Ancestral light and chloroplast regulation form the foundations for C₄ gene expression

Steven J. Burgess, Ignasi Granero-Moya, Mathieu J. Grangé-Guermente, Chris Bournnell,
Matthew J. Terry¹ and Julian M. Hibberd*

Department of Plant Sciences, Downing Street, University of Cambridge, Cambridge CB2 3EA, UK

¹Biological Sciences, University of Southampton, Highfield Campus, Southampton, SO17 1BJ, UK

SJB - sjb287@cam.ac.uk

IG-M - ignasi.granero@e-campus.uab.cat

MJGG - mjg239@cam.ac.uk

CB - cmb211@cam.ac.uk

MJT - mjt@soton.ac.uk

JMH (corresponding) - jmh65@cam.ac.uk

Number of pages: 18

Number of Figures: 3

Number of Colour Figures: 3

Word Count: 1947

Number of Supplementary Figures: 6

Number of Supplementary Tables: 7

Introductory paragraph

C₄ photosynthesis acts as a carbon concentrating mechanism that leads to large increases in photosynthetic efficiency. The C₄ pathway is found in over sixty plant lineages¹ but the molecular enablers of this evolution are poorly understood. In particular, it is unclear how non-photosynthetic proteins in the ancestral C₃ system have repeatedly become strongly expressed and integrated into photosynthesis gene regulatory networks in C₄ leaves. Here, we provide clear evidence that in C₃ leaves, genes encoding key enzymes of the C₄ pathway are already co-regulated with photosynthesis genes and are controlled by both light and chloroplast-to-nucleus signalling. In C₄ leaves this regulation becomes increasingly dependent on the chloroplast. We propose that regulation of C₄ cycle genes by light and the chloroplast in the ancestral C₃ state has facilitated the repeated evolution of the complex and convergent C₄ trait.

Main

In C₃ leaves, photosynthesis genes are regulated directly by light-responsive networks² as well as by retrograde signalling from the chloroplast³. Although it has been proposed that during the evolution of the C₄ trait genes encoding components of the C₄ pathway are incorporated into photosynthesis networks⁴⁻⁶, and a small number of C₄ genes are known to respond to light⁷, how this comes about is unclear. For example, proteins used in the C₄ pathway fulfil a number of disparate roles in C₃ plants⁸, none of which are associated with the core photosynthetic process. It has therefore been puzzling how genes encoding these proteins become integrated into the gene regulatory networks that underpin photosynthesis during the transition from C₃ to C₄ metabolism. To address this we undertook a systems analysis of genes important for the C₄ pathway in the C₄ species *Gynandropsis gynandra* (formerly designated as *Cleome gynandra*)⁹. In order to construct a model of the evolutionary events leading to formation of the C₄ photosynthesis, we compared these results with analysis of genes from *A. thaliana*, which we used as a proxy for the ancestral C₃ state.

RNA-SEQ analysis of etiolated *G. gynandra* seedlings after 0 or 6 h of illumination identified 2671 transcripts as being differentially expressed between the two conditions (Table S1). Of these, 1754 were induced and 926 repressed after exposure to light. 54 genes appeared to undergo differential splicing, resulting in transcripts that fell into both up- and down-regulated datasets (Table S1). It was notable that Biological Process Gene Ontology (GO) terms^{10,11} representing photosynthesis and plastid organisation dominated the light-induced samples (Table S2), whilst those relating to auxin and brassinosteroid responses were enriched in the light-repressed dataset (Table S2). These data are consistent with analysis of the dark to light response in the C₃ model *A. thaliana* (Figure S1)¹² and support the notion that there is considerable conservation in the basic responses to light in these C₃ and C₄ species.

