* **8**, R24 (2007).

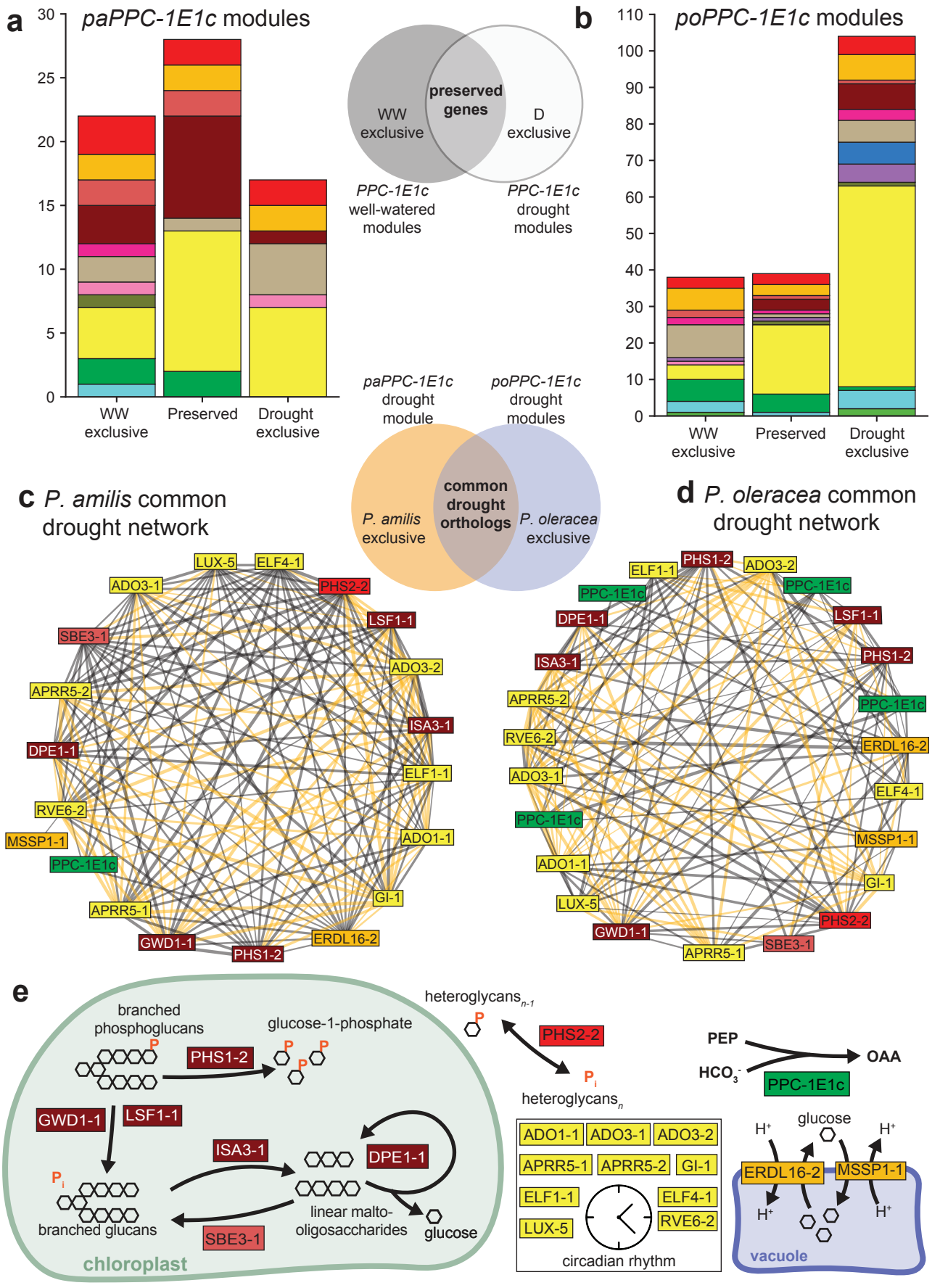
119. McLeay, R. C. & Bailey, T. L. Motif enrichment analysis: a unified framework and an evaluation on ChIP data. *Bmc Bioinformatics* **11**, 165 (2010).

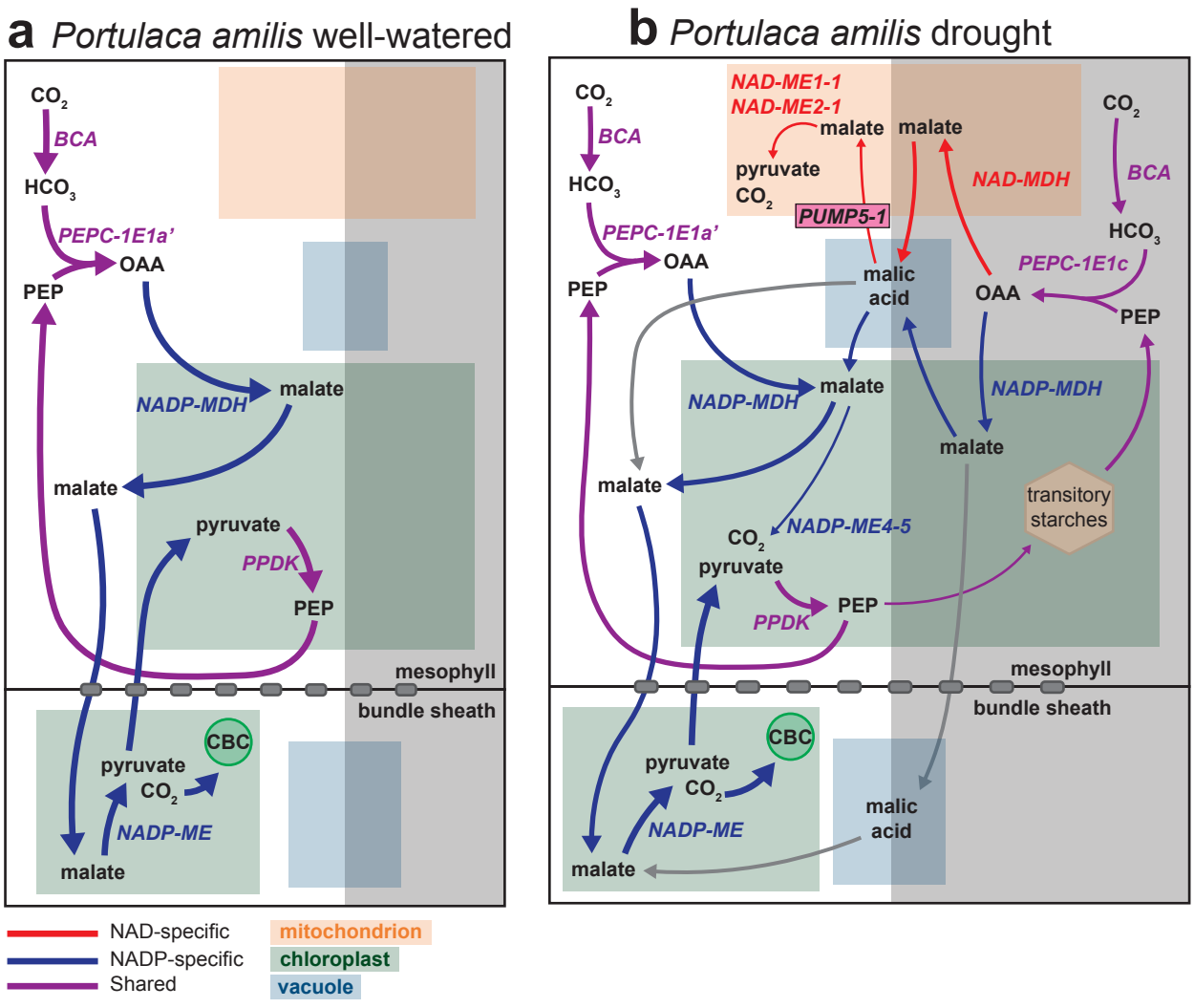
120. Zhao, T. & Schranz, M. E. Network-based microsynteny analysis identifies major differences and genomic outliers in mammalian and angiosperm genomes. *Proc. Nat. Acad. Sci. USA* **116**, 2165–2174 (2019).

