The Plant Cell, Vol. 23: 2087–2105, June 2011, www.plantcell.org

RESEARCH ARTICLES

Evolution of C_4 Photosynthesis in the Genus *Flaveria*: How Many and Which Genes Does It Take to Make C_4? [W]

Udo Gowik,[a,1] Andrea Bräutigam,[b] Katrin L. Weber,[b] Andreas P.M. Weber,[b] and Peter Westhoff[a]

[a] Institute of Plant Molecular and Developmental Biology, Heinrich-Heine-University, 40225 Duesseldorf, Germany
[b] Institute of Plant Biochemistry, Heinrich-Heine-University, 40225 Duesseldorf, Germany

Selective pressure exerted by a massive decline in atmospheric CO_2 levels 55 to 40 million years ago promoted the evolution of a novel, highly efficient mode of photosynthetic carbon assimilation known as C_4 photosynthesis. C_4 species have concurrently evolved multiple times in a broad range of plant families, and this multiple and parallel evolution of the complex C_4 trait indicates a common underlying evolutionary mechanism that might be elucidated by comparative analyses of related C_3 and C_4 species. Here, we use mRNA-Seq analysis of five species within the genus *Flaveria*, ranging from C_3 to C_3-C_4 intermediate to C_4 species, to quantify the differences in the transcriptomes of closely related plant species with varying degrees of C_4-associated characteristics. Single gene analysis defines the C_4 cycle enzymes and transporters more precisely and provides new candidates for yet unknown functions as well as identifies C_4 associated pathways. Molecular evidence for a photorespiratory CO_2 pump prior to the establishment of the C_4 cycle-based CO_2 pump is provided. Cluster analysis defines the upper limit of C_4-related gene expression changes in mature leaves of *Flaveria* as 3582 alterations.

INTRODUCTION

C_4 plants are characterized by high rates of photosynthesis and efficient use of water and nitrogen resources. High photosynthetic rates are achieved by addition of a new metabolic pathway, the C_4 cycle, in which the initial product of CO_2 fixation is a four-carbon (C_4) organic acid rather than a three-carbon (C_3) organic acid. In most C_4 species, C_4 photosynthesis involves two different cell types, mesophyll and bundle sheath cells. Only few species have been described that carry out a C_4 cycle within a single cell (Edwards et al., 2004). As shown in Figure 1A, in an NADP-dependent malic enzyme type C_4 plant, CO_2 is initially fixed in the mesophyll cells by phosphoenolpyruvate carboxylase (PEPC), which converts three-carbon phosphoenolpyruvate (PEP) into four-carbon oxaloacetate (OAA). OAA is converted into a transport form (malate or aspartate) by malate dehydrogenase (MDH) or aspartate aminotransferase (Asp-AT), respectively, and is then transported to the bundle sheath. Following decarboxylation of malate by NADP-malic enzyme (NADP-ME), the CO_2 is refixed by ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco), producing 3-phosphoglycerate that is further converted to triose phosphate. The pyruvate produced from malate (or its aminated form, Ala) is transferred back to the mesophyll where PEP is regenerated by pyruvate orthophosphate dikinase (PPDK) (Hatch, 1987).

C_4 plants show drastically reduced rates of photorespiration because CO_2 is concentrated at the site of Rubisco and is able to outcompete molecular oxygen, which, when used by Rubisco, results in photorespiration. Close contact between mesophyll and bundle sheath cells is vital for C_4 photosynthesis, and the leaf structure of C_4 plants is altered compared with most C_3 plants. The bundle sheath cells are enlarged, the interveinal distance is reduced, and the leaf thickness is limited to maximize the contact of mesophyll and bundle sheath cells (Dengler and Nelson, 1999). This pattern is called Kranz anatomy (Haberlandt, 1904). To guarantee the high flux of metabolites between the two cell types, they are connected via numerous plasmodesmata (Botha, 1992). The CO_2 pump ensures high rates of photosynthesis even when CO_2 concentrations are low in the intercellular air spaces of the leaf. Therefore, C_4 plants are able to limit the opening of their stomata and minimize water loss due to transpiration. As the CO_2 pump delivers saturating concentrations of CO_2 to the site of Rubisco, high photosynthetic rates are maintained with less Rubisco than is required in C_3 species. This is reflected in higher nitrogen use efficiency (Long, 1999).

While the basic biochemistry of the C_4 cycle is well understood, our knowledge about other genes and proteins needed for efficient C_4 photosynthesis is limited. For example, we have not identified yet all the transporters that ensure the increased inter- and intracellular metabolic fluxes nor the genes that regulate and maintain the alterations in cell and overall leaf morphology.

C_4 photosynthesis evolved several times independently during the evolution of higher plants. It originated at least 32 times in

[1] Address correspondence to gowik@uni-duesseldorf.de.
The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) are: Udo Gowik (gowik@uni-duesseldorf.de) and Peter Westhoff (west@uni-duesseldorf.de).
[W] Online version contains Web-only data.
www.plantcell.org/cgi/doi/10.1105/tpc.111.086264

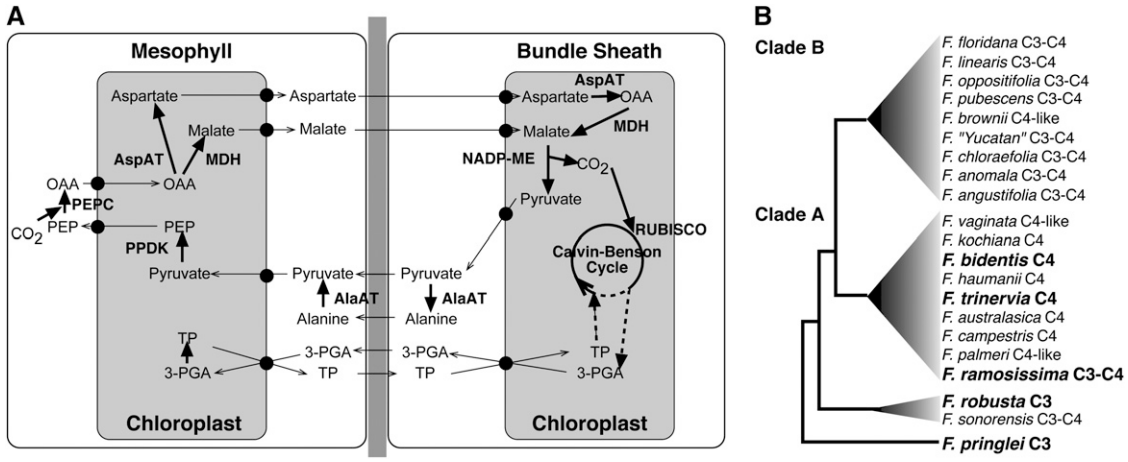

Figure 1. The Genus *Flaveria* as a Model Organism to Study C_4 Evolution.

(A) Schematic view of the NADP-ME type C_4 pathway as it can be found in C_4 *Flaveria* species modified from Gowik and Westhoff (2011). See the text for abbreviations and a detailed description of the pathway.
(B) Phylogeny of the genus *Flaveria* according to McKown et al. (2005).

eudicots and 16 times in monocots (Sage, 2004; Muhaidat et al., 2007). These multiple independent origins of C_4 photosynthesis suggest that the evolution of a C_3 into a C_4 species must have been relatively easy in genetic terms (Westhoff and Gowik, 2010). Recently, the C_4 syndrome has been investigated at the systems level by comparing the transcriptome of a C_4 to a closely related C_3 species (Bräutigam et al., 2011). Approximately 600 transcripts were differentially expressed at a significant level. While many of the transcriptional changes could be placed into a C_4 context, the question of which and how many of the changes are related to the C_4 syndrome rather than to the evolutionary distance of the two species remained open.

To get an insight how many and which genes were altered during C_4 evolution, we performed a comparative transcriptome analysis of leaves of closely related C_3, C_4, and C_3-C_4 intermediate species of the genus *Flaveria*. This genus is very valuable for investigating the evolution of the C_4 pathway because, in addition to having closely related C_3 and C_4 species of the NADP-ME type, it also contains a large range of C_3-C_4-intermediate species differing in the degree of "C_4ness" (Figure 1B).

Since no species of the genus *Flaveria* is a model organism with a known genome sequence and consequently no microarrays are available, we used massively parallel pyrosequencing of mRNAs (RNA-Seq) to analyze the leaf transcriptomes of C_3, C_4, and C_3-C_4 intermediate *Flaveria* species. This digital gene expression analysis (DGE) was based on generating random sequence tags that were proportional to the abundance of the corresponding transcripts in a particular sample and was shown to be useful for comparing steady state transcript levels in related nonmodel species (Bräutigam and Gowik, 2010; Bräutigam et al., 2011). The leaf transcriptomes of *Flaveria bidentis* (C_4) and *Flaveria pringlei* (C_3) were analyzed by the 454-FLX technology, and the newer 454-TITANIUM technology was used to sequence the leaf transcriptomes of *Flaveria trinervia* (C_4), *Flaveria robusta* (C_3), and *Flaveria ramosissima* (C_3-C_4).

RESULTS

Carbon Isotope Discrimination in the Different *Flaveria* Species

Plants discriminate against $^{13}CO_2$ during CO_2 uptake because of the different diffusivity of $^{13}CO_2$ and $^{12}CO_2$ and the preference of Rubisco for $^{12}CO_2$. In C_4 plants, this effect is less pronounced due to the CO_2 concentration mechanism. Thus, C_3 and C_4 plants can be distinguished by the carbon isotope composition of their dry matter (O'Leary, 1981). To confirm the photosynthetic types under greenhouse conditions, the carbon isotope ratios of the five *Flaveria* species investigated in this study were analyzed by determining the $\delta^{13}C$ values of dried leaf material (see Supplemental Table 1 online). The $\delta^{13}C$ values of the C_4 species are 12 to 15‰ higher (less negative) than the $\delta^{13}C$ values of the C_3 species. The $\delta^{13}C$ value of the C_3-C_4 intermediate species *F. ramosissima* is C_3 like (see Supplemental Table 1 online). These results echo earlier results (Monson et al., 1986; Edwards and Ku, 1987; Monson et al., 1988) showing that most C_3-C_4 intermediate *Flaverias*, including *F. ramosissima*, exhibit C_3 like carbon isotope ratios, although *F. ramosissima* fixes almost 50% of CO_2 via the C_4 pathway. Hence, the *Flaverias* under investigation behave as expected under our conditions.

454 Sequencing of *Flaveria* Leaf Transcriptomes

To identify differences in transcript abundance related to aspects of the C_4 syndrome, the leaf transcriptomes of *F. bidentis* (C_4) and *F. pringlei* (C_3) were compared. Also, in a second experiment, the leaf transcriptomes of *F. trinervia* (C_4), *F. robusta* (C_3), and *F. ramosissima* (C_3-C_4) were similarly compared. The analysis of gene expression in five species rather than a species pair reduced the probability of detecting species specific rather than C_4-specific differences. One sequencing run on a GS FLX system was conducted on the cDNA libraries from both *F. bidentis* and

F. pringlei, leading to 135,855,412 and 114,292,070 nucleotides of raw sequences, respectively. For *F. trinervia*, *F. robusta*, and *F. ramosissima*, the more advanced 454 TITANIUM technology was used, leading to 285,219,596, 308,800,825 and 333,275,756 nucleotides of raw data. After quality control and processing, this resulted in 527,596 clean reads from *F. bidentis* and 448,627 clean reads from *F. pringlei* with a mean read length of 229 nucleotides for both species, 974,217 clean reads from *F. trinervia*, 871,850 clean reads from *F. robusta*, and 1,096,348 clean reads from *F. ramosissima* with mean read lengths of 286, 349, and 297 bp, respectively (Table 1).

Clean reads were aligned to a minimal set of coding sequences of the *Arabidopsis thaliana* transcriptome (http://www.Arabidopsis.org/), as described by Bräutigam et al. (2011) to minimize erroneous read mapping to genes that have arisen from segmental or tandem gene duplications in the *Brassicacean* lineage (Bräutigam and Gowik, 2010). The alignment was performed in protein space using the BLAST-like alignment tool BLAT (Kent, 2002), and the best hit for each 454 read was retained. Between 66.6 and 72.7% of the reads from each *Flaveria* species could be mapped onto the *Arabidopsis* transcriptome (Table 2). The quantitative data for all genes detected can be found in Supplemental Data Set 1 online.

ESTs corresponding to 55, 55, 58, 60, and 61% of the *Arabidopsis* transcripts included in the minimal coding sequences set were identified in the individual leaf cDNA libraries of *F. bidentis*, *F. pringlei*, *F. trinervia*, *F. robusta*, and *F. ramosissima*, respectively. This indicated that the leaf transcriptomes of the *Flaveria* species were sampled to a comparable extent. We examined the coverage of different functional gene classes to test whether the data sets and the mappings for the different species within each experiment were comparable. For most functional classes transcripts representing more than 50% of the genes were detected (Table 3) in each of the five species. The classes putative lipid transfer protein, defense, and function unknown were the only ones underrepresented in the *F. robusta*/*F. ramosissima*/*F. trinervia* experiment as well as in the *F. pringlei*/*F. bidentis* experiment. The coverage of the individual functional classes was comparable for all species (Table 3).

Differential gene expression within each experiment (*F. bidentis* versus *F. pringlei* and *F. trinervia* versus *F. robusta*) was determined using Poisson statistics (Audic and Claverie, 1997) followed by a Bonferroni correction to account for multiple parallel testing. Among the 13,574 transcripts captured in the *F. bidentis*/*F. pringlei* experiment, the abundance of 463 transcripts differed significantly ($P < 0.01$) between the C_3 and the C_4 plant. Two hundred transcripts were more abundant and 263 transcripts less abundant in the C_4 plant *F. bidentis* compared with the C_3 plant *F. pringlei* (Table 2; see Supplemental Data Set 1 online). The combined ESTs of *F. trinervia* (C_4) and *F. robusta* (C_3) correspond to 14,304 transcripts. A total of 410 transcripts were significantly ($P < 0.01$) more and 585 transcripts less abundant in the C_4 plant.