62

63 Notably, in C₄ *G. gynandra* 15 of 18 genes associated with the C₄ 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 C₄ cycle in *G. gynandra* are therefore regulated by networks that respond
68 either directly or indirectly to light, and a small number of C₄ 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 C₄ genes to light,
72 we conducted greening experiments with *G. gynandra* in the presence of Norflurazon (NF) which
73 inhibits carotenoid biosynthesis³ and therefore chloroplast development (Figure S2) resulting in the
74 reduced expression of many nuclear-encoded chloroplast genes³. Consistent with studies of C₃ *A.*
75 *thaliana*, light-induction of *G. gynandra* genes for *RIBULOSE BISPHOSPHATE CARBOXYLASE*
76 *OXYGENASE SMALL SUBUNIT (RBCS)* and *LIGHT HARVESTING COMPLEX A (LHCA)* was
77 chloroplast dependent (Figure 2A). Moreover, the response of twelve C₄ genes to light was
78 perturbed by NF indicating that a large proportion of the C₄ pathway in *G. gynandra* is regulated by
79 the chloroplast. Five C₄ 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 C₄ 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 C₄ 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

previously been linked to plastid retrograde signalling, also shows perturbed expression for many C_4 genes. Specifically, genes encoding *PEPC*, *CA*, *NADPME*, *NADP-MDH*, *PPDK*, *PCK* and *AspAT* all showed altered expression in the *Zmhcf136* mutant (Table S6)¹³. Combined with the report that that NF treatment of maize reduces transcript levels of *PPDK*¹⁴, these data support the notion that regulation by the chloroplast is important in other C_4 lineages. Overall, these data argue strongly for many genes of C_4 photosynthesis being subject to regulation by chloroplast signalling, light signalling or a combination of both.

We next sought to assess the extent to which genes orthologous to those recruited into C_4 photosynthesis are subject to light regulation in the ancestral C_3 state. Using publically available Chromatin Immuno-Precipitation (ChIP), microarray¹⁵ (Figure S4) and mutant data from C_3 *A. thaliana*, direct or indirect interactions were inferred based on whether or not mutant data were supported by ChIP-SEQ signals. This indicated that C_4 genes in *A. thaliana*, which are not used in photosynthesis, are subject to both direct and indirect regulation by phytochromes that mediate red light-signalling pathways¹⁶, but also by HY5 a master regulator of photomorphogenesis¹⁷, and by phytochrome interacting factors (PIFs) that act to repress gene expression in the dark¹⁸ (Figure 3A). To our knowledge, interactions between phyB and C_4 genes are only available from analysis of insertional mutants, whereas for phyA ChIP-SEQ data is also available (Figure 3A). For example, *AtTPT* is exclusively regulated by phyA (Table S5)¹⁹, whereas *AtCA2* is exclusively repressed by phyB (Table S5)²⁰. Six C_4 genes (*AtCA1*, *AtPPC2*, *AtBASS2*, *AtRP1*, *AtMDH1* and *AtPCK1*) are regulated by both phyA and phyB, as individual knockout of these phytochromes resulted in changes in gene expression (Table S5)^{19,20}. A further five genes (*AtPEPCK1*, *AtNADME2*, *AtALAAT2*, *AtPPDK* and *AtPPA6*) are under redundant regulation by phyA or phyB, being mis-expressed in *phyAphyB* double mutants²¹ (Table S5). Downstream of phytochromes, six C_4 genes are activated by HY5 (*AtPPT1*, *AtCA1*, *AtTPT*, *AtBASS2*, *AtPPA6* and *AtMDH1*) and four are repressed (*AtDIC1*, *AtPPDK*, *AtPEPCK1* and *AtPCK1*; Figure 3A and Table S5)^{17,22}. PIFs act antagonistically with HY5 by competitively binding E-box sequences²³. Seven C_4 genes (*AtASP1*, *AtBASS2*, *AtPPA6*, *AtPCK1*, *AtCA1*, *AtDIC1* and *AtPPC2*) are directly or indirectly regulated by either PIF1, 3 or 4 (Figure 3A)^{24–26}. Overall, these data indicate that 16 of the 18 *A. thaliana* genes orthologous to those involved in the C_4 cycle in *G. gynandra* are under direct or indirect regulation by light-signalling components (Figure 3A).

As many C_4 genes from *G. gynandra* showed chloroplast regulation (Figure 2) we investigated the extent to which C_4 orthologues from C_3 *A. thaliana* are regulated by retrograde signalling from the chloroplast. Consistent with previous reports²⁷, *AtLHCA* and *AtRBCS* were responsive to the chloroplast with NF largely abolishing their light induction (Figure S5A). Surprisingly, seven genes of the core C_4 cycle (*AtCA1*, *AtPPC2*, *AtNADME1*, *AtNADME2*, *AtPPA6*, *AtRP1* and *AtTPT*) were

either partially or completely dependent on signals from the chloroplast for the light-induction response (Figure S5B).

