The Plant Cell, Vol. 25: 2522–2535, July 2013, www.plantcell.org

Evolution of C_4 Photosynthesis in the Genus *Flaveria*: Establishment of a Photorespiratory CO_2 Pump[W]

Stefanie Schulze,[a,b] Julia Mallmann,[a] Janet Burscheidt,[a] Maria Koczor,[a] Monika Streubel,[a] Hermann Bauwe,[c] Udo Gowik,[a,b] and Peter Westhoff [a,b,1]

[a] Heinrich-Heine-Universität, Department Biologie, 40225 Duesseldorf, Germany
[b] Cluster of Excellence on Plant Sciences "From Complex Traits towards Synthetic Modules," 40225 Duesseldorf, Germany
[c] Universität Rostock, Abteilung Pflanzenphysiologie, 18059 Rostock, Germany

ORCID ID: 0000-0002-4621-1490 (P.W.).

C_4 photosynthesis is nature's most efficient answer to the dual activity of ribulose-1,5-bisphosphate carboxylase/oxygenase and the resulting loss of CO_2 by photorespiration. Gly decarboxylase (GDC) is the key component of photorespiratory CO_2 release in plants and is active in all photosynthetic tissues of C_3 plants, but only in the bundle sheath cells of C_4 plants. The restriction of GDC to the bundle sheath is assumed to be an essential and early step in the evolution of C_4 photosynthesis, leading to a photorespiratory CO_2 concentrating mechanism. In this study, we analyzed how the P-protein of GDC (GLDP) became restricted to the bundle sheath during the transition from C_3 to C_4 photosynthesis in the genus *Flaveria*. We found that C_3 *Flaveria* species already contain a bundle sheath–expressed *GLDP* gene in addition to a ubiquitously expressed second gene, which became a pseudogene in C_4 *Flaveria* species. Analyses of C_3-C_4 intermediate *Flaveria* species revealed that the photorespiratory CO_2 pump was not established in one single step, but gradually. The knowledge gained by this study sheds light on the early steps in C_4 evolution.

INTRODUCTION

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), the key enzyme of CO_2 fixation in plants, is a bispecific enzyme. It not only operates as a carboxylase but also as an oxygenase. The product of the oxygenase reaction is the two-carbon compound 2-phosphoglycolate that has to be recycled in a process called photorespiration (Ogren and Bowes, 1971; Ogren, 1984). During photorespiration, CO_2 is released, leading to a net loss of photoassimilated CO_2. The loss of CO_2 becomes problematic especially under hot and arid conditions when stomata have to close to avoid water loss (reviewed in Sage, 2004). Under these conditions, CO_2 uptake is drastically reduced, and the relation between photosynthetic CO_2 fixation and photorespiratory CO_2 release becomes unfavorable.

The release of CO_2 is catalyzed by Gly decarboxylase (GDC) in the mitochondria of plant cells (Oliver and Raman, 1995). GDC is a multiprotein system comprising the four proteins P-, L-, T-, and H-protein (gene designations *GLDP*, *GLDL*, *GLDT*, and *GLDH*, respectively) with the P-protein being the actual decarboxylase (Oliver and Raman, 1995). GDC is not only essential in photorespiration but also necessary for C_1 metabolism that presumably takes place in all cells of a plant and provides one-carbon compounds for a number of biosynthetic pathways (Hanson and Roje, 2001). This was experimentally shown with a *GLDP* double knockout mutant of *Arabidopsis thaliana* that possesses no active GDC and cannot survive even under elevated CO_2 (i.e., nonphotorespiratory conditions) (Engel et al., 2007).

C_4 photosynthesis is one of nature's answers to cope with the oxygenase activity of Rubisco. It is essentially a CO_2 pump that concentrates CO_2 at the site of Rubisco. In the vast majority of C_4 species, the CO_2-concentrating mechanism requires the close metabolic interaction of two different cells: mesophyll and bundle sheath cells. The bundle sheath cells typically form a wreath-like layer around the vasculature and harbor Rubisco and the other Calvin-Benson cycle enzymes. Bundle sheath cells are surrounded by the mesophyll cells, which are devoid of Rubisco but contain phospho*enol*pyruvate carboxylase, an oxygen-insensitive carboxylase (Hattersley, 1984; Dengler and Nelson, 1999). The atmospheric CO_2, after conversion to bicarbonate, is initially fixed by phospho*enol*pyruvate carboxylase in the mesophyll, resulting in a four-carbon compound, malate and/or Asp, after which this photosynthetic pathway is named C_4 photosynthesis. The C_4 compound diffuses along its concentration gradient via the plasmodesmata into the bundle sheath cells where it becomes decarboxylated by NADP/NAD malic enzymes or phospho*enol*-pyruvate carboxykinase (Hatch et al., 1975). The released CO_2 is finally channeled through Rubisco into the Calvin-Benson cycle. Due to the elevated CO_2 concentration at the site of Rubisco, its oxygenase reaction is largely abolished, and photorespiration is drastically reduced in C_4 plants (Hatch, 1987; Sage, 2004). This includes lower activities of enzymes of the photorespiratory pathway, most of which are restricted to the bundle sheath cells (Li et al., 2010).

C_4 photosynthesis has independently evolved up to 66 times within the angiosperms (Sage et al., 2012). This polyphyletic origin of C_4 photosynthesis suggests that the evolution of a C_3 into a C_4 species must have been relatively easy in genetic

[1] Address correspondence to west@uni-duesseldorf.de.
The author 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) is: Peter Westhoff (west@uni-duesseldorf.de).
[W] Online version contains Web-only data.
www.plantcell.org/cgi/doi/10.1105/tpc.113.114520

terms. The genus *Flaveria* (Powell, 1978) is an attractive model in which to study the transition from C_3 to C_4 photosynthesis. The genus includes not only true C_3 and C_4 species but also a large number of C_3-C_4 intermediate species with a differing degree of "C_4-ness" (Edwards and Ku, 1987; McKown et al., 2005). The evolutionary analysis of the kinetic and regulatory characteristics of C_4 phospho*enol*pyruvate carboxylase and of the determinants for the mesophyll-specific transcription of its gene may serve as an example of how this genus can be exploited for dissecting the evolutionary trajectory from C_3 to C_4 photosynthesis (Stockhaus et al., 1997; Gowik et al., 2004; Akyildiz et al., 2007).

The compartmentation of GDC in the bundle sheath cells and, hence, its virtual absence in mesophyll cells is assumed to constitute a very early and essential step in the evolution toward C_4 photosynthesis (Sage, 2004; Bauwe, 2011; Sage et al., 2012). Restriction of GDC to the bundle sheath cells results in the photorespiratory CO_2 only being released in the bundle sheath. This results in the establishment of a photorespiratory CO_2 pump creating a CO_2-enriched environment for the Rubisco of the bundle sheath, but not of the mesophyll cells. Since Gly, a two-carbon compound, serves as a transport metabolite, this photorespiratory CO_2 concentrating mechanism is also termed C_2 photosynthesis. Immunolocalization experiments with the C_3-C_4 intermediate *Moricandia arvensis* indicated that the compartmentation of GDC activity in the bundle sheath cells was caused by the cell-specific restriction of only one of its components, the P-protein (Rawsthorne et al., 1988; Morgan et al., 1993).

This study seeks to answer the question of how the photorespiratory CO_2 pump was established during the evolution of C_4 photosynthesis in the genus *Flaveria*. From previous work, we knew that C_4 *Flaveria* species contain only one functional *GLDP* gene (named *GLDPA*) that appears to be active only in the bundle sheath (Engelmann et al., 2008). We wanted to know how this expression specificity evolved, bearing in mind that C_3 *Flaveria* species contain several *GLDP* genes (Bauwe et al., 1995), which should be active in all photosynthetic tissues. In recent work, we showed that the regulation of *GLDP* expression is complex, involving transcription from two promoters oriented in tandem and most probably also posttranscriptional control via differential RNA stability (Wiludda et al., 2012). The data presented here demonstrate that the photorespiratory CO_2 pump in the genus *Flaveria* was established step by step and that it involved pseudogenization of one, and in addition, a relaxation of the bundle sheath specificity of another already existing *GLDP* gene.

RESULTS

The *GLDP* Gene Family of the Genus *Flaveria*

To get an overview of the structure of the *GLDP* gene family in C_3, C_4, and C_3-C_4 intermediate *Flaveria* species, we conducted a phylogenetic analysis. We used published sequences of cDNA or genomic clones (Kopriva and Bauwe, 1994; Bauwe et al., 1995; Bauwe and Kopriva, 1995; Chu, 1996) and de novo assembled sequences derived from RNaseq of different *Flaveria* species using 454 (Gowik et al., 2011) or Illumina (J. Mallmann, U. Gowik, and P. Westhoff, unpublished data) sequencing, respectively. The sequences were aligned using ClustalX (see Supplemental Data Set 1 online), and gene trees were constructed using the maximum likelihood method as described in Methods. Figure 1A shows that the *GLDP* genes of the various *Flaveria* species group into three clusters, A, B, and C. All nine analyzed *Flaveria* species contained one gene of each group with the exception of *Flaveria pringlei* in which two group A and two group B genes were found (Figure 1A). It is known that *F. pringlei* is a tetraploid (Cameron et al., 1989), probably arisen by allopolyploidization with the C_3-C_4 intermediate species *Flaveria angustifolia* (Kopriva et al., 1996; McKown et al., 2005). Additionally, *GLDP* cDNA sequences from *Flaveria* were compared with those from eudicot species with fully sequenced genomes. It turned out that the *Flaveria GLDP*s were more closely related to each other than to any other *GLDP* sequence (see Supplemental Figure 1 and Supplemental Data Set 2 online).

The phylogenetic analysis revealed that group A and B *GLDP* genes are more closely related to each other than to group C genes. Group A *GLDP* genes contain *GLDPA* of *Flaveria trinervia* (C_4; formerly *gdcsPA*; Cossu and Bauwe, 1998; accession number Z99767) and *GLDPA1* and *GLDPA2* of *F. pringlei* (C_3; formerly *gdcsPB* and *gdcsPA*; Kopriva and Bauwe, 1994; Bauwe et al., 1995; Bauwe and Kopriva, 1995; accession numbers Z36879 and Z54239), all of which have been characterized by sequencing of genomic clones. Group B *GLDP* genes are exemplified by the *GLDPB** pseudogene of *F. trinervia* (C_4; formerly *gdcsPB*; Cossu and Bauwe, 1998; accession number Z99768) and the *GLDPB1* and *GLDPB2* genes of *F. pringlei* (C_3; formerly *gdcsPE* and *gdcsPD*; Chu, 1996; accession numbers KC545951 and KC545950, respectively). The *GLDP* group C contains the *GLDPC* gene of *F. pringlei* (C_3; formerly *gdcsPC*; Chu, 1996; accession number KC545949).

