Kümpers, Britta M.C. and Burgess, Steven J. and Reyna-Llorens, Ivan and Smith-Unna, Richard and Boursnell, Chris and Hibberd, Julian M. (2017) Shared characteristics underpinning C 4 leaf maturation derived from analysis of multiple C 3 and C 4 species of Flaveria. Journal of Experimental Botany, 68 (2). pp. 177-189. ISSN 1460-2431

Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/41477/2/erw488.pdf

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the Creative Commons Attribution licence and may be reused according to the conditions of the licence. For more details see: http://creativecommons.org/licenses/by/2.5/

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk

Journal of Experimental Botany, Vol. 68, No. 2 pp. 177–189, 2017
doi:10.1093/jxb/erw488 Advance Access publication 6 January 2017
This paper is available online free of all access charges (see http://jxb.oxfordjournals.org/open_access.html for further details)

RESEARCH PAPER

Shared characteristics underpinning C_4 leaf maturation derived from analysis of multiple C_3 and C_4 species of *Flaveria*

Britta M.C. Kümpers*,†, Steven J. Burgess*, Ivan Reyna-Llorens, Richard Smith-Unna, Chris Boursnell and Julian M. Hibberd‡

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

* These authors contributed equally to this work.
† Present address: Centre for Plant Integrative Biology, School of Biosciences, University of Nottingham, Loughborough LE12 5RD, UK.
‡ Correspondence: jmh65@cam.ac.uk

Received 4 November 2016; Editorial decision 7 December 2016; Accepted 13 December 2016

Editor: Susanne von Caemmerer, Australian National University

Abstract

Most terrestrial plants use C_3 photosynthesis to fix carbon. In multiple plant lineages a modified system known as C_4 photosynthesis has evolved. To better understand the molecular patterns associated with induction of C_4 photosynthesis, the genus *Flaveria* that contains C_3 and C_4 species was used. A base to tip maturation gradient of leaf anatomy was defined, and RNA sequencing was undertaken along this gradient for two C_3 and two C_4 *Flaveria* species. Key C_4 traits including vein density, mesophyll and bundle sheath cross-sectional area, chloroplast ultrastructure, and abundance of transcripts encoding proteins of C_4 photosynthesis were quantified. Candidate genes underlying each of these C_4 characteristics were identified. Principal components analysis indicated that leaf maturation and the photosynthetic pathway were responsible for the greatest amount of variation in transcript abundance. Photosynthesis genes were over-represented for a prolonged period in the C_4 species. Through comparison with publicly available data sets, we identify a small number of transcriptional regulators that have been up-regulated in diverse C_4 species. The analysis identifies similar patterns of expression in independent C_4 lineages and so indicates that the complex C_4 pathway is associated with parallel as well as convergent evolution.

Key words: C_4 leaf anatomy, C_4 photosynthesis, convergent evolution, *Flaveria*, gene expression, parallel evolution, RNA-seq.

Introduction

Photosynthesis powers life on earth by harvesting energy from sunlight to fix carbon dioxide. In seed plants, three types of photosynthesis known as C_3, C_4, and Crassulacean acid metabolism (CAM) have been defined. C_3 photosynthesis is considered the ancestral form and is found in the majority of species. All three photosynthetic pathways use Ribulose-1,5-Bisphosphate Carboxylase Oxygenase (RuBisCO) to catalyse CO_2 fixation in the Calvin–Benson–Bassham (CBB) cycle and, in so doing, two molecules of the three-carbon molecule 3-phosphoglycerate are formed. However, RuBisCO is not completely substrate specific, catalysing an oxygenation reaction that produces 2-phosphoglycolate as well as 3-phosphoglycerate (Bowes *et al.*, 1971). As phosphoglycolate is toxic and removes carbon from the CBB cycle, it is rapidly metabolized by the photorespiratory pathway to ensure it does not accumulate, but also to retrieve carbon. Photorespiration

© The Author 2017. Published by Oxford University Press on behalf of the Society for Experimental Biology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited

uses energy inputs and is not completely effective in terms of carbon retrieval so some CO_2 is lost (Bauwe *et al.*, 2010). The rate of oxygenation at the active site of RuBisCO increases with temperature, and so in lower latitudes multiple plant lineages have evolved either CAM or C_4 photosynthesis (Sage *et al.*, 2011), which in both cases initially fix CO_2 by an alternative carboxylase, and then subsequently release high concentrations of CO_2 around RuBisCO to limit oxygenation.

Almost all C_4 plants use a two-celled system in which phospho*enol*pyruvate carboxylase (PEPC) initially fixes carbon in mesophyll (M) cells, and then, after release of CO_2 in the bundle sheath (BS) cells, RuBisCO re-fixes CO_2 in the CBB cycle. The production of CO_2 within the BS leads to high concentrations of CO_2 around RuBisCO and minimizes oxygenation. PEPC generates oxaloacetate, which is rapidly metabolized to malate and/or aspartate prior to their diffusion to the BS. Three C_4 acid decarboxylase enzymes known as NADP-malic enzyme (NADP-ME), NAD-malic enzyme (NAD-ME), and phospho*enol*pyruvate carboxykinase (PEPCK) have been shown to release CO_2 around RuBisCO and, for many years, based on their relative abundance, these decarboxylase enzymes were used to define three C_4 biochemical subtypes. More recently it has been shown that this division into three types is less rigid than previously thought (Furbank, 2011; Bellasio and Griffiths, 2014; Wang *et al.*, 2014). The use of distinct decarboxylases in separate C_4 lineages indicates that C_4 photosynthesis is underpinned by convergent evolution. In addition to the 60 groups of plants known to have evolved the C_4 pathway (Sage *et al.*, 2011), there are some genera containing both C_3 and C_4 species. Probably the most studied of these clades with closely related C_3 and C_4 species is the genus *Flaveria* found in the Asteraceae (Drincovich *et al.*, 1998; Gowik *et al.*, 2011; Schulze *et al.*, 2013).

Most species of *Flaveria* are native to Central and North America and grow as annual or perennial herbs or shrubs with decussate leaves (Powell, 1978). C_4 species of *Flaveria* show high activities of the NADP-ME C_4 acid decarboxylase in chloroplasts of the BS, and leaf anatomy conforms to the Atriplicoid type (McKown and Dengler, 2007). Phylogenetic reconstruction of this group has been conducted using morphological as well as life history and gene sequence data for 21 of the 23 known species (McKown *et al.*, 2005; Lyu *et al.*, 2015). The consensus from this work is that the ancestral condition in *Flaveria* is C_3 photosynthesis (McKown *et al.*, 2005; Lyu *et al.*, 2015) and that the occurrence of multiple C_3 and C_4 species within the same genus provides an interesting system to study processes associated with the C_4 phenotype (see Supplementary Fig. S1 at *JXB* online).

In this study, we used two pairs of C_3 and C_4 *Flaveria* species, and set out to link the gradual maturation of C_3 and C_4 characteristics in leaves to underlying alterations in transcript abundance. By linking the development of the C_4 phenotype to changes in gene expression in two C_4 species, and comparing these findings with equivalent data from two C_3 species in the same genus, we aimed to identify common traits associated with C_4 photosynthesis, and to remove species-specific characteristics from our data sets. Using the maturing leaf as a dynamic system, we show that in both C_4 species studied, the induction of Kranz anatomy occurs along a base to tip developmental gradient in leaves of >2 cm length. We sampled this maturation gradient and undertook RNA sequencing to correlate the underlying patterns of gene expression with anatomical development.

Materials and methods

Plant growth

Flaveria bidentis (L.) Kuntze, *Flaveria pringlei* Gandoger, *Flaveria robusta* Rose, and *Flaveria trinervia* (Spreng.) C. Mohr were grown in a glasshouse on the roof of the Department of Plant Sciences, Cambridge. Temperature was maintained above 20 °C and supplemental lighting was provided to ensure at least 250–350 µmol m^{-2} s^{-1} photon flux density for 16 h d^{-1}. Seeds were sown directly onto soil (Levington's M3 potting compost; Scotts Miracle-Gro Company, Godalming, UK) in covered pots as seeds need high humidity for germination. Covers were removed when seedlings were 1 cm high.

Analysis of leaf anatomy

Samples of 4 mm^2 to 1 cm^2 were fixed in 4% (w/v) formaldehyde at 4 °C overnight and then placed on ice and dehydrated prior to being placed in 100% (v/v) ethanol, followed by 1:1 ethanol/Technovit mix and then 100% Technovit 7100 (Heraeus Kulzer, Germany). Samples were subsequently left in Technovit solution plus hardener I (1 g of hardener per 100 ml) for at least 1 h. Disposable plastic resin embedding moulds (Agar Scientific, UK) were filled with a mixture of Technovit plus hardener I and II (15 ml of Technovit plus hardener I were mixed with 1 ml of hardener II). Samples were arranged within the embedding moulds which were then covered with unstretched Parafilm® M to seal them from air, and left to harden overnight. Samples were removed, heated to 80 °C, and trimmed for sectioning. Sections of 2 µm thickness were produced using a Thermo Scientific Microm HM340E microtome. Ribbons were mounted onto SuperFrost®white microscope slides (VWR, Leuven, The Netherlands), left to dry, and then stained with 0.1% (w/v) toluidine blue. All sections were analysed with a BX41 light microscope (Olympus, Center Valley, PA, USA), usually using the bright-field setting.

To clear leaves, they were placed into 70% (v/v) ethanol and heated to 80 °C. The next day samples were placed in 5% (w/v) NaOH for ~15 min to clear leaves further, and then mounted in water and analysed by light microscopy.

