Title:

Wide sampling of natural diversity identifies novel molecular signatures of C_4 photosynthesis

Authors:

Steven Kelly[1§*], Sarah Covshoff[2§], Samart Wanchana[3*§], Vivek Thakur[3], W. Paul Quick[3,4], Yu Wang[5], Martha Ludwig[6], Richard Bruskiewich[3], Alisdair R. Fernie[7], Rowan F. Sage[8], Zhijian Tian[9], Zixiang Yan[9], Jun Wang[9], Yong Zhang[10], Xin-Guang Zhu[11,12], Gane Ka-Shu Wong[9,10,13,*], Julian M. Hibberd[2,*]

Affiliations:

[1]Department of Plant Sciences, South Parks Road, University of Oxford, Oxford, OX1 3RB, UK.

[2]Department of Plant Sciences, Downing Street, University of Cambridge, Cambridge, CB2 3EA, UK.

[3]C_4 Rice Center, International Rice Research Institute (IRRI), DAPO Box 7777, Metro Manila, Philippines.

[4]Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK.

[5]Institute for Genomic Biology, University of Illinois at Urbana Champagne, IL USA, 61801.

[6]School of Chemistry & Biochemistry, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia.

[7]Max-Planck-Insitüt für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, D-14476 Potsdam-Golm, Germany.

[8]Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 3B2, Canada.

[9]BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China.

[10]Department of Biological Sciences, University of Alberta, Edmonton AB, T6G 2E9, Canada.

[11]CAS-MPG Partner Institute for Computational Biology, Chinese Academy of Sciences, Shanghai, China 200031.

[12]State Laboratory for Hybrid Rice Research, Changsha, Hunan, China.

[13]Department of Medicine, University of Alberta, Edmonton AB, T6G 2E1, Canada.

*For correspondence: email - jmh65@cam.ac.uk, gane@ualberta.ca, steven.kelly@plants.ox.ac.uk

Introductory paragraph

Much of biology is associated with convergent traits, and it is challenging to determine the extent to which underlying molecular mechanisms are shared across phylogeny. By analyzing plants representing eighteen independent origins of C_4 photosynthesis, we quantified the extent to which this convergent trait utilises identical molecular mechanisms. We demonstrate that biochemical changes that characterise C_4 species are recovered by this process, and expand the paradigm by four metabolic pathways not previously associated with C_4 photosynthesis. Furthermore, we show that expression of many genes that distinguish C_3 and C_4 species respond to low CO_2, providing molecular evidence that reduction in atmospheric CO_2 was a driver for C_4 evolution. Thus the origin and architecture of complex traits can be derived from transcriptome comparisons across natural diversity.

Main text:

The evolution of complex traits has produced great diversity in form and function across the living world. A large number of similar complex traits have evolved independently in multiple disparate lineages indicating that common responses to environmental selection can result in convergent phenotypes[1,2]. The C_4 photosynthetic pathway, with at least 65 independent origins distributed across the angiosperms[3], is considered one of the most remarkable examples of evolutionary convergence in eukaryotes. Thus the C_4 pathway represents an attractive trait with which to determine whether phylogenetically diverse species can be examined to discover the shared molecular basis of complex convergent phenotypes.

Using a comparative approach, we analyzed gene expression of 30 C_4 and 17 C_3 species representing 18 independent evolutionary origins of C_4 photosynthesis (Fig. 1A,B). This set of species includes representatives from all seven orders within the eudicotyledons known to have evolved C_4 photosynthesis (Fig. 1a, Supplemental File 1)[4]. This sampling expands upon previous transcriptome studies with C_3 and C_4 eudicot plants in *Cleomaceae*, *Asteraceae* and *Portulaceae*[1–3]. RNA was isolated from leaves of all species, sequenced and subjected to *de novo* transcriptome assembly. Collectively these samples comprise 850 million reads that were assembled into 1.5 million contigs of which 1.1 million were assigned to orthogroups (Fig. 1B, Supplemental File 1) using a machine learning approach[5]. Analysis of the correlation in mRNA abundance estimates between species revealed that species did not cluster according to their photosynthetic type but rather according to phylogenetic relationship (Supplemental File 2). That is, C_4 species of *Flaveria* are more similar to C_3 *Flaveria* than to C_4 species from other genera, and thus variation in gene expression underlying phenotypic convergence is not the primary determinant of differences in mRNA abundance between C_3 and C_4 species.