To independently confirm the DGE results, quantitative RT-PCR experiments were performed on three leaf RNA isolates of *F. bidentis* and *F. pringlei*, the RNA used for 454 sequencing and two independent isolates. The results obtained with cDNA used for RNA-Seq as well as the mean values from three experiments strongly correlated with the results from DGE ($R^2 = 0.95$ and 0.86, respectively; see Supplemental Figure 1 online), indicating the reliability of the expression ratios estimated by RNA-Seq.

A Number of Functional Classes Differ between C_4 and C_3 Species

Transcripts of genes known to be involved in the C_4 cycle, the photosynthetic electron transport and CO_2 fixation, and photorespiration showed pronounced differences between C_3 and C_4 (Figure 2). A high percentage of the genes contained in the functional class of potential C_4 cycle genes showed strong and significant upregulation in the C_4 plants in both the *F. bidentis*/*F. pringlei* and the *F. trinervia*/*F. robusta* comparison. Other classes with high percentages of significantly more highly expressed genes in both experiments are glycolysis and the oxidative pentose phosphate pathway, whereas genes related to nitrogen metabolism and the shikimate pathway were significantly downregulated in the C_4 species.

To complement this analysis, we searched for functional classes showing significant differential expression between the C_3 and C_4 species using overrepresentation analysis by the PageMan Software (Usadel et al., 2006). This software considers the changes of all genes within a functional class. These are compared with the changes of all genes observed within the whole experiment to predict functional classes that exhibit

Table 1. Results of the 454 Sequencing Runs

Data	*F. trinervia* (454 T)	*F. robusta* (454 T)	*F. ramosissima* (454 T)	*F. bidentis* (454 F)	*F. pringlei* (454 F)
Raw reads	966,609	860,886	1,084,773	515,931	444,438
Raw nucleotides	285,219,596	308,800,825	333,275,756	135,855,412	114,292,070
Raw mean length	295	358	307	263	257
Clean and processed reads	974,217	871,850	1,096,348	527,596	448,627
Clean nucleotides	278,160,925	299,789,666	326,109,740	120,768,247	102,606,567
Clean and processed mean length	286	349	297	229	229
Reads mapped on TAIR9 coding sequences	648,969	634,109	759,862	368,342	326,108
Reads mapped (%)	66.6	72.7	69.3	69.8	72.7

F. trinervia, *F. ramosissima*, and *F. robusta* cDNA libraries were sequenced using 454 TITANIUM (454 T), whereas *F. bidentis* and *F. pringlei* cDNA libraries were sequenced using 454 FLX (454 F) chemistry. The raw reads from 454 sequencing were processed (exclusion of low-quality reads, elimination of adaptor sequences, and separation of sequence reads joined via concatemerization) to obtain clean reads.

Table 2. Mapping Results for the 454 Reads

Species Comparison	Transcriptomes	No. of Loci	Percentage of Total Loci in TAIR9
	TAIR 9 (minimalized transcriptome)	21,972	
	F. trinervia	12,817	58.3
	F. robusta	13,264	60.4
	F. ramosissima	13,534	61.6
Ft/Fro comparison more abundant ($P < 0.01$)	*F. trinervia* + *F.robusta*	14,304	65.1
	F. trinervia	410	1.9
	F. robusta	585	2.7
Ft/Fra comparison more abundant ($P < 0.01$)	*F. trinervia* + *F. ramosissima*	14,371	65.4
	F. trinervia	344	1.6
	F. ramosissima	503	2.3
Fra/Fro comparison more abundant ($P < 0.01$)	*F. ramosissima* + *F. robusta*	14,547	66.2
	F. ramosissima	385	1.8
	F. robusta	369	1.7
	F. bidentis	12,164	55.4
	F. pringlei	12,254	55.8
	F. bidentis + *F. pringlei*	13,574	61.8
Fb/Fp comparison more abundant ($P < 0.01$)	*F. bidentis*	200	0.9
	F. pringlei	263	1.2

Reads were mapped to a minimal version of the *Arabidopsis* transcriptome in the protein space using BLAT. Total numbers of transcripts detected by at least one read and the numbers of significantly differentially abundant transcripts ($P < 0.01$) in all possible species-by-species comparisons are given along with the corresponding percentage of the total loci. Ft, *F. trinervia*; Fra, *F. ramosissima*; Fro, *F. robusta*; Fb, *F. bidentis*; Fp, *F. pringlei*.

differential expression compared with all the other remaining functional classes. In both experiments, next to the not assigned class and classes related to photosynthesis, several classes associated with protein metabolism, especially the ribosomal proteins, show differential expression profiles compared with all other functional classes (Figure 3). Since Pageman and Mapman are designed for *Arabidopsis*, a C_3 plant, no functional C_4 class is annotated in these tools.

C_4 Cycle Genes Are Strongly Upregulated in the C_4 *Flaverias*

Transcripts encoding the proteins necessary for the NADP-ME type of C_4 photosynthesis were significantly upregulated in the C_4 plants *F. trinervia* and *F. bidentis* compared with the C_3 plants *F. robusta* and *F. pringlei.* The biggest difference, with a 180/125-fold higher transcript abundance, was PPDK followed by the PEPC with a 134/47-fold upregulation (Figure 4A; see Supplemental Table 2 online). Also, the abundance of transcripts for NADP-ME and the chloroplastidic MDH was 6/7 and 14/23 times higher in the C_4 plants. With an absolute abundance ranging from 3515/3,678 (MDH) to 34,365/15,887 (PEPC) reads per million (rpm), all these transcripts belong to the most abundant transcripts in *F. trinervia* and *F. bidentis*. Also, Ala aminotransferases and one transcript encoding an ASP-AT were upregulated in the C_4 plants. This confirms that amino acids, in addition to malate, are also used as transport metabolites in the C_4 *Flaverias*.

The adenosine monophosphate kinase gene was found to be strongly upregulated in both C_4 species. In the *F. bidentis*/*F. pringlei* experiment, we also identified two significantly upregulated inorganic pyrophosphatases (see Supplemental Table 2 online).

Transcripts encoding regulatory factors for C_4 cycle proteins, the PEPC kinase (PEPC-K) and the PPDK regulatory protein (PPDK-RP), were upregulated as well (4- to 46-fold for PEPC-K and 2- to 13-fold for PPDK-RP), although their absolute abundance is clearly lower than that of the C_4 enzymes (73 to 176 rpm for PEPC-K and 120 to 274 rpm for PPDK-RP).

The genes encoding the enzymes necessary for the NAD-ME or phosphoenolpyruvate carboxykinase (PEP-CK) C_4 subtype, such as the mitochondrial NAD-dependent malate dehydrogenase (mNAD-MDH), the mitochondrial NAD-dependent malic enzyme (mNAD-ME), cytoplasmatic or mitochondrial ASP-ATs, or the cytoplasmic PEP-CK, show only low or moderate expression and the C_4- to C_3-associated differences were small and not significant (Figure 4A; see Supplemental Table 2 online), indicating that the true C_4 *Flaveria* species exclusively use the NADP-ME pathway as reported earlier (Drincovich et al., 1998). This is supported by the extractable protein activities and steady state metabolite amounts (see Supplemental Table 3 and Supplemental Table 4 online). PEP-CK activity is increased in C_4 albeit much less compared with the major decarboxylation enzyme NADP-ME.

In *F. ramosissima* (C_3-C_4), Next to the NADP-ME Type C_4 Cycle, Typical NAD-ME Type C_4 Genes Also Are Upregulated

In the C_3-C_4 intermediate plant *F. ramosissima*, the transcripts of the genes related to the NADP-ME type C_4 photosynthesis showed intermediate amounts compared with the C_3 plant *F. robusta* and the C_4 plant *F. trinervia* (Figure 4A; see Supplemental Table 2 online). The amounts of all these transcripts were significantly higher than in the C_3 plant *F. robusta*, implying that in *F. ramosissima*, the C_4 cycle is working to a certain extent and that *F. ramosissima* is a true intermediate based on its transcriptional profile. By contrast, the transcript abundance for the Ala aminotransferase gene in *F. ramosissima* was higher than in the C_4 species *F. trinervia.*

Table 3. Qualitative Patterns of Transcript Abundance in the Leaves of *F.trinervia*, *F. ramosissima*, *F. robusta*, *F. bidentis*, and *F. pringlei*

Functional Class	Genes Detected in:				
	F. trinervia	*F. robusta*	*F. ramosissima*	*F. bidentis*	*F. pringlei*
All	58.33%	60.37%	61.60%	55.36%	55.77%
C_4 candidate	91.30%	91.30%	91.30%	91.30%	91.30%
Photosynthesis, photosystem I	86.96%	95.65%	95.65%	91.30%	86.96%
Photosynthesis, photosystem II	87.76%	89.80%	93.88%	91.84%	89.80%
Photosynthesis, cytochrome b6/f	37.50%	62.50%	50.00%	75.00%	50.00%
Photosynthesis, cef	75.00%	85.71%	89.29%	82.14%	82.14%
Photosynthesis, ATPase	91.67%	91.67%	91.67%	91.67%	66.67%
Photosynthesis, other	90.00%	90.00%	90.00%	90.00%	90.00%
Calvin cycle	94.12%	100.00%	94.12%	94.12%	94.12%
Tricarboxylic acid cycle	92.00%	92.00%	92.00%	92.00%	92.00%
Mitochondrial electron transfer/ATPase	70.23%	72.52%	80.92%	70.99%	66.41%
OPPP	90.48%	95.24%	90.48%	76.19%	90.48%
Glycolysis	93.48%	91.30%	93.48%	93.48%	93.48%
Other central carbon metabolism	80.77%	84.62%	76.92%	65.38%	76.92%
Photorespiration	100.00%	100.00%	100.00%	100.00%	100.00%
Starch	93.02%	93.02%	93.02%	88.37%	95.35%
Suc	91.43%	97.14%	85.71%	82.86%	88.57%
Lipids	85.48%	86.77%	90.32%	82.90%	86.45%
Nitrogen metabolism	90.32%	90.32%	96.77%	90.32%	90.32%
Shikimate pathway	84.62%	92.31%	84.62%	92.31%	84.62%
One-carbon compound metabolism	100.00%	100.00%	100.00%	100.00%	100.00%
Amino acid metabolism	93.65%	93.65%	95.24%	90.48%	88.89%
Nucleotide metabolism	92.23%	96.12%	95.15%	94.17%	93.20%
Pigment synthesis	91.67%	93.75%	97.92%	91.67%	91.67%
Cofactor synthesis	89.02%	93.90%	93.90%	90.24%	90.24%
Secondary metabolism	68.83%	74.03%	76.19%	67.10%	71.00%
Enzyme, other	68.33%	73.75%	75.83%	61.67%	62.50%
Enzyme, putative	76.07%	80.48%	82.12%	72.54%	76.70%
Sulfur assimilation	100.00%	100.00%	100.00%	100.00%	100.00%
Cell wall synthesis	60.57%	63.09%	64.04%	53.31%	59.31%
Minor sugar metabolism	72.00%	73.60%	77.07%	70.40%	68.80%
β 1.3 Glucan metabolism	55.00%	60.00%	67.50%	55.00%	62.50%
Vesicle trafficking	82.29%	85.94%	86.46%	81.77%	82.81%
Cytoskeleton	76.72%	80.95%	82.01%	73.54%	69.31%
Other cellular processes	76.71%	80.76%	81.67%	73.91%	74.90%
Chloroplast process	88.97%	94.12%	94.12%	90.44%	90.44%
Mitochondrial process	73.44%	71.88%	70.31%	68.75%	62.50%
Peroxisomal process	93.75%	93.75%	93.75%	100.00%	93.75%
Protein synthesis	84.84%	87.43%	88.72%	85.58%	83.92%
Protein modification	83.33%	91.67%	100.00%	83.33%	100.00%
Protein degradation	60.93%	62.31%	63.70%	58.65%	59.64%
Heat shock/protein folding	82.11%	82.63%	84.74%	80.00%	80.53%
Transport	85.77%	86.39%	87.64%	81.02%	82.02%
Redox regulation	89.29%	89.88%	90.48%	85.71%	89.88%
Metal handling	60.00%	60.00%	66.67%	56.67%	56.67%
Storage protein	56.25%	62.50%	62.50%	56.25%	56.25%
Transcriptional regulation	58.75%	57.11%	60.79%	52.12%	53.20%
Epigenetic regulation	80.88%	80.88%	82.35%	70.59%	76.47%
Posttranslational modification (phosphate)	81.11%	82.12%	84.76%	76.45%	77.20%
Other regulatory processes	67.21%	68.70%	69.30%	64.98%	62.89%
Hormone metabolism	68.24%	68.24%	76.35%	62.16%	64.19%
Hormone signaling	80.77%	80.77%	84.62%	76.92%	84.62%
Hormone-responsive genes	50.38%	56.49%	56.49%	54.96%	48.85%
Defense	39.01%	45.05%	47.80%	35.71%	36.26%
Putative lipid transfer proteins	21.11%	24.44%	24.44%	22.22%	21.11%
Not classified	73.73%	76.27%	77.12%	68.64%	74.58%
Function unknown	42.06%	44.22%	44.67%	39.44%	39.29%

Different functional classes are represented by comparable fractions of genes detected in the leaf cDNA libraries from the different species.