To determine the spatial expression patterns of the different *GLDP* genes, we used data sets from Illumina RNaseq experiments of roots, stems, and leaves of *Flaveria bidentis* (C_4) and *Flaveria robusta* (C_3). Figure 1B shows that transcripts of group C *GLDP* genes could be detected predominantly in roots and stems in both species, while the *GLDP* group A and group B genes accumulated preferentially in leaves. Accordingly, group A and group B *GLDP* genes are mainly relevant for photorespiration, whereas group C *GLDP* genes most likely are not involved in photorespiration but in the maintenance of basal C_1 metabolism (Hanson and Roje, 2001) in stems and roots.

The Group A *GLDP* Genes of the C_3 *Flaveria* Species *F. pringlei* and *F. robusta* Are Expressed Specifically in Bundle Sheath Cells

RNA in situ hybridization experiments had shown that *GLDP* transcripts accumulate in the leaves of *F. trinervia* and *F. bidentis* (both C_4) only in bundle sheath cells (Engelmann et al., 2008). Promoter-reporter gene studies were in line with these observations and conclusions. When 1571 bp of 5′ flanking sequences of the *F. trinervia GLDPA* gene (including the 5′ untranslated region upstream of the AUG codon) were fused to the β-glucuronidase (GUS) reporter gene and transformed into *F. bidentis* (C_4), reporter gene activity was only observed in the

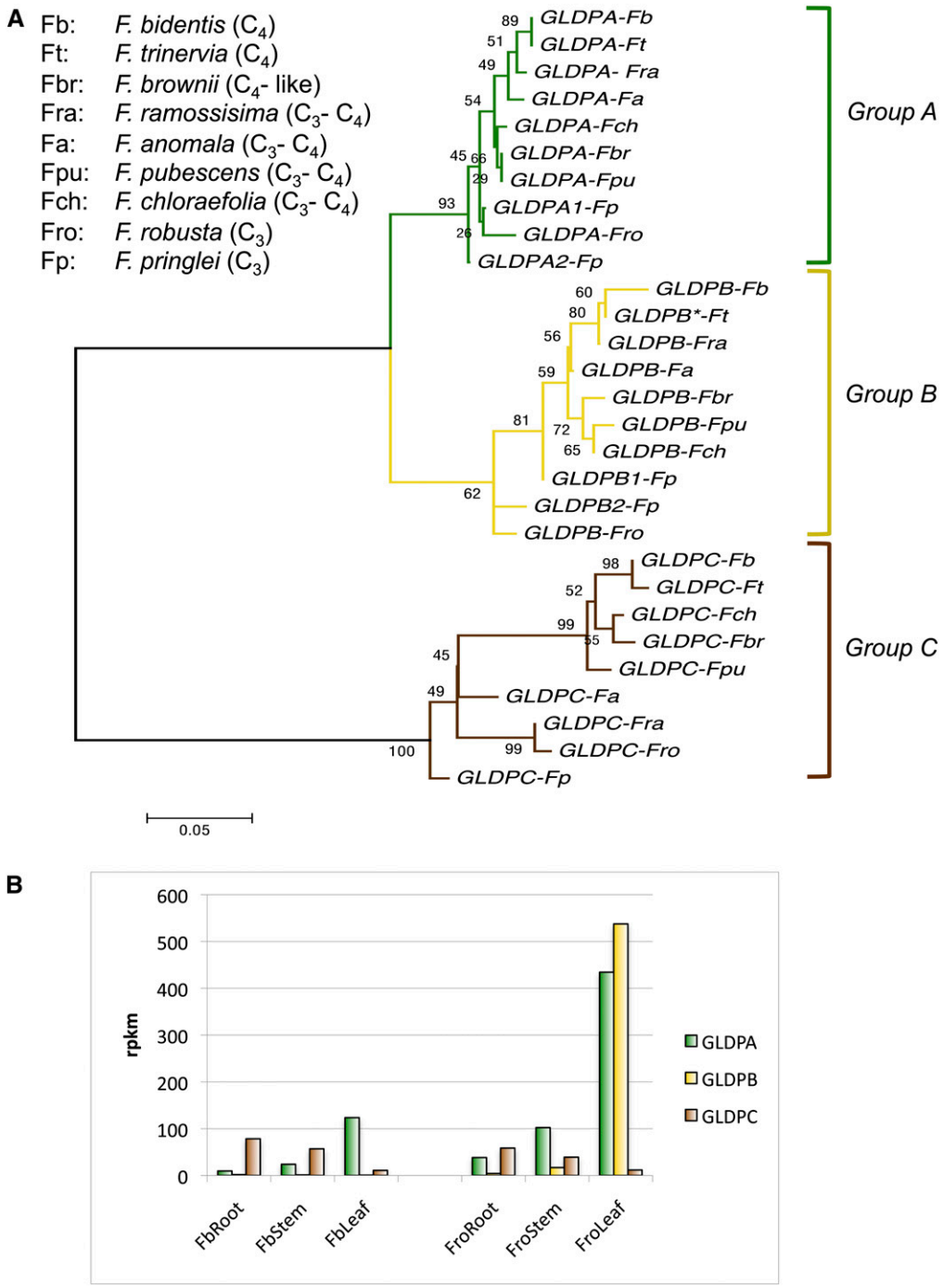

Figure 1. Molecular Phylogenetic Analysis of *GLDP* Genes of the Genus *Flaveria* and *GLDP* Transcript Abundance in Organs.

(A) Maximum likelihood tree of *GLDP* sequences in *Flaveria*. The tree was constructed with MEGA5 (Tamura et al., 2011) using the Tamura 3 parameter model. The tree is drawn to scale, with branch length measured in the number of substitutes per site. The tree is based on 203 nucleotide positions, starting at the ATG, which were aligned using ClustalX 2.0.8 (Higgins and Sharp, 1988, 1989; Thompson et al., 1997; Larkin et al., 2007) (alignment is shown in Supplemental Data Set 1 online). Bootstrap values (1000 replicates) are shown next to the branches (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50% of bootstrap replicates are collapsed.

(B) Abundance of *GLDPA*, *GLDPB*, and *GLDPC* transcripts in stems, roots, and leaves of *F. bidentis* (Fb; C_4) and *F. robusta* (Fro; C_3) as measured by mapping RNaseq Illumina reads on the respective cDNAs and expressed in reads per kilobase per million (rpkm).

bundle sheath cells and to a small degree in the vascular bundle, but not in the mesophyll tissue (Engelmann et al., 2008). To gain insight into how bundle sheath–specific *GLDP* expression evolved in the genus *Flaveria*, we analyzed the expression specificity of the *GLDPA* genes of *F. pringlei* (C_3) and *F. robusta* (C_3) in both the C_4 plant *F. bidentis* and the C_3 species *Arabidopsis*.

The 5′ flanking sequences of the *GLDPA* genes of *F. pringlei* (*GLDPA1*, 2217 bp; *GLDPA2*, 2040 bp; Bauwe et al., 1995) and *F. robusta* (1154 bp; accession number KC545947) were fused to the GUS reporter gene and transformed into *F. bidentis* plants. Figure 2A demonstrates that the 5′ flanking sequences of the *GLDPA1* gene of *F. pringlei* and the *GLDPA* gene of *F. robusta* were able to drive expression of the GUS reporter gene predominantly in the bundle sheath cells and to a lesser extent in the vasculature (compared with Engelmann et al., 2008; Wiludda et al., 2012). No GUS activity could be detected in the mesophyll cells (Figure 2A, top panel). The promoter strengths of the two *GLDPA* 5′ flanking sequences were comparable (Figure 2B) and were in the same range as that of the *GLDPA* promoter of *F. trinervia* (compared with Engelmann et al., 2008). The 5′ flanking sequence of *GLDPA2* of *F. pringlei* exhibited a similar expression behavior in *F. bidentis* as the other two *GLDPA* 5′ flanking sequences (see Supplemental Figure 2A online, top panel).

Transformation systems for C_3 *Flaveria* species are not available. Since the Brassicacean C_3 species *Arabidopsis* faithfully recapitulates the expression profile of the *GLDPA* 5′ flanking region of the C_4 species *F. trinervia* (Engelmann et al., 2008), this species was also used for the analysis of the 5′ flanking sequences of the *GLDPA* genes of *F. pringlei* and *F. robusta.* Figure 2A shows that the spatial expression pattern of the *GLDPA1* 5′ flanking region of the C_3 *Flaveria* species *F. pringlei* in transgenic *Arabidopsis* resembles that observed for the C_4 plant *F. bidentis*. Promoter activity could not be detected in the mesophyll cells but only in the bundle sheath and the vasculature. In addition, the promoter strength in *Arabidopsis* was also comparable to that in *F. bidentis* (Figure 2B). Thus, the expression profile of the *GLDPA1* promoter of the C_3 species *F. pringlei* in *Arabidopsis* is indistinguishable from that of the *GLDPA* promoter of *F. trinervia* in this distantly related C_3 species (compared with Engelmann et al., 2008). The 5′ flanking regions of the *GLDPA* gene of *F. robusta* and of the *GLDPA2* gene of *F. pringlei* also function essentially as bundle sheath/vasculature-specific promoters in *Arabidopsis*; however, they additionally exhibit a faint activity in the mesophyll tissue (Figure 2A; see Supplemental Figure 2A online).

Taken together, the expression specificities and quantities encoded by the 5′ flanking sequences of the group A *GLDP* genes of the two C_3 *Flaveria* species are almost indistinguishable from that of the 5′ flanking region of the *GLDPA* gene of the C_4 species *F. trinervia*. This suggests that the last common ancestor, leading to extant C_3 and C_4 *Flaveria* species, already had a bundle sheath specific *GLDP* gene.

The *GLDPB2* Gene of the C_3 Species *F. pringlei* Is Active in All Photosynthetic Tissues of Both *F. bidentis* (C_4) and *Arabidopsis* (C_3)

The group B gene *GLDPB** of *F. trinervia* (C_4) is known to be a pseudogene due to an insertion into the first exon leading to an interruption of the *GLDPB* reading frame (Cossu and Bauwe, 1998). Moreover, the 5′ flanking region (1981 bp) of *GLDPB** did not show any promoter activity when analyzed in *Arabidopsis* (see Supplemental Figure 3 online). To assess the expression specificity of the 5′ flanking sequences of group B *GLDP* genes of C_3 *Flaveria* species, 2733 bp of the 5′ flanking region of the *GLDPB2* gene of *F. pringlei* were fused to the GUS reporter gene and analyzed in transgenic *F. bidentis* (C_4) and *Arabidopsis* (C_3). Figure 3 shows that this 5′ flanking region drives the expression of the reporter gene in all photosynthetic leaf tissues in both species.

