View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Apollo

1 **Ancestral light and chloroplast regulation form the foundations for C_4 gene expression**

2

3 Steven J. Burgess, Ignasi Granero-Moya, Mathieu J. Grangé-Guermente, Chris Boursnell,
4 Matthew J. Terry[1] and Julian M. Hibberd*

5

6 Department of Plant Sciences, Downing Street, University of Cambridge, Cambridge CB2 3EA, UK
7 [1]Biological Sciences, University of Southampton, Highfield Campus, Southampton, SO17 1BJ, UK

8

9

10

11 SJB - sjb287@cam.ac.uk
12 IG-M - ignasi.granero@e-campus.uab.cat
13 MJGG - mjg239@cam.ac.uk
14 CB - cmb211@cam.ac.uk
15 MJT - mjt@soton.ac.uk
16 JMH (corresponding) - jmh65@cam.ac.uk

17

18

19 Number of pages: 18
20 Number of Figures: 3
21 Number of Colour Figures: 3
22 Word Count: 1947
23 Number of Supplementary Figures: 6
24 Number of Supplementary Tables: 7

25 **Introductory paragraph**

26 C_4 photosynthesis acts as a carbon concentrating mechanism that leads to large increases in
27 photosynthetic efficiency. The C_4 pathway is found in over sixty plant lineages[1] but the molecular
28 enablers of this evolution are poorly understood. In particular, it is unclear how non-photosynthetic
29 proteins in the ancestral C_3 system have repeatedly become strongly expressed and integrated
30 into photosynthesis gene regulatory networks in C_4 leaves. Here, we provide clear evidence that in
31 C_3 leaves, genes encoding key enzymes of the C_4 pathway are already co-regulated with
32 photosynthesis genes and are controlled by both light and chloroplast-to-nucleus signalling. In C_4
33 leaves this regulation becomes increasingly dependent on the chloroplast. We propose that
34 regulation of C_4 cycle genes by light and the chloroplast in the ancestral C_3 state has facilitated the
35 repeated evolution of the complex and convergent C_4 trait.

36

37 **Main**

38 In C_3 leaves, photosynthesis genes are regulated directly by light-responsive networks[2] as well as
39 by retrograde signalling from the chloroplast[3]. Although it has been proposed that during the
40 evolution of the C_4 trait genes encoding components of the C_4 pathway are incorporated into
41 photosynthesis networks[4–6], and a small number of C_4 genes are known to respond to light[7], how
42 this comes about is unclear. For example, proteins used in the C_4 pathway fulfil a number of
43 disparate roles in C_3 plants[8], none of which are associated with the core photosynthetic process. It
44 has therefore been puzzling how genes encoding these proteins become integrated into the gene
45 regulatory networks that underpin photosynthesis during the transition from C_3 to C_4 metabolism.
46 To address this we undertook a systems analysis of genes important for the C_4 pathway in the C_4
47 species *Gynandropsis gynandra* (formerly designated as *Cleome gynandra*)[9]. In order to construct
48 a model of the evolutionary events leading to formation of the C_4 photosynthesis, we compared
49 these results with analysis of genes from *A. thaliana,* which we used as a proxy for the ancestral
50 C_3 state.

51

52 RNA-SEQ analysis of etiolated *G. gynandra* seedlings after 0 or 6 h of illumination identified 2671
53 transcripts as being differentially expressed between the two conditions (Table S1). Of these, 1754
54 were induced and 926 repressed after exposure to light. 54 genes appeared to undergo differential
55 splicing, resulting in transcripts that fell into both up- and down-regulated datasets (Table S1). It
56 was notable that Biological Process Gene Ontology (GO) terms[10,11] representing photosynthesis
57 and plastid organisation dominated the light-induced samples (Table S2), whilst those relating to
58 auxin and brassinosteroid responses were enriched in the light-repressed dataset (Table S2).
59 These data are consistent with analysis of the dark to light response in the C_3 model *A. thaliana*
60 (Figure S1)[12] and support the notion that there is considerable conservation in the basic responses
61 to light in these C_3 and C_4 species.

62

63 Notably, in C_4 *G. gynandra* 15 of 18 genes associated with the C_4 pathway showed increased
64 transcript abundance after exposure to light (Figure 1; Table S3). Of these, transcripts from
65 *PHOSPHOENOLPYRUVATE CARBOXYLASE (GgPPC2)* and *PYRUVATE,ORTHOPHOSPHATE*
66 *DIKINASE (GgPPDK)* were amongst the ten most abundant in the dataset (Table S4). The majority
67 of genes involved in the C_4 cycle in *G. gynandra* are therefore regulated by networks that respond
68 either directly or indirectly to light, and a small number of C_4 genes are as responsive to light as
69 those of the light-dependent reactions of photosynthesis.

70

71 To quantify the importance of chloroplast-to-nucleus signals for the response of C_4 genes to light,
72 we conducted greening experiments with *G. gynandra* in the presence of Norflurazon (NF) which
73 inhibits carotenoid biosynthesis[3] and therefore chloroplast development (Figure S2) resulting in the
74 reduced expression of many nuclear-encoded chloroplast genes[3]. Consistent with studies of C_3 *A.*
75 *thaliana,* light-induction of *G. gynandra* genes for *RIBULOSE BISPHOSPHATE CARBOXYLASE*
76 *OXYGENASE SMALL SUBNUT (RBCS*) and *LIGHT HARVESTING COMPLEX A* (*LHCA*) was
77 chloroplast dependent (Figure 2A). Moreover, the response of twelve C_4 genes to light was
78 perturbed by NF indicating that a large proportion of the C_4 pathway in *G. gynandra* is regulated by
79 the chloroplast. Five C_4 genes (*DICARBOXYLIC ACID TRANSPORTER 1 (GgDIC1),*
80 *SODIUM:HYDROGEN ANTIPORTER 1 (GgNHD1), BILE SODIUM ACID TRANSPORTER 2*
81 *(GgBASS2), GgPPDK* and *PYROPHOSPHORLYASE 6* (*GgPPA6))* were almost entirely
82 dependent on the chloroplast*,* as NF abolished increases in transcript abundance in response to
83 light (Figure 2B). Partial dependence on the chloroplast for light-induced expression was also
84 detected for four genes (*BETA CARBONIC ANHYDRASE 1, BETA CARBONIC ANHYDRASE 2,*
85 *GgPPC2,* and *PHOSPHOENOLPYRUVATE TRANSPORTER 1 (GgPPT1)),* which showed a
86 reduced response to light in the presence of NF (Figure 2B). In addition, other C_4 genes showed
87 more complex behaviours consistent with the operation of plastid signals. This included *GgTPT,*
88 whose transcripts increased on exposure to light, but decreased in the presence of NF (Figure 2B)
89 suggesting regulation is not only dependent on plastid signals in the light, but also on plastid
90 signals that operate independently of a light requirement. Furthermore, *MALATE*
91 *DEHYDROGENASE (GgMDH)* and *PHOSPHOENOLPYRUVATE CARBOXYKINASE (GgPCK1)*
92 were less abundant after light-induction, but this effect was reduced in the presence of NF,
93 indicating the process is chloroplast-dependent (Figure 2B). As NF is known to lead to plastid
94 damage, we sought additional lines of evidence that link C_4 gene expression with plastid to nucleus
95 signalling. Lincomycin (Lin), which inhibits plastid translation, led to results that were very
96 consistent with those obtained from NF, with 11 of the 12 genes perturbed by NF also showing a
97 clear response to Lin (Figure S3). In addition, Lin also affected expression of *NADME1* and *ASP1*
98 and for some genes (*CA1, CA2, PPC2, PPT1*) demonstrated a more complete dependence on the
99 chloroplast than was observed with NF (Figure S3). The *hcf136* mutant from maize, which has