To quantify leaf anatomical characteristics, Photoshop CS5 was used. The program was calibrated with scale bars and the lasso tool was used to measure cell area for both BS and M cells. Measurements of M cells were taken between BS cells, with 30 BS and 30 M cells being measured per leaf section. This was done for all six leaf stages and for three biological replicates for each species. Images derived from leaf sections were assembled using Photoshop. The background of sections was averaged and in some cases the contrast was increased to improve visibility of cell borders. Vein density was determined using images of cleared leaves. In mature sections of the leaf there was no ambiguity in collecting these data, whereas in the most basal sections of some leaves, the estimates may be underestimates of the extent of venation as immature veins are not always visible. Three independent leaves were measured per species. The length of the veins was measured using Q-Capture Pro7, and vein density was expressed in mm mm^{-2}.

Samples for TEM were processed by the Multi-Imaging Centre in the Department of Physiology, Development and Neuroscience (Cambridge). Sections of ~1 mm thickness of all six stages along a developing leaf were used for embedding. Tissues were fixed in 4% (w/v) glutaraldehyde in 0.1 M HEPES buffer with a pH of 7.4 for 12 h at 4 °C. Subsequently they were rinsed in 0.1 M HEPES buffer five times, and then treated with 1% osmium ferricyanide at room

temperature for 2 h and rinsed in deionized water five times before being treated with 2% (w/v) uranyl acetate in 0.05 M maleate buffer with a pH of 5.5 for 2 h at room temperature. They were rinsed again in deionized water and dehydrated in an ascending series of ethanol solutions from 70% to 100% (v/v). This was followed by treatment with two changes of dry acetonitrile and infiltration with Quetol 651 epoxy resin. Sections 50–70 nm thick were cut with a Leica Ultracut UCT and stained with lead citrate and uranyl acetate. Images were taken with a Tecnai G2 transmission electron microscope (FEI, Hillsboro, USA) and operated at 120 kV using an AMT XR60B digital camera (Advanced Microscopy Techniques, Woburn, USA) running Deben software (Deben UK Limited, Bury St. Edmunds, UK). Images were taken at ×1700, ×3500, and ×5000 magnification.

RNA extraction and sequencing

Deep sequencing was carried out to analyse transcriptomes associated with the leaf developmental gradient defined by the anatomical analysis. Opposite leaves from the third leaf pair after the cotyledons were harvested in the morning between 10:00 h and 11:00 h (4 h after the start of the photoperiod in the glasshouse). The leaves were harvested when they had reached 1.8–2 cm length (from base to tip). Leaves were measured on an ice-cold glass plate and cut into six portions of equal length (3–3.3 mm). Each sample was then immediately placed into liquid nitrogen. Sections of 12 leaves were collected before extraction of RNA using the Qiagen Plant RNeasy Mini Kit. The optional DNA digestion step with RNase-free DNase was performed. Three biological replicates were analysed for each of the four *Flaveria* species. Total leaf RNA was sent to the Genomics & Transcriptomics Laboratory (GTL) in Düsseldorf (Germany) for paired-end sequencing. RNA samples were prepared following the TrueSeq RNA sample preparation v2 guide (Revision F) and sequenced using an Illumina/Solexa HiSeq 2000 machine. Prior to assembly, reads were trimmed using Trimmomatic v0.32 (Bolger *et al.*, 2014), corrected using BayesHammer v2.5.1 (Nikolenko *et al.*, 2013), and then digitally normalized using the khmer package (Crusoe *et al.*, 2015). SOAPdenovo-trans (Xie *et al.*, 2014), IDBA_tran v1.0.13 (Peng *et al.*, 2013), and SGA v0.10.12 (Simpson and Durbin, 2012) were used for *de novo* assembly. For each species, the individual assemblies were combined and clustered using CD-HIT-EST (Simpson and Durbin, 2012; Li and Godzik, 2006). Annotation of contigs used Conditional Reciprocal Best Blast (github.com/cboursnell/crb-blast) to the *Arabidopsis thaliana* peptide sequences as previously described (Aubry *et al.*, 2014). For principal components analysis, *Z*-scores were calculated from transcripts per million (TPM) reads from for each gene across the four species. Expression patterns of genes of interest were analysed using a custom-made R script to extract the data of interest and to visualize expression patterns. Gene Ontology (GO) term analysis was performed according to Burgess *et al.* (2016).

Results and Discussion

Leaf maturation in C_3 and C_4 Flaveria *species*

We first confirmed that fully expanded mature leaves of C_3 *F. pringlei* and *F. robusta* as well as C_4 *F. bidentis* and *F. trinervia* showed characteristic C_3 and C_4 anatomy under our growth conditions. In all cases, the third leaf pair was chosen as the first and second leaf pairs have previously been described as juvenile (McKown and Dengler, 2009). Mature third leaves of C_3 and C_4 *Flaveria* ranged from 6 cm to 10 cm in length, but the C_3 species had narrower leaves with an entire margin whereas the C_4 leaves were wider and the margin was slightly dentate (Fig. 1A). Both C_4 species had closer veins than the C_3 species (Fig. 1B; Supplementary Fig. S2). In addition,

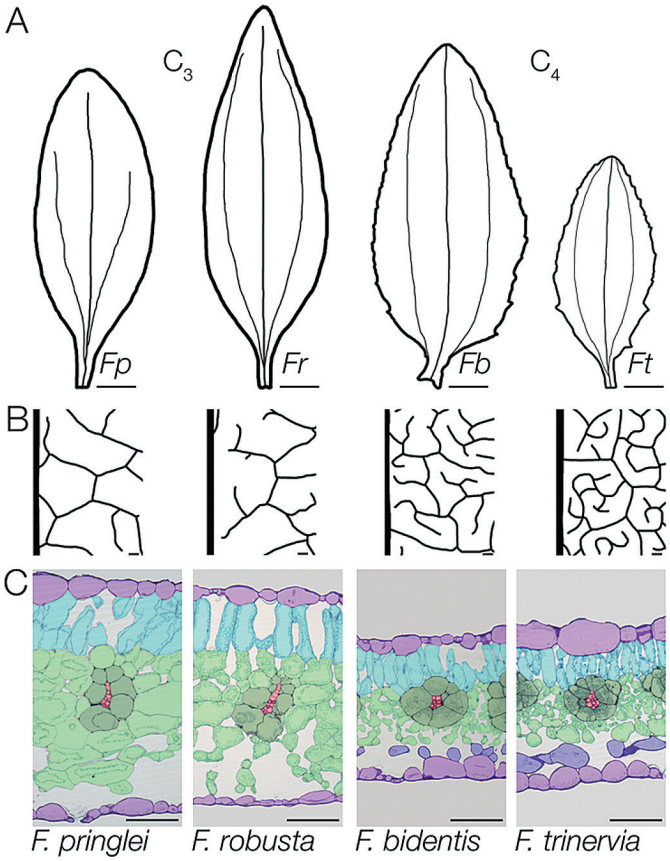

Fig. 1. Characteristics of mature leaves from C_3 and C_4 species of *Flaveria*. (A) Representative outlines of mature leaves showing the variation between species and the dentate nature of the two C_4 species. (B) Vein traces taken in the middle of the leaf to the right of the mid-vein indicating denser venation in mature leaves of both C_4 species. (C) Transverse sections showing the reduced mesophyll cross-sectional area and the closer veins with more symmetric bundle sheaths in the C_4 compared with the C_3 species. Each cell type is false colour-coded: veins (red), bundle sheath (dark green), spongy mesophyll (light green), palisade mesophyll (turquoise), parenchyma layer with no or few chloroplasts (blue), and epidermis (lilac). Species abbreviations are as follows: *Flaveria pringlei* (*Fp*), *Flaveria robusta* (*Fr*), *Flaveria bidentis* (*Fb*), *Flaveria trinervia* (*Ft*). Scale bars represent 1 cm (A) or 100 μm (B and C).

analysis of transverse leaf sections indicated that C_3 leaves were thicker because of larger cells and more cell layers (Fig. 1C). Clear differences in BS and M cell arrangement were visible between the C_3 and C_4 pairs, with the BS in both C_4 species being more uniform in shape than in the C_3 species (Fig. 1C). In addition, compared with the C_3 species, M cells in the C_4 leaves were smaller (Supplementary Fig. S2) and appeared to show increased contact with the BS. In agreement with previous analysis (McKown and Dengler, 2007), leaves of the two C_4 and two C_3 species typically contained five and eight ground cell layers between the adaxial and abaxial epidermis, respectively.