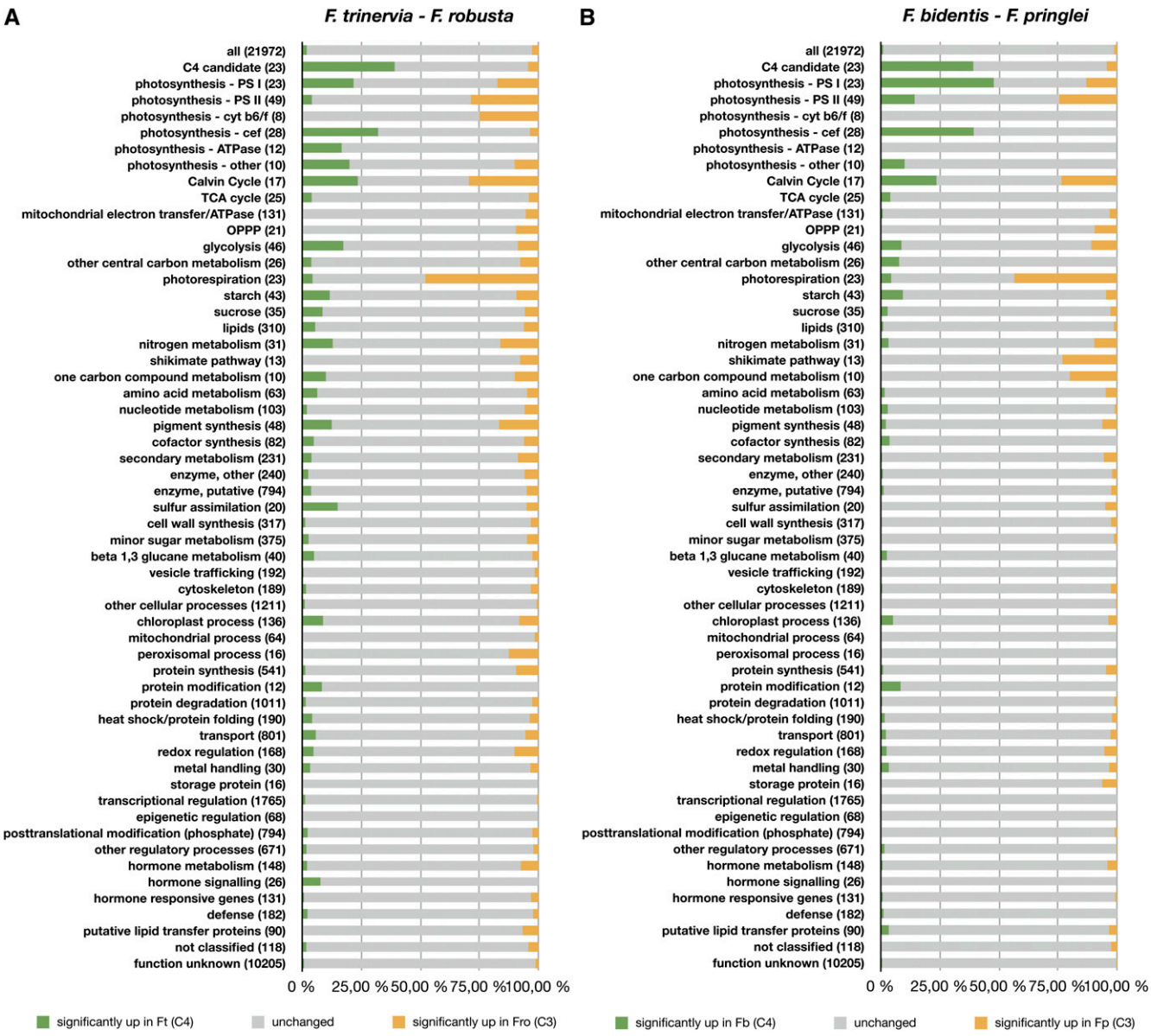

Figure 2. The Quantitative Patterns of Transcript Accumulation in C_3 and C_4 *Flaverias* Are Distinct.

(A) Comparison of *F. trinervia* (Ft, C_4) and *F. robusta* (Fro, C_3).
(B) Comparison of *F. bidentis* (Fb, C_4) and *F. pringlei* (Fp, C_3). Shown are the percentages of genes with significantly higher abundance of transcripts in the C_4 species (green bars), percentages of genes unchanged (gray bars, including genes not detected), and percentages of genes with significantly lower abundance of transcripts in C_4 species (yellow bars). Percentages are based on the total number of genes in each annotation class (values in parentheses on the *y* axis). TCA, tricarboxylic acid.

Additionally, cytoplasmic and mitochondrial ASP-AT genes and an mNAD-MDH were upregulated significantly in *F. ramosissima* compared with the C_3 and the C_4 species. Two mNAD-MEs were upregulated in *F. ramosissima*, whereas the differences were significant only for one gene in comparison with the C_4 plant *F. trinervia* (Figure 4A; see Supplemental Table 2 online). Accordingly, the extractable NAD-ME activity in the leaves of *F. ramosissima* was significantly higher than in the other four *Flaveria* species (Figure 4B; see Supplemental Table 3 online). We further analyzed the steady state amounts of metabolites, including those associated with the C_4 pathway. In *F. ramosissima*, the Ala level was comparable to those found in the two C_4 species; however, the Asp level exceeded those of all other *Flaverias* (Figure 4C; see Supplemental Table 4 online).

C_4-Related Transport

C_4 photosynthesis requires the transport of large amounts of metabolites across the chloroplast envelope, and this transport

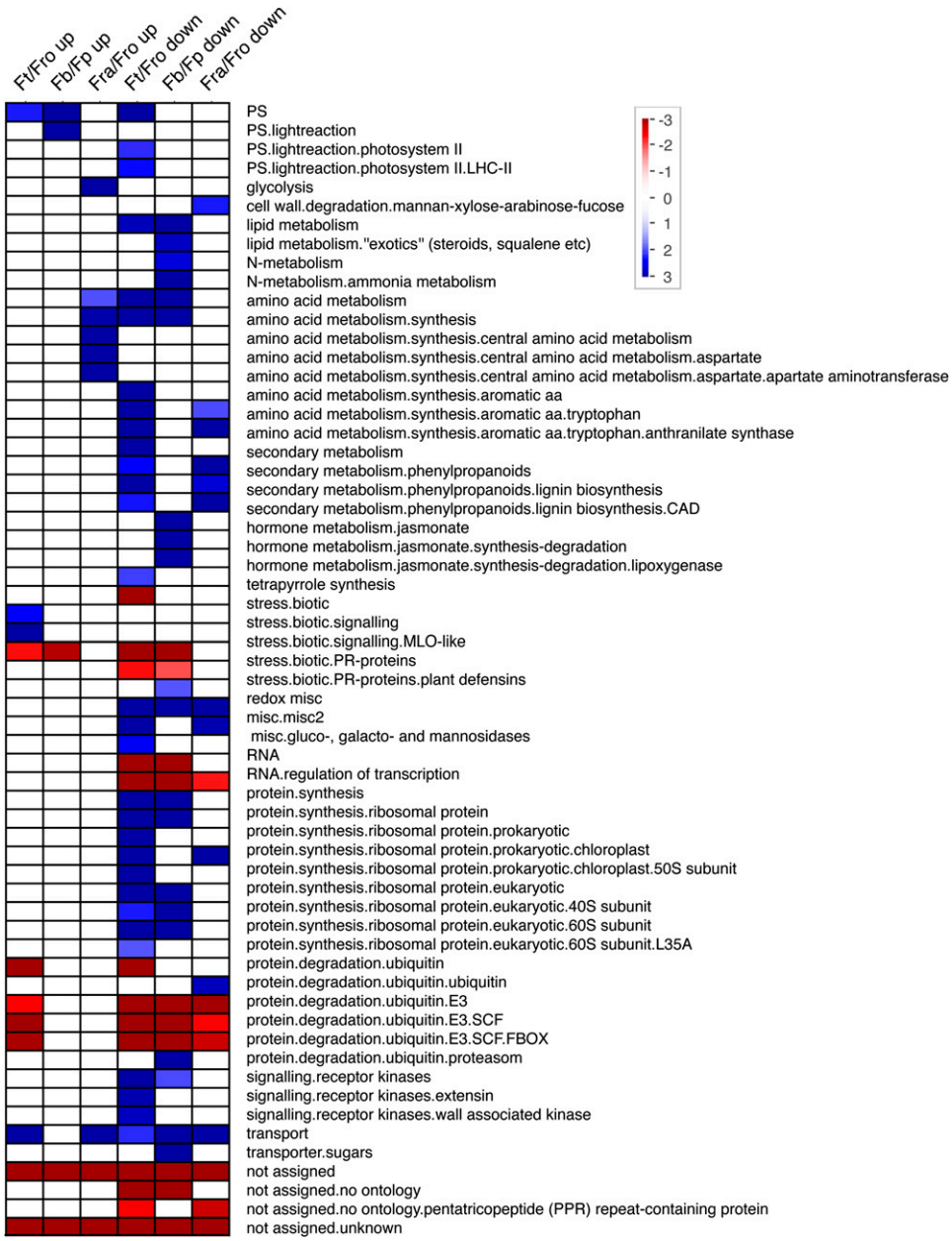

Figure 3. Overrepresentation Analyses of Up- and Downregulated Genes within Functional Gene Classes Defined by MapMan Bins.

Fisher's exact test followed by the Bonferroni correction was used to identify functional categories enriched in up- or downregulated genes when transcript abundances in *F. trinervia* (Ft, C_4) and *F. robusta* (Fro, C_3), *F. bidentis* (Fb, C_4) and *F. pringlei* (Fp, C_3), or *F. ramosissima* (Fra, C_3-C_4) and *F. robusta* (Fro, C_3) were compared. Blue, up- or downregulated genes are significantly overrepresented; red, up- or downregulated genes are significantly underrepresented. aa, amino acid; LHC, light-harvesting complex; PS, photosynthesis.

is not necessary in C_3 plants (Bräutigam et al., 2008; Weber and von Caemmerer, 2010; Bräutigam and Weber, 2011). Our experiments confirmed the importance of the plastidic phosphoenolpyruvate phosphate translocator and the triosephosphate phosphate translocator for the C_4 pathway since they were upregulated in the C_4 *Flaverias* (Figure 4A; see Supplemental Table 2 online), confirming earlier results from other C_4 species (Weber and von Caemmerer, 2010; Bräutigam and Weber, 2011).

The *Flaveria* species belong to the group of pyruvate sodium symporter C_4 plants (Aoki et al., 1992). A gene annotated as bile acid sodium symporter was dramatically upregulated in the C_4 compared with the C_3 *Flaveria* species. BASS 2 protein is a pyruvate sodium symporter (T. Furumoto, T. Yamaguchi, Y. Ohshima-Ichie, M. Nakamura, Y. Tsuchida-Iwata, M. Shimamura, J. Ohnishi, S. Hata, U. Gowik, P. Westhoff, A. Bräutigam, A. Weber, and K. Izui, unpublished data). To avoid massive sodium

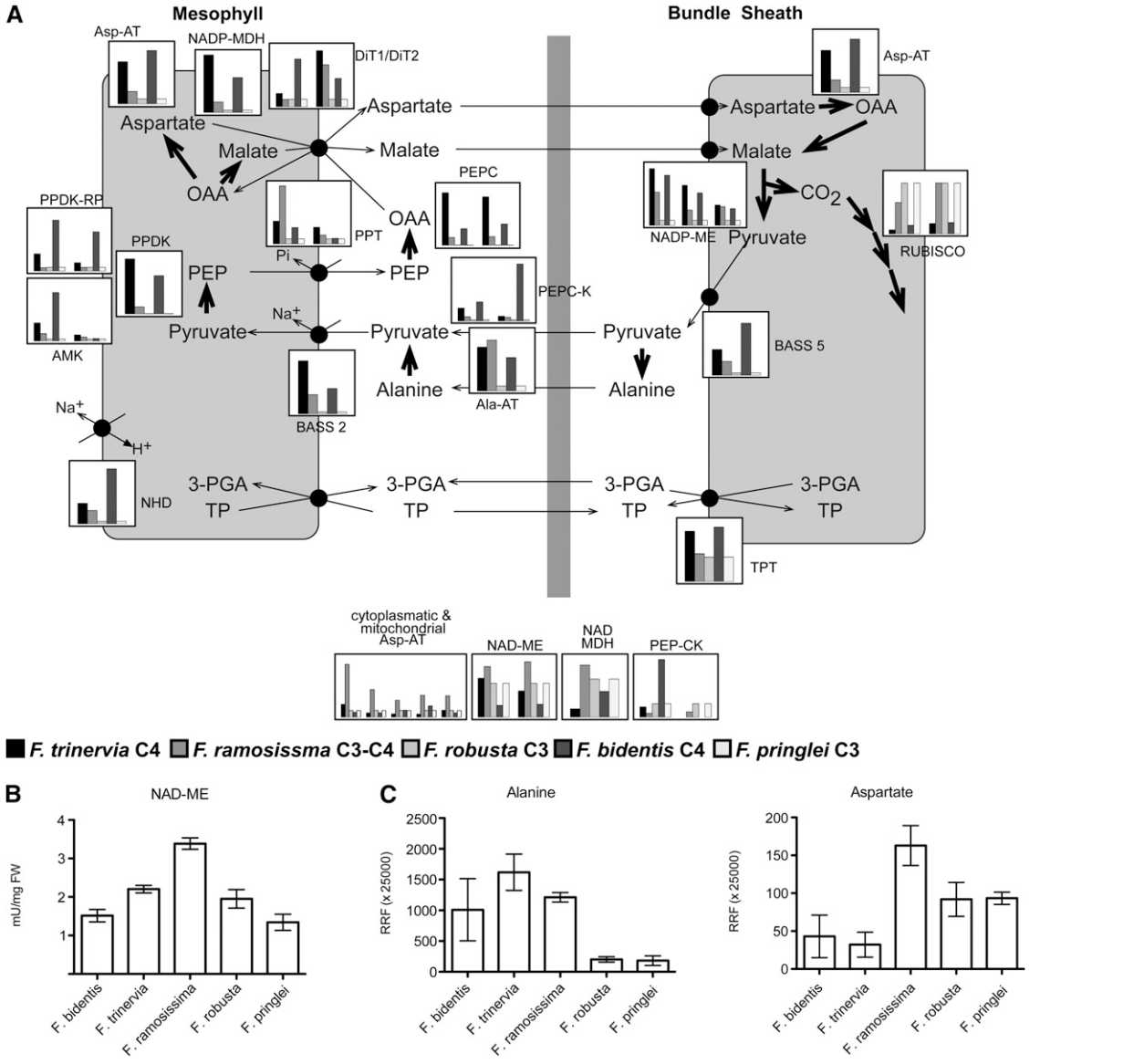

Figure 4. Differences in C_4 Pathway Gene Expression for *F. trinervia* (C_4), *F. ramosissima* (C_3-C_4), *F. robusta* (C_3), *F. bidentis* (C_4), and *F. pringlei* (C_3).

(A) Schematic view of the NADP-ME type C_4 pathway. Relative transcript abundances are given in small inset boxes. The transcript levels for *F. trinervia*, *F. ramosissima*, and *F. robusta* were normalized by setting the *F. robusta* transcript level to one, and the *F. bidentis* and *F. pringlei* transcript levels were normalized by setting the *F. pringlei* transcript level to one for each gene.