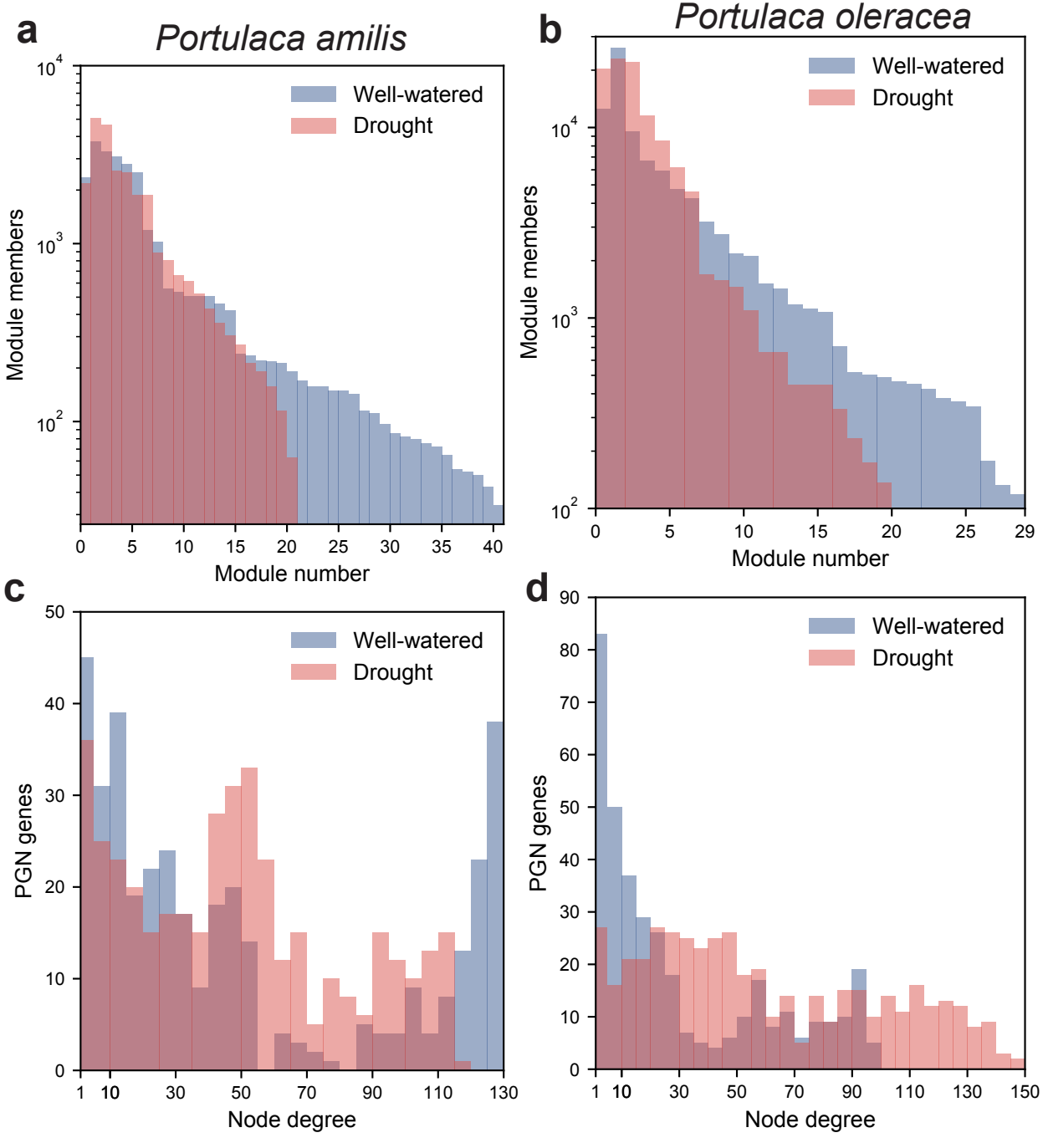

a
BCA
PEPC
MDH/ASP
carboxylation
4-carbon acid
mesophyll
bundle sheath
decarboxylation
(ASP + MDH)
ME
CBC
b
Night
Day
BCA
PEPC
MDH
stomata open
stomata closed
carboxylation
CBC
ME
decarboxylation
malic acid
mesophyll
c
Molluginaceae (C3, C3-C4, C4)
Montiaceae (C3, C3+CAM)
Halophytaceae (C3+CAM)
Basellaceae (C3+CAM)
Didieraceae (C3+CAM)
Talinaceae (C3+CAM)
Cactaceae (CAM)
Anacampserotaceae (C3+CAM, CAM)
18.6 Ma
23.0 Ma
13.6 Ma
6.2 Ma
17.5 Ma
2.2 Ma
15.2 Ma
10.7 Ma
OL clade
Cryptopetala clade
Oleracea clade
(P. oleracea)
Umbraticola clade
Pilosa clade
(P. amilis)
Portulaca
C4 biochem.
NADP
NAD
NAD C3-C4
d
ΔH+ (μeq / g FM)
30
25
20
15
10
5
0
-5
NS
***
***
WW
D
WW
D
P. amilis
P. oleracea
e
P. amilis
well-watered
8 day drought