Most C_4 genes in *G. gynandra* and their orthologs in *A. thaliana* can be categorised into four groups depending on their response to illumination in the presence or absence of chloroplast signalling inhibitors (Figure 3B), and so we ordered the categories according to increasing influence of the chloroplast over this process. Although it is possible that the regulation of C_4 cycle genes in *A. thaliana* has diverged since the last common ancestor of *A. thaliana* and *G. gynandra*, we use the data to propose a model for the evolution of C_4 photosynthesis. The key features are first, that contrary to previous suggestions^{4,5} many C_4 genes appear to be operate within light regulatory networks in the C_3 state and second, that there is a degree of “fine-tuning” in terms of light and chloroplast regulation which may explain increases in gene expression during evolution of the C_4 pathway. Seven C_4 genes in *A. thaliana* responded to white light (Figure 3B), and there was also evidence that a further nine are integrated to some extent into light-signalling networks (Figure 3A). We further suggest that gain of light-activated chloroplast-dependent expression has occurred for eight genes (*GgCA2*, *GgASP1*, *GgPCK1*, *GgDIC1*, *GgNHD1*, *GgBASS2*, *GgPPDK*, and *GgPPT1*) in *G. gynandra*. (Figure S6). These data strongly imply that evolution of chloroplast function in C_4 leaves is underpinned by a shared molecular mechanism that is required for establishment of the C_4 carbon pump.

Ancestral regulation opens up the possibility that increases in C_4 gene expression could be driven by changes in transcription factor abundance, however we found no evidence for significant changes in phyA, phyB, HY5 or PIFs in available data for *G. gynandra*⁵. As many genes encoding components of the C_4 cycle are encoded by multi-gene families it is possible that ancestral regulation by either light or the chloroplast predisposes certain members of these gene families to recruitment into C_4 photosynthesis (Table S7). It is notable that in *A. thaliana* the *PPC* gene family contains four members, of which only one (*AtPPC2*) is light-responsive²⁷. It may be no coincidence that it is the ortholog of *AtPPC2* that was recruited into the C_4 pathway in *G. gynandra*. This situation would require a boosting of pre-existing signals rather than a gain of new ones, and would support the notion that some genes are primed for involvement in C_4 photosynthesis²⁸. However, this is unlikely to be the only important factor and we suggest that during the course of C_4 evolution, ancestral characteristics including sub-cellular location of proteins, transcript abundance, and light and chloroplast regulation, impact on the recruitment of specific genes from multi-gene families.

The C_4 pathway is commonly described as one of the most remarkable examples of convergent evolution in biology. Our finding that a large number of C_4 genes in the C_3 model *A. thaliana* are regulated by light and the chloroplast indicates that their complete integration into photosynthesis

networks in C₄ leaves is not as large a step as has previously been thought. If this is true for additional C₃ lineages that are sister to C₄ origins, it would help explain why C₄ photosynthesis has appeared in over sixty lineages of plants. It would also expand the extent to which the convergent evolution of this highly complex trait is based on parallel evolution at the molecular level^{6,29,30}.

Acknowledgements

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

Contributions of authors

SJB, MJT and JMH designed the study. SJB, IG-M and MJG-G carried out the experimental work. SJB and CB conducted the bioinformatics. SJB, MJT and JMH interpreted results and wrote the paper.

Materials and Methods

Plant Material and Growth Conditions

Both species were grown as described previously¹. 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. gynandra*. 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 software². 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 Trimmomatic³ and Salmon version 0.4.2⁴ 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 blast⁷. Genes were annotated with the database
230 org.At.tair.db⁸ and divided into functional categories using Gene Ontology (GO) terms^{9,10}.
231 Enrichment was calculated using the Bioconductor packages topGO¹¹. Plots were generated with
232 custom scripts in RStudio¹² using the package ggplot2¹³.

