Evolution of \( \text{C}_4 \) Photosynthesis in the Genus \textit{Flaveria}: How Many and Which Genes Does It Take to Make \( \text{C}_4 \)?

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

\( \text{a} \) Institute of Plant Molecular and Developmental Biology, Heinrich-Heine-University, 40225 Duesseldorf, Germany
\( \text{b} \) Institute of Plant Biochemistry, Heinrich-Heine-University, 40225 Duesseldorf, Germany

Selective pressure exerted by a massive decline in atmospheric \( \text{CO}_2 \) levels 55 to 40 million years ago promoted the evolution of a novel, highly efficient mode of photosynthetic carbon assimilation known as \( \text{C}_4 \) photosynthesis. \( \text{C}_4 \) species have concurrently evolved multiple times in a broad range of plant families, and this multiple and parallel evolution of the complex \( \text{C}_4 \) trait indicates a common underlying evolutionary mechanism that might be elucidated by comparative analyses of related \( \text{C}_3 \) and \( \text{C}_4 \) species. Here, we use mRNA-Seq analysis of five species within the genus \textit{Flaveria}, ranging from \( \text{C}_3 \) to \( \text{C}_4 \)-\( \text{C}_3 \) intermediate to \( \text{C}_4 \) species, to quantify the differences in the transcriptomes of closely related plant species with varying degrees of \( \text{C}_4 \)-associated characteristics. Single gene analysis defines the \( \text{C}_4 \) cycle enzymes and transporters more precisely and provides new candidates for yet unknown functions as well as identifies \( \text{C}_4 \) associated pathways. Molecular evidence for a photorespiratory \( \text{CO}_2 \) pump prior to the establishment of the \( \text{C}_4 \) cycle-based \( \text{CO}_2 \) pump is provided. Cluster analysis defines the upper limit of \( \text{C}_4 \)-related gene expression changes in mature leaves of \textit{Flaveria} as 3582 alterations.

INTRODUCTION

\( \text{C}_4 \) 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 \( \text{C}_4 \) cycle, in which the initial product of \( \text{CO}_2 \) fixation is a four-carbon (\( \text{C}_4 \)) organic acid rather than a three-carbon (\( \text{C}_3 \)) organic acid. In most \( \text{C}_4 \) species, \( \text{CO}_2 \) photosynthesis involves two different cell types, mesophyll and bundle sheath cells. Only few species have been described that carry out a \( \text{C}_4 \) cycle within a single cell (Edwards et al., 2004). As shown in Figure 1A, in an \( \text{NADP} \)-dependent malic enzyme type \( \text{C}_4 \) plant, \( \text{CO}_2 \) 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 \( \text{NADP} \)-malic enzyme (\( \text{NADP-ME} \)), the \( \text{CO}_2 \) 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 (PDPK) (Hatch, 1987).

\( \text{C}_4 \) plants show drastically reduced rates of photorespiration because \( \text{CO}_2 \) 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 \( \text{C}_4 \) photosynthesis, and the leaf structure of \( \text{C}_4 \) plants is altered compared with most \( \text{C}_3 \) 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 \( \text{CO}_2 \) pump ensures high rates of photosynthesis even when \( \text{CO}_2 \) concentrations are low in the intercellular air spaces of the leaf. Therefore, \( \text{C}_4 \) plants are able to limit the opening of their stomata and minimize water loss due to transpiration. As the \( \text{CO}_2 \) pump delivers saturating concentrations of \( \text{CO}_2 \) to the site of Rubisco, high photosynthetic rates are maintained with less Rubisco than is required in \( \text{C}_3 \) species. This is reflected in higher nitrogen use efficiency (Long, 1999).

While the basic biochemistry of the \( \text{C}_4 \) cycle is well understood, our knowledge about other genes and proteins needed for efficient \( \text{C}_4 \) 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.