It follows that C_3 *Flaveria* species contain at least two *GLDP* genes, one each from groups A and B, which differ in their

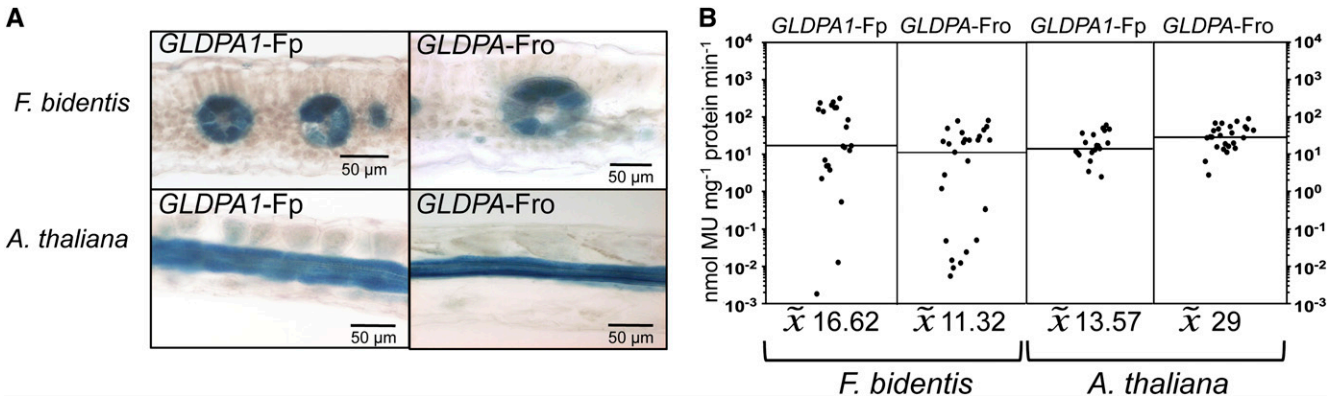

Figure 2. Functional Analysis of the 5′ Flanking Sequences of the *GLDPA* Genes of the C_3 Species *F. pringlei* and *F. robusta*.

The 5′ flanking sequences of *GLDPA1*-Fp (2217 bp) and *GLDPA*-Fro (1154 bp) were fused to the GUS reporter gene and analyzed in transgenic *F. bidentis* (C_4) and transgenic *Arabidopsis* (C_3).

(A) Histochemical localization of GUS activity in leaf sections of transgenic plants.

(B) GUS activities in leaves of transgenic plants. Each dot indicates an independent transgenic line. The black line represents the median. MU, 4-methylumbelliferone.

expression patterns in leaves. While group A *GLDP* genes are specifically/preferentially expressed in the bundle sheath, group B *GLDP* genes are active in all photosynthetic tissues. During evolution of C_4, the ubiquitously expressed group B *GLDP* gene was converted into a pseudogene; hence, *GLDP* expression in leaves became bundle sheath specific.

The 5′ Flanking Sequences of the Two *GLDP* Genes of *Arabidopsis* Do Not Direct Any Tissue Specificity in Leaves

To investigate whether a bundle sheath–specific *GLDP* gene might be a common feature of C_3 plants, we analyzed the promoters of the two *GLDP* genes of *Arabidopsis*. We fused the 1852- and 1451-bp 5′ flanking sequences of both genes, *GLDP1* (*AT4G33010*) and *GLDP2* (*AT2G26080*), to the GUS reporter gene and transformed these constructs into *Arabidopsis*. Figure 4 demonstrates that both 5′ flanking regions drive GUS expression in all photosynthetic leaf tissues, supporting the genetic findings that the two *GLDP* genes act redundantly (Engel et al., 2007). A bundle sheath–specific *GLDP* gene is therefore not a common feature of dicotyledonous C_3 plants.

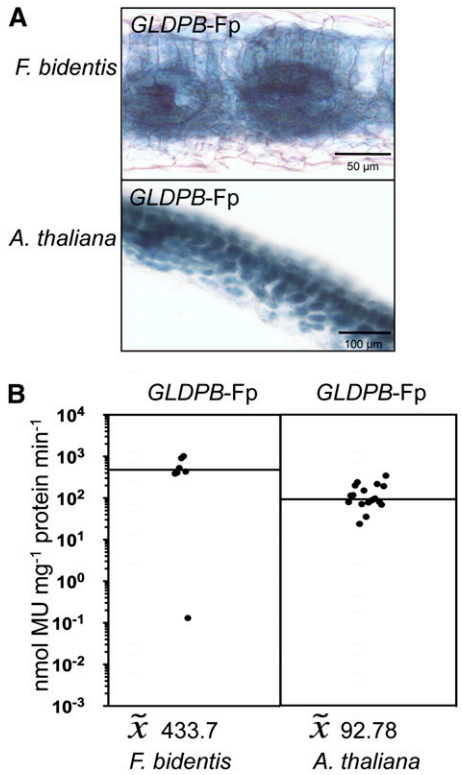

Figure 3. Functional Analysis of the 5′ Flanking Sequence of a Group B *GLDP* Gene of the C_3 Species *F. pringlei*.

The 5′ flanking sequence of *GLDPB2*-Fp (2733 bp) as a representative of group B *GLDP* genes of C_3 *Flaveria* was fused to the GUS reporter gene and analyzed in transgenic *F. bidentis* (C_4) and transgenic *Arabidopsis* (C_3).

(A) Histochemical localization of GUS activity in leaf sections of transgenic plants.

(B) GUS activities in leaves of transgenic plants. Each dot indicates an independent transgenic line. The black line represents the median. MU, 4-methylumbelliferone.

Expression of Group B *GLDP* Genes in Different *Flaveria* Species Is Negatively Correlated to the Degree of C_4-Ness, While Group A *GLDP* Genes Show a Maximum of Expression in C_3-C_4 Intermediate Species

The genus *Flaveria* offers the unique opportunity to study the steps taken during evolution from C_3 to C_4 photosynthesis, particularly because of its large number of C_3-C_4 intermediate species. According to McKown et al. (2005), the genus contains three main phylogenetic groups. The first diverging group includes the three C_3 *Flaveria* species (*F. pringlei*, *Flaveria cronquistii* and *F. robusta*) and the C_3-C_4 intermediate *Flaveria sonorensis*. Clade B contains seven C_3-C_4 intermediate species and the C_4-like species *Flaveria brownii*. All true C_4 *Flaveria* species belong to clade A, which also contains several C_4-like species (for instance, *Flaveria palmeri*) and the C_3-C_4 intermediate *Flaveria ramosissima*. It is thought that C_3-C_4 intermediate (C_2) photosynthesis evolved twice in *Flaveria*, once in the predecessor of the extant basal species *F. sonorensis*, and once in the line leading to the last common ancestor of clade A and B species (McKown et al., 2005).

To obtain insight into changes in the expression of group A and B *GLDP* genes, we used RNaseq to compare their RNA amounts in the leaves of different *Flaveria* species, ranging from C_3 (*F. pringlei* and *F. robusta* [basal *Flaveria* species]), via C_3-C_4 intermediate species with varying degrees of C_4-ness (*F. chloraefolia*, *F. pubescens*, *F. anomala* [all belonging to clade B], and *F. ramossisima* [clade A]), and the C_4-like species *F. brownii* (clade B), to the fully fledged C_4 species *F. bidentis* and *F. trinervia* (both belonging to clade A).

Figure 5 illustrates that the amounts of group B *GLDP* transcripts are lower in C_3-C_4 intermediates than in the C_3 reference. Even lower levels of group B *GLDP* transcripts were found in the C_4-like species *F. brownii*, and no group B *GLDP* transcripts were detected in the two C_4 *Flaveria* species. By contrast, group A *GLDP* transcripts increased continuously from C_3, to C_3-C_4 intermediates, to the C_4-like species, but dropped to lower levels in the true C_4 species, as expected. These findings indicate a progressive increase of C_2 cycle activity in both *Flaveria* clades until the establishment of a fully functional C_4 pathway in clade A species.

The Tandem Promoter Structure of the *GLDPA* Gene of *F. trinervia* (C_4) Is Evolutionary Conserved in *GLDPA*-Type Genes

The transcriptional regulation of the *GLDPA* gene of *F. trinervia* (C_4) appears to be rather complex, since its 5′ flanking sequence was shown to contain two transcriptional start sites, each of which is preceded by a promoter (Wiludda et al., 2012). The proximal promoter, defined by region 7 (P_{R7}; Figure 6) of the 5′ flanking sequence, is responsible for the expression in the bundle sheath and the vasculature, while the distal promoter, defined by region 2 (P_{R2}; Figure 6), is active in all green leaf

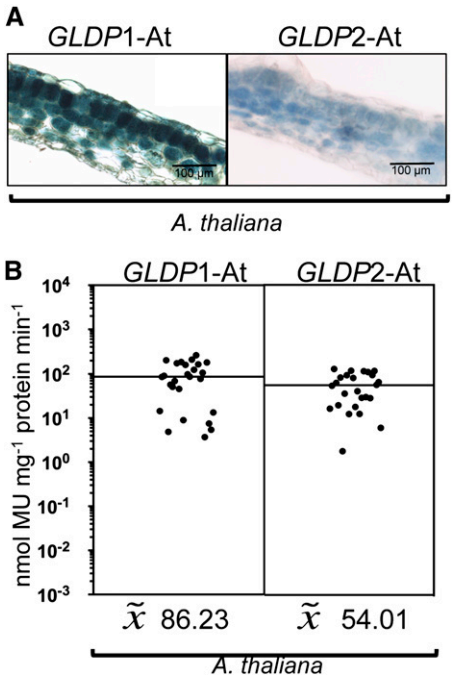

Figure 4. Functional Analysis of the 5′ Flanking Sequences of *GLDP1* and *GLDP2* Genes of *Arabidopsis*.

The 5′ flanking sequences of *GLDP1* (1852 bp) and *GLDP2* (1451 bp) of *Arabidopsis* (C_3) were fused to the GUS reporter gene and analyzed in transgenic *Arabidopsis* (C_3).

(A) Histochemical localization of GUS activity in leaf sections of transgenic plants.

(B) GUS activities in leaves of transgenic plants. Each dot indicates an independent transgenic line. The black line represents the median. MU, 4-methylumbelliferone.

tissues (Wiludda et al., 2012). The most reasonable explanation for this dual promoter structure is the need for small amounts of *GLDP* in the mesophyll cells of C_4 plants for maintaining C_1 metabolism, after the complete switch off of the *GLDPB* genes (Wiludda et al., 2012). We wanted to know whether this promoter organization is specific for C_4 *Flaveria* species or whether it is an ancient feature of *GLDPA* genes and occurred already in C_3 *Flaveria* species.