100 previously been linked to plastid retrograde signalling, also shows perturbed expression for many
101 C_4 genes. Specifically, genes encoding *PEPC*, *CA*, *NADPME, NADP-MDH, PPDK, PCK* and
102 *AspAT* all showed altered expression in the *Zmhcf136* mutant (Table S6)[13]. Combined with the
103 report that that NF treatment of maize reduces transcript levels of *PPDK*[14], these data support the
104 notion that regulation by the chloroplast is important in other C_4 lineages. Overall, these data argue
105 strongly for many genes of C_4 photosynthesis being subject to regulation by chloroplast signalling,
106 light signalling or a combination of both.
107
108 We next sought to assess the extent to which genes orthologous to those recruited into C_4
109 photosynthesis are subject to light regulation in the ancestral C_3 state. Using publically available
110 Chromatin Immuno-Precipitation (ChIP), microarray[15] (Figure S4) and mutant data from C_3 *A.*
111 *thaliana*, direct or indirect interactions were inferred based on whether or not mutant data were
112 supported by ChIP-SEQ signals. This indicated that C_4 genes in *A. thaliana,* which are not used in
113 photosynthesis, are subject to both direct and indirect regulation by phytochromes that mediate red
114 light-signalling pathways[16], but also by HY5 a master regulator of photomorphogenesis[17], and by
115 phytochrome interacting factors (PIFs) that act to repress gene expression in the dark[18] (Figure
116 3A). To our knowledge, interactions between phyB and C_4 genes are only available from analysis
117 of insertional mutants, whereas for phyA ChIP-SEQ data is also available (Figure 3A). For
118 example, *AtTPT* is exclusively regulated by phyA (Table S5)[19], whereas *AtCA2* is exclusively
119 repressed by phyB (Table S5)[20]. Six C_4 genes (*AtCA1*, *AtPPC2*, *AtBASS2*, *AtRP1*, *AtMDH1* and
120 *AtPCK1*) are regulated by both phyA and phyB, as individual knockout of these phytochromes
121 resulted in changes in gene expression (Table S5)[19,20]. A further five genes (*AtPEPCk1, AtNAD-*
122 *ME2, AtALAAT2, AtPPDK* and *AtPPA6*) are under redundant regulation by phyA or phyB, being
123 mis-expressed in *phyAphyB* double mutants[21] (Table S5). Downstream of phytochromes, six C_4
124 genes are activated by HY5 (*AtPPT1, AtCA1, AtTPT, AtBASS2, AtPPA6* and *AtMDH1*) and four
125 are repressed (*AtDIC1, AtPPDK, AtPEPCk1* and *AtPCK1*; Figure 3A and Table S5)[17,22]. PIFs act
126 antagonistically with *HY5* by competitively binding E-box sequences[23]. Seven C_4 genes (*AtASP1*,
127 *AtBASS2*, *AtPPA6*, *AtPCK1*, *AtCA1*, *AtDIC1* and *AtPPC2*) are directly or indirectly regulated by
128 either PIF1, 3 or 4 (Figure 3A)[24–26]. Overall, these data indicate that 16 of the 18 *A. thaliana* genes
129 orthologous to those involved in the C_4 cycle in *G. gynandra* are under direct or indirect regulation
130 by light-signalling components (Figure 3A).
131
132 As many C_4 genes from *G. gynandra* showed chloroplast regulation (Figure 2) we investigated the
133 extent to which C_4 orthologues from C_3 *A. thaliana* are regulated by retrograde signalling from the
134 chloroplast. Consistent with previous reports[27], *AtLHCA* and *AtRBCS* were responsive to the
135 chloroplast with NF largely abolishing their light induction (Figure S5A). Surprisingly, seven genes
136 of the core C_4 cycle (*AtCA1*, *AtPPC2, AtNADME1, AtNADME2, AtPPA6, AtRP1* and *AtTPT*) were

137 either partially or completely dependent on signals from the chloroplast for the light-induction
138 response (Figure S5B).
139
140 Most C_4 genes in *G. gynandra* and their orthologs in *A. thaliana* can be categorised into four
141 groups depending on their response to illumination in the presence or absence of chloroplast
142 signalling inhibitors (Figure 3B), and so we ordered the categories according to increasing
143 influence of the chloroplast over this process. Although it is possible that the regulation of C_4 cycle
144 genes in *A. thaliana* has diverged since the last common ancestor of *A. thaliana* and *G. gynandra*,
145 we use the data to propose a model for the evolution of C_4 photosynthesis. The key features are
146 first, that contrary to previous suggestions[4,5] many C_4 genes appear to be operate within light
147 regulatory networks in the C_3 state and second, that there is a degree of "fine-tuning" in terms of
148 light and chloroplast regulation which may explain increases in gene expression during evolution of
149 the C_4 pathway. Seven C_4 genes in *A. thaliana* responded to white light (Figure 3B), and there was
150 also evidence that a further nine are integrated to some extent into light-signalling networks (Figure
151 3A). We further suggest that gain of light-activated chloroplast-dependent expression has occurred
152 for eight genes (*GgCA2, GgASP1, GgPCK1, GgDIC1, GgNHD1, GgBASS2, GgPPDK,* and
153 *GgPPT1*) in *G. gynandra*. (Figure S6). These data strongly imply that evolution of chloroplast
154 function in C_4 leaves is underpinned by a shared molecular mechanism that is required for
155 establishment of the C_4 carbon pump.
156
157 Ancestral regulation opens up the possibility that increases in C_4 gene expression could be driven
158 by changes in transcription factor abundance, however we found no evidence for significant
159 changes in phyA, phyB, HY5 or PIFs in available data for *G. gynandra*[5]. As many genes encoding
160 components of the C_4 cycle are encoded by multi-gene families it is possible that ancestral
161 regulation by either light or the chloroplast predisposes certain members of these gene families to
162 recruitment into C_4 photosynthesis (Table S7). It is notable that in *A. thaliana* the *PPC* gene family
163 contains four members, of which only one (*AtPPC2*) is light-responsive[27]. It may be no coincidence
164 that it is the ortholog of *AtPPC2* that was recruited into the C_4 pathway in *G. gynandra*. This
165 situation would require a boosting of pre-existing signals rather than a gain of new ones, and would
166 support the notion that some genes are primed for involvement in C_4 photosynthesis[28]. However,
167 this is unlikely to be the only important factor and we suggest that during the course of C_4
168 evolution, ancestral characteristics including sub-cellular location of proteins, transcript abundance,
169 and light and chloroplast regulation, impact on the recruitment of specific genes from multi-gene
170 families.
171
172 The C_4 pathway is commonly described as one of the most remarkable examples of convergent
173 evolution in biology. Our finding that a large number of C_4 genes in the C_3 model *A. thaliana* are
174 regulated by light and the chloroplast indicates that their complete integration into photosynthesis

175 networks in C_4 leaves is not as large a step as has previously been thought. If this is true for
176 additional C_3 lineages that are sister to C_4 origins, it would help explain why C_4 photosynthesis has
177 appeared in over sixty lineages of plants. It would also expand the extent to which the convergent
178 evolution of this highly complex trait is based on parallel evolution at the molecular level[6,29,30].
179

180 **Acknowledgements**
181 The work was funded by the European Union *3to4* project and Biotechnology and Biological
182 Sciences Research Council (BBSRC) grant BB/J011754/1. Ignasi Granero-Moya was supported by
183 the Amgen Foundation. Research on chloroplast signalling by MJT was supported by BBSRC
184 grant (BB/J018139/1).
185

186 **Contributions of authors**
187 SJB, MJT and JMH designed the study. SJB, IG-M and MJG-G carried out the experimental work.
188 SJB and CB conducted the bioinformatics. SJB, MJT and JMH interpreted results and wrote the
189 paper.