In all four *Flaveria* species, leaves of 1.8–2 cm length showed a basipetal maturation gradient with differentiating tissues at the base and fully differentiated tissues at the tip (Fig. 2A–C; Supplementary Fig. S3). This base-to-tip maturation programme is typical of dicotyledons, but it is notable that in *Flaveria* maturation occurred in larger leaves than in

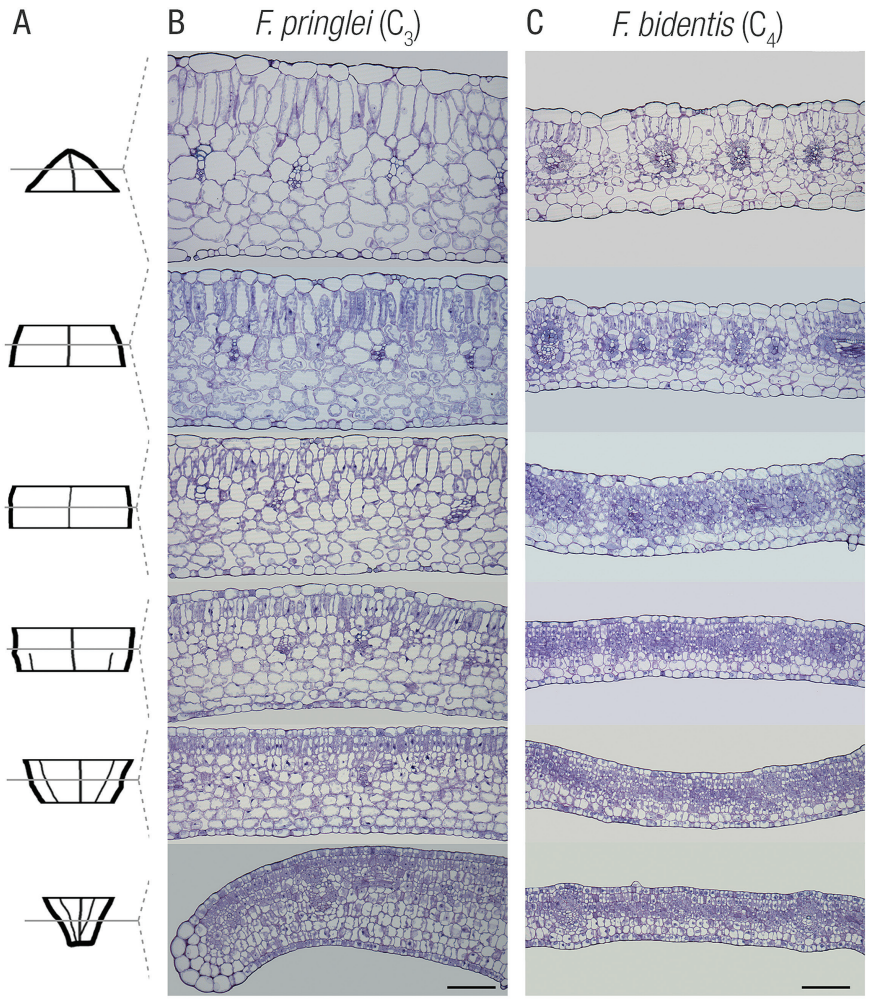

Fig. 2. Within-leaf sampling indicates the gradual maturation gradient in leaf anatomy for C_3 and C_4 *Flaveria* speciecs. (A) Representative leaf outline illustrating leaf sampling. Representative transverse sections from C_3 *Flaveria pringlei* (B) and C_4 *Flaveria bidentis* (C) from base (bottom) to tip (top). Note the gradual expansion of cells, increased vacuolization, and clearer delineation of both mesophyll and bundle sheath cells from base to tip. Scale bars represent 100 μm.

Gynandropsis gynandra, where the maturation gradient was detected in leaflets of 0.4 cm length (Aubry *et al.*, 2014). This prolonged development of leaf maturation in *Flaveria* thus provides an excellent system to analyse the induction of C_4 photosynthesis as the larger leaf allows these leaves to be divided into more stages for functional analysis.

To better understand differences in leaf maturation in C_3 and C_4 *Flaveria*, leaves from each species were divided into six portions, and RNA was extracted, quantified, integrity determined (Supplementary Fig. S4) and then subjected to deep sequencing. On average, 20 million reads were recovered from each replicate (Supplementary Table S1). After *de novo* assembly of these reads, the average number of annotated transcripts per species was 12 475 (Supplementary Table S1).

Read mapping was used to quantify transcript abundance in TPM (Supplementary Data S1). By grouping the two base, mid, and tip stages, it was possible to analyse transcript behaviour quickly across the maturation gradient. This revealed that on average 40% of annotated transcripts showed descending behaviours, meaning that they were expressed more strongly at the base of the leaf than at the tip (Supplementary Table S1). Using the 8000 annotations found in all four species, correlation analysis showed that patterns of gene expression first clustered by species and then by photosynthetic type, meaning that the correlation between neighbouring developmental stages was highest within species and that the correlation was higher between species of the same photosynthetic type than between species of different photosynthetic types (Supplementary Fig. S5). A gradient from base to tip was clear in all four species, but was more pronounced in the two C_4 species (Supplementary Fig. S5).

The six portions from the base to the tip of the leaf showed a clear induction of genes known to encode components of the C_4 cycle (Fig. 3). An important step in the evolution of C_4 photosynthesis is thought to be a co-ordinate alteration in expression of genes encoding the photorespiratory cycle and nitrogen metabolism (Mallman *et al.*, 2014). Consistent with a reduction in photorespiration in the C_4 species, genes of photorespiration increased along the developmental gradient in C_3 species but to a lesser extent in the C_4 *Flaveria* species. The main exception was *DIT1*, a transporter gene which is involved in the C_4 cycle and is more abundant in both C_4

species (Supplementary Data S1). Around 200 genes are consistently up-regulated in the upper tip tissue using C_4 compared with C_3 photosynthesis (Fig. 4A; Supplementary Data S2) and, in young tissue that is still undergoing differentiation, the number of genes up-regulated in the C_4 species is around double this number. These estimates are significantly lower than those reported from analysis of whole leaves at various developmental stages from C_3 *Tarenaya hassleriana* and C_4 *G. gynandra* in the Cleomaceae (Külahoglu *et al.*, 2014), where between 2500 and 3500 genes were estimated to be up-regulated in the C_4 species. The reduction in differential gene expression in the C_4 compared with C_3 *Flaveria* species in this study compared with the estimates from the Cleomaceae may be due to the use of two C_3 and two C_4 species, removing species-specific differences, but also to the reduced phylogenetic distance between the *Flaveria* species (Sage *et al.*, 2011). Principal components analysis indicated that the first two dimensions contributed to 37% and 19% of the variation, respectively, and that whereas the first dimension was associated with the leaf maturation signal from all species, the second dimension was associated with the photosynthetic pathway being used (Fig. 4B). This finding implies that C_4 metabolism has a large impact on the differences in gene expression associated with C_3 and C_4 leaves within one genus. When MapMan categories associated with the top decile of differentially expressed genes were assessed it was notable that photosynthesis genes were over-represented for a prolonged period along the gradient in the C_4 species (Fig. 4C). This was also the case for genes related to DNA and transport. In contrast, the MapMan category associated with RNA was over-represented for longer in the C_3 species (Fig. 4C). Equivalent analysis for GO terms indicated that cell wall loosening was over-represented for longer in both C_4 species compared with the C_3 species (Supplementary Fig. S6). Overall, these data indicate that the maturation of *Flaveria* leaves captures changes in transcript abundance associated with induction of C_3 or C_4 photosynthesis, and provide insight into the broader alterations in gene expression. We next aimed to relate the underlying anatomical changes associated with leaf maturation in these C_3 and C_4 species of *Flaveria* to changes in gene expression determined from the deep sequencing.

Kranz anatomy during leaf maturation

Using cleared leaves, veins were traced to visualize their development. In the most basal part of the leaf, major veins were present but, in the two C_4 species, the highest order veins were still being laid down. Some of these developing higher order veins could be detected by the specific patterns of periclinal cell divisions associated with their production (Supplementary Fig. S7). Vein density was determined in each of the six sections along the leaf developmental gradient, and this showed markedly different developmental patterns in the C_3 and C_4 pairs of *Flaveria* (Fig. 5A). Vein density was lower at the base in C_4 compared with C_3 leaves of *Flaveria*, but, during the transition from the base to the middle of the leaf, density increased dramatically in the C_4 species (Fig. 5B). In both C_3 and C_4 pairs, vein density then decreased from the middle to the tip of the leaf. The reduction in vein density as leaves matured is probably associated with expansion of M and BS cells. In the two most basal sections of the leaf, cell division was still ongoing and procambial strands were initiated first with a periclinal and then an anticlinal division, to derive BS cells (Supplementary Fig. S7). These basal portions of the leaf therefore appear most interesting to interrogate for candidate genes associated with these processes. The *de novo*

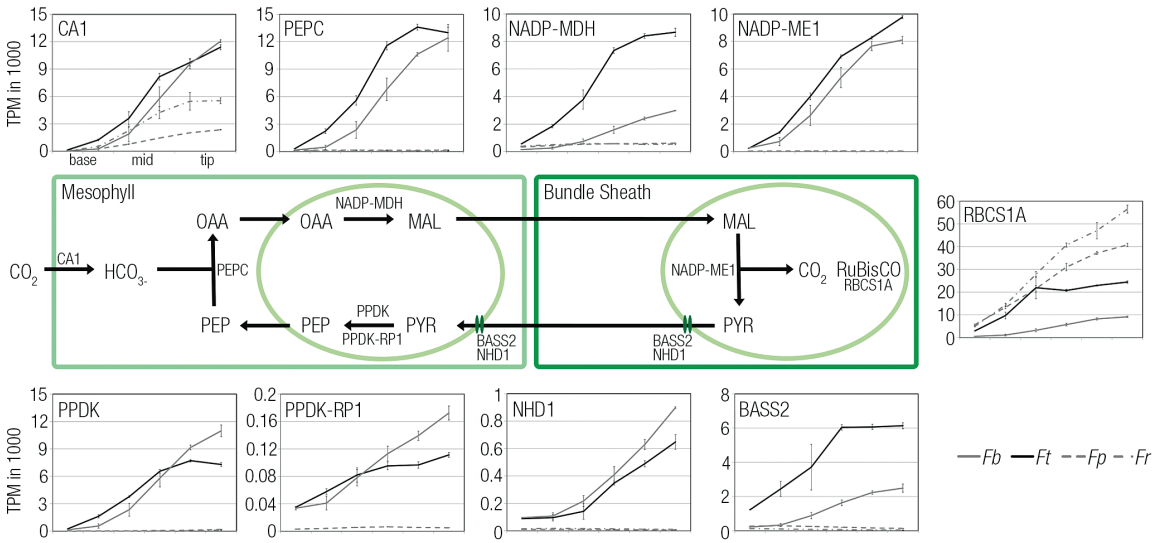

Fig. 3. Transcripts encoding the C_4 cycle increase dramatically during leaf maturation of the C_4 species. The schematic of major components of the C_4 pathway located in mesophyll and bundle sheath cells of C_4 leaves is surrounded by plots of transcript abundance for key genes of the NADP-ME subtype in transcripts per million (TPM) for the four species. *RBCS1A* transcripts are, as expected, higher in the C_3 species. All C_4 transcripts other than *CARBONIC ANHYDRASE1* (*CA1*) are barely detectable in the C_3 species. *PEPC1*, *PHOSPHOENOLPYRUVATE CARBOXYLASE 1*; *NADP-MDH*, *NADP-DEPENDENT MALATE DEHYDROGENASE*; *NADP-ME1*, *NADP-DEPENDENT MALIC ENZYME1*; *PPDK*, *PYRUVATE,ORTHOPHOSPHATE DIKINASE; PPDK-RP1*, *PYRUVATE,ORRTHOPHOSPHATE DIKINASE REGULATORY PROTEIN*; *NHD1*, *SODIUM:HYDROGEN ANTIPORTER*; *BASS2*, *BILE ACID SODIUM SYMPORTER FAMILY2.* The two C_4 species are shown with solid lines, and the two C_3 species with dashed lines.

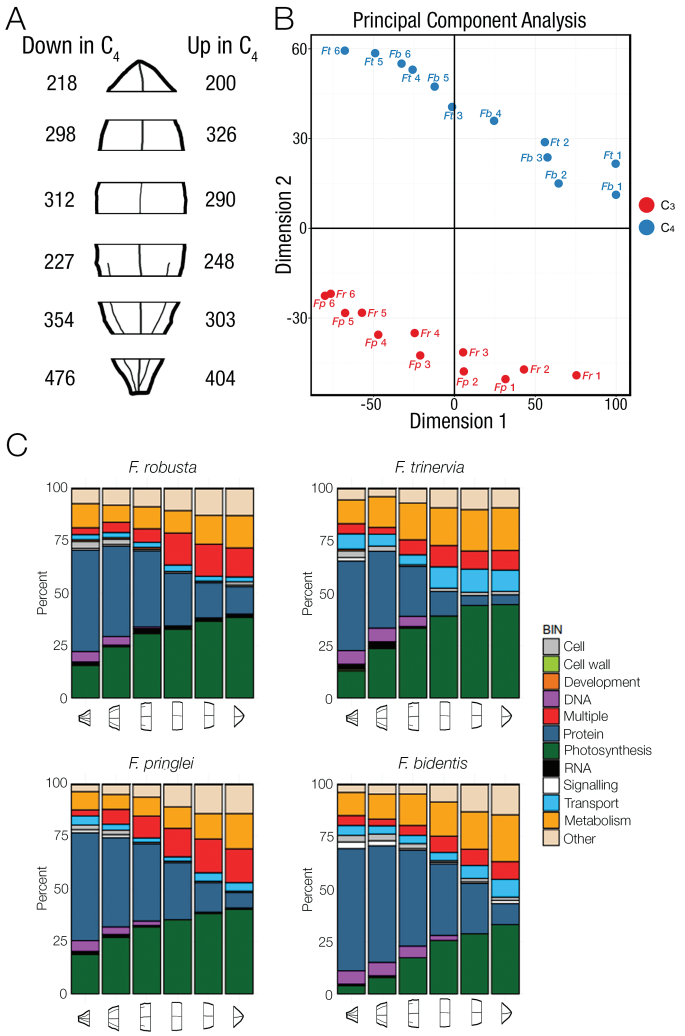

Fig. 4. Overview of the global changes in gene expression along the C_3 and C_4 *Flaveria* maturation gradients. (A) RNA-seq was conducted on six regions of *Flaveria* leaves. The numbers of differentially expressed genes in each section of the leaf are depicted to the left (up in both C_3 *F. pringlei* and *F. robusta*) and to the right (up in both C_4 *F. bidentis* and *F. trinervia*) of the leaf schematic. (B) Principal components analysis of the differentially expressed genes in all four species shows that the first dimension is associated with leaf maturation, but the second dimension is associated with the photosynthetic pathway. C_3 species are depicted in red, C_4 species in blue, and each stage is numbered from 1 (base) to 6 (tip). (C) Summary of MapMan categories associated with the top decile of differentially expressed genes in each species. *Fp*, *Flaveria pringlei*; *Fr*, *Flaveria robusta*; *Fb*, *Flaveria bidentis*; *Ft*, *Flaveria trinervia*.

assembled transcriptomes from the four *Flaveria* species were therefore analysed for homologues of genes known to impact on vein formation in others species.

Six genes that effect vein formation in *A. thaliana* showed different behaviours in the C_4 *Flaveria* species compared with the C_3 species (Fig. 6). In all cases, absolute transcript abundance tended to be higher at the base of the C_4 leaves compared with the base of the C_3 species. However, the most consistent differences in expression were found for *A. thaliana HOMEOBOX-GENE-8* (*ATHB8*). *SHR* and *SCR* have been implicated in the development of C_4 Kranz anatomy (Slewinski *et al.*, 2012; Slewinski, 2013). *SHR* was detected in three of the species with a descending pattern, but with no clear

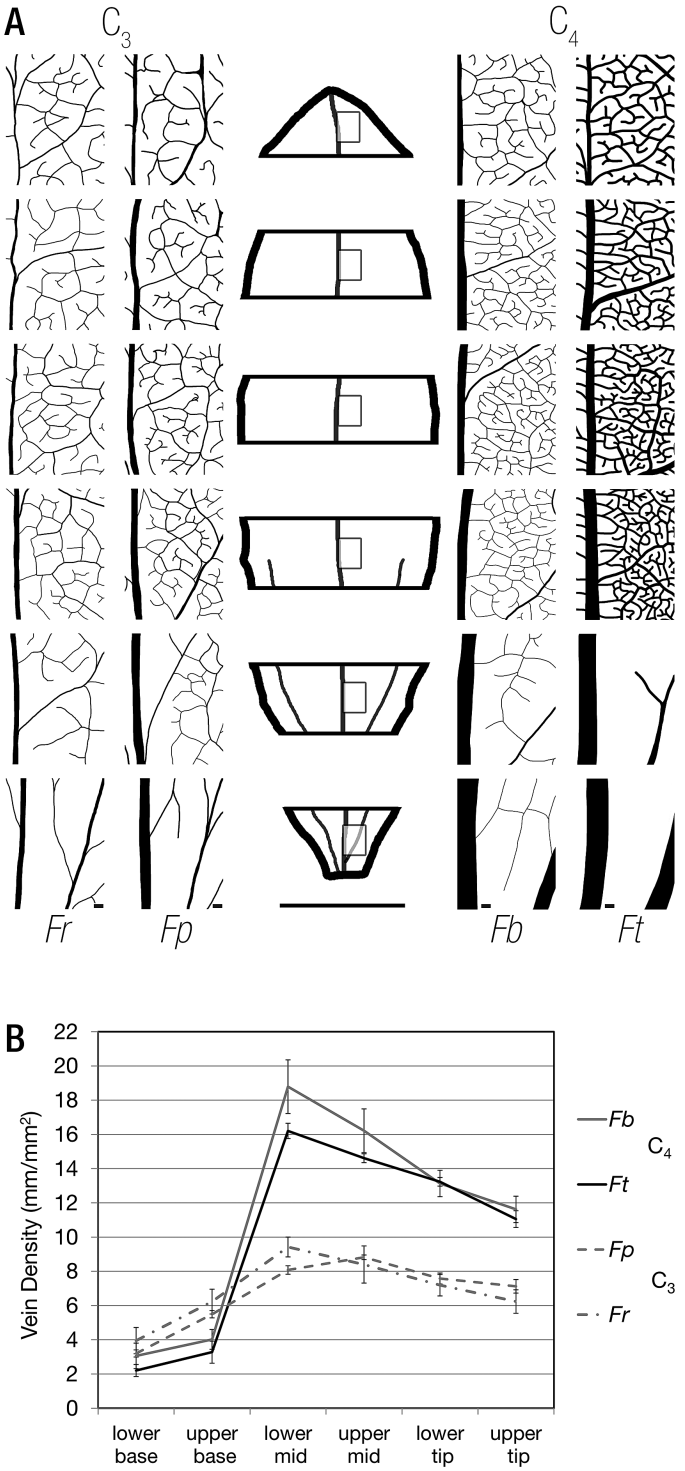

Fig. 5. Vein density increases dramatically between the upper base and lower mid leaf sections of the C_4 species compared with the C_3 species. (A) Representative vein traces from base to tip of all four species illustrating the maturation of vein density. *Fp*, *Flaveria pringlei*; *Fr*, *Flaveria robusta*; *Fb*, *Flaveria bidentis*; *Ft*, *Flaveria trinervia*. The leaf outline in the centre indicates that vein density measurements were obtained from the centre of each section to the right of the midrib. (B) Quantitation of vein density in each section for the four species. In all species, vein density was highest in the middle and decreased towards the tip of leaves. A rapid increase in vein density occurs between the upper base and lower mid in both C_4 species. Error bars represent 1 SE. Scale bars represent 100 μm for the vein traces and 0.5 cm for the leaf outline.

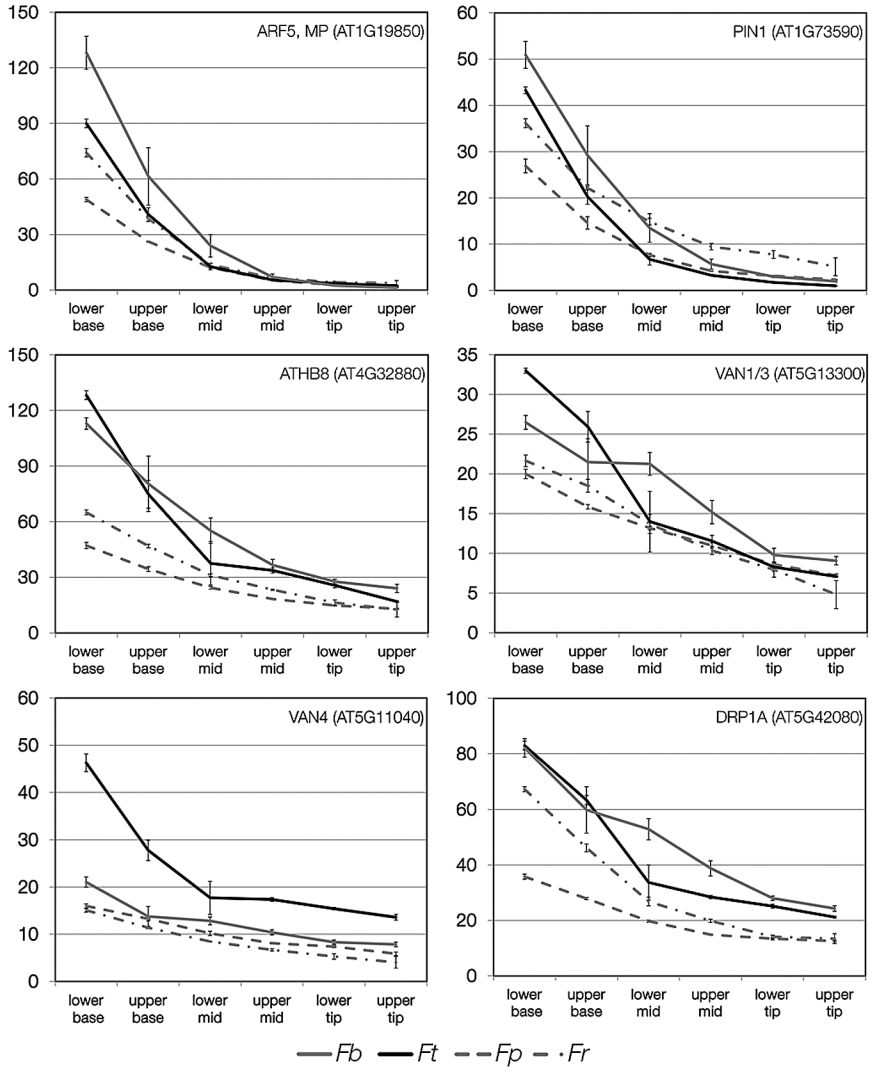

Fig. 6. Patterns of transcript abundance in the four *Flaveria* species for genes known to be involved in vein formation. Transcript abundance is shown in transcripts per million (TPM), and the parts of the leaf are annotated on the *x*-axis. The two C_4 species are shown with solid lines, and the two C_3 species with dashed lines. *Fp*, *Flaveria pringlei*; *Fr*, *Flaveria robusta*; *Fb*, *Flaveria bidentis*; *Ft*, *Flaveria trinervia*.

difference between C_3 and C_4. *SCR* also showed a descending pattern, but again with no clear difference between photosynthetic types. Of Scarecrow-like genes, only *SCARECROW-LIKE 14* was clearly higher in C_4 and showed a parabolic expression pattern, being most highly expressed at the base and the tip (Supplementary Data S3). *SCARECROW-LIKE 14* was recently identified from cross-species selection scans in the grasses (Huang *et al.*, 2016). Transcripts encoding the auxin response factors ARF3, ARF8, and also IAA7 showed differences in the timing of expression in C_4 compared with C_3 species (Supplementary Data S3).

The same six regions from leaf base to tip used to define vein density were next used to quantify maturation of BS and M cells (Fig. 7). Leaf thickness was higher in the C_3 compared with the C_4 species, and cell expansion continued for longer in the C_3 leaves (Fig. 7A). The BS in the C_3 species was less regular than in the C_4 species, and individual cell size within the BS was more variable (Fig. 7A). However, it was notable that the cross-sectional area of BS cells in the C_3 species was larger than that of C_4 species, particularly towards the tip. It has been proposed that a large BS cell size is a key early event associated with the evolution of C_4 photosynthesis (Christin *et al.*, 2013; Griffiths *et al.*, 2013; Williams *et al.*, 2013), and so it appears that within *Flaveria* this is a pre-existing trait. It was noticeable that the cross-sectional area of M cells from the two C_3 species was around five times greater than that of M cells of the C_4 species (Fig. 7B). This strongly implies that compared with C_3 species, the reduced leaf depth of C_4 species is associated with an inhibition of M cell expansion as well as a reduced number of cell layers. A number of genes that have been annotated as having roles in cell proliferation or expansion showed behaviours that may explain the reduced cell expansion in leaves of both C_4 species. Most notable was *DWF4*, a 22α-hydroxylase that catalyses the rate-limiting step of brassinosteroid synthesis and so controls cell expansion. In both C_3 species, *DWF4* transcript abundance increased along the leaf gradient but in the C_4 species its transcripts were barely detectable (Supplementary Data S4). These data imply that *DWF4* is a strong candidate for the reduced expansion associated with maturation of the C_4 leaf. In sorghum,

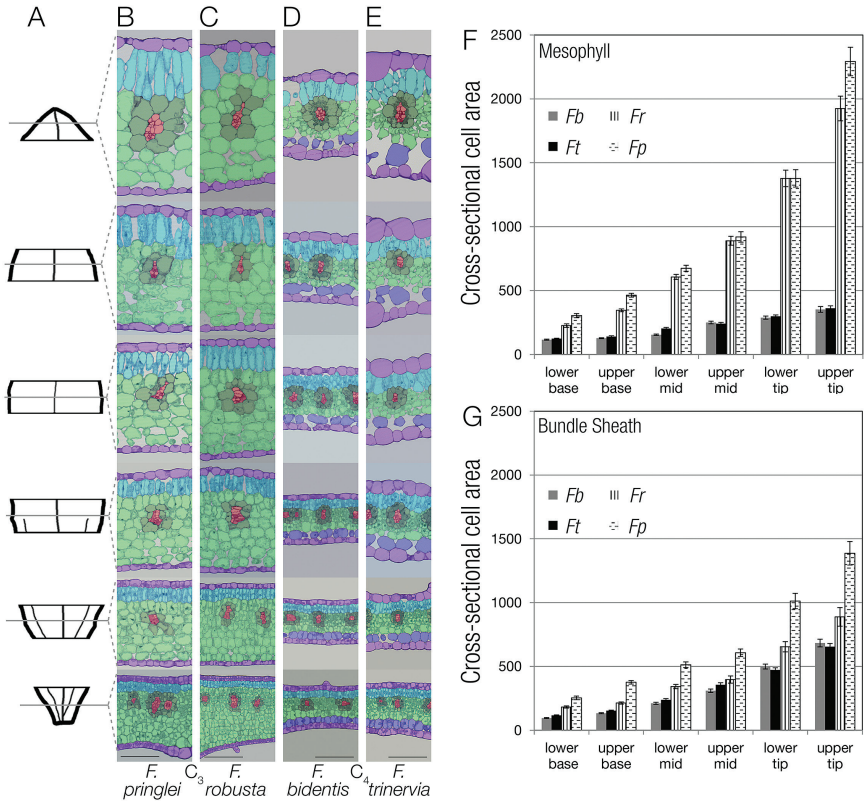

Fig. 7. During leaf maturation mesophyll cell expansion is reduced in the C_4 compared with the C_3 species. (A) Representative leaf outline illustrating leaf sampling. (B–E) Representative transverse sections from the base to tip of leaves from C_3 *F. pringlei* (B) and *F. robusta* (C), as well as *F. bidentis* (D) and *F. trinervia* (E). (F, G) Quantification of cross-sectional cell areas for mesophyll (F) and bundle sheath (G) cells along the leaf maturation gradient. Each cell type is false colour-coded: veins (red), bundle sheath (dark green), spongy mesophyll (light green), palisade mesophyll (turquoise), parenchyma layer with no or few chloroplasts (blue), and epidermis (lilac). *Fp*, *Flaveria pringlei*; *Fr*, *Flaveria robusta*; *Fb*, *Flaveria bidentis*; *Ft*, *Flaveria trinervia*. Scale bars (B–E) represent 100 μm.

mutation of CYP450D2, a Cyt P450 involved in brassinosteroid synthesis, leads to perturbed vein spacing (Rizal *et al.*, 2015). Although the orthologue to CYP450D2 is not present in *A. thaliana*, transcripts encoding CYP90D1 and CYP90C1 appear to be slightly higher in the base of the C_4 compared with the C_3 *Flaveria* species (Supplementary Data S4).

