

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

Supplementary Information File Contents

Figure S1: Light induction of C₄ orthologs in *A. thaliana*.

Figure S2: *G. gynandra* seedlings grown in the presence of NF.

Figure S3: Chloroplast regulation of C₄ genes in *G. gynandra*.

Figure S4: Location of phyA ChIP-SEQ binding sites over C₄ genes.

Figure S5: Chloroplast regulation of C₄ genes in *A. thaliana*.

Figure S6: Gain of chloroplast-dependent C₄ gene expression in *G. gynandra*.

Table S1: Summary statistics from RNA-Seq analysis.

Table S2: Top ten GO terms enriched in the *G. gynandra* light-induced and light-repressed gene lists (p<0.05).

Table S3-S7 Legends

Supplementary References

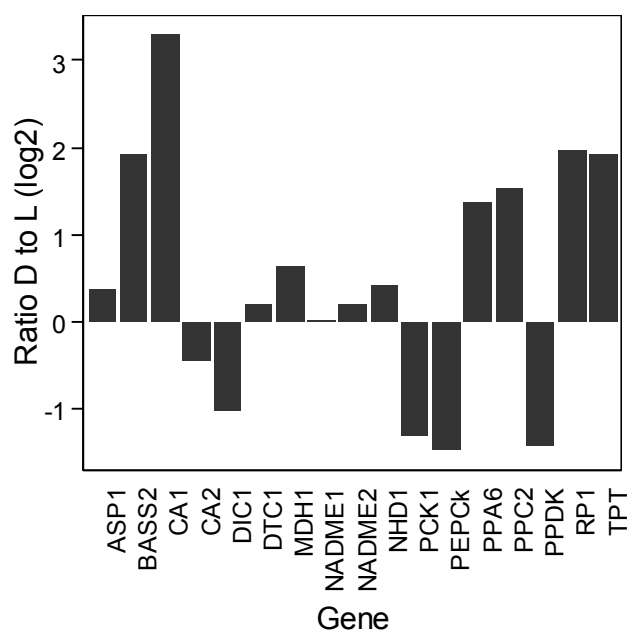


Figure S1: Light induction of C_4 orthologs in *A. thaliana*. Microarray data are derived from Charron et al. (2009)¹⁴. Log fold change in transcript abundance is shown comparing samples collected before and after 6 h WL light treatment. ASP1, ASPARTATE AMINOTRANSFERASE 1; BASS2, BILE SODIUM ACID TRANSPORTER 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 ANTIPTORTER 1; PCK1, PHOSPHOENOLPYRUVATE CARBOXYKINASE 1; PPCK1, PHOSPHOENOLPYRUVATE CARBOXYLASE KINASE 1; PPA6, PYROPHOSPHORLYASE 6; PPC2, PHOSPHOENOLPYRUVATE CARBOXYLASE 2; PPK, PYRUVATE, ORTHOPHOSPHATE DIKINASE; RP1, PPK REGULATORY PROTEIN 1; TPT, TRIOSE PHOSPHATE TRANSPORTER.