Materials and Methods

Plant Material and Growth Conditions

Both species were grown as described previously[1]. Briefly, *Arabidopsis thaliana* Col-0 seeds were surface-sterilized, and seedlings then grown at 22 °C on plates containing 0.5 strength Murashige and Skoog salts and 0.8% (w/v) agar with or without 5 µM (*A. thaliana*) or 50 µM (*G. gynandra*) NF. Lincomycin was provided at 0.5mM to both *A. thaliana* and *G. gyndandra*. As NF was dissolved in ethanol, to control for any differences in gene expression resulting from addition of the solvent, an equal amount of ethanol was added to samples not treated with NF (5 µL, *A. thaliana*; 50 µL, *G. gynandra*). *A. thaliana* seeds were stratified for 72 h by dark incubation at 4 °C and then germination was induced by exposure to 100 µmol m^{-2} s^{-1} white light for 1 h. Germination of *G. gynandra* was stimulated by incubation in the dark at 32 °C on damp tissue paper for 36 h.

To carry out de-etiolation experiments, seedlings were placed at 22 °C in the dark to promote hypocotyl extension. For Quantitative Reverse Transcriptase Polymerase Chain Reactions (QRT-PCR) and RNA-Sequencing (RNA-SEQ), samples were taken after three and seven days respectively. At 0 h, material was harvested under green light, and after 6 or 24 h under white light. Tissue was flash-frozen in liquid nitrogen and stored at -80 °C prior to processing.

RNA extraction, QRT-PCR and RNA-SEQ

Samples were ground in a mortar and pestle and RNA extraction was carried out with the RNeasy® Plant Mini Kit (74904; QIAGEN) according to the manufacturer's instructions, with the exception that the wash with PE buffer was repeated five times to remove residual guanidium thiocyanate contamination.

For QRT-PCR, 0.2 µg RNA was incubated with Superscript® II Reverse Transcriptase (18064-022; ThermoFisher Scientific) to generate cDNA. QRT-PCR using SYBR® Green Jumpstart™ *Taq* ReadyMix™ (S4438-100RXN; Sigma-Aldrich) in a Rotor-Gene Q system (QIAGEN) was then used to quantify transcript abundance. Relative expression was calculated by comparison to *ACTIN7* and data processed using REST 2009 software[2]. For each gene, four technical and three biological replicates were performed.

RNA sequencing was performed by the Department of Biochemistry Sequencing Services at the University of Cambridge, UK. 0.2 µg RNA was used for library preparation using the TruSeq RNA Library Preparation Kit v2 (RS-122-2001; Illumina). Samples were analysed on a NextSeq500 (Illumina) Mid Output 150 cycle run. Data processing was performed using custom scripts. Briefly, reads were processed using Trimmomatic[3] and Salmon version 0.4.2[4] was used to align reads to the *G. gynandra* transcriptome and perform quantification analysis. Differential expression analysis was performed using baySeq[5,6] with a False Discovery Rate (FDR) set at <0.05. *G. gynandra*

228 transcripts were annotated with respect to orthologous genes from *A. thaliana* by mapping to the *A.*
229 *thaliana* genome using reciprocal best blast[7]. Genes were annotated with the database
230 org.At.tair.db[8] and divided into functional categories using Gene Ontology (GO) terms[9,10].
231 Enrichment was calculated using the Bioconductor packages topGO[11]. Plots were generated with
232 custom scripts in RStudio[12] using the package ggplot2[13].
233

234 **References**

235 1. Sage, R. F., Christin, P.-A. & Edwards, E. J. The C4 plant lineages of planet Earth. *J. Exp.*
236 *Bot.* **62,** 3171–3181 (2011).
237 2. Jiao, Y., Lau, O. S. & Deng, X. W. Light-regulated transcriptional networks in higher plants.
238 *Nat. Rev. Genet.* **8,** 217–230 (2007).
239 3. Chan, K. X., Phua, S. Y., Crisp, P., McQuinn, R. & Pogson, B. J. Learning the languages of
240 the chloroplast: Retrograde signaling and beyond. *Annu. Rev. Plant Biol.* (2015).
241 doi:10.1146/annurev-arplant-043015-111854
242 4. Aubry, S., Kelly, S., Kümpers, B. M. C., Smith-Unna, R. D. & Hibberd, J. M. Deep
243 evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in
244 two independent origins of C4 photosynthesis. *PLOS Genet.* **10,** e1004365 (2014).
245 5. Kulahoglu, C. *et al.* Comparative transcriptome atlases reveal altered gene expression
246 modules between two Cleomaceae C3 and C4 plant species. *Plant Cell* **26,** 3243–3260
247 (2014).
248 6. Christin, P.-A. *et al.* Parallel recruitment of multiple genes into C4 photosynthesis. *Genome*
249 *Biol. Evol.* **5,** 2174–87 (2013).
250 7. Hibberd, J. M. & Covshoff, S. The regulation of gene expression required for C4
251 photosynthesis. *Annu Rev Plant Biol* **61,** 181–207 (2010).
252 8. Eastmond, P. J. *et al.* Arabidopsis uses two gluconeogenic gateways for organic acids to
253 fuel seedling establishment. *Nat. Commun.* **6,** (2015).
254 9. Marshall, D. M. *et al.* Cleome, a genus closely related to Arabidopsis, contains species
255 spanning a developmental progression from C3 to C4 photosynthesis. *Plant J.* **51,** 886–896
256 (2007).
257 10. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25,** 25–29
258 (2000).
259 11. The Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids*
260 *Res.* **43,** D1049–D1056 (2015).
261 12. Ma, L. *et al.* Light control of Arabidopsis development entails coordinated regulation of
262 genome expression and cellular pathways. *Plant Cell* **13,** 2589–2608 (2001).
263 13. Covshoff, S. *et al.* Deregulation of maize C4 photosynthetic development in a mesophyll
264 cell-defective mutant. *Plant Physiol* **146,** 1469–1481 (2008).
265 14. Tamada, Y. *et al.* Effect of photooxidative destruction of chloroplasts on the expression of
266 nuclear genes for C4 photosynthesis and for chloroplast biogenesis in maize. *J. Plant*
267 *Physiol.* **160,** 3–8 (2003).
268 15. Charron, J.-B. F., He, H., Elling, A. A. & Deng, X. W. Dynamic landscapes of four histone
269 modifications during deetiolation in Arabidopsis. *Plant Cell* **21,** 3732–3748 (2009).
270 16. Rockwell, N. C., Su, Y.-S. & Lagarias, J. C. Phytochrome Structure and Signaling
271 Mechanisms. *Annu. Rev. Plant Biol.* **57,** 837–858 (2006).

272 17. Lee, J. *et al.* Analysis of transcription factor HY5 genomic binding sites revealed its
273 hierarchical role in light regulation of development. *Plant Cell* **19,** 731–749 (2007).
274 18. Leivar, P. & Quail, P. H. PIFs: pivotal components in a cellular signaling hub. *Trends Plant*
275 *Sci.* **16,** 19–28 (2011).
276 19. Chen, F. *et al.* Arabidopsis phytochrome A directly targets numerous promoters for
277 individualized modulation of genes in a wide range of pathways. *Plant Cell* **26,** 1949–1966
278 (2014).
279 20. Tepperman, J. M. *et al.* Expression profiling of phyB mutant demonstrates substantial
280 contribution of other phytochromes to red-light-regulated gene expression during seedling
281 de-etiolation. *Plant J.* **38,** 725–739 (2004).
282 21. Shikata, H. *et al.* Phytochrome controls alternative splicing to mediate light responses in
283 Arabidopsis. *Proc. Natl. Acad. Sci. USA* **111,** 18781–18786 (2014).
284 22. Zhang, H. *et al.* Genome-wide mapping of the HY5-mediated genenetworks in Arabidopsis
285 that involve both transcriptional and post-transcriptional regulation. *Plant J.* **65,** 346–358
286 (2011).
287 23. Toledo-Ortiz, G. *et al.* The HY5-PIF regulatory module coordinates light and temperature
288 control of photosynthetic gene transcription. *PLOS Genet.* **10,** e1004416 (2014).
289 24. Oh, E., Zhu, J.-Y. & Wang, Z.-Y. Interaction between BZR1 and PIF4 integrates
290 brassinosteroid and environmental responses. *Nat. Cell Biol.* **14,** 802–809 (2012).
291 25. Oh, E. *et al.* Genome-Wide Analysis of Genes Targeted by PHYTOCHROME
292 INTERACTING FACTOR 3-LIKE5 during Seed Germination in Arabidopsis. *Plant Cell* **21,**
293 403–419 (2009).
294 26. Zhang, Y. *et al.* A Quartet of PIF bHLH Factors Provides a Transcriptionally Centered
295 Signaling Hub That Regulates Seedling Morphogenesis through Differential Expression-
296 Patterning of Shared Target Genes in Arabidopsis. *PLoS Genet.* **9,** e1003244 (2013).
297 27. Ruckle, M. E., DeMarco, S. M. & Larkin, R. M. Plastid Signals Remodel Light Signaling
298 Networks and Are Essential for Efficient Chloroplast Biogenesis in Arabidopsis. *Plant Cell*
299 **19,** 3944–3960 (2007).
300 28. Kajala, K. *et al.* Multiple Arabidopsis genes primed for recruitment into C4 photosynthesis.
301 *Plant J.* **69,** 47–56 (2012).
302 29. Brown, N. J. *et al.* Independent and parallel recruitment of preexisting mechanisms
303 underlying C4 photosynthesis. *Science (80-.).* **331,** 1436–1439 (2011).
304 30. John, C. R., Smith-Unna, R. D., Woodfield, H., Covshoff, S. & Hibberd, J. M. Evolutionary
305 convergence of cell-specific gene expression in independent lineages of C4 grasses. *Plant*
306 *Physiol.* **165,** 62–75 (2014).
307

308 **Figure 1: Light induction of C_4 genes in *G. gynandra***. RNA-SEQ data are plotted as Transcript
309 Per Million (TPM). Transcript abundance is shown from samples collected from dark-grown
310 seedlings (D) and after transfer to light (L) for 6 h. *CA1, CARBONIC ANHYDRASE 1; CA2,*
311 *CARBONIC ANHYDRASE 2; PPC2, PHOSPHOENOLPYRUVATE CARBOXYLASE 2; PPCK1,*
312 *PHOSPHOENOLPYRUVATE CARBOXYLASE KINASE 1; ASP1, ASPARTATE AMINO*
313 *TRANSFERASE 1; PCK1, PHOSPHOENOLPYRUVATE CARBOXYKINASE 1; DIC1,*
314 *DICARBOXYLIC ACID TRANSPORTER 1; MDH, MITOCHONDRIAL MALATE*
315 *DEHYDROGENASE; NAD-ME1, NAD-DEPENDENT MALIC ENZYME 1; NAD-ME2, NAD-*
316 *DEPENDENT MALIC ENZYME 2; ALAAT2, ALANINE AMINO TRANSFERASE 2; NHD1,*
317 *SODIUM:HYDROGEN ANTIPORTER 1; BASS2, BILE SODIUM ACID TRANSPORTER 2; PPDK,*
318 *PYRUVATE,ORTHOPHOSPHATE DIKINASE; RP1; PPDK REGULATORY PROTEIN 1; PPA6,*
319 *PYROPHOSPHORLYASE 6; PPT1, PHOSPHOENOLPYRUVATE TRANSPORTER 1; TPT,*
320 *TRIOSE PHOSPHATE TRANSPORTER*. Data are shown as means and standard errors of three
321 biological replicates, the calculated significance values for differential expression are indicated:
322 FDR 0.05-0.01 (*) and FDR<0.01 (**).
323
324 **Figure 2: Chloroplast regulation of C_4 genes in *G. gynandra*.** Quantitative real time polymerase
325 chain reactions were used to determine the impact of light and chloroplast signalling on C_4
326 transcript abundance in *G. gynandra*. Data are derived from samples collected from dark-grown
327 seedlings ± 50 µM Norflurazon (NF) before (light green) and after transfer to light (dark green) for
328 24 h. **A.** *LHCA, LIGHT HARVESTING COMPLEX A; RBCS, RIBULOSE-1,5-BISPHOSPHATE*
329 *CARBOXYLASE/OXYGENASE SMALL SUBUNIT 1A;* **B.** *CA1, CARBONIC ANHYDRASE 1; CA2,*
330 *CARBONIC ANHYDRASE 2; PPC2, PHOSPHOENOLPYRUVATE CARBOXYLASE 2; ASP1,*
331 *ASPARTATE AMINO TRANSFERASE 1; PCK1, PHOSPHOENOLPYRUVATE CARBOXYKINASE*
332 *1; DIC1, DICARBOXYLIC ACID TRANSPORTER 1; MDH, MITOCHONDRIAL MALATE*
333 *DEHYDROGENASE; NAD-ME1, NAD-DEPENDENT MALIC ENZYME 1; NAD-ME2, NAD-*
334 *DEPENDENT MALIC ENZYME 2; ALAAT2, ALANINE AMINO TRANSFERASE 2; NHD1,*
335 *SODIUM:HYDROGEN ANTIPORTER 1; BASS2, BILE SODIUM ACID TRANSPORTER 2; PPDK,*
336 *PYRUVATE,ORTHOPHOSPHATE DIKINASE; RP1; PPDK REGULATORY PROTEIN 1; PPA6,*
337 *PYROPHOSPHORLYASE 6; PPT1, PHOSPHOENOLPYRUVATE TRANSPORTER 1; TPT,*
338 *TRIOSE PHOSPHATE TRANSPORTER.* Data are shown as means with error bars representing
339 upper and lower bounds of a 95% confidence interval from three biological replicates. Differential
340 expression (p<0.1) compared to 0 h without NF is indicated by (*).
341
342 **Figure 3: Proposed model of changes in light regulation during evolution of the C_4 pathway**
343 **in *G. gynandra*. A.** Composite model of the regulation of C_4 orthologs in *A. thaliana* by known light
344 signalling components based on available chromatin immunoprecipitation datasets and analysis of
345 mutants. Symbols: *phyA, PHYTOCHROME A; phyB, PHYTOCHROME B; CA1, CARBONIC*

346 *ANHYDRASE 1; PPC2, PHOSPHOENOLPYRUVATE CARBOXYLASE 2; TPT, TRIOSE*
347 *PHOSPHATE TRANSPORTER, BASS2, BILE SODIUM ACID TRANSPORTER 2; CA2,*
348 *CARBONIC ANHYDRASE 2; PPCK1, PHOSPHOENOLPYRUVATE CARBOXYLASE KINASE 1;*
349 *NAD-ME2, NAD-DEPENDENT MALIC ENZYME 2; PPA6, PYROPHOSPHORLYASE 6; PPDK,*
350 *PYRUVATE,ORTHOPHOSPHATE DIKINASE; HY5, ELONGATED HYPOCOTYL 5; PPT1,*
351 *PHOSPHOENOLPYRUVATE TRANSPORTER 1; DIC1, DICARBOXYLIC ACID TRANSPORTER*
352 *1;* PIF4, PHYTOCHROME INTERACTING FACTOR 4*.* Direct and indirect repression are
353 represented by solid and dashed blue lines respectively, whereas direct and indirect activation are
354 represented by solid and dashed red arrows respectively. **B.** Diagram illustrating the proposed
355 regulation of C_4 genes in *G. gynandra* (C_4) and orthologous genes in *A. thaliana* (C_3). Categories
356 include genes unresponsive to light or chloroplast signalling, light signalling alone, light and
357 chloroplast signalling or signalling entirely dependent on the chloroplast (which includes genes for
358 which light activation is lost after chloroplast damage, as well as when further inhibition in the
359 absence of light is observed, such as after Lin treatment in the dark). The groupings are based on
360 genes showing statistically significant differences in gene expression ($p<0.1$) in the presence or
361 absence of NF and Lin during a dark to light transition (Figure 2, Figure S3, Figure S5). For *G.*
362 *gynandra* genes, the category chosen was based on the strongest response to either NF or Lin
363 provided a similar response was observed in the other treatment. *ALAAT2* is omitted from the
364 figure as its expression profile did not fit the defined profiles for light or chloroplast regulation
365 (Figure 2, Figure S3, Figure S5).

366 **Supplementary Figure Legends**

367

368 **Figure S1:** Light induction of C_4 orthologs in *A. thaliana*. Microarray data are derived from Charron
369 et al. (2009)[14]. Log fold change in transcript abundance is shown comparing samples collected
370 before and after 6 h WL light treatment. *ASP1, ASPARTATE AMINOTRANSFERASE 1; BASS2,*
371 *BILE SODIUM ACID TRANSPORTER 2; CA1, CARBONIC ANHYDRASE 1; CA2, CARBONIC*
372 *ANHYDRASE 2; DIC1, DICARBOXYLIC ACID TRANSPORTER 1; DTC1,*
373 *DICARBOXYLATE/TRICARBOXYLATE CARRIER; MDH, MITOCHONDRIAL MALATE*
374 *DEHYDROGENASE; NAD-ME1, NAD-DEPENDENT MALIC ENZYME 1; NAD-ME2, NAD-*
375 *DEPENDENT MALIC ENZYME 2; NHD1, SODIUM:HYDROGEN ANTIPORTER 1; PCK1,*
376 *PHOSPHOENOLPYRUVATE CARBOXYKINASE 1; PPCK1, PHOSPHOENOLPYRUVATE*
377 *CARBOXYLASE KINASE 1; PPA6, PYROPHOSPHORLYASE 6; PPC2,*
378 *PHOSPHOENOLPYRUVATE CARBOXYLASE 2; PPDK, PYRUVATE,ORTHOPHOSPHATE*
379 *DIKINASE; RP1; PPDK REGULATORY PROTEIN 1; TPT, TRIOSE PHOSPHATE*
380 *TRANSPORTER*.

381

382 **Figure S2:** Photographs of *G. gynandra* seedlings grown in the presence of varying concentrations
383 of NF (values presented above image with the units µM). Representative images are derived from
384 samples collected from 3 day-old dark-grown seedlings after transfer to light for 24 h.

385

386 **Figure S3: Chloroplast regulation of C_4 genes in *G. gynandra*.** Quantitative real time
387 polymerase chain reactions were used to determine the impact of light and chloroplast signalling
388 on C_4 transcript abundance in *G. gynandra*. Data are derived from samples collected from dark-
389 grown seedlings ± 0.5 mM Lincomycin before (light green) and after transfer to light (dark green)
390 for 24 h. **A.** *LHCA, LIGHT HARVESTING COMPLEX A; RBCS, RIBULOSE-1,5-BISPHOSPHATE*
391 *CARBOXYLASE/OXYGENASE SMALL SUBUNIT 1A;* **B.** *CA1, CARBONIC ANHYDRASE 1; CA2,*
392 *CARBONIC ANHYDRASE 2; PPC2, PHOSPHOENOLPYRUVATE CARBOXYLASE 2; ASP1,*
393 *ASPARTATE AMINO TRANSFERASE 1; PCK1, PHOSPHOENOLPYRUVATE CARBOXYKINASE*
394 *1; DIC1, DICARBOXYLIC ACID TRANSPORTER 1; MDH, MITOCHONDRIAL MALATE*
395 *DEHYDROGENASE; NAD-ME1, NAD-DEPENDENT MALIC ENZYME 1; NAD-ME2, NAD-*
396 *DEPENDENT MALIC ENZYME 2; ALAAT2, ALANINE AMINO TRANSFERASE 2; NHD1,*
397 *SODIUM:HYDROGEN ANTIPORTER 1; BASS2, BILE SODIUM ACID TRANSPORTER 2; PPDK,*
398 *PYRUVATE,ORTHOPHOSPHATE DIKINASE; RP1; PPDK REGULATORY PROTEIN 1; PPA6,*
399 *PYROPHOSPHORLYASE 6; PPT1, PHOSPHOENOLPYRUVATE TRANSPORTER 1; TPT,*
400 *TRIOSE PHOSPHATE TRANSPORTER.* Data are shown as means with error bars representing
401 standard error from five biological replicates. Differential expression ($p<0.1$) compared to 0 h
402 without NF is indicated by (*).

403

404

405 **Figure S4:** Location of phyA ChIP-SEQ binding sites over C_4 genes **A.** *AtCA1* **B.** *AtPPC2* **C.**
406 *AtBASS2*. Data was obtained from[15] and visualized using the Integrative Genome Viewer
407 (IGV)[16,17].

408

409 **Figure S5: Chloroplast regulation of C_4 genes in *A. thaliana*.** Quantitative real time polymerase
410 chain reactions were used to determine the impact of light and chloroplast signalling on C_4
411 transcript abundance in *A. thaliana*. Data are derived from samples collected from dark-grown
412 seedlings ± 5 μM Norflurazon (NF) before (light green) and after transfer to light (dark green) for 24
413 h. **A.** *LHCA, LIGHT HARVESTING COMPLEX* A; *RBCS, RIBULOSE-1,5-BISPHOSPHATE*
414 *CARBOXYLASE/OXYGENASE SMALL SUBUNIT 1A*; **B.** *CA1, CARBONIC ANHYDRASE 1; CA2,*
415 *CARBONIC ANHYDRASE 2; PPC2, PHOSPHOENOLPYRUVATE CARBOXYLASE 2; ASP1,*
416 *ASPARTATE AMINO TRANSFERASE 1; PCK1, PHOSPHOENOLPYRUVATE CARBOXYKINASE*
417 *1; DIC1, DICARBOXYLIC ACID TRANSPORTER 1; MDH, MITOCHONDRIAL MALATE*
418 *DEHYDROGENASE; NAD-ME1, NAD-DEPENDENT MALIC ENZYME 1; NAD-ME2, NAD-*
419 *DEPENDENT MALIC ENZYME 2; ALAAT2, ALANINE AMINO TRANSFERASE 2; NHD1,*
420 *SODIUM:HYDROGEN ANTIPORTER 1; BASS2, BILE SODIUM ACID TRANSPORTER 2; PPDK,*
421 *PYRUVATE,ORTHOPHOSPHATE DIKINASE; RP1; PPDK REGULATORY PROTEIN 1; PPA6,*
422 *PYROPHOSPHORLYASE 6; PPT1, PHOSPHOENOLPYRUVATE TRANSPORTER 1; TPT,*
423 *TRIOSE PHOSPHATE TRANSPORTER.* Data are shown as means with error bars representing
424 upper and lower bounds of a 95% confidence interval from three biological replicates. Differential
425 expression ($p<0.1$) compared to 0 h without NF is indicated by (*).

426

427 **Figure S6: Gain of chloroplast-dependent C_4 gene expression in *G. gynandra*.** Schematic
428 illustrating the extent of light-activated, chloroplast-dependent C_4 gene expression in the ancestral
429 C_3 state, based on data from *A. thaliana,* as well the C_4 state based on data from *G. gynandra*.

430 **Supplementary Table Legends**

431

432 **Table S1:** Summary of RNA-Seq analysis comparing samples extracted from 7 day old, etiolated
433 *G. gynandra* seedlings before and after 6 h illumination.

Replicates	3		
No. of PE reads	127,543,218		
Average reads per library	21,257,203		
Detected genes	26945		
Detected Transcripts	57296		
	Induced	Repressed	Total
No. of genes DE (FDR<0.05)	1597	887	2433
No, of transcripts DE (FDR<0.05)	1767	940	2707

434

435 **Table S2:** Top ten GO terms enriched in the *G. gynandra* light-induced and light-repressed gene
436 lists (p<0.05), following 6 h illumination of 7 day old, etiolated seedlings in light.

GO ID	GO Term	Genes Annotated	Fisher p-value
Light induced			
GO:0015979	Photosynthesis	333	< 1e-30
GO:0019684	photosynthesis, light reaction	253	< 1e-30
GO:0009657	plastid organization	346	5.10E-30
GO:0009668	plastid membrane organization	181	1.20E-21
GO:0010027	thylakoid membrane organization	181	1.20E-21
GO:0019288	isopentenyl diphosphate biosynthetic process	204	2.30E-19
GO:0019682	glyceraldehyde-3-phosphate metabolic process	205	3.30E-19
GO:0006090	pyruvate metabolic process	207	6.30E-19
GO:0009240	isopentenyl diphosphate biosynthetic process	207	6.30E-19
GO:0046490	isopentenyl diphosphate metabolic process	207	6.30E-19
Light repressed			
GO:0009825	multidimensional cell growth	87	0.00012
GO:0009741	response to brassinosteroid	83	0.00166
GO:0009733	response to auxin	272	0.0035
GO:0046685	response to arsenic-containing substance	26	0.00516
GO:0052482	defense response by cell wall thickening	15	0.00666
GO:0052544	defense response by callose deposition in cell wall	15	0.00666
GO:0048443	stamen development	114	0.00828
GO:0048466	androecium development	114	0.00828
GO:0042547	cell wall modification involved in multidimensional cell growth	16	0.00901
GO:0009958	positive gravitropism	17	0.01187

437

438 **Table S3:** Expression Data from RNA-SEQ analysis of seven day old, etiolated *G. gynanda*
439 seedlings before and after 6 hours of white light treatment. The Arabidopsis Genome Initiative
440 (AGI) identifier, locus and transcript identifiers from the *G. gynandra* draft genome (unpublished)
441 are provided alongside expression values as counts and statistical outputs from baySEQ[5,6].
442
443 **Table S4:** Expression Data from RNA-SEQ analysis of seven day old, etiolated *G. gynandra*
444 seedlings before and after 6 hours of white light treatment. The Arabidopsis Genome Initiative
445 (AGI) identifier, locus and transcript identifiers from the *G. gynandra* draft genome (unpublished)
446 are provided alongside expression values as TPMs.
447
448 **Table S5:** Meta-analysis of available gene expression and ChIP-SEQ datasets relating to
449 *Arabidopsis thaliana* light signalling components.
450
451 **Table S6:** Abundance of C_4 genes in mesophyll and bundle sheath cells comparing wild type and
452 the *Zmhcf136* mutant. Expression data are taken from[18].
453
454 **Table S7**: Abundance of members of multigene families containing enzymes orthologous to core
455 C_4 cycle proteins in *Arabidopsis thaliana*. Expression data was obtained from the eFP genome
456 browser[19] and correspond to values from the vegetative rosette stage.

References for Methods

1. McCormac, A. C. & Terry, M. J. Light-signalling pathways leading to the co-ordinated expression of HEMA1 and Lhcb during chloroplast development in Arabidopsis thaliana. *Plant J.* **32,** 549–559 (2002).

2. Pfaffl, M. W., Horgan, G. W. & Dempfle, L. Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* **30,** e36–e36 (2002).

3. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

4. Patro, R., Duggal, G. & Kingsford, C. Accurate, fast, and model-aware transcript expression quantification with Salmon. *bioRxiv* doi:10.1101/021592 (2015). doi:10.1101/021592

5. Hardcastle, T. J. & Kelly, K. A. baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* **11,** doi:10.1186/1471–2105–11–422 (2010).

6. Hardcastle, T. J. baySeq: Empirical Bayesian analysis of patterns of differential expression in count data. R package version 2.4.1. (2012).

7. Aubry, S., Kelly, S., Kümpers, B. M. C., Smith-Unna, R. D. & Hibberd, J. M. Deep evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in two independent origins of C4 photosynthesis. *PLOS Genet.* **10,** e1004365 (2014).

8. Carlson, M. org.At.tair.db: Genome wide annotation for Arabidopsis. (2015).

9. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25,** 25–29 (2000).

10. The Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43,** D1049–D1056 (2015).

11. Alexa, A. & Rahnenfuhrer, J. topGO: topGO: Enrichment analysis for Gene Ontology. (2010).

12. RStudio Team. RStudio: Integrated Development Environment for R. http://www.rstudio.com/ (2015)

13. Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer New York, 2009). at <http://had.co.nz/ggplot2/book>

14. Charron, J.-B. F., He, H., Elling, A. A. & Deng, X. W. Dynamic landscapes of four histone modifications during deetiolation in Arabidopsis. *Plant Cell* **21,** 3732–3748 (2009).

15. Chen, F. *et al.* Arabidopsis phytochrome A directly targets numerous promoters for individualized modulation of genes in a wide range of pathways. *Plant Cell* **26,** 1949–1966 (2014).

16. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings Bioinforma.* **14 ,** 178–192 (2013).

17. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29,** 24–26 (2011).

18. Covshoff, S. *et al.* Deregulation of maize C4 photosynthetic development in a mesophyll cell-defective mutant. *Plant Physiol* **146,** 1469–1481 (2008).

19. Winter, D. *et al.* An ‘Electronic Fluorescent Pictograph’ browser for exploring and analyzing

499 large-scale biological data sets. *PLoS One* **2,** e718 (2007).

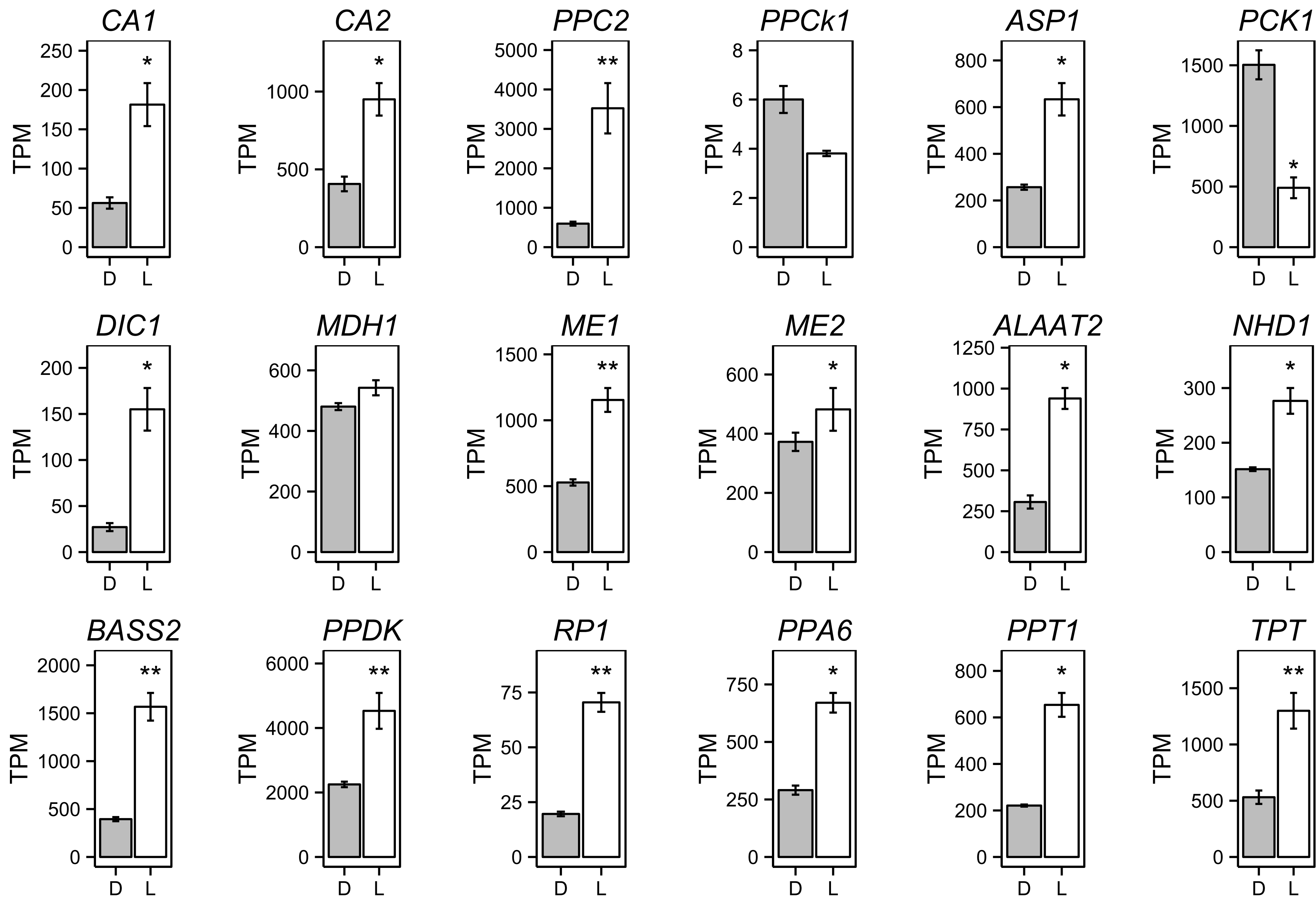
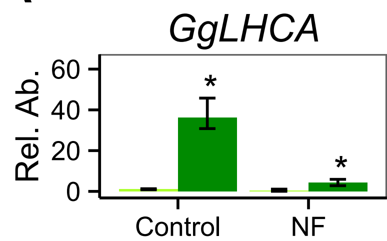
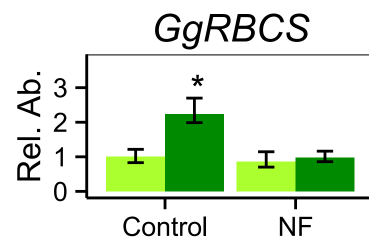


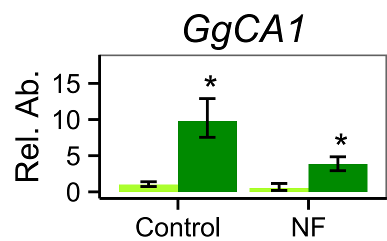
Figure 1

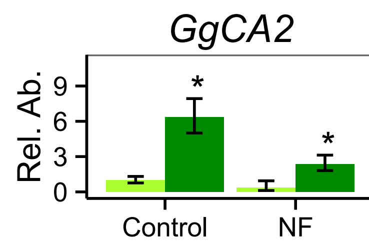
A

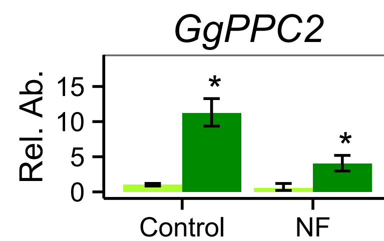


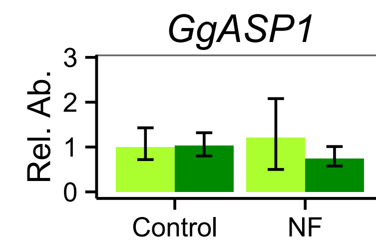


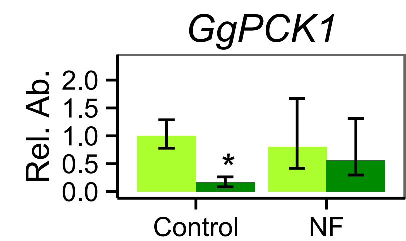
B

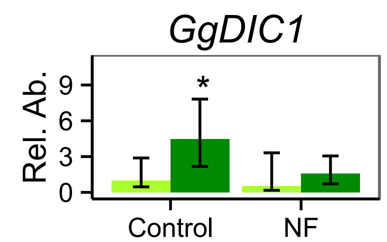


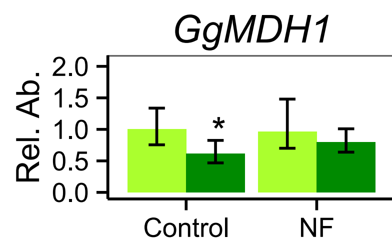


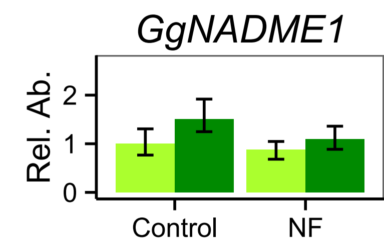


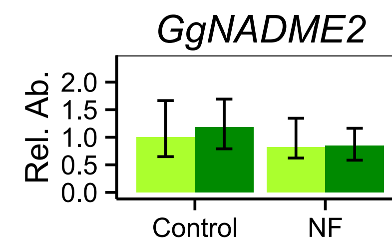


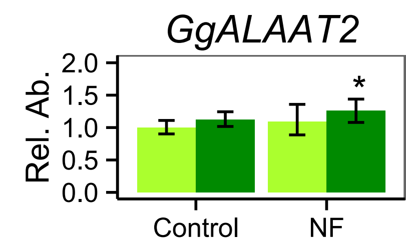


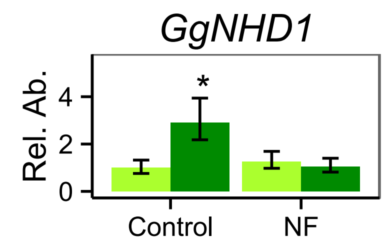


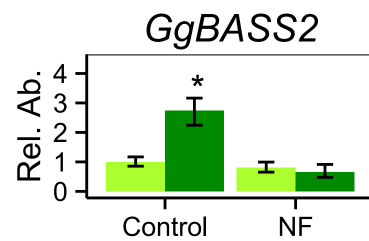


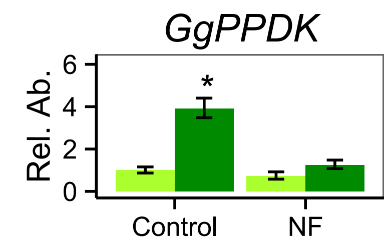


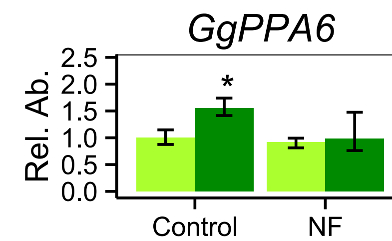


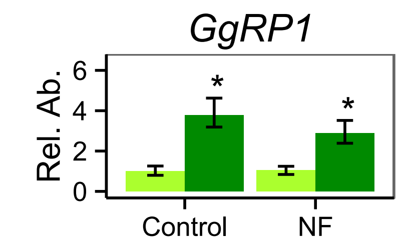


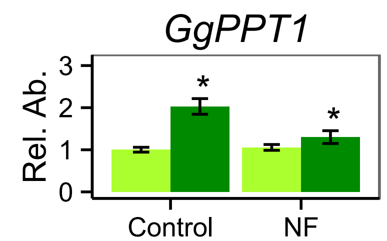












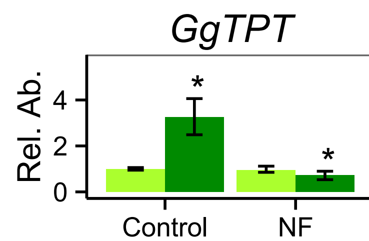
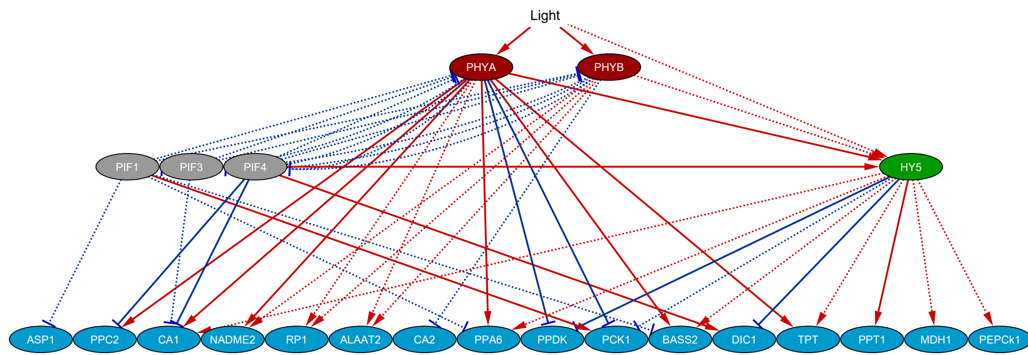


Figure 2

A



B

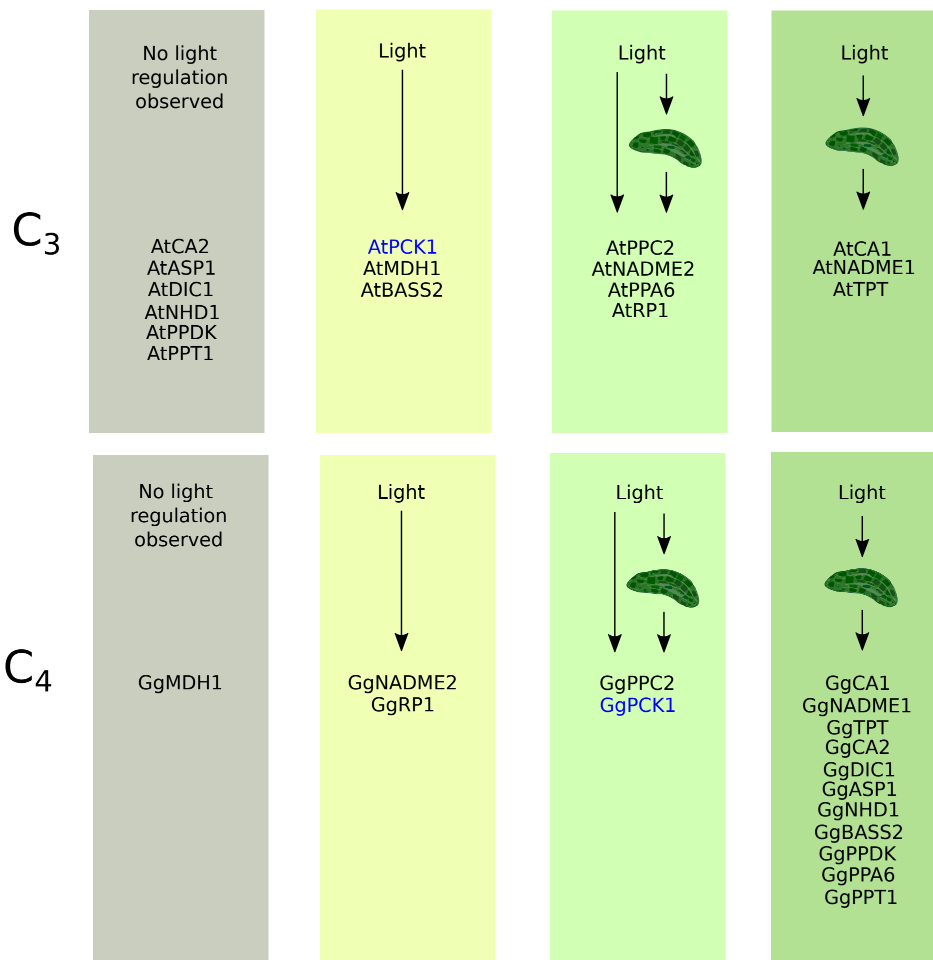


Figure 3