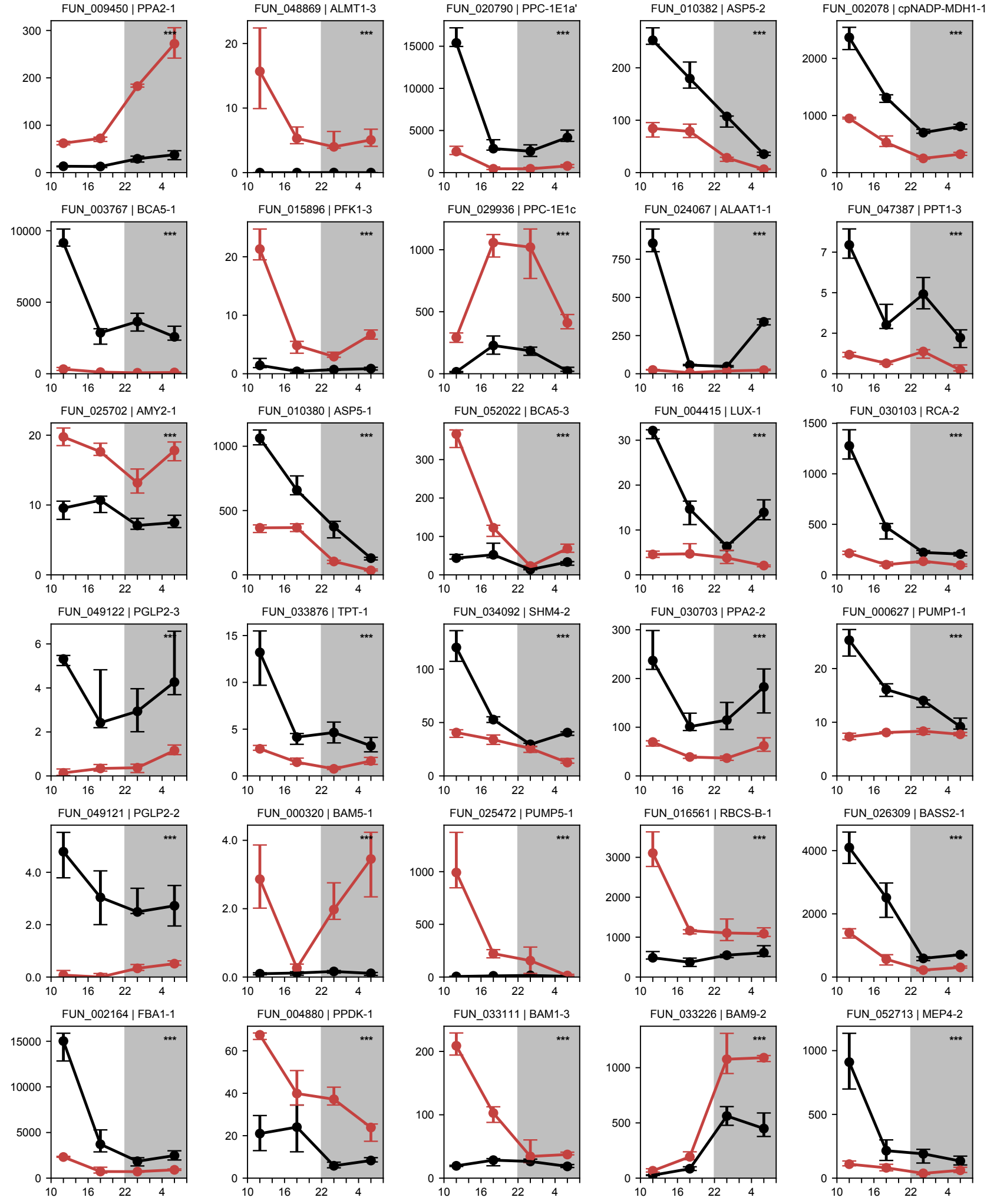

a
Length (Mbp)
cpNAD-ME4-4
cpNAD-ME4-3
cpNAD-ME4-2
cpNAD-ME4-1
PPDK-1
pNAD-MDH1-1
BCA5-2
BCA5-1
BCA1-1
PPCK1-1
PPT2-1
cpNADP-MDH1-1
ASP2-1
RP1-1
mtNAD-MDH-2
mtNAD-MDH-1
ADK3-2
PPT1-2
ASP5-2
ASP5-1
PPA2-1
BCA1-3
pNAD-MDH-1
cpNAD-MDH-1
BASS5-1
PPC-1E1b
cpNADP-ME4-5
ASP1-1
cNAD-MDH1-1
PPC-1E1a'
PPDK-2
ALAAT1-1
PPC-2E
cpNAD-MDH-2
mtNAD-ME1-1
PPA6-1
cNADP-ME1-1
ASP3-1
ADK5-1
PPA2-2
PPC-1E1c
cpNADP-MDH-2
PPA6-1
cNADP-ME1-1
ASP3-1
ADK5-1
RP1-2
PCK1-1
BCA1-4
mtNADME2-1
BASS2-1
NHD2-1
PPC-1E2
BCA1-2
PPT1-3
PPA1-2
ADK3-1
BCA5-3
NHD1-1
ALAAT1-2
PPA1-3
b
Fraction of genome
Non-repetitive
simple repeats
complex repeats
nonLTR
Unknown
SINE
Retroposon
RC
LTR
LINE
DNA
Fraction of genome
Kimura substitution level
c
Portulaca amilis v1 (C4+CAM)
Zea mays V4 (C4)
Arabidopsis thaliana TAIR10 (C3)
Solanum lycopersicum iTAGv2.3 (C3)
Beta vulgaris RefBeet-1.2 (C3)
Ananas comosus v3 (CAM)
Adj. R2 = 0.143
pslope = 0.0017
Gene models (1000s)
log10[genome size (bp)]

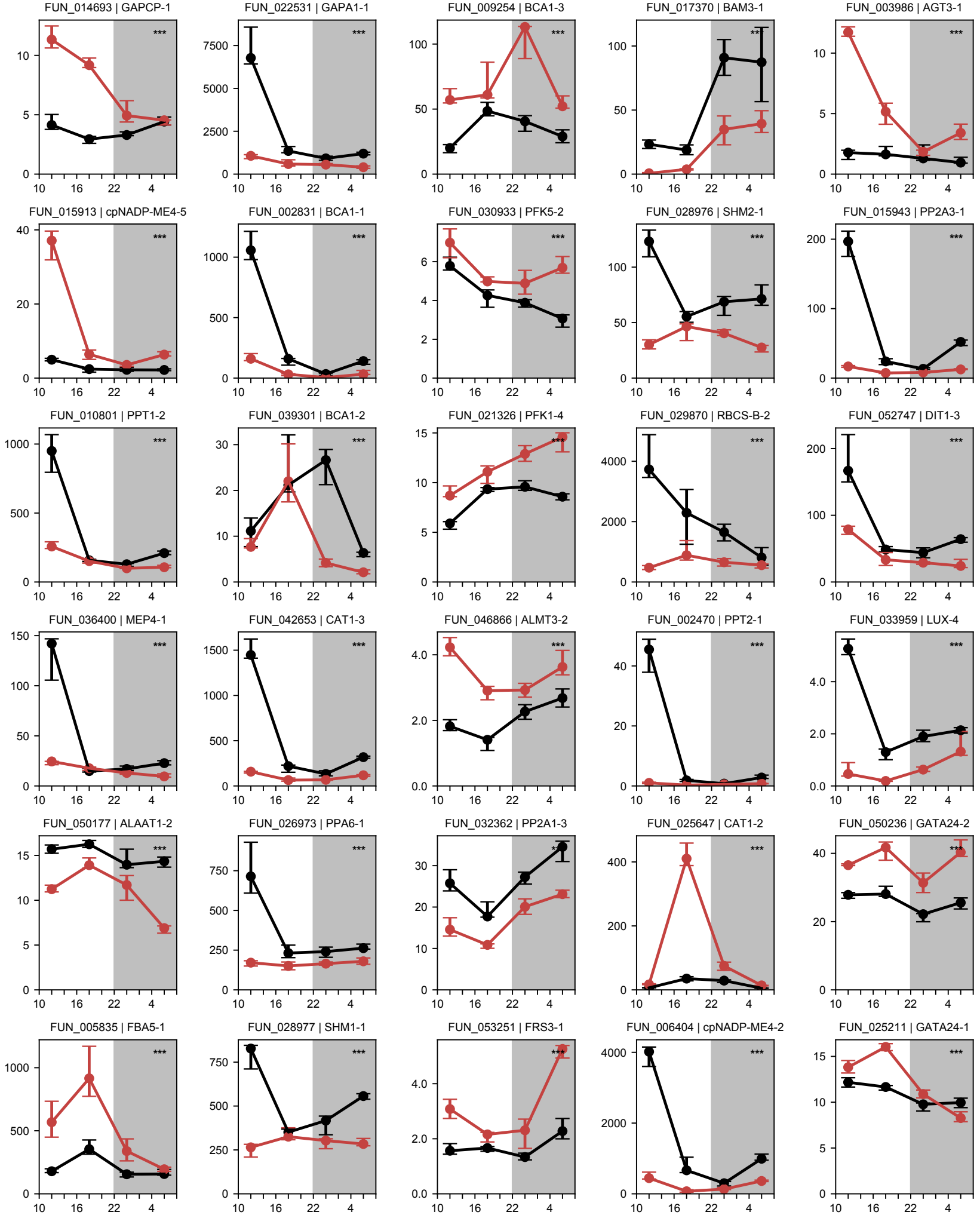

Transitory starch pathway
Starch/sugar transport
Starch synthesis
Starch degradation
Proton transport
Photorespiration
PEPC regulation
PEP generation
Metabolite transport
Decarboxylation
Circadian/Light response
Carboxylation
Calvin cycle
Aminotransferase
Gene functional category
a
P. amilis well-watered PGN
b
P. amilis drought PGN
Degree
1
10
50
100
Module
paWW1
paWW3
paWW5
paWW2
paWW4
paWW6
paWW7
paD1
paD6
paD2
paD5
paD4
paD3
paD8
paD0
paD13
z
z-score
time
c
paPPC-1E1a' (FUN_020790)
well-watered (paWW1)
drought (paD6)
TPM
d
paPPC-1E1c (FUN_029936)
well-watered (paWW5)
drought (paD4)
e
P. oleracea well-watered PGN
poWW2
poWW5
poWW1
poWW4
poWW7
poWW3
poWW8
poWW15
poWW12
f
P. oleracea drought PGN
poD4
poD2
poD3
poD5
poD1
poD6
poD0
poD9