The comparison of the 5′ flanking sequences of the *GLDPA1* and *GLDPA* genes from the C_3 species *F. pringlei* and *F. robusta* with those from the C_4 species *F. trinervia* and *F. bidentis* (accession number KC545946; Figure 6) revealed a high degree of sequence conservation in regions 2 and 7, which define the distal and proximal promoters, respectively. High levels of sequence similarity were also observed for regions 1 and 3, while the regions in between (i.e., regions 4 to 6) are much less conserved and may even be drastically shortened as in *F. robusta* (Figure 6).

To test experimentally whether regions 2 and 7 of the 5′ flanking regions of the *GLDPA1* and *GLDPA* genes of *F. pringlei* and *F. robusta*, respectively, function as promoters, they were fused to the GUS reporter gene, and the constructs were transformed into *F. bidentis* (*GLDPA1* of *F. pringlei*) and *Arabidopsis* (*GLDPA1* of *F. pringlei* and *GLDPA* of *F. robusta*).

Figure 7 illustrates that region 7 of the *GLDPA1* 5′ flanking region of *F. pringlei* directs bundle sheath expression in *F. bidentis*, while region 2 shows promoter activity in all green leaf tissues. Similar expression profiles were observed when the two regions from the *GLDPA1* gene of *F. pringlei* or from the *GLDPA* gene of *F. robusta* were analyzed in transgenic *Arabidopsis* (Figure 7; see Supplemental Figure 4 online). It follows that the *GLDPA1* and *GLDPA* genes of *F. pringlei* and *F. robusta*, respectively, possess the same dual promoter structure as found in the 5′ flanking sequence of the *GLDPA* gene of the C_4 species *F. trinervia*.

The Amounts of Transcripts Derived from the Distal Subpromoter of the *GLDPA* Gene of C_3 *Flaveria* Species Are Negligible

When assayed by promoter-reporter gene fusions, the distal *GLDPA* subpromoter of *F. trinervia* (defined by region 2) turned out to be much stronger than the proximal subpromoter (defined by region 7; Wiludda et al., 2012). By contrast, when the two promoters are in their natural configuration as in the 5′ flanking region of the *F. trinervia GLDPA* gene, their RNA output was opposite. Only traces of transcripts originating from the distal transcriptional start site were detectable, while RNAs starting at the proximal site made up the vast majority of all *GLDPA* transcripts (Wiludda et al., 2012). To test if the same was true for the *GLDPA* promoters of C_3 and C_3-C_4 intermediate *Flaveria* species, we used the RNaseq data available for C_3, C_3-C_4, and C_4 *Flaveria* species and mapped the reads against the genomic sequence of *GLDPA* from *F. trinervia*.

We found that the vast majority of all *GLDPA* transcripts of *F. trinervia* (C_4), but also of *F. bidentis* (C_4), originated from the proximal transcriptional start site (TSS_{R7}) and that only small but clearly detectable amounts arose from the distal site (TSS_{R2}; Figures 8A and 8B). By contrast, transcripts derived from the distal transcriptional start site were drastically reduced (*F. pringlei*) or even absent (*F. robusta*) in the *GLDPA* transcript populations of the C_3 *Flaveria* species (Figures 8C and 8D). In C_3-C_4 intermediate species, the amounts of transcripts derived from the distal transcription start site were in between those of the C_4 and C_3 *Flaveria* species (see Supplemental Figures 5A to 5E online). Taken together, the data indicate that RNA output from the distal promoter of the *GLDPA* 5′ flanking region is negligible in the C_3 species but that it rose to small but clearly detectable amounts during C_4 evolution. Moreover, it has to be inferred that the promoter structures of the *GLDPA* 5′ flanking regions were functionally conserved during evolution within clades A and B of the genus *Flaveria*.

Splicing of Transcripts Derived from the Distal Transcriptional Start Site Changed during C_4 Evolution

Transcripts originating from the distal transcriptional start site of the *GLDPA* gene of *F. trinervia* contain a large intron of ~1000 nucleotides that has to be spliced in order to generate functional *GLDPA* mRNAs (Wiludda et al., 2012). These spliced mRNAs

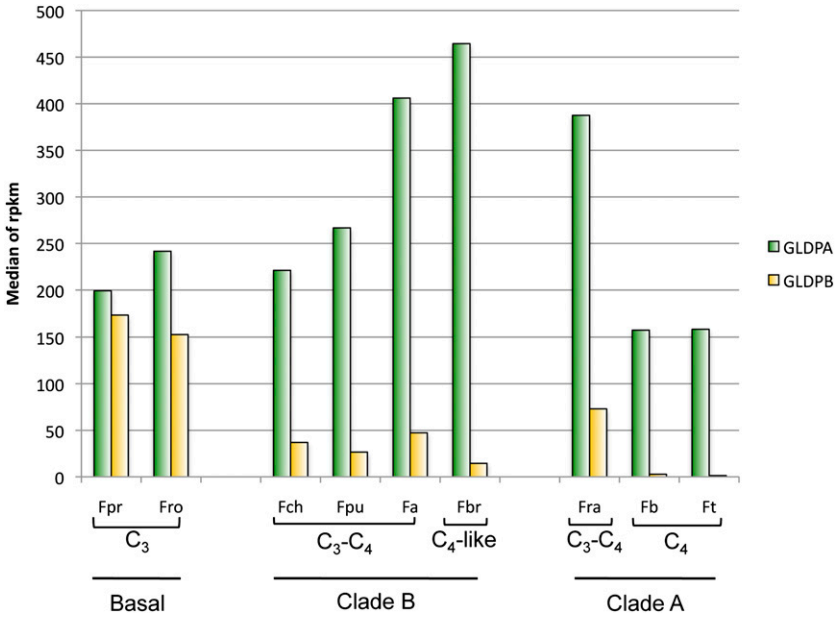

Figure 5. Transcript Abundance of Group A and Group B *GLDP* Genes in C_3, C_3-C_4 Intermediate, and C_4 *Flaveria* Species.

The transcript abundance was calculated as the median of four Illumina RNaseq experiments and is expressed in reads per kilobase per million (rpkm). Transcript abundances of group A *GLDP* genes are displayed in green, and transcript abundances of group B are displayed in yellow. Fp, *F. pringlei*; Fro, *F. robusta*; Fch, *F. chloraefolia*; Fpu, *F. pubescens*; Fa, *F. anomala*; Fra, *F. ramossisima*; Fbr, *F. brownii*; Fb, *F. bidentis*; Ft, *F. trinervia.*

can be detected in *F. trinervia*, although only in small amounts. The large *GLDPA* 5′ intron is spliced out much less efficiently than the gene-internal introns (Wiludda et al., 2012). Unspliced *GLDPA* transcripts of *F. trinervia* appear to be unstable, possibly due to the presence of many open reading frames within the 5′ intron, which could be involved in activating the nonsense-mediated mRNA decay pathway (Wiludda et al., 2012). Since RNA output from the distal promoter was barely detectable in the C_3 species (Figures 8C and 8D) but detectable in small amounts in the C_3-C_4 intermediates (see Supplemental Figures 5A to 5E

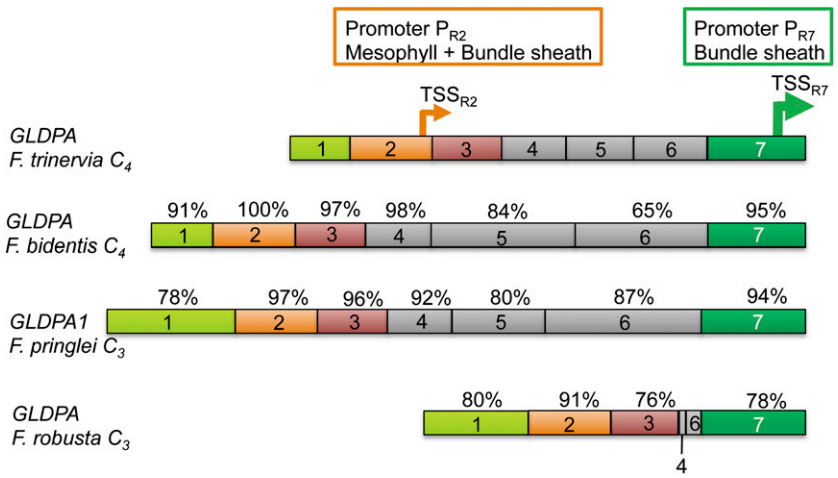

Figure 6. Schematic Comparison of the 5′ Flanking Sequences of Group A *GLDP* Genes from C_4 and C_3 *Flaveria* Species.

The 5′ flanking sequence of the *GLDPA* gene of *F. trinervia* was divided into seven functionally characterized regions (Engelmann et al., 2008). Regions 2 (orange) and 7 (dark green) contain the two subpromoters P_{R2} and P_{R7}. Region 1 (light green) enhances the activities of both subpromoters. Region 3 (red) is required to suppress the mesophyll activity of the subpromoter P_{R2}, but only in *Arabidopsis*. Regions 4 to 6 (gray) are not required for promoter activity. The 5′ flanking sequences of the *GLDPA* genes from *F. bidentis* (C_4), *F. pringlei* (C_3), and *F. robusta* (C_3) are color-labeled according to their homologous regions in the *GLDPA* 5′ flanking region of *F. trinervia*. Similarities are given as percentage of identical nucleotide positions relative to the corresponding regions of *GLDPA*-Ft.

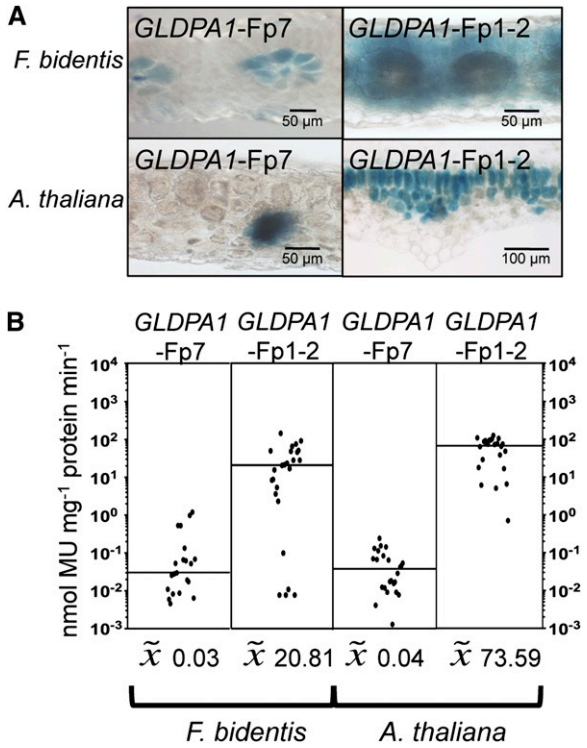

Figure 7. Functional Analysis of the Subpromoters P_{R7} and P_{R2} of the *GLDPA1* Gene of *F. pringlei*.