(B) Activity of NAD-ME in the extractable enzyme fractions of leaves from all five species (+ SE; $n = 3$). FW, fresh weight.

(C) Ala and Asp amounts in the leaves of all five species (+ SE; $n = 3$).

imbalance across the chloroplast envelope, BASS 2 acts in concert with a sodium proton antiporter (NHD), tying pyruvate import to the proton gradient (T. Furumoto, T. Yamaguchi, Y. Ohshima-Ichie, M. Nakamura, Y. Tsuchida-Iwata, M. Shimamura, J. Ohnishi, S. Hata, U. Gowik, P. Westhoff, A. Bräutigam, A. Weber, and K. Izui, unpublished data). In addition to BASS 2, a NHD was highly upregulated in both C_4 species compared with the C_3 *Flaverias*. The chloroplast dicarboxylate transporter 1 (DiT1) catalyzes the exchange of malate and OAA in addition to malate and 2-oxoglutarate and is expressed in the mesophyll of the NADP-ME grasses maize (*Zea mays*) and sorghum (*Sorghum bicolor*; Kinoshita et al., 2011). DiT1 as well as the chloroplast DiT2 were significantly upregulated in the C_4 plants *F. bidentis* and *F. trinervia* compared with the C_3 species (Figure 4A; see Supplemental Table 2 online). An additional gene belonging to the bile acid sodium symporter family, BASS 4 was upregulated to a comparable extent in the C_4 species (see Supplemental Table 2 online).

Several other transport proteins of unknown function displayed a C_4 accumulation pattern. These candidate C_4 transporters included a magnesium/proton exchanger, a high affinity potassium transporter, and the three chloroplastic cation exchangers CAX1, CAX3, and CAX4, which were all more highly expressed in the C_4 plants. Some transporter protein genes were significantly downregulated in the C_4 species: two for sugar transporters and the two for the water channel proteins TIP2;2 and PIP2B, respectively. A transcript encoding a putative voltage-dependent anion channel 1 (VDAC1) was also less abundant in the C_4 plants.

Photorespiration Is Downregulated in C_4 but Upregulated in the C_3-C_4 Intermediate Species

The highest percentage of genes downregulated in the C_4 species in both experiments belonged to the photorespiration class (Figure 2). Nearly all of the genes within this class were downregulated in the C_4 species, and for nearly 50% of them, the differences are statistically significant (Figure 5A). This was also true for the genes related to the reassimilation of photorespiratory ammonium by the plastidic Gln synthase and the ferredoxin-dependent Glu synthase but not for the transporters DiT1 and DiT2, which catalyze the 2-oxoglutarate/Glu exchange across the plastid membrane (see above). Flux through the photorespiratory pathway is reduced in C_4 plants compared with C_3 plants (Leegood, 2002; Sage, 2004), and, at least for this pathway, transcript abundance mirrors flux (Bräutigam et al., 2011; this article).

Surprisingly, the C_3-C_4 intermediate *F. ramosissima* did not show intermediate characteristics. By contrast, transcript abundances for most genes related to photorespiration in the C_3-C_4 intermediate species *F. ramosissima* were higher than in the C_3 species *F. robusta*, and for more than one half of them, this difference was statistically significant. In addition to the transcript amounts, both the steady state amount of Gly as well as the steady state amount of Ser increased, while glycolate and glycerate amounts remained comparable to the C_3 and C_4 species (Figure 5B; see Supplemental Table 4 online).

Photosynthetic Electron Transport and Calvin-Benson Cycle Were Modified during C_4 Evolution

Within the Calvin-Benson cycle class, most genes showed lower transcript abundance in the C_4 than in the C_3 plants. The strongest differences were found for the genes encoding the small subunit of the Rubisco, which were downregulated 4.5- to 12.5-fold in the C_4 *Flaverias*. In the C_3-C_4 intermediate *F. ramosissima*, the transcript abundance of most Calvin-Benson cycle genes was C_3 like with the exception of the small subunit of Rubisco, which was significantly downregulated, mirroring earlier investigations on Rubisco protein amounts in C_3-C_4 intermediate *Flaveria* species (Wessinger et al., 1989).

The classes with genes involved in the photosynthetic electron transport showed heterogeneous characteristics (see Supplemental Figure 2 online). Photosystem I genes were upregulated to a higher percentage in the C_4 plants, whereas more of the photosystem II genes were downregulated in both C_4 species. The class of genes related to the cyclic electron transfer was one of the classes containing the highest fraction of significantly upregulated genes in *F. trinervia* compared with *F. robusta* as well as in *F. bidentis* compared with *F. pringlei*. Since most genes encoding the ATPase and the cytochrome b_6f complex are encoded on the chloroplast genome, they were not analyzed in the experiments. In *F. ramosissima*, several genes related to the cyclic electron transfer as well as of photosystem I showed intermediate abundance compared with the C_3 and C_4 species, and others were at the level found in the C_3 plant. Most transcripts related to photosystem II show intermediate characteristics as many of them are downregulated compared with the C_3 plant *F. robusta* but not as much as much as in the C_4 plant *F. trinervia.*

Chloroplast Biogenesis and Maintenance Is Altered in C_4 Species

Several genes involved in chloroplast biogenesis and maintenance were differentially expressed between the C_3 and C_4 *Flaverias* (see Supplemental Table 5 online). Among these are genes encoding proteins of so far unknown function, which are predicted as being localized in the plastids, making them candidates for further analysis.

HCF101 and HCF107 are involved in the biogenesis of photosystem I and photosystem II, respectively (Lezhneva et al., 2004; Sane et al., 2005), and were found to be upregulated in the C_4 species. Several DnaJ proteins with unknown function behaved similarly. Plastidic DnaJ proteins are involved in the stabilization of thylakoid membrane complexes like photosystem II (Chen et al., 2010). Several proteases belonging to Clp (ClpR1 and ClpP5) and FtsH (FtsH8, VAR1, and VAR2) complexes were upregulated, too. While the Clp complex is essential for chloroplast biogenesis (Kim et al., 2009), the FtsH complex is mainly involved in the maintenance of photosystem II function (Kato et al., 2009).

The two chloroplast RNA binding proteins CSP41A and CSP41B were downregulated in the C_4 species. These proteins play a role in the expression of plastid genes and may be involved in the biogenesis of plastidial ribosomes (Beligni and Mayfield, 2008; Bollenbach et al., 2009). Several proteins involved in chloroplast division, namely, FtsZ1, FtsZ2, Arc5, and Cpn60B (Gao et al., 2003), are downregulated in the C_4 species. Although chloroplast division is largely completed in mature leaves, protein turnover appeared upregulated in both C_4 species compared with the respective C_3 species.

The C_4 Syndrome Alters Nitrogen Metabolism, Amino Acid Metabolism, and Translation

C_4 plants need less Rubisco in their leaves than C_3 species to perform the same amount of CO_2 fixation leading to a better nitrogen use efficiency by C_4 compared with C_3 species (Black, 1973; Ku et al., 1979; Oaks, 1994; Brown, 1999; Osborne and Freckleton, 2009; Ghannoum et al., 2011). Protein synthesis was altered in the C_4 *Flaveria* species, since this MapMan bin and several of its sub-bins are enriched in downregulated transcripts compared with all other MapMan bins using the overrepresentation analysis of the PageMan

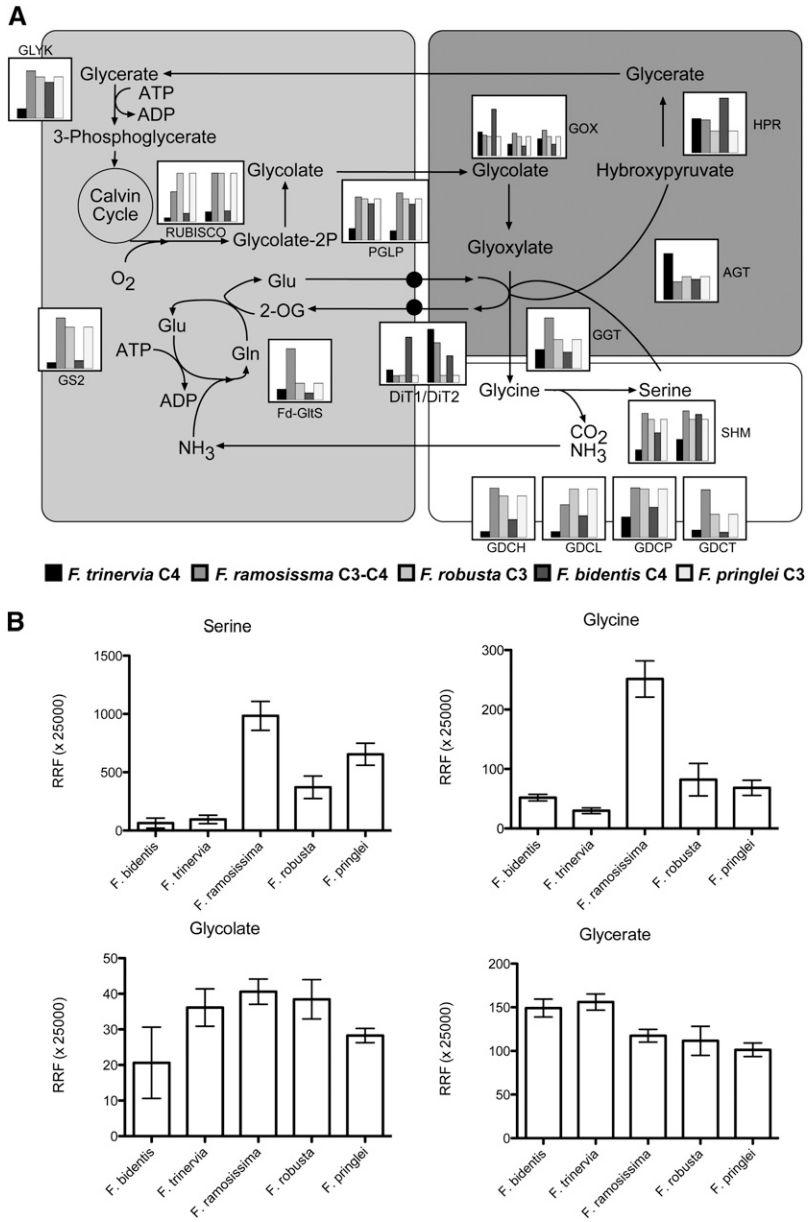

Figure 5. Photorespiration Is Altered between *F. trinervia* (C_4), *F. ramosissima* (C_3-C_4), *F. robusta* (C_3), *F. bidentis* (C_4), and *F. pringlei* (C_3).

(A) Schematic view of the photorespiratory pathway. Relative transcript abundances are given in small inset boxes. The transcript levels for *F. trinervia*, *F. ramosissima*, and *F. robusta* were normalized by setting the *F. robusta* transcript level to one, and the *F. bidentis* and *F. pringlei* transcript levels were normalized by setting the *F. pringlei* transcript level to one for each gene.

(B) Amounts of important photorespiratory metabolites in the leaves of all five species (± SE; $n = 3$).

software (Figure 3). Downregulated transcripts representing cytosolic ribosomes were enriched in all C_4 species, while transcripts associated with plastidic ribosomes were only overrepresented in *F. trinervia*. No enrichment was detected for downregulated components of mitochondrial ribosomes, indicating that there was no general effect on translation but specific for ribosomes translating photosynthetic and photorespiratory transcripts.

In *F. ramosissima*, the abundance of transcripts related to the eukaryotic ribosomal proteins was similar to C_3 levels, whereas the transcripts related to the plastidic ribosomal proteins showed amounts that are intermediate between the C_3 species *F. robusta* and the C_4 species *F. trinervia*.

In accordance with these findings, elemental analysis showed that the C_4 *Flaverias* exhibit higher carbon to nitrogen ratios (7.8 to 8.6) than the C_3 species (5.5 to 5.7) (see Supplemental Table 6

online). *F. ramosissima* had an intermediate carbon to nitrogen ratio (6.8 to 6.9).

Consequently, the genes involved in amino acid synthesis were downregulated in the C_4 *Flaveria* species, since downregulated transcripts were overrepresented within the bins "amino acid metabolism" and "amino acid metabolism synthesis." In the *F. bidentis*/*F. pringlei* experiment, the bins "nitrogen metabolism" and "ammonia metabolism" were enriched in downregulated genes (Figure 3).

Expression Changes Related to C_4

To discover additional genes that might be associated with the C_4 trait, all *Flaveria* transcriptome data were clustered and tested for C_4-related patterns. Hierarchical clustering showed that the two C_4 species are more similar to each other than to the other three analyzed *Flaveria* species with respect to their overall leaf transcript profile (Figure 6A). K-means clustering identified 20 clusters with species-related gene expression changes, which are unrelated to a C_4 pattern (see Supplemental Figure 3 online). Six clusters show patterns related to C_4 photosynthesis, either high in C_4 versus low in C_3 (three clusters) or high in C_3 versus low in C_4 (three clusters) (Figure 6B). The clusters vary in regard to *F. ramosissima* expression as exemplified for C_4 transcripts (intermediate) or photorespiratory transcripts (higher in *F. ramosissima*) above. Taken together, the C_4 clusters contained 3582 transcripts (Figure 6B; see Supplemental Data Set 2 online). A total of 1418 of these genes were in clusters with C_4 upregulated genes, whereas 2164 genes were downregulated during C_4 evolution.

Early Evolutionary Changes

Clusters one and two contained 1213 genes, which were upregulated in the two C_4 species and C_4-like or intermediate in the C_3-C_4 intermediate species *F. ramosissima*. The genes encoding the core C_4 enzymes and known or putative C_4 transporters were all part of cluster one (see Supplemental Data Set 2 online). Additional functional classes that were enriched within cluster one and two were minor carbohydrate metabolism, glycolysis, the tricarboxylic acid cycle, abscisic acid metabolism, posttranslational modification of proteins, and phosphoinositol and light signaling (Figure 6c; see Supplemental Figure 4 online). No cluster was formed that contains transcripts downregulated both in the C_4 species and in the intermediate. The number of these transcripts was thus small. We suggest that the changes in the C_4 species and the intermediate were C_4 changes in the narrow sense.