\( \text{C}_4 \) photosynthesis evolved several times independently during the evolution of higher plants. It originated at least 32 times in eudicots and 16 times in monocots (Sage, 2004; Muhaiedat et al., 2007). These multiple independent origins of \( \text{C}_4 \) photosynthesis suggest that the evolution of a \( \text{C}_3 \) into a \( \text{C}_4 \) species must have been relatively easy in genetic terms (Westhoff and Gowik, 2010). Recently, the \( \text{C}_4 \) syndrome has been investigated at the systems level by comparing the transcriptome of a \( \text{C}_3 \) to a closely related \( \text{C}_3 \) species (Bräutigam et al., 2011). Approximately 600 transcripts were differentially expressed at a significant level. While many of

---

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 (westhof@uni-duesseldorf.de).

© 2011 American Society of Plant Biologists. All rights reserved.
the transcriptional changes could be placed into a C4 context, the question of which and how many of the changes are related to the C4 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 C4 evolution, we performed a comparative transcriptome analysis of leaves of closely related C3, C4, and C3-C4 intermediate species of the genus *Flaveria*. This genus is very valuable for investigating the evolution of the C4 pathway because, in addition to having closely related C3 and C4 species of the NADP-ME type, it also contains a large range of C3-C4-intermediate species differing in the degree of “C4ness” (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 C3, C4, and C3-C4 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 (Brautigam and Gowik, 2010; Brautigam et al., 2011). The leaf transcriptomes of *Flaveria bidentis* (C4) and *Flaveria pringlei* (C3) were analyzed by the 454-FLX technology, and the newer 454-TITANIUM technology was used to sequence the leaf transcriptomes of *Flaveria trinervia* (C4), *Flaveria robusta* (C3), and *Flaveria ramosissima* (C3-C4).

**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 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 to sequence the leaf transcriptomes of *Flaveria trinervia* (C$_4$), *Flaveria robusta* (C$_3$), and *Flaveria ramosissima* (C$_3$-C$_4$).
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 Brassicaceaean 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. pringlei/F. bidentis experiment as well as in the F. trinervia/F. robusta 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 C3 and the C4 plant. Two hundred transcripts were more abundant and 263 transcripts less abundant in the C4 plant F. bidentis compared with the C3 plant F. pringlei (Table 2; see Supplemental Data Set 1 online). The combined ESTs of F. trinervia (C4) and F. robusta (C3) correspond to 14,304 transcripts. A total of 410 transcripts were significantly (P < 0.01) more and 585 transcripts less abundant in the C4 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 (R² = 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 C4 and C3 Species

Transcripts of genes known to be involved in the C4 cycle, the photosynthetic electron transport and CO₂ fixation, and photorespiration showed pronounced differences between C3 and C4 (Figure 2). A high percentage of the genes contained in the functional class of potential C₄ cycle genes showed strong and significant upregulation in the C₄ plants in both the F. bidentis/F. pringlei and the F. trinervia/F. robusta comparison. Other classes with high percentages of significantly 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 C₄ species.

To complement this analysis, we searched for functional classes showing significant differential expression between the C₃ and C₄ 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 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 C₃ plant, no functional C₄ class is annotated in these tools.

Table 1. Results of the 454 Sequencing Runs

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

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 concatenation) to obtain clean reads.
C₄ Cycle Genes Are Strongly Upregulated in the C₄ Flaverias

Transcripts encoding the proteins necessary for the NADP-ME type of C₄ photosynthesis were significantly upregulated in the C₄ plants *F. trinervia* and *F. bidentis* compared with the C₃ 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 the NADP-ME and the chloroplastic MDH was 6/7 and 14/23 times higher in the C₄ 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 C₄ plants. This confirms that amino acids, in addition to malate, are also used as transport metabolites in the C₄ Flaverias.

The adenosine monophosphate kinase gene was found to be strongly upregulated in both C₄ 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 C₄ 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 C₄ 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) C₃ 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 cytoplasmatic PEP-CK, show only low or moderate expression and the C₃- to C₄-associated differences were small and not significant (Figure 4A; see Supplemental Table 2 online), indicating that the true C₄ *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 C₄ albeit much less compared with the major decarboxylation enzyme NADP-ME.

