Article: Discoveries

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

2

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

4

5 [1] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA; [2] Present
6 address: Nanostring Technologies, Seattle, WA, USA; [3] Department of Biology, University of Central
7 Florida, Orlando, FL, USA

8

9 Corresponding authors: Ian S. Gilman (ian.gilman@yale.edu), Erika J. Edwards
10 (erika.edwards@yale.edu)

Abstract

C_4 and Crassulacean acid metabolism (CAM) have been considered as largely independent photosynthetic 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 photosynthesis, 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 paralogs for primary carbon fixation and, although it is likely that nocturnal CAM malate stores are shuttled into diurnal C_4 decarboxylation pathways, we find some evidence that metabolite cycling may occur at low levels. C_4 likely evolved in *Portulaca* through co-option of redundant paralogs 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.

30 **Introduction**

31 Modularity and evolvability have become central research areas in biology over the past three decades
32 (Wagner and Altenberg 1996; Clune et al. 2013). Modularization generates simple building blocks that
33 facilitate rapid and complex adaptations through the combination of smaller functions and the ability of
34 simpler units to explore mutational space with minimal pleiotropic effects (Wagner and Altenberg 1996;
35 Lenski et al. 2003; Kashtan et al. 2005, 2007; Clune et al. 2013). Earlier literature on modularity focused
36 on genetic algorithms or digital organisms (Lenski et al. 2003; Kastan et al. 2005, 2007) and animal
37 morphological development (e.g., Davidson and Erwin 2006), but modern -omics strategies have enabled
38 detailed resolution of gene regulation in diverse study systems. Modeling results have shown that carbon
39 metabolic networks have high potential for exaptation into evolutionary innovations in bacteria (Barve
40 and Wagner 2013); and empirical research has identified deeply conserved carbon metabolic modules
41 from anaplerotic pathways that have been repeatedly recruited into plant photosynthetic adaptations
42 known as carbon concentrating mechanisms (CCMs) (Heyduk et al. 2019). CCMs provide classic
43 examples of rapid, adaptive evolution through the exaptation of existing metabolic modules that are
44 subsequently refined for new functions.

45 Since the early Oligocene, CCMs—typically taking the form of C_4 photosynthesis or
46 Crassulacean acid metabolism (CAM)—have evolved many dozens, if not hundreds, of times
47 independently (Keeley and Rundel 2003; Edwards et al. 2010; Christin et al. 2011; Sage 2016; Edwards
48 2019). Certain facets of CCMs, such as the main biochemical pathways, are evolutionarily accessible to
49 all green plant lineages because they belong to deeply conserved photosynthetic and respiratory gene
50 networks (Bräutigam et al. 2017; Heyduk et al. 2019). It is thought that by coupling these gene networks
51 to light responses (C_4) (Hibberd and Covshoff 2010) and circadian oscillators (CAM) (Hartwell 2006),
52 CCMs create two-stage carbon fixation pathways that address the most fundamental ecophysiological
53 tradeoff in land plants: balancing CO_2 influx with water loss to transpiration.

54 Most plants assimilate carbon using C_3 photosynthesis, in which CO_2 is directly fixed by the
55 enzyme RuBisCO during the day in mesophyll cells, and transformed into glucose in the Calvin-Benson
56 cycle. However, RuBisCO can also bind to O_2, triggering a set of costly reactions called photorespiration,
57 which is exacerbated by hot, dry conditions (reviewed in Peterhansel et al. 2010). C_4 and CAM increase
58 the efficiency of photosynthesis by increasing the concentration of CO_2 around RuBisCO. This is
59 achieved by first assimilating carbon with phosphoenolpyruvate (PEP) carboxylase (PEPC), which is
60 either spatially (C_4) or temporally (CAM) decoupled from final fixation of CO_2 by RuBisCO (fig. 1A and
61 B). In both CCMs, CO_2 is first converted into bicarbonate by beta carbonic anhydrase (BCA), and then
62 bicarbonate and PEP are transformed into oxaloacetate (OAA) by PEPC (fig. 1A and B). This occurs in
63 the mesophyll of both C_4 and CAM species, but exclusively during the night in CAM (fig. 1B),

64 whereafter OAA is reduced to malate by malate dehydrogenase (MDH) and stored in the vacuole
65 overnight as malic acid. During the day, CAM plants limit gas exchange with the external environment,
66 and instead decarboxylate stored malate with malic enzymes (MEs) to release CO_2 in the mesophyll—
67 therefore increasing the concentration of CO_2 around RuBisCO. Most C_4 species also convert OAA to
68 malate (termed NADP-type biochemistry), but others first transform it to aspartate (NAD-type
69 biochemistry) via aspartate aminotransferase (ASP). These 4-carbon acids are transported to the bundle
70 sheath cells of C_4 species, and are decarboxylated into CO_2 by MEs. In the case of NAD-type C_4 species,
71 ASP reversibly transforms aspartate back into OAA in bundle sheath mitochondria, where it is reduced to
72 form malate by NAD-MDH and finally decarboxylated by NAD-ME. By separating initial CO_2 fixation
73 in the mesophyll from ultimate assimilation in the bundle sheath, C_4 increases maximum photosynthetic
74 rates and water use efficiency while minimizing energy and carbon lost to photorespiration (Ogren 1984;
75 Sage 2004; Taylor et al. 2014). CAM does not always reduce photorespiration (Niewiadomska and
76 Borland 2007; Lüttge 2011), but greatly increases water use efficiency (Szarek and Ting 1975; Winter et
77 al. 2005) and reduces the harmful effects of high heat and radiation caused by oxidative stress
78 (Niewiadomska and Borland 2007).
79 C_4 and CAM boost the efficiency of photosynthesis in different ways and are therefore
80 hypothesized to have evolved in response to different stressors. C_4 has been primarily considered as an
81 adaptation to high rates of photorespiration (Ehleringer et al. 1997; Sage et al. 2012), and CAM to water
82 limitation (Raven and Spicer 1996; Keeley and Rundel 2003), though photorespiratory stress and drought
83 stress are often experienced simultaneously (Arakaki et al. 2011). Given the overlapping environmental
84 contexts of CAM and C_4 evolution, their repeated assembly of a similar set of biochemical modules
85 (Christin et al. 2014) (fig. 1A and B), and their apparent ease of evolution, it is perhaps surprising that use
86 of both C_4 and CAM (hereafter C_4+CAM) has only been reported in three lineages: *Portulaca*
87 (Portulacaceae) (Koch and Kennedy 1980), *Spinifex* (Poaceae) (Ho et al. 2019), and *Trianthema*
88 (Aizoaceae) (Winter et al. 2020). On the other hand, it may be precisely the large overlap in CCM genes,
89 gene networks, and metabolites that may make C_4 and CAM mutually exclusive in most lineages (Sage
90 2002). Plants that use both C_4 and CAM seemingly must regulate the same gene networks in often
91 contrasting patterns that may result in futile metabolite cycling and inefficient transport. For example,
92 daytime decarboxylation activity in the mesophyll during CAM would directly oppose the carboxylation
93 activity of C_4.
94 *Portulaca*, a geographically widespread clade of over 100 annual and perennial herbs and small
95 shrubs, predominantly use C_4, but facultatively exhibit CAM in response to drought (Koch and Kennedy
96 1980, 1982; Lara et al. 2003, 2004; Guralnick et al. 2007; Christin et al. 2014; Holtum et al. 2017; Ferrari
97 et al. 2020b, 2020c; Winter et al. 2019). It has been hypothesized that full C_4 and CAM cycles cannot co-

98 operate in the same mesophyll cells, and so the CAM cycle must be either spatially separated from the C_4
99 cycle, that is, relegated to a subpopulation of leaf cells such as water storage or bundle sheath (Sage 2002;
100 Guralnick et al. 2002, 2007); or alternatively, that *Portulaca* might operate a novel two cell CAM system,
101 whereby CAM malate eventually becomes decarboxylated in the bundle sheath (Lara et al. 2003, 2004).
102 Facultative CAM, the reversible induction of CAM in response to drought stress, is assumed to be
103 ancestral in *Portulaca* because it has been observed in every major sub-clade assayed (Holtum et al. 2017;
104 Winter et al. 2019) as well as in *Portulaca*'s closest relatives (Guralnick and Jackson 2001; Edwards and
105 Diaz 2006; Guralnick et al. 2007; Brillhaus et al. 2016; Holtum et al. 2018; Hancock et al. 2019; Holtum
106 et al. 2021) (fig. 1C). Transcriptomic studies of *P. oleracea* (Ferrari et al. 2020b, 2020c), and in a closely
107 related facultative CAM species, *Talinum* (Talinaceae) (Brillhaus et al. 2016), suggest that the ancestrally
108 inherited CAM pathway uses both NAD- and NADP-enzymes for CAM. The biochemical diversity of C_4
109 pathways in major *Portulaca* clades and the observation of intermediate photosynthetic phenotypes (C_3-
110 C_4) in the Cryptopetala clade (Ocampo et al. 2013; Voznesenskaya et al. 2017; Winter et al. 2019) imply
111 multiple origins of C_4 within a facultative CAM context (fig. 1C). Multiple C_4 origins are further
112 evidenced by major differences in leaf ultrastructure among *Portulaca* lineages (Ocampo et al. 2013;
113 Voznesenskaya et al. 2010, 2017). Most C_4 plants have high leaf vein density and large bundle sheath
114 cells—termed Kranz anatomy—to increase the efficiency of C_4 (Slack and Hatch 1967; Sage et al. 2012),
115 and these changes in leaf morphology occur early in C_4 evolution (Williams et al. 2012; Christin et al.
116 2013).
117 A detailed survey of photosynthetic gene activity has only been completed in one *Portulaca*
118 species, *P. oleracea* (Ferrari et al. 2020b, 2020c), which focused on CCM gene identification in an NAD-
119 type C_4 species. Similar experiments on other *Portulaca* species are needed to answer fundamental
120 questions about the function and evolution of C_4+CAM, and the integration of new metabolic modules
121 more generally. For example, facultative CAM is hypothesized to be ancestral to *Portulaca* based on
122 physiological studies and phylogenetic relationships, but expression of CAM-specific orthologs in
123 multiple species has not yet been documented. And although multiple carboxylation/decarboxylation
124 pathways can co-exist in some C_4 plants (Furbank 2011), it is not obvious why or how different C_4
125 biochemistries would evolve in *Portulaca* lineages that shared an ancestral, mixed CAM biochemistry
126 type, nor do we know the extent to which these CCMs have been integrated in various lineages.
127 Here, we present a chromosome-level genome assembly of *Portulaca amilis*, the first sequenced
128 genome of any C_4+CAM plant. We also analyzed transcriptomic data during a CAM-induction
129 experiment to understand the synergistic evolution of multiple metabolic modules. Concurrent
130 transcriptomic data for *P. oleracea* provided a comparison to assess the similarity of C_4+CAM across a
131 deep split in the *Portulaca* phylogeny and discriminate between ancestral, inherited and lineage specific

132 gene networks. Our experiments confirm hypotheses based on data from the 1KP Initiative (Matasci et al.
133 2014) that *Portulaca* species share one CAM-specific PEPC ortholog *(PPC-1E1c*) and another C_4-
134 specific ortholog *(PPC-1E1a'*) (Christin et al. 2014). Our analyses suggest that the nocturnal CAM gene
135 network is more highly conserved compared to the C_4 pathways of the two *Portulaca* species. The
136 components of this deeply conserved portion of the CAM pathway link PEPC to starch catabolism via
137 core circadian clock elements, lending evidence to previous hypotheses of CAM evolution (Bräutigam et
138 al. 2017). These nocturnal CAM networks were mostly preserved regardless of water status, but were
139 upregulated under drought. We found some evidence for CAM-specific decarboxylation pathways, but
140 the diurnal portion of CAM appears to have been largely integrated into the C_4 metabolism, and we
141 predict that nocturnal acids produced by CAM are likely shuttled into the diurnal pool of bundle sheath
142 C_4-acids. Finally, although *P. amilis* and *P. oleracea* C_4 gene co-expression networks share some central
143 C_4 orthologs, exclusive use of orthologs from most core gene families highlights the diversity of
144 C_4+CAM systems and provides further evidence that C_4 evolved largely independently in multiple
145 *Portulaca* lineages.
146
147 **Results**
148 *CAM induction experiment*
149 Significant diel fluctuations in titratable leaf acidity indicative of CAM activity were observed after seven
150 and eight days without water in *P. amilis* and *P. oleracea*, respectively (fig. 1D). Diel fluctuations in
151 titratable acidity were not significantly different from zero when well-watered (one sample *t*-test, $p > 0.05$
152 for both species), but were significantly different under drought ($p < 0.001$ for both species; fig. 1D).
153 There was no significant difference between the magnitude of diel fluctuations between species
154 (independent *t*-test, $p = 0.12$).
155
156 *Genome and transcriptome assembly and annotation*
157 The *Portulaca amilis* v1 genome assembly was 403.89 Mb in length and extremely contiguous (L50/N50
158 = 5 scaffolds/42.598 Mb; L90/N90 = 9 scaffolds/39.183 Mb), with nine primary scaffolds representing
159 the nine expected chromosomes of the haploid *P. amilis* genome (fig. 2A) based on the karyotype of a
160 close relative (*P. grandiflora*; Sharma and Bhattacharyya 1956). The scaffolds were 96.9% complete as
161 measured by the BUSCO v5 embryophyta database (n = 1,614 BUSCOs), and 92.9% of BUSCOs were
162 complete and single copy—further supporting the assembly as an accurate haploid representation of the
163 *P. amilis* genome. Only 4.0% of embryophyta BUSCOs were duplicated, 1.5% were fragmented, and
164 1.6% were missing; eukaryotic BUSCOs were similarly complete (C:98.0% [S:82.7%, D:15.3%], F:1.2%,
165 M:0.8%, n:255). Slightly less than half of the assembly was masked as repetitive elements (46.96%),

166 which primarily consisted of LTRs (41.96%), DNA elements (29.36%), and LINEs (13.85%) (fig. 2B). A
167 repeat landscape of these elements did not show any sudden bursts of element activity or multiple peaks
168 (fig. 2B), which can be signs of genomic upheaval due to hybridization or polyploidy events (McClintock
169 1984; Dion-Côté et al. 2014). The genome was annotated by integrating RNA-seq data from the CAM
170 induction experiment and multiple *ab initio* gene prediction pipelines to arrive at well-supported
171 consensus gene models (see Materials and Methods). The final annotation contained 53,094 gene models
172 with 58,732 unique coding sequences. These counts likely slightly overestimate the number of true gene
173 models and coding sequences due to incorrect splitting of some gene models (fig. 2C).
174 The *P. oleracea de novo* transcriptome assembly contained 413,658 transcripts corresponding to
175 230,086 "unigenes" (hereafter, "genes"). The assembly was 94.7% complete as measured by the BUSCO
176 v5 embryophyta database, but only 16.4% of BUSCOs were single copy, while the remaining 78.3% were
177 duplicated; another 3.2% were categorized as fragmented, and 2.1% were missing. These fractions were
178 mirrored in the eukaryotic BUSCOs (C:98.9% [S:16.9%, D:82.0%], F:0.0%, M:1.1%, n:255), and the
179 high rate of duplicate BUSCOs was expected as most members of the *P. oleracea* species complex are
180 polyploid (Ocampo and Columbus 2012). Less than half of the transcripts (198,026) were predicted to be
181 coding sequences, which corresponded to 45,285 genes.
182 A total of 21,621 *P. amilis* and 44,532 *P. oleracea* genes met the filtering criteria for differential
183 abundance analysis with sleuth (Pimentel et al. 2017), and over half of these genes (*P. amilis*, 62%; *P.*
184 *oleracea*, 51.0%) were found to be significantly differentially abundant between well-watered and
185 drought conditions following Benjamini Hochberg corrections (Benjamini and Hochberg 1995) for false
186 discovery ($q < 0.01$). Although we could not explicitly test cycling parameters of transcripts given our
187 sampling strategy, *P. oleracea* and *P. amilis* appeared to have slightly different rhythms of transcript
188 abundance. For transcripts with peak morning abundance, the final nocturnal sampling time points had
189 abundances much closer to morning levels in *P. oleracea*, resulting in many "U"-shaped abundance
190 profiles. This suggests either a slightly higher cycling frequency in *P. oleracea* or a phase shift earlier
191 relative to *P. amilis*.
192
193 *Co-expression network analysis*
194 After filtering and outlier detection within WGCNA (Langfelder and Horvath 2008), the final *P. amilis*
195 and *P. oleracea* co-expression datasets contained 26,760 genes for 24 samples and 26,337 genes for 22
196 samples, respectively. Simultaneous network construction and module detection partitioned genes into 41
197 well-watered and 21 drought *P. amilis* modules and 29 well-watered and 20 drought *P. oleracea* modules
198 (supplementary fig. S1A and B), where modules are mutually exclusive clusters of highly correlated
199 genes. In both species, the largest drought and well-watered modules were similar in size (i.e., number of

genes), but the well-watered networks contained many more small modules than the drought networks (supplementary fig. S1A and B). To focus on photosynthetic gene networks (PGNs), we used over 500 (*P. amilis*) and over 600 (*P. oleracea*) manually curated photosynthesis-associated nuclear genes to extract subnetworks by subsetting WGCNA adjacency matrices, and categorized the resulting genes into 14 functional groups (fig. 3).

The well-watered and drought *P. amilis* PGNs were similar in size and density, but had different distributions of node degree (the number of connections of a given gene) (supplementary fig. S1C). The well-watered *P. amilis* PGN had a lower median degree but higher mean degree compared to the drought PGN (supplementary table S1) due to a small number of extremely highly connected genes belonging to modules paWW1 and paWW3 (fig. 3A and supplementary fig. 1C). 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 (supplementary table S1). The 104 genes of paWW1 represented all functional groups except starch synthesis, and paWW1 was characterized by expression typical of the C_4 pathway: very high abundance early in the morning 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'*PNMthe ortholog of the C_4-specific *PPC* used by *P. oleracea* (Christin et al. 2014; Ferrari et al. 2020b, 2020c) (fig. 3C). The paWW1 module also contained many light response transcription factors, and the majority of photorespiratory genes (fig. 3A). 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) also contained *poPPC-1E1a'*, however the *P. oleracea* transcriptome contained five *poPPC-1E1a'* transcripts; four were recovered in poWW2 but one (TRINITY_DN22457_c0_g2) was recovered in poWW1. Both poWW2 and poWW1 had expression profiles consistent with C_4 photosynthesis and had similar functional compositions, but poWW2 had a more "U"-shaped normalized pattern with higher late dark period abundance (fig. 3E). In contrast to *P. amilis*, the largest *P. oleracea* module was in the drought PGN (poD4; fig. 3F), which maintained C_4-like expression and contained all five *poPPC-1E1a'* transcripts but no genes with decarboxylating functions, the largest fraction of which were found in poD3 (fig. 3F).