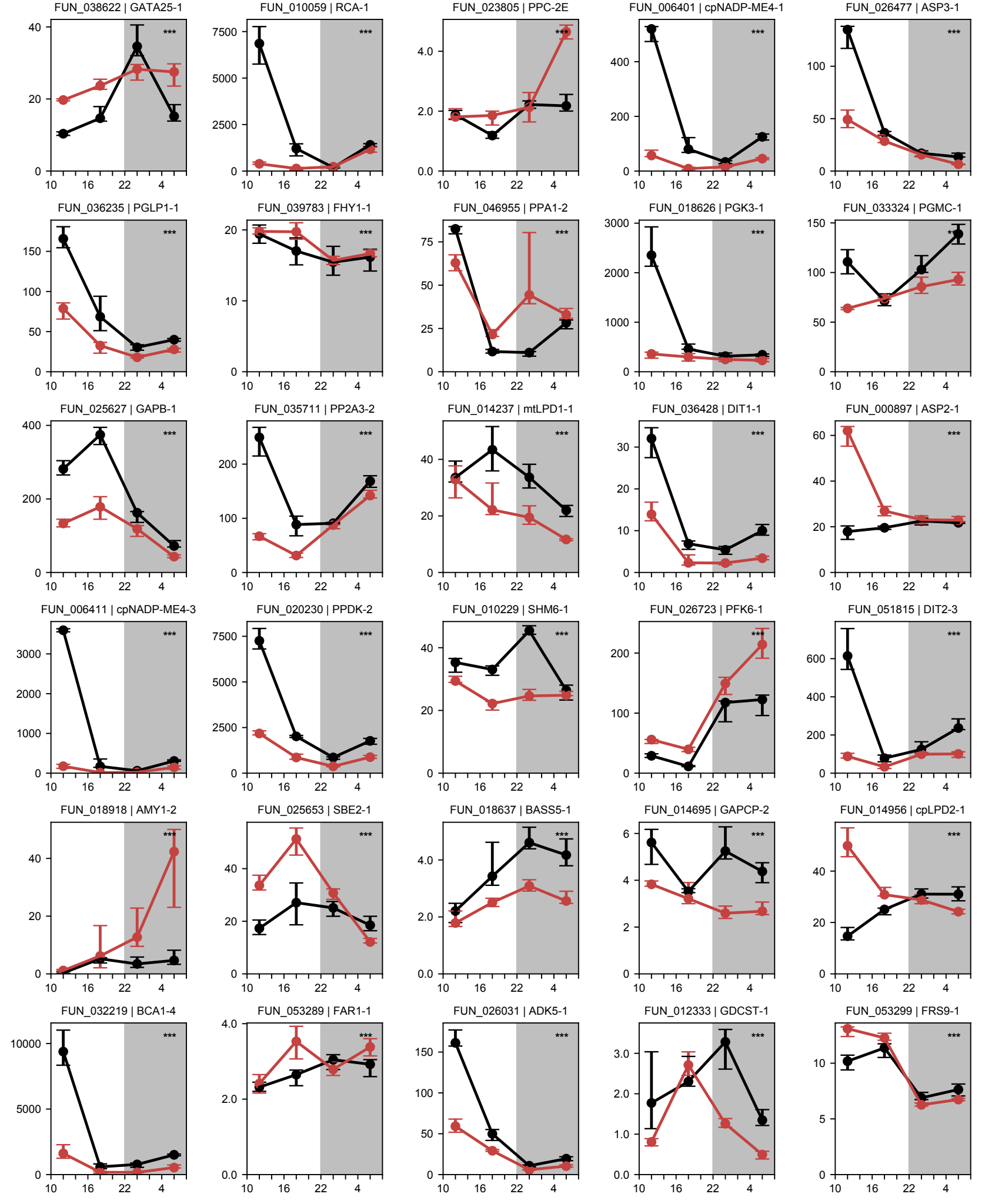

a
paPPC-1E1c modules
30
25
20
15
10
5
0
WW exclusive
Preserved
Drought exclusive
WW exclusive
preserved genes
D exclusive
PPC-1E1c well-watered modules
PPC-1E1c drought modules
b
poPPC-1E1c modules
110
100
90
80
70
60
50
40
30
20
10
0
WW exclusive
Preserved
Drought exclusive
paPPC-1E1c drought module
poPPC-1E1c drought modules
P. amilis exclusive
common drought orthologs
P. oleracea exclusive
c P. amilis common drought network
LUX-5
ELF4-1
ADO3-1
PHS2-2
LSF1-1
SBE3-1
ADO3-2
APRR5-2
ISA3-1
DPE1-1
ELF1-1
RVE6-2
ADO1-1
MSSP1-1
PPC-1E1c
APRR5-1
GI-1
GWD1-1
ERDL16-2
PHS1-2
d P. oleracea common drought network
PHS1-2
ADO3-2
ELF1-1
PPC-1E1c
PPC-1E1c
LSF1-1
DPE1-1
PHS1-2
ISA3-1
APRR5-2
PPC-1E1c
RVE6-2
ERDL16-2
ADO3-1
ELF4-1
PPC-1E1c
MSSP1-1
ADO1-1
GI-1
LUX-5
PHS2-2
GWD1-1
SBE3-1
APRR5-1
e
branched phosphoglucans
glucose-1-phosphate
PHS1-2
GWD1-1
LSF1-1
ISA3-1
DPE1-1
branched glucans
linear malto-oligosaccharides
glucose
SBE3-1
chloroplast
heteroglycans n-1
PHS2-2
heteroglycans n
PEP
OAA
HCO3-
PPC-1E1c
ADO1-1
ADO3-1
ADO3-2
APRR5-1
APRR5-2
GI-1
ELF1-1
ELF4-1
LUX-5
RVE6-2
circadian rhythm
glucose
H+
ERDL16-2
MSSP1-1
vacuole

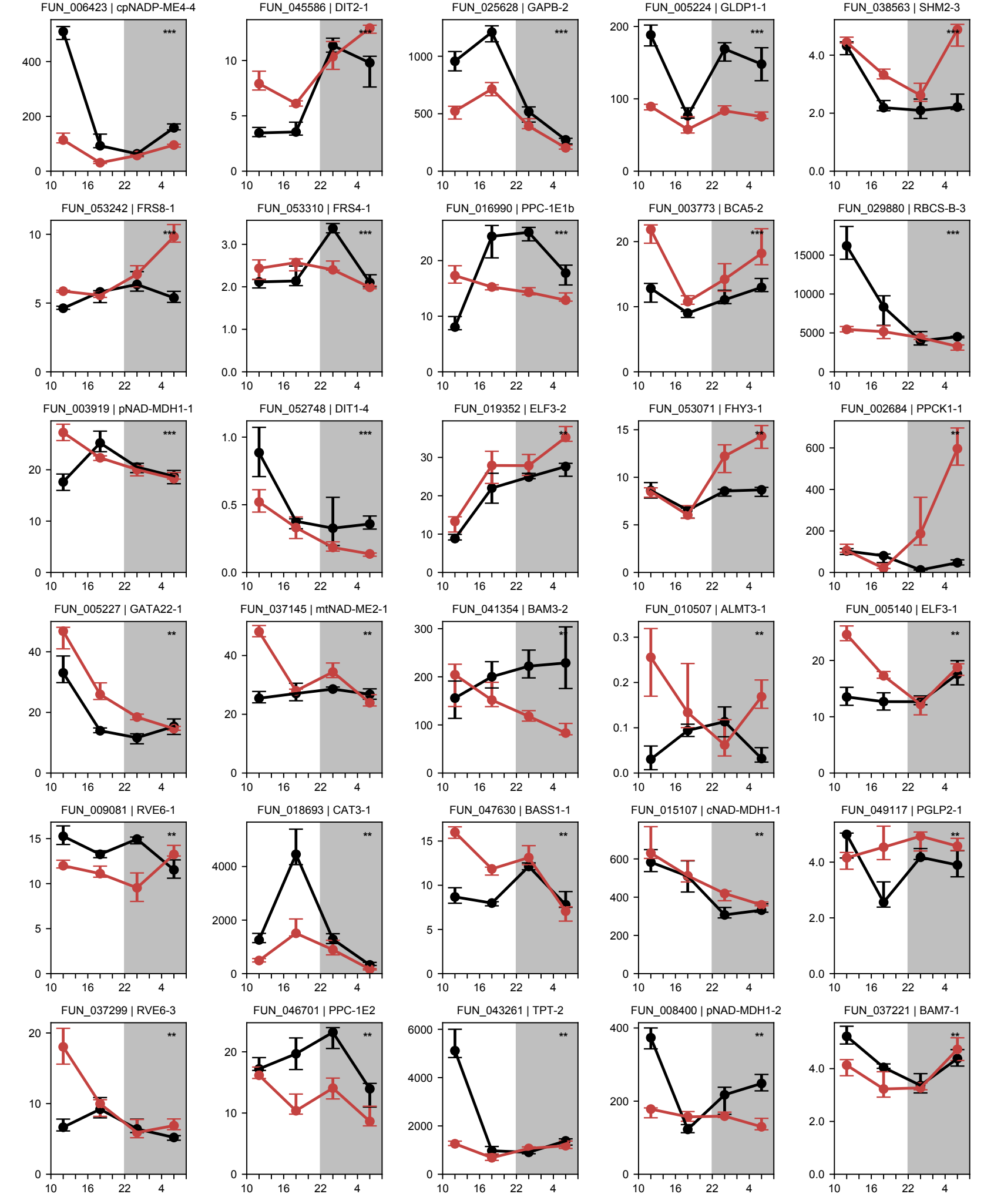

a Portulaca amilis well-watered
b Portulaca amilis drought
CO2
BCA
HCO3
PEPC-1E1a'
OAA
PEP
NADP-MDH
malate
pyruvate
PPDK
mesophyll
bundle sheath
CBC
CO2
NADP-ME
NAD-ME1-1
NAD-ME2-1
PUMP5-1
NAD-MDH
malic acid
PEPC-1E1c
transitory starches
NADP-ME4-5
NAD-specific
NADP-specific
Shared
mitochondrion
chloroplast
vacuole