In *F. ramosissima* (C₃-C₄), Next to the NADP-ME Type C₄ Cycle, Typical NAD-ME Type C₃ Genes Also Are Upregulated

In the C₃-C₄ intermediate plant *F. ramosissima*, the transcripts of the genes related to the NADP-ME type C₄ photosynthesis showed intermediate amounts compared with the C₃ plant *F. robusta* and the C₄ plant *F. trinervia* (Figure 4A; see Supplemental Table 2 online). The amounts of all these transcripts were significantly higher in the C₄ plant *F. robusta*, implying that in *F. ramosissima*, the C₄ 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 C₄ species *F. trinervia*.

Additionally, cytoplasmatic and mitochondrial ASP-AT genes and an mNAD-MDH were upregulated significantly in *F. ramosissima* compared with the C₃ and the C₄ species. Two mNAD-MEs were upregulated in *F. ramosissima*, whereas the differences were significant only for one gene in comparison with the C₄ plant *F. trinervia* (Figure 4A; see Supplemental Table 2 online). Accordingly, the extractable NAD-ME activity in the
Table 3. Qualitative Patterns of Transcript Abundance in the Leaves of *F.* trinervia, *F.* ramosissima, *F.* robusta, *F.* bidentis, and *F.* pringlei

<table>
<thead>
<tr>
<th>Functional Class</th>
<th>Genes Detected in:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>F.</em> trinervia</td>
</tr>
<tr>
<td>All</td>
<td>58.33%</td>
</tr>
<tr>
<td>C4 candidate</td>
<td>91.30%</td>
</tr>
<tr>
<td>Photosynthesis, photosystem I</td>
<td>86.96%</td>
</tr>
<tr>
<td>Photosynthesis, photosystem II</td>
<td>87.76%</td>
</tr>
<tr>
<td>Photosynthesis, cytochrome b6/f</td>
<td>37.50%</td>
</tr>
<tr>
<td>Photosynthesis, cef</td>
<td>75.00%</td>
</tr>
<tr>
<td>Photosynthesis, ATPase</td>
<td>91.67%</td>
</tr>
<tr>
<td>Photosynthesis, other</td>
<td>90.00%</td>
</tr>
<tr>
<td>Calvin cycle</td>
<td>94.12%</td>
</tr>
<tr>
<td>Tricarboxylic acid cycle</td>
<td>92.00%</td>
</tr>
<tr>
<td>Mitochondrial electron transfer/ATPase</td>
<td>70.23%</td>
</tr>
<tr>
<td>OPPP</td>
<td>90.48%</td>
</tr>
<tr>
<td>Glycolysis</td>
<td>93.48%</td>
</tr>
<tr>
<td>Other central carbon metabolism</td>
<td>80.77%</td>
</tr>
<tr>
<td>Photosynthesis</td>
<td>100.00%</td>
</tr>
<tr>
<td>Starch</td>
<td>93.02%</td>
</tr>
<tr>
<td>Suc</td>
<td>91.43%</td>
</tr>
<tr>
<td>Lipids</td>
<td>85.48%</td>
</tr>
<tr>
<td>Nitrogen metabolism</td>
<td>90.32%</td>
</tr>
<tr>
<td>Shikimate pathway</td>
<td>84.62%</td>
</tr>
<tr>
<td>One-carbon compound metabolism</td>
<td>100.00%</td>
</tr>
<tr>
<td>Amino acid metabolism</td>
<td>93.65%</td>
</tr>
<tr>
<td>Nucleotide metabolism</td>
<td>92.23%</td>
</tr>
<tr>
<td>Pigment synthesis</td>
<td>91.67%</td>
</tr>
<tr>
<td>Cofactor synthesis</td>
<td>89.02%</td>
</tr>
<tr>
<td>Secondary metabolism</td>
<td>68.03%</td>
</tr>
<tr>
<td>Enzyme, other</td>
<td>68.33%</td>
</tr>
<tr>
<td>Enzyme, putative</td>
<td>76.07%</td>
</tr>
<tr>
<td>Sulfur assimilation</td>
<td>100.00%</td>
</tr>
<tr>
<td>Cell wall synthesis</td>
<td>60.57%</td>
</tr>
<tr>
<td>Minor sugar metabolism</td>
<td>72.00%</td>
</tr>
<tr>
<td>β1.3 Glucan metabolism</td>
<td>55.00%</td>
</tr>
<tr>
<td>Vesicle trafficking</td>
<td>82.29%</td>
</tr>
<tr>
<td>Cytoskeleton</td>
<td>76.72%</td>
</tr>
<tr>
<td>Other cellular processes</td>
<td>76.71%</td>
</tr>
<tr>
<td>Chloroplast process</td>
<td>88.97%</td>
</tr>
<tr>
<td>Mitochondrial process</td>
<td>73.44%</td>
</tr>
<tr>
<td>Peroxisomal process</td>
<td>93.75%</td>
</tr>
<tr>
<td>Protein synthesis</td>
<td>84.84%</td>
</tr>
<tr>
<td>Protein modification</td>
<td>83.33%</td>
</tr>
<tr>
<td>Protein degradation</td>
<td>60.93%</td>
</tr>
<tr>
<td>Heat shock/protein folding</td>
<td>82.11%</td>
</tr>
<tr>
<td>Transport</td>
<td>85.77%</td>
</tr>
<tr>
<td>Redox regulation</td>
<td>89.29%</td>
</tr>
<tr>
<td>Metal handling</td>
<td>60.00%</td>
</tr>
<tr>
<td>Storage protein</td>
<td>56.25%</td>
</tr>
<tr>
<td>Transcriptional regulation</td>
<td>58.75%</td>
</tr>
<tr>
<td>Epigenetic regulation</td>
<td>80.88%</td>
</tr>
<tr>
<td>Posttranslational modification (phosphate)</td>
<td>81.11%</td>
</tr>
<tr>
<td>Other regulatory processes</td>
<td>67.21%</td>
</tr>
<tr>
<td>Hormone metabolism</td>
<td>68.24%</td>
</tr>
<tr>
<td>Hormone signaling</td>
<td>80.77%</td>
</tr>
<tr>
<td>Hormone-responsive genes</td>
<td>50.38%</td>
</tr>
<tr>
<td>Defense</td>
<td>39.01%</td>
</tr>
<tr>
<td>Putative lipid transfer proteins</td>
<td>21.11%</td>
</tr>
<tr>
<td>Not classified</td>
<td>73.73%</td>
</tr>
<tr>
<td>Function unknown</td>
<td>42.06%</td>
</tr>
</tbody>
</table>