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* (Christin et al. 2014; Ferrari et al. 2020b, 2020c), *P. amilis* used the *PPC* paralog *paPPC-1E1c* for CAM (fig. 3D), and this gene belonged to smaller modules with abundance

234 profiles that peaked across the light-dark transition, and were dominated by starch degradation and core
235 circadian clock elements (paWW5 and paD4; figs. 3A, B, and D). Extracting the genes that remained co-
236 expressed (preserved) across well-watered and drought conditions resulted in 28 genes of the *paPPC-*
237 *1E1c* modules (42%) (fig. 4A). With the exception of one photorespiratory gene (*paCAT1-2*,
238 FUN_025647), all had functions related to carboxylation (*paPPC-1E1c* and *paBCA1-2*), starch
239 metabolism/catabolism, or circadian clock and light response. At least one paralog of all members of the
240 starch degradation pathway were preserved, except amylase (*AMY*) and beta-amylase (*BAM*), but one
241 *BAM* homolog (*paBAM3-2*, FUN_041354) was recovered exclusively in the drought module, along with
242 one paralog of aluminum activated malate transporter 4 (*paALMT4-2*, FUN_033599), which imports
243 malate into the vacuole. Contrary to our expectations, no *MDH* homologs were recovered in this
244 nocturnal CAM network, and most were recovered in drought modules paD6, paD2, and paD5 with C_4-
245 like expression. Of the 11 *paMDH* homologs, two chloroplastic paralogs (*paNAD-MDH-1*, FUN_011866,
246 paWW2, and paD13; *paNAD-MDH-2*, FUN_029296, paWW4, and paD3) exhibited nocturnal peak
247 abundance under both well-watered and drought conditions, but their peak abundances were orders of
248 magnitude lower than other, more highly abundant copies. And although there was high preservation of
249 the phosphorylative starch degradation pathway, little of the downstream pathway that builds PEP from
250 simpler sugars was recovered in the nocturnal CAM network.
251 The *P. oleracea* transcriptome contained multiple transcripts of *poPPC-1E1c* that were recovered
252 in two well-watered modules (poWW4 and poWW8) and three drought modules (poD2, poD5, and
253 poD9). All modules had expression profiles consistent with CAM expression, and peaked over the day-
254 night transition (fig. 3E and F). The number of preserved genes between these sets of well-watered and
255 drought modules (39) was similar to that of *P. amilis*, but the fraction of preserved genes in *P. oleracea*
256 was much lower (21.5%) because the *poPPC-1E1c* drought PGN (143 genes) was over three times the
257 size of *P. amilis* (45 genes). The preserved genes were functionally similar to those preserved in the *P.*
258 *amilis* nocturnal CAM network, and largely consisted of genes involved with starch
259 metabolism/catabolism, and circadian clock elements (fig. 4B). Among the genes preserved with *poPPC-*
260 *1E1c* was a cytoplasmic *poNAD-MDH1* (TRINITY_DN7462_c3_g1), which was the most highly
261 expressed *NAD-MDH* under drought. Furthermore, many of the genes that link phosphorylative starch
262 degradation to PEP generation were preserved in the nocturnal CAM network, including *AMY*, fructose-
263 bisphosphate aldolase (*FBA*), glyceraldehyde-3-phosphate dehydrogenase (*GAP*), and
264 phosphofructokinase (*PFK*). Although not preserved, the *poPPC-1E1c* drought modules importantly
265 included copies of phosphoglucomutase (*PGM*) and enolase (*ENO*), the final two steps in PEP generation
266 from starch or soluble sugars.

267 We used the *paPPC-1E1c* and *poPPC-1E1c* drought modules to identify the set of genes common
268 to both nocturnal CAM networks (fig. 4C and D). By taking the intersection of the two sets of *PPC-1E1c*
269 drought module genes (i.e., preserved and drought exclusive genes), we uncovered 21 *P. amilis* and 24 *P.*
270 *oleracea* orthologs (some genes had multiple paralogs in the *P. oleracea* network). These nocturnal CAM
271 networks contained one gene involved in starch synthesis (starch branching enzyme 3, *SBE3*), five
272 phosphorylative starch degradation genes, and alpha-glucan phosphorylase 2 (*PHS2*), which is also part
273 of phosphorylative starch metabolism; it reversibly transforms larger heteroglycans into smaller
274 heteroglycans and glucose 1-phosphate (fig. 4E). These common nocturnal CAM networks covered much
275 of the phosphorylative starch degradation pathway that forms the basis of PEP generation in most CAM
276 plants (Weiss et al. 2011; Borland et al. 2016), but the interactions among many genes were different
277 between species (fig. 4C and D). The largest group of shared genes were core circadian clock elements:
278 *Adagio/ZEITLUPE* (*ADO*/*ZTL*) *1* and *3*, two *Arabidopsis pseudo-response regulator 5* (*APRR5*) paralogs,
279 *EARLY FLOWERING* (*ELF*) *1* and *4*, *GIGANTEA* (*GI*), one *LUX ARRHYTHMO* (*LUX*) homolog, and
280 one *REVEILLE 6* (*RVE6*) paralog (fig 4C-E). Although the two species' nocturnal CAM PGNs shared
281 many key orthologs, there were some notable differences. The *P. amilis* CAM PGN was much smaller
282 (45 genes) than that of *P. oleracea* (143 genes), but the latter did not contain an *ALMT* homolog, while
283 the most highly expressed *paALMT* homolog (*ALMT4-2*) was recovered in the *P. amilis* CAM PGN. The
284 *P. oleracea* nocturnal CAM PGN contained three paralogs of *PPC kinase 1* (*PPCK1*) that were all highly
285 abundant, differentially expressed between well-watered and drought, and had peak abundance early in
286 the dark period (supplementary figs. S4). PPCK increases the activity of PEPC, which is inhibited by
287 malate, forming a negative feedback loop (Nimmo 2000). The single *PPCK* ortholog in *P. amilis*
288 (FUN_002684) was also significantly more abundant under drought (q = 0.001) and had peak nocturnal
289 expression (supplementary figs. S3), but it was recovered in drought module paD1, which increased in
290 abundance throughout the night (fig. 3B), unlike *paPPC-1E1c*. PPCK activity is typically essential for
291 highly active CCMs, and a CAM-like expression pattern in *paPPCK1* would suggest low *paPPC-1E1a'*
292 activity during C_4. However, we found a derived amino acid residue, glutamic acid (E), in *paPPC-1E1a'*
293 at the same location where an aspartic acid (D) residue (D509) was demonstrated to reduce malate
294 inhibition in *Kalanchoë* and *Phalaenopsis* orchids (Yang et al. 2017). In *Kalanchoë* and *Phalaenopsis*, the
295 D509 residue was shown to be derived from either arginine (R), lysine (K), or histidine (H), the latter of
296 which is present at position 509 in *poPPC-1E1a'*.
297 We considered two modules each in *P. amilis* (paWW1 + paWW3 and paD5 + paD6) and *P.*
298 *oleracea* (poWW2 + poWW1 and poD3 + poD4) when calculating the preservation of the C_4 pathway
299 because these pairs of modules had C_4-like expression patterns and were strongly connected (fig. 3A and
300 E). In the *P. amilis* C_4 PGNs, 86 genes (38%) were preserved across well-watered and drought conditions

301 (supplementary fig. S4). This preserved set contained representatives of all core C_4 (*ALAAT*, *ASP*, *BCA*,
302 *ME*, *MDH*, *PPC-1E1a'*) and PEP regeneration gene families (adenylate kinase (*ADK*), sodium/pyruvate
303 cotransporter (*BASS*), sodium-hydrogen antiporter (*NDH*), soluble inorganic pyrophosphatase (*PPA*),
304 pyruvate, phosphate dikinase (*PPDK*), and PEP/phosphate translocator (*PPT*)). It also contains members
305 of the entire transitory starch pathway that builds and transports various sugars and carbohydrates (e.g.,
306 *ENO*, *GAP*, phosphoglycerate kinase (*PGK*), *PGM*, and triose phosphate/phosphate translocator (*TPT*)).
307 In addition to *paBASS2-1* (FUN_026309), which imports pyruvate into the chloroplast as a substrate for
308 PEP generation, we recovered two paralogs of methyl erythritol phosphate/*RETICULATA-RELATED 4*
309 (*MEP4*/*RER4*; *A. thaliana* ortholog AT5G12470). MEP transporters have been implicated as mesophyll
310 pyruvate importers and bundle sheath pyruvate exporters in maize (Friso et al. 2010; Chang et al. 2012),
311 and thus may serve a similar function to BASS. The preserved modules also contained key enzymatic
312 regulators (e.g., serine/threonine-protein phosphatase (*PP2A*) subunits and RuBisCO activase (*RCA*)) and
313 transcription factors and clock elements (e.g., *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*)).
314 *Portulaca amilis* and *P. oleracea* span a fairly deep node in the *Portulaca* phylogeny (fig. 1C)
315 and it was unclear which facets of C_4 they might share. A similar fraction of C_4 module genes were
316 preserved between well-watered and drought C_4 PGNs in *P. oleracea* (117; 41.2%), and these largely
317 consisted of the same gene families (58.3% overlap with preserved *P. amilis* C_4 gene families). To
318 identify the common orthologs used in the two species, we took the intersection of the genes in their well-
319 watered C_4 PGNs. The *P. amilis* and *P. oleracea* well-watered C_4 PGNs contained 186 and 165
320 homologs, respectively, which represented 258 unique ortholog groups, as many single copy genes in *P.*
321 *amilis* have multiple paralogs in the *P. oleracea* transcriptome. Roughly one quarter (26.7%) of these
322 ortholog groups (69/258) were recovered in the common C_4 networks, while 117 (45.3%) and 72 (27.9%)
323 were exclusive to *P. amilis* and *P. oleracea*, respectively. The common networks included many of the
324 most highly expressed copies of key C_4 genes, such as *PPC-1E1a'*, *BCA1-1*, *BCA1-4*, *ASP1-1*, *ASP5-2*,
325 *mtNAD-ME1-1*, *cpNADP-ME4-4*, and *cpNADP-ME4-5*; supplementary fig. S5). They also contained the
326 PEP generation pathway, including *ADK2-1*, *BASS2-1*, *MEP4-2*, *PPT1-2* and *PPT2-1*, and *PPDK-2*. The
327 photorespiratory pathway was also largely shared, including glycine decarboxylase subunits H (*GDH1-1*),
328 P (*GLDP1-1*), and T (*GDCST-2*), two homologs of serine hydroxymethyltransferase (*SHM1-1* and
329 *SHM2-1*), phosphoglycolate phosphatase 1 (*PGLP1-1*) and 2 *(PGLP2-3*), plastidal glycolate/glycerate
330 translocator (*PLGG-1*) and three copies of dicarboxylate transporter 1 *(DIT1-1*, *DIT1-3*, and *DIT1-4*).
331 Indicative of the unique evolutionary histories of *P. amilis* and *P. oleracea*, each species' C_4 PGN
332 contained exclusive homologs of most C_4 gene families, some of which represent species-specific
333 duplication events. Both species exhibited exclusive use of orthologs of *ADK*, *ALAAT*, *ASP*, *BCA*, *MDH*,
334 *ME*, *NDH*, *PP2A*, *PPA*, mitochondrial uncoupling protein (*PUMP*), RuBisCO, and *RCA*, among others.

335 Some exclusive homologs fit biochemical expectations. For example, exclusive to *P. amilis* (NADP-type
336 C_4) were three chloroplastic *NADP-ME4* paralogs (FUN_006401, FUN_006404, and FUN_006411)
337 which did not have orthologs in the *P. oleracea* transcriptome. These three copies were on the same
338 strand and physically close in the genome (within ~85kb); they possibly represent tandem duplication
339 events of *paNADP-ME4-4* (FUN_006423), which is ~50kb downstream of FUN_006411 (fig. 2A). In *P.*
340 *oleracea*, two *ASP* (TRINITY_DN71325_c0_g1 and TRINITY_DN346_c1_g3) and three *ALAAT*
341 (TRINITY_DN43865_c0_g1, TRINITY_DN43865_c0_g2, and TRINITY_DN81553_c0_g1) homologs
342 were found in orthogroups that had no ortholog in the *P. amilis* genome. Many of the genes that were
343 exclusive to either C_4 PGN were transcription factors involved in the circadian clock and light response,
344 or regulatory enzymes. For example, unique orthologs of the *FAR-RED IMPAIRED RESPONSE* family
345 were part of each C_4 PGN, while only the *P. amilis* C_4 PGN contained an *APRR* homolog, and only the *P.*
346 *oleracea* C_4 PGN contained a *LATE ELONGATED HYPOCOTYL* (*LHY*) transcription factor.
347 We also used the intersection of the drought C_4 PGNs to identify potential ancestral elements of
348 the daytime portion of the CAM cycle. We expected this portion of the CAM cycle (decarboxylation of
349 malate leading to the release of CO_2) to be nearly indistinguishable from the C_4 cycle (fig. 1A and B), but
350 to only exhibit this expression pattern when elicited by drought stress. To be considered part of the
351 ancestral CAM cycle, elements needed to meet the following criteria: 1) be shared between the two
352 drought C_4 PGNs and 2) be significantly more abundant under drought. Eleven orthologs met these
353 potentially CAM-induced criteria, and only one had a function directly related to daytime CO_2 release
354 from nocturnal malate stores. Chloroplastic malic enzyme *NADP-ME4-5* (FUN_015913;
355 TRINITY_DN81025_c0_g1) was shared between both species' C_4 drought PGNs and upregulated under
356 drought ($q < 0.01$). *PUMP5-1*—also known as dicarboxylate transporter 1 (*DIC1*)—of both species also
357 appeared CAM-induced, but was not significantly differentially abundant when considering the entire
358 time series in *P. oleracea* ($q = 0.24$). The *P. amilis* C_4 PGNs contained a drought exclusive mitochondrial
359 malic enzyme (*NAD-ME2-1*; FUN_037145) that was significantly upregulated ($q < 0.01$), and a second,
360 preserved ortholog (*NAD-ME1-1*; FUN_021996) that was significantly upregulated under drought ($q <$
361 0.05). Coupled with upregulated *PUMP5-1* expression, this may suggest that the NAD decarboxylation
362 pathway, which occurs mainly in the mitochondria, may still be elicited by CAM in *P. amilis* despite
363 NADP-type C_4. However, three photorespiratory genes were also in this group: alanine-glyoxylate
364 aminotransferase 3 (*AGT3*), *SHM2-3*, and dihydrolipoyl dehydrogenase/glycine cleavage system L
365 subunit 2 (*LPD2*). Therefore, an alternative hypothesis is that increased *PUMP5-1* abundance in concert
366 with photorespiratory genes indicated greater photorespiratory and oxidative stress under drought.
367 Finally, and unexpectedly, one *P. amilis* paralog of the RuBisCO small subunit B (*paRBCS-B-1*;

368 FUN_016561) fit a CAM induced abundance pattern ($q < 0.001$) and was expressed on the same order of
369 magnitude as other *RBCS* homologs.
370
371 *Motif identification and enrichment*
372 Enriched motifs were detected in at least one region (i.e., 3'UTR, 5'UTR, introns, or upstream promoter)
373 for gene models in many *P. amilis* modules. However, no enriched motifs were identified for paWW1 or
374 paWW3—which contained the bulk of C_4 genes when well-watered. In contrast, the 5'UTRs and
375 upstream promoter regions of the preserved *paPPC-1E1c* module were enriched in 10 unique motifs ($p <$
376 0.01), all of which had circadian or light response functions (Table 1). These motifs were all annotated to
377 Myb-related families, except bHLH74, which contains a G-box motif (CACGTG), and plays a role
378 governing flowering time in response to blue light (Liu et al. 2013). All enriched motifs contained either
379 an Evening Element (EE; AAATATCT) or GATA (HGATAR) motif. EE and GATA motifs often serve
380 as binding sites for core circadian oscillator transcription factors such as RVEs, CCA1, LHY, and
381 APRR1/TIMING OF CAB 1 (TOC1) (Reyes et al. 2004; Gendron et al. 2012; McClung 2019).
382 There were no enriched motifs in or near genes preserved with *paPPC-1E1a'*, and the *P. amilis*
383 C_4 PGN genes were only enriched in four motifs: one Myb (MA1028.1), two bZIP (MA0968.1,
384 MA1033.1), and one AP2/ERF (MA1261.1) factor (supplementary table S2). While these motifs were
385 significantly enriched ($p < 0.01$), they were much less widespread than the enriched motifs found in the
386 preserved CAM genes. At the low end, the KAN4 motif (MA1033.1) was present in just five 5'UTRs
387 (8.5% of sequences), and the most widespread (bZIP68, MA0968.1) was present in 12 sequences
388 (20.3%). Motifs found in the preserved *paPPC-1E1c* genes were present in 20.7-58.6% of the sequences,
389 and all 10 enriched motifs were found at least once in the 5'UTR and promoter region of *paPPC-1E1c*.
390
391 **Discussion**
392 Plant photosynthetic adaptations can serve as models for exploring many central themes of biology, such
393 as the evolution of modularity and evolvability, gene and genome duplication, convergent evolution, and
394 the evolution of novel phenotypes. Network analyses provide powerful tools for identifying and
395 comparing modules of genes related to functions of interest against a background of tens of thousands of
396 genes with complicated temporal expression patterns. C_4+CAM photosynthesis has only been reported in
397 three plant lineages, *Portulaca* (Koch and Kennedy 1980), *Spinifex* (Ho et al. 2019), and *Trianthema*
398 (Winter et al. 2020), but carbon concentrating mechanisms (CCMs) have evolved many dozens of times
399 (Keeley and Rundel 2003; Edwards et al. 2010; Christin et al. 2011; Sage 2016; Edwards 2019) and the
400 use of multiple CCMs has been documented in aquatic plant lineages (Zhang et al. 2013; Huang et al.
401 2018). More generally, CCM evolution is a classic example of the repeated assembly of a new function

402 via the integration of multiple, existing gene networks (Heyduk et al. 2019). When coupled with patterns
403 of transcript abundance and placed in an evolutionary context, our network results provide an initial
404 interpretation of how these metabolic pathways interact in *Portulaca,* and how they were evolutionarily
405 assembled.
406
407 *The evolution of C_4+CAM in Portulaca*
408 A detailed map of the transitions from facultative CAM to the diversity of C_4+CAM (including C_3-C_4)
409 pathways of extant *Portulaca* is not possible given a comparison of just two species, but we can identify
410 key commonalities and differences. There is good evidence that facultative CAM is ancestral in
411 *Portulaca*, and evolved by coupling elements of the TCA cycle (i.e., PEPC and BCA) to the
412 phosphorylative starch degradation pathway via a core circadian oscillator. CAM in both species was
413 restricted to a minimal set of tightly regulated nocturnal pathways that link *PPC-1E1c* to the transitory
414 starch pathway, with only parts of the decarboxylation pathways distinguishable from C_4. This aligns with
415 the hypothesis proposed by Bräutigam et al. (2017) that the initial steps in CAM evolution do not require
416 any "rewiring" of gene networks, but rather occur by upregulating nocturnal organic acid production that
417 occurs in all plants to some degree. Subsequent canalization of these fluxes occurs by bringing these
418 networks under circadian control. If C_4 evolution began with such a partial CAM system in *Portulaca*,
419 where the nighttime portion of the cycle included derived gene networks but the daytime portion did not,
420 CAM would have likely imposed few pleiotropic constraints on C_4 evolution given sufficient redundant
421 paralogs. In this scenario, recruitment of genes into C_4 would proceed as in other C_4 lineages (reviewed in
422 Heyduk et al. 2019), and no modification of existing CAM decarboxylation pathways in the mesophyll
423 would be necessary. However, transcriptomic data presented here and elsewhere (Brilhaus et al. 2016;
424 Ferrari et al. 2020b) imply that CAM-specific decarboxylation pathways did exist in the common ancestor
425 to *Portulaca*. It therefore seems likely that the initial steps towards C_4 evolved within a narrow window of
426 CAM evolution, before other aspects of CAM were fixed, such as diurnal stomatal closure or changes to
427 leaf ultrastructure that greatly reduced intercellular diffusion of metabolites.
428 Congruent with previous studies (Christin et al. 2014; Ferrari et al. 2020b, 2020c), we found that
429 C_4 and CAM in *Portulaca* use different paralogs for initial carboxylation (e.g., PEPC, ME, ASP,
430 MDH)—thus avoiding potential pleiotropic constraints on individual genes. Furthermore, sufficient
431 paralogs of NAD- and NADP-type enzymes were present in the most recent common ancestor of *P.*
432 *amilis* and *P. oleracea* to allow differential enzyme co-option during their independent C_4 origins.
433 However, use of *PPC-1E1a'* for C_4 in both species suggests a bias in gene recruitment that would indicate
434 some "preconditioning" for C_4. Preconditioning of a particular paralog could take multiple forms, such as