Regions 1 and 2 (*GLDPA*-Fp1-2; 641 bp) and 7 (*GLDPA*-Fp7; 318 bp) of the 5′ flanking region of the *GLDPA* gene of *F. pringlei* (C_3) were fused to the GUS reporter gene and analyzed in transgenic *F. bidentis* (C_4) and transgenic *Arabidopsis* (C_3).

(A) Histochemical localization of GUS activity in leaf sections of transgenic plants.

(B) GUS activities in leaves of transgenic plants. Each dot indicates an independent transgenic line. The black line represents the median. MU, 4-methylumbelliferone.

online), we wanted to know when during C_4 evolution the 5′ intron and its splicing were established.

Inspection of the RNaseq data did not reveal any correctly spliced *GLDPA* RNAs derived from the distal transcriptional start site in the two C_3 species or the C_3-C_4 intermediates. By contrast, in the C_4-like species *F. brownii*, small amounts of spliced *GLDPA* RNA accumulate (see Supplemental Figure 6 online). As expected, splicing of the 5′ intron occurs also in the C_4 species *F. bidentis* (see Supplemental Figure 6 online). A functional spliced 5′ intron was therefore not only observed in the C_4 *Flaveria* species *F. bidentis* and *F. trinervia*, but also in the C_4-like species *F. brownii*. This indicates that its presence is a typical feature of C_4 and C_4-like *Flaveria* species.

Interestingly, the splice acceptor sites differ in the two C_4 species (see Supplemental Figure 6 online). While the splice acceptor site of the *GLDPA* gene of *F. trinervia* is located ~16 nucleotides downstream of the first ATG codon of the *GLDPA* open reading frame, the acceptor sites found in *F. bidentis* and also in *F. brownii* are positioned ~15 nucleotides upstream (see Supplemental Figure 6 online). We conclude that the splicing of the 5′ intron was established at least two times independently during C_4 speciation in clade A of *Flaveria*.

DISCUSSION

The photorespiratory CO_2 pump is considered to be an essential and early step in the evolutionary trajectory toward C_4 photosynthesis (Sage, 2004; Bauwe, 2011; Sage et al., 2012). The establishment of this pump required that the Gly-decarboxylating step of the photorespiratory pathway, which is performed by GDC, became restricted to the bundle sheath cells. Earlier studies with C_3-C_4 intermediate species indicated that modifications to the gene encoding the P-protein, the actual GDC, were responsible for this reallocation of GDC activity from the mesophyll to the bundle sheath and that this event occurred rather early during C_4 evolution (Monson et al., 1984; Morgan et al., 1993; Sage, 2004; Bauwe, 2011; Sage et al., 2012). Two questions immediately arose from this evolutionary scenario: By which gene regulatory mechanism did GLDP become restricted to the bundle sheath cells, and how was this compartmentation achieved in time? Work presented here, using the genus *Flaveria* as an evolutionary model, provides conclusive answers to both questions.

C_3 *Flaveria* Species Contain a Bundle Sheath Cell–Specific *GLDP* Gene

Two scenarios can be imagined how a bundle sheath–specific expression of *GLDP* evolved during the transition from C_3 to C_4. First, a ubiquitously expressed *GLDP* gene changed its expression behavior to become bundle sheath specific. Alternatively, a *GLDP* gene with the requested bundle sheath specificity of expression was already present in C_3 species and an additional, ubiquitously expressed *GLDP* gene became inactivated during C_4 evolution. The first scenario is best exemplified by the evolution of mesophyll expression specificity in the *ppcA* phospho*enol*pyruvate carboxylase gene of the genus *Flaveria* (Stockhaus et al., 1997; Gowik et al., 2004; Akyildiz et al., 2007).

By contrast, the evolution of bundle sheath specificity of *GLDP* expression in *Flaveria* followed the second scenario. The *GLDPA* genes of the C_3 *Flaveria* species *F. pringlei* and *F. robusta* already contained promoter sequences driving bundle sheath–specific gene expression in both C_4 and C_3 plants (Figure 2; see Supplemental Figure 4 online). The other leaf-expressed *GLDP* genes (i.e., those of group B) were expressed in all photosynthetically active tissues of C_3 *Flaveria* species, but turned into pseudogenes in the C_4 species (Cossu and Bauwe, 1998; Figure 5; see Supplemental Figure 3 online).

The occurrence of bundle sheath–specific *GLDP* genes is not a universal feature of C_3 plants. The C_3 species *Arabidopsis*, for instance, contains two *GLDP* genes both of which are expressed similarly in all leaf chlorenchyma cells as concluded from their promoter activities (Figure 4). The presence of a *GLDP* gene with bundle sheath specificity of expression in C_3 *Flaveria* species should therefore be viewed as part of a preconditioning syndrome that distinguishes C_3 taxa that evolved C_4

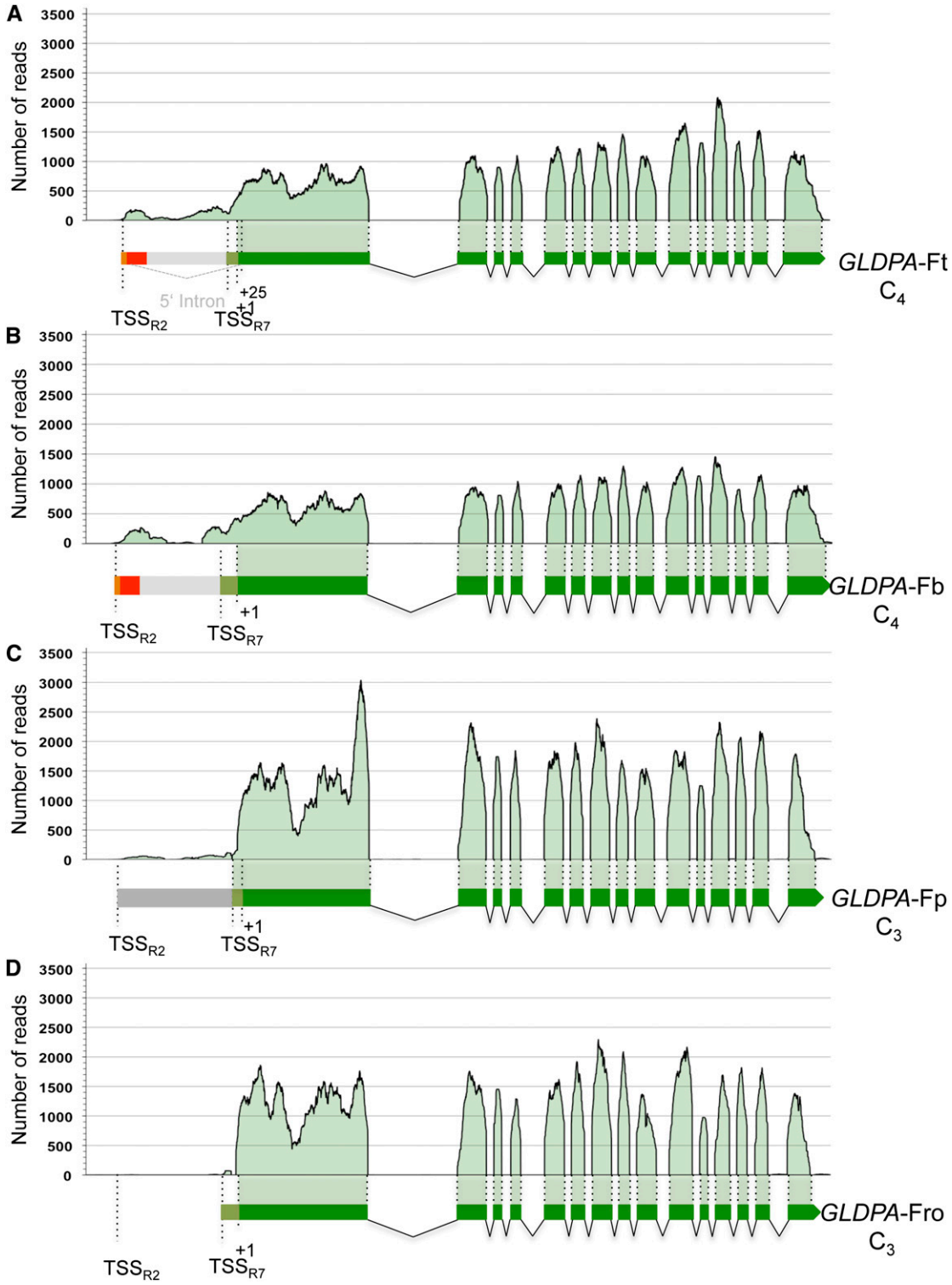

Figure 8. Transcript Coverage of the *GLDPA* Genes of *F. trinervia*, *F. bidentis*, *F. pringlei*, and *F. robusta* Leaf RNaseq Experiments.

Illumina reads obtained from sequencing the leaf transcriptomes of *F. trinervia* (C_4) **(A)**, *F. bidentis* (C_4) **(B)**, *F. pringlei* (C_3) **(C)**, and *F. robusta* (C_3) **(D)** were mapped on the sequence of the *GLDPA* gene of *F. trinervia* including its 5′ flanking sequence. Region 2 (orange); region 3 (red); regions 4, 5, and 6 (gray); and region 7 (light green) are indicated in the 5′ flanking region, and exon parts of the coding region are indicated in dark green. The numbers of reads covering each position of the gene sequence were counted.

photosynthesis from others that did not (Sage, 2004; Sage et al., 2011, 2012). Such a preconditioning phase has been proposed as an inherent, most likely necessary step in C_4 evolution. One can envision that the presence of a bundle sheath–specific *GLDP* gene, in addition to a ubiquitously expressed *GLDP* gene, in a C_3 species facilitated the evolution of C_3-C_4 intermediate photosynthesis. This is because a knockout or the drastic downregulation of the ubiquitously expressed *GLDP* gene would suffice to initiate the establishment of a photorespiratory CO_2 pump as a precondition to evolve the C_4 pathway (Sage, 2004; Bauwe, 2011; Sage et al., 2012). As deduced from the phylogeny (McKown et al., 2005), a photorespiratory CO_2 pump must have evolved twice in *Flaveria*, once in the lineage leading to the C_3-C_4 intermediate *F. sonorensis*, and once in the lineage leading to the C_3-C_4 intermediate and C_4 species of clades A and B.