Different functional classes are represented by comparable fractions of genes detected in the leaf cDNA libraries from the different species.
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 C4 pathway. In *F. ramosissima*, the Ala level was comparable to those found in the two C4 species; however, the Asp level exceeded those of all other *Flaverias* (Figure 4C; see Supplemental Table 4 online).

### C4-Related Transport

C4 photosynthesis requires the transport of large amounts of metabolites across the chloroplast envelope, and this transport is not necessary in C3 plants (Braëutigam et al., 2008; Weber and von Caemmerer, 2010; Braëutigam and Weber, 2011). Our experiments confirmed the importance of the plastidic phosphoenolpyruvate phosphate translocator and the triosephosphate

---

**Figure 2.** The Quantitative Patterns of Transcript Accumulation in C3 and C4 *Flaveria* Are Distinct.

(A) Comparison of *F. trinervia* (Ft, C4) and *F. robusta* (Fro, C3).

(B) Comparison of *F. bidentis* (Fb, C4) and *F. pringlei* (Fp, C3). Shown are the percentages of genes with significantly higher abundance of transcripts in the C4 species (green bars), percentages of genes unchanged (gray bars, including genes not detected), and percentages of genes with significantly lower abundance of transcripts in C4 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.
phosphate translocator for the C₄ pathway since they were upregulated in the C₄ *Flaveria* (Figure 4A; see Supplemental Table 2 online), confirming earlier results from other C₄ species (Weber and von Caemmerer, 2010; Bräutigam and Weber, 2011). The *Flaveria* species belong to the group of pyruvate sodium symporter C₄ plants (Aoki et al., 1992). A gene annotated as bile acid sodium symporter was dramatically upregulated in the C₄ compared with the C₃ *Flaveria* species. BASS 2 protein is a pyruvate sodium symporter (T. Furumoto, T. Yamaguchi, Y. Ohshima-Ichie, M. Nakamura, Y. Iwata, M. Shimamura, Y. Takahashi, J. Ohnishi, S. Hata, U. Gowik, P. Westhoff, A. Bräutigam, A. Weber, and K. Izui, unpublished data). To avoid massive sodium imbalance across the chloroplast envelope, BASS 2 acts in concert with a sodium proton antiporter (NHD).


