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

Steven J. Burgess, Ignasi Granero-Moya, Mathieu J. Grangé-Guermente, Chris Boursnell, Matthew J. Terry$^1$ and Julian M. Hibberd$^*$

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

$^1$Biological 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) terms¹⁰,¹¹ 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.
Notably, in C₄ G. gynandra 15 of 18 genes associated with the C₄ pathway showed increased transcript abundance after exposure to light (Figure 1; Table S3). Of these, transcripts from PHOSPHOENOLPYRUVATE CARBOXYLASE (GgPPC2) and PYRUVATE,ORTHOPHOSPHATE DIKINASE (GgPPDK) were amongst the ten most abundant in the dataset (Table S4). The majority of genes involved in the C₄ cycle in G. gynandra are therefore regulated by networks that respond either directly or indirectly to light, and a small number of C₄ genes are as responsive to light as those of the light-dependent reactions of photosynthesis.

To quantify the importance of chloroplast-to-nucleus signals for the response of C₄ genes to light, we conducted greening experiments with G. gynandra in the presence of Norflurazon (NF) which inhibits carotenoid biosynthesis and therefore chloroplast development (Figure S2) resulting in the reduced expression of many nuclear-encoded chloroplast genes. Consistent with studies of C₃ A. thaliana, light-induction of G. gynandra genes for RIBULOSE BISPHOSPHATE CARBOXYLASE OXYGENASE SMALL SUBNUT (RBCS) and LIGHT HARVESTING COMPLEX A (LHCA) was chloroplast dependent (Figure 2A). Moreover, the response of twelve C₄ genes to light was perturbed by NF indicating that a large proportion of the C₄ pathway in G. gynandra is regulated by the chloroplast. Five C₄ genes (DICARBOXYLIC ACID TRANSPORTER 1 (GgDIC1), SODIUM:HYDROGEN ANTIPORTER 1 (GgNHD1), BILE SODIUM ACID TRANSPORTER 2 (GgBASS2), GgPPDK and PYROPHOSPHORLYASE 6 (GgPPA6)) were almost entirely dependent on the chloroplast, as NF abolished increases in transcript abundance in response to light (Figure 2B). Partial dependence on the chloroplast for light-induced expression was also detected for four genes (BETA CARBONIC ANHYDRASE 1, BETA CARBONIC ANHYDRASE 2, GgPPC2, and PHOSPHOENOLPYRUVATE TRANSPORTER 1 (GgPPT1)), which showed a reduced response to light in the presence of NF (Figure 2B). In addition, other C₄ genes showed more complex behaviours consistent with the operation of plastid signals. This included GgTPT, whose transcripts increased on exposure to light, but decreased in the presence of NF (Figure 2B) suggesting regulation is not only dependent on plastid signals in the light, but also on plastid signals that operate independently of a light requirement. Furthermore, MALATE DEHYDROGENASE (GgMDH) and PHOSPHOENOLPYRUVATE CARBOXYKINASE (GgPCK1) were less abundant after light-induction, but this effect was reduced in the presence of NF, indicating the process is chloroplast-dependent (Figure 2B). As NF is known to lead to plastid damage, we sought additional lines of evidence that link C₄ gene expression with plastid to nucleus signalling. Lincomycin (Lin), which inhibits plastid translation, led to results that were very consistent with those obtained from NF, with 11 of the 12 genes perturbed by NF also showing a clear response to Lin (Figure S3). In addition, Lin also affected expression of NADME1 and ASP1 and for some genes (CA1, CA2, PPC2, PPT1) demonstrated a more complete dependence on the 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 C₄ 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 PPDK¹⁴, these data support the notion that regulation by the chloroplast is important in other C₄ lineages. Overall, these data argue strongly for many genes of C₄ 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 C₄ photosynthesis are subject to light regulation in the ancestral C₃ state. Using publically available Chromatin Immuno-Precipitation (ChIP), microarray¹⁵ (Figure S4) and mutant data from C₃ A. thaliana, direct or indirect interactions were inferred based on whether or not mutant data were supported by ChIP-SEQ signals. This indicated that C₄ 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 C₄ 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 C₄ 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 C₄ 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 C₄ 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 C₄ cycle in G. gynandra are under direct or indirect regulation by light-signalling components (Figure 3A).