A prolonged rate of cell division in the C_4 *Flaveria* species is also supported by transcript profiles. A number of genes implicated in cell division are found at higher levels in basal sections of C_4 than C_3 leaves including *PLE* (Müller *et al.*, 2004), *MOR1* (Whittington *et al.*, 2001), *ATK1* (Marcus *et al.*, 2003), and *ATK3* (Mitsui *et al.*, 1994) which are involved in microtubule organization, and *SCC3* (Chelysheva *et al.*, 2005) and *SYN1* (Cai *et al.*, 2003) which are implicated in sister chromatid pairing (Supplementary Data S4). Transcripts derived from genes associated with DNA replication and repair including *SOG1* (Yoshiyama *et al.*, 2009), *TOP6B* (Gilkerson and Callis, 2014), *MutS* homologues *MSH* (*MSH3*, *MSH4*, and *MSH7*) (Culligan and Hays, 2000), and multiple components of the mini-chromosome maintenance (MCM) complex (*MCM2*, *MCM3*, *MCM5*, and *MCM6*) (Tuteja *et al.*, 2011) were also more abundant in the C_4 species at the base of the leaf (Supplementary Data S4). Good candidate regulators of cell division include MYB3R4 which forms a complex with E2FB and *RETINOBLASTOMA-RELATED PROTEIN* (RBR1) to activate cell cycle progression (Haga *et al.*, 2011), and *DEL1* which maintains cell division by repressing entry into the endocycle (Vlieghe *et al.*, 2005), both of which are higher in the base of C_4 leaves and maintain expression longer into the leaf gradient.

A reduction along the leaf gradient in the abundance of transcripts encoding CPP SYNTHASE 1 (CPS1), which catalyses the first step of gibberellin synthesis, and a spike in transcripts of *GA2ox2* (Supplementary Data S4), which degrades gibberellin, between the upper base and lower mid in C_4 leaves is in keeping with findings that a decrease in gibberellin levels controls the transition between cell division and expansion in maize (Nelissen *et al.*, 2012). A corresponding pattern cannot be seen in C_3 leaves, which may mean that the transition between cell division had already occurred at the base of C_3 lineages, or is mediated by another mechanism. We did not undertake analysis of sink-to-source transitions within the *Flaveria* leaves. However, as sugar signalling impacts on the transition from cell division to expansion, it is possible that the altered maturation of C_4 compared with C_3 leaves is related to sink-to-source transition as well as the C_4 paradigm *per se* (Li *et al.*, 2010).

Chloroplast maturation within the leaf gradient

Species that primarily use NADP-ME to decarboxylate malate in the C_4 BS commonly develop dimorphic chloroplasts with

M cells containing stacked thylakoids (grana) and the BS cells containing fewer grana (Edwards and Walker, 1983; Edwards and Voznesenskaya, 2011). To provide insight into the dynamics of chloroplast development and maturation in C_3 and C_4 *Flaveria* species, TEM was used to investigate chloroplast structure in each of the six stages along the leaf gradient. Dimorphic chloroplasts were observed in the C_4 species but not in the C_3 species (Fig. 8A–D). Towards the base of the C_4 leaves, both BS and M chloroplasts contained 2–3 granal lamellae per stack (Supplementary Fig. S8). In more mature parts of the leaf, levels of stacking were reduced in BS cells but increased in M cells (Fig. 8A–D). In both C_3 species, a developmental gradient in granal stacking was also visible from base to tip, with more mature parts of the leaf showing increased levels of granal stacking in both cell types (Supplementary Fig. S8). Thus, a key difference between the C_3 and C_4 leaves was the reduction in granal stacking in C_4 BS cells from the middle of the leaf.

The control of photosynthesis gene expression and chloroplast development has previously been linked to a pair of transcription factors known as GOLDEN-LIKE 1 and 2 (GLK1 and GLK2) (Supplementary Data S5). Unfortunately, *de novo* assembly did not generate a complete set of contigs for either GLK1 or GLK2 from all four species, so we were unable to determine if their abundance correlated with the observed changes in chloroplast development. However, other candidate genes that can be associated with chloroplast maturation were identified. For example, cytokinin is known to play a role in chloroplast maturation, and increased abundance of transcripts encoding the cytokinin receptors AHK2 and AHK3, along with the downstream effector ARR1 in C_4 species suggests that enhanced cytokinin signalling may play a role in these alterations to the C_4 chloroplast. Additionally, it was also notable that the behaviour of two nuclear-encoded chloroplast RNA polymerases (RpoTp and RpoTmp) (Swiatecka-Hagenbruch *et al.*, 2008) showed clear differences between the C_3 and C_4 *Flaveria* species. Transcripts encoding RpoTp, which controls *ycf1* that is involved in plastid protein import, were very low in both C_4 species. Further indications to changes in chloroplast protein import related to components of the TIC/TOC complex (Supplementary Data S5). Transcripts encoding the outer envelope receptor proteins TOC34 and TOC159 were higher in the base and peaked later in the C_4 than the C_3 species. Additionally, TOC75, the major channel constituent for protein import into the chloroplast (Li and Chiu, 2010), was lower at the base of C_4 leaves. These data would be consistent with protein import playing a role in the establishment of dimorphic chloroplasts. Chloroplast maturation also correlated with increased abundance of nuclear-encoded chloroplast sigma factors, which direct the nuclear-encoded chloroplast RNA polymerase (NEP) to promoters.

Previous studies have demonstrated that C_4 plants alter photosystem activity to adjust ADP/ATP and NADPH ratios to ensure proper functioning of the carbon pump. Consistent with this was the C_4-specific up-regulation of SIG3, which initiates transcription of *psbN* (Zghidi *et al.*, 2007) that is involved in repair of PSII complexes (Torabi *et al.*, 2014). It was also notable that transcripts encoding components of the

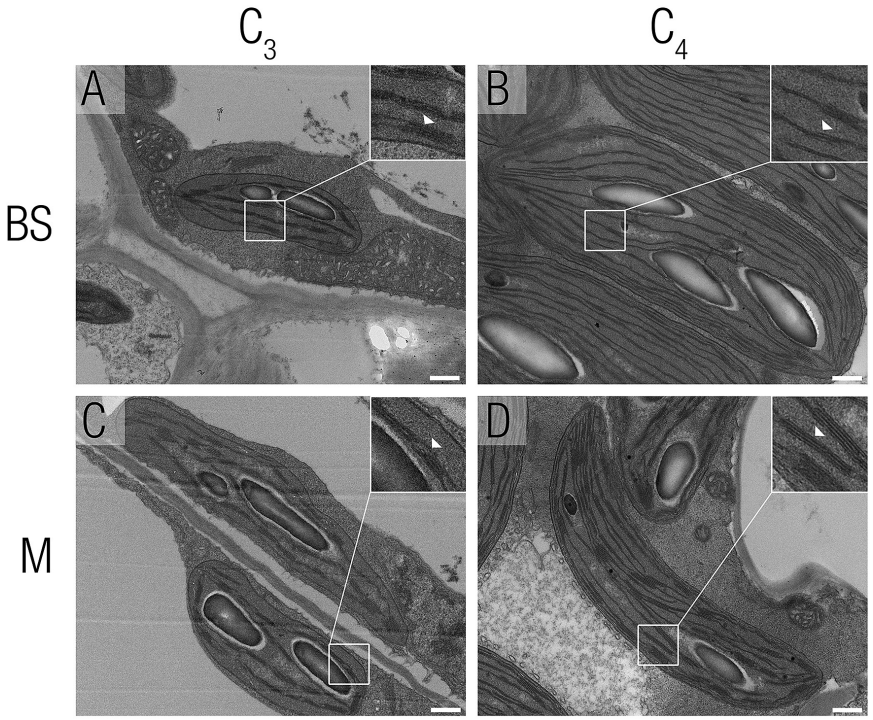

Fig. 8. Chloroplast ultrastructure of mesophyll and bundle sheath cells from C_3 *F. pringlei* and C_4 *F. bidentis.* Transmission electron microscopy was used to investigate chloroplast ultrastructure. (A, B) Bundle sheath chloroplasts from C_3 *Flaveria pringlei* and C_4 *Flaveria bidentis*. (C, D) Mesophyll chloroplasts from C_3 *F. pringlei* and C_4 *F. bidentis*. Little granal stacking was seen in bundle sheath chloroplasts from *F. bidentis* (B) whereas it could be observed in C_3 bundle sheath and in all mesophyll chloroplasts. Insets show a close-up of thylakoids, with white arrowheads indicating thylakoid stacks (A, C, D) or the lack of extensive stacking (B). Scale bars represent 500 nm.

NDH complex (*NDF1*, *NDF6*, and *PIFI*), as well as genes involved in cyclic electron transport (*PGR5* and *PGRL1A*), were more abundant in both C_4 species, and these differences became more apparent from base to tip (Supplementary Data S5). Chloroplast dimorphism also coincided with an increase in transcript abundance of *CRR1*, which is proposed to be involved in NDH complex assembly.

Analysis of electron micrographs also revealed that chloroplast length in the BS increased from base to tip in the C_4 species more than in the C_3 species (Fig. 9). Two members of the *REDUCED CHLOROPLAST COVERAGE* (*REC*) gene family, *REC2* and *REC3*, knockout of which causes a reduction in chloroplast size (Larkin *et al.*, 2016), were up-regulated in the C_4 species. Further, *FtsZ1-1*, which is involved in chloroplast division, was significantly down-regulated in the C_4 compared with the C_3 species (Supplementary Data S5). As reduced division of chloroplasts leads to increased size (Pyke and Leech, 1994; Pyke, 1997), the reduced expression of chloroplast division genes in C_4 plants could lead to their increased size in BS cells.