435 increased expression (Moreno-Villena et al. 2018) or adaptive amino acid substitutions (Christin et al.
436 2014), which would lead to multiple parallel recruitments of PPC-1E1a' into C_4.
437 Shared use of photorespiratory modules found here and proto-C_4 anatomy in C_3-C_4 intermediate
438 taxa (Voznesenskaya et al. 2010, 2017) suggest that *Portulaca* may share further steps towards C_4, such
439 as C_2 metabolism. C_2 metabolism refers to the enhanced refixation of photorespired CO_2 by preferential
440 expression of various photorespiratory enzymes in mesophyll and bundle sheath (Sage et al. 2012;
441 Mallmann et al. 2014). C_2 metabolism activity defines Type I C_3-C_4 intermediates, such as the members
442 of the Cryptopetala clade (Voznesenskaya et al. 2010, 2017), and like other CCMs, requires the
443 recruitment of existing gene networks for specialized expression patterns. Our comparison of well-
444 watered C_4 gene networks revealed many shared orthologs of photorespiratory genes, including subunits
445 of glycine decarboxylase, SHM, and PLGG that may point to a single origin of C_2 in *Portulaca*. However,
446 these genes tended to belong to smaller gene families than many C_4 and CAM genes, and it is difficult to
447 distinguish between a scenario in which these paralogs were co-opted into C_2 metabolism once near the
448 stem of extant *Portulaca*, or more recently along multiple branches in parallel. C_2 metabolism is
449 hypothesized to typically be followed by rapid evolution of a C_4 system (Sage et al. 2012; Mallman et al.
450 2014), and therefore represents a limit of what we expect to be shared among *Portulaca* lineages with
451 distinct morphologies and biochemistries. Broader sampling across the *Portulaca* phylogeny is needed to
452 precisely delimit the shared initial steps towards C_4 and subsequent parallel refinement of C_4+CAM.
453
454 *Integration of C_4 and CAM in Portulaca*
455 *Portulaca amilis* and *P. oleracea* have canonical C_4 pathways, but constitutively operate very weak CAM
456 cycles (at least at the transcriptional level) that are upregulated under drought. The core elements of C_4
457 and CAM were moderately preserved in both species, and where parts of the C_4 and CAM pathways
458 could theoretically overlap, we often found alternative paralogs. Increased CAM activity was associated
459 with a reduction in size of C_4 PGNs, but there was no evidence that this was due to dual roles of genes in
460 nocturnal CAM physiology under drought. The primary reason for reduced correlations between C_4 genes
461 was significant downregulation under water limitation. If transcript abundances of C_4 and CAM related
462 genes can serve as a proxy for enzymatic and metabolic activity, conflicting intracellular physiological
463 cues from CAM likely played a minor role in altering C_4 activity. Our investigation into possible drought
464 induced diurnal CAM genes found evidence that diurnal elements of the CAM pathway could be
465 identified against the strong background activity of C_4 decarboxylation, but only at low levels. Drought-
466 induced upregulation of chloroplastic *NADP-ME4-5* in both species and mitochondrial *paNAD-ME1-1*
467 and *paNAD-ME2-1* in *P. amilis* suggests that C_4 did not co-opt the entirety of the diurnal portion of the
468 CAM cycle for decarboxylation. If *Portulaca* had fully recruited the CAM decarboxylation pathways into

469 the C_4 system, we would not expect upregulation of any decarboxylation pathway under drought. In
470 contrast, constitutive C_4-like expression of the same mitochondrial orthologs in *P. oleracea* (*poNAD-*
471 *ME1-1* and *poNAD-ME2-1*) is consistent with use in C_4, suggesting that *P. oleracea* may have recruited
472 its primary ancestral CAM decarboxylating pathway into its C_4 system.
473 During the day, a CAM-specific decarboxylation pathway could result in futile carboxylation and
474 decarboxylation fluxes if present in the mesophyll. However, a comparison of *ME* abundance levels
475 suggests that any CAM related decarboxylation would be orders of magnitude below C_4 pathway activity.
476 Given the strong diffusion gradients between mesophyll and bundle sheath of C_4 species (Leegood 1985),
477 we suspect that the evolution of C_4 in *Portulaca* has resulted in two-cell CAM, as has been previously
478 hypothesized (Lara et al. 2004), where malate produced by CAM in the mesophyll is largely shuttled to
479 the bundle sheath to be decarboxylated, but it is unclear if overnight storage occurs in the mesophyll,
480 bundle sheath, or both (fig. 5). While likely functionally insignificant, the signal of a second
481 decarboxylation pathway suggests that the CAM decarboxylation modules were not entirely repurposed to
482 the bundle sheath during the evolution of C_4. Rather, an entirely new decarboxylation pathway was
483 assembled in *P. amilis*, and some decarboxylation elements (i.e., *PUMP5-1*, *NADP-ME4-5*) were not
484 recruited into C_4 in *P. oleracea* (supplementary fig. S6).
485 Our network results confirm that the nocturnal portion of CAM is a fixed gene network—rather
486 than the result of drought-induced rewiring—that is both transcriptionally and post-transcriptionally
487 regulated under drought to increase CAM activity. Gas exchange measurements in other *Portulaca*
488 species (Holtum et al. 2017; Winter et al. 2019; Ferrari et al. 2020b, 2020c) show substantial net CO_2 loss
489 during the night when well-watered. Therefore, well-watered expression levels of the preserved CAM-
490 related genes are either not high enough to re-fix all nocturnally respired CO_2 or must have their products
491 regulated at a post-transcriptional level. Evidence for the latter is given by expression patterns of PPCK,
492 which increases the activity of PEPC (Nimmo 2000). *PPCK* was upregulated in both species under
493 drought, and during the dark period in particular, which implies a role in increasing nocturnal PEPC
494 activity for CAM under drought. We were surprised to find just a single *PPCK* ortholog in *P. amilis*, and
495 are not aware of other C_4 species without high morning expression of *PPCK*. The glutamic acid (E)
496 residue at position 509 in *paPPC-1E1a'* warrants further investigation because an aspartic acid residue at
497 this position has been shown to increase PEPC activity by reducing malate inhibition (Yang et al. 2017).
498 Glutamic acid is functionally similar to aspartic acid, and may similarly reduce malate inhibition.
499 Previous transcriptomic studies (Brillhaus et al. 2016; Ferrari et al. 2020b) suggest that *Portulaca*
500 inherited a CAM cycle that used both NAD- and NADP-type enzymes, but it was not clear how the
501 evolution of divergent C_4 biochemistries would affect CAM biochemistry. Water limitation studies in
502 *Portulaca* have shown similar CAM responses (i.e., titratable acidity and net gas exchange) across

503 species, including in C_3-C_4 *P. cryptopetala* (Winter et al. 2019). This provides evidence that carbon
504 concentrating mechanisms function mostly independently of one another in *Portulaca*, and are not
505 strongly affected by C_4 (or C_2) biochemistry type. Prior to our CAM induction experiment in *P. amilis*,
506 transcriptomic experiments had been restricted to *P. oleracea* (Ferrari et al. 2020b, 2020c), focused on
507 other traits, such as flowering in *P. grandiflora* (Maguvu et al. 2020), or used transcriptomic data (i.e.,
508 1KP Initiative) collected at one time point from well-watered plants (Christin et al. 2014) that could not
509 link orthologs to metabolic modules. The addition of *P. amilis* (NAD-type C_4 species) allowed us to test
510 the expectation of differential loss and gain of mesophyll- and bundle sheath-specific reactions in *P.*
511 *amilis* and *P. oleracea* due to their different C_4 biochemistries (fig. 5). Our findings show that these two
512 C_4+CAM systems appear to have evolved through differential gene family expansion and paralog use.
513 At the transcriptional level, both *P. amilis* and *P. oleracea* used NADP- and NAD-type
514 biochemistry, but expression was biased toward NADP-type enzymes in *P. amilis* and, to a lesser extent,
515 toward NAD-type enzymes in *P. oleracea*. We found only one *NADP-MDH* ortholog in each species, but
516 it was the most abundant *MDH* in both *P. amilis* and *P. oleracea*. In our data, chloroplastic *NADP-MDH-*
517 *1* was significantly upregulated in *P. oleracea* under drought, as previously reported for some *NADP-*
518 *MDH* transcripts in Ferrari et al. (2020b). To our knowledge, there has been no protein labelling of MDH
519 in *Portulaca*; NADP-MDH activity in the mesophyll of *P. oleracea* is theoretically compatible with
520 predominantly NAD-type C_4+CAM, but protein assays are needed to confirm the functional significance
521 of these findings. Of the ten *P. oleracea ME* transcripts analyzed, six were NAD- and four were NADP-
522 type, with *NADP-ME1-1* and *NADP-ME4-4* being the most highly abundant *ME* of all, although multiple
523 *NAD-ME*s were expressed at the same order of magnitude. Western blots by Voznesenskaya et al. (2010)
524 showed faint labelling of NAD-ME in *P. amilis* and strong labelling in *P. oleracea* during the day when
525 well-watered, and strong and no labelling of NADP-ME in *P. amilis* and *P. oleracea*, respectively.
526 Moderate expression levels of *NADP-ME* in *P. oleracea* were also found by Ferrari et al. (2020b) when
527 well-watered, which implies that strong post-transcriptional regulation must occur in *P. oleracea* such
528 that minimal protein is created. The small amount of NAD-ME labeling in *P. amilis* western blots is
529 compatible with low levels of NAD-type CAM, as suggested here by *paPUMP5-1* and mitochondrial
530 malic enzymes *paNAD-ME1-1* and *paNAD-ME2-1*. As an NAD-type C_4 species, *P. oleracea* was
531 expected to constitutively express *NAD-ME* and *PUMP* in the bundle sheath to move malate and aspartate
532 across the mitochondrial membrane (fig. 5A). We observed a potential CAM-specific *NADP-ME* in both
533 species, consistent with experiments by Ferrari et al. (2020) that found one *NAD-ME* transcript
534 significantly upregulated in the morning under drought. When well-watered, *poPUMP1-1* had C_4-like
535 expression, but it was significantly downregulated under drought in an equal magnitude but inverse
536 pattern of the upregulated *poPUMP5-1*. Although *poPUMP5-1* was not significantly differentially

537 abundant under drought when considering the entire time series, its striking change in expression makes it
538 a compelling candidate as a member of the ancestral *Portulaca* CAM cycle that warrants further
539 investigation across the clade more broadly.
540
541 *Portulaca as a model for metabolic evolution*
542 The natural history of *Portulaca* provides a compelling model system for studying many evolutionary and
543 ecophysiological questions, and its unique metabolism and experimental practicality make it a powerful
544 tool for genetic engineering. Forward engineering of both C_4 (Schuler et al. 2016) and CAM (Borland et
545 al. 2014) have become priorities for agriculture as arid lands expand and populations grow. *Portulaca*
546 comprises over 100 species that are found in nearly every country and have invaded some continents
547 multiple times (Ocampo and Columbus 2012). Despite the small genome size of *P. amilis*, duplication
548 events deeper in the history of the Portulacineae (Wang et al. 2019) have played large roles in the
549 evolution of C_4+CAM, with implications for neofunctionalization and the evolution of modularity. With
550 the generation of the first *Portulaca* genome, we provide a resource for studying these past events and for
551 the functional genetic work to understand and engineer CCMs.
552 CAM is found in every major clade of the Portulacineae to varying degrees (Ting and Hanscom
553 1997; Guralnick and Jackson 2001; Guralnick et al. 2007; Edwards and Ogburn 2012; Holtum et al. 2018;
554 Hancock et al. 2019) (fig. 1C), and stronger and constitutive CAM have evolved in *Portulaca*'s closest
555 relatives, Anacampserotaceae and Cactaceae (Nobel 2003; Guralnick et al. 2007; Edwards and Donoghue
556 2006). These three clades overlap in climate space (Edwards and Ogburn 2012) and in many parts of their
557 geographic ranges (Eggli 2004), exhibit succulent morphologies capable of photosynthesis in their stems
558 (Guralnick et al. 2007; Ogburn and Edwards 2009), and likely share the same ancestral CAM cycle. It
559 remains an open question why only *Portulaca* evolved C_4+CAM, while stronger CAM evolved in its
560 relatives. C_4 and CAM consume more energy than C_3 photosynthesis, and we lack models to predict
561 environmental conditions that would select for C_4 in a CAM plant (or vice versa). As mentioned above,
562 photorespiration plays a central role in the transition from C_3 to C_4 photosynthesis through C_2
563 metabolism. Selective pressures to evolve C_2 in the facultative CAM ancestor of *Portulaca* may have
564 been no different than in C_3 plants as CAM does not necessarily reduce photorespiration, and in some
565 cases can exacerbate it (Niewiadomska and Borland 2007). Future progress on these topics will be made
566 through continued study of the C_3-C_4 Cryptopetala clade that uses C_2 metabolism.
567 Despite an increased general interest in CCMs over recent years, and *Portulaca*'s C_4+CAM
568 system in particular, we have only a limited understanding of the benefits of combining C_4 and CAM. The
569 advantages of facultative CAM vary between lineages, as does carbon gain, which is typically <10% of
570 the daytime photosynthetic activity (Herrera 2009; Winter 2019). Facultative CAM increases water use

571 efficiency and can prevent reduced growth and reproductive output under drought (Herrera 2009; Winter
572 et al. 2019) while also reducing photoinhibition (radiation induced damage to photosystems) (Herrera
573 2009). Measures of chlorophyll *a* in *P. oleracea* over a water limitation experiment provide evidence that
574 facultative CAM reduced photoinhibition (Ferrari et al. 2020b). A photoprotective role of CAM is further
575 corroborated by rapid recovery of photosynthetic rate upon rewatering (Holtum et al. 2017; Winter et al.
576 2019), indicating little or no damage was done to photosystems. Gas exchange measurements also
577 demonstrate that *Portulaca* continue to grow despite several days of drought stress (Holtum et al. 2017;
578 Winter et al. 2019). Furthermore, drought stress may increase CO_2 leakiness of bundle sheath in C_4 plants,
579 a problem augmented when carboxylation activity decreases relative to decarboxylation (Ghannnoum,
580 2009). Facultative CAM may provide some carbon and photoprotection during drought that accelerates
581 the transition to full photosynthetic capacity upon rewatering and allows for growth and reproduction
582 despite large water deficits.
583 The annotated *P. amilis* genome and metabolic modules identified here, combined with recent
584 resources for transgenic research in *Portulaca* (Ferrari et al. 2020a) signal a significant shift toward the
585 development of an experimental, functional genomics-based research program in C_4+CAM. The
586 publication of the *Kalanchoë* genome (Yang et al. 2017) and subsequent functional genetic work have
587 confirmed previous hypotheses and made unexpected discoveries, such as the role of PEPC1 in regulation
588 of the core circadian clock (Boxall et al. 2020). Unlike CAM in *Kalanchoë*, C_4+CAM has explicit
589 structure in both time and space that must be mapped to confirm the use of various mesophyll- and bundle
590 sheath-specific enzymes. On one level, we suggest that *P. amilis* is a better candidate than *P. oleracea* for
591 further model system development, in part because *P. oleracea* exists within the context of a larger
592 species complex of over a dozen described species and subspecies with uncertain relationships and
593 multiple whole genome duplication events (Ocampo and Columbus 2012). *P. amilis* has an
594 indistinguishable life history from *P. oleracea* (i.e., 3-5 week life cycle, thousands of seeds produced,
595 self-fertilization), but does not suffer from the same taxonomic challenges and variations in ploidy. On
596 the other hand, we view the *Portulaca* clade as providing not one, but at least three, evolutionary origins
597 of C_4+CAM systems, and there is much to be learned in understanding their similarities and differences.
598
599 **Conclusions**
600 The evolution of modularity is a common theme in studies of evolutionary development and the
601 origination of novel phenotypes. We have demonstrated how network-based analyses can distinguish
602 complex and highly overlapping metabolic modules, and establish their orthology between species. A
603 C_4+CAM system provides a unique opportunity to understand how a single module that is typically
604 recruited into two distinct functions is regulated when both of the new functions are operating in a single

605 organism. Remarkably, their integration appears possible in part via additional modularization, with little
606 overlap in C_4 and CAM gene networks. Assignment of paralogs to functions and generation of a high-
607 quality genome for *P. amilis* were necessary first steps for future functional genetic work on C_4+CAM.
608 Perhaps the most pertinent questions surround the utility of C_4+CAM, which could be addressed in part
609 by knock-down or knock-out *PPC-1E1c* lines to study the fitness of a purely C_4 *Portulaca*. In addition to
610 functional genetic experiments, similar CAM-induction experiments are needed throughout *Portulaca* to
611 establish homology between metabolic pathways across the multiple origins; especially in the C_3-C_4
612 Cryptopetala clade, which will provide greater clarity on which elements of C_4+CAM are ancestrally
613 shared across multiple origins in *Portulaca*.
614
615 **Materials and Methods**
616 *Plant materials and CAM-induction experiment*
617 Specimens of *Portulaca amilis* were collected in Casselberry, Florida, U.S.A. and propagated in the Plant
618 Environmental Center at Brown University. All studied *Portulaca* are selfing, and many are primarily
619 cleistogamous (Geesink 1969, Kim and Carr 1990); cleistogamous flowers self-pollinate in the
620 developing flower bud without the need of a vector, and most buds progress directly to fruit set without
621 opening. S2 seeds were germinated at Yale's Marsh Botanical Garden. One S2 *P. amilis* was removed
622 from soil, gently washed to remove soil, flash frozen whole in liquid N_2, and shipped overnight to
623 Dovetail Genomics (Scotts Valley, CA, U.S.A.) for whole genome sequencing and assembly. An S4
624 generation was collected and vouchered at the Yale University Herbarium in the Peabody Museum of
625 Natural History.
626 S3 *P. amilis* were germinated in a growth chamber with the following environmental conditions:
627 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
628 bottom watered and the mean photosynthetically active radiation (PAR) at plant level was ~385 mol $\cdot$ m^{-2}
629 $\cdot$ s^{-1} during the 14 hour photoperiod. After plants had reached maturity and branched sufficiently (~4
630 weeks), eight plants were selected for CAM-induction: six biological replicates and two indicator plants.
631 Indicator plants were monitored daily for fluctuations in titratable acidity to determine CAM induction,
632 and were not included in any further analyses. To induce CAM, water was withheld from plants, and
633 significant increases in nocturnal titratable acidity were observed in *P. amilis* and *P. oleracea* indicator
634 plants after seven and eight days, respectively. Leaf tissue was collected from all six biological replicates
635 for titratable acidity assays at 16:00 and 04:00 (the following day) and bulk RNA at 10:00, 16:00, 22:00,
636 and 04:00 (the following day) on day 0 (referred to as "well-watered") and days 8 and 9 (referred to
637 "drought").

638 Titratable acidity was assessed by boiling 0.2-0.3g fresh leaf tissue in 60mL 20% EtOH until the
639 volume was reduced by half (~30mL). Distilled water was then added to return the volume to 60mL. This
640 process was repeated, and the supernatant was then covered and allowed to cool to room temperature.
641 Samples were then titrated to a pH of 7.0 using 0.002M NaOH, and recorded in units of μEq H^+ per gram
642 fresh mass—calculated as volume titrant (μL) × titrant molarity (M) / tissue mass (g).
643 Total leaf RNA was extracted using Option 2 (CTAB/TRIzol/sarkosyl) from Jordon-Thaden et al.
644 (2015). Sample purity was measured using a NanoDrop (Invitrogen by Thermo Fisher Scientific,
645 Carlsbad, CA), and a representative set of 12 samples were further assayed on using a 2100 Bioanalyzer
646 (Agilent Technologies, Santa Clara, CA) for fragment length distribution and RNA integrity (RIN); RIN
647 scores ≥ 7 were considered high quality. Library preparation included poly(A) tailing to pull down
648 mRNA, and sequencing was done on an Illumina HiSeq (2x100bp paired-end reads), with an expected 25
649 million reads per sample. Four of the 96 total libraries failed during sequencing; all were *P. oleracea*
650 samples (one 10:00 drought, one 16:00 drought, and two 04:00 drought).
651
652 *Genome and transcriptome assembly and annotation*
653 The *Portulaca amilis* v1 genome was assembled *de novo* by Dovetail Genomics, using one Chicago (see
654 details in Putnam et al. 2016) and two Dovetail Hi-C (see details in Lieberman-Aiden et al. 2009)
655 libraries. The Chicago library was sequenced on an Illumina HiSeqX, resulting in 178 million 2x150bp
656 paired-end reads that yielded 169.1x physical coverage of the genome (1-100kb pairs). The two Dovetail
657 Hi-C libraries were also sequenced on an Illumina HiSeqX; libraries one and two contained 202 million
658 and 342 million 2x150bp paired-end reads, respectively, which together provided 52,626.57x physical
659 coverage of the genome (10-10000kb pairs). Meraculous 2 (Chapman et al. 2011) was used to generate an
660 assembly from 791 million paired-end shotgun read pairs (mean insert size ~550bp and ~350bp) with
661 kmer size 91. The Meraculous assembly and shotgun, Chicago, and Hi-C reads were then used for
662 iterative scaffolding using HiRise v2.1.6-072ca03871cc (Lieberman-Aiden et al. 2009; Putnam et al.
663 2016) to create the final assembly consisting of 4,053 scaffolds.
664 The *P. amilis* v1 genome was annotated using funannotate v1.6.0 (Palmer and Stajich, 2019), a
665 modular software package for genomic scaffold cleaning, genomic element prediction (i.e., genes, UTRs,
666 tRNAs, etc.), functional annotation, and comparative genomics. All annotation steps after genome
667 assembly were run within funannotate unless otherwise noted, and a step-by-step walkthrough of the
668 entire annotation process can be found on GitHub (https://github.com/isgilman/).
669 Using funannotate's 'clean' function, scaffolds were checked for duplicates with a "leave one
670 out" strategy where, starting with the shortest scaffold, one scaffold was left out and mapped back to the
671 remaining scaffolds with minimap2 (Heng, 2018). 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 (Smit and Hubley [online]) and RepeatMasker v4.0.9-p2 (Smit et al. [online]). 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 (Abrusán et al. 2009). 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) (Bao et al. 2015) and transposable element data from Dfam v3.0 (Hubley et al. 2016). 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 RNAseq reads were filtered 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 (Schmieder and Edwards, 2011). To capture the greatest diversity of transcripts and provide sufficient depth to call rare isoforms, we pooled all RNAseq reads from the six biological replicates across four time points and two conditions (well-watered and drought), totaling ~1.58 billion reads from mature leaf tissue. RNAseq reads were then normalized using the 'insilico_read_normalization.pl' utility script within Trinity v2.8.5 (Grabherr et al. 2011). Funannotate's 'train' function first builds a coordinate-sorted BAM file and associated index file using HISAT2 v2.1.0 (Him et al. 2019) for an initial genome-guided transcriptome assembly with Trinity v2.5.1. The resulting initial transcripts then were fed through the PASA pipeline v2.3.3 (Haas et al. 2003) to generate gene models and predicted coding sequences using TransDecoder v5.5.0 (https://transdecoder.github.io). Kallisto v0.45.1 (Bray et al. 2016) was then used to quantify abundance data, and PASA gene models with with abundance at least 10% of the most abundance gene model at each respective locus were retained for training the *ab initio* predictors Augustus v3.2.3 (Stank et al. 2008), SNAP v2013_11_29 (Korf 2004), glimmerHMM v3.0.4 (Majoros et al. 2004), and GeneMark-ET (Besemer and Borodovsky, 2005).