As many C₄ genes from G. gynandra showed chloroplast regulation (Figure 2) we investigated the extent to which C₄ orthologues from C₃ 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 C₄ 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 C₄ 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 C₄ 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 C₄ photosynthesis. The key features are first, that contrary to previous suggestions⁴,⁵ many C₄ genes appear to operate within light regulatory networks in the C₃ 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 C₄ pathway. Seven C₄ 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 C₄ leaves is underpinned by a shared molecular mechanism that is required for establishment of the C₄ carbon pump.

Ancestral regulation opens up the possibility that increases in C₄ 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. gynandra⁵. As many genes encoding components of the C₄ 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 C₄ 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 C₄ 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 C₄ photosynthesis²⁸. However, this is unlikely to be the only important factor and we suggest that during the course of C₄ 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 C₄ pathway is commonly described as one of the most remarkable examples of convergent evolution in biology. Our finding that a large number of C₄ genes in the C₃ model A. thaliana are regulated by light and the chloroplast indicates that their complete integration into photosynthesis
networks in $\text{C}_4$ leaves is not as large a step as has previously been thought. If this is true for 
additional $\text{C}_3$ lineages that are sister to $\text{C}_4$ origins, it would help explain why $\text{C}_4$ 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 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 \(\mu\)M (*A. thaliana*) or 50 \(\mu\)M (*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 \(\mu\)L, *A. thaliana*; 50 \(\mu\)L, *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 \(\mu\)mol 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 \(\mu\)g 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 \(\mu\)g 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*
transcripts were annotated with respect to orthologous genes from *A. thaliana* by mapping to the *A. thaliana* genome using reciprocal best blast. Genes were annotated with the database org.At.tair.db and divided into functional categories using Gene Ontology (GO) terms. Enrichment was calculated using the Bioconductor packages topGO. Plots were generated with custom scripts in RStudio using the package ggplot2.
References


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, PYROPHOSPHORYLASE 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, PYROPHOSPHORYLASE 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
ANHYDRASE 1; PPC2, PHOSPHOENOLPYRUVATE CARBOXYLASE 2; TPT, TRIOSE PHOSPHATE TRANSPORTER, BASS2, BILE SODIUM ACID TRANSPORTER 2; CA2, CARBONIC ANHYDRASE 2; PPCK1, PHOSPHOENOLPYRUVATE CARBOXYLASE KINASE 1; NAD-ME2, NAD-DEPENDENT MALIC ENZYME 2; PPA6, PYROPHOSPHORYLASE 6; PPDK, PYRUVATE,ORTHOPHOSPHATE DIKINASE; HY5, ELONGATED HYPOCOTYL 5; PPT1, PHOSPHOENOLPYRUVATE TRANSPORTER 1; DIC1, DICARBOXYLIC ACID TRANSPORTER 1; PIF4, PHYTOCHROME INTERACTING FACTOR 4. Direct and indirect repression are represented by solid and dashed blue lines respectively, whereas direct and indirect activation are represented by solid and dashed red arrows respectively. B. Diagram illustrating the proposed regulation of C₄ genes in G. gynandra (C₄) and orthologous genes in A. thaliana (C₃). Categories include genes unresponsive to light or chloroplast signalling, light signalling alone, light and chloroplast signalling or signalling entirely dependent on the chloroplast (which includes genes for which light activation is lost after chloroplast damage, as well as when further inhibition in the absence of light is observed, such as after Lin treatment in the dark). The groupings are based on genes showing statistically significant differences in gene expression (p<0.1) in the presence or absence of NF and Lin during a dark to light transition (Figure 2, Figure S3, Figure S5). For G. gynandra genes, the category chosen was based on the strongest response to either NF or Lin provided a similar response was observed in the other treatment. ALAAT2 is omitted from the figure as its expression profile did not fit the defined profiles for light or chloroplast regulation (Figure 2, Figure S3, Figure S5).
Supplementary Figure Legends

Figure S1: Light induction of C_4 orthologs in A. thaliana. Microarray data are derived from Charron et al. (2009). 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 DEHYDRGENASE; 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, PYROPHOSPHORYLASE 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 µM). 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 C_4 genes in G. gynandra. Quantitative real time polymerase chain reactions were used to determine the impact of light and chloroplast signalling on C_4 transcript abundance in G. gynandra. Data are derived from samples collected from dark-grown seedlings ± 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 DEHYDRGENASE; 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, PYROPHOSPHORYLASE 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.