Convergence in transcript abundance between independent C_4 lineages

By combining our data from the four *Flaveria* species with publicly available data sets, we next sought to investigate whether separate lineages of C_4 plants shared common changes in transcript abundance compared with their C_3 congeners. A recent study compared transcriptomes derived from leaf developmental gradients of *T. hassleriana* (C_3) and *G. gynandra* (C_4) from the Cleomaceae (Külahoglu *et al.*, 2014). In contrast to our work, whole leaves were staged by age (Külahoglu *et al.*, 2014) and changes in transcript abundance compared with non-photosynthetic tissues such as roots. This analysis identified a set of 33 genes that could be annotated with unique *A. thaliana* identifiers. This set of genes showed root expression in the C_3 species *T. hassleriana* but leaf expression in the C_4 species *G. gynandra* (Külahoglu *et al.*, 2014), and so it was proposed that these were key genes that neofunctionalize to become involved in C_4 photosynthesis. Of these 33 genes, homologues to 15 were detected in all four *Flaveria* species, but only three showed up-regulation in the C_4 compared with C_3 species in our study and three showed higher expression in the two C_3 species (Supplementary Data S6).

Previously, transcript abundance in developing leaves of the C_4 dicot *G. gynandra* (Aubry *et al.*, 2014) has been compared with that in the C_4 monocot maize (Li *et al.*, 2010; Pick *et al.*, 2011). Despite the very wide phylogenetic distance between these species, 18 transcription factors that showed the same patterns of behaviour in both C_4 *G. gynandra* and C_4 maize were identified (Aubry *et al.*, 2014). Ten of these

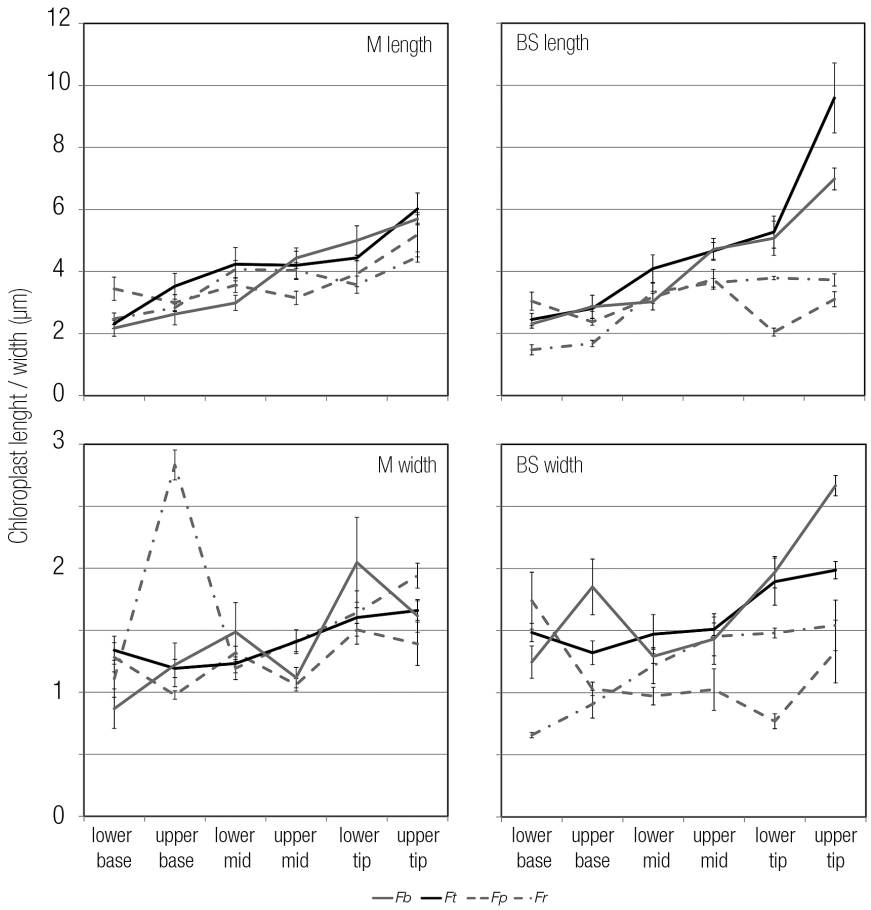

Fig. 9. Quantitation of chloroplast length and width from mesophyll and bundle sheath cells. Length and width were taken from thin sections used in TEM. For each species, five chloroplasts were measured per stage and cell type. Error bars represent 1 SE. The two C_4 species are shown with solid lines, and the two C_3 species with dashed lines.

18 candidates were also detected in all four *Flaveria* species. Five of those showed ascending behaviour and higher transcript abundance in the two C_4 species. These were RAD-LIKE6 (RL6, AT1G75250); AT2G05160, a CCCH-type zing finger family protein with an RNA-binding domain; AT2G21530, a SMAD/FHA domain-containing protein; RNA Polymerase Sigma Subunit C (SIGC, AT3G53920), and BEL1-Like Homeodomain1 (BLH1, AT5G67030) (Supplementary Data S7). The five genes have therefore been shown to exhibit the same behaviour along leaf developmental gradients in C_4 species in the monocots and dicots [both rosids (Cleomaceae) and asterids (Asteraceae)]. This strongly supports the notion that independent lineages of C_4 plants have recruited homologous *trans*-factors to induce the C_4 system.

Summary and conclusions

Most previous analysis derived from comparisons of RNA-seq of C_3 and C_4 species have relied on either sampling one congeneric C_3 and C_4 species pair, or analysis of mature leaf tissues. The number of genes that are differentially expressed between any two species can be very high simply due to species-specific differences, which do not relate to their photosynthetic pathway. In this work, we identify genes that are consistently differentially expressed in multiple C_3 and C_4 species from the same genus along a leaf maturation gradient. In the *Flaveria* genus, the data set therefore indicates that ~200 genes are consistently up-regulated in mature tissue using C_4 compared with C_3 photosynthesis. In young tissue that is still undergoing differentiation, the number of genes up-regulated in the C_4 species is around double this number. Leaf maturation and the photosynthetic pathway were responsible for the greatest amount of variation in transcript abundance. In *Flaveria*, the evolution of C_4 photosynthesis is not associated with an expansion of BS cell size, but rather a reduction in the cross-sectional area of M cells in C_4 compared with C_3 species. Overall, the analysis therefore provides quantitation of changes in gene expression associated with C_3 and C_4 photosynthesis, candidate genes underlying the alterations in leaf characteristics associated with these pathways, and, through comparison with independent C_4 lineages outside of the Asteraceae, insight into the extent to which parallel evolution underlies the convergent and complex C_4 trait.

Supplementary data

Supplementary data are available at *JXB* online.

Fig. S1. Phylogeny of the *Flaveria* genus.

Fig. S2. Mature leaf vein density and cell size.

Fig. S3. Leaf anatomy maturation gradient in C_3 and C_4 *Flaveria*.

Fig. S4. RNA quality from sections of each species.

Fig. S5. TPM correlation matrix.

Fig. S6. GO term analysis presented as heatmaps for each *Flaveria* species.

Fig. S7. Cell division and vein development in the base samples of *Flaveria*.

Fig. S8. Chloroplast maturation in mesophyll and bundle sheath cells for each species.

Data S1. TPM values for all four species for all stages and all replicates.

Data S2. Summary of transcripts that were up-regulated in both C_3 species or both C_4 species at each stage.

Data S3. Transcript abundance of SHR, SCR, and SCR-like genes as well as auxin response genes.

Data S4. Transcript abundance of genes associated with cell division.

Data S5. Transcript abundance of genes associated with chloroplast maturation and development.

Data S6. Comparison of data derived from this study with genes identified by Külahoglu *et al.* (2014).

Data S7. Comparison of data derived from this study with genes identified by Aubry *et al.* (2014).

Table S1. Number of reads and transcripts, per stage and species.

Author contributions

BMCK and JMH designed the study. BMCK grew the plants, undertook the anatomical analysis, and isolated the RNA; RS-U and CB undertook the *de novo* transcriptome assembly and annotation of contigs; BMCJ, SJB, and IR-L analysed the RNA-seq data; BMCK, SJB, and JMH wrote the paper.

Acknowledgements

Seeds for all *Flaveria* species used were kindly provided by Udo Gowik (Universität Düsseldorf). The European Union *3to4* project and the Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/J011754/1 funded the project.

References

Aubry S, Kelly S, Kümpers BM, Smith-Unna RD, Hibberd JM. 2014. Deep evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in two independent origins of C_4 photosynthesis. PLoS Genetics **10,** e1004365.

Bauwe H, Hagemann M, Fernie AR. 2010. Photorespiration: players, partners and origin. Trends in Plant Science **15,** 330–336.

Bellasio C, Griffiths H. 2014. The operation of two decarboxylases, transamination, and partitioning of C_4 metabolic processes between mesophyll and bundle sheath cells allows light capture to be balanced for the maize C_4 pathway. Plant Physiology **164,** 466–480.

Bowes G, Ogren WL, Hageman RH. 1971. Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. Biochemical and Biophysical Research Communications **45,** 716–722.

Burgess SJ, Granero-Moya I, Grangé-Guermente MJ, Boursnell C, Terry MJ, Hibberd JM. 2016. Ancestral light and chloroplast regulation form the foundations for C4 gene expression. Nature Plants **2,** 16161.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30,** 2114–2120.

Cai X, Dong F, Edelmann RE, Makaroff CA. 2003. The Arabidopsis SYN1 cohesin protein is required for sister chromatid arm cohesion and homologous chromosome pairing. Journal of Cell Science **116,** 2999–3007.

Chelysheva L, Diallo S, Vezon D, *et al.* 2005. AtREC8 and AtSCC3 are essential to the monopolar orientation of the kinetochores during meiosis. Journal of Cell Science **118,** 4621–4632.

Christin P-A, Osborne CP, Chatelet DS, Columbus JT, Besnard G, Hodkinson TR, Garrison LM, Vorontsova MS, Edwards EJ. 2013. Anatomical enablers and the evolution of C_4 photosynthesis in grasses. Proceedings of the National Academy of Sciences, USA **110,** 1381–1386.