Funannotate's 'predict' function was then used to produce an initial set of gene models from the scaffolds using high abundance PASA models, *ab initio* prediction software, and evidence from transcripts and predicted proteins. First, transcript alignment information from the coordinate-sorted BAM file and protein alignments using Diamond v0.9.24 (Buchfink et al. 2015) and Exonerate v2.4.0 (Slater and Birney, 2005) were used to generate a 'hints' file for Augustus and GeneMark-ET. Augustus models with at least 90% exon overlap with protein alignments were selected as high-quality models (HiQ). Gene models from the above predictors, HiQ models, transcripts, and predicted proteins were then

706 used as sources of evidence to create consensus models in EVidenceModeler v1.1.1 (Haas et al. 2008)
707 with weights Augustus:1, HiQ:1, GeneMark-ET:1, PASA:10, SNAP:1, glimmerHMM:1, proteins:1,
708 transcripts:1.The resulting consensus gene models were filtered to remove any proteins less than 50
709 amino acids, models that began or ended in assembly gaps, and transposable elements. Next, tRNA gene
710 models were identified with tRNAscan-SE v2.0 (Chan and Lowe, 2019). Final consensus gene models
711 were constructed by funannotate's 'update' function which added and updated UTRs and refined exon
712 boundaries of alternatively spliced models.
713 To assemble the *de novo* transcriptome for *P. oleracea*, we again filtered all paired-end raw
714 RNAseq reads by quality (phred < 25) and trimmed to remove adapters, poly-A/T tails longer than 12 bp,
715 and low-quality tails (phred < 20) using PRINSEQ v0.20.4. As with *P. amilis*, we pooled all *P. oleracea*
716 RNAseq reads across time points, conditions, and replicates. Pooled reads were input into Trinity v2.11.0
717 and assembled *de novo*. We reduced the redundancy of the resulting contigs using CD-HIT v4.8.1 (Fu et
718 al. 2012) by collapsing contigs with at least 98% similarity (-c 98).
719
720 *RNAseq differential abundance analysis*
721 We used sleuth (Pimentel et al. 2017) to assess differential abundance of RNAseq reads quantified by
722 Kallisto for each species independently. For *P. amilis*, we mapped reads to the predicted mRNA
723 transcripts from the genome annotation, and for *P. oleracea* we used the redundancy reduced *de novo*
724 transcriptome. All analyses were conducted at the gene level by providing a transcript-to-gene map to
725 sleuth generated from the final set of gene models and setting 'gene_mode' = TRUE. Abundance data
726 were quantified in transcripts per million (TPM) and differential abundance was assessed using a
727 likelihood ratio test between the reduced design matrix, where abundance was purely a function of
728 sampling time ('y ~ time_point'), and the full design matrix that included time and water treatment ('y ~
729 time_point + treatment'). *p*-values for unique transcripts were aggregated for each gene model and
730 transformed into *q*-values by adjusting for false discovery rate (Benjamini Hochberg correction).
731
732 *Co-expression network analysis*
733 We used Weighted Gene Correlation Network Analysis (WGCNA) (Langfelder and Horvath 2008) to
734 construct co-expression networks and identify modules, and NetworkX (Hagberg et al. 2008) to visualize
735 and measure network features. WGCNA was run separately for each species and condition, resulting in
736 four networks. Using the normalized abundance data for all biological replicates from sleuth, we first
737 divided samples into well-watered and drought conditions for separate network construction (hereafter
738 referred to as the well-watered and drought datasets, networks, modules, etc). We then removed genes
739 from each dataset with maximum expression less than 5 TPM. Remaining genes with excessive missing

740 data or too low variance were removed with the WGCNA function 'goodSampleGenes'. Outlier samples
741 were detected using hierarchical clustering with 'hclust' (stats v3.6.2; R-core team), and removed from
742 each dataset. We determined the soft threshold powers for each network separately using
743 'pickSoftThreshold'; soft threshold power was 8 for both well-watered and drought *P. amilis* networks,
744 and 14 and 7 for *P. oleracea* well-watered and drought networks, respectively. We constructed networks
745 and identified modules using 'blockwiseModules' with corType="bicor", maxPOutliers=0.05,
746 TOMType="signed", deepSplit=2, and mergeCutHeight=0.25 for all networks. With corType="bicor",
747 we allowed modules to contain both correlated and anti-correlated expression time profiles. We set
748 minModuleSize=30 for *P. amilis* networks and minModuleSize=100 for *P. oleracea* networks.
749 In order to extract photosynthetic gene networks (PGNs; see Results), we subsetted the resulting
750 adjacency matrices from WGCNA using over 500 (*P. amilis*) and over 600 (*P. oleracea*) manually
751 curated genes with functions broadly related to photosynthesis and removed edges with adjacency < 0.25.
752 This resulted in networks containing 368-466 genes (supplementary table S1). We binned these genes into
753 functional categories reflecting major metabolic pathways (e.g., Calvin-Benson cycle, photorespiration,
754 and starch degradation) and enzymatic reactions or functions (e.g., aminotransferase and PEPC
755 regulation). As many enzymes catalyze reversible reactions or function in multiple pathways, these
756 categorizations are intended to reflect the primary role each protein plays and we recognize that
757 alternative categorizations are possible. For example, RuBisCO is the central protein in both the Calvin-
758 Benson cycle and photorespiration, and MDH functions in both the initial carboxylation pathway that
759 generates malate from CO_2 and in the decarboxylation pathway that liberates CO_2 from malate in NAD-
760 type C_4 biochemistry, albeit to a lesser extent.
761
762 *Motif identification and enrichment*
763 Putative *cis*-regulatory motif identification and enrichment was done using the MEME Suite (Bailey et al.
764 2009). First, 5'- and 3'UTRs, introns, and 2kb upstream regions were extracted from all *P. amilis* gene
765 models. The MEME tool (Bailey and Elkan 1994) was used to identify *de novo* motifs for each co-
766 expression module and region using an e-value threshold of 0.05 and a minimum motif size of 6bp. We
767 downloaded the JASPAR CORE 2018 motif database (Khan et al. 2018), supplemented it with the *de*
768 *novo* motifs identified in Wai et al. (2019) related to CAM, and finally added our *de novo P. amilis* motifs
769 after checking for redundancy with Tomtom (Gupta et al. 2007). We then used AME (--scoring avg --
770 method fisher --hit-lo-fraction 0.25 --evalue-report-threshold 10.0) (McLeay and Bailey, 2010) with
771 random sets of control sequences to measure motif enrichment in every module-region combination, as
772 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 tireless 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.

Figure legends

Figure 1. Simplified C_4 (A) and CAM (B) pathways showing shared carboxylation and decarboxylation pathways. ASP, aspartate aminotransferase; BCA, beta carbonic anhydrase; CBC, Calvin-Benson cycle; MDH, malate dehydrogenase; ME, malic enzyme; PEPC, PEP carboxylase. Cartoon dendrogram of the Portullugo (Portulacineae + Molluginaceae), highlighting major *Portulaca* clades (C), which are colored by primary C_4 biochemical pathway. Mean age estimates for focal *Portulaca* clades are from Ocampo and Columbus (2012). Night-day differences in titratable acidity of experimental *P. amilis* (orange) and *P. oleracea* plants (purple) (D). WW, well-watered; D, drought. Asterisks connecting lines indicate significant differences between well-watered and drought treatments ($p < 0.001$). Images of *P. amilis* when well-watered and after 8 days without water (E).

Figure 2. Idiogram of the nine primary scaffolds of the *P. amilis* genome with key C_4 and CAM genes highlighted (A). Breakdown of repetitive elements of the *P. amilis* genome (B). The horizontal bar chart shows the fractions of major classes of repetitive elements and the repeat landscape (histogram) shows the relative abundances of repetitive elements versus the Kimura substitution level from the consensus. Scatterplot of number of gene models versus the log10-transformed genome size for 107 angiosperm genomes (C) with notable benchmarking and CCM genomes highlighted; data is from Zhao and Schranz (2019). Line in (C) shows results of ordinary least squares regression.

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

807 one gene and node sizes represent their degrees (the number of other genes in the network that they are
808 correlated or anticorrelated to). Each module's size and functional composition is shown in the histogram,
809 and the z-score normalized expression of each module's constituent genes is shown along the horizontal
810 axis. An example of the *z*-score normalized expression for module paWW1 is enlarged in (A) to show
811 individual gene expression with the median highlighted in bold. Transcripts per million (TPM)
812 normalized expression profiles of two focal *PPC* transcripts (*paPPC-1E1a'* and *paPPC-1E1c*) are shown
813 in (C) and (D), respectively.
814
815 Figure 4. *PPC-1E1c* PGN preservation and subnetworks of common genes. Stacked bar charts showing
816 the number and functional composition of genes that were exclusive to the well-watered or drought *PPC-*
817 *1E1c* PGNs, or preserved across treatments in *P. amilis* (A) and *P. oleracea* (B). Subnetworks induced by
818 genes common to both *P. amilis* (C) and *P. oleracea* (D) drought *PPC-1E1c* PGNs; drought PGNs
819 include preserved and drought exclusive genes. Orange edges represent correlations among genes shared
820 between species, while black edges are unique to each respective species. Metabolic pathways of the
821 shared *PPC-1E1c* subnetworks (E), showing PEPC, sugar transport, and starch metabolism. Colors of bar
822 plot units (A and B) and gene boxes (C, D, and E) indicate functional categories as in fig. 3.
823
824 Figure 5. Hypothesized biochemical cycles of C_4+CAM in *P. amilis* when well-watered (A) and
825 droughted (B). Red, blue, and purple lines indicate NAD-specific, NADP-specific, and shared reactions,
826 respectively, and grey lines show novel, possible pathways for malate that are unique to C_4+CAM
827 photosynthesis. Line thicknesses are indicative of relative fluxes through pathways. ALAAT, alanine
828 aminotransferase; ASP, aspartate aminotransferase; BCA, beta carbonic anhydrase; CBC, Calvin-Benson
829 cycle; NAD-MDH, NAD-dependent malate dehydrogenase; NADP-MDH, NADP-dependent malate
830 dehydrogenase; NAD-ME, NAD-dependent malic enzyme; NADP-ME, NADP-dependent malic enzyme;
831 PEP, phosphoenolpyruvate; PEPC, PEP carboxylase; PPDK, pyruvate, phosphate dikinase; OAA,
832 oxaloacetate.
833
834 **Tables**
835 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 1 (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)/Basic helix-loop-helix protein 74 (bHLH74)	AT1G10120

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

References

1. Abrusán G, Grundmann N, DeMester L, Makalowski W. 2009. TEclass—a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* 25:1329–1330.
2. Arakaki M, Christin P-A, Nyffeler R, Lendel A, Eggli U, Ogburn RM, Spriggs E, Moore MJ, Edwards EJ. 2011. Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc Nat Acad Sci USA.* 108:8379–8384.
3. Bailey TL, Elkan C. 1994. 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. AAAI Press, Menlo Park, California. p. 28-36.
4. Bailey TL, Bodén M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37:W202-W208.

849 5. Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in
850 eukaryotic genomes. *Mob DNA.* 6:11.
851 6. Barve A, Wagner A. 2013. A latent capacity for evolutionary innovation through exaptation in
852 metabolic systems. *Nature* 500:203–206.
853 7. Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful
854 Approach to Multiple Testing. *J Royal Statistical Soc B (Meth). 57*: 289-300.
855 8. Besemer J., Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes,
856 eukaryotes and viruses. *Nucleic Acids Res.* 33:W451–4.
857 9. Borland AM, Hartwell J, Weston DJ, Schlauch KA, Tschaplinski TJ, Tuskan GA, Yang X, Cushman
858 JC. 2014. Engineering crassulacean acid metabolism to improve water-use efficiency. *Trends in*
859 *Plant Science* 19:327–338.
860 10. Borland AM, Guo H-B, Yang X, Cushman JC. 2016. Orchestration of carbohydrate processing for
861 crassulacean acid metabolism. *Curr Opinion in Plant Bio.* 31:118–124.
862 11. Boxall SF, Kadu N, Dever LV, Knerova J, Waller JL, Gould PD, Hartwell J. 2020. Kalanchoë PPC1
863 is Essential for Crassulacean Acid Metabolism and the Regulation of Core Circadian Clock and
864 Guard Cell Signaling Genes. *Plant Cell* 32:1136–1160.
865 12. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq
866 quantification. *Nat Biotechnol*. 34:525–527.
867 13. Brilhaus D, Bräutigam A, Mettler-Altmann T, Winter K, Weber APM. 2016. Reversible Burst of
868 Transcriptional Changes during Induction of Crassulacean Acid Metabolism in *Talinum triangulare*.
869 *Plant Phys.* 170:102–122.
870 14. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND.
871 *Nature Methods* 12:59–60.
872 15. Chan PP, Lowe TM. 2019. tRNAscan-SE: Searching for tRNA genes in genomic sequences.
873 *Methods in Mol Bio*. p. 1–14.
874 16. Chang Y-M, Liu W-Y, Shih AC-C, Shen M-N, Lu C-H, Lu M-YJ, Yang H-W, Wang T-Y, Chen
875 SC-C, Chen SM, et al. 2012. Characterizing Regulatory and Functional Differentiation between
876 Maize Mesophyll and Bundle Sheath Cells by Transcriptomic Analysis. *Plant Phys.* 160:165–177.
877 17. Chapman JA, Ho I, Sunkara S, Luo S, Schroth GP, Rokhsar DS. 2011. Meraculous: De Novo
878 Genome Assembly with Short Paired-End Reads. *Plos One* 6:e23501.
879 18. Christin P-A, Arakaki M, Osborne CP, Bräutigam A, Sage RF, Hibberd JM, Kelly S, Covshoff S,
880 Wong GK-S, Hancock L, et al. 2014. Shared origins of a key enzyme during the evolution of C_4 and
881 CAM metabolism. *J Exp Bot.* 65:3609–3621.

882 19. Christin P-A, Sage TL, Edwards EJ, Ogburn RM, Khoshravesh R, Sage RF. 2011. Complex
883 evolutionary transitions and the significance of C_3-C_4 intermediate forms of photosynthesis in
884 Molluginaceae. *Evolution* 65:643–660.
885 20. Christin P-A, Osborne CP, Sage RF, Arakaki M, Edwards EJ. 2011. C_4 eudicots are not younger
886 than C_4 monocots. *J Exp Bot.* 62:3171–3181.
887 21. Clune J, Mouret J-B, Lipson H. 2013. The evolutionary origins of modularity. *Proc Royal Soc B Biol*
888 *Sci.* 280:20122863.
889 22. Davidson EH, Erwin DH. 2006. Gene Regulatory Networks and the Evolution of Animal Body
890 Plans. *Science* 311:796–800.
891 23. Dion-Côté A-M, Renaut S, Normandeau E, Bernatchez L. 2014. RNA-seq Reveals Transcriptomic
892 Shock Involving Transposable Elements Reactivation in Hybrids of Young Lake Whitefish Species.
893 *Mol Biol Evol.* 31:1188–1199.
894 24. Edwards E. J., Diaz M. 2006. Ecological physiology of *Pereskia guamacho*, a cactus with leaves.
895 *Plant Cell Environ.* 29:247–256.
896 25. Edwards EJ, Donoghue MJ. 2006. *Pereskia* and the Origin of the Cactus Life-Form. *Am Nat.*
897 167:777–793.
898 26. Edwards EJ, Ogburn RM. 2012. Angiosperm Responses to a Low-CO_2 World: CAM and C_4
899 Photosynthesis as Parallel Evolutionary Trajectories. *Int J Plant Sci.* 173:724–733.
900 27. Edwards EJ, Osborne CP, Strömberg CAE, Smith SA, Consortium CG, Bond WJ, Christin P-A,
901 Cousins AB, Duvall MR, Fox DL, et al. 2010. The origins of C_4 grasslands: integrating evolutionary
902 and ecosystem science. *Science* 328:587–591.
903 28. Edwards EJ. 2019. Evolutionary trajectories, accessibility and other metaphors: the case of C_4 and
904 CAM photosynthesis. *New Phytologist* 223:1742–1755.
905 29. Eggli U. 2004. Illustrated Handbook of Succulent Plants: Dicotyledons. Berlin: Springer.
906 30. Ehleringer JR, Cerling TE, Helliker BR. 1997. C_4 photosynthesis, atmospheric CO_2, and climate.
907 *Oecologia* 112:285–299.
908 31. Ferrari RC, Bittencourt PP, Nagumo PY, Oliveira WS, Rodrigues MA, Hartwell J, Freschi L. 2020a.
909 Developing *Portulaca oleracea* as a model system for functional genomics analysis of C_4/CAM
910 photosynthesis. *Func Plant Biol*. 48:666-682.
911 32. Ferrari RC, Bittencourt PP, Rodrigues MA, Moreno-Villena JJ, Alves FRR, Gastaldi VD, Boxall SF,
912 Dever LV, Demarco D, Andrade SCS, et al. 2020b. C_4 and Crassulacean Acid Metabolism within a
913 single leaf: deciphering key components behind a rare photosynthetic adaptation. *New Phytologist*
914 225:1699–1714.