Comparison of the transcriptomes from C_3 and C_4 species identified 149 genes that showed altered transcript abundance between C_3 and C_4 species in all 18 lineages: 113 that were more abundant and 36 less abundant in all C_4 species (Table 1, Supplemental File 3, Supplemental File 4). This set includes many genes encoding components of C_4 photosynthesis that are known to change during evolution of the C_4 pathway (Fig. 2A, Supplemental File 5, and Supplemental File 6). Four transcription factors were more abundant in all C_4 species (*PAT1, ZML2, SHR* and a

bHLH transcription factor of unknown function). Both *PAT1* and *ZML2* act to induce the expression of genes encoding photosynthesis proteins downstream of phytochrome and cryptochrome signalling respectively[6,7] while *SHR* is a validated regulator of C_4 Kranz anatomy in *Zea mays*[8-10]. Thus three of the four transcription factors have previously been identified as playing a role in the regulation of photosynthesis gene expression or leaf anatomy, both of which are altered during the evolution of C_4 photosynthesis. The uncharacterised bHLH domain transcription factor has no known functional role, but has previously been described as being upregulated in the bundle sheath (BS) cells of the C_3 plant *Arabidopsis thaliana*[11]. It is therefore possible that this bHLH transcription factor plays an ancestral role in the BS of C_3 species that has become enhanced in all C_4 lineages. Wide sampling of natural diversity therefore indicates that there is convergence in the recruitment of key regulators of gene expression in independent lineages of C_4 species.

Transcripts encoding 16 proteins comprising four metabolic pathways that have not previously been associated with C_4 photosynthesis were detected as differentially abundant between C_4 and C_3 species (Fig. 2B, Supplemental File 7). These pathways described below encompass: a novel carbon concentrating pathway involving the GABA shunt; metabolism associated with regeneration of phospho*enol*pyruvate (PEP), the primary CO_2 acceptor in the C_4 pathway; modifications to pyruvate metabolism that prevent diversion of pyruvate from the C_4 cycle into non-photosynthetic pathways such as lipid and branched amino acid biosynthesis; and a photorespiratory pathway previously associated with chlorophyte algae (Supplemental File 7).

The abundance of transcripts encoding a key component of the γ-aminobutyric acid (GABA) shunt was increased in all C_4 compared with C_3 species. In the conventional model for NAD-ME type C_4 photosynthesis, aspartate synthesised in mesophyll (M) cells is shuttled to mitochondria in the BS where it is transaminated to oxaloacetate by aspartate amino transferase (ASP1), reduced to malate by NAD-dependent malate dehydrogenase (NAD-MDH), and then decarboxylated to pyruvate which can then return to the M (Fig. 2a). Although *ASP1* transcripts were more abundant in all C_4 species that we studied, this was not the case for the later steps in the NAD-ME pathway (Supplemental File 7). Instead, we propose that oxaloacetate is used to feed the tricarboxylic acid (TCA) cycle in BS cells. Here, 2-oxoglutarate synthesised by the TCA cycle can be converted to glutamate and decarboxylated by glutamate decarboxylase (GAD4) to GABA, resulting in release

of CO_2 and return of carbon skeletons as succinate to the TCA cycle (Fig. 2, Supplemental File 5, Supplemental File 7). This proposed pathway provides both a novel mechanism to transfer CO_2 to the BS using CO_2 that was fixed by phospho*enol*pyruvate carboxylase (PEPC) in the M, and a source of ATP in the BS by using NADH generated by the running the TCA cycle for oxidative phosphorylation (Fig. 2B, Supplemental File 5, Supplemental File 6). Two orthogonal approaches provide evidence that this cycle functions to concentrate CO_2 in the C_4 BS. First, computational modelling revealed that the additional ATP this pathway provides to BS cells resulted in an increased CO_2 assimilation rate irrespective of C_4 subtype under low light conditions (Supplemental File 8). Moreover, when PEPCK is used for decarboxylation this increase in CO_2 assimilation rate was maintained under high light (Supplemental File 8). Second, biochemical evidence for this pathway has in fact been reported previously - after ^{14}C labelled glutamate was fed to isolated BS strands in *Zea mays*[12], radiolabel is rapidly released and redistributed to other metabolites in a manner that is most parsimoniously explained by glutamate decarboxylase mediated decarboxylation followed by re-fixation of labelled CO^2 by RuBisCO. Thus, sampling the natural diversity of C4 species uncovered an adjunct CO_2 concentrating pathway that is supported by biochemical data and a metabolic model of C_4 photosynthesis.