Late Evolutionary Changes

Cluster three contains transcripts that were more highly expressed in the C_4 species compared with the C_3 species but not in the C_3-C_4 intermediate species *F. ramosissima*. In this cluster, photosynthesis and light reaction transcripts as well as transcripts related to abscisic acid, auxin, and ethylene metabolism, several families of transcription factors and phosphorelay signaling were enriched. Clusters four, five, and six contained genes that were downregulated in the C_4 compared with the C_3 species. In the C_3-C_4 intermediate species *F. ramosissima*, genes from the three clusters were mainly expressed on C_3 level (Figure 6B). Within these clusters, genes related to major carbohydrate metabolism (including the Calvin Benson cycle and photorespiration) and minor carbohydrate metabolism, tricarboxylic acid cycle, C1 metabolism, and tetrapyrrole synthesis were enriched (Figure 6C). The cluster analysis confirmed the overrepresentation analysis based on the species by species comparisons with respect to the protein synthesis and nitrogen metabolism and indicated these changes are late changes. A total of 2369 changes were late and we suggested that these changes were C_4 changes in the wider sense.

Regulatory Genes

In clusters one and two, the C_4 clusters in the narrow sense, we found 151 genes encoding transcriptional regulators and 35 genes related to signaling pathways (see Supplemental Data Set 3 online). Among the transcriptional regulators, we identified two plastidal Sigma70-like factors, SIG1 and SIG5, which were furthermore significantly upregulated in the *F. trinervia*/*F. robusta* experiment. Plastidial sigma factors are encoded in the nuclear genome and control plastid gene expression by guiding RNA polymerase to the promoter (Lerbs-Mache, 2011). In *F. ramosissima*, SIG5 showed an abundance that was intermediate compared with the *F. robusta* and *F. trinervia*, whereas the transcript abundance of SIG1 was comparable to the C_4 plant *F. trinervia* in the C_3-C_4 intermediate. In the C_4 plant *Cleome gynandra*, a different Sigma70-like factor, SIG6, was upregulated significantly compared with the C_3 plant *Cleome spinosa* (Bräutigam et al., 2011). Thus, it might be possible that also the different abundance of plastidic sigma factors in C_3 and C_4 species differentially regulate chloroplast gene expression and thus alter the abundance of the complexes of the photosynthetic electron transfer chain observed in these species. Another transcription factor that was expressed significantly differential in the *F. trinervia* and *F. robusta* was the auxin response factor *ARF2*. With 185 to 516 rpm, the *ARF2* gene was highly expressed for a regulatory factor in the leaves of all five *Flaveria* species. In the C_3-C_4 intermediate *F. ramosissima*, the abundance of *ARF2* transcripts was intermediate compared with *F. trinervia* and *F. robusta*. Homozygous *Arabidopsis ARF2* mutants show a pleiotropic phenotype. Among others, the leaf size is enlarged caused by an increase of both cell division and cell expansion (Okushima et al., 2005; Gonzalez et al., 2010). Thus, one can assume that *ARF2* is involved in the establishment and maintenance of the typical C_4 leaf anatomy. *GOLDEN2 LIKE* (*GLK*) transcription factors are known to be involved in the chloroplast dimorphism in mesophyll and bundle sheath cells of maize (Waters et al., 2009). *GLK2* was a member of cluster two, indicating that changes to the GLK proteins played also an important role in the development of the C_4 pathway in *Flaveria*. Interestingly, in *Cleome*, the GLK2 counterpart GLK1 was upregulated in the C_4 species (Bräutigam et al., 2011).

A total of 183 transcription factors and 91 genes related to signaling were found in clusters three to six, the C_4 clusters in the wider sense (see Supplemental Data Set 3 online). Most strikingly, one can find 73 signaling receptor kinase genes, including

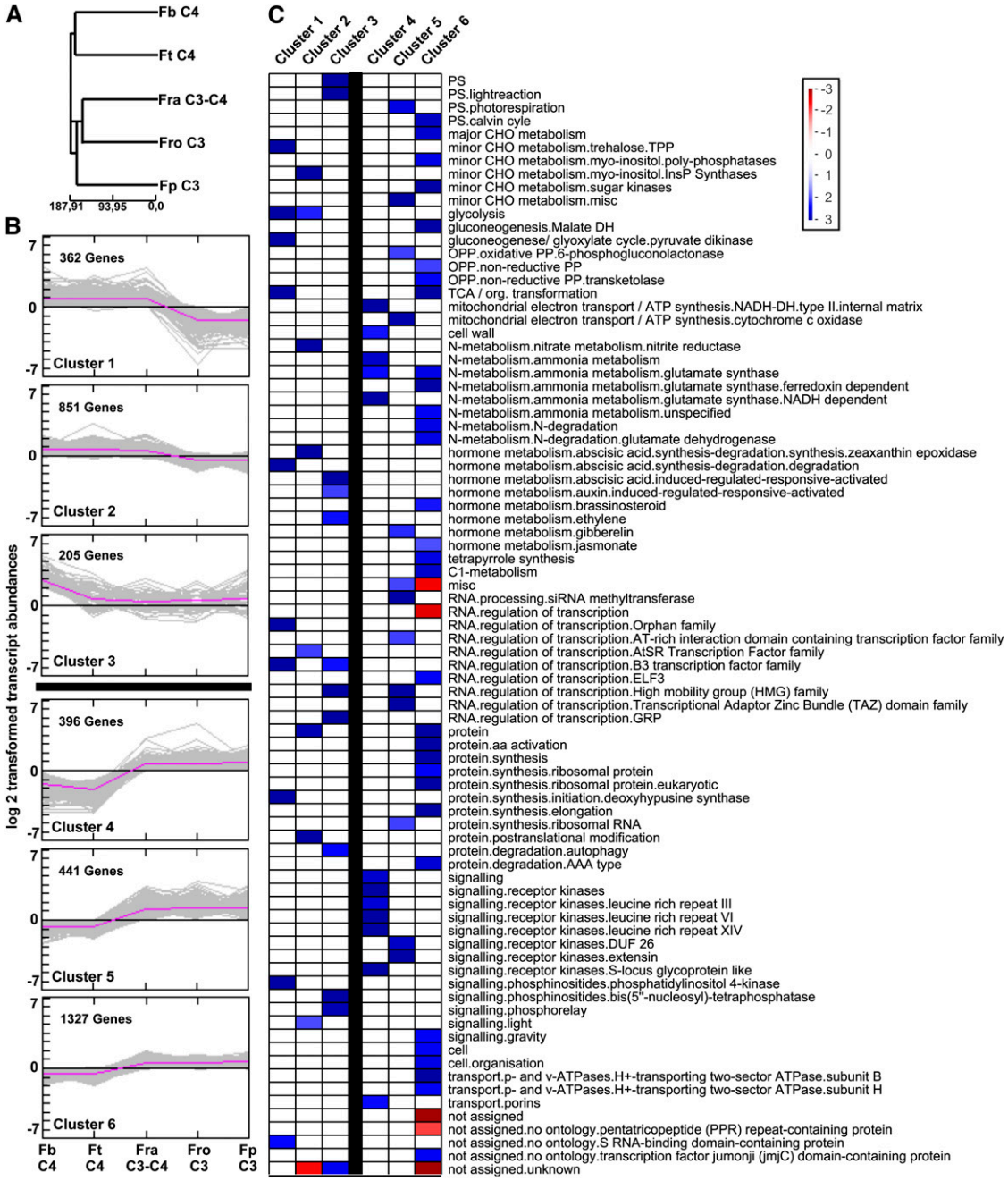

Figure 6. Cluster Analysis of Transcript Abundances in *F. bidentis* (C_4), *F. trinervia* (C_4), *F. ramosissima* (C_3-C_4), *F. robusta* (C_3), and *F. pringlei* (C_3).

(A) Hierarchical sample clustering of all expressed transcripts. The tree was calculated with the MEV program using the HCL module with the Euclidean distance criterion and the average linkage method. According to their transcript profiles, the two C_4 species are more closely related to each other than to the other three *Flaveria* species.

(B) C_4-related clusters. K-means analysis was used to define 26 clusters identifying different expression profiles. The six clusters with a C_4-related pattern are shown. All 26 clusters can be found in Supplemental Figure 2 online.

(C) Functional category (MapMan bins) enrichment among the six C_4-related clusters. Enrichment of genes belonging to distinct functional categories was analyzed with the Wilcoxon statistic followed by the Benjamini-Hochberg correction. Blue, significantly overrepresented; red, significantly underrepresented. The complete enrichment analysis for all 26 clusters is shown in Supplemental Figure 4 online. aa, amino acid; CHO, carbohydrate; PS, photosynthesis; TCA, tricarboxylic acid cycle.

CLAVATA1 and ERECTA in clusters four, five, and six, meaning that they were downregulated in the C_4 species. CLAVATA1 and ERECTA are known to be involved in cell and also organ differentiation by mediating cell–cell communication (Shiu and Bleecker, 2001; van Zanten et al., 2009). Although the function of the majority of the other proteins is unknown, their cumulative appearance suggests a relationship to the different types of photosynthesis or leaf architecture (see Supplemental Data Set 3 online).

DISCUSSION

Comparison of *Flaveria* Leaf Transcriptomes by Next-Generation Sequencing

We used 454 sequencing to analyze the leaf transcriptomes of five *Flaveria* species exhibiting different modes of photosynthesis and identified ESTs corresponding to between 55 and 61% of the *Arabidopsis* transcripts included in the minimal coding sequences set we used for mapping in the individual leaf cDNA libraries of the five species. Approximately 60% of the known 33,282 *Arabidopsis* genes show a detectable expression in the aboveground part of *Arabidopsis* seedlings (Weber et al., 2007). Assuming that comparable fractions of genes were expressed in the leaves of the investigated *Flaveria* species, a large proportion of the leaf transcriptomes of all five species was captured. This assumption was supported by the fact that the number of captured transcripts only slightly increased in the *F. trinervia*/*F. ramosissima*/*F. robusta* experiment compared with the *F. bidentis*/*F. pringlei* experiment, although nearly twice as many reads were available for the former.

The coverage of the individual functional gene classes was >50% for most classes and comparable for all species, indicating that the leaf transcriptomes of the *Flaveria* species were sampled to a comparable extent. Two complementary analyses were conducted using these data: (1) a gene-by-gene comparison using statistical tests based on the two experiments and (2) a global analysis using clustering tools.

The gene-by-gene comparison resulted in 463 differentially expressed genes in the *F. bidentis*/*F. pringlei* experiment (corresponding to 3.4% of the transcripts detected within these two species) and 995 genes in the *F. trinervia*/*F. robusta* experiment (corresponding to 6.9% of the genes detected with this experiment). Since the more advanced GS TITANIUM sequencing technology, which was used for the *F. trinervia*/*F. robusta* experiment, created more reads and, thus, more statistical power for the Audic and Claverie algorithm, more differences were identified in this second experiment compared with the GS FLX experiment conducted on *F. bidentis* and *F. pringlei*. The transcript abundance of 213 genes was significantly different in both C_3 to C_4 comparisons, and many genes changed in the same direction without reaching a significant level. Only 31 genes exhibited opposing significant differential transcript abundances in both experiments. This was equivalent to 0.21% of the transcripts detected within the *F. trinervia*/*F. robusta* experiment, indicating that the vast majority of differences in transcript abundances found in this study is related to the different modes of photosynthesis rather than to the phylogenetic distance of the analyzed *Flaveria* species.

Leaf Transcriptomes Changed during C_4 Evolution

The cluster analysis resulted in six clusters with a C_4-related pattern. Taken together, these C_4 clusters contained 3582 transcripts. A total of 1418 of these genes were in clusters with C_4 upregulated genes, whereas 2164 genes are downregulated during C_4 evolution. These numbers are the current best estimate for transcript abundance changes related to C_4. Until a functional C_4 cycle is introduced into a C_3 plant, it will remain unknown how many of these transcript changes are necessary and sufficient to establish a C_4 cycle. Based on the multiple concurrent and parallel successful evolution of the C_4 trait in many plant families, it is likely that many of the changes will either be controlled by common gene regulatory networks (Westhoff and Gowik, 2010) or may have evolved after successful establishment of the C_4 cycle. In this experiment, the evolutionary progression can be established by comparing the intermediate species with the C_4 and the C_3 species based on PageMan analysis. While all known core C_4 genes were changed early during C_4 evolution, other major changes happened after the establishment of the C_4 cycle. In case of the nitrogen metabolism, amino acid synthesis and transcriptional machinery, which were reduced in the C_4 species, this is logically consistent, since first the highly abundant transcripts of the functional classes “Calvin-Benson cycle” or “photorespiration” had to be reduced. These reductions on the other hand require the existence of a fully functional C_4 cycle.

The majority of C_4-related genes are regulated at least in part at the level of transcript abundance (see above; Bräutigam et al., 2011). While the simple overrepresentation analysis based on the species by species comparisons suggests that changes within in the regulatory genes are statistically underrepresented the cluster analysis discovers a multitude of regulatory genes with C_4-related transcript patterns. They may be involved in the development and maintenance of C_4 leaves and are prime candidates for further analysis.

The Transcription of C_4 Cycle Genes Was Altered during C_4 Evolution in *Flaveria*

The C_4 *Flaveria* species are assigned to the NADP-ME C_4 photosynthesis type (Ku et al., 1991). This is reflected by our study. Transcript data, extractable enzyme activities, and the metabolite levels confirmed that the two C_4 *Flaveria* species *F. trinervia* and *F. bidentis* exclusively use the NADP-ME C_4 pathway. In addition to the classical NADP-ME genes, we found a strong upregulation of an Ala and an Asp-AT, indicating that the C_4 *Flaverias* also use amino acids as transport metabolites. The protein encoded by the contig of the upregulated ASP-AT, as well as its *Arabidopsis* counterpart (AT4G31990), is predicted to be localized to the chloroplast (ChloroP, AtASP5, 0.547; FtASP5, 0.539). This confirmed earlier results showing that C_4 species *F. bidentis* and *F. trinervia* use Asp to a variable extent as transport metabolite (Moore et al., 1986; Meister et al., 1996), whereas the majority of ASP-AT activity is localized to the chloroplasts in mesophyll as well as in bundle sheath cells (Moore et al., 1984; Meister et al., 1996).

Two additional enzymes are key in C_4 photosynthesis. We detected the strong and significant upregulation of an adenosine

monophosphate kinase gene in both C_4 species. This gene was also found to be upregulated in the C_4 plant *C. gynandra* (Bräutigam et al., 2011) and is thought to be involved in the processing of the adenosine monophosphate produced by the PPDK (Hatch and Slack, 1968). In the *F. bidentis*/*F. pringlei* experiment, we also identified two significantly upregulated inorganic pyrophosphatases (see Supplemental Table 2 online), which were also upregulated in *C. gynandra* (Bräutigam et al., 2011). The upregulation at the transcript level is consistently detected in different species and different genera, reinforcing physiological analysis in that the processing of the AMP and pyrophosphate generated by PPDK is an integral part of the C_4 cycle (Slack et al., 1969).

The next-generation sequencing analysis provided a model for the transport processes at the mesophyll chloroplast envelope. In addition to translocators for PEP and inorganic phosphate, triose phosphates, 3-phosphoglycerate, inorganic phosphate, pyruvate, sodium ions, and protons (BASS 2/NHD), we found a strong upregulation of the chloroplast DiT1 and DiT2. Thus, the pattern of DiT1 expression in the different *Flaveria* species was similar to the pattern of other genes directly involved in the C_4 cycle, supporting the assumption that DiT1 is indeed involved in the C_4 photosynthesis as the OAA/malate shuttle of mesophyll chloroplasts. The upregulation of plastidic Asp-AT in the C_4 *Flaverias* pointed to a role for a second dicarboxylate transporter, DiT2. DiT2 has a broader substrate spectrum than DiT1 and prefers Asp (Renné et al., 2003). Upregulation of DiT2 in the C_4 *Flaverias* prompted the hypothesis that DiT2 was involved in the exchange of Asp across mesophyll and bundle sheath chloroplast envelopes as part of the C_4 cycle. The DiT genes were not upregulated in the C_4 species *C. gynandra*. This coincides with the proposed function for the DiTs, since *C. gynandra* is a NAD-ME type plant and does not have to shuttle OAA, malate, or Asp across its chloroplast envelope.

Bundle sheath chloroplasts play a key role in NADP-ME C_4 photosynthesis. Next to the inorganic phosphate, which must exhibit high activities also in the bundle sheath chloroplasts, no further C_4-related bundle sheath chloroplast transporter is known to date. No candidates exist for the malate importer or the pyruvate exporter. *Flaveria* contains an additional gene belonging to the bile acid sodium symporter family, BASS 4, which was upregulated to an extent comparable with other C_4 genes. In contrast with the chloroplastic pyruvate transporter BASS 2, this gene was not upregulated in the C_4 species *C. gynandra* when compared with the C_3 species *C. spinosa*. Since neither pyruvate export nor malate import at NAD-ME bundle sheath chloroplasts is required, it is tempting to hypothesize that this transporter might be involved in either pyruvate export from or malate import into the bundle sheath chloroplast.

The comparison also revealed a number of transport proteins with unknown or predicted functions only. These may play accessory roles in transport by creating or dissipating gradients needed for or caused by C_4 related transport, much as adenosine monophosphate kinase and pyrophosphatase are needed to balance metabolism. Two of these transport proteins, VDAC and AVP1, were also significantly altered in the C_4 plant *C. gynandra* when compared with the C_3 plant *C. spinosa* (Bräutigam et al., 2011), indicating a potential relevance of these genes for C_4 photosynthesis. The AVP1 transcripts are significantly more abundant in *F. trinervia* and *C. gynandra* than in the respective C_3 species (Bräutigam et al., 2011). Also, in *F. bidentis*/*F. pringlei*, AVP1 was upregulated, although the difference was not significant. In the C_3-C_4 intermediate plant *F. ramosissima*, AVP1 abundance was intermediate compared with *F. trinervia* and *F. robusta*. *Arabidopsis* AVP1 mutants show defects in leaf and root development since AVP1 affects polar auxin transport (Li et al., 2005).

Up to Three Distinct CO_2 Concentration Mechanisms Operate in the C_3-C_4 Intermediate *F. ramosissima*

In *F. ramosissima*, the transcripts of the genes related to the NADP-ME type C_4 photosynthesis showed intermediate levels compared with the C_3 plant *F. robusta* and the C_4 plant *F. trinervia.* This implies that in *F. ramosissima*, the C_4 cycle is working to a certain extent and that *F. ramosissima* is a true intermediate based on its transcriptional profile. This is in agreement with earlier results showing that, in *F. ramosissima*, more than 40% of the CO_2 is directly fixed into the C_4 acids malate and Asp (Ku et al., 1991). Based on the transcriptional profile, *F. ramosissima* is intermediate with regard to a NADP-ME type C_4 cycle. This is also reflected in the changes of photosynthetic electron transport chain gene expression.

F. ramosissima is also intermediate with regard to protein synthesis. The only downregulated transcripts related to the Calvin-Benson cycle were those of Rubisco, while in C_4 species, the majority of Calvin-Benson cycle transcripts were downregulated. Unlike C_4 species, which had a downregulated photorespiratory cycle, *F. ramosissima* accumulated more photorespiratory transcripts. In consequence, only plastidic but not cytosolic elements of the protein biosynthesis machinery were downregulated, and no changes in amino acid metabolism were detected. Hence, *F. ramosissima* was not capable of fully reaping the nitrogen benefits of C_4 photosynthesis, as indicated by its intermediate C/N ratio.

We found significant upregulation of genes related to the NAD-ME type C_4 pathway like cytoplasmic and mitochondrial ASP-AT genes, an mNAD-MDH, or two mNAD-ME. This transcript profile provokes the hypothesis that a (partial?) NAD-ME type C_4 cycle is active in addition to the NADP-ME type C_4 pathway in the C_3-C_4 intermediate. This was confirmed by the extractable NAD-ME activity that was significantly higher in the leaves of *F. ramosissima* than in the other four *Flaveria* species. Also, an analysis of the steady state metabolite levels suggested a similar conclusion. While the Ala level in *F. ramosissima* was comparable to those found in the two C_4 species, the Asp level exceeded those of all other *Flaverias*. Since Ala and Asp are the predominant transport metabolites in NAD-ME plants (Hatch, 1987), these findings supported the hypothesis of a NAD-ME type C_4 cycle in *F. ramosissima*. This finding was surprising since all true C_4 *Flaveria* species belong exclusively to the NADP-ME C_4 plants (Drincovich et al., 1998; this article). It is not clear if this reflects plasticity in the photosynthetic metabolism of *Flaverias* during C_4 evolution that was lost after a fully developed NADP-ME cycle was established or if *F. ramosissima* developed the NAD-ME cycle after splitting from the *Flaveria* lineage leading to true C_4

species. This will be clarified in the future by analyzing further C_3-C_4 intermediate *Flaveria* species.

This study also provided comprehensive molecular evidence for a photorespiratory CO_2 concentration mechanism in the C_3-C_4 intermediate species, which was previously hypothesized to represent a biochemical CO_2 pump (Rawsthorne et al., 1988a, 1988b) and might have been an intermediate step toward the evolution of the C_4 pathway (Bauwe et al., 1987; Sage, 2004; Bauwe, 2011).

Photorespiratory genes were expressed at a higher level than even in the C_3 species and, importantly, also the steady state levels of Gly and Ser, the transport metabolites of the photorespiratory CO_2 pump, were higher in *F. ramosissima* compared with the C_3 and C_4 *Flaveria* species. Hence, a photorespiratory CO_2 pump may still operate in *F. ramosissima*. Based on the available data, three distinct CO_2 concentrating mechanisms, the NADP-ME-, the NAD-ME-type, and the photorespiratory Gly shuttle, operate in parallel in *F. ramosissima* (Figure 7).

To produce a consistent model (Figure 7), it is critical to consider the ammonia balance between the cell types. The photorespiratory CO_2 pump moves two molecules of Gly to the bundle sheath cells where they are decarboxylated, leading to one molecule each of Ser, CO_2, and ammonium. Without compensation, this would lead to a massive accumulation of ammonia in the bundle sheath cells, even if the resulting Ser is transported back to the mesophyll cells for phosphoglycerate regeneration as proposed in Figure 7 and as supported by the high steady state Ser levels in *F. ramosissima* leaves. In case of *F. ramosissima* with a working C_4 cycle, ammonia can be balanced by adjusting the ratios of the transport metabolites Ala/pyruvate and Asp/malate. For less advanced C_3-C_4 intermediates, which solely rely on the photorespiratory CO_2 concentration mechanism, the imbalance also needs to be solved. Means to transport ammonia from bundle sheath to mesophyll cells, like a Glu-oxoglutarate shuttle, an Ala-pyruvate shuttle, or an aspartate-malate shuttle, would be required in these less advanced intermediates. One of the latter two might have been a starting point for the evolution of a metabolite transport framework needed for the C_4 cycle. If an Ala-pyruvate shuttle and an Asp-malate shuttle would exist in parallel in a single species, only minor alterations to these pathways, in a way that malate and Asp are transported from the mesophyll to the bundle sheath cells and Ala and pyruvate are transferred back would be necessary to establish a C_4-like CO_2 transport pathway that could replace the photorespiratory Gly/Ser pump.

Comparison to C_4 Photosynthesis in the Genus *Cleome*: Common Themes of C_4 Evolution

Two comparative transcriptome studies on closely related C_3 and C_4 species from the dicot genera *Flaveria* and *Cleome*

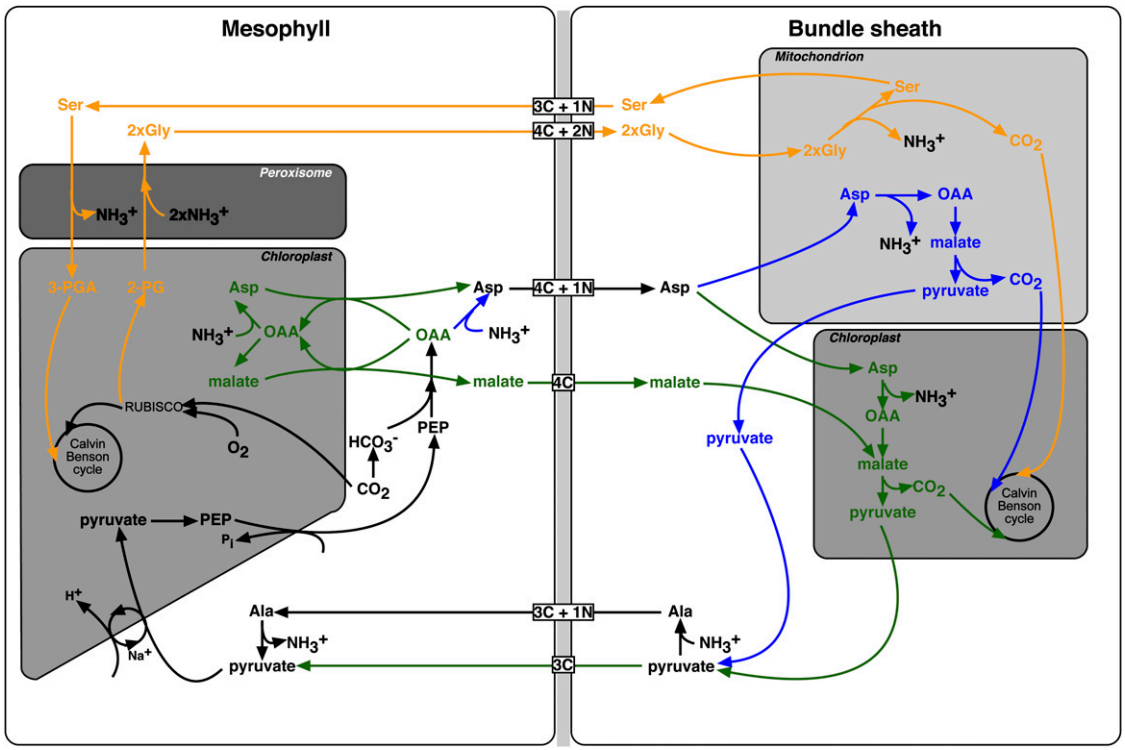

Figure 7. Schematic of the CO_2 Concentrating and Photorespiratory Pathways in the C_3-C_4 Intermediate Species *F. ramosissima.*

Three distinct CO_2 concentrating mechanisms, the NADP-ME type (green), the NAD-ME type (blue) C_4 pathway, and the photorespiratory Gly shuttle (orange), operate in parallel in this C_3-C_4 intermediate. *F. ramosissima*, with a working C_4 cycle, can compensate for the massive ammonia imbalance introduced by the photorespiratory CO_2 pump, by adjusting the ratios of the transport metabolites Ala/pyruvate and Asp/malate.

(Bräutigam et al., 2011) estimate the common differences between C_3 and C_4 leaf metabolism: in both species, C_4 cycle–related transcripts for enzyme reactions and transport processes are upregulated, although the types of enzymes and transporter partially reflect the preferred decarboxylation enzymes NADP-ME and NAD-ME. Next to the C_4 cycle, the biggest differences between C_3 and C_4 species were found for genes related to photosynthesis in both studies. The Calvin-Benson cycle and the photorespiratory pathway were downregulated. Photosystem I and the cyclic electron transfer were upregulated, whereas photosystem II was downregulated at least in *Flaveria*. This may reflect different ATP and NADPH demands related to the different modes of photosynthesis. For some C_4 NADP-ME type grass species, it was shown that photosystem II expression in the bundle sheath cells is suppressed (Meierhoff and Westhoff, 1993; Majeran and van Wijk, 2009) to suppress oxygen evolution near Rubisco. However, this is most likely not the reason for reduced photosystem II transcript levels in the C_4 and C_3-C_4 intermediate *Flaverias* as it was shown earlier that at least in the C_4 plant *F. trinervia* transcripts for most photosystem II subunits were found to be equally abundant in mesophyll and bundle sheath cells, whereas the photosystem II subunits involved in oxygen evolution seem to be downregulated posttrancriptionally in the bundle sheath (Höfer et al., 1992).

The translational machinery of the cytosol and the chloroplast, nitrogen, and amino acid metabolism were downregulated in the C_4 species, which likely reflects the lowered demands of highly abundant Calvin-Benson cycle and photorespiratory proteins. Also, we found several genes not obviously related to the C_4 pathway whose expression was altered similarly in both genera like the transporters AVP1 and VDAC1, making them prime candidates for further research. To complete our picture of common and essential C_4-related alterations in gene expression, the analysis of further closely related C_3 C_4 pairs, especially C_4 plants of the PEP-CK type, would be highly desirable.

The comparison of the transcriptomes of closely related C_3 and C_4 species from genera like *Flaveria*, in this study, or *Cleome* (Bräutigam et al., 2011) elucidate the evolutionary trajectories of C_4 evolution and reveal how many and which genes were altered during the transitions from C_3 into C_4 plants.

We analyzed the transcriptomes of five nonmodel species belonging to the genus *Flaveria* exhibiting a C_3, C_4, or C_3-C_4 intermediate mode of photosynthesis. We discovered quantitative differences in transcript abundance between the species with a different mode of photosynthesis and identified metabolic pathways and functional groups of genes that are altered as a whole, such as photosynthesis or nitrogen and amino acid metabolism, and the translational machinery of the cytosol and the chloroplast. The next step toward recreating C_4 in a C_3 dicot would be to analyze how these differences and the alterations in cell and tissue architecture are established during leaf development by, for example, comparative developmental time series monitoring the transcriptomes of C_3 and C_4 leaves throughout a developmental gradient.

A transcriptome analysis of four developmental stages as well as mesophyll and bundle sheath cells of a maize leaf has recently been published (Li et al., 2010). This study provides detailed insight into the photosynthetic development of a maize leaf and will surely help in understanding the regulatory and metabolic networks involved in grass leaf development. We feel that a combination of two approaches, (1) the analysis of different developmental stages from (2) closely related species exhibiting different modes of photosynthesis, will uncover the differences in leaf development of C_3 and C_4 plants.

The combination of genome information of a closely related C_3 C_4 species pair with the quantitative expression information already available would create a powerful tool to define *cis*-regulatory changes. The genus *Flaveria* is a prime candidate for such a sequencing project since four decades of research have made *Flaveria* the most intensely studied dicot model system with respect to C_4 evolution.

METHODS

Plant Material, RNA Isolation, and cDNA Synthesis

Flaveria bidentis, *Flaveria pringlei*, *Flaveria ramosissima*, *Flaveria robusta*, and *Flaveria trinervia* plants were grown in 17-cm pots on soil (C-400 with Cocopor [Stender Erden] fertilized with 3 g/L Osmocote exact standard 3 to 4 M [Scotts]). Plants were grown in the greenhouse from May to August with additional light for 16 h per day until 50 to 60 cm height and before the onset of flowering. For the isolation of total RNA, the second and fourth visible leaves from the top of 10 plants each were harvested in the middle of the light period and immediately frozen in liquid nitrogen. Total RNA was isolated as described earlier (Westhoff et al., 1991). Poly(A^+) RNA was enriched by two consecutive rounds of oligo(dT) purification with the Oligotex mRNA Midi kit (Qiagen). For *F. bidentis* and *F. pringlei*, cDNA was prepared with the SMART PCR cDNA synthesis kit (Clontech-Takara Bio) using 1 μg poly(A^+) RNA as starting material and 11 (*F. bidentis*) or 12 (*F. pringlei*) PCR cycles. In case of *F. ramosissima*, *F. robusta*, and *F. trinervia*, the SMARTer PCR cDNA synthesis kit (Clontech-Takara Bio), with 300 ng of poly(A^+) RNA as starting material and 15 PCR cycles was used. The cDNA was purified with the QIAquick PCR purification kit (Qiagen). Purity and integrity of total RNA, poly(A^+) RNA, and cDNA was verified spectroscopically with a NanoDrop ND-1000, with the Agilent 2100 Bioanalyzer, and by agarose gel electrophoresis.

454 Sequencing

To avoid an irregular sequence distribution, the cDNAs were ligated to concatemers prior to nebulization. Preparation of sequencing libraries and 454 sequencing (Roche Applied Science) followed standard procedures. One sequencing run on a GS FLX system was conducted on the cDNA libraries from *F. bidentis* and *F. pringlei*. For *F. trinervia*, *F. robusta*, and *F. ramosissima* cDNA libraries, the more advanced 454 TITANIUM technology was used for bioinformatic separation of sequence reads joined via concatemerization with the SMART adaptor sequences. The 454 sequencing and processing of raw data were performed by GATC Biotech.

Mapping/Statistics

Clean reads were aligned to a minimal set of coding sequences of the TAIR 9 release of the *Arabidopsis thaliana* genome (http://www.Arabidopsis.org/) using BLAT (Kent, 2002). The minimal set of coding sequences was used to reduce the effects of the different evolutionary histories of the *Arabidopsis* and *Flaveria* genomes; for example, splitting of reads on paralogous genes that exist in the one but not in the other species (Bräutigam and Gowik, 2010). It was generated by excluding genes shown to have arisen by segmental or tandem duplication (Bowers

et al., 2003; Haberer et al., 2004; Thomas et al., 2006) within the *Arabidopsis* genome, whereas the gene with the lowest AGI number was retained in each case. The minimal transcriptome contains 21,971 of the 27,379 protein coding genes listed by the TAIR 9 release of the *Arabidopsis* genome (http://www.Arabidopsis.org/). Alignments were performed in the protein space, the best hit for each 454 read was retained using an in-house PERL script, and hit counts were then transformed to reads per million to normalize for the number of reads available for each species.

Log2 expression ratios were calculated after adding a pseudocount to the number of reads in each species to account for any zero counts. Differentially expressed transcripts were identified using the Poisson statistics developed by Audic and Claverie (1997) as implemented in Discovery Space (Robertson et al., 2007) on the non-normalized read counts followed by a Bonferroni correction to account for the accumulation of α-type errors when conducting multiple pairwise comparisons.

Quantitative RT-PCR

Prior to quantitative RT-PCR, total RNA was treated with DNase I (Roche Applied Science) followed by phenol/chloroform extraction and ethanol precipitation. Reverse transcription was performed with the QuantiTect reverse transcription kit (Qiagen) using 1 μg of RNA. Quantitative PCR was performed with an ABI 5700 real-time PCR system (Applied Biosystems) using the QuantiTect SYBER Green PCR kit (Qiagen) and a 50-fold dilution of the cDNAs. Initial denaturation and activation of the enzyme was performed at 95°C for 15 min followed by 40 cycles of 94°C for 15 s, 60°C for 30 s, and 72°C for 30 s. Primers were designed to produce PCR products of ~200 bp and to have a temperature of 60°C. The specificity of PCR reactions was verified by melting curve analysis and agarose gel electrophoresis. To estimate the efficiency of the PCR reactions, four consecutive 10-fold dilutions of the cDNAs were tested with each primer pair. Only reactions with efficiencies >90% were considered for further analysis. The $\Delta\Delta$CT method, using the UBQ3 homologous of *F. bidentis* and *F. pringlei* as reference genes, was used to quantify relative transcript abundance. All quantitative PCR reactions were performed in triplicates.

Overrepresentation Analysis and K-Means Clustering

To identify functional MapMan categories with significant differences in the C_3 and C_4 species, we used all expressed genes found for the five species and applied Fisher's exact test followed by a Bonferroni correction as implemented in PageMan software (Usadel et al., 2006).

K-means clustering was performed for all expressed genes with the MEV program (http://www.tm4.org/mev). The number of clusters (26 clusters) was defined by the Figures of Merit application. The K-means analysis was performed using the K-means/K-medians clustering module (KMC) with the Euclidean distance criterion. The Wilcoxon statistics, including the Benjamini-Hochberg correction as implemented in PageMan (Usadel et al., 2006), were used to test the clusters for enrichment of genes of functional MapMan categories.

Polar Metabolites and Enzyme Activity Analysis

For metabolite analysis and enzyme activity determination, mature leaves were collected in the middle of the light period and immediately frozen in liquid nitrogen. Three independent biological replicates were used. The tissues were ground in a mortar and a 50-mg fresh weight aliquot was extracted using the procedure described by Lee and Fiehn (2008). Ribitol was used as an internal standard for data normalization. For gas chromatography electron impact ionization time-of-flight mass spectrometry analysis, samples were processed and analyzed according to Lee and Fiehn (2008). Enzyme activities were determined according to Häusler et al. (2001).

Element Analysis and δ^{13}C Determination

For element and stable isotope analysis, a fine powder of mature leaves from two independent cultivations of the different *Flaveria* species was lyophilized and analyzed using an Isoprime 100 isotope ratio mass spectrometer coupled to an Elementar elemental analyzer (ISOTOPE cube; Elementar Analysensysteme) following the manufacturer's recommendations. The calibration for δ^{13}C measurements followed the two-point method described by Coplen et al. (2006).

Accession Number

The read data have been submitted to the National Center for Biotechnology Information Short Read Archive under accession number SRP006166.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Determination of Differences in Transcript Abundance in *F. bidentis* and *F. pringlei* via RNA-Seq and qPCR Lead to Similar Results.

Supplemental Figure 2. Transcript Levels of Genes Related to the Light Reaction of Photosynthesis Are Altered in *F. trinervia* (C_4), *F. ramosissima* (C_3-C_4), *F. robusta* (C_3), *F. bidentis* (C_4), and *F. pringlei* (C_3).

Supplemental Figure 3. Cluster Analysis of Transcript Abundances in *F. bidentis* (C_4), *F. trinervia* (C_4), *F. ramosissima* (C_3-C_4), *F. robusta* (C_3), and *F. pringlei* (C_3).

Supplemental Figure 4. Functional Category (MapMan Bins) Enrichment among the 26 Clusters from K-Means Clustering.

Supplemental Table 1. Carbon Isotope Ratios in Dried Leaf Material of *F. bidentis*, *F. trinervia*, *F. ramosissima*, *F. robusta*, and *F. pringlei*.

Supplemental Table 2. Transcript Abundance of C_4 Cycle Genes and C_4-Related Transporters.

Supplemental Table 3. Quantitation of C_4 Marker Enzyme Activities in Leaf Extracts of *F. bidentis*, *F. trinervia*, *F. ramosissima*, *F. robusta*, and *F. pringlei.*

Supplemental Table 4. Relative Abundance of Predominant Metabolites Detected by GC-EI-TOF in the Leaves of *F. bidentis*, *F. trinervia*, *F. ramosissima*, *F. robusta*, and *F. pringlei.*

Supplemental Table 5. Transcript Abundance of Genes Related to Chloroplast Biogenesis and Maintenance.

Supplemental Table 6. Carbon-to-Nitrogen Ratios in the Leaves of *F. bidentis*, *F. trinervia*, *F. ramosissima*, *F. robusta*, and *F. pringlei*.

Supplemental Data Set 1. Excel Worksheet Providing Quantitative Information for All Reads Mapped onto the Reference Transcriptome from *Arabidopsis thaliana*.

Supplemental Data Set 2. All Genes Exhibiting a C_4-Related Expression Pattern (Clusters One to Six).

Supplemental Data Set 3. All Regulatory Genes Exhibiting a C_4-Related Expression Pattern (Clusters One to Six).

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft through the Collaborative Research Center SFB590 and the Research Group FOR1186.

AUTHOR CONTRIBUTIONS

U.G., A.B., A.P.M.W., and P.W. designed the research. U.G., A.B., and K.L.W. performed the research. U.G., A.B., A.P.M.W., and P.W. analyzed data. U.G., A.B., and A.P.M.W. wrote the article.

Received April 11, 2011; revised May 24, 2011; accepted June 15, 2011; published June 24, 2011.

REFERENCES

Aoki, N., Ohnishi, J., and Kanai, R. (1992). Two different mechanisms for transport of pyruvate into mesophyll chloroplasts of C_4 plants–A comparative study. Plant Cell Physiol. **33:** 805–809.

Audic, S., and Claverie, J.M. (1997). The significance of digital gene expression profiles. Genome Res. **7:** 986–995.

Bauwe, H. (2011). Photorespiration: The bridge to C4 photosynthesis. In C4 Photosynthesis and Related CO_2 Concentrating Mechanisms, A.S. Raghavendra and R.F. Sage, eds (Dordrecht, The Netherlands: Springer), pp. 81–108.

Bauwe, H., Keerberg, O., Bassüner, R., Pärnik, T., and Bassüner, B. (1987). Reassimilation of carbon dioxide by Flaveria (Asteraceae) species representing different types of photosynthesis. Planta **172:** 214–218.

Beligni, M.V., and Mayfield, S.P. (2008). *Arabidopsis thaliana* mutants reveal a role for CSP41a and CSP41b, two ribosome-associated endonucleases, in chloroplast ribosomal RNA metabolism. Plant Mol. Biol. **67:** 389–401.

Black, C.C.J. (1973). Photosynthetic carbon fixation in relation to net CO_2 uptake. Annu. Rev. Plant Physiol. **24:** 253–286.

Bollenbach, T.J., Sharwood, R.E., Gutierrez, R., Lerbs-Mache, S., and Stern, D.B. (2009). The RNA-binding proteins CSP41a and CSP41b may regulate transcription and translation of chloroplast-encoded RNAs in Arabidopsis. Plant Mol. Biol. **69:** 541–552.

Botha, C.E.J. (1992). Plasmodesmatal distribution, structure and frequency in relation to assimilation in C_3 and C_4 grasses in southern Africa. Planta **187:** 348–358.

Bowers, J.E., Chapman, B.A., Rong, J.K., and Paterson, A.H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422:** 433–438.

Bräutigam, A., and Gowik, U. (2010). What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. Plant Biol (Stuttg) **12:** 831–841.

Bräutigam, A., Hoffmann-Benning, S., and Weber, A.P. (2008). Comparative proteomics of chloroplast envelopes from C3 and C4 plants reveals specific adaptations of the plastid envelope to C4 photosynthesis and candidate proteins required for maintaining C4 metabolite fluxes. Plant Physiol. **148:** 568–579. Erratum. Plant Physiol. **148:** 1734.

Bräutigam, A., et al. (2011). An mRNA blueprint for C4 photosynthesis derived from comparative transcriptomics of closely related C3 and C4 species. Plant Physiol. **155:** 142–156.

Bräutigam, A., and Weber, A.P.M. (2011). Transport processes: Connecting the reactions of C4 photosynthesis. In C4 Photosynthesis and Related CO_2 Concentrating Mechanisms, A.S. Raghavendra and R.F. Sage, eds (Dordrecht, The Netherlands: Springer), pp. 199–219.