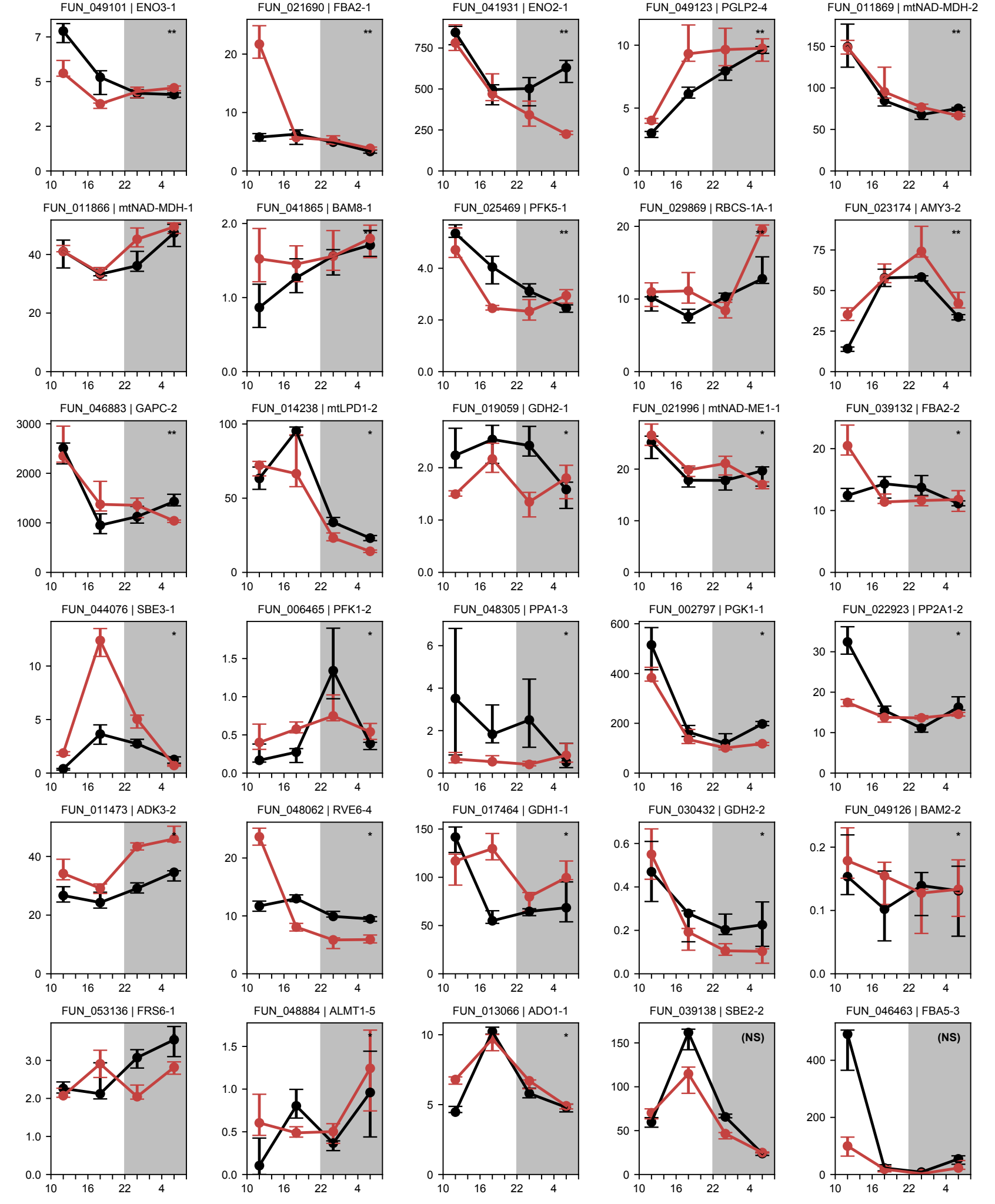

a
Portulaca amilis
Well-watered
Drought
Module members
Module number
b
Portulaca oleracea
Well-watered
Drought
Module members
Module number
c
Well-watered
Drought
PGN genes
Node degree
d
Well-watered
Drought
PGN genes
Node degree

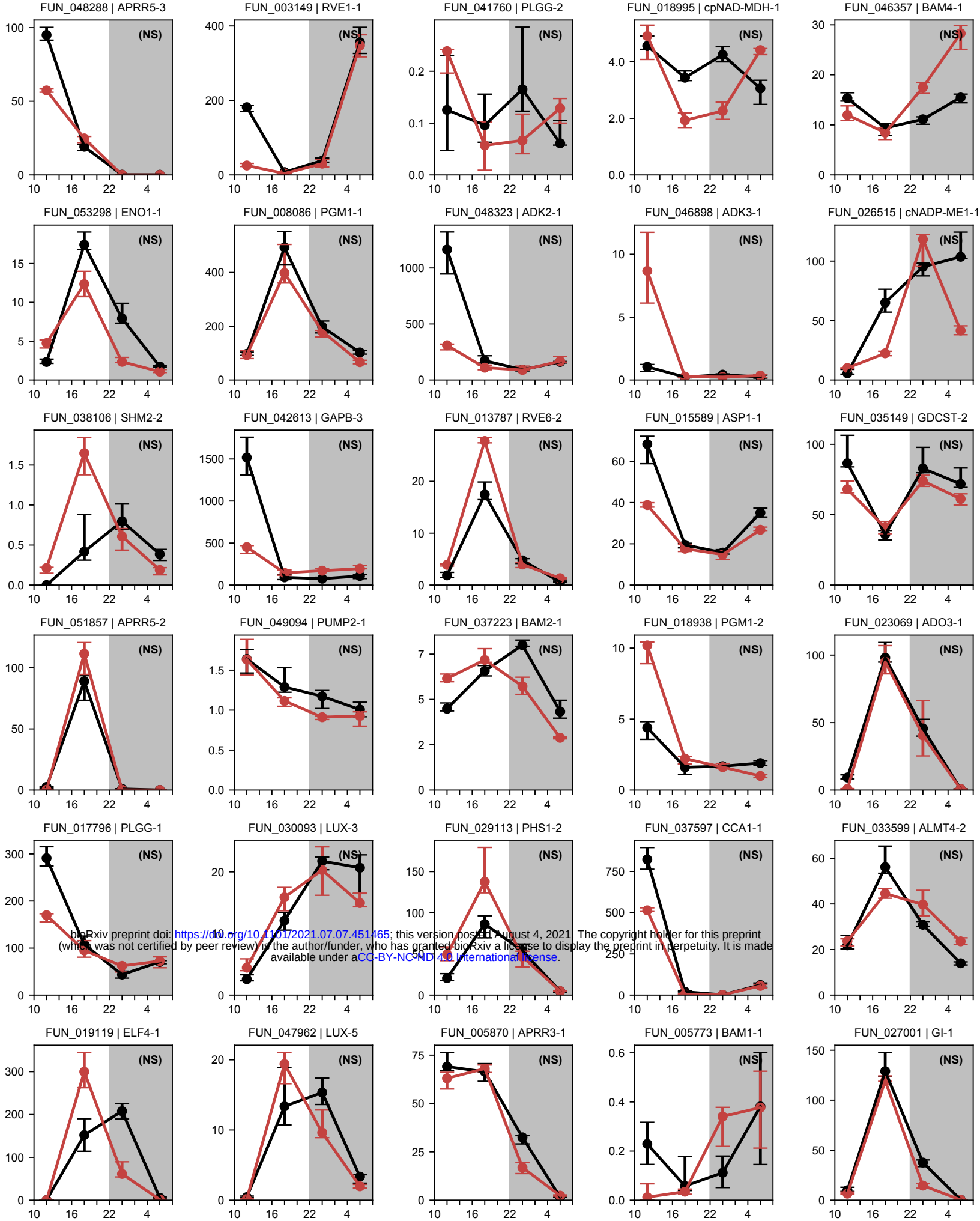

FUN_009450 | PPA2-1
FUN_048869 | ALMT1-3
FUN_020790 | PPC-1E1a'
FUN_010382 | ASP5-2
FUN_002078 | cpNADP-MDH1-1
FUN_003767 | BCA5-1
FUN_015896 | PFK1-3
FUN_029936 | PPC-1E1c
FUN_024067 | ALAAT1-1
FUN_047387 | PPT1-3
FUN_025702 | AMY2-1
FUN_010380 | ASP5-1
FUN_052022 | BCA5-3
FUN_004415 | LUX-1
FUN_030103 | RCA-2
FUN_049122 | PGLP2-3
FUN_033876 | TPT-1
FUN_034092 | SHM4-2
FUN_030703 | PPA2-2
FUN_000627 | PUMP1-1
FUN_049121 | PGLP2-2
FUN_000320 | BAM5-1
FUN_025472 | PUMP5-1
FUN_016561 | RBCS-B-1
FUN_026309 | BASS2-1
FUN_002164 | FBA1-1
FUN_004880 | PPDK-1
FUN_033111 | BAM1-3
FUN_033226 | BAM9-2
FUN_052713 | MEP4-2
***
10
16
22
4


FUN_014693 | GAPCP-1
FUN_022531 | GAPA1-1
FUN_009254 | BCA1-3
FUN_017370 | BAM3-1
FUN_003986 | AGT3-1
FUN_015913 | cpNADP-ME4-5
FUN_002831 | BCA1-1
FUN_030933 | PFK5-2
FUN_028976 | SHM2-1
FUN_015943 | PP2A3-1
FUN_010801 | PPT1-2
FUN_039301 | BCA1-2
FUN_021326 | PFK1-4
FUN_029870 | RBCS-B-2
FUN_052747 | DIT1-3
FUN_036400 | MEP4-1
FUN_042653 | CAT1-3
FUN_046866 | ALMT3-2
FUN_002470 | PPT2-1
FUN_033959 | LUX-4
FUN_050177 | ALAAT1-2
FUN_026973 | PPA6-1
FUN_032362 | PP2A1-3
FUN_025647 | CAT1-2
FUN_050236 | GATA24-2
FUN_005835 | FBA5-1
FUN_028977 | SHM1-1
FUN_053251 | FRS3-1
FUN_006404 | cpNADP-ME4-2
FUN_025211 | GATA24-1
***
10
16
22
4

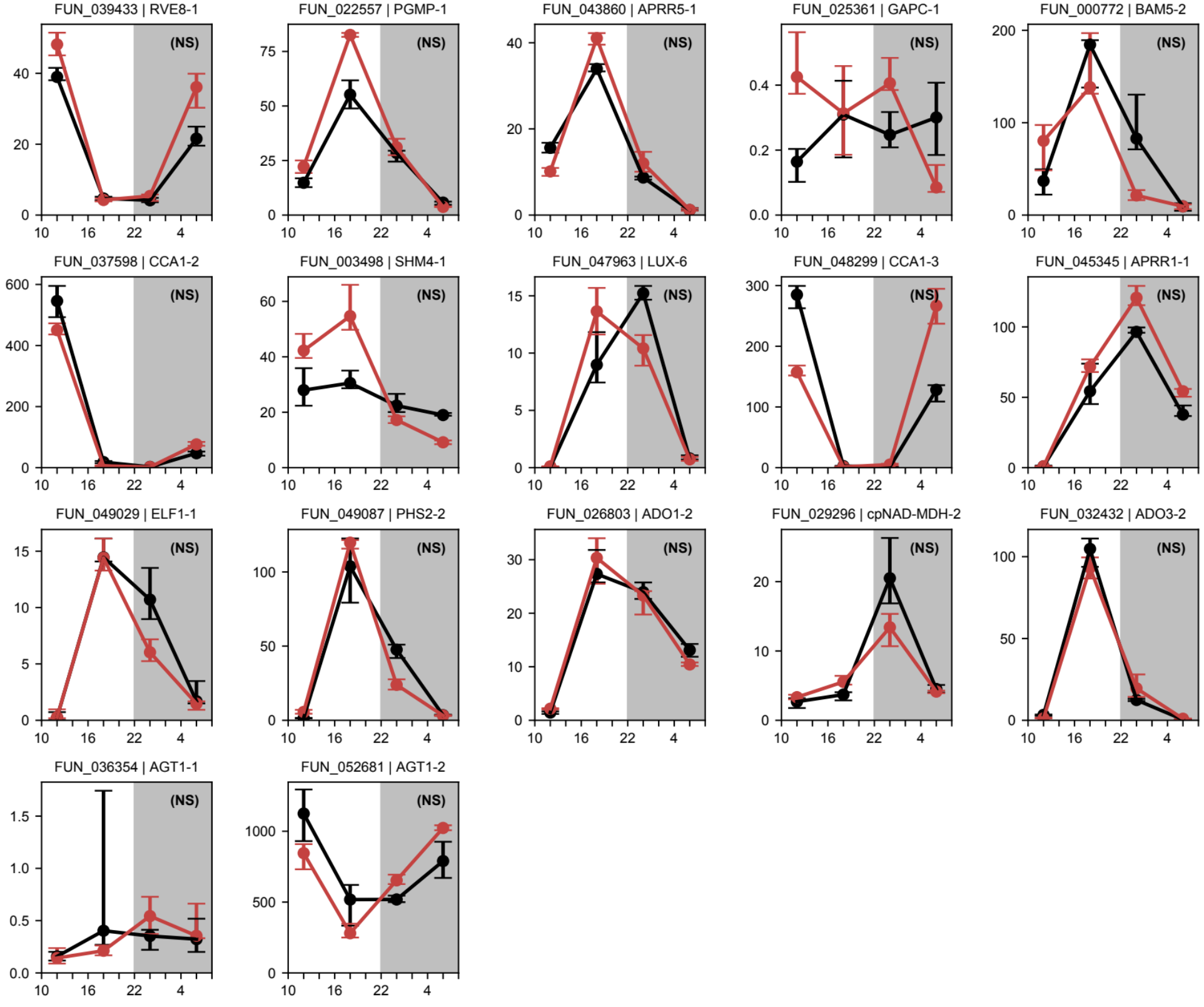

FUN_038622 | GATA25-1
FUN_010059 | RCA-1
FUN_023805 | PPC-2E
FUN_006401 | cpNADP-ME4-1
FUN_026477 | ASP3-1
FUN_036235 | PGLP1-1
FUN_039783 | FHY1-1
FUN_046955 | PPA1-2
FUN_018626 | PGK3-1
FUN_033324 | PGMC-1
FUN_025627 | GAPB-1
FUN_035711 | PP2A3-2
FUN_014237 | mtLPD1-1
FUN_036428 | DIT1-1
FUN_000897 | ASP2-1
FUN_006411 | cpNADP-ME4-3
FUN_020230 | PPDK-2
FUN_010229 | SHM6-1
FUN_026723 | PFK6-1
FUN_051815 | DIT2-3
FUN_018918 | AMY1-2
FUN_025653 | SBE2-1
FUN_018637 | BASS5-1
FUN_014695 | GAPCP-2
FUN_014956 | cpLPD2-1
FUN_032219 | BCA1-4
FUN_053289 | FAR1-1
FUN_026031 | ADK5-1
FUN_012333 | GDCST-1
FUN_053299 | FRS9-1
***
10
16
22
4

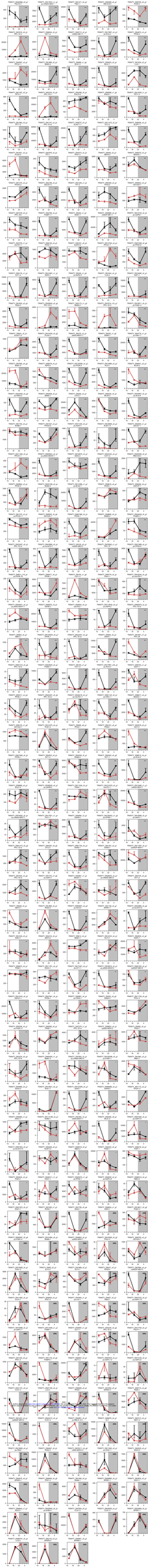

FUN_006423 | cpNADP-ME4-4
FUN_045586 | DIT2-1
FUN_025628 | GAPB-2
FUN_005224 | GLDP1-1
FUN_038563 | SHM2-3
FUN_053242 | FRS8-1
FUN_053310 | FRS4-1
FUN_016990 | PPC-1E1b
FUN_003773 | BCA5-2
FUN_029880 | RBCS-B-3
FUN_003919 | pNAD-MDH1-1
FUN_052748 | DIT1-4
FUN_019352 | ELF3-2
FUN_053071 | FHY3-1
FUN_002684 | PPCK1-1
FUN_005227 | GATA22-1
FUN_037145 | mtNAD-ME2-1
FUN_041354 | BAM3-2
FUN_010507 | ALMT3-1
FUN_005140 | ELF3-1
FUN_009081 | RVE6-1
FUN_018693 | CAT3-1
FUN_047630 | BASS1-1
FUN_015107 | cNAD-MDH1-1
FUN_049117 | PGLP2-1
FUN_037299 | RVE6-3
FUN_046701 | PPC-1E2
FUN_043261 | TPT-2
FUN_008400 | pNAD-MDH1-2
FUN_037221 | BAM7-1