915 33. Ferrari RC, Cruz BC, Gastaldi VD, Storl T, Ferrari EC, Boxall SF, Hartwell J, Freschi L. 2020c.
916 Exploring C_4-CAM plasticity within the *Portulaca oleracea* complex. *Scientific Reports* 10:14237–
917 14.
918 34. Friso G, Majeran W, Huang M, Sun Q, van Wijk KJ. 2010. Reconstruction of Metabolic Pathways,
919 Protein Expression, and Homeostasis Machineries across Maize Bundle Sheath and Mesophyll
920 Chloroplasts: Large-Scale Quantitative Proteomics Using the First Maize Genome Assembly. *Plant*
921 *Phys.* 152:1219–1250.
922 35. Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation
923 sequencing data. *Bioinformatics* 28:3150–3152.
924 36. Geesink R. 1969. An account of the genus *Portulaca* in Indo-Australia and the Pacific
925 (Portulacaceae). *Blumea* 17.
926 37. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
927 Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data
928 without a reference genome. *Nat Biotechnol*. 29:644–652.
929 38. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. 2007. Quantifying similarity between
930 motifs. *Genome Biol.* 8:R24.
931 39. Guralnick LJ, Jackson MD. 2001. The occurrence and phylogenetics of Crassulacean acid
932 metabolism in the Portulacaceae. *Int J Plant Sci.* 162:257–262.
933 40. Guralnick LJ, Edwards G, Ku MSB, Hockema B, Franceschi VR. 2002. Photosynthetic and
934 anatomical characteristics in the C_4-Crassulacean Acid Metabolism-cycling plant, *Portulaca*
935 grandiflora. *Func Plant Biol*. 29:763–773.
936 41. Guralnick LJ, Cline A, Smith M, Sage RF. 2007. Evolutionary physiology: the extent of C_4 and
937 CAM photosynthesis in the genera *Anacampseros* and *Grahamia* of the Portulacaceae. *J Exp Bot.*
938 59:1735–1742.
939 42. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Hannick LI, Maiti R, Ronning CM,
940 Rusch DB, Town CD, et al. 2003. Improving the *Arabidopsis* genome annotation using maximal
941 transcript alignment assemblies. *Nucleic Acids Res*. 31:5654–5666.
942 43. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008.
943 Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to
944 Assemble Spliced Alignments. *Genome Biol.* 9:R7.
945 44. Hagberg AA, Schult DA, Swart PJ. 2008. Exploring network structure, dynamics, and function using
946 NetworkX. In: Varoquaux G, Vaught T, Millman J editors. Proceedings of the 7th Python in Science
947 Conference (SciPy2008). Pasadena, CA, USA

948 45. Hancock LP, Holtum JAM, Edwards EJ. 2019. The Evolution of CAM Photosynthesis in Australian
949 *Calandrinia* Reveals Lability in C_3+CAM Phenotypes and a Possible Constraint to the Evolution of
950 Strong CAM. *Integr Comp Biol.* 59:517–534.
951 46. Hartwell J. 2006. The Circadian Clock in CAM Plants. In: Hall AJW, McWatters H editors. Annual
952 Plant Reviews Volume 21: Endogenous Plant Rhythms. Kundli, Insi. Blackwell Publishing Ltd. p.
953 211–236.
954 47. Hibberd JM, Covshoff S. 2010. The Regulation of Gene Expression Required for C_4 Photosynthesis.
955 *Ann Rev Plant Biol.* 61:181–207.
956 48. Heng Li. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–
957 3100.
958 49. Herrera A. 2009. Crassulacean Acid Metabolism and fitness under water deficit stress: if not for
959 carbon gain, what is facultative CAM good for? *Ann Bot.* 103:645–653.
960 50. Heyduk K, Moreno-Villena JJ, Gilman IS, Christin P-A, Edwards EJ. 2019. The genetics of
961 convergent evolution: insights from plant photosynthesis. *Nat Rev Gen.* 313:1–9.
962 51. Ho C-L, Chiang J-M, Lin T-C, Martin CE. 2019. First report of C_4/CAM-cycling photosynthetic
963 pathway in a succulent grass, *Spinifex littoreus* (Brum. f.) Merr., in coastal regions of Taiwan. *Flora*
964 254:194-202.
965 52. Holtum JAM, Hancock LP, Edwards EJ, Winter K. 2017. Optional use of CAM photosynthesis in
966 two C_4 species, *Portulaca cyclophylla* and *Portulaca digyna*. *J Plant Phys.* 214:91–96.
967 53. Holtum JAM, Hancock LP, Edwards EJ, Winter K. 2018. Crassulacean Acid Metabolism in the
968 Basellaceae (Caryophyllales). *Plant Biol.* 20:409–414.
969 54. Holtum JAM, Hancock LP, Edwards EJ, Winter K. 2021. CAM photosynthesis in desert blooming
970 *Cistanthe* of the Atacama, Chile. *Func Plant Biol*. 48:691-702.
971 55. Huang WM, Shao H, Zhou SN, Zhou Q, Fu WL, Zhang T, Jiang H-S, Li W, Gontero B, Maberly
972 SC. 2018. Different CO_2 acclimation strategies in juvenile and mature leaves of Ottelia alismoides.
973 *Photosynth Res.* 138:219–232.
974 56. Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, Smit AFA, Wheeler TJ. 2016. The
975 Dfam database of repetitive DNA families. *Nucleic Acids Res*. 44:D81–9.
976 57. Jordon-Thaden IE, Chanderbali AS, Gitzendanner MA, Soltis DE. 2015. Modified CTAB and
977 TRIzol Protocols Improve RNA Extraction from Chemically Complex Embryophyta. *Apps Plant*
978 *Sci.* 3:1400105–1400122.
979 58. Kashtan N, Alon U. 2005. Spontaneous evolution of modularity and network motifs. *Proc Nat Acad*
980 *Sci USA.* 102:13773–13778.

981 59. Kashtan N, Noor E, Alon U. 2007. Varying environments can speed up evolution. *Proc Nat Acad Sci*
982 *USA*. 104:13711–13716.
983 60. Keeley JE, Rundel PW. 2003. Evolution of CAM and C_4 Carbon-Concentrating Mechanisms. *Int J*
984 *Plant Sci.* 164:S55–S77.
985 61. Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, Bessy A,
986 Chèneby J, Kulkarni SR, Tan G, et al. 2018. JASPAR 2018: update of the open-access
987 database of transcription factor binding profiles and its web framework. *Nucleic Acids Res*.
988 46:D260–D266.
989 62. Kim I, Carr GD. 1990. Reproductive biology and uniform culture of *Portulaca* in Hawaii. *Pac Sci*.
990 44:123–129.
991 63. Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59.
992 64. Langfelder P, Horvath S. 2008. WGCNA: An R package for weighted correlation network analysis.
993 *BMC Bioinformatics* 9:559.
994 65. Lara MV, Disante KB, Podestá FE, Andreo CS, Drincovich MF. 2003. Induction of a Crassulacean
995 Acid like Metabolism in the C_4 succulent plant, *Portulaca oleracea* L.: physiological and
996 morphological changes are accompanied by specific modifications in phosphoenolpyruvate
997 carboxylase. *Photosynth Res* 77:241.
998 66. Lara MV, Drincovich MF, Andreo CS. 2004. Induction of a Crassulacean Acid-like Metabolism in
999 the C_4 Succulent Plant, *Portulaca oleracea* L.: Study of Enzymes Involved in Carbon Fixation and
1000 Carbohydrate Metabolism. *Plant Cell Phys.* 45: 618–626.
1001 67. Leegood RC. 1985. The intercellular compartmentation of metabolites in leaves of *Zea mays* L.
1002 *Planta* 164:163–171.
1003 68. Lenski RE, Ofria C, Pennock RT, Adami C. 2003. The evolutionary origin of complex features.
1004 *Nature* 423:139–144.
1005 69. Liu Y, Li X, Li K, Liu H, Lin C. 2013. Multiple bHLH Proteins form Heterodimers to Mediate
1006 CRY2-Dependent Regulation of Flowering-Time in *Arabidopsis*. *Plos Genet.* 9:e1003861.
1007 70. Lüttge, U. 2011. Photorespiration in Phase III of Crassulacean Acid Metabolism: Evolutionary and
1008 Ecophysiological Implications. In: Lüttge U, Beyschlag W, Büdel B, Francis D editors. Progress in
1009 Botany 72. Springer-Verlag Berlin Heidelberg. p. 371-384.
1010 71. Maguvu TE, Higuchi Y, Sugiura S, Ito H, Shibata M. 2021. De novo Whole Transcriptome Analysis
1011 of Wingpod Purslane, *Portulaca umbraticola*, and Its Application to the Discovery of Genes Related
1012 to Flower Senescence and Opening Rhythm. *Hortic J.* 70:97-107.
1013 72. Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source ab initio
1014 eukaryotic gene-finders. *Bioinformatics* 20:2878–2879.

1015 73. Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U. 2014.
1016 The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*. *eLife*
1017 3:e02478.
1018 74. Matasci N, Hung L-H, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N, Warnow T,
1019 Ayyampalayam S, Barker M, et al. 2014. Data access for the 1,000 Plants (1KP) project.
1020 *GigaScience* 3:17.
1021 75. McLeay RC, Bailey TL. 2010. Motif Enrichment Analysis: a unified framework and an evaluation
1022 on ChIP data. *BMC Bioinformatics* 11:165.
1023 76. McClintock B. 1984. The significance of responses of the genome to challenge. *Science* 226:792-
1024 801.
1025 77. McClung C. 2019. The Plant Circadian Oscillator. *Biology* 8:14–17.
1026 78. Moreno-Villena JJ, Dunning LT, Osborne CP, Christin P-A. 2018. Highly Expressed Genes Are
1027 Preferentially Co-Opted for C_4 Photosynthesis. *Mol Biol Evo.* 35:94–106.
1028 79. Niewiadomska E, Borland A. 2007. Crassulacean Acid Metabolism: a cause or consequence of
1029 oxidative stress in planta? In: Luttge U, Beyschlag W, Murata J editors. Progress in Botany 69.
1030 Springer-Verlag Berlin Heidelberg. p. 247-266.
1031 80. Nimmo HG. 2000. The regulation of phosphoenolpyruvate carboxylase in CAM plants. *Trends Plant*
1032 *Sci* 5:75–80.
1033 81. Nobel PS. 2003. Environmental Biology of Agaves and Cacti. Cambridge University Press
1034 82. Ocampo G, Koteyeva NK, Voznesenskaya EV, Edwards GE, Sage TL, Sage RF, Columbus JT.
1035 2013. Evolution of leaf anatomy and photosynthetic pathways in Portulacaceae. *Am J Bot*.
1036 100:2388–2402.
1037 83. Ogburn RM, Edwards EJ. 2009. Anatomical variation in Cactaceae and relatives: Trait lability and
1038 evolutionary innovation. *Am J Bot.* 96:391–408.
1039 84. Ogren WL. 1984. Photorespiration: Pathways, Regulation, and Modification. *Ann Rev Plant Phys.*
1040 35:415–442.
1041 85. Palmer J, Stajich J. [online]. nextgenusfs/funannotate: funannotate v1.6.0 (Version 1.6.0). Zenodo.
1042 http://doi.org/10.5281/zenodo.3354704.
1043 86. Peterhansel C, Horst I, Niessen M, Blume C, Kebeish R, Kürkcüoglu S, Kreuzaler F. 2010.
1044 Photorespiration. *The Arabidopsis Book* 8:e0130.
1045 87. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. 2017. Differential analysis of RNA-seq
1046 incorporating quantification uncertainty. *Nature Methods* 14:687–690.

1047 88. Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley
1048 PD, Sugnet CW, et al. 2016. Chromosome-scale shotgun assembly using an in vitro method for
1049 long-range linkage. *Genome Res.* 26:342–350.
1050 89. Raven JA, Spicer RA. 1996. The evolution of Crassulacean Acid Metabolism. In: Winter K, Smith
1051 JAC, editors. Ecological Studies 114: Crassulacean Acid Metabolism: Biochemistry, Ecophysiology,
1052 and Evolution. Springer-Verlag Berlin Heidelberg. p. 360–385.
1053 90. Reyes JC, Muro-Pastor MI, Florencio FJ. 2004. The GATA Family of Transcription Factors in
1054 Arabidopsis and Rice. *Plant Phys.* 134:1718–1732.
1055 91. Sage RF. 2002. Are Crassulacean Acid Metabolism and C_4 photosynthesis incompatible? *Func Plant*
1056 *Biol.* 29:775–785.
1057 92. Sage RF. 2004. The evolution of C_4 photosynthesis. *New Phytologist*. 161: 341-370.
1058 93. Sage RF. 2016. A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery:
1059 species number, evolutionary lineages, and Hall of Fame. *J Exp Bot.* 67:4039–4056.
1060 94. Sage RF, Sage TL, Kocacinar F. 2012. Photorespiration and the Evolution of C_4 Photosynthesis.
1061 *Plant Biol.* 63:19–47.
1062 95. Schuler ML, Mantegazza O, Weber APM. 2016. Engineering C_4 photosynthesis into C_3 chassis in
1063 the synthetic biology age. *Plant J.* 87:51–65
1064 96. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets.
1065 *Bioinformatics* 27:863–864.
1066 97. Sharma AK, Bhattacharyya NK. 1956. Cytogenetics of some members of Portulacaceae and related
1067 families. *Caryologia* 8:257–274.
1068 98. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing
1069 genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–
1070 3212.
1071 99. Slack CR, Hatch MD. 1967. Comparative studies on the activity of carboxylases and other enzymes
1072 in relation to the new pathway of photosynthetic carbon dioxide fixation in tropical grasses. *Biochem*
1073 *J.* 103(3):660-665.
1074 100. Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison.
1075 *BMC Bioinformatics* 6:31.
1076 101. Smit AFA, Hubley R. [online]. RepeatModeler Open-1.0. http://www.repeatmasker.org [2008-
1077 2015].
1078 102. Smit AFA, Hubley R, Green P. [online]. RepeatMasker Open-4.0. http://www.repeatmasker.org
1079 [2013-2015].

1080 103. Szarek SR, Ting IP. Photosynthetic Efficiency of CAM Plants in Relation to C_3 and C_4 Plants. In:
1081 Environmental and Biological Control of Photosynthesis. Marcelle, R, editor. Springer Netherlands.
1082 p. 289-297.
1083 104. Taylor SH, Ripley BS, Martin T, De-Wet L, Woodward FI, Osborne CP. 2014. Physiological
1084 advantages of C_4 grasses in the field: a comparative experiment demonstrating the importance of
1085 drought. *Global Change Biol.* 20:1992–2003.
1086 105. Ting IP, Hanscom Z. 1977. Induction of Acid Metabolism in *Portulacaria afra*. *Plant Phys.* 59:511–
1087 514.
1088 106. Voznesenskaya EV, Koteyeva NK, Edwards GE, Ocampo G. 2010. Revealing diversity in structural
1089 and biochemical forms of C_4 photosynthesis and a C_3-C_4 intermediate in genus Portulaca L.
1090 (Portulacaceae). *J Exp Bot.* 61:3647–3662.
1091 107. Voznesenskaya, EV, Koteyeva NK, Edwards GE, Ocampo G. Unique photosynthetic phenotypes in
1092 *Portulaca* (Portulacaceae): C_3-C_4 intermediates and NAD-ME C_4 species with Pilosoid-type Kranz
1093 anatomy. *J Exp Bot.* 68, 225–239 (2017).
1094 108. Wagner GP, Altenberg L. 1996. Perspective: Complex Adaptations And The Evolution Of
1095 Evolvability. *Evolution* 50:967–976.
1096 109. Wang N, Yang Y, Moore MJ, Brockington SF, Walker JF, Brown JW, Liang B, Feng T, Edwards C,
1097 Mikenas J, et al. 2019. Evolution of Portulacineae Marked by Gene Tree Conflict and Gene Family
1098 Expansion Associated with Adaptation to Harsh Environments. *Mol Biol Evo.* 36:112–126.
1099 110. Williams BP, Johnston IG, Covshoff S, Hibberd JM. 2013. Phenotypic landscape inference reveals
1100 multiple evolutionary paths to C_4 photosynthesis. *eLife* 2:e00961.
1101 111. Winter K, Aranda J, Holtum JAM. 2005. Carbon Isotope Composition and Water-use Efficiency in
1102 Plants with Crassulacean Acid Metabolism. *Func Plant Biol.* 32:381–388.
1103 112. Winter K, Sage RF, Edwards EJ, Virgo A, Holtum JAM. 2019. Facultative Crassulacean Acid
1104 Metabolism in a C_3–C_4 intermediate. *J Exp Bot.* 108:8379–9.
1105 113. Winter K. 2019. Ecophysiology of constitutive and facultative CAM photosynthesis. *J Exp Bot.*
1106 2:16178–14.
1107 114. Winter K, Garcia M, Virgo A, Ceballos J, Holtum JAM. 2021. Does the C_4 plant *Trianthema*
1108 *portulacastrum* (Aizoaceae) exhibit weakly expressed crassulacean acid metabolism (CAM)? *Func*
1109 *Plant Biol*. 48:655-665.
1110 115. Yang X, Hu R, Yin H, Jenkins J, Shu S, Tang H, Liu D, Weighill DA, Yim WC, Ha J, et al. 2017.
1111 The *Kalanchoë* genome provides insights into convergent evolution and building blocks of
1112 crassulacean acid metabolism. *Nat. Comm.* 8:1–15.

116. Zhang Y, Yin L, Jiang H-S, Li W, Gontero B, Maberly SC. 2013. Biochemical and biophysical CO_2 concentrating mechanisms in two species of freshwater macrophyte within the genus *Ottelia* (Hydrocharitaceae). *Photosynth Res.* 121:285–297.

117. Zhao T, Schranz ME. 2019. Network-based microsynteny analysis identifies major differences and genomic outliers in mammalian and angiosperm genomes. *Proc Nat Acad Sci USA.* 116:2165–2174.

Figure 1.

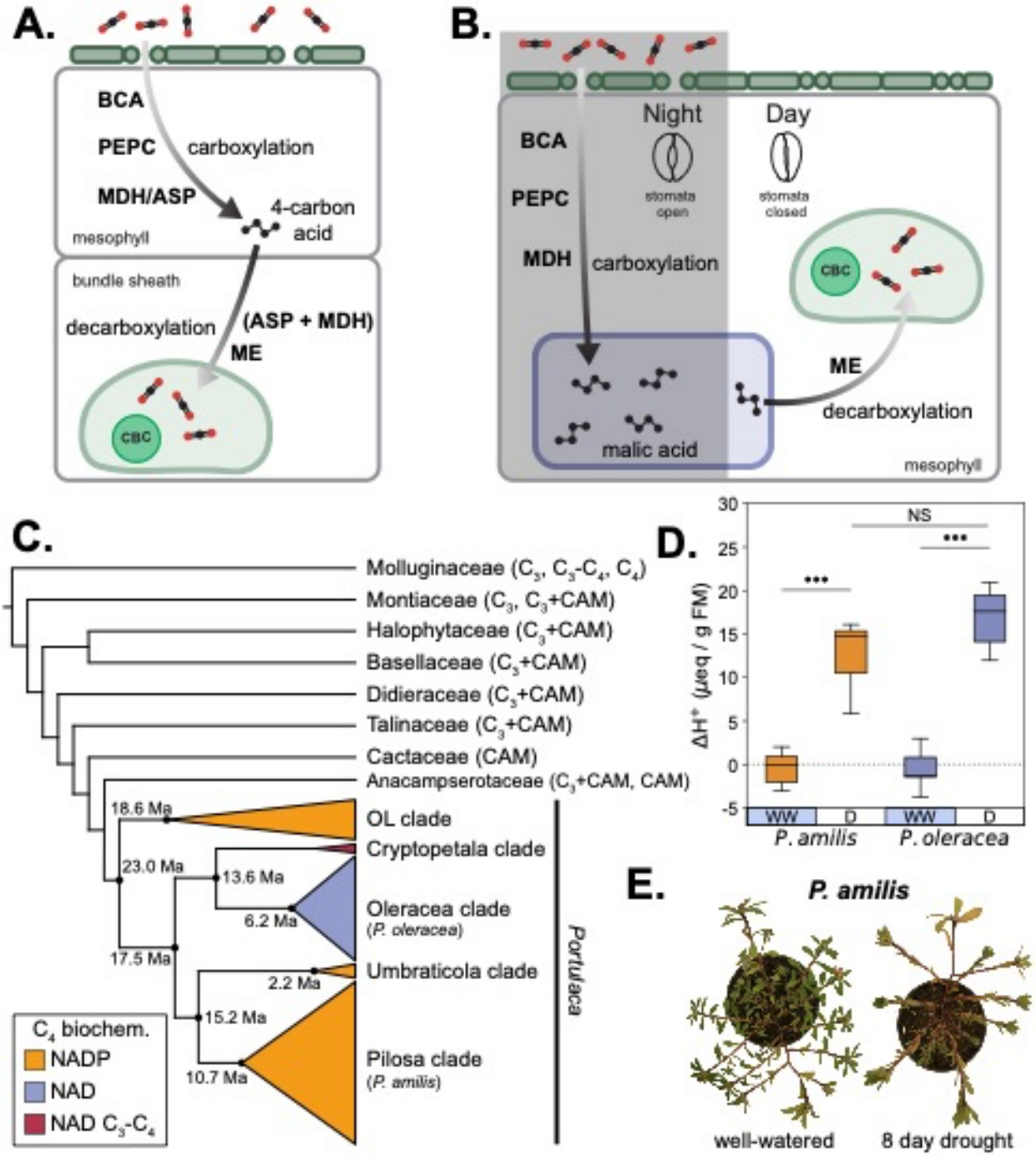


Figure 2

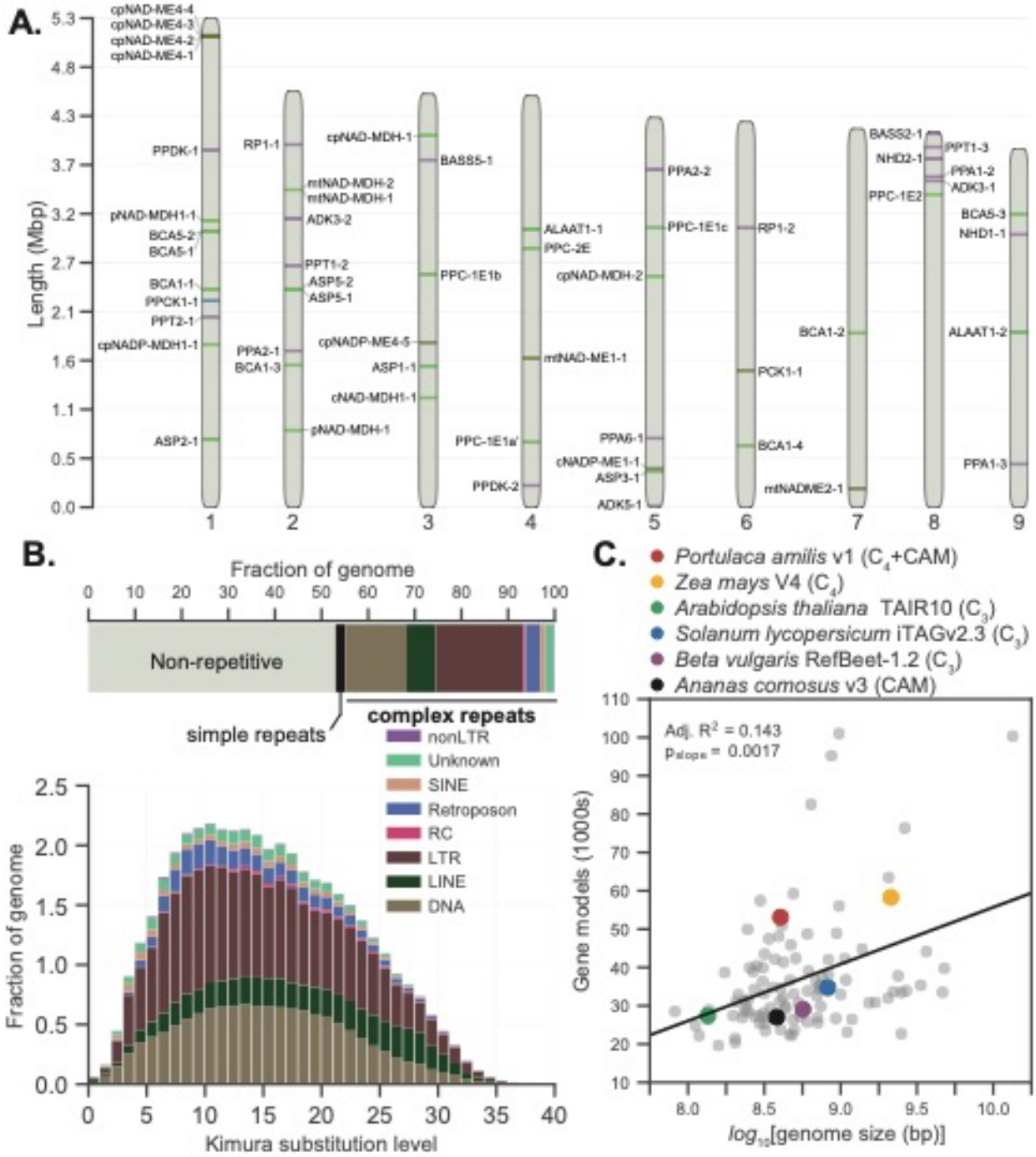

A.
Length (Mbp)
cpNAD-ME4-4
cpNAD-ME4-3
cpNAD-ME4-2
cpNAD-ME4-1
PPDK-1
pNAD-MDH1-1
BCA5-2
BCA5-1
BCA1-1
PPCK1-1
PPT2-1
cpNADP-MDH1-1
ASP2-1
RP1-1
mtNAD-MDH-2
mtNAD-MDH-1
ADK3-2
PPT1-2
ASP5-2
ASP5-1
PPA2-1
BCA1-3
pNAD-MDH-1
cpNAD-MDH-1
BASS5-1
cpNADP-ME4-5
ASP1-1
cNAD-MDH1-1
PPC-1E1b
ALAAT1-1
PPC-2E
cpNAD-MDH-2
mtNAD-ME1-1
PPC-1E1a'
PPDK-2
PPA2-2
PPC-1E1c
PPA6-1
cNADP-ME1-1
ASP3-1
ADK5-1
RP1-2
PCK1-1
BCA1-4
mtNADME2-1
BCA1-2
BASS2-1
NHD2-1
PPC-1E2
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)]

Figure 3.

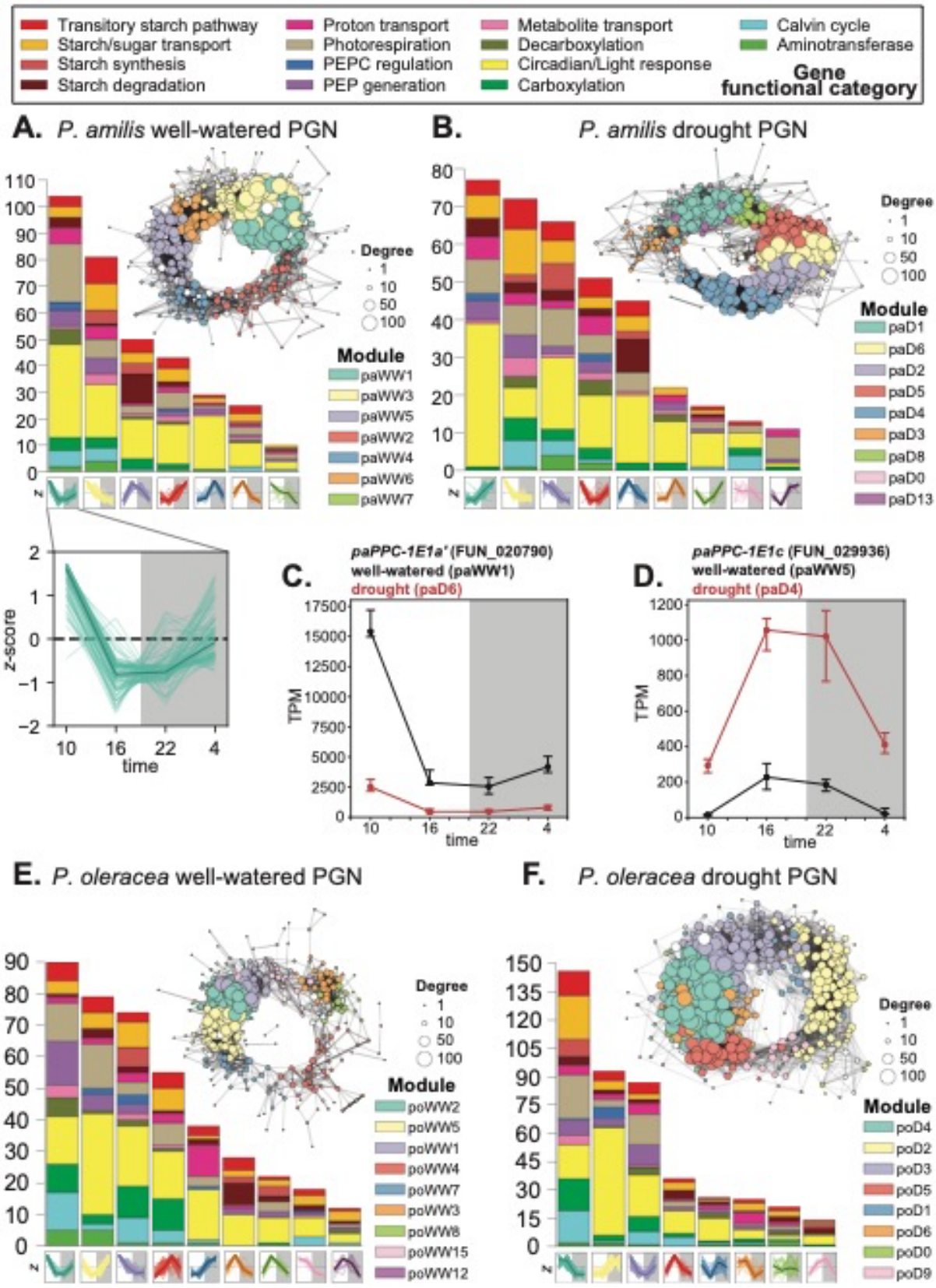
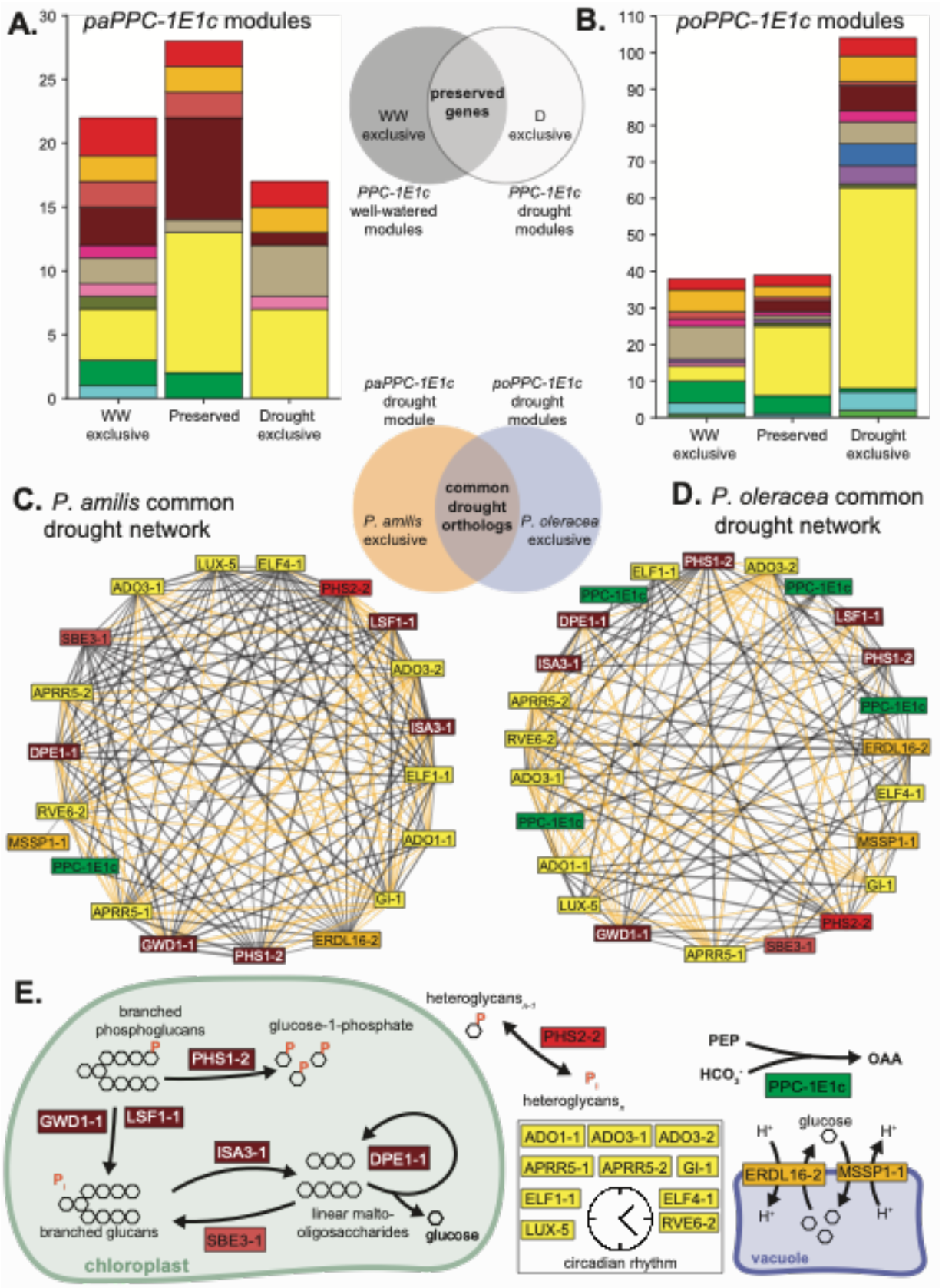
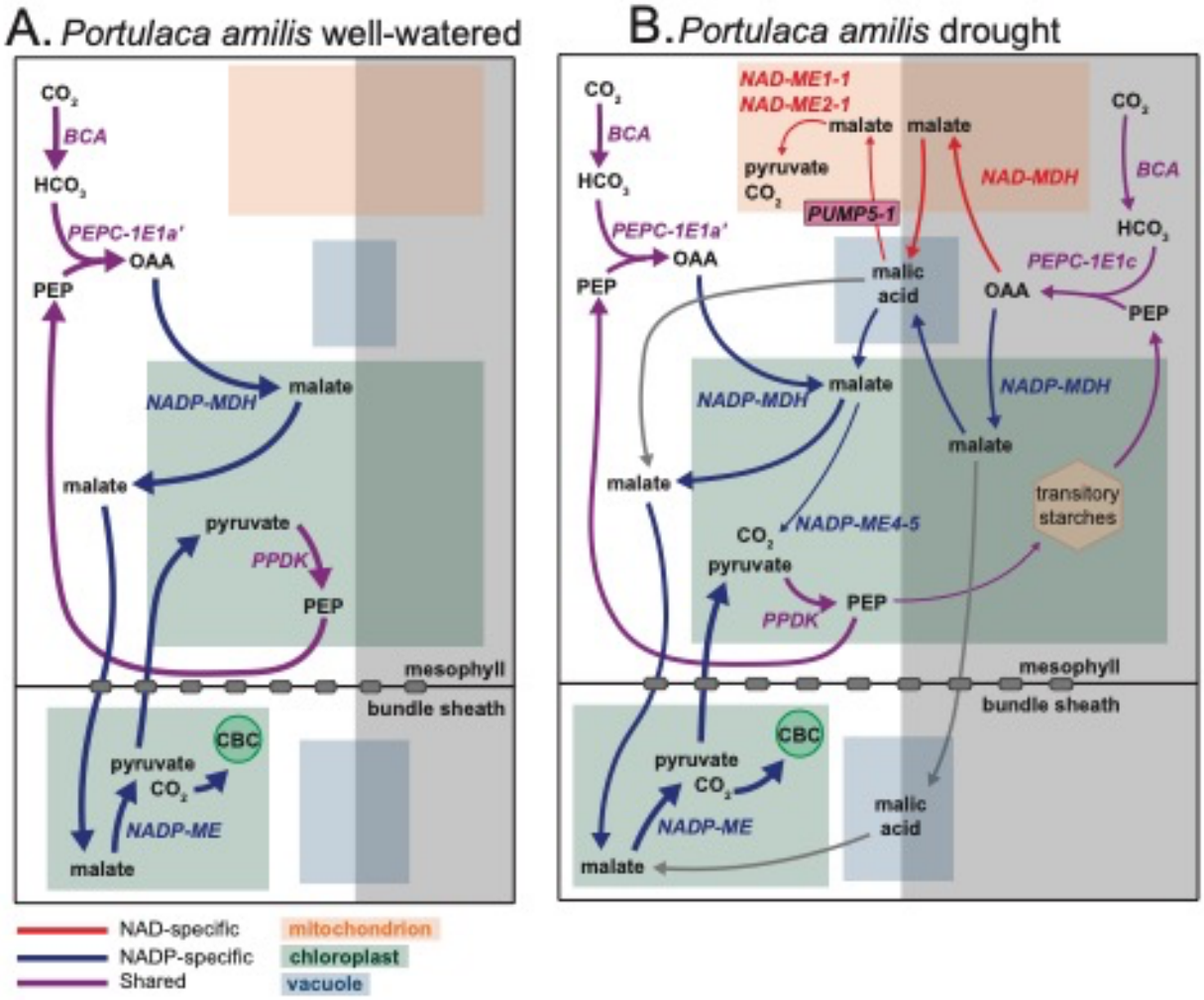

Transitory starch pathway
Starch/sugar transport
Starch synthesis
Starch degradation
Proton transport
Photorespiration
PEPC regulation
PEP generation
Metabolite transport
Decarboxylation
Circadian/Light response
Carboxylation
Calvin cycle
Aminotransferase
Gene functional category
A. P. amilis well-watered 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-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
F. P. oleracea drought PGN
poWW2
poWW5
poWW1
poWW4
poWW7
poWW3
poWW8
poWW15
poWW12
poD4
poD2
poD3
poD5
poD1
poD6
poD0
poD9

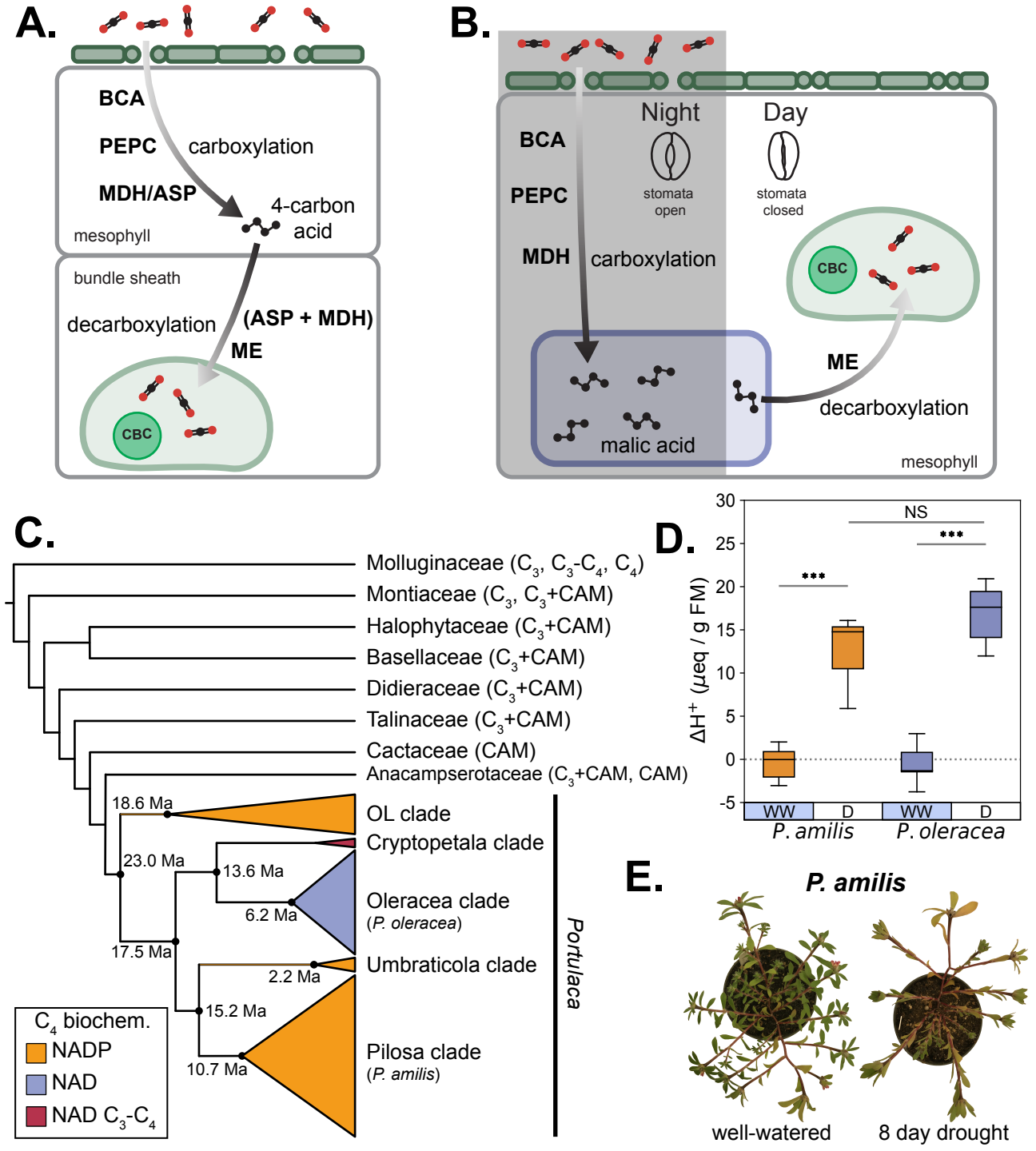
Figure 4.

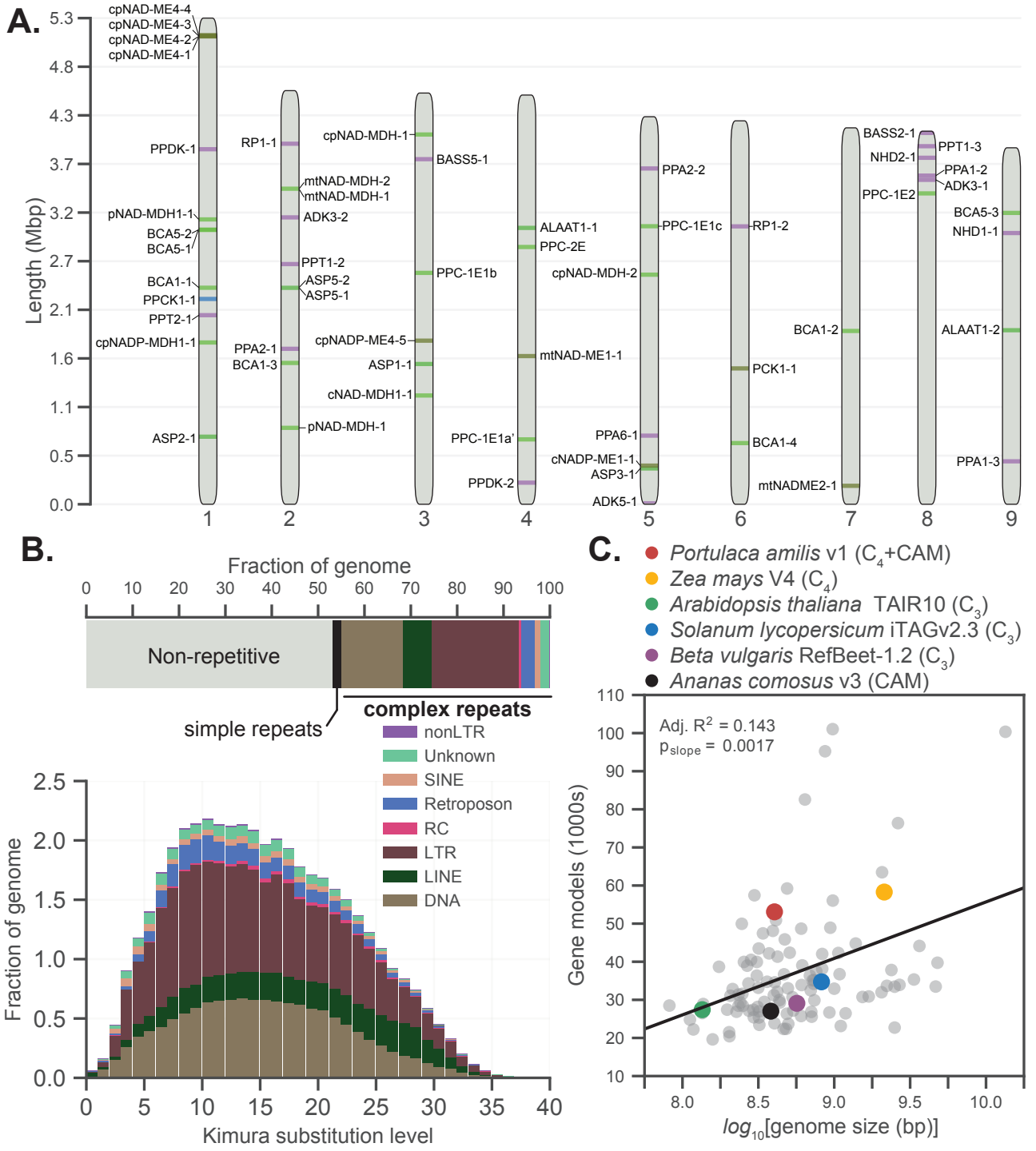

A.
paPPC-1E1c modules
WW exclusive
Preserved
Drought exclusive
preserved genes
WW exclusive
D exclusive
PPC-1E1c well-watered modules
PPC-1E1c drought modules
B.
poPPC-1E1c modules
WW exclusive
Preserved
Drought exclusive
paPPC-1E1c drought module
poPPC-1E1c drought modules
P. amilis exclusive
common drought orthologs
P. oleracea exclusive
C. P. amilis common drought network
D. P. oleracea common drought network
E.
branched phosphoglucans
glucose-1-phosphate
PHS1-2
GWD1-1
LSF1-1
ISA3-1
DPE1-1
SBE3-1
branched glucans
linear malto-oligosaccharides
glucose
chloroplast
heteroglycans
PHS2-2
PEP
HCO3-
OAA
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
ERDL16-2
MSSP1-1
vacuole

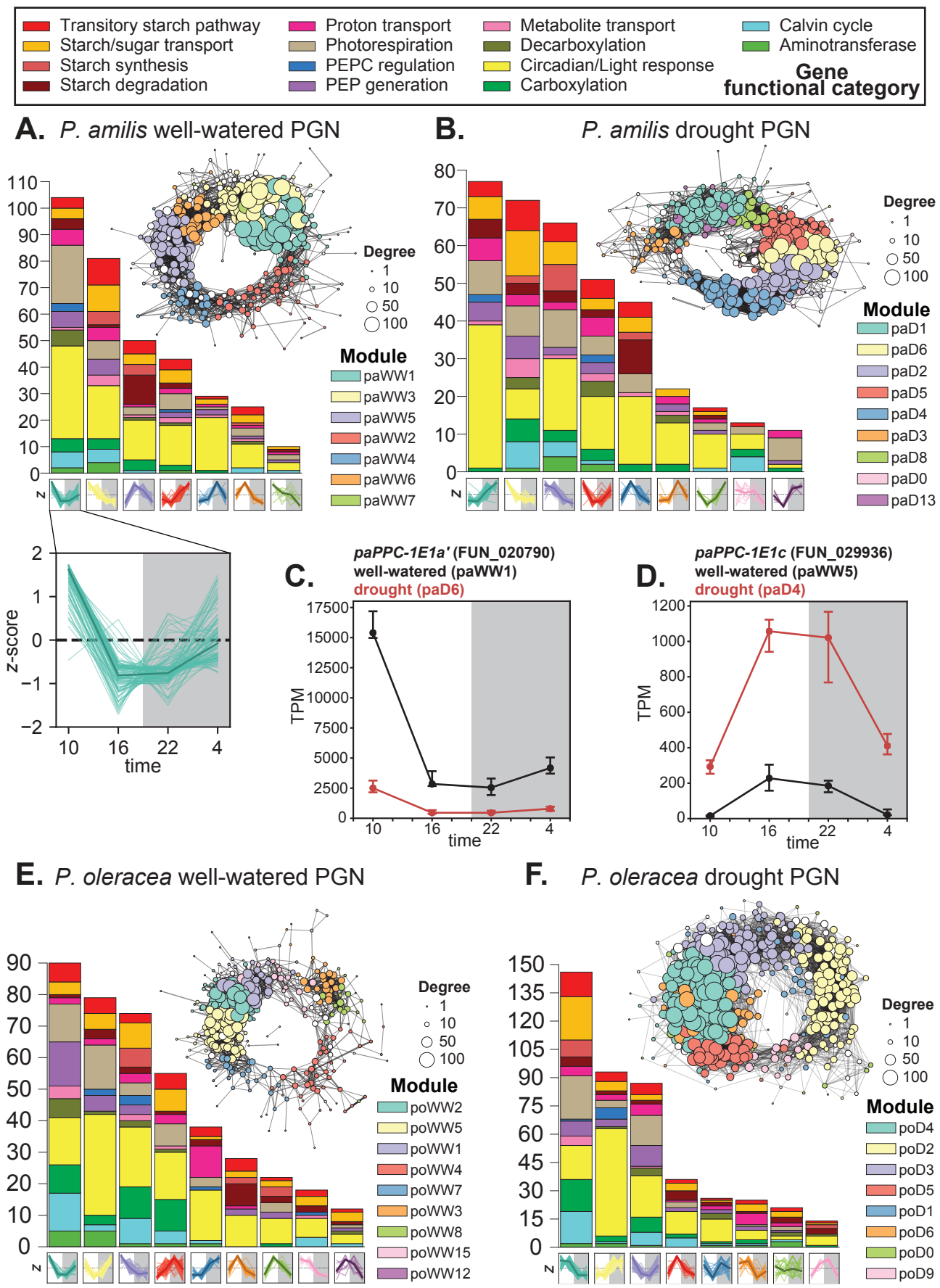
1126 Figure 5.

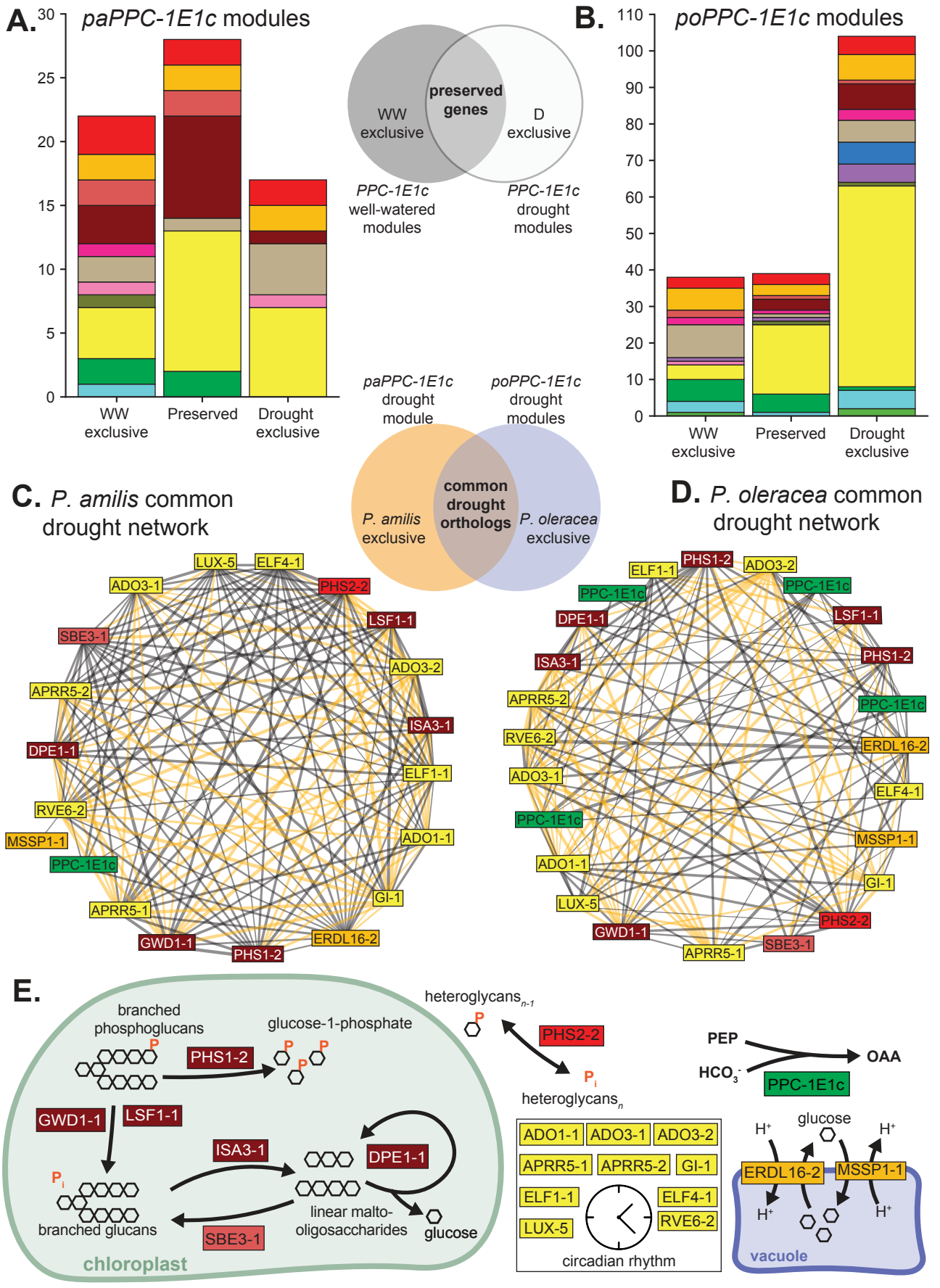


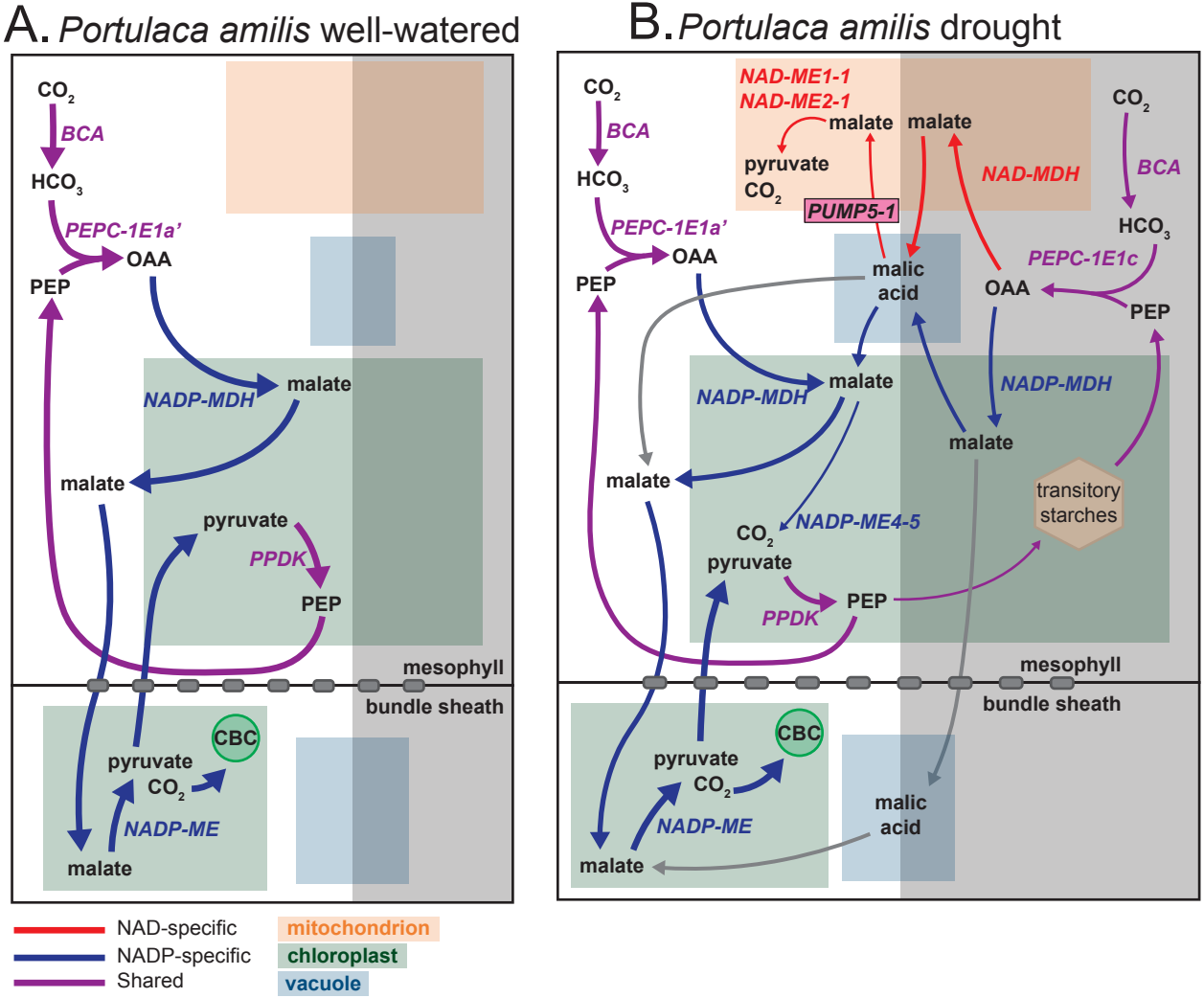
1127

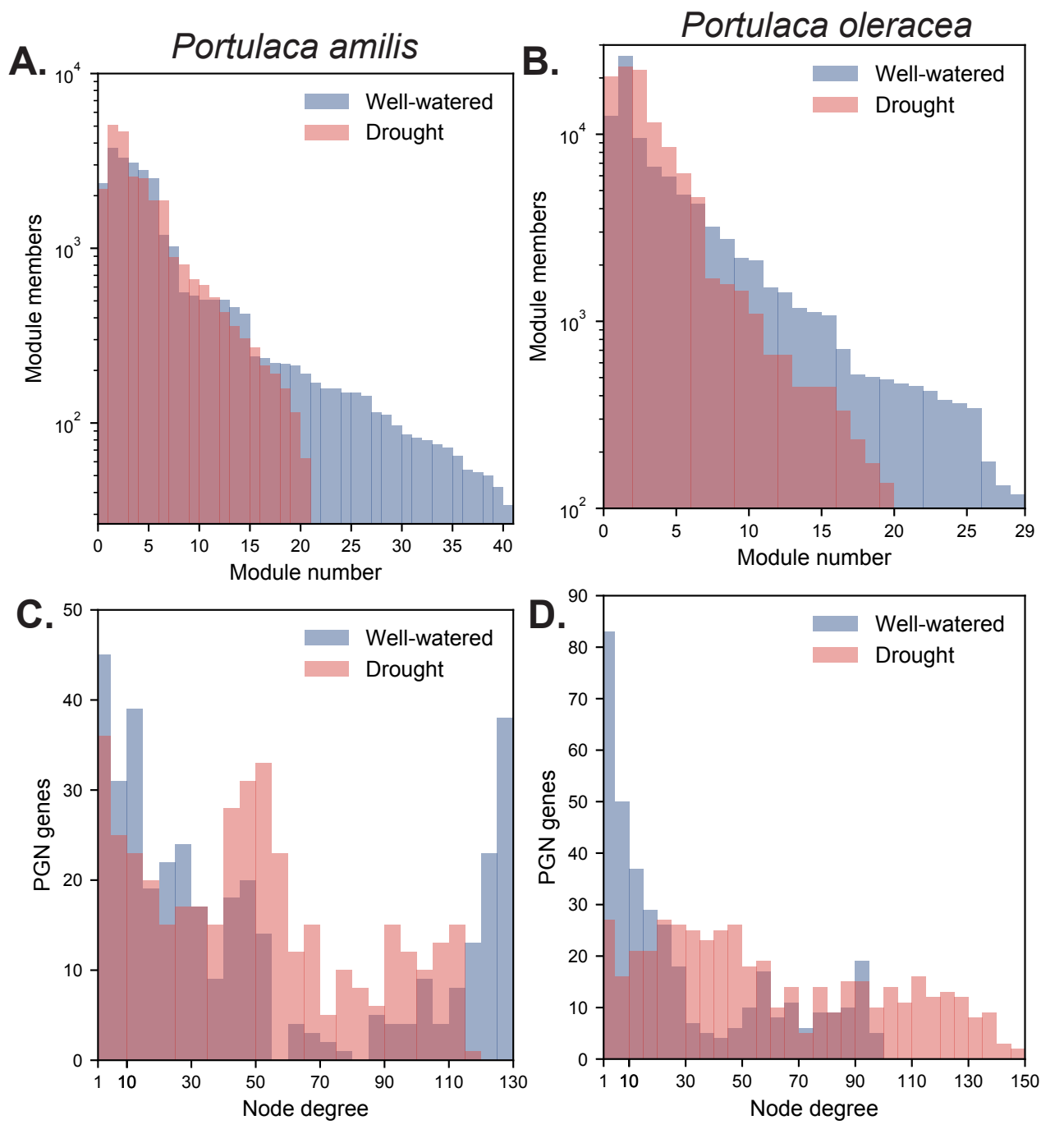

A.
BCA
PEPC
MDH/ASP
carboxylation
4-carbon acid
mesophyll
bundle sheath
decarboxylation
(ASP + MDH)
ME
CBC
B.
Night
Day
BCA
PEPC
MDH
carboxylation
stomata open
stomata closed
CBC
ME
decarboxylation
malic acid
mesophyll
C.
Molluginaceae (C3, C3-C4, C4)
Montiaceae (C3, C3+CAM)
Halophytaceae (C3+CAM)
Basellaceae (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

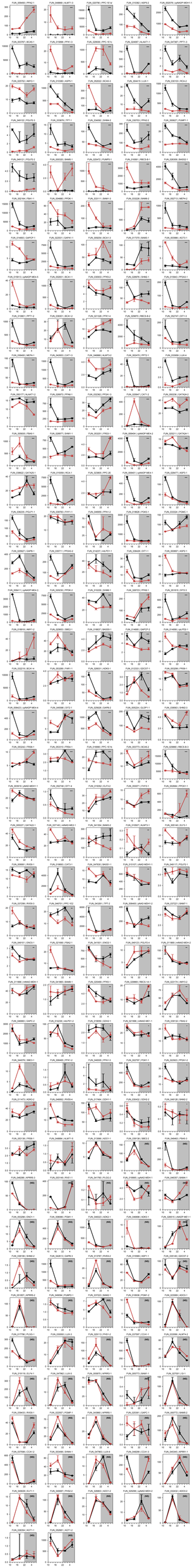

A.
Length (Mbp)
cpNAD-ME4-4
cpNAD-ME4-3
cpNAD-ME4-2
cpNAD-ME4-1
PPDK-1
pNAD-MDH1-1
BCA5-2
BCA5-1
BCA1-1
PPCK1-1
PPT2-1
cpNADP-MDH1-1
ASP2-1
RP1-1
mtNAD-MDH-2
mtNAD-MDH-1
ADK3-2
PPT1-2
ASP5-2
ASP5-1
PPA2-1
BCA1-3
pNAD-MDH-1
cpNAD-MDH-1
BASS5-1
PPC-1E1b
cpNADP-ME4-5
ASP1-1
cNAD-MDH1-1
PPC-1E1a'
PPDK-2
ALAAT1-1
PPC-2E
cpNAD-MDH-2
mtNAD-ME1-1
PPA6-1
cNADP-ME1-1
ASP3-1
ADK5-1
PPA2-2
PPC-1E1c
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)]

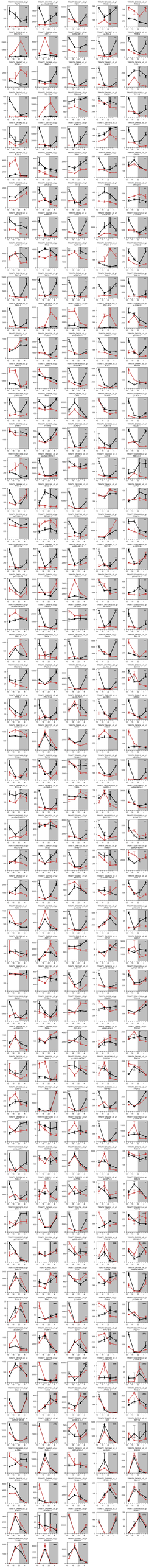

Transitory starch pathway
Starch/sugar transport
Starch synthesis
Starch degradation
Proton transport
Photorespiration
PEPC regulation
PEP generation
Metabolite transport
Decarboxylation
Circadian/Light response
Carboxylation
Calvin cycle
Aminotransferase
Gene functional category
A. P. amilis well-watered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

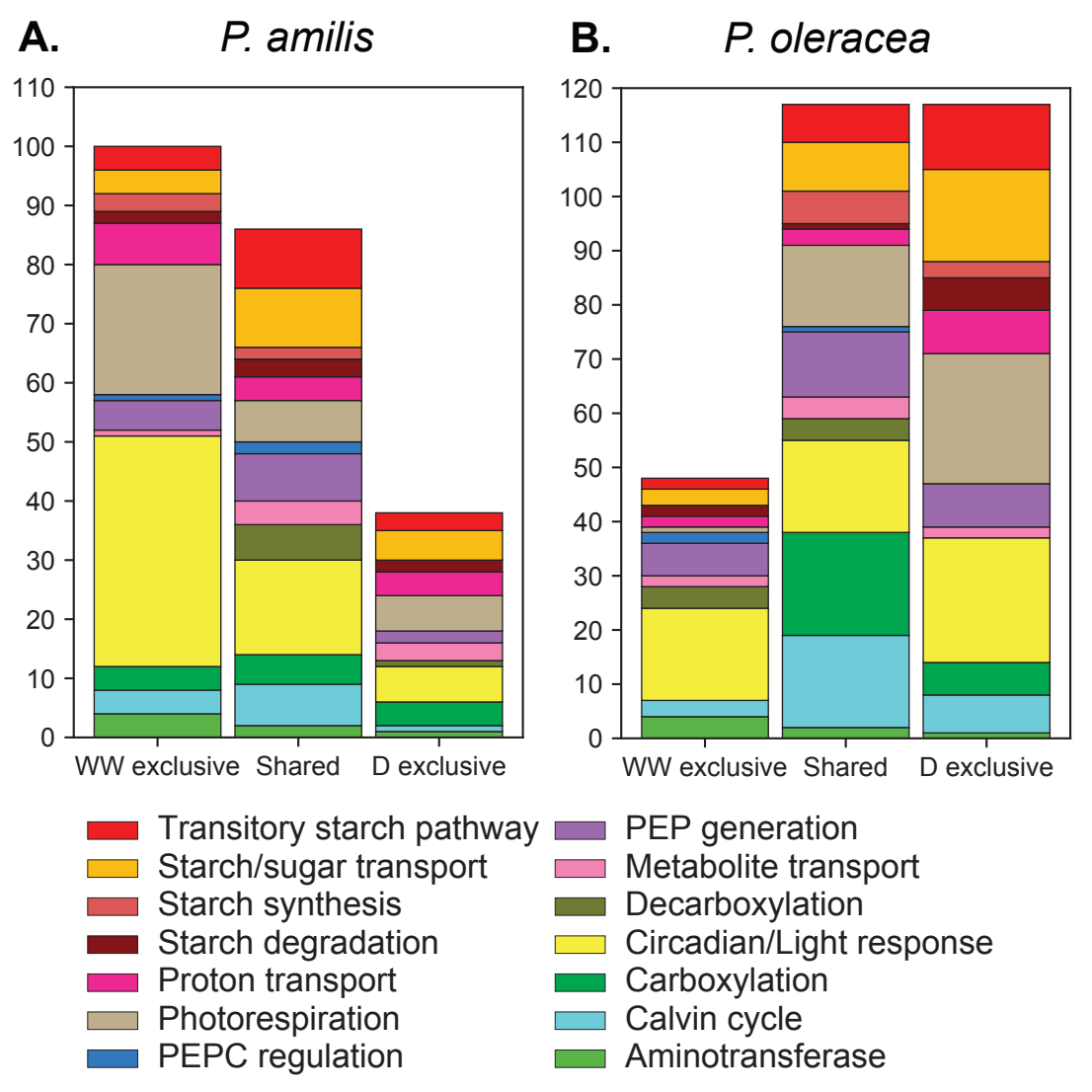

A.
paPPC-1E1c modules
30
25
20
15
10
5
0
WW exclusive
Preserved
Drought exclusive
WW exclusive
preserved genes
D exclusive
PPC-1E1c well-watered modules
PPC-1E1c drought modules
B.
poPPC-1E1c modules
110
100
90
80
70
60
50
40
30
20
10
0
WW exclusive
Preserved
Drought exclusive
paPPC-1E1c drought module
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
GI-1
APRR5-1
ERDL16-2
GWD1-1
PHS1-2
D. P. oleracea common drought network
PHS1-2
ELF1-1
ADO3-2
PPC-1E1c
PPC-1E1c
DPE1-1
LSF1-1
ISA3-1
PHS1-2
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
Pi
heteroglycans n
PEP
HCO3-
OAA
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
CO2
BCA
HCO3
PEPC-1E1a'
OAA
PEP
NADP-MDH
malate
malate
pyruvate
PPDK
PEP
mesophyll
bundle sheath
CBC
pyruvate
CO2
NADP-ME
malate
B. Portulaca amilis drought
NAD-ME1-1
NAD-ME2-1
malate
malate
pyruvate
CO2
NAD-MDH
CO2
BCA
HCO3
PUMP5-1
PEPC-1E1a'
OAA
PEP
malic acid
OAA
PEPC-1E1c
PEP
NADP-MDH
malate
NADP-MDH
malate
malate
transitory starches
CO2
NADP-ME4-5
pyruvate
PPDK
PEP
mesophyll
bundle sheath
CBC
pyruvate
CO2
NADP-ME
malate
malic acid
NAD-specific
NADP-specific
Shared
mitochondrion
chloroplast
vacuole

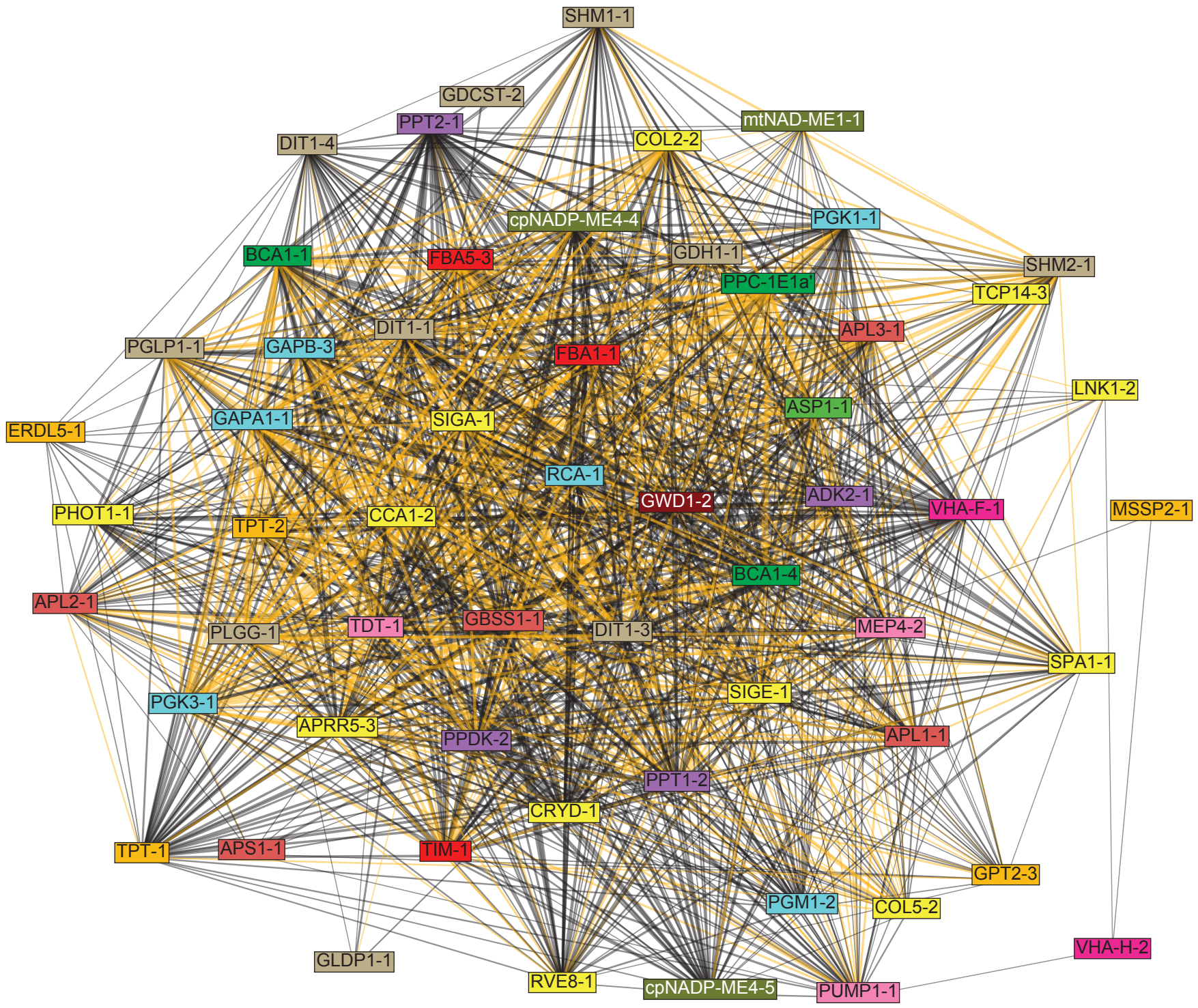

A.
Portulaca amilis
Well-watered
Drought
Module members
Module number
B.
Portulaca oleracea
Well-watered
Drought
Module members
Module number
C.
Well-watered
Drought
PGN genes
Node degree
D.
Well-watered
Drought
PGN genes
Node degree



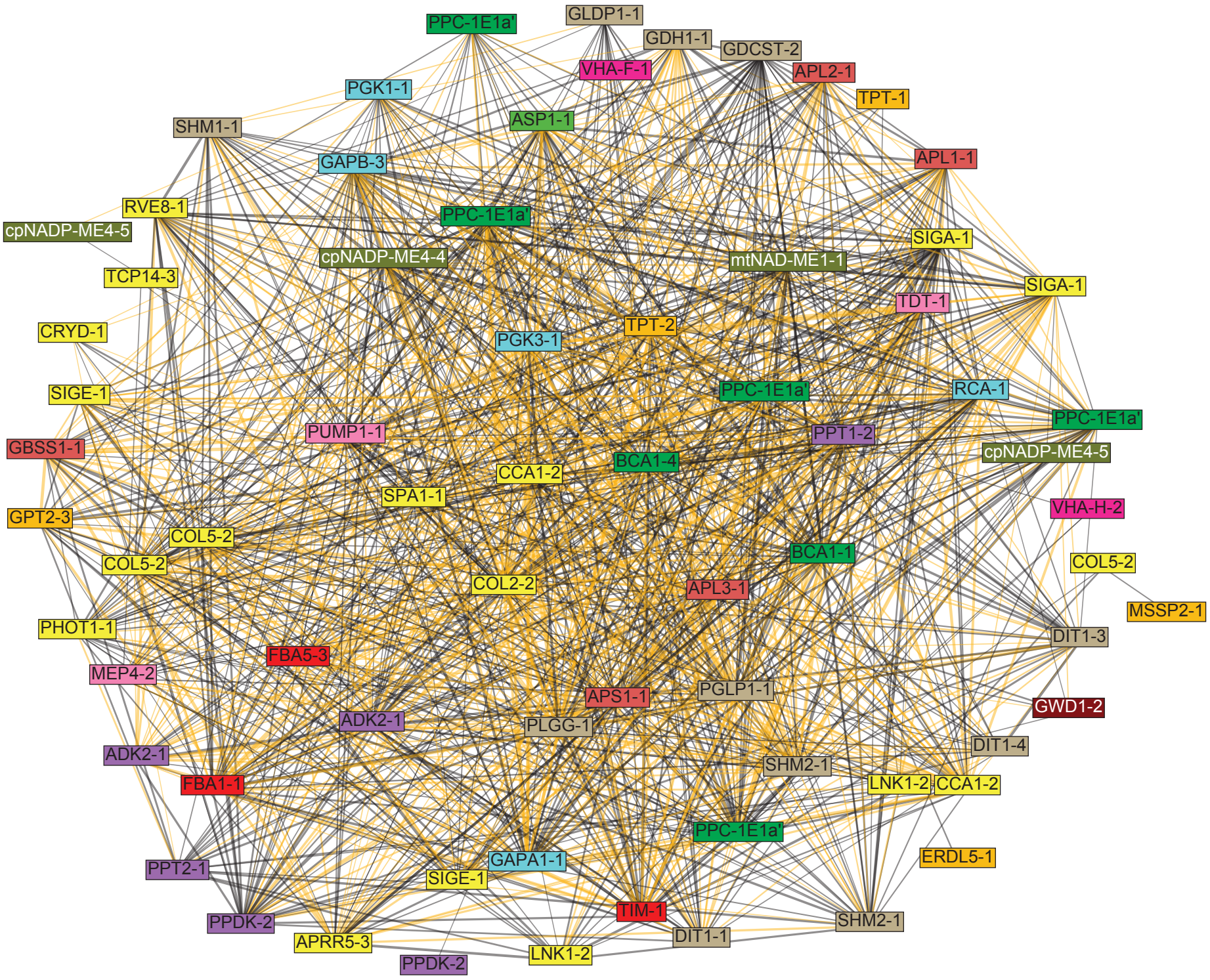



A.
P. amilis
B.
P. oleracea
110
100
90
80
70
60
50
40
30
20
10
0
120
WW exclusive
Shared
D exclusive
Transitory starch pathway
Starch/sugar transport
Starch synthesis
Starch degradation
Proton transport
Photorespiration
PEPC regulation
PEP generation
Metabolite transport
Decarboxylation
Circadian/Light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

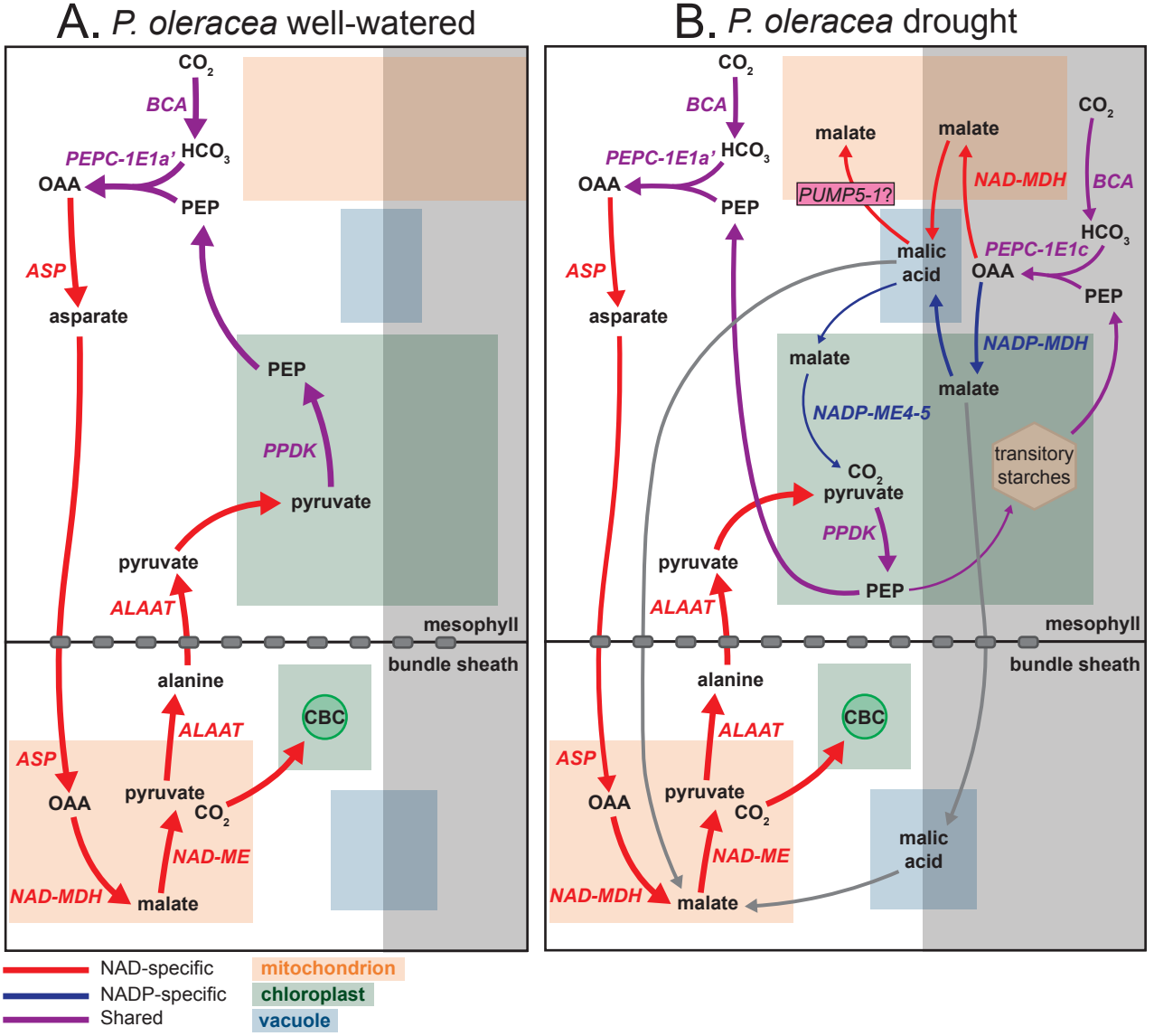
A. *P. amilis* well-watered C4 PGN



B. *P. oleracea* well-watered C4 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
PEP
PPDK
pyruvate
pyruvate
ALAAT
mesophyll
bundle sheath
alanine
ALAAT
CBC
ASP
OAA
pyruvate
CO2
NAD-ME
NAD-MDH
malate
malate
malate
NAD-MDH
CO2
BCA
PUMP5-1?
HCO3
malic acid
PEPC-1E1c
OAA
PEP
NADP-MDH
malate
malate
NADP-ME4-5
transitory starches
CO2
pyruvate
PPDK
PEP
malic acid
NAD-specific
NADP-specific
Shared
mitochondrion
chloroplast
vacuole