Brown, R.H. (1999). Agronomic implications of C4 photosynthesis. In C4 Plant Biology, R.F. Sage and R.K. Monson, eds (San Diego, CA: Academic Press), pp. 473–507.

Chen, K.-M., Holmström, M., Raksajit, W., Suorsa, M., Piippo, M., and Aro, E.-M. (2010). Small chloroplast-targeted DnaJ proteins are involved in optimization of photosynthetic reactions in *Arabidopsis thaliana*. BMC Plant Biol. **10:** 43.

Coplen, T.B., Brand, W.A., Gehre, M., Gröning, M., Meijer, H.A., Toman, B., and Verkouteren, R.M. (2006). New guidelines for delta13C measurements. Anal. Chem. **78:** 2439–2441.

Dengler, N.G., and Nelson, T. (1999). Leaf structure and development in C4 plants. In C4 Plant Biology, R.F. Sage and R.K. Monson, eds (San Diego, CA: Academic Press), pp. 133–172.

Drincovich, M.F., Casati, P., Andreo, C.S., Chessin, S.J., Franceschi, V.R., Edwards, G.E., and Ku, M.S.B. (1998). Evolution of C4 photosynthesis in Flaveria species. Isoforms of NADP-malic enzyme. Plant Physiol. **117:** 733–744.

Edwards, G.E., Franceschi, V.R., and Voznesenskaya, E.V. (2004). Single-cell C(4) photosynthesis versus the dual-cell (Kranz) paradigm. Annu. Rev. Plant Biol. **55:** 173–196.

Edwards, G.E., and Ku, M.S.B. (1987). Biochemistry of C_3-C_4 intermediates. In The Biochemistry of Plants, Vol. 10, M.D. Hatch and N.K. Boardman, eds (New York: Academic Press), pp. 275–325.

Gao, H.B., Kadirjan-Kalbach, D., Froehlich, J.E., and Osteryoung, K.W. (2003). ARC5, a cytosolic dynamin-like protein from plants, is part of the chloroplast division machinery. Proc. Natl. Acad. Sci. USA **100:** 4328–4333.

Ghannoum, O., Evans, J.R., and Caemmerer, S. (2011). Nitrogen and water use efficiency of C4 plants. In C4 Photosynthesis and Related CO_2 Concentrating Mechanisms, A.S. Raghavendra and R.F. Sage, eds (Dordrecht, The Netherlands: Springer), pp. 129–146.

Gonzalez, N., et al. (2010). Increased leaf size: Different means to an end. Plant Physiol. **153:** 1261–1279.

Gowik, U., and Westhoff, P. (2011). The path from C3 to C4 photosynthesis. Plant Physiol. **155:** 56–63.

Haberer, G., Hindemitt, T., Meyers, B.C., and Mayer, K.F.X. (2004). Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of Arabidopsis. Plant Physiol. **136:** 3009–3022.

Haberlandt, G. (1904). Physiologische Pflanzenanatomie, 3rd ed. (Leipzig, Germany: Verlag von Wilhelm Engelmann).

Hatch, M.D. (1987). C_4 photosynthesis: A unique blend of modified biochemistry, anatomy and ultrastructure. Biochim. Biophys. Acta **895:** 81–106.

Hatch, M.D., and Slack, C.R. (1968). A new enzyme for the interconversion of pyruvate and phosphopyruvate and its role in the C4 dicarboxylic acid pathway of photosynthesis. Biochem. J. **106:** 141–146.

Häusler, R.E., Rademacher, T., Li, J., Lipka, V., Fischer, K.L., Schubert, S., Kreuzaler, F., and Hirsch, H.J. (2001). Single and double overexpression of $C(_4)$-cycle genes had differential effects on the pattern of endogenous enzymes, attenuation of photorespiration and on contents of UV protectants in transgenic potato and tobacco plants. J. Exp. Bot. **52:** 1785–1803.

Höfer, M.U., Santore, U.J., and Westhoff, P. (1992). Differential accumulation of the 10-, 16- and 23-kDa peripheral components of the water-splitting complex of photosystem II in mesophyll and bundle-sheath chloroplasts of the dicotyledonous C_4 plant *Flaveria trinervia* (Spreng.) C. Mohr. Planta **186:** 304–312.

Kato, Y., Miura, E., Ido, K., Ifuku, K., and Sakamoto, W. (2009). The variegated mutants lacking chloroplastic FtsHs are defective in D1 degradation and accumulate reactive oxygen species. Plant Physiol. **151:** 1790–1801.

Kent, W.J. (2002). BLAT—The BLAST-like alignment tool. Genome Res. **12:** 656–664.

Kim, J., Rudella, A., Ramirez Rodriguez, V., Zybailov, B., Olinares, P.D.B., and van Wijk, K.J. (2009). Subunits of the plastid ClpPR protease complex have differential contributions to embryogenesis, plastid biogenesis, and plant development in *Arabidopsis*. Plant Cell **21:** 1669–1692.

Kinoshita, H., Nagasaki, J., Yoshikawa, N., Yamamoto, A., Takito, S., Kawasaki, M., Sugiyama, T., Miyake, H., Weber, A.P.M., and

Taniguchi, M. (2011). The chloroplastic 2-oxoglutarate/malate transporter has dual function as the malate valve and in carbon/nitrogen metabolism. Plant J. **65:** 15–26.

Ku, M.S.B., Schmitt, M.R., and Edwards, G.E. (1979). Quantitative determination of RuBP carboxylase oxygenase protein in leaves of several C3 and C4 plants. J. Exp. Bot. **30:** 89–98.

Ku, M.S.B., Wu, J., Dai, Z., Scott, R.A., Chu, C., and Edwards, G.E. (1991). Photosynthetic and photorespiratory characteristics of *flaveria* species. Plant Physiol. **96:** 518–528.

Lee, Y., and Fiehn, O. (2008). High quality metabolomic data for *Chlamydomonas reinhardtii*. Plant Methods **4:** 7.

Leegood, R.C. (2002). C(4) photosynthesis: Principles of CO(2) concentration and prospects for its introduction into C(3) plants. J. Exp. Bot. **53:** 581–590.

Lerbs-Mache, S. (2011). Function of plastid sigma factors in higher plants: Regulation of gene expression or just preservation of constitutive transcription? Plant Mol. Biol. **76:** 235–249.

Lezhneva, L., Amann, K., and Meurer, J. (2004). The universally conserved HCF101 protein is involved in assembly of [4Fe-4S]-cluster-containing complexes in *Arabidopsis thaliana* chloroplasts. Plant J. **37:** 174–185.

Li, J., et al. (2005). Arabidopsis H^+-PPase AVP1 regulates auxin-mediated organ development. Science **310:** 121–125.

Li, P., et al. (2010). The developmental dynamics of the maize leaf transcriptome. Nat. Genet. **42:** 1060–1067.

Long, S.P. (1999). Environmental responses. In C4 Plant Biology, R.F. Sage and R.K. Monson, eds (San Diego, CA: Academic Press), pp. 215–249.

Majeran, W., and van Wijk, K.J. (2009). Cell-type-specific differentiation of chloroplasts in C4 plants. Trends Plant Sci. **14:** 100–109.

McKown, A.D., Moncalvo, J.-M., and Dengler, N.G. (2005). Phylogeny of Flaveria (Asteraceae) and inference of C4 photosynthesis evolution. Am. J. Bot. **92:** 1911–1928.

Meierhoff, K., and Westhoff, P. (1993). Differential biogenesis of photosystem II in mesophyll and bundle-sheath cells of monocotyledonous NADP-malic enzyme-type C_4 plants: The non-stoichiometric abundance of the subunits of photosystem II in the bundle-sheath chloroplasts and the translational activity of the plastome-encoded genes. Planta **191:** 23–33.

Meister, M., Agostino, A., and Hatch, M.D. (1996). The roles of malate and aspartate in C_4 photosynthetic metabolism of *Flaveria bidentis* (L). Planta **199:** 262–269.

Monson, R.K., Moore, B.d., Ku, M.S.B., and Edwards, G.E. (1986). Co-function of C_3- and C_4-photosynthetic pathways in C_3, C_4 and C_3-C_4 intermediate *Flaveria* species. Planta **168:** 493–502.

Monson, R.K., Teeri, J.A., Ku, M.S.B., Gurevitch, J., Mets, L.J., and Dudley, S. (1988). Carbon-isotope discrimination by leaves of *Flaveria* species exhibiting different amounts of C_3- and C_4-cycle co-function. Planta **174:** 145–151.

Moore, B.d., Cheng, S.H., and Edwards, G.E. (1986). The influence of leaf development on the expression of C_4 metabolism in *Flaveria trinervia*, a C_4 dicot. Plant Cell Physiol. **27:** 1159–1167.

Moore, B.d., Ku, M.S.B., and Edwards, G.E. (1984). Isolation of leaf bundle sheath protoplasts from C4 dicot species and intracellular localization of selected enzymes. Plant Sci. Lett. **35:** 127–138.

Muhaidat, R., Sage, R.F., and Dengler, N.G. (2007). Diversity of Kranz anatomy and biochemistry in C4 eudicots. Am. J. Bot. **94:** 362–381.

Oaks, A. (1994). Efficiency of nitrogen utilization in C_3 and C_4 cereals. Plant Physiol. **106:** 407–414.

Okushima, Y., Mitina, I., Quach, H.L., and Theologis, A. (2005). AUXIN RESPONSE FACTOR 2 (ARF2): A pleiotropic developmental regulator. Plant J. **43:** 29–46.

O'Leary, M.H. (1981). Carbon isotope fractionation in plants. Phytochemistry **20:** 553–567.

Osborne, C.P., and Freckleton, R.P. (2009). Ecological selection pressures for C-4 photosynthesis in the grasses. Philos. Trans. R. Soc. Lond. B Biol. Sci. **276:** 1753–1760.

Rawsthorne, S., Hylton, C.M., Smith, A.M., and Woolhouse, H.W. (1988a). Distribution of photorespiratory enzymes between bundle-sheath and mesophyll cells in leaves of the C_3-C_4 intermediate species *Moricandia arvensis* (L.) DC. Planta **176:** 527–532.

Rawsthorne, S., Hylton, C.M., Smith, A.M., and Woolhouse, H.W. (1988b). Photorespiratory metabolism and immunogold localization of photorespiratory enzymes in leaves of C3 and C3-C4 intermediate species of Moricandia. Planta **173:** 298–308.

Renné, P., Dressen, U., Hebbeker, U., Hille, D., Flügge, U.-I., Westhoff, P., and Weber, A.P. (2003). The Arabidopsis mutant *dct* is deficient in the plastidic glutamate/malate translocator DiT2. Plant J. **35:** 316–331.

Robertson, N., Oveisi-Fordorei, M., Zuyderduyn, S.D., Varhol, R.J., Fjell, C., Marra, M., Jones, S., and Siddiqui, A. (2007). DiscoverySpace: An interactive data analysis application. Genome Biol. **8:** R6.

Sage, R.F. (2004). The evolution of C4 photosynthesis. New Phytol. **161:** 341–370.

Sane, A.P., Stein, B., and Westhoff, P. (2005). The nuclear gene HCF107 encodes a membrane-associated R-TPR (RNA tetratricopeptide repeat)-containing protein involved in expression of the plastidial psbH gene in Arabidopsis. Plant J. **42:** 720–730.

Shiu, S.-H., and Bleecker, A.B. (2001). Plant receptor-like kinase gene family: Diversity, function, and signaling. Sci. STKE **2001:** re22.

Slack, C.R., Hatch, M.D., and Goodchild, D.J. (1969). Distribution of enzymes in mesophyll and parenchyma-sheath chloroplasts of maize leaves in relation to the C4-dicarboxylic acid pathway of photosynthesis. Biochem. J. **114:** 489–498.

Thomas, B.C., Pedersen, B., and Freeling, M. (2006). Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome Res. **16:** 934–946.

Usadel, B., Nagel, A., Steinhauser, D., Gibon, Y., Bläsing, O.E., Redestig, H., Sreenivasulu, N., Krall, L., Hannah, M.A., Poree, F., Fernie, A.R., and Stitt, M. (2006). PageMan: An interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. BMC Bioinformatics **7:** 535.

van Zanten, M., Snoek, L.B., Proveniers, M.C.G., and Peeters, A.J.M. (2009). The many functions of ERECTA. Trends Plant Sci. **14:** 214–218.

Waters, M.T., Wang, P., Korkaric, M., Capper, R.G., Saunders, N.J., and Langdale, J.A. (2009). GLK transcription factors coordinate expression of the photosynthetic apparatus in *Arabidopsis*. Plant Cell **21:** 1109–1128.

Weber, A.P., Weber, K.L., Carr, K., Wilkerson, C., and Ohlrogge, J.B. (2007). Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. Plant Physiol. **144:** 32–42.

Weber, A.P.M., and von Caemmerer, S. (2010). Plastid transport and metabolism of C3 and C4 plants—Comparative analysis and possible biotechnological exploitation. Curr. Opin. Plant Biol. **13:** 257–265.

Wessinger, M.E., Edwards, G.E., and Ku, M.S.B. (1989). Quantity and kinetic properties of ribulose 1,5-bisphosphate carboxylase in C_3, C_4, and C_3-C_4 intermediate species of *Flaveria* (Asteraceae). Plant Cell Physiol. **30:** 665–671.

Westhoff, P., and Gowik, U. (2010). Evolution of C4 photosynthesis—Looking for the master switch. Plant Physiol. **154:** 598–601.

Westhoff, P., Offermann-Steinhard, K., Höfer, M., Eskins, K., Oswald, A., and Streubel, M. (1991). Differential accumulation of plastid transcripts encoding photosystem II components in the mesophyll and bundle-sheath cells of monocotyledonous NADP-malic enzyme-type C_4 plants. Planta **184:** 377–388.

Evolution of C_4 Photosynthesis in the Genus *Flaveria*: How Many and Which Genes Does It Take to Make C_4?

Udo Gowik, Andrea Bräutigam, Katrin L. Weber, Andreas P.M. Weber and Peter Westhoff

Plant Cell 2011;23;2087-2105; originally published online June 24, 2011;
DOI 10.1105/tpc.111.086264