Culligan KM, Hays JB. 2000. Arabidopsis MutS homologs—AtMSH2, AtMSH3, AtMSH6, and a novel AtMSH7—form three distinct protein heterodimers with different specificities for mismatched DNA. The Plant Cell **12,** 991–1002.

Crusoe MR, Alameldin HF, Awad S, *et al.* 2015. The khmer software package: enabling efficient nucleotide sequence analysis. F1000Research **4,** 900.

Drincovich MF, Casati P, Andreo CS, Chessin SJ, Franceschi VR, Edwards GE, Ku MS. 1998. Evolution of C4 photosynthesis in *flaveria* species. Isoforms of NADP-malic enzyme. Plant Physiology **117,** 733–744.

Edwards GE, Voznesenskaya EV. 2011. C_4 photosynthesis: Kranz forms and single-cell C4 in terrestrial plants. In: **Raghavendra A, Sage RF**, eds. C4 photosynthesis and related CO_2 concentrating mechanisms. Advances in photosynthesis and respiration, Vol. **32**. Dordrecht, The Netherlands: Springer, 29–61.

Edwards GE, Walker D. 1983. C3, C4: mechanisms, and cellular and environmental regulation, of photosynthesis. Oxford: Blackwell Scientific Publications.

Furbank RT. 2011. Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types? Journal of Experimental Botany **62,** 3103–3108.

Gilkerson J, Callis J. 2014. A genetic screen for mutants defective in IAA1-LUC degradation in Arabidopsis thaliana reveals an important requirement for TOPOISOMERASE6B in auxin physiology. Plant Signaling and Behavior **9,** e972207.

Gowik U, Bräutigam A, Weber KL, Weber AP, Westhoff P. 2011. Evolution of C_4 photosynthesis in the genus *Flaveria*: how many and which genes does it take to make C_4? The Plant Cell **23,** 2087–2105.

Griffiths H, Weller G, Toy LF, Dennis RJ. 2013. You're so vein: bundle sheath physiology, phylogeny and evolution in C_3 and C_4 plants. Plant, Cell and Environment **36,** 249–261.

Haga N, Kobayashi K, Suzuki T, *et al*. 2011. Mutations in MYB3R1 and MYB3R4 cause pleiotropic developmental defects and preferential down-regulation of multiple G2/M-specific genes in Arabidopsis. Plant Physiology **157,** 706–717.

Huang P, Studer AJ, Schnable JC, Kellogg EA, Brutnell TP. 2017 Cross species selection scans identify components of C4 photosynthesis in the grasses. Journal of Experimental Botany **68,** 127–135.

Külahoglu C, Denton AK, Sommer M, *et al*. 2014. Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae C_3 and C_4 plant species. The Plant Cell **26,** 3243–3260.

Larkin RM, Stefano G, Ruckle ME, Stavoe AK, Sinkler CA, Brandizzi F, Malmstrom CM, Osteryoung KW. 2016. REDUCED CHLOROPLAST COVERAGE genes from *Arabidopsis thaliana* help to establish the size of the chloroplast compartment. Proceedings of the National Academy of Sciences, USA **113,** E1116–E1125.

Li HM, Chiu CC. 2010. Protein transport into chloroplasts. Annual Review of Plant Biology **61,** 157–180.

Li P, Ponnala L, Gandotra N, *et al*. 2010. The developmental dynamics of the maize leaf transcriptome. Nature Genetics **42,** 1060–1067.

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics **22,** 1658–1659.

Lyu MJ, Gowik U, Kelly S, *et al*. 2015. RNA-Seq based phylogeny recapitulates previous phylogeny of the genus Flaveria (Asteraceae) with some modifications. BMC Evolutionary Biology **15,** 116.

Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber AP, Westhoff P, Gowik U. 2014. The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*. eLife **3,** e02478.

Marcus AI, Li W, Ma H, Cyr RJ. 2003. A kinesin mutant with an atypical bipolar spindle undergoes normal mitosis. Molecular Biology of the Cell **14,** 1717–1726.

McKown AD, Dengler NG. 2007. Key innovations in the evolution of Kranz anatomy and C_4 vein pattern in *Flaveria* (Asteraceae). American Journal of Botany **94,** 382–399.

McKown AD, Dengler NG. 2009. Shifts in leaf vein density through accelerated vein formation in C_4 *Flaveria* (Asteraceae). Annals of Botany **104,** 1085–1098.

McKown AD, Moncalvo JM, Dengler NG. 2005. Phylogeny of *Flaveria* (Asteraceae) and inference of C4 photosynthesis evolution. American Journal of Botany **92,** 1911–1928.

Mitsui H, Nakatani K, Yamaguchi-Shinozaki K, Shinozaki K, Nishikawa K, Takahashi H. 1994. Sequencing and characterization of the kinesin-related genes katB and katC of *Arabidopsis thaliana*. Plant Molecular Biology **25,** 865–876.

Müller S, Smertenko A, Wagner V, Heinrich M, Hussey PJ, Hauser MT. 2004. The plant microtubule-associated protein AtMAP65-3/PLE is essential for cytokinetic phragmoplast function. Current Biology **14,** 412–417.

Nelissen H, Rymen B, Jikumaru Y, Demuynck K, Van Lijsebettens M, Kamiya Y, Inzé D, Beemster GT. 2012. A local maximum in gibberellin levels regulates maize leaf growth by spatial control of cell division. Current Biology **22,** 1183–1187.

Nikolenko SI, Korobeynikov AI, Alekseyev MA. 2013. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. BMC Genomics **14** Suppl 1, S7.

Peng Y, Leung HC, Yiu SM, Lv MJ, Zhu XG, Chin FY. 2013. IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. Bioinformatics **29,** 326–i334.

Pick TR, Bräutigam A, Schlüter U, *et al*. 2011. Systems analysis of a maize leaf developmental gradient redefines the current C4 model and provides candidates for regulation. The Plant Cell **23,** 4208–4220.

Powell AM. 1978. Systematics of Flaveria (Flaveriinae-Asteraceae). Annals of the Missouri Botanical Garden **65,** 590–636.

Pyke K. 1997. The genetic control of plastid division in higher plants. American Journal of Botany **84,** 1017.

Pyke KA, Leech RM. 1994. A genetic analysis of chloroplast division and expansion in Arabidopsis thaliana. Plant Physiology **104,** 201–207.

Rizal G, Thakur V, Dionora J, *et al*. 2015. Two forward genetic screens for vein density mutants in sorghum converge on a cytochrome P450 gene in the brassinosteroid pathway. The Plant Journal **84,** 257–266.

Sage RF, Christin PA, Edwards EJ. 2011. The C(4) plant lineages of planet Earth. Journal of Experimental Botany **62,** 3155–3169.

Schulze S, Mallmann J, Burscheidt J, Koczor M, Streubel M, Bauwe H, Gowik U, Westhoff P. 2013. Evolution of C_4 photosynthesis in the genus *flaveria*: establishment of a photorespiratory CO_2 pump. The Plant Cell **25,** 2522–2535.

Simpson JT, Durbin R. 2012. Efficient *de novo* assembly of large genomes using compressed data structures. Genome Research **22,** 549–556.

Slewinski TL. 2013. Using evolution as a guide to engineer Kranz-type C_4 photosynthesis. Frontiers in Plant Science **4,** 1–13.

Slewinski TL, Anderson AA, Zhang C, Turgeon R. 2012. Scarecrow plays a role in establishing Kranz anatomy in maize leaves. Plant and Cell Physiology **53,** 2030–2037.

Swiatecka-Hagenbruch M, Emanuel C, Hedtke B, Liere K, Börner T. 2008. Impaired function of the phage-type RNA polymerase RpoTp in transcription of chloroplast genes is compensated by a second phage-type RNA polymerase. Nucleic Acids Research **36,** 785–792.

Torabi S, Umate P, Manavski N, *et al*. 2014. PsbN is required for assembly of the photosystem II reaction center in Nicotiana tabacum. The Plant Cell **26,** 1183–1199.

Tuteja N, Tran NQ, Dang HQ, Tuteja R. 2011. Plant MCM proteins: role in DNA replication and beyond. Plant Molecular Biology **77,** 537–545.

Vlieghe K, Boudolf V, Beemster GT, *et al*. 2005. The *DP-E2F-like* gene *DEL1* controls the endocycle in Arabidopsis thaliana. Current Biology **15,** 59–63.

Wang Y, Bräutigam A, Weber AP, Zhu XG. 2014. Three distinct biochemical subtypes of C4 photosynthesis? A modelling analysis. Journal of Experimental Botany **65,** 3567–3578.

Whittington AT, Vugrek O, Wei KJ, Hasenbein NG, Sugimoto K, Rashbrooke MC, Wasteneys GO. 2001. MOR1 is essential for organizing cortical microtubules in plants. Nature **411,** 610–613.

Williams BP, Johnston IG, Covshoff S, Hibberd JM. 2013. Phenotypic landscape inference reveals multiple evolutionary paths to C_4 photosynthesis. eLife **2,** e00961.

Xie Y, Wu G, Tang J, *et al*. 2014. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics **30,** 1660–1666.

Yoshiyama K, Conklin PA, Huefner ND, Britt AB. 2009. Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage. Proceedings of the National Academy of Sciences, USA **106,** 12843–8.

Zghidi W, Merendino L, Cottet A, Mache R, Lerbs-Mache S. 2007. Nucleus-encoded plastid sigma factor SIG3 transcribes specifically the psbN gene in plastids. Nucleic Acids Research **35,** 455–464.