**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, C₄) and *F. robusta* (Fro, C₃), *F. bidentis* (Fb, C₄) and *F. pringlei* (Fp, C₃), or *F. ramosissima* (Fra, C₃-C₄) and *F. robusta* (Fro, C₃) 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.

In addition to BASS 2, a NHD was highly upregulated in both C4 species compared with the C3 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 C4 plants F. bidentis and F. trinervia compared with the C3 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 C4 species (see Supplemental Table 2 online).

Several other transport proteins of unknown function displayed a C4 accumulation pattern. These candidate C4 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 C₃ plants. Some transporter protein genes were significantly downregulated in the C₄ 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 C₄ plants.

**Photosynthesis Is Downregulated in C₄ but Upregulated in the C₃-C₄ Intermediate Species**

The highest percentage of genes downregulated in the C₄ species in both experiments belonged to the photorespiration class (Figure 2). Nearly all of the genes within this class were downregulated in the C₄ species, and for nearly 50% of them, the differences are statistically significant (Figure 5A). This was also true for the genes related to the reassembly of respiratory 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 C₄ plants compared with C₃ plants (Leegood, 2002; Sage, 2004), and, at least for this pathway, transcript abundance mirrors flux (Bräutigam et al., 2011; this article).

Surprisingly, the C₃-C₄ intermediate F. ramosissima did not show intermediate characteristics. By contrast, transcript abundances for most genes related to photorespiration in the C₃-C₄ intermediate species F. ramosissima were higher than in the C₃ 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 C₃ and C₄ species (Figure 5B; see Supplemental Table 4 online).

**Photosynthetic Electron Transport and Calvin-Benson Cycle Were Modified during C₄ Evolution**

Within the Calvin-Benson cycle class, most genes showed lower transcript abundance in the C₄ than in the C₃ 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 C₄ Flaverias. In the C₃-C₄ intermediate F. ramosissima, the transcript abundance of most Calvin-Benson cycle genes was C₃-like with the exception of the small subunit of Rubisco, which was significantly downregulated, mirroring earlier investigations on Rubisco protein amounts in C₃-C₄ 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 C₄ plants, whereas more of the photosystem II genes were downregulated in both C₂ 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 b₆f 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 C₃ and C₄ species, and others were at the level found in the C₃ plant. Most transcripts related to photosystem II show intermediate characteristics as many of them are downregulated compared with the C₃ plant F. robusta but not as much as much as in the C₄ plant F. trinervia.

**Chloroplast Biogenesis and Maintenance Is Altered in C₄ Species**

Several genes involved in chloroplast biogenesis and maintenance were differentially expressed between the C₃ and C₄ 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 C₄ 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 C₄ 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 C₄ species. Although chloroplast division is largely completed in mature leaves, protein turnover appeared upregulated in both C₄ species compared with the respective C₃ species.

**The C₄ Syndrome Alters Nitrogen Metabolism, Amino Acid Metabolism, and Translation**

C₄ plants need less Rubisco in their leaves than C₃ species to perform the same amount of CO₂ fixation leading to a better nitrogen use efficiency by C₄ compared with C₃ species (Black, 1973; Ku et al., 1979; Oaks, 1994; Brown, 1999; Osborne and Freckleton, 2009; Ghanem et al., 2011). Protein synthesis was altered in the C₄ 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 software (Figure 3). Downregulated transcripts representing cytosolic ribosomes were enriched in all C₄ 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 C₃ levels, whereas the transcripts related to the plastidic ribosomal proteins showed amounts that are intermediate between the C₃ species *F. robusta* and the C₄ species *F. trinervia*.

In accordance with these findings, elemental analysis showed that the C₄ Flaverias exhibit higher carbon to nitrogen ratios (7.8 to 8.6) than the C₃ 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).

Figure 5. Photorespiration is altered between *F. trinervia* (C₄), *F. ramosissima* (C₃-C₄), *F. robusta* (C₃), *F. bidentis* (C₄), and *F. pringlei* (C₃).