<table>
<thead>
<tr>
<th>Replicates</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of PE reads</td>
<td>127,543,218</td>
</tr>
<tr>
<td>Average reads per library</td>
<td>21,257,203</td>
</tr>
<tr>
<td>Detected genes</td>
<td>26945</td>
</tr>
<tr>
<td>Detected Transcripts</td>
<td>57296</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. of genes DE (FDR&lt;0.05)</th>
<th>Induced</th>
<th>Repressed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of transcripts DE (FDR&lt;0.05)</td>
<td>1597</td>
<td>887</td>
<td>2433</td>
</tr>
<tr>
<td>No. of transcripts DE (FDR&lt;0.05)</td>
<td>1767</td>
<td>940</td>
<td>2707</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>GO ID</th>
<th>GO Term</th>
<th>Genes Annotated</th>
<th>Fisher p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Light induced</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0015979</td>
<td>Photosynthesis</td>
<td>333</td>
<td>&lt; 1e-30</td>
</tr>
<tr>
<td>GO:0019684</td>
<td>photosynthesis, light reaction</td>
<td>253</td>
<td>&lt; 1e-30</td>
</tr>
<tr>
<td>GO:0009657</td>
<td>plastid organization</td>
<td>346</td>
<td>5.10E-30</td>
</tr>
<tr>
<td>GO:0009668</td>
<td>plastid membrane organization</td>
<td>181</td>
<td>1.20E-21</td>
</tr>
<tr>
<td>GO:0010027</td>
<td>thylakoid membrane organization</td>
<td>181</td>
<td>1.20E-21</td>
</tr>
<tr>
<td>GO:0019288</td>
<td>isopentenyl diphosphate biosynthetic process</td>
<td>204</td>
<td>2.30E-19</td>
</tr>
<tr>
<td>GO:0019682</td>
<td>glyceraldehyde-3-phosphate metabolic process</td>
<td>205</td>
<td>3.30E-19</td>
</tr>
<tr>
<td>GO:0006090</td>
<td>pyruvate metabolic process</td>
<td>207</td>
<td>6.30E-19</td>
</tr>
<tr>
<td>GO:009240</td>
<td>isopentenyl diphosphate biosynthetic process</td>
<td>207</td>
<td>6.30E-19</td>
</tr>
<tr>
<td>GO:0046490</td>
<td>isopentenyl diphosphate metabolic process</td>
<td>207</td>
<td>6.30E-19</td>
</tr>
<tr>
<td><strong>Light repressed</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0009825</td>
<td>multidimensional cell growth</td>
<td>87</td>
<td>0.00012</td>
</tr>
<tr>
<td>GO:0009741</td>
<td>response to brassinosteroid</td>
<td>83</td>
<td>0.00166</td>
</tr>
<tr>
<td>GO:009733</td>
<td>response to auxin</td>
<td>272</td>
<td>0.0035</td>
</tr>
<tr>
<td>GO:0046685</td>
<td>response to arsenic-containing substance</td>
<td>26</td>
<td>0.00516</td>
</tr>
<tr>
<td>GO:0052482</td>
<td>defense response by cell wall thickening</td>
<td>15</td>
<td>0.00666</td>
</tr>
<tr>
<td>GO:0052544</td>
<td>defense response by callose deposition in cell wall</td>
<td>15</td>
<td>0.00666</td>
</tr>
<tr>
<td>GO:0048443</td>
<td>stamen development</td>
<td>114</td>
<td>0.00828</td>
</tr>
<tr>
<td>GO:0048466</td>
<td>androecium development</td>
<td>114</td>
<td>0.00828</td>
</tr>
<tr>
<td>GO:0042547</td>
<td>cell wall modification involved in multidimensional cell growth</td>
<td>16</td>
<td>0.00901</td>
</tr>
<tr>
<td>GO:0009958</td>
<td>positive gravitropism</td>
<td>17</td>
<td>0.01187</td>
</tr>
</tbody>
</table>
Table S3: Expression Data from RNA-SEQ analysis of seven day old, etiolated *G. gynanda* 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 C\(_4\) genes in mesophyll and bundle sheath cells comparing wild type and the Zmhcf136 mutant. Expression data are taken from\(^18\).

Table S7: Abundance of members of multigene families containing enzymes orthologous to core C\(_4\) cycle proteins in *Arabidopsis thaliana*. Expression data was obtained from the eFP genome browser\(^19\) and correspond to values from the vegetative rosette stage.
References for Methods


Figure 1
Figure 3