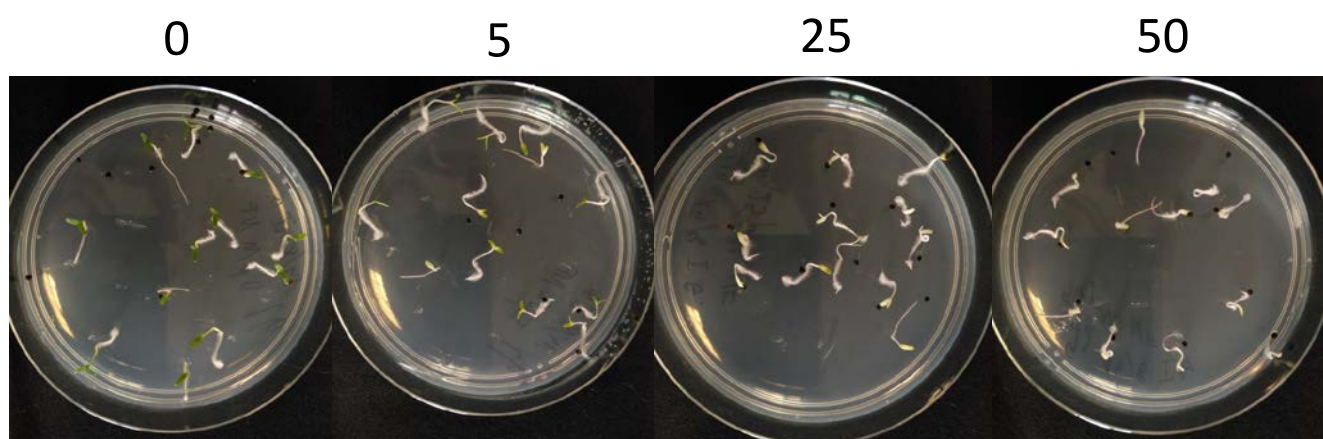


Figure S2: Photographs of *G. gynandra* seedlings grown in the presence of varying concentrations of NF (values presented above image with the units μM). Representative images are derived from samples collected from 3 day-old dark-grown seedlings after transfer to light for 24h

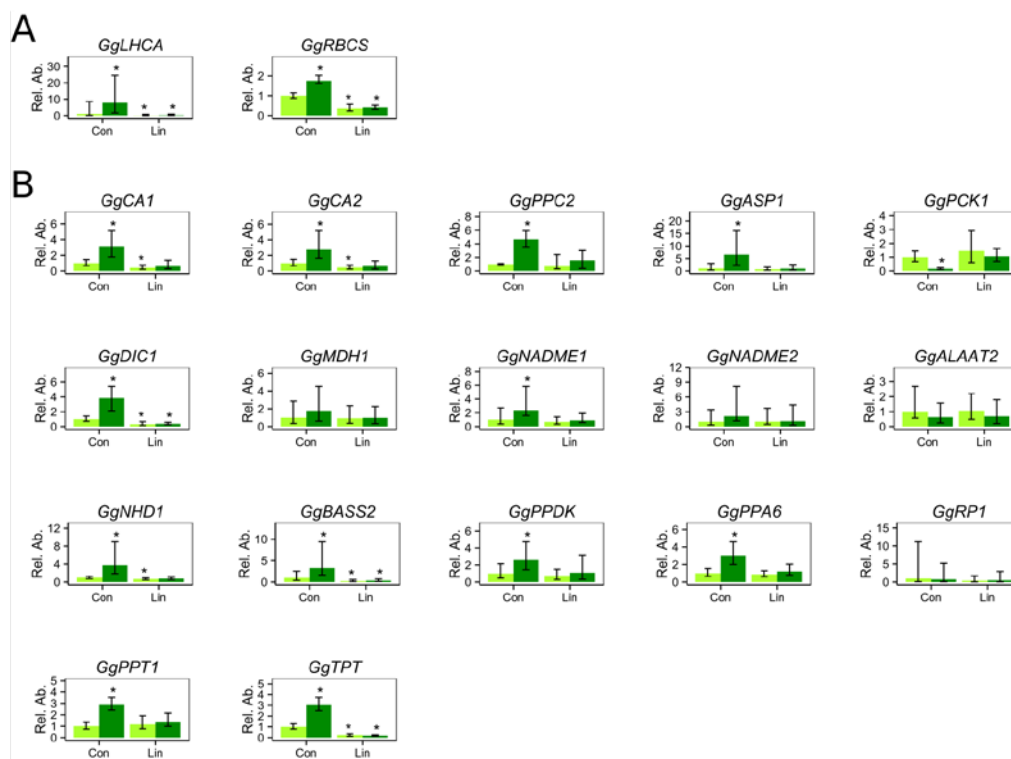


Figure S3: Chloroplast regulation of *C₄* genes in *G. gynandra*. Quantitative real time polymerase chain reactions were used to determine the impact of light and chloroplast signalling on *C₄* transcript abundance in *G. gynandra*. Data are derived from samples collected from dark-grown seedlings $\pm$ 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 ANTIPOINTER 1; BASS2, BILE SODIUM ACID TRANSPORTER 2; PPK, PYRUVATE, ORTHOPHOSPHATE DIKINASE; RP1, PPK 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 (*).

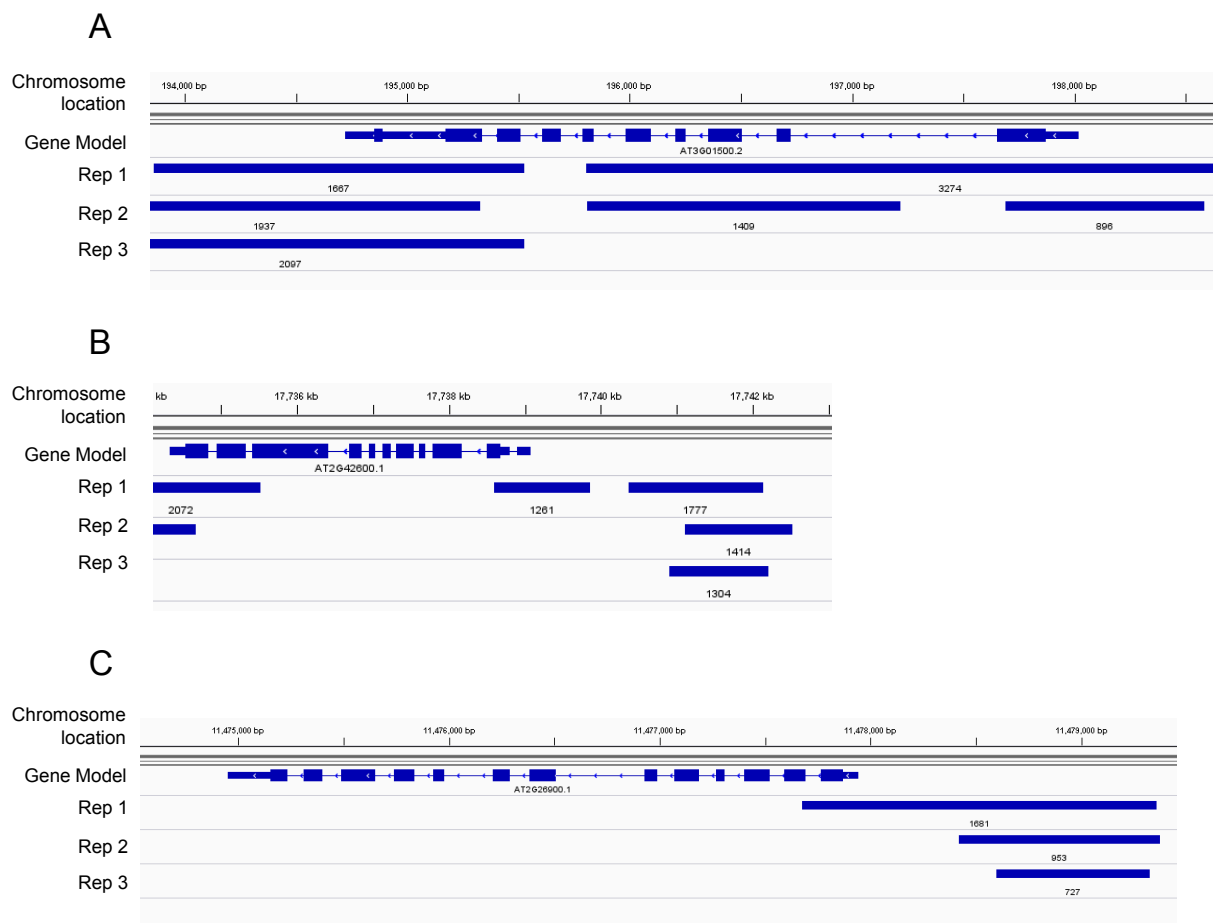


Figure S4: Location of phyA ChIP-SEQ binding sites over C_4 genes **A.** *AtCA1* **B.** *AtPPC2* **C.** *AtBASS2*. Data was obtained from¹⁵ and visualized using the Integrative Genome Viewer (IGV).