(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).
Consequently, the genes involved in amino acid synthesis were downregulated in the C4 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 C4

To discover additional genes that might be associated with the C4 trait, all Flaveria transcriptome data were clustered and tested for C4-related patterns. Hierarchical clustering showed that the two C4 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 C4 pattern (see Supplemental Figure 3 online). Six clusters show patterns related to C4 photosynthesis, either high in C4 versus low in C3 (three clusters) or high in C4 versus low in C4 (three clusters) (Figure 6B). The clusters vary in regard to F. ramosissima expression as exemplified for C4 transcripts (intermediate) or photorespiratory transcripts (higher in F. ramosissima) above. Taken together, the C4 clusters contained 3582 transcripts (Figure 6B; see Supplemental Data Set 2 online). A total of 1418 of these genes were in clusters with C4 upregulated genes, whereas 2164 genes were downregulated during C4 evolution.

Early Evolutionary Changes

Clusters one and two contained 1213 genes, which were upregulated in the two C4 species and C4-like or intermediate in the C3-C4 intermediate species F. ramosissima. The genes encoding the core C4 enzymes and known or putative C4 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 C4 species and in the intermediate. The number of these transcripts was thus small. We suggest that the changes in the C4 species and the intermediate were C4 changes in the narrow sense.

Late Evolutionary Changes

Cluster three contains transcripts that were more highly expressed in the C4 species compared with the C3 species but not in the C2-C4 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 phosphorylation signaling were enriched. Clusters four, five, and six contained genes that were downregulated in the C3 compared with the C2 species. In the C3-C4 intermediate species F. ramosissima, genes from the three clusters were mainly expressed on C3 level (Figure 6B). Within these clusters, genes related to major carbohydrate metabolism (including the Calvin Benson cycle and photosynthesis) 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 C4 changes in the wider sense.

Regulatory Genes

In clusters one and two, the C4 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 C4 plant F. trinervia in the C2-C4 intermediate. In the C4 plant Cleome gynandra, a different Sigma70-like factor, SIG6, was upregulated significantly compared with the C3 plant Cleome spinosa (Bräutigam et al., 2011). Thus, it might be possible that also the different abundance of plastidic sigma factors in C3 and C4 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 C3-C4 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 C4 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 C4 pathway in Flaveria. Interestingly, in Cleome, the GLK2 counterpart GLK1 was upregulated in the C4 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 C3 clusters in the wider sense (see Supplemental Data Set 3 online). Most strikingly, one can find 73 signaling receptor kinase genes, including CLAVATA1 and ERECTA in clusters four, five, and six, meaning
Figure 6. Cluster Analysis of Transcript Abundances in F. bidentis (C4), F. trinervia (C4), F. ramosissima (C3-C4), F. robusta (C3), and F. pringlei (C3).

(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 C4 species are more closely related to each other than to the other three Flaveria species.

(B) C4-related clusters. K-means analysis was used to define 26 clusters identifying different expression profiles. The six clusters with a C4-related pattern are shown. All 26 clusters can be found in Supplemental Figure 2 online.

(C) Functional category (MapMan bins) enrichment among the six C4-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.
that they were downregulated in the C₄ 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 C₃ to C₄ 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 C₄ Evolution

The cluster analysis resulted in six clusters with a C₄-related pattern. Taken together, these C₄ clusters contained 3582 transcripts. A total of 1418 of these genes were in clusters with C₄ upregulated genes, whereas 2164 genes are downregulated during C₄ evolution. These numbers are the current best estimate for transcript abundance changes related to C₄. Until a functional C₄ cycle is introduced into a C₃ plant, it will remain unknown how many of these transcript changes are necessary and sufficient to establish a C₄ cycle. Based on the multiple concurrent and parallel successful evolution of the C₄ 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 C₄ cycle. In this experiment, the evolutionary progression can be established by comparing the intermediate species with the C₄ and the C₃ species based on PageMan analysis. While all known core C₄ genes were changed early during C₄ evolution, other major changes happened after the establishment of the C₄ cycle. In case of the nitrogen metabolism, amino acid synthesis and transcriptional machinery, which were reduced in the C₄ species, this is logically consistent, since first the highly abundant transcripts of the functional classes “Calvin-Benson cycle” or “pho- torespiration” had to be reduced. These reductions on the other hand require the existence of a fully functional C₄ cycle.