233

References

1. Sage, R. F., Christin, P.-A. & Edwards, E. J. The C₄ plant lineages of planet Earth. *J. Exp. Bot.* **62**, 3171–3181 (2011).
2. Jiao, Y., Lau, O. S. & Deng, X. W. Light-regulated transcriptional networks in higher plants. *Nat. Rev. Genet.* **8**, 217–230 (2007).
3. Chan, K. X., Phua, S. Y., Crisp, P., McQuinn, R. & Pogson, B. J. Learning the languages of the chloroplast: Retrograde signaling and beyond. *Annu. Rev. Plant Biol.* (2015). doi:10.1146/annurev-arplant-043015-111854
4. 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).
5. Kulahoglu, C. *et al.* Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae C₃ and C₄ plant species. *Plant Cell* **26**, 3243–3260 (2014).
6. Christin, P.-A. *et al.* Parallel recruitment of multiple genes into C₄ photosynthesis. *Genome Biol. Evol.* **5**, 2174–87 (2013).
7. Hibberd, J. M. & Covshoff, S. The regulation of gene expression required for C₄ photosynthesis. *Annu Rev Plant Biol* **61**, 181–207 (2010).
8. Eastmond, P. J. *et al.* Arabidopsis uses two gluconeogenic gateways for organic acids to fuel seedling establishment. *Nat. Commun.* **6**, (2015).
9. Marshall, D. M. *et al.* Cleome, a genus closely related to Arabidopsis, contains species spanning a developmental progression from C₃ to C₄ photosynthesis. *Plant J.* **51**, 886–896 (2007).
10. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
11. The Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015).
12. Ma, L. *et al.* Light control of Arabidopsis development entails coordinated regulation of genome expression and cellular pathways. *Plant Cell* **13**, 2589–2608 (2001).
13. Covshoff, S. *et al.* Dereglulation of maize C₄ photosynthetic development in a mesophyll cell-defective mutant. *Plant Physiol* **146**, 1469–1481 (2008).
14. Tamada, Y. *et al.* Effect of photooxidative destruction of chloroplasts on the expression of nuclear genes for C₄ photosynthesis and for chloroplast biogenesis in maize. *J. Plant Physiol.* **160**, 3–8 (2003).
15. 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).
16. Rockwell, N. C., Su, Y.-S. & Lagarias, J. C. Phytochrome Structure and Signaling 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

Figure 1: Light induction of C₄ genes in *G. gynandra*. RNA-SEQ data are plotted as Transcript Per Million (TPM). Transcript abundance is shown from samples collected from dark-grown seedlings (D) and after transfer to light (L) for 6 h. *CA1*, CARBONIC ANHYDRASE 1; *CA2*, CARBONIC ANHYDRASE 2; *PPC2*, PHOSPHOENOLPYRUVATE CARBOXYLASE 2; *PPCK1*, PHOSPHOENOLPYRUVATE CARBOXYLASE KINASE 1; *ASP1*, ASPARTATE AMINO TRANSFERASE 1; *PCK1*, PHOSPHOENOLPYRUVATE CARBOXYKINASE 1; *DIC1*, DICARBOXYLIC ACID TRANSPORTER 1; *MDH*, MITOCHONDRIAL MALATE DEHYDROGENASE; *NAD-ME1*, NAD-DEPENDENT MALIC ENZYME 1; *NAD-ME2*, NAD-DEPENDENT MALIC ENZYME 2; *ALAAT2*, ALANINE AMINO TRANSFERASE 2; *NHD1*, SODIUM:HYDROGEN ANTIporter 1; *BASS2*, BILE SODIUM ACID TRANSPORTER 2; *PPDK*, PYRUVATE, ORTHOPHOSPHATE DIKINASE; *RP1*, PPDK REGULATORY PROTEIN 1; *PPA6*, PYROPHOSPHORLYASE 6; *PPT1*, PHOSPHOENOLPYRUVATE TRANSPORTER 1; *TPT*, TRIOSE PHOSPHATE TRANSPORTER. Data are shown as means and standard errors of three biological replicates, the calculated significance values for differential expression are indicated: FDR 0.05-0.01 (*) and FDR<0.01 (**).