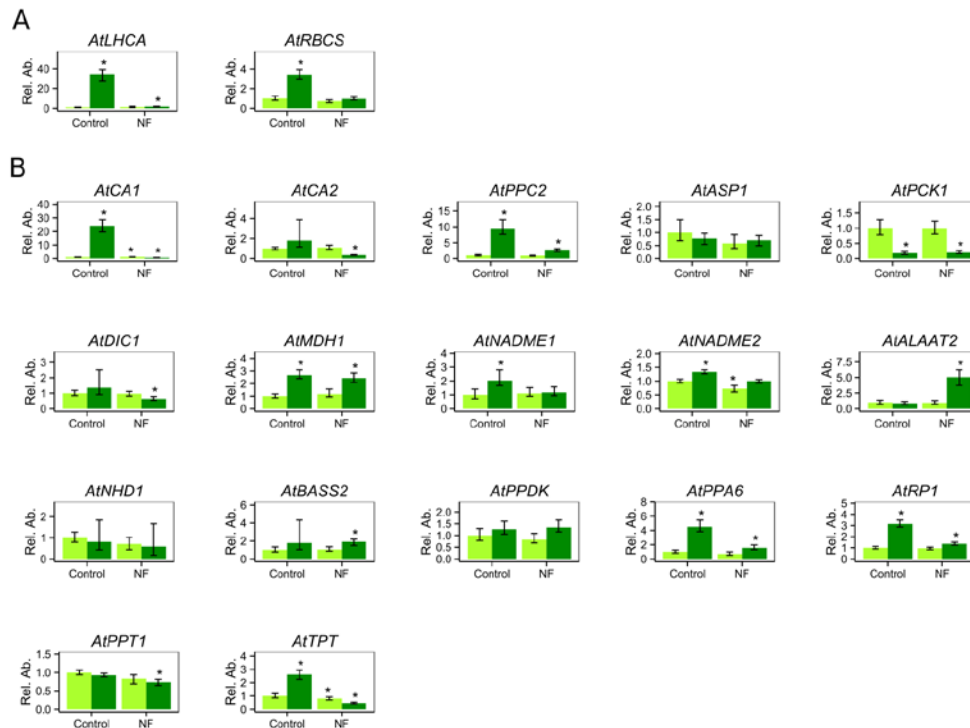


Figure S5: Chloroplast regulation of C_4 genes in *A. thaliana*. Quantitative real time polymerase chain reactions were used to determine the impact of light and chloroplast signalling on C_4 transcript abundance in *A. thaliana*. Data are derived from samples collected from dark-grown seedlings $\pm$ 5 μ M 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 (*).

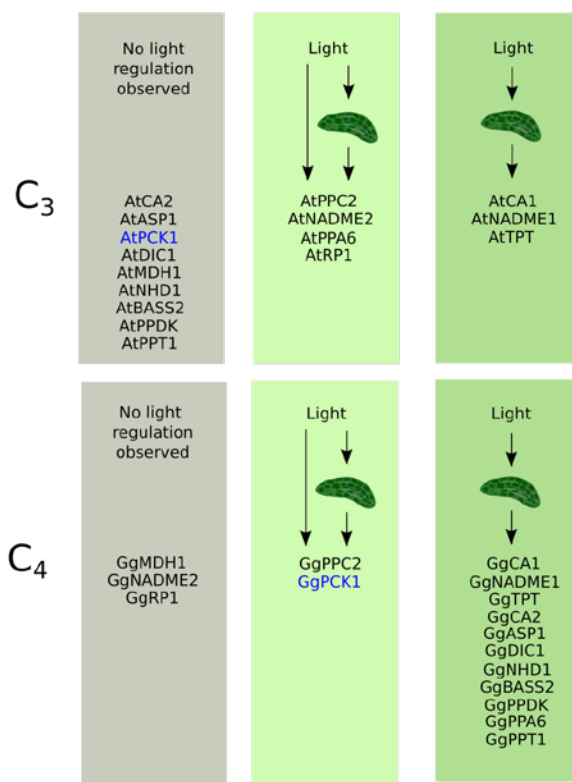


Figure S6: Gain of chloroplast-dependent C₄ gene expression in *G. gynandra*. Schematic illustrating the extent of light-activated, chloroplast-dependent C₄ gene expression in the ancestral C₃ state, based on data from *A. thaliana*, as well the C₄ state based on data from *G. gynandra*.

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	204	2.30E-19
	process		
GO:0019682	glyceraldehyde-3-phosphate metabolic	205	3.30E-19
	process		
GO:0006090	pyruvate metabolic process	207	6.30E-19
	isopentenyl diphosphate biosynthetic		
GO:0009240	process	207	6.30E-19
	isopentenyl diphosphate metabolic		
GO:0046490	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	26	0.00516
	substance		
GO:0052482	defense response by cell wall	15	0.00666
	thickening		
GO:0052544	defense response by callose	15	0.00666
	deposition in cell wall		
GO:0048443	stamen development	114	0.00828
GO:0048466	androecium development	114	0.00828
GO:0042547	cell wall modification involved in	16	0.00901
	multidimensional cell growth		
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 from⁷.

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 browser⁸ and correspond to values from the vegetative rosette stage.

Supplementary References

1. 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).
2. 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).
3. 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).
4. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
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. Covshoff, S. *et al.* Deregulation of maize C4 photosynthetic development in a mesophyll cell-defective mutant. *Plant Physiol* **146**, 1469–1481 (2008).
8. Winter, D. *et al.* An ‘Electronic Fluorescent Pictograph’ browser for exploring and analyzing large-scale biological data sets. *PLoS One* **2**, e718 (2007).