The majority of C₃-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 C₄-related transcript patterns. They may be involved in the development and maintenance of C₄ leaves and are prime candidates for further analysis.

The Transcription of C₄ Cycle Genes Was Altered during C₄ Evolution in Flaveria

The C₄ Flaveria species are assigned to the NADP-ME C₄ 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 C₄ Flaveria species F. trinervia and F. bidentis exclusively use the NADP-ME C₄ pathway. In addition to the classical NADP-ME genes, we found a strong upregulation of an Ala and an Asp-AT, indicating that the C₄ 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 C₄ 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 C₄ photosynthesis. We detected the strong and significant upregulation of an adenosine
monophosphate kinase gene in both C4 species. This gene was also found to be upregulated in the C4 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 C4 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 is similar to that of other genes directly involved in the C4 cycle, supporting the assumption that DIT1 is indeed involved in the C4 photosynthesis as the OAA/malate shuttle of mesophyll chloroplasts. The upregulation of plastidic Asp-AT in the C4 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 C4 Flaverias prompted the hypothesis that DIT2 was involved in the exchange of Asp across mesophyll and bundle sheath chloroplast envelopes as part of the C4 cycle. The DIT genes were not upregulated in the C3 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 C4 photosynthesis. Next to the inorganic phosphate, which must exhibit high activities also in the bundle sheath chloroplasts, no further C4-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 C4 genes. In contrast with the chloroplastic pyruvate transporter BASS 2, this gene was not upregulated in the C3 species C. gynandra when compared with the C3 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 C4 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 C4 plant C. gynandra when compared with the C3 plant C. spinosa (Bräutigam et al., 2011), indicating a potential relevance of these genes for C4 photosynthesis. The AVP1 transcripts are significantly more abundant in F. trinervia and C. gynandra than in the respective C3 species (Bräutigam et al., 2011). Also, in F. bidentis/F. pringlei, AVP1 was upregulated, although the difference was not significant. In the C3-C4 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 CO2 Concentration Mechanisms Operate in the C3-C4 Intermediate F. ramosissima

In F. ramosissima, the transcripts of the genes related to the NADP-ME type C4 photosynthesis showed intermediate levels compared with the C3 plant F. robusta and the C4 plant F. trinervia. This implies that in F. ramosissima, the C4 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 CO2 is directly fixed into the C4 acids malate and Asp (Ku et al., 1991). Based on the transcriptional profile, F. ramosissima is intermediate with regard to a NADP-ME type C4 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 C4 species, the majority of Calvin-Benson cycle transcripts were downregulated. Unlike C4 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 C4 photosynthesis, as indicated by its intermediate C/N ratio.

We found significant upregulation of genes related to the NADP-ME type C4 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 C4 cycle is active in addition to the NADP-ME type C4 pathway in the C3-C4 C4 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 C4 species, the Asp level exceeded those of all other Flaveria species. 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 C4 cycle in F. ramosissima. This finding was surprising since all true C4 Flaveria species belong exclusively to the NADP-ME C4 plants (Drinovich et al., 1998; this article). It is not clear if this reflects plasticity in the photosynthetic metabolism of Flaveria during C4 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 C4
species. This will be clarified in the future by analyzing further C3-C4 intermediate Flaveria species.

This study also provided comprehensive molecular evidence for a photorespiratory CO₂ concentration mechanism in the C3-C4 intermediate species, which was previously hypothesized to represent a biochemical CO₂ pump (Rawsthorne et al., 1988a, 1988b) and might have been an intermediate step toward the evolution of the C₄ pathway (Bauwe et al., 1987; Sage, 2004; Bauwe, 2011).