To maintain flux through the C_4 pathway, PEP supply is critical as it is the entry point of the cycle. Transcripts encoding a chloroplastic phosphoglucomutase (PGM) and an enolase (ENO) were more abundant in all C_4 compared with C_3 species (Fig. 2B, Supplemental File 7). These proteins facilitate conversion of Calvin-Benson cycle intermediates to PEP (Fig. 2B), providing an additional route for transfer of photo-assimilated carbon to the M, and consequently regeneration of the initial carbon acceptor. Transcripts encoding pyruvate kinase (PK), which catalyzes the reverse reaction, were less abundant in C_4 compared with C_3 species (Fig. 2B). A reduction in the amount of the cognate protein would limit futile cycling between PEP and pyruvate during C_4 photosynthesis. The third pathway detected in our analysis indicates pyruvate metabolism has been modified to prevent diversion of pyruvate from C_4 photosynthesis into non-photosynthetic pathways such as lipid and branched amino acid biosynthesis. Transcripts encoding pyruvate dehydrogenase kinase (PDK) were more abundant (Fig. 2B), and aceto-lactate synthase (ALS) less abundant, in all C_4 compared with C_3 leaves (Supplemental File 7). As PDK deactivates

pyruvate dehydrogenase by phosphorylation and ALS channels pyruvate into the synthesis of branched-chain amino acids, these alterations would support the core C_4 cycle by reducing loss of pyruvate from photosynthetic pools.

The fourth pathway that we propose is modified in all C_4 compared with C_3 species has been previously associated with algae rather than land plants. Chloroplasts of chlorophyte algae enclose their RuBisCO in structures called pyrenoids. These structures facilitate an increased CO_2 concentration around RuBisCO resulting in reduced photorespiration[13]. These algae also lack the peroxisome-based photorespiratory pathway that evolved in the common ancestor of embryophytes and charophyte algae[14]. Although the ancestral chlorophyte photorespiratory pathway involving glycolate dehydrogenase (GlcDH) and an alanine:glyoxylate amino transferase (ALAAT2) is still active in C_3 plants, flux of glycolate through this pathway is low compared with the peroxisome-based pathway[15]. Our analysis indicates that in all 18 C_4 lineages there is a concerted increase in abundance of transcripts encoding key components of the chlorophyte algal pathway, specifically GlcDH and ALAAT2 (Fig. 2B, Supplemental File 7). Two scenarios may explain this change during C_4 evolution. First, the chlorophyte photorespiratory pathway plays a role in C_4 photosynthesis. Second, GlcDH plays a role in converting some dihydroxyacetone phosphate (DHAP) produced in the Calvin-Benson cycle of BS cells to pyruvate *via* the methylglyoxal pathway enabling the use of some DHAP to maintain the C_4 cycle (Supplemental File 9). In this latter scenario, ALAAT2 would still process photorespiratory glyoxylate that had been produced in the peroxisome by glycolate oxidase. Both proposed scenarios require a source of NAD^+ and in this context it is noteworthy that both transcripts encoding Complex I of the respiratory electron transfer chain and the plant uncoupling mitochondrial protein 1 (PUMP1) were upregulated in C_4 species (Fig. 2B). Together, these would increase regeneration of NAD^+ and de-couple some proton flux through Complex I from ATP synthesis. Moreover, this increase in NAD^+ would also support photorespiratory glycine decarboxylase and utilise NADH from the TCA cycle (Fig. 2B).

We also evaluated whether transcriptome sampling across a deep phylogeny could be used to clarify selective forces promoting C_4 evolution. Low atmospheric CO_2 has been proposed to be a key driver of C_4 evolution[16], and analysis of *A. thaliana* identified genes responsive to low CO_2 in plants[17]. Thirty-one of the 113 genes that were more abundant in all C_4 species sampled showed

increased expression in *A. thaliana* grown under low CO_2 (Fig. 3, Supplemental File 10). As the probability of such an overlap is low ($p = 4 \times 10^{-9}$), these data indicate that there is a significant association between genes that are expressed highly in C_4 species and those that are more abundant in C_3 *A. thaliana* grown under low CO_2. There is also a significant association between genes that are more abundant under low CO_2 and those that are less abundant in all C_4 plants ($p = 1 \times 10^{-5}$, Fig. 3). However, nine of these twelve genes encode components of the photorespiratory pathway and the remaining three are unknown proteins predicted to localize to the chloroplast and are thus implicated in photorespiratory processes (Supplemental File 11). These results are consistent with the hypothesis that low atmospheric CO_2 concentration induced changes in gene expression that facilitated C_4 evolution. The molecular mechanisms that underpin this response remain to be identified. In the future it will be informative to investigate the extent to which other ecological drivers such as heat, drought and salinity alter gene expression and potentially target genes that are then recruited into this complex trait. In addition, the data also reveal that a set of genes that are more abundant in C_4 species are preferentially expressed in the BS cells of C_3 species and thus indicate that neofunctionalisation of BS cells utilized pathways already present in this cell type (Supplemental File 5, Supplemental File 12). Thus these data also provide molecular support for the hypothesis that expansion and specialization of the C_3 BS is an early and key step in the evolution of the C_4 phenotype[19,18].

Finally, the data identify four metabolic pathways previously unknown to be important for C_4 function, and identify a role for the GABA shunt pathway in concentrating CO_2 and generating ATP in the BS of all C_4 species. The re-emergence in plants of the peroxisome-based photorespiratory pathway from algae is to our knowledge, the first documented example of an evolutionary reversion being a key component of the advent of complexity and convergence in eukaryotic biology. We envisage that this approach of comparative transcriptome sampling of non-model species will now be used to provide insight into molecular signatures associated with complex traits across the tree of life. The fact that we recapitulated previous knowledge of C_4 photosynthesis, but also significantly extended the functional model of C_4 metabolism, implies there is much more to be discovered about this pathway.

194

Figures

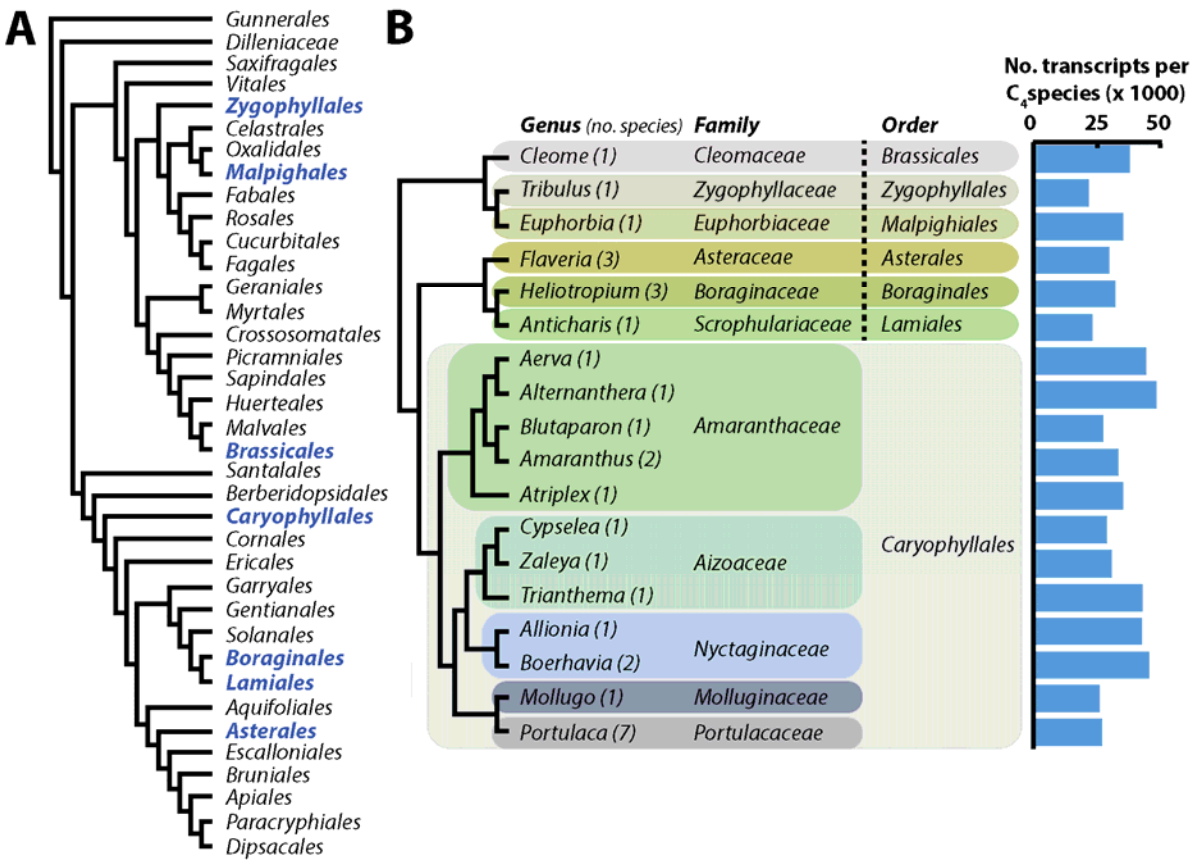


Fig. 1: Within the eudicotyledons C_4 photosynthesis has evolved in eleven families from 7 different orders. (A) Phylogenetic tree showing the seven orders containing C_4 species (blue). (B) Phylogenetic tree showing the relationship between the eighteen genera (eleven families) encompassing thirty C_4 species sampled in this study. Numbers after the genus name indicate the number of C_4 species sampled. The mean number of *de novo* transcripts per species is indicated for each genus.

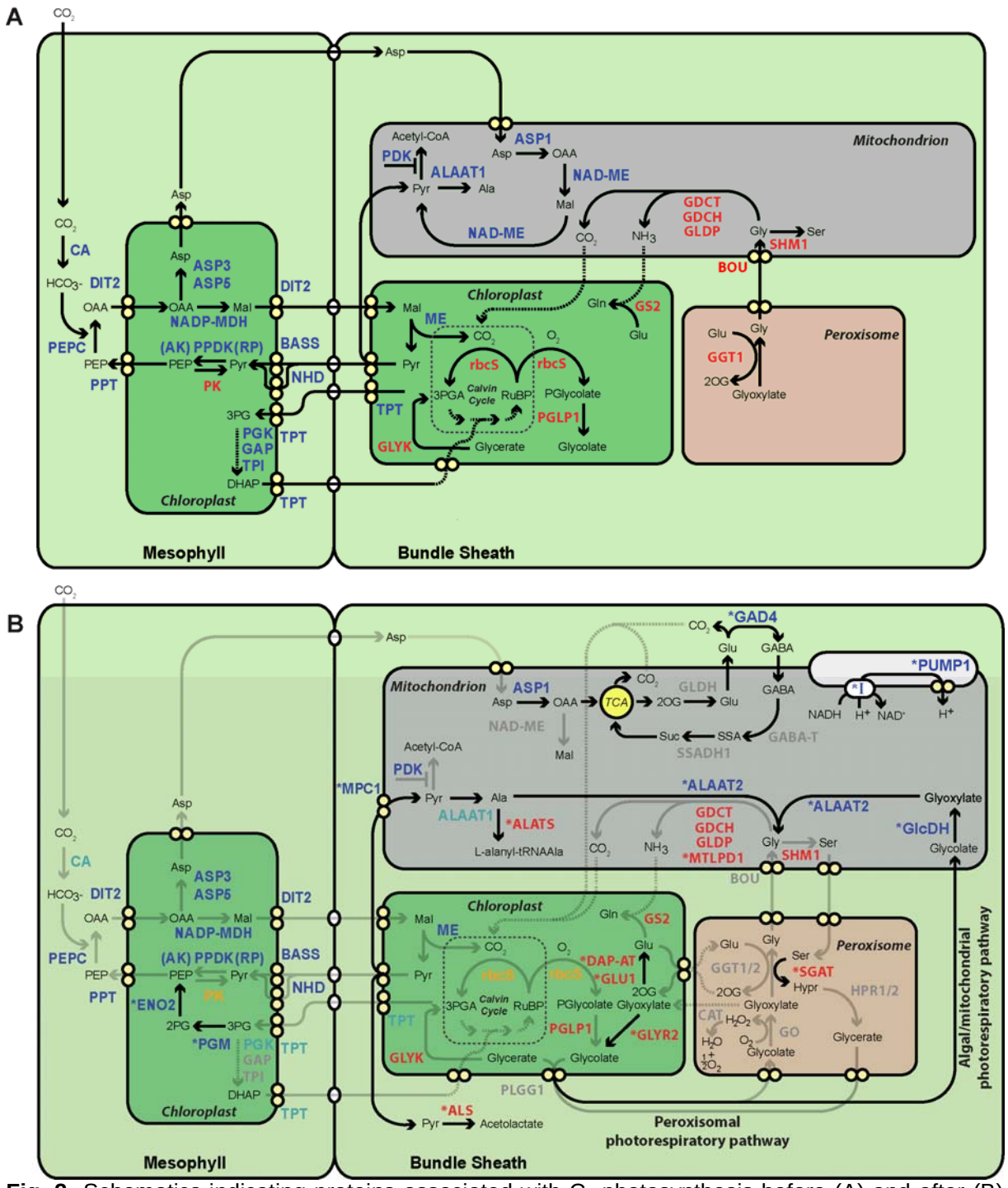
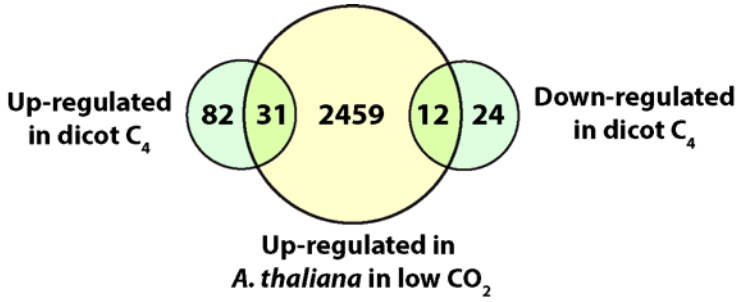


Fig. 2: Schematics indicating proteins associated with C_4 photosynthesis before (A) and after (B) this study. (A) Enzymes and transporters previously known to be upregulated (blue) and down-regulated (red) in C_4 compared with C_3 leaves are depicted according subcellular location and whether they are preferentially expressed in either mesophyll of bundle sheath cells. (B) After sequencing 30 C_4 and 17 C_3 species spanning the seven eudicotyledon orders known to have evolved C_4 photosynthesis an additional 16 genes consistently up (blue) or down regulated (red) in

212 C_4 compared with C_3 leaves were identified. Light blue and orange indicate transcripts for genes
213 that are significantly upregulated and downregulated in all C_4 species when compared to all C_3, but
214 they fail to achieve significance with computational occlusion and resampling. For abbreviations of
215 gene names see Supplemental File 10. TCA, tricarboxylic cycle.
216



217

218 **Fig. 3:** Overlap between genes that are upregulated in response to low CO_2 in *Arabidopsis*
219 *thaliana* and that are up or downregulated in C_4 species.
220

Category	Cell	Chloroplast	Mitochondrion
Metabolic components	15 (8)	29 (10)	6 (4)
Signalling components	18 (2)	7 (1)	0 (0)
Transporters	4 (1)	9 (1)	2 (0)
Transcription regulators	4 (1)	0 (2)	0 (0)
Post transcription regulators	3 (1)	0 (0)	0 (0)
Other	3 (0)	2 (2)	0 (0)
Unknown	9 (3)	1 (0)	1 (0)

221 n= 113, n = 36 No. up-regulated (No. down-regulated)

222 **Table 1:** Summary of functional categories of genes differentially expressed between C_4 and C_3
223 leaves.

References

1 Stern, D. L. The genetic causes of convergent evolution. *Nature reviews. Genetics* **14**, 751-764, doi:10.1038/nrg3483 (2013).

2 Morris, S. C. Evolution: like any other science it is predictable. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **365**, 133-145, doi:10.1098/rstb.2009.0154 (2010).

3 Sage, R. F. A portrait of the C4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *Journal of experimental botany* **67**, 4039-4056, doi:10.1093/jxb/erw156 (2016).

4 Sage, R. F., Christin, P. A. & Edwards, E. J. The C(4) plant lineages of planet Earth. *Journal of experimental botany* **62**, 3155-3169, doi:10.1093/jxb/err048 (2011).