To support the notion that the presence of bundle sheath–specific *GLDP* genes in C_3 species could be a general enhancer of C_4 evolution, one could analyze *GLDP* genes from C_3 species closely related to clades with several C_4 origins, such as the Amaranthaceae or the PACMAD clade of grasses. The grasses are probably the oldest angiosperm lineage in which C_4 species evolved (Edwards et al., 2010) and therefore illustrate how fully optimized C_4 species finally look with respect to metabolic organization and the underlying transcriptional regulation. While the genome of the C_3 grass rice (*Oryza sativa*) contains two *GLDP* genes (Goff et al., 2002; LOC_Os01g51410 and LOC_OS06g40940), the genomes of the C_4 grasses maize (*Zea mays*; Schnable et al., 2009) (GRMZM2G104310), sorghum (*Sorghum bicolor*; Paterson et al., 2009) (Sb08g003440), and *Setaria italica* (Bennetzen et al., 2012) (Si000068m) harbor only one *GLDP* copy. If the C_4 grasses represent the terminal stage of C_4 evolution and if they pursued a similar evolutionary path as *Flaveria*, one may speculate that their ubiquitously expressed *GLDP* gene(s) have been lost from the genomes after pseudogenization.

The Photorespiratory CO_2 Pump in *Flaveria* Is Established Gradually

It was proposed that the photorespiratory CO_2 pump was established by an abrupt loss of GLDP in the mesophyll cells (Sage, 2004; Sage et al., 2012). This hypothesis does not comply with our RNA profiling studies, which included a representative set of species from all different phylogenetic groups of *Flaveria* ranging from those with C_3 through C_3-C_4 intermediate and C_4-like to C_4 photosynthesis. Our investigations demonstrated that GDC disappeared from the mesophyll cells not abruptly but gradually. We showed that the ubiquitously expressed group B *GLDP* genes are downregulated in the C_3-C_4 intermediate *Flaveria* species belonging to clade A as well as to clade B compared with the C_3 species. However, they are not completely switched off (Figure 5). Clade A and B *Flaveria* species derived from a common ancestor that was most likely a C_3-C_4 intermediate (McKown et al., 2005). The reduction in the expression of group B *GLDP* genes has therefore been initiated during the establishment of the photorespiratory CO_2 pump in this last common ancestor of clade A and B *Flaveria* species. After the separation of clade A and B species, expression of these genes was progressively further reduced in clade B species (i.e., the C_4-like species *F. brownii*) or even completely switched off in the C_4 species of clade A.

These results are in line with a recent study that modeled the evolution of the C_4 pathway based on combining a biochemical model with a population genetic framework (Heckmann et al., 2013). The model shows that all C_3-C_4 intermediates, including those from clade B, can be treated as true evolutionary intermediates with the potential to develop a fully functional C_4 cycle. The model predicts that the photorespiratory CO_2 concentrating cycle must have originated early during the evolution of C_4 photosynthesis and that the cycle was established gradually.

Is it plausible that the photorespiratory CO_2 pump was not established abruptly (Sage, 2004), but step by step? One can imagine that the capacities to decarboxylate large amounts of Gly efficiently and recapture the correspondingly large amounts of photorespiratory CO_2 were not ab initio present in the bundle sheaths of C_3 ancestors of contemporary C_4 plants. Indeed, bundle sheath cells of present C_3 species with "Proto-Kranz" anatomy are still relatively poor in chloroplasts and mitochondria (Muhaidat et al., 2011; Sage et al., 2012). If such a C_3 species were to abruptly lose all its Gly decarboxylation activity in the mesophyll, it would most probably not be viable anymore. A gradual reduction of Gly decarboxylation in the mesophyll cells could initiate a series of steps organized in a positive feedback loop (Bauwe, 2011; Muhaidat et al., 2011; Sage et al., 2012). Gly had to diffuse to the bundle sheath for decarboxylation, thereby creating a higher CO_2 concentration around Rubisco in the bundle sheath. The Rubisco in the bundle sheath would become more engaged in CO_2 fixation than the mesophyll enzyme, thus creating a selection pressure to enhance the number of bundle sheath chloroplasts and the amount of Rubisco in these cells. Even more Gly decarboxylation activity could then be shifted to the bundle sheath cells; concomitantly, the number of mitochondria would increase. This would lead to further CO_2 enrichment in the bundle sheath and allow an increase in the amount of Rubisco in this compartment, in which the enzyme would operate under the favorable condition of an elevated CO_2 concentration. Due to the surrounding mesophyll, a second beneficial outcome of this positive feedback loop would have been a higher CO_2 refixation capacity than before. Other evolutionary adaptations would have occurred in parallel, for instance, the optimization of organelle positioning in the bundle sheath cells (i.e., centripetal mitochondria and centrifugal chloroplasts), the upregulation of inter- and intracellular Gly and Ser transport, and/or changes in overall leaf anatomy.

The Tandem Promoter Structure of *GLDPA* Is Conserved in Both C_3 and C_4 *Flaveria* Species, but the RNA Output Is Not

The 5′ flanking sequences of the *GLDPA* genes of C_3 *Flaveria* species are very similar to those of their counterparts in the C_4 species. The tandem promoter structure is highly conserved (Figure 6) and both promoters of the C_3 species direct the same expression specificities in transgenic *Arabidopsis* and *F. bidentis* as the corresponding regions from the C_4 species *F. trinervia* (compared with Wiludda et al., 2012; Figure 7; see Supplemental Figure 4 online). By contrast, the *GLDPA* transcript profiles differ

between the C_3 and C_4 *Flaveria* species. While the RNA output from the distal *GLDPA* promoter is negligible in C_3 *Flaveria* species, RNAs derived from this promoter accumulate to small amounts in C_4 *Flaveria* species (Figure 8).

These findings could suggest that the distal *GLDPA* promoter is silent in the context of the authentic 5′ flanking region in the C_3 *Flaveria* species, whereas its counterpart in the C_4 *Flaveria* species is active. This would imply that the distal promoter is cryptic in the C_3 species and became activated only in the course of C_4 evolution. The promoter activation could be brought about by changes in the rates of transcriptional initiation, pausing, or elongation (reviewed in Shearwin et al., 2005; Levine, 2011; Palmer et al., 2011). Alternatively, the different RNA output from the distal promoter in C_3 versus C_4 *Flaveria* species may not be regulated transcriptionally but posttranscriptionally at the level of transcript stability. Indeed, tentative evidence indicates that the 5′ intron is involved in regulating the accumulation of stable transcripts from the distal promoter of the *GLDPA* gene of *F. trinervia* (C_4) (Wiludda et al., 2012), possibly via nonsense-mediated mRNA decay (Kertesz et al., 2006; Hori and Watanabe, 2007; Brogna and Wen, 2009). How much each of these regulatory levels contributes to the differences in RNA output from these two types of orthologous promoters remains to be investigated.

Why do the distal promoters of C_3 and C_4 *Flaveria* species differ in their RNA output? We proposed recently (Wiludda et al., 2012) that expression from the distal *GLDPA* promoter must be leaky in C_4 *Flaveria* species because each plant cell must be capable of synthesizing C_1 compounds regardless of whether or not it photorespires. A complete shutdown of GDC in the mesophyll cells of C_4 plants would thus be fatal (Bauwe, 2011). The promoter could be silent in C_3 and C_3-C_4 intermediate species because these plants possess a group B *GLDP* gene that is active in all leaf chlorenchyma cells (Figure 5); consequently, no selective pressure would favor leakiness in expression, as is observed in C_4 species.

From Evolutionary Analysis to Synthetic Experimental Evolution

Flaveria is the youngest genus with respect to C_4 evolution (Christin et al., 2011a), and the large number of C_3-C_4 intermediate species (Edwards and Ku, 1987; McKown et al., 2005) suggests that C_4 evolution is still going on in *Flaveria*. We used *Flaveria* to analyze the evolutionary trajectory toward the establishment of a photorespiratory CO_2 pump and its further integration into the C_4 pathway. The model derived from these studies shows that the photorespiratory CO_2 pump evolved step by step and that this gradual evolution was eased by the presence of duplicated *GLDP* genes differing in expression specificity (Figure 9).

We do not know whether this evolutionary scenario is unique for *Flaveria* or whether it represents a general model for C_3-to-C_4 transitions. It would be worthwhile, therefore, to study phylogenetically diverse genera that contain both C_3 and C_4 species, and ideally also C_3-C_4 intermediates, as for instance *Mollugo* (Christin et al., 2011b), *Cleome* (Marshall et al., 2007), or *Heliotropium* (Muhaidat et al., 2011). An alternative approach could pursue synthetic experimental evolution (Morange, 2009) using C_3 model plants such as *Arabidopsis* that can easily be manipulated by genetic engineering and are suitable for multiple rounds of mutation and selection due to their fast life cycles. The optimization of photosynthesis by placing a C_4 pathway into current C_3 species is on the agenda of crop biologists (Hibberd et al., 2008; von Caemmerer et al., 2012). Setting up a photorespiratory CO_2 pump by synthetic experimental evolution is an important and necessary component of this endeavor.

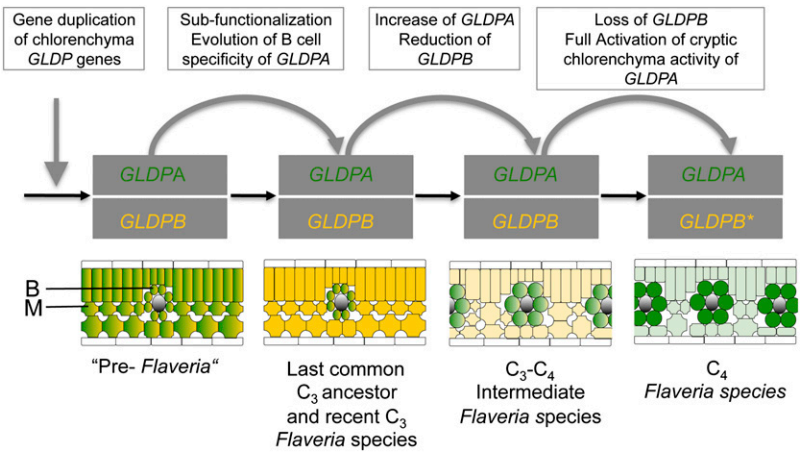

Figure 9. Model for the Evolution of Bundle Sheath–Specific *GLDP* Expression in the Genus *Flaveria*.

The duplication of a photorespiratory *GLDP* gene in early, ancestral C_3 *Flaveria* species led to two ubiquitously expressed *GLDP* genes with identical expression patterns in all chlorenchyma tissues (M, mesophyll; B, bundle sheath) of the leaf. Subfunctionalization remodeled the expression of the group A *GLDP* genes to become bundle sheath specific and led to an ancestral C_3 species with the same spatial *GLDP* expression pattern as today's C_3 species. During transition to C_3-C_4 intermediate photosynthesis, the expression level of group B *GLDP* genes was reduced and the distal *GLDPA* subpromoter became activated. Finally, group B *GLDP* genes were inactivated by pseudogenization. Green, group A *GLDP* spatial expression; yellow, group B *GLDP* spatial expression.