Photorespiratory genes were expressed at a higher level than even in the C₃ species and, importantly, also the steady state levels of Gly and Ser, the transport metabolites of the photorespiratory CO₂ pump, were higher in *F. ramosissima* compared with the C₃ and C₄ *Flaveria* species. Hence, a photorespiratory CO₂ pump may still operate in *F. ramosissima*. Based on the available data, three distinct CO₂ 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 CO₂ pump moves two molecules of Gly to the bundle sheath cells where they are decarboxylated, leading to one molecule each of Ser, CO₂, 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 C₄ cycle, ammonia can be balanced by adjusting the ratios of the transport metabolites Ala/pyruvate and Asp/malate. For less advanced C₃-C₄ intermediates, which solely rely on the photorespiratory CO₂ 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 C₄ 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 C₄-like CO₂ transport pathway that could replace the photorespiratory Gly/Ser pump.

**Comparison to C₄ Photosynthesis in the Genus Cleome:**

Common Themes of C₄ Evolution

Two comparative transcriptome studies on closely related C₃ and C₄ species from the dicot genera *Flaveria* and *Cleome*

![Figure 7. Schematic of the CO₂ Concentrating and Photorespiratory Pathways in the C₃-C₄ Intermediate Species F. ramosissima.](image-url)
(Bräutigam et al., 2011) estimate the common differences between C₃ and C₄ leaf metabolism: in both species, C₄ 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 C₄ cycle, the biggest differences between C₃ and C₄ species were found for genes related to photosynthesis in both studies. The Calvin-Benson cycle and the photoreospiratory 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 C₄ 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 C₄ and C₃-C₄ intermediate Flaverias as it was shown earlier that at least in the C₄ 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 posttranscriptionally 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 C₄ species, which likely reflects the lowered demands of highly abundant Calvin-Benson cycle and photoreospiratory proteins. Also, we found several genes not obviously related to the C₄ 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 C₄-related alterations in gene expression, the analysis of further closely related C₃-C₄ pairs, especially C₄ plants of the PEP-CK type, would be highly desirable.

The comparison of the transcriptomes of closely related C₃ and C₄ species from genera like Flaveria, in this study, or Cleome (Bräutigam et al., 2011) elucidate the evolutionary trajectories of C₄ evolution and reveal how many and which genes were altered during the transitions from C₃ into C₄ plants.

We analyzed the transcriptomes of five nonmodel species belonging to the genus Flaveria exhibiting a C₃, C₄, or C₃-C₄ 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 C₄ in a C₃ 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 C₃ and C₄ 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 C₃ and C₄ plants.

The combination of genome information of a closely related C₃ C₄ 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 C₄ evolution.

METHODS

Plant Material, RNA Isolation, and cDNA Synthesis

Flaveria bidentis, Flaveria pringlei, Flavernia 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 spectrosopically 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 concatamers 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 concatenamerization 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.

\[ \text{Log}_2 \text{expression ratios were calculated after adding a pseudocount to the number of reads in each species to account for any zero counts. Differentially expressed transcripts were identified using the Poisson statistics developed by Audic and Claverie (1997) as implemented in Discovery Space (Robertson et al., 2007) on the non-normalized read counts followed by a Bonferroni correction to account for the accumulation of } \alpha \text{-type errors when conducting multiple pairwise comparisons.} \]

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 \( \mu \)g 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 C3 and C4 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 \( ^{13} \text{C} \) 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 Analysysysteme) following the manufacturer’s recommendations. The calibration for \( ^{13} \text{C} \) 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 SRR006166.

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* (C3), *F. ramosissima* (C3-C4), *F. robusta* (C3), *F. bidentis* (C4), and *F. pringlei* (C3).

Supplemental Figure 3. Cluster Analysis of Transcript Abundances in *F. bidentis* (C3), *F. trinervia* (C3), *F. ramosissima* (C3-C4), *F. robusta* (C3), and *F. pringlei* (C3).

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 C4 Cycle Genes and C4-Related Transporters.

Supplemental Table 3. Quantitation of C4 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 C4-Related Expression Pattern (Clusters One to Six).

Supplemental Data Set 3. All Regulatory Genes Exhibiting a C4-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

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

REFERENCES


Evolution of C4 Photosynthesis in the Genus Flaveria: How Many and Which Genes Does It Take to Make C4?
Udo Gowik, Andrea Bräutigam, Katrin L. Weber, Andreas P.M. Weber and Peter Westhoff

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

This information is current as of July 6, 2017