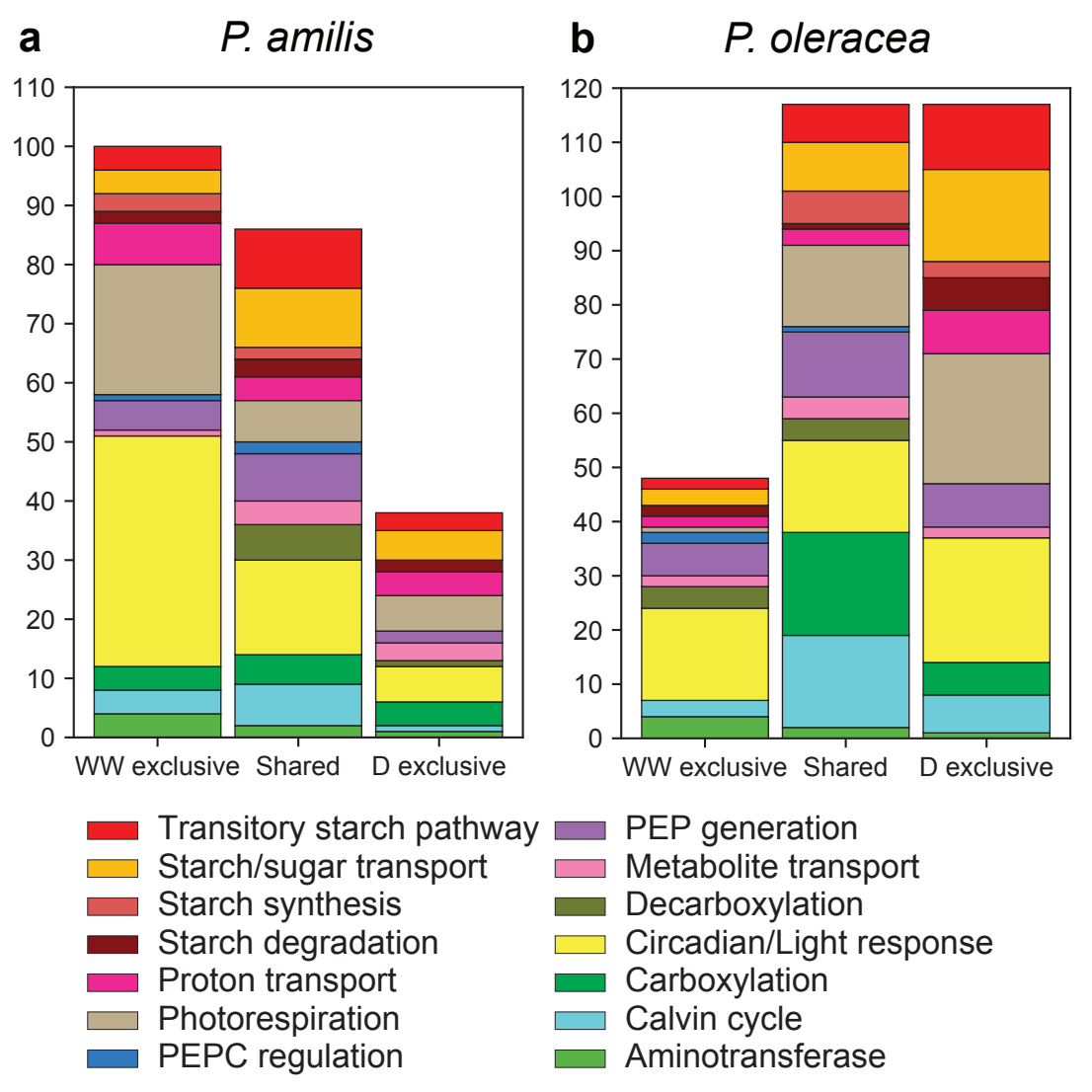

FUN_049101 | ENO3-1
FUN_021690 | FBA2-1
FUN_041931 | ENO2-1
FUN_049123 | PGLP2-4
FUN_011869 | mtNAD-MDH-2
FUN_011866 | mtNAD-MDH-1
FUN_041865 | BAM8-1
FUN_025469 | PFK5-1
FUN_029869 | RBCS-1A-1
FUN_023174 | AMY3-2
FUN_046883 | GAPC-2
FUN_014238 | mtLPD1-2
FUN_019059 | GDH2-1
FUN_021996 | mtNAD-ME1-1
FUN_039132 | FBA2-2
FUN_044076 | SBE3-1
FUN_006465 | PFK1-2
FUN_048305 | PPA1-3
FUN_002797 | PGK1-1
FUN_022923 | PP2A1-2
FUN_011473 | ADK3-2
FUN_048062 | RVE6-4
FUN_017464 | GDH1-1
FUN_030432 | GDH2-2
FUN_049126 | BAM2-2
FUN_053136 | FRS6-1
FUN_048884 | ALMT1-5
FUN_013066 | ADO1-1
FUN_039138 | SBE2-2
FUN_046463 | FBA5-3
(NS)
(NS)



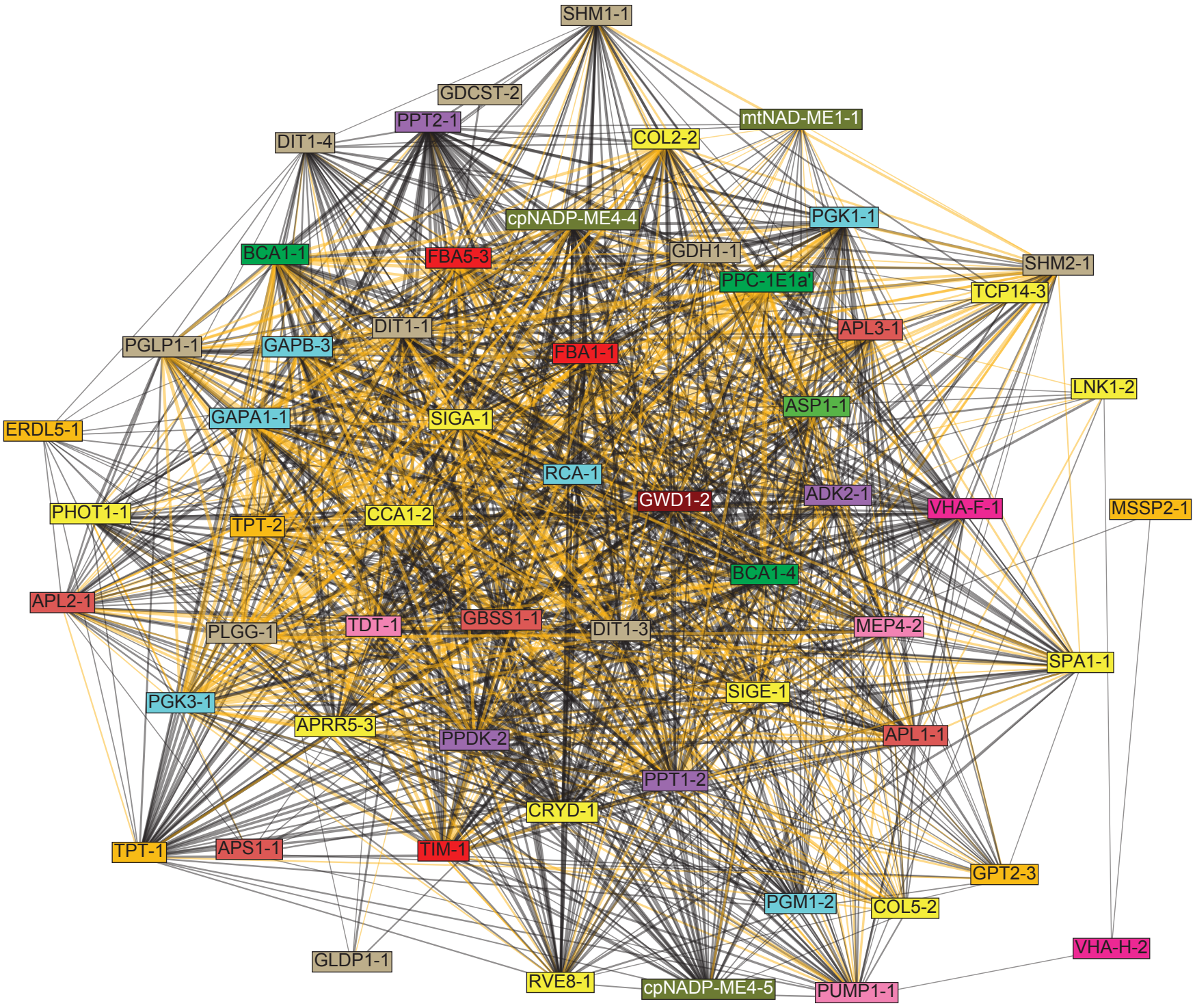
bioRxiv preprint doi: https://doi.org/10.1101/2021.07.07.451465; this version posted August 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


FUN_039433 | RVE8-1
FUN_022557 | PGMP-1
FUN_043860 | APRR5-1
FUN_025361 | GAPC-1
FUN_000772 | BAM5-2
FUN_037598 | CCA1-2
FUN_003498 | SHM4-1
FUN_047963 | LUX-6
FUN_048299 | CCA1-3
FUN_045345 | APRR1-1
FUN_049029 | ELF1-1
FUN_049087 | PHS2-2
FUN_026803 | ADO1-2
FUN_029296 | cpNAD-MDH-2
FUN_032432 | ADO3-2
FUN_036354 | AGT1-1
FUN_052681 | AGT1-2
(NS)
10
16
22
4