Figure 2: Chloroplast regulation of C₄ genes in *G. gynandra*. Quantitative real time polymerase chain reactions were used to determine the impact of light and chloroplast signalling on C₄ transcript abundance in *G. gynandra*. Data are derived from samples collected from dark-grown seedlings ± 50 µM Norflurazon (NF) before (light green) and after transfer to light (dark green) for 24 h. **A.** *LHCA*, LIGHT HARVESTING COMPLEX A; *RBCS*, RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE SMALL SUBUNIT 1A; **B.** *CA1*, CARBONIC ANHYDRASE 1; *CA2*, CARBONIC ANHYDRASE 2; *PPC2*, PHOSPHOENOLPYRUVATE CARBOXYLASE 2; *ASP1*, ASPARTATE AMINO TRANSFERASE 1; *PCK1*, PHOSPHOENOLPYRUVATE CARBOXYKINASE 1; *DIC1*, DICARBOXYLIC ACID TRANSPORTER 1; *MDH*, MITOCHONDRIAL MALATE DEHYDROGENASE; *NAD-ME1*, NAD-DEPENDENT MALIC ENZYME 1; *NAD-ME2*, NAD-DEPENDENT MALIC ENZYME 2; *ALAAT2*, ALANINE AMINO TRANSFERASE 2; *NHD1*, SODIUM:HYDROGEN ANTIporter 1; *BASS2*, BILE SODIUM ACID TRANSPORTER 2; *PPDK*, PYRUVATE, ORTHOPHOSPHATE DIKINASE; *RP1*, PPDK REGULATORY PROTEIN 1; *PPA6*, PYROPHOSPHORLYASE 6; *PPT1*, PHOSPHOENOLPYRUVATE TRANSPORTER 1; *TPT*, TRIOSE PHOSPHATE TRANSPORTER. Data are shown as means with error bars representing upper and lower bounds of a 95% confidence interval from three biological replicates. Differential expression (p<0.1) compared to 0 h without NF is indicated by (*).

Figure 3: Proposed model of changes in light regulation during evolution of the C₄ pathway in *G. gynandra*. **A.** Composite model of the regulation of C₄ orthologs in *A. thaliana* by known light signalling components based on available chromatin immunoprecipitation datasets and analysis of 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; PPKK,
 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 C₄ genes in *G. gynandra* (C₄) and orthologous genes in *A. thaliana* (C₃). 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).

Supplementary Figure Legends

Figure S1: Light induction of C₄ orthologs in *A. thaliana*. Microarray data are derived from Charron et al. (2009)¹⁴. Log fold change in transcript abundance is shown comparing samples collected before and after 6 h WL light treatment. *ASP1*, ASPARTATE AMINOTRANSFERASE 1; *BASS2*, BILE SODIUM ACID TRANSPORTER 2; *CA1*, CARBONIC ANHYDRASE 1; *CA2*, CARBONIC ANHYDRASE 2; *DIC1*, DICARBOXYLIC ACID TRANSPORTER 1; *DTC1*, DICARBOXYLATE/TRICARBOXYLATE CARRIER; *MDH*, MITOCHONDRIAL MALATE DEHYDROGENASE; *NAD-ME1*, NAD-DEPENDENT MALIC ENZYME 1; *NAD-ME2*, NAD-DEPENDENT MALIC ENZYME 2; *NHD1*, SODIUM:HYDROGEN ANTIporter 1; *PCK1*, PHOSPHOENOLPYRUVATE CARBOXYKINASE 1; *PPCK1*, PHOSPHOENOLPYRUVATE CARBOXYLASE KINASE 1; *PPA6*, PYROPHOSPHORLYASE 6; *PPC2*, PHOSPHOENOLPYRUVATE CARBOXYLASE 2; *PPDK*, PYRUVATE, ORTHOPHOSPHATE DIKINASE; *RP1*, *PPDK* REGULATORY PROTEIN 1; *TPT*, TRIOSE PHOSPHATE TRANSPORTER.