METHODS

Sequence Alignments and Phylogenetic Analyses

Sequences used to construct the phylogenetic tree of the *GLDP* genes of the *Flaveria* species were obtained from either known sequences (*GLDPA*-Ft [Cossu and Bauwe, 1998], *GLDPB**-Ft [Cossu and Bauwe, 1998], *GLDPA1*-Fp [Bauwe et al., 1995], *GLDPA2*-Fp [Bauwe et al., 1995], *GLDPB1*-Fp [Chu, 1996], *GLDPB2*-Fp [Chu, 1996], and *GLDPC*-Fp [Chu, 1996]) or from contigs assembled from either 454 (Gowik et al., 2011) or Illumina (J. Mallmann, U. Gowik, and P. Westhoff, unpublished data) sequencing. Alignments of sequences for phylogenetic analyses were performed with ClustalX 2.0.8 (Higgins and Sharp, 1988, 1989; Thompson et al., 1997; Larkin et al., 2007). A full alignment of the sequences used is available as Supplemental Data Set 1 online. Phylogenetic analyses were done with the program MEGA 5 (Tamura et al., 2011) using the maximum likelihood method with the Tamura 3 parameter model. Bootstrapping was performed 1000 times. Sequence comparisons for similarity studies between the promoters were done with the Genomatix DiAlign Web interface (Morgenstern et al., 1996; Morgenstern et al., 1998; Morgenstern, 1999).

Mapping and Quantification of Reads

Illumina reads from sequencing the leaf transcriptomes of *Flaveria pringlei*, *Flaveria robusta*, *Flaveria chloraefolia*, *Flaveria pubescens*, *Flaveria anomala*, *Flaveria ramossisima*, *Flaveria brownii*, *Flaveria bidentis*, and *Flaveria trinervia* were obtained in four independent experiments. All plant species were grown next to each other in the greenhouse, RNA was isolated as described (Westhoff et al., 1991), and construction of sequence libraries and sequencing followed Illumina protocols. Between 30 and 58 million reads could be obtained per species and experiment (J. Mallmann, U. Gowik, and P. Westhoff, unpublished data). The reads were mapped against the sequences of the *GLDPA* gene of *F. trinervia* and the *GLDPB** gene of *F. trinervia* to obtain an overall distribution of reads along the 5′ flanking and coding regions of the genes. Mapping was performed with the CLC Genomics server version 3.2.1 by CLC Bio with the "map reads against reference" tool for high-throughput sequencing.

To determine the abundance of *GLDPA*, *GLDPB*, and *GLDPC* transcripts in roots, stems, and leaves of *F. bidentis* and *F. robusta*, the coding sequences of *GLDPA*-Ft, *GLDPB*-Fro, and *GLDPC*-Fro (the latter two obtained from full-length contigs from assembled Illumina reads) were used as references and Illumina reads from sequencing the root, stem, and leaf transcriptomes of *F. bidentis* and *F. robusta* were mapped on these. Leaf, stem, and root transcriptomes of *F. robusta* and *F. bidentis* were sequenced in a single experiment as described above. Between 37 and 42 million reads could be obtained per species and tissue. Plants were grown next to each other in the greenhouse. The sequences for *GLDPB*-Fro and *GLDPC*-Fro are available as Supplemental Data Set 3 online.

Isolation of *GLDP* 5′ Flanking Regions by Vectorette PCR

The 5′ flanking regions of *GLDP* genes from *F. robusta* and *F. bidentis* were isolated by vectorette PCR (Siebert et al., 1995) as implemented in the Genome Walking method. Libraries of genomic DNA were prepared as described in the GenomeWalker Universal Kit manual from Clontech. DNA for library construction was isolated from leaves of the respective *Flaveria* species with the Qiagen DNeasy plant mini kit.

Cloning of Promoter-Reporter Gene Constructs

The 5′ regions of *GLDPA*-Fp and *GLDPB*-Fp genes were fused to the *GUS* (*uidA*) gene (Novel and Novel, 1973) and cloned into the pBin19 plant transformation vector (Bevan, 1984). All other constructs were constructed using restriction sites added with PCR to the respective sequences. Following sequencing for confirmation of the sequence, these fragments were inserted in pBI121 (Jefferson et al., 1987; Chen et al., 2003). The sequences of the oligonucleotides used can be found in Supplemental Table 1 online.

Transformation of *Arabidopsis thaliana* and *F. bidentis*

Transformation of *Arabidopsis* was performed following the floral dip protocol (Clough and Bent, 1998) as modified by Logemann et al. (2006). Strain GV3101 of *Agrobacterium tumefaciens* (Holsters et al., 1980; Koncz and Schell, 1986) provided the helper plasmid for the transformations. *F. bidentis* was transformed as described by Chitty et al. (1994) using *Agrobacterium* strain AGL1 (Hood et al., 1986; Lazo et al., 1991).

In Situ Detection of GUS and Fluorimetric Activity Measurements

Fluorimetric measurements of GUS activity were performed according to Jefferson et al. (1987) and Kosugi et al. (1990). In the case of *F. bidentis*, the fifth leaf of a 40- to 50-cm tall T0 plant was harvested for the analysis, and in the case of *Arabidopsis*, three rosette leafs were harvested from T1 plants that were around 4 weeks old. For both species, leaves were harvested before the onset of flowering. Histochemical GUS staining and light microscopy were performed as described by Engelmann et al. (2008).

Accession Numbers

Sequence data from this article are available in the GenBank/EMBL data libraries under the following accession numbers: Z36879 (*F. pringlei GLDPA2*; formerly *gdcsPA*), Z54239 (*F. pringlei GLDPA1*; formerly *gdcsPB*), KC545949 (*F. pringlei GLDPC*; formerly *gdcsPC*), KC545950 (*F. pringlei GLDPB2*; formerly *gdcsPD*), KC545951 (*F. pringlei GLDPB1*; formerly *gdcsPE*), Z99767 (*F. trinervia GLDPA*; formerly *gdcsPA*), Z99768 (*F. trinervia GLDPB**; formerly *gdcsPB pseudogene*), KC545946 (5′ flanking sequence of *F. bidentis GLDPA*), and KC545947 (5′ flanking sequence of *F. robusta GLDPA*). The *Arabidopsis* genes *GLDP1* (AT4G33010) and *GLDP2* (AT2G26080) can be found at The Arabidopsis Information Resource (www.Arabidopsis.org).

Supplemental Data

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

Supplemental Figure 1. Molecular Phylogenetic Analysis of *GLDP* Genes.

Supplemental Figure 2. Functional Analysis of the 5′ Flanking Sequence of the *GLDPA2* Gene of *F. pringlei*.

Supplemental Figure 3. Functional Analysis of the 5′ Flanking Sequence of the *GLDPB** Gene of *F. trinervia*.

Supplemental Figure 4. Functional Analysis of the Proximal (P_{R7}) and Distal (P_{R2}) Subpromoters of the *GLDPA* Gene of the C_3 Species *F. robusta*.

Supplemental Figure 5. Transcript Coverage of the *GLDPA* Genes of the C_3-C_4 Intermediates *F. chloraefolia*, *F. pubescens*, *F. anomala*, *F. ramosissima*, and *F. brownii*.

Supplemental Figure 6. Splice Variants of *GLDPA* Transcripts Derived from the Distal Transcriptional Start Site.

Supplemental Table 1. List of Oligonucleotides Used to Amplify Promoter Sequences for GUS Constructs and the 5′ Splice Site of Transcripts Derived from Distal Transcriptional Start Site of *F. bidentis*.

Supplemental Data Set 1. ClustalX Alignment of the *GLDP* Sequences Used for the Phylogenetic Analysis Shown in Figure 1A.

Supplemental Data Set 2. ClustalX Alignment of the *GLDP* Sequences Used for the Phylogenetic Analysis in Supplemental Figure 1 online.

Supplemental Data Set 3. Sequences of the *GLDPB* and *GLDPG* Genes of *F. robusta*.

ACKNOWLEDGMENTS

We thank Borjana Arsova for carefully reading this article and giving useful comments. This work was funded by the Deutsche Forschungsgemeinschaft through the Sonderforschungsbereich SFB 590, the Forschergruppe FOR1186 (Promics), and the Excellence Cluster EXC 1028.

AUTHOR CONTRIBUTIONS

S.S., H.B., U.G., and P.W. designed the research. S.S., J.M., J.B., M.K., and M.S. performed the research. S.S. and U.G. analyzed the data. S.S., H.B., U.G., and P.W. wrote the article.

Received June 4, 2013; revised June 20, 2013; accepted June 28, 2013; published July 11, 2013.

REFERENCES

Akyildiz, M., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., and Westhoff, P. (2007). Evolution and function of a *cis*-regulatory module for mesophyll-specific gene expression in the C_4 dicot *Flaveria trinervia*. Plant Cell **19:** 3391–3402.

Bauwe, H. (2011). Photorespiration: The bridge to C_4 photosynthesis. In C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms, A.S. Raghavendra and R.F. Sage, eds (Dordrecht, The Netherlands: Springer), pp. 81–108.

Bauwe, H., Chu, C.C., Kopriva, S., and Nan, Q. (1995). Structure and expression analysis of the *gdcsPA* and *gdcsPB* genes encoding two P-isoproteins of the glycine-cleavage system from *Flaveria pringlei*. Eur. J. Biochem. **234:** 116–124.

Bauwe, H., and Kopriva, S. (1995). The *gdcsPA* gene from *Flaveria pringlei* (Asteraceae). Plant Physiol. **107:** 655.

Bennetzen, J.L., et al. (2012). Reference genome sequence of the model plant *Setaria*. Nat. Biotechnol. **30:** 555–561.

Bevan, M. (1984). Binary *Agrobacterium* vectors for plant transformation. Nucleic Acids Res. **12:** 8711–8721.

Brogna, S., and Wen, J. (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. Nat. Struct. Mol. Biol. **16:** 107–113.

Cameron, R.G., Bassett, C.L., Bouton, J.H., and Brown, R.H. (1989). Transfer of C_4 photosynthetic characters through hybridization of *Flaveria* species. Plant Physiol. **90:** 1538–1545.

Chen, P.Y., Wang, C.K., Soong, S.C., and To, K.Y. (2003). Complete sequence of the binary vector pBI121 and its application in cloning T-DNA insertion from transgenic plants. Mol. Breed. **11:** 287–293.

Chitty, J.A., Furbank, R.T., Marshall, J.S., Chen, Z., and Taylor, W.C. (1994). Genetic transformation of the C_4 plant, *Flaveria bidentis*. Plant J. **6:** 949–956.

Christin, P.A., Osborne, C.P., Sage, R.F., Arakaki, M., and Edwards, E.J. (2011a). C_4 eudicots are not younger than C_4 monocots. J. Exp. Bot. **62:** 3171–3181.