5 Aubry, S., Kelly, S., Kumpers, B. M., 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 genetics* **10**, e1004365, doi:10.1371/journal.pgen.1004365 (2014).

6 Torres-Galea, P., Hirtreiter, B. & Bolle, C. Two GRAS proteins, SCARECROW-LIKE21 and PHYTOCHROME A SIGNAL TRANSDUCTION1, function cooperatively in phytochrome A signal transduction. *Plant physiology* **161**, 291-304, doi:10.1104/pp.112.206607 (2013).

7 Shaikhali, J. *et al.* The CRYPTOCHROME1-dependent response to excess light is mediated through the transcriptional activators ZINC FINGER PROTEIN EXPRESSED IN INFLORESCENCE MERISTEM LIKE1 and ZML2 in Arabidopsis. *The Plant cell* **24**, 3009-3025, doi:10.1105/tpc.112.100099 (2012).

8 Wang, P., Kelly, S., Fouracre, J. P. & Langdale, J. A. Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C4 Kranz anatomy. *The Plant journal : for cell and molecular biology* **75**, 656-670, doi:10.1111/tpj.12229 (2013).

9 Slewinski, T. L. Using evolution as a guide to engineer kranz-type c4 photosynthesis. *Frontiers in plant science* **4**, 212, doi:10.3389/fpls.2013.00212 (2013).

10 Slewinski, T. L. *et al.* Short-root1 plays a role in the development of vascular tissue and kranz anatomy in maize leaves. *Molecular plant* **7**, 1388-1392, doi:10.1093/mp/ssu036 (2014).

11 Aubry, S., Smith-Unna, R. D., Boursnell, C. M., Kopriva, S. & Hibberd, J. M. Transcript residency on ribosomes reveals a key role for the Arabidopsis thaliana bundle sheath in sulfur and glucosinolate metabolism. *The Plant journal : for cell and molecular biology* **78**, 659-673, doi:10.1111/tpj.12502 (2014).

12 Weissmann, S. *et al.* Interactions of C4 Subtype Metabolic Activities and Transport in Maize Are Revealed through the Characterization of DCT2 Mutants. *The Plant cell* **28**, 466-484, doi:10.1105/tpc.15.00497 (2016).

13 Giordano, M., Beardall, J. & Raven, J. A. CO2 concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. *Annual review of plant biology* **56**, 99-131, doi:10.1146/annurev.arplant.56.032604.144052 (2005).

14 Finet, C., Timme, R. E., Delwiche, C. F. & Marletaz, F. Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. *Current biology : CB* **20**, 2217-2222, doi:10.1016/j.cub.2010.11.035 (2010).

15 Niessen, M. *et al.* Mitochondrial glycolate oxidation contributes to photorespiration in higher plants. *Journal of experimental botany* **58**, 2709-2715, doi:10.1093/jxb/erm131 (2007).

16 Ehleringer, J. R., Sage, R. F., Flanagan, L. B. & Pearcy, R. W. Climate change and the evolution of C(4) photosynthesis. *Trends in ecology & evolution* **6**, 95-99, doi:10.1016/0169-5347(91)90183-X (1991).

17 Li, Y., Xu, J., Haq, N. U., Zhang, H. & Zhu, X. G. Was low CO2 a driving force of C4 evolution: Arabidopsis responses to long-term low CO2 stress. *Journal of experimental botany* **65**, 3657-3667, doi:10.1093/jxb/eru193 (2014).

18 Williams, B. P., Johnston, I. G., Covshoff, S. & Hibberd, J. M. Phenotypic landscape inference reveals multiple evolutionary paths to C4 photosynthesis. *eLife* **2**, e00961, doi:10.7554/eLife.00961 (2013).

19 Christin, P. A. *et al.* Anatomical enablers and the evolution of C4 photosynthesis in grasses. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 1381-1386, doi:10.1073/pnas.1216777110 (2013).

Acknowledgements:

This work was funded by a Bill and Melinda Gates Foundation and a Department for International Development award to IRRI and sub-awards to SK, XZ, RFS and JMH. The 1000 Plants (1KP) initiative, led by GKSW, is funded by the Alberta Ministry of Advanced Education, Alberta Innovates Technology Futures (AITF), Innovates Centre of Research Excellence (iCORE), Musea Ventures, BGI-Shenzhen and China National Genebank (CNGB), and DSERC Discovery grants to RFS. SK is a Royal Society University Research Fellow. This work was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no 637765.

List of Supplementary Materials:

Supplemental File 1: Summary of C_3 and C_4 species sampled in this study. Each species is classified by its photosynthetic pathway, its family and order. The number of sequenced reads, assembled contigs, annotated contigs, and assembly N50 for each species obtained are also listed.

Supplemental File 2: Phylogenetic position accounts for more variance in mRNA abundance than photosynthetic pathway. The heat map depicts the Spearman's ranked correlation coefficient (ρ) between each species pair computed from global mRNA abundance estimates. The hierarchical cluster (tree to the left of the heat map) was computed directly from the correlation coefficients by converting these correlation coefficients to distance estimates. The distance estimate between two species A and B is evaluated as 1 - ρ . i.e. $d(A,B) = 1 - \rho$. A tree is then inferred from these

distance estimates using the minimum evolution principle. Names of C_4 species are shown in blue.

Supplemental File 3: Summary of genes showing differential expression between C_4 and C_3 leaves. The likelihood of differential expression is provided considering all species. Also provided is the proportion of resampling tests in which the gene was detected as consistently differentially regulated between C_3 and C_4 species. The mean and standard deviation of the expression estimate is provided for the C_3 and C_4 cohort as well as the number of samples that were identified as outliers and masked prior to differential expression testing. The expected count for each gene for each species is also provided.

Supplemental File 4: The subset of genes from Supplemental File 3 that received 100% support from computational occlusion and resampling.

Supplemental File 5: Additional information.

Supplemental File 6: Genes previously reported to be differentially expressed in C_4 compared with C_3 leaves.

Supplemental File 8: Four additional metabolic pathways identified in this study. In each case the relative expression is given for C_3 (grey bars) and C_4 (green bars) species. * indicates a likelihood of differential expression ≥ 0.95 and 100% support from computational occlusion and resampling. n.s. indicates a non-significant difference between the C_3 and C_4 expression levels. (A) GABA shunt. (B) Phospho*enol*pyruvate regeneration. (C) First step of branched chain amino acid biosynthesis. (D) Chlorophyte photorespiratory pathway.

Supplemental File 8: Modelling the addition of the GABA shunt to the C4 photosynthesis.

Supplemental File 9: Schematic illustrating an alternative hypothesis for the function of the glycolate dehydrogenase gene that is potentially a lactate dehydrogenase. This pathway would

337 convert dihydroxyacetone phosphate to pyruvate *via* methylglyoxal.

338

339 **Supplemental File 10:** Genes that were up-regulated in all C_4 species and also up-regulated in
340 response to low atmospheric CO_2 in the C_3 plant *Arabidopsis thaliana*.

341

342 **Supplemental File 11**: Genes that were down-regulated in all C_4 species and also up-regulated in
343 response to low atmospheric CO_2 in the C_3 plant *Arabidopsis thaliana*.

344

345 **Supplemental file 12:** The overlap between the genes differentially regulated in all C_4 species with
346 other datasets. A) Comparison of transcripts upregulated in all C_4 species with those preferentially
347 expressed in *Arabidopsis thaliana* bundlesheath cells. B) Comparison of genes identified as
348 differentially abundant in all C_4 species compared to those identified as differentially abundant
349 between C_3 and C_4 species of *Flaveria*. C) Analysis of the cell type specific expression in *Zea*
350 *mays* and *Setaria italica* of the orthologues of the genes identified as up-regulated in all C_4 species
351 in this study.

352

353 **Supplemental File 13:** Genes that were up-regulated in all C_4 species and also up-regulated in in
354 bundle sheath cells of the C_3 plant *Arabidopsis thaliana*.

355

356 **Supplemental File 14:** Full names and accession numbers for all genes shown in Fig. 2.

357

358 **Supplementary Methods:** Detailed descriptions of the data sources and methods.

359

360

361 **Author contributions**

362 SK, GKSW, RB, RFS, SC and JMH conceived the work. SC and RFS acquired the plant collection
363 and mRNA, while SK, VT, SW, WPQ, XGZ, YW, GKSW, ML, RB, JW, YZ, ZY, ZT, ARF and RFS
364 conducted the analysis. SK designed and developed the bioinformatic analyses. SK and JMH
365 interpreted the data and wrote the paper.