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

Figure S3: Chloroplast regulation of C₄ genes in *G. gynandra*. Quantitative real time polymerase chain reactions were used to determine the impact of light and chloroplast signalling on C₄ transcript abundance in *G. gynandra*. Data are derived from samples collected from dark-grown seedlings $\pm$ 0.5 mM Lincomycin before (light green) and after transfer to light (dark green) for 24 h. **A.** *LHCA*, LIGHT HARVESTING COMPLEX A; *RBCS*, RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE SMALL SUBUNIT 1A; **B.** *CA1*, CARBONIC ANHYDRASE 1; *CA2*, CARBONIC ANHYDRASE 2; *PPC2*, PHOSPHOENOLPYRUVATE CARBOXYLASE 2; *ASP1*, ASPARTATE AMINO TRANSFERASE 1; *PCK1*, PHOSPHOENOLPYRUVATE CARBOXYKINASE 1; *DIC1*, DICARBOXYLIC ACID TRANSPORTER 1; *MDH*, MITOCHONDRIAL MALATE DEHYDROGENASE; *NAD-ME1*, NAD-DEPENDENT MALIC ENZYME 1; *NAD-ME2*, NAD-DEPENDENT MALIC ENZYME 2; *ALAAT2*, ALANINE AMINO TRANSFERASE 2; *NHD1*, SODIUM:HYDROGEN ANTIporter 1; *BASS2*, BILE SODIUM ACID TRANSPORTER 2; *PPDK*, PYRUVATE, ORTHOPHOSPHATE DIKINASE; *RP1*, *PPDK* REGULATORY PROTEIN 1; *PPA6*, PYROPHOSPHORLYASE 6; *PPT1*, PHOSPHOENOLPYRUVATE TRANSPORTER 1; *TPT*, TRIOSE PHOSPHATE TRANSPORTER. Data are shown as means with error bars representing standard error from five biological replicates. Differential expression ($p < 0.1$) compared to 0 h without NF is indicated by (*).

Figure S4: Location of phyA ChIP-SEQ binding sites over C₄ genes **A.** *AtCA1* **B.** *AtPPC2* **C.** *AtBASS2*. Data was obtained from¹⁵ and visualized using the Integrative Genome Viewer (IGV)^{16,17}.

Figure S5: Chloroplast regulation of C₄ genes in *A. thaliana*. Quantitative real time polymerase chain reactions were used to determine the impact of light and chloroplast signalling on C₄ transcript abundance in *A. thaliana*. Data are derived from samples collected from dark-grown seedlings ± 5 µM Norflurazon (NF) before (light green) and after transfer to light (dark green) for 24 h. **A.** *LHCA*, *LIGHT HARVESTING COMPLEX A*; *RBCS*, *RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE SMALL SUBUNIT 1A*; **B.** *CA1*, *CARBONIC ANHYDRASE 1*; *CA2*, *CARBONIC ANHYDRASE 2*; *PPC2*, *PHOSPHOENOLPYRUVATE CARBOXYLASE 2*; *ASP1*, *ASPARTATE AMINO TRANSFERASE 1*; *PCK1*, *PHOSPHOENOLPYRUVATE CARBOXYKINASE 1*; *DIC1*, *DICARBOXYLIC ACID TRANSPORTER 1*; *MDH*, *MITOCHONDRIAL MALATE DEHYDROGENASE*; *NAD-ME1*, *NAD-DEPENDENT MALIC ENZYME 1*; *NAD-ME2*, *NAD-DEPENDENT MALIC ENZYME 2*; *ALAAT2*, *ALANINE AMINO TRANSFERASE 2*; *NHD1*, *SODIUM:HYDROGEN ANTIporter 1*; *BASS2*, *BILE SODIUM ACID TRANSPORTER 2*; *PPDK*, *PYRUVATE, ORTHOPHOSPHATE DIKINASE*; *RP1*; *PPDK REGULATORY PROTEIN 1*; *PPA6*, *PYROPHOSPHORLYASE 6*; *PPT1*, *PHOSPHOENOLPYRUVATE TRANSPORTER 1*; *TPT*, *TRIOSE PHOSPHATE TRANSPORTER*. Data are shown as means with error bars representing upper and lower bounds of a 95% confidence interval from three biological replicates. Differential expression (p<0.1) compared to 0 h without NF is indicated by (*).

Figure S6: Gain of chloroplast-dependent C₄ gene expression in *G. gynandra*. Schematic illustrating the extent of light-activated, chloroplast-dependent C₄ gene expression in the ancestral C₃ state, based on data from *A. thaliana*, as well the C₄ state based on data from *G. gynandra*.

Supplementary Table Legends

Table S1: Summary of RNA-Seq analysis comparing samples extracted from 7 day old, etiolated *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

Table S2: Top ten GO terms enriched in the *G. gynandra* light-induced and light-repressed gene 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
<i>Light induced</i>			
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	$6.30E-19$
<i>Light repressed</i>			
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

Table S3: Expression Data from RNA-SEQ analysis of seven day old, etiolated *G. gynandra* seedlings before and after 6 hours of white light treatment. The Arabidopsis Genome Initiative (AGI) identifier, locus and transcript identifiers from the *G. gynandra* draft genome (unpublished) are provided alongside expression values as counts and statistical outputs from baySEQ^{5,6}.