Christin, P.A., Sage, T.L., Edwards, E.J., Ogburn, R.M., Khoshravesh, R., and Sage, R.F. (2011b). Complex evolutionary transitions and the significance of C_3-C_4 intermediate forms of photosynthesis in Molluginaceae. Evolution **65:** 643–660.

Chu, C. (1996). Molecular Structure and Expression Patterns of Glycine Decarboxylase Genes from *Flaveria pringlei* C_3 and *Flaveria anomala* C_3-C_4. PhD dissertation (Halle-Wittenberg, Germany: Martin-Luther University).

Clough, S.J., and Bent, A.F. (1998). Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J. **16:** 735–743.

Cossu, R., and Bauwe, H. (1998). The electronic Plant Gene Register. Plant Physiol. **116:** 445–446.

Dengler, N.G., and Nelson, T. (1999). Leaf structure and development in C_4 plants. In C_4 Plant Biology, R.F. Sage and R.K. Monson, eds (San Diego, CA: Academic Press), pp. 133–172.

Edwards, E.J., et al. (2010). The origins of C_4 grasslands: Integrating evolutionary and ecosystem science. Science **328:** 587–591.

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

Engel, N., van den Daele, K., Kolukisaoglu, U., Morgenthal, K., Weckwerth, W., Parnik, T., Keerberg, O., and Bauwe, H. (2007). Deletion of glycine decarboxylase in *Arabidopsis* is lethal under nonphotorespiratory conditions. Plant Physiol. **144:** 1328–1335.

Engelmann, S., Wiludda, C., Burscheidt, J., Gowik, U., Schlue, U., Koczor, M., Streubel, M., Cossu, R., Bauwe, H., and Westhoff, P. (2008). The gene for the P-subunit of glycine decarboxylase from the C_4 species *Flaveria trinervia*: Analysis of transcriptional control in transgenic *Flaveria bidentis* (C_4) and *Arabidopsis* (C_3). Plant Physiol. **146:** 1773–1785.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution **39:** 783–791.

Goff, S.A., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). Science **296:** 92–100.

Gowik, U., Brautigam, A., Weber, K.L., Weber, A.P., and Westhoff, P. (2011). Evolution of C_4 photosynthesis in the genus *Flaveria*: How many and which genes does it take to make C_4? Plant Cell **23:** 2087–2105.

Gowik, U., Burscheidt, J., Akyildiz, M., Schlue, U., Koczor, M., Streubel, M., and Westhoff, P. (2004). *cis*-Regulatory elements for mesophyll-specific gene expression in the C_4 plant *Flaveria trinervia*, the promoter of the C_4 phospho*enol*pyruvate carboxylase gene. Plant Cell **16:** 1077–1090.

Hanson, A.D., and Roje, S. (2001). One-carbon metabolism in higher plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **52:** 119–137.

Hatch, M., Kagawa, T., and Craig, S. (1975). Subdivision of C_4-pathway species based on differing C_4 acid decarboxylating systems and ultrastructural features. Aust. J. Plant Physiol. **2:** 111–128.

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

Hattersley, P.W. (1984). Characterization of C_4 type leaf anatomy in grasses (Poaceae). Mesophyll: Bundle sheath area ratios. Ann. Bot. (Lond.) **53:** 163–180.

Heckmann, D., Schulze, S., Denton, A., Gowik, U., and Westhoff, P., Weber, A.P., and Lercher, M.J. (2013). Predicting C_4 photosynthesis evolution: Modular, individually adaptive steps on a Mount Fuji fitness landscape. Cell **153:** 1579–1588.

Hibberd, J.M., Sheehy, J.E., and Langdale, J.A. (2008). Using C_4 photosynthesis to increase the yield of rice-rationale and feasibility. Curr. Opin. Plant Biol. **11:** 228–231.

Higgins, D.G., and Sharp, P.M. (1988). CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. Gene **73:** 237–244.

Higgins, D.G., and Sharp, P.M. (1989). Fast and sensitive multiple sequence alignments on a microcomputer. Comput. Appl. Biosci. **5:** 151–153.

Holsters, M., et al. (1980). The functional organization of the nopaline *A. tumefaciens* plasmid pTiC58. Plasmid **3:** 212–230.

Hood, E.E., Helmer, G.L., Fraley, R.T., and Chilton, M.D. (1986). The hypervirulence of *Agrobacterium tumefaciens* A281 is encoded in a region of pTiBo542 outside of T-DNA. J. Bacteriol. **168:** 1291–1301.

Hori, K., and Watanabe, Y. (2007). Context analysis of termination codons in mRNA that are recognized by plant NMD. Plant Cell Physiol. **48:** 1072–1078.

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: Beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. **6:** 3901–3907.

Kertesz, S., Kerenyi, Z., Merai, Z., Bartos, I., Palfy, T., Barta, E., and Silhavy, D. (2006). Both introns and long 3′-UTRs operate as *cis*-acting elements to trigger nonsense-mediated decay in plants. Nucleic Acids Res. **34:** 6147–6157.

Koncz, C., and Schell, J. (1986). The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. Mol. Gen. Genet. **204:** 383–396.

Kopriva, S., and Bauwe, H. (1994). P-protein of glycine decarboxylase from *Flaveria pringlei*. Plant Physiol. **104:** 1077–1078.

Kopriva, S., Chu, C.C., and Bauwe, H. (1996). Molecular phylogeny of *Flaveria* as deduced from the analysis of nucleotide sequences encoding the H-protein of the glycine cleavage system. Plant Cell Environ. **19:** 1028–1036.

Kosugi, S., Ohashi, Y., Nakajima, K., and Arai, Y. (1990). An improved assay for β-glucuronidase in transformed cells: Methanol almost completely suppresses a putative endogenous β-glucuronidase activity. Plant Sci. **70:** 133–140.

Larkin, M.A., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics **23:** 2947–2948.

Lazo, G.R., Stein, P.A., and Ludwig, R.A. (1991). A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. Biotechnology (NY) **9:** 963–967.

Levine, M. (2011). Paused RNA polymerase II as a developmental checkpoint. Cell **145:** 502–511.

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

Logemann, E., Birkenbihl, R.P., Ulker, B., and Somssich, I.E. (2006). An improved method for preparing *Agrobacterium* cells that simplifies the *Arabidopsis* transformation protocol. Plant Methods **2:** 16.

Marshall, D.M., Muhaidat, R., Brown, N.J., Liu, Z., Stanley, S., Griffiths, H., Sage, R.F., and Hibberd, J.M. (2007). *Cleome*, a genus closely related to *Arabidopsis*, contains species spanning a developmental progression from C_3 to C_4 photosynthesis. Plant J. **51:** 886–896.

McKown, A.D., Moncalvo, J.M., and Dengler, N.G. (2005). Phylogeny of *Flaveria* (Asteraceae) and inference of C_4 photosynthesis evolution. Am. J. Bot. **92:** 1911–1928.

Monson, R.K., Edwards, G.E., and Ku, M.S.B. (1984). C_3-C_4 intermediate photosynthesis in plants. Bioscience **34:** 563–574.

Morange, M. (2009). Synthetic biology: A bridge between functional and evolutionary biology. Biol. Theory **4:** 368–377.

Morgan, C.L., Turner, S.R., and Rawsthorne, S. (1993). Coordination of the cell-specific distribution of the four subunits of glycine decarboxylase and of serine hydroxymethyltransferase in leaves of C_3-C_4 intermediate species from different genera. Planta **190:** 468–473.

Morgenstern, B. (1999). DIALIGN 2: Improvement of the segment-to-segment approach to multiple sequence alignment. Bioinformatics **15:** 211–218.

Morgenstern, B., Dress, A., and Werner, T. (1996). Multiple DNA and protein sequence alignment based on segment-to-segment comparison. Proc. Natl. Acad. Sci. USA **93:** 12098–12103.

Morgenstern, B., Frech, K., Dress, A., and Werner, T. (1998). DIALIGN: Finding local similarities by multiple sequence alignment. Bioinformatics **14:** 290–294.

Muhaidat, R., Sage, T.L., Frohlich, M.W., Dengler, N.G., and Sage, R.F. (2011). Characterization of C_3-C_4 intermediate species in the genus *Heliotropium* L. (Boraginaceae): Anatomy, ultrastructure and enzyme activity. Plant Cell Environ. **34:** 1723–1736.

Novel, G., and Novel, M. (1973). Mutants d'*Escherichia coli* K 12 affectés pour leur croissance sur méthyl-β-D-glucuronide: Localisation du gène de structure de la β-D-glucuronidase (uid A). Mol. Gen. Genet. **120:** 319–335.

Ogren, W.L. (1984). Photorespiration - Pathways, regulation, and modification. Annu. Rev. Plant Physiol. Plant Mol. Biol. **35:** 415–442.

Ogren, W.L., and Bowes, G. (1971). Ribulose diphosphate carboxylase regulates soybean photorespiration. Nat. New Biol. **230:** 159–160.

Oliver, D.J., and Raman, R. (1995). Glycine decarboxylase: Protein chemistry and molecular biology of the major protein in leaf mitochondria. J. Bioenerg. Biomembr. **27:** 407–414.

Palmer, A.C., Egan, J.B., and Shearwin, K.E. (2011). Transcriptional interference by RNA polymerase pausing and dislodgement of transcription factors. Transcription **2:** 9–14.

Paterson, A.H., et al. (2009). The *Sorghum bicolor* genome and the diversification of grasses. Nature **457:** 551–556.

Powell, A. (1978). Systematics of *Flaveria* (Flaveriinae-Asteraceae). Ann. Mo. Bot. Gard. **65:** 590–636.

Rawsthorne, S., Hylton, C.M., Smith, A.M., and Woolhouse, H.W. (1988). Photorespiratory metabolism and immunogold localization of photorespiratory enzymes in leaves of C_3 and C_3-C_4 intermediate species of *Moricandia*. Planta **173:** 298–308.

Sage, R.F. (2004). The evolution of C_4 photosynthesis. New Phytol. **161:** 341–370.

Sage, R.F., Christin, P.A., and Edwards, E.J. (2011). The C_4 plant lineages of planet Earth. J. Exp. Bot. **62:** 3155–3169.

Sage, R.F., Sage, T.L., and Kocacinar, F. (2012). Photorespiration and the evolution of C_4 photosynthesis. Annu. Rev. Plant Biol. **63:** 19–47.

Schnable, P.S., et al. (2009). The B73 maize genome: Complexity, diversity, and dynamics. Science **326:** 1112–1115.

Shearwin, K.E., Callen, B.P., and Egan, J.B. (2005). Transcriptional interference–A crash course. Trends Genet. **21:** 339–345.

Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A., and Lukyanov, S.A. (1995). An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res. **23:** 1087–1088.

Stockhaus, J., Schlue, U., Koczor, M., Chitty, J.A., Taylor, W.C., and Westhoff, P. (1997). The promