Title:

Wide sampling of natural diversity identifies novel molecular signatures of C₄ photosynthesis

Authors:

Steven Kelly^{1§*}, Sarah Covshoff^{2§}, Samart Wanchana^{3*§}, Vivek Thakur³, W. Paul Quick^{3,4}, Yu Wang⁵, Martha Ludwig⁶, Richard Bruskiewich³, Alisdair R. Fernie⁷, Rowan F. Sage⁸, Zhijian Tian⁹, Zixiang Yan⁹, Jun Wang⁹, Yong Zhang¹⁰, Xin-Guang Zhu^{11,12}, Gane Ka-Shu Wong^{9,10,13,*}, Julian M. Hibberd^{2,*}

Affiliations:

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

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

³C₄ Rice Center, International Rice Research Institute (IRRI), DAPO Box 7777, Metro Manila, Philippines.

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

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

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

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

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

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

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

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

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

¹³Department 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 C₄ photosynthesis, we quantified the extent to which this convergent trait utilises identical molecular mechanisms. We demonstrate that biochemical changes that characterise C₄ species are recovered by this process, and expand the paradigm by four metabolic pathways not previously associated with C₄ photosynthesis. Furthermore, we show that expression of many genes that distinguish C₃ and C₄ species respond to low CO₂, providing molecular evidence that reduction in atmospheric CO₂ was a driver for C₄ evolution. Thus the origin and architecture of complex traits can be derived from transcriptome comparisons across natural diversity.

49 Main text:

50 The evolution of complex traits has produced great diversity in form and function across the
51 living world. A large number of similar complex traits have evolved independently in multiple
52 disparate lineages indicating that common responses to environmental selection can result in
53 convergent phenotypes^{1,2}. The C₄ photosynthetic pathway, with at least 65 independent origins
54 distributed across the angiosperms³, is considered one of the most remarkable examples of
55 evolutionary convergence in eukaryotes. Thus the C₄ pathway represents an attractive trait with
56 which to determine whether phylogenetically diverse species can be examined to discover the
57 shared molecular basis of complex convergent phenotypes.

58 Using a comparative approach, we analyzed gene expression of 30 C₄ and 17 C₃ species
59 representing 18 independent evolutionary origins of C₄ photosynthesis (Fig. 1A,B). This set of
60 species includes representatives from all seven orders within the eudicotyledons known to have
61 evolved C₄ photosynthesis (Fig. 1a, Supplemental File 1)⁴. This sampling expands upon previous
62 transcriptome studies with C₃ and C₄ eudicot plants in *Cleomaceae*, *Asteraceae* and *Portulacaceae*¹⁻
63 ³. RNA was isolated from leaves of all species, sequenced and subjected to *de novo* transcriptome
64 assembly. Collectively these samples comprise 850 million reads that were assembled into 1.5
65 million contigs of which 1.1 million were assigned to orthogroups (Fig. 1B, Supplemental File 1)
66 using a machine learning approach⁵. Analysis of the correlation in mRNA abundance estimates
67 between species revealed that species did not cluster according to their photosynthetic type but
68 rather according to phylogenetic relationship (Supplemental File 2). That is, C₄ species of *Flaveria*
69 are more similar to C₃ *Flaveria* than to C₄ species from other genera, and thus variation in gene
70 expression underlying phenotypic convergence is not the primary determinant of differences in
71 mRNA abundance between C₃ and C₄ species.

72 Comparison of the transcriptomes from C₃ and C₄ species identified 149 genes that showed
73 altered transcript abundance between C₃ and C₄ species in all 18 lineages: 113 that were more
74 abundant and 36 less abundant in all C₄ species (Table 1, Supplemental File 3, Supplemental File
75 4). This set includes many genes encoding components of C₄ photosynthesis that are known to
76 change during evolution of the C₄ pathway (Fig. 2A, Supplemental File 5, and Supplemental File
77 6). Four transcription factors were more abundant in all C₄ 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 C₄ 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 C₄ 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 C₃ plant *Arabidopsis thaliana*¹¹. It is therefore possible that this bHLH transcription factor plays an ancestral role in the BS of C₃ species that has become enhanced in all C₄ 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 C₄ species.

Transcripts encoding 16 proteins comprising four metabolic pathways that have not previously been associated with C₄ photosynthesis were detected as differentially abundant between C₄ and C₃ 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 phosphoenolpyruvate (PEP), the primary CO₂ acceptor in the C₄ pathway; modifications to pyruvate metabolism that prevent diversion of pyruvate from the C₄ 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 C₄ compared with C₃ species. In the conventional model for NAD-ME type C₄ 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 C₄ 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 CO₂ 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 CO₂ to the BS using CO₂ that was fixed by phosphoenolpyruvate 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 CO₂ in the C₄ BS. First, computational modelling revealed that the additional ATP this pathway provides to BS cells resulted in an increased CO₂ assimilation rate irrespective of C₄ subtype under low light conditions (Supplemental File 8). Moreover, when PEPC is used for decarboxylation this increase in CO₂ assimilation rate was maintained under high light (Supplemental File 8). Second, biochemical evidence for this pathway has in fact been reported previously - after ¹⁴C 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 CO₂ by RuBisCO. Thus, sampling the natural diversity of C₄ species uncovered an adjunct CO₂ concentrating pathway that is supported by biochemical data and a metabolic model of C₄ photosynthesis.

To maintain flux through the C₄ 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 C₄ compared with C₃ 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 C₄ compared with C₃ species (Fig. 2B). A reduction in the amount of the cognate protein would limit futile cycling between PEP and pyruvate during C₄ photosynthesis. The third pathway detected in our analysis indicates pyruvate metabolism has been modified to prevent diversion of pyruvate from C₄ 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 C₄ compared with C₃ 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 C₄ cycle by reducing loss of pyruvate from photosynthetic pools.

The fourth pathway that we propose is modified in all C₄ compared with C₃ 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 CO₂ concentration around RuBisCO resulting in reduced photorespiration¹³. These algae also lack the peroxisome-based photorespiratory pathway that evolved in the common ancestor of embryophytes and charophyte algae¹⁴. Although the ancestral chlorophyte photorespiratory pathway involving glycolate dehydrogenase (GlcDH) and an alanine:glyoxylate amino transferase (ALAAT2) is still active in C₃ plants, flux of glycolate through this pathway is low compared with the peroxisome-based pathway¹⁵. Our analysis indicates that in all 18 C₄ 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 C₄ evolution. First, the chlorophyte photorespiratory pathway plays a role in C₄ 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 C₄ 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 C₄ 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 C₄ evolution. Low atmospheric CO₂ has been proposed to be a key driver of C₄ evolution¹⁶, and analysis of *A. thaliana* identified genes responsive to low CO₂ in plants¹⁷. Thirty-one of the 113 genes that were more abundant in all C₄ species sampled showed

increased expression in *A. thaliana* grown under low CO₂ (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 C₄ species and those that are more abundant in C₃ *A. thaliana* grown under low CO₂. There is also a significant association between genes that are more abundant under low CO₂ and those that are less abundant in all C₄ 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 CO₂ concentration induced changes in gene expression that facilitated C₄ 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 C₄ species are preferentially expressed in the BS cells of C₃ 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 C₃ BS is an early and key step in the evolution of the C₄ phenotype^{19,18}.

Finally, the data identify four metabolic pathways previously unknown to be important for C₄ function, and identify a role for the GABA shunt pathway in concentrating CO₂ and generating ATP in the BS of all C₄ 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 C₄ photosynthesis, but also significantly extended the functional model of C₄ metabolism, implies there is much more to be discovered about this pathway.

195 **Figures**

196

197 **Fig. 1:** Within the eudicotyledons C₄ photosynthesis has evolved in eleven families from 7 different
198 orders. (A) Phylogenetic tree showing the seven orders containing C₄ species (blue). (B)
199 Phylogenetic tree showing the relationship between the eighteen genera (eleven families)
200 encompassing thirty C₄ species sampled in this study. Numbers after the genus name indicate the
201 number of C₄ species sampled. The mean number of *de novo* transcripts per species is indicated
202 for each genus.

203

204

Fig. 2: Schematics indicating proteins associated with C₄ photosynthesis before (A) and after (B) this study. (A) Enzymes and transporters previously known to be upregulated (blue) and downregulated (red) in C₄ compared with C₃ leaves are depicted according subcellular location and whether they are preferentially expressed in either mesophyll or bundle sheath cells. (B) After sequencing 30 C₄ and 17 C₃ species spanning the seven eudicotyledon orders known to have evolved C₄ photosynthesis an additional 16 genes consistently up (blue) or down regulated (red) in

C₄ compared with C₃ leaves were identified. Light blue and orange indicate transcripts for genes that are significantly upregulated and downregulated in all C₄ species when compared to all C₃, but they fail to achieve significance with computational occlusion and resampling. For abbreviations of gene names see Supplemental File 10. TCA, tricarboxylic cycle.

Fig. 3: Overlap between genes that are upregulated in response to low CO₂ in *Arabidopsis thaliana* and that are up or downregulated in C₄ species.

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)

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

Table 1: Summary of functional categories of genes differentially expressed between C₄ and C₃ leaves.

224 References

- 225 1 Stern, D. L. The genetic causes of convergent evolution. *Nature reviews. Genetics*
226 **14**, 751-764, doi:10.1038/nrg3483 (2013).
- 227 2 Morris, S. C. Evolution: like any other science it is predictable. *Philosophical*
228 *transactions of the Royal Society of London. Series B, Biological sciences* **365**, 133-
229 145, doi:10.1098/rstb.2009.0154 (2010).
- 230 3 Sage, R. F. A portrait of the C4 photosynthetic family on the 50th anniversary of its
231 discovery: species number, evolutionary lineages, and Hall of Fame. *Journal of*
232 *experimental botany* **67**, 4039-4056, doi:10.1093/jxb/erw156 (2016).
- 233 4 Sage, R. F., Christin, P. A. & Edwards, E. J. The C(4) plant lineages of planet Earth.
234 *Journal of experimental botany* **62**, 3155-3169, doi:10.1093/jxb/err048 (2011).
- 235 5 Aubry, S., Kelly, S., Kumpers, B. M., Smith-Unna, R. D. & Hibberd, J. M. Deep
236 evolutionary comparison of gene expression identifies parallel recruitment of trans-
237 factors in two independent origins of C4 photosynthesis. *PLoS genetics* **10**,
238 e1004365, doi:10.1371/journal.pgen.1004365 (2014).
- 239 6 Torres-Galea, P., Hirtreiter, B. & Bolle, C. Two GRAS proteins, SCARECROW-
240 LIKE21 and PHYTOCHROME A SIGNAL TRANSDUCTION1, function
241 cooperatively in phytochrome A signal transduction. *Plant physiology* **161**, 291-304,
242 doi:10.1104/pp.112.206607 (2013).
- 243 7 Shaikhali, J. *et al.* The CRYPTOCHROME1-dependent response to excess light is
244 mediated through the transcriptional activators ZINC FINGER PROTEIN
245 EXPRESSED IN INFLORESCENCE MERISTEM LIKE1 and ZML2 in Arabidopsis.
246 *The Plant cell* **24**, 3009-3025, doi:10.1105/tpc.112.100099 (2012).
- 247 8 Wang, P., Kelly, S., Fouracre, J. P. & Langdale, J. A. Genome-wide transcript
248 analysis of early maize leaf development reveals gene cohorts associated with the
249 differentiation of C4 Kranz anatomy. *The Plant journal : for cell and molecular*
250 *biology* **75**, 656-670, doi:10.1111/tpj.12229 (2013).
- 251 9 Slewinski, T. L. Using evolution as a guide to engineer kranz-type c4
252 photosynthesis. *Frontiers in plant science* **4**, 212, doi:10.3389/fpls.2013.00212
253 (2013).
- 254 10 Slewinski, T. L. *et al.* Short-root1 plays a role in the development of vascular tissue
255 and kranz anatomy in maize leaves. *Molecular plant* **7**, 1388-1392,
256 doi:10.1093/mp/ssu036 (2014).
- 257 11 Aubry, S., Smith-Unna, R. D., Bournsnel, C. M., Kopriva, S. & Hibberd, J. M.
258 Transcript residency on ribosomes reveals a key role for the Arabidopsis thaliana
259 bundle sheath in sulfur and glucosinolate metabolism. *The Plant journal : for cell*
260 *and molecular biology* **78**, 659-673, doi:10.1111/tpj.12502 (2014).
- 261 12 Weissmann, S. *et al.* Interactions of C4 Subtype Metabolic Activities and Transport
262 in Maize Are Revealed through the Characterization of DCT2 Mutants. *The Plant*
263 *cell* **28**, 466-484, doi:10.1105/tpc.15.00497 (2016).
- 264 13 Giordano, M., Beardall, J. & Raven, J. A. CO2 concentrating mechanisms in algae:
265 mechanisms, environmental modulation, and evolution. *Annual review of plant*
266 *biology* **56**, 99-131, doi:10.1146/annurev.arplant.56.032604.144052 (2005).
- 267 14 Finet, C., Timme, R. E., Delwiche, C. F. & Marletaz, F. Multigene phylogeny of the
268 green lineage reveals the origin and diversification of land plants. *Current biology : CB*
269 **20**, 2217-2222, doi:10.1016/j.cub.2010.11.035 (2010).
- 270 15 Niessen, M. *et al.* Mitochondrial glycolate oxidation contributes to photorespiration
271 in higher plants. *Journal of experimental botany* **58**, 2709-2715,
272 doi:10.1093/jxb/erm131 (2007).

- 273 16 Ehleringer, J. R., Sage, R. F., Flanagan, L. B. & Pearcy, R. W. Climate change and
274 the evolution of C(4) photosynthesis. *Trends in ecology & evolution* **6**, 95-99,
275 doi:10.1016/0169-5347(91)90183-X (1991).
- 276 17 Li, Y., Xu, J., Haq, N. U., Zhang, H. & Zhu, X. G. Was low CO₂ a driving force of C₄
277 evolution: Arabidopsis responses to long-term low CO₂ stress. *Journal of*
278 *experimental botany* **65**, 3657-3667, doi:10.1093/jxb/eru193 (2014).
- 279 18 Williams, B. P., Johnston, I. G., Covshoff, S. & Hibberd, J. M. Phenotypic landscape
280 inference reveals multiple evolutionary paths to C₄ photosynthesis. *eLife* **2**, e00961,
281 doi:10.7554/eLife.00961 (2013).
- 282 19 Christin, P. A. *et al.* Anatomical enablers and the evolution of C₄ photosynthesis in
283 grasses. *Proceedings of the National Academy of Sciences of the United States of*
284 *America* **110**, 1381-1386, doi:10.1073/pnas.1216777110 (2013).
- 285
- 286

287 **Acknowledgements:**

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

295

296 **List of Supplementary Materials:**

297 **Supplemental File 1:** Summary of C₃ and C₄ species sampled in this study. Each species is
298 classified by its photosynthetic pathway, its family and order. The number of sequenced reads,
299 assembled contigs, annotated contigs, and assembly N50 for each species obtained are also
300 listed.

301

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

distance estimates using the minimum evolution principle. Names of C₄ species are shown in blue.

Supplemental File 3: Summary of genes showing differential expression between C₄ and C₃ 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 C₃ and C₄ species. The mean and standard deviation of the expression estimate is provided for the C₃ and C₄ 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 C₄ compared with C₃ leaves.

Supplemental File 8: Four additional metabolic pathways identified in this study. In each case the relative expression is given for C₃ (grey bars) and C₄ (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 C₃ and C₄ expression levels. (A) GABA shunt. (B) Phosphoenolpyruvate 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

convert dihydroxyacetone phosphate to pyruvate *via* methylglyoxal.

Supplemental File 10: Genes that were up-regulated in all C₄ species and also up-regulated in response to low atmospheric CO₂ in the C₃ plant *Arabidopsis thaliana*.

Supplemental File 11: Genes that were down-regulated in all C₄ species and also up-regulated in response to low atmospheric CO₂ in the C₃ plant *Arabidopsis thaliana*.

Supplemental file 12: The overlap between the genes differentially regulated in all C₄ species with other datasets. A) Comparison of transcripts upregulated in all C₄ species with those preferentially expressed in *Arabidopsis thaliana* bundle sheath cells. B) Comparison of genes identified as differentially abundant in all C₄ species compared to those identified as differentially abundant between C₃ and C₄ species of *Flaveria*. C) Analysis of the cell type specific expression in *Zea mays* and *Setaria italica* of the orthologues of the genes identified as up-regulated in all C₄ species in this study.

Supplemental File 13: Genes that were up-regulated in all C₄ species and also up-regulated in bundle sheath cells of the C₃ plant *Arabidopsis thaliana*.

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

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

Author contributions

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