Table S4: Expression Data from RNA-SEQ analysis of seven day old, etiolated *G. gynandra* seedlings before and after 6 hours of white light treatment. The Arabidopsis Genome Initiative (AGI) identifier, locus and transcript identifiers from the *G. gynandra* draft genome (unpublished) are provided alongside expression values as TPMs.

Table S5: Meta-analysis of available gene expression and ChIP-SEQ datasets relating to *Arabidopsis thaliana* light signalling components.

Table S6: Abundance of C₄ genes in mesophyll and bundle sheath cells comparing wild type and the *Zmhcf136* mutant. Expression data are taken from¹⁸.

Table S7: Abundance of members of multigene families containing enzymes orthologous to core C₄ cycle proteins in *Arabidopsis thaliana*. Expression data was obtained from the eFP genome browser¹⁹ and correspond to values from the vegetative rosette stage.

457 References for Methods

- 458 1. McCormac, A. C. & Terry, M. J. Light-signalling pathways leading to the co-ordinated
459 expression of HEMA1 and Lhcb during chloroplast development in *Arabidopsis thaliana*.
460 *Plant J.* **32**, 549–559 (2002).
- 461 2. Pfaffl, M. W., Horgan, G. W. & Dempfle, L. Relative expression software tool (REST©) for
462 group-wise comparison and statistical analysis of relative expression results in real-time
463 PCR. *Nucleic Acids Res.* **30**, e36–e36 (2002).
- 464 3. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence
465 data. *Bioinformatics* **30**, 2114–2120 (2014).
- 466 4. Patro, R., Duggal, G. & Kingsford, C. Accurate, fast, and model-aware transcript expression
467 quantification with Salmon. *bioRxiv* doi:10.1101/021592 (2015). doi:10.1101/021592
- 468 5. Hardcastle, T. J. & Kelly, K. A. baySeq: Empirical Bayesian methods for identifying
469 differential expression in sequence count data. *BMC Bioinformatics* **11**, doi:10.1186/1471–
470 2105–11–422 (2010).
- 471 6. Hardcastle, T. J. baySeq: Empirical Bayesian analysis of patterns of differential expression
472 in count data. R package version 2.4.1. (2012).
- 473 7. Aubry, S., Kelly, S., Kümpers, B. M. C., Smith-Unna, R. D. & Hibberd, J. M. Deep
474 evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in
475 two independent origins of C4 photosynthesis. *PLOS Genet.* **10**, e1004365 (2014).
- 476 8. Carlson, M. org.At.tair.db: Genome wide annotation for Arabidopsis. (2015).
- 477 9. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29
478 (2000).
- 479 10. The Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids*
480 *Res.* **43**, D1049–D1056 (2015).
- 481 11. Alexa, A. & Rahnenfuhrer, J. topGO: topGO: Enrichment analysis for Gene Ontology.
482 (2010).
- 483 12. RStudio Team. RStudio: Integrated Development Environment for R.
484 <http://www.rstudio.com/> (2015)
- 485 13. Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer New York, 2009). at
486 <<http://had.co.nz/ggplot2/book>>
- 487 14. Charron, J.-B. F., He, H., Elling, A. A. & Deng, X. W. Dynamic landscapes of four histone
488 modifications during deetiolation in Arabidopsis. *Plant Cell* **21**, 3732–3748 (2009).
- 489 15. Chen, F. *et al.* Arabidopsis phytochrome A directly targets numerous promoters for
490 individualized modulation of genes in a wide range of pathways. *Plant Cell* **26**, 1949–1966
491 (2014).
- 492 16. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV):
493 high-performance genomics data visualization and exploration. *Briefings Bioinforma.* **14**,
494 178–192 (2013).
- 495 17. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
- 496 18. Covshoff, S. *et al.* Dereglulation of maize C4 photosynthetic development in a mesophyll
497 cell-defective mutant. *Plant Physiol* **146**, 1469–1481 (2008).
- 498 19. Winter, D. *et al.* An ‘Electronic Fluorescent Pictograph’ browser for exploring and analyzing

Figure 1

A **B** **Figure 2**

A

B

Figure 3