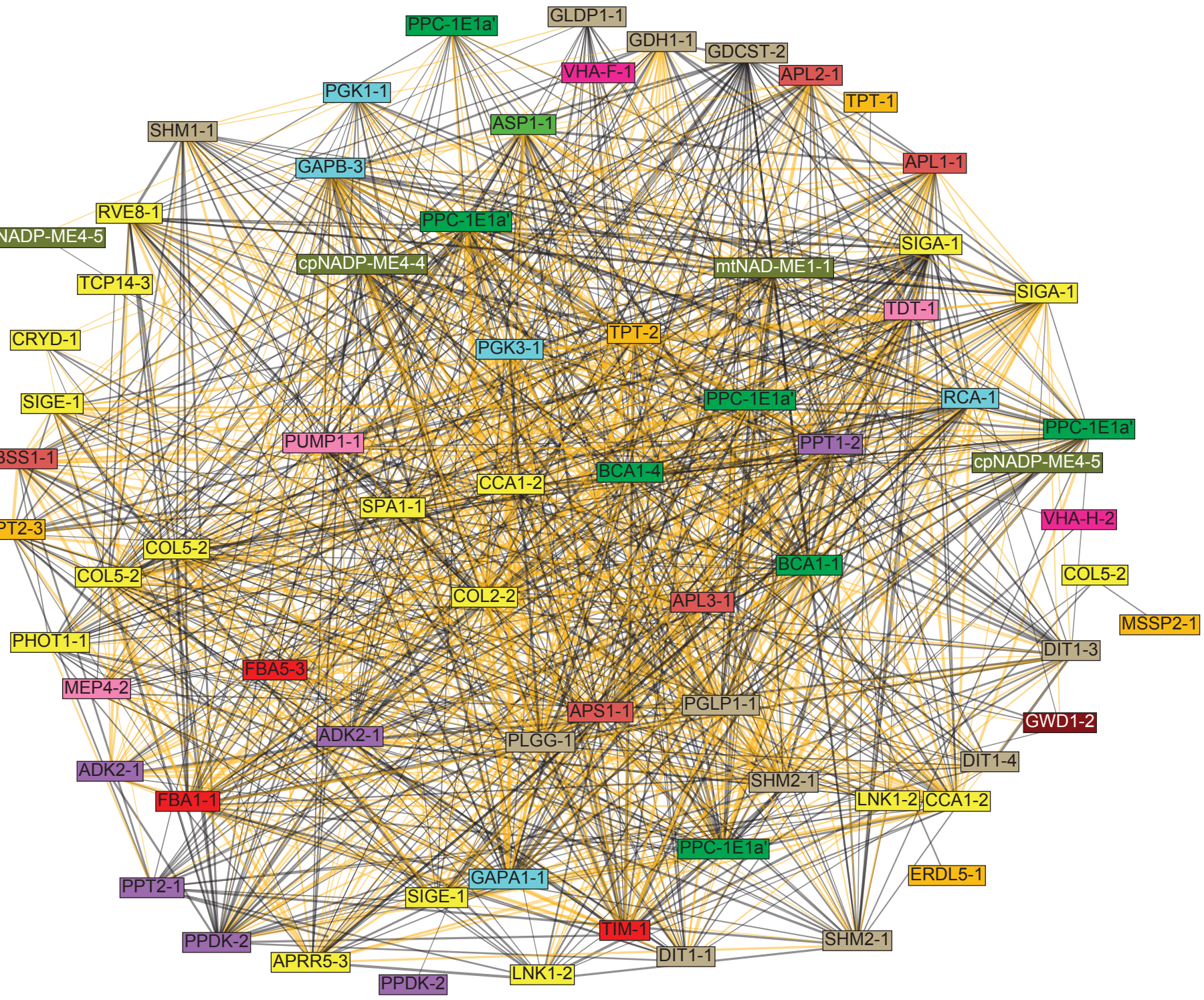



a
P. amilis
b
P. oleracea
110
100
90
80
70
60
50
40
30
20
10
0
120
110
100
90
80
70
60
50
40
30
20
10
0
WW exclusive
Shared
D exclusive
WW exclusive
Shared
D exclusive
Transitory starch pathway
Starch/sugar transport
Starch synthesis
Starch degradation
Proton transport
Photorespiration
PEPC regulation
PEP generation
Metabolite transport
Decarboxylation
Circadian/Light response
Carboxylation
Calvin cycle
Aminotransferase

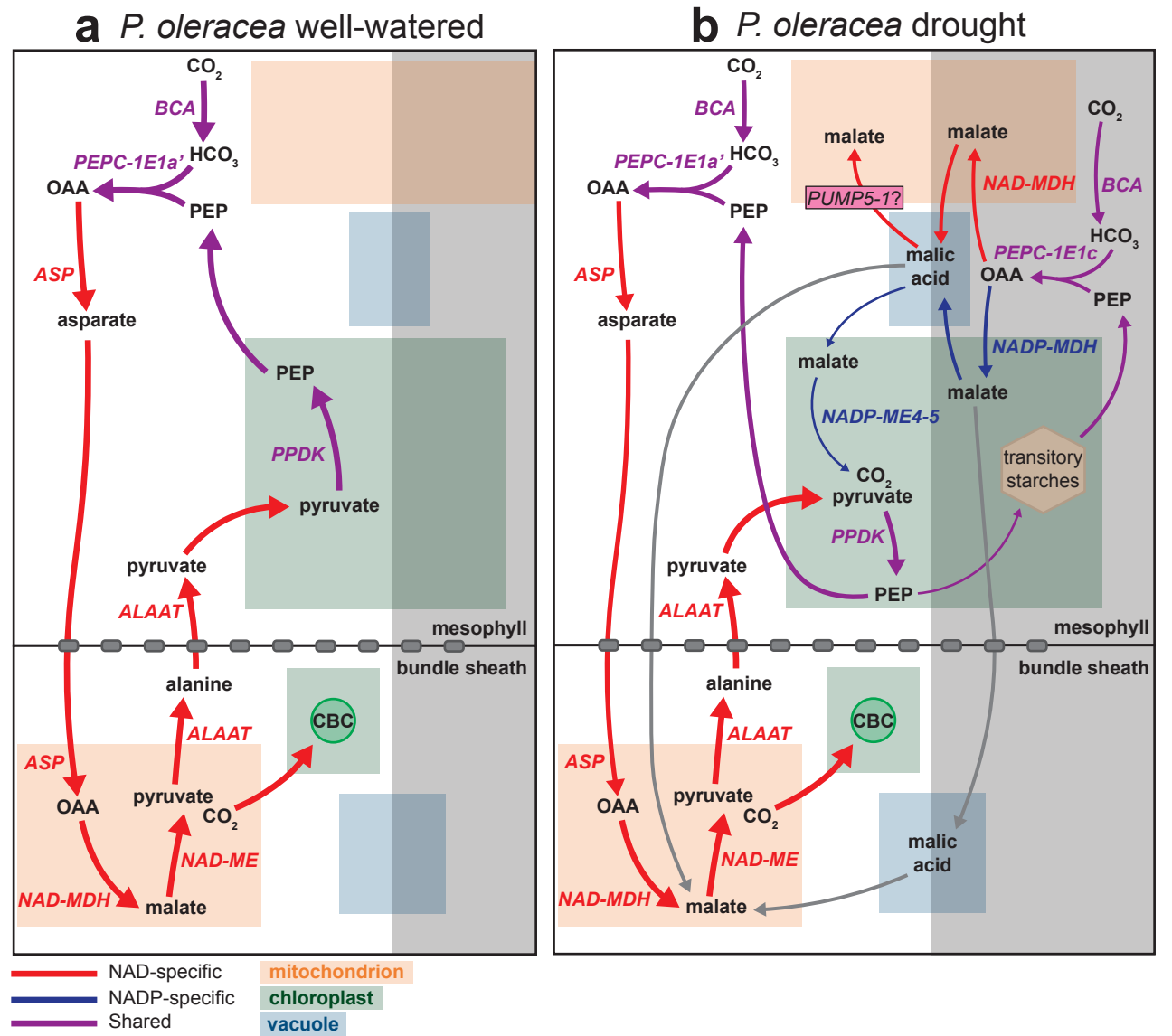
a *P. amilis* well-watered C_4 PGN



b *P. oleracea* well-watered C_4 PGN



Transitory starch pathway
Starch/sugar transport
Starch synthesis
Starch degradation
Proton transport
Photorespiration
PEPC regulation
PEP generation
Metabolite transport
Decarboxylation
Circadian/Light response
Carboxylation
Calvin cycle
Aminotransferase


a P. oleracea well-watered
b P. oleracea drought
CO2
BCA
HCO3
PEPC-1E1a'
OAA
PEP
ASP
asparate
PPDK
pyruvate
ALAAT
alanine
mesophyll
bundle sheath
CBC
NAD-ME
NAD-MDH
malate
PUMP5-1?
malic acid
PEPC-1E1c
NADP-MDH
NADP-ME4-5
transitory starches
NAD-specific
NADP-specific
Shared
mitochondrion
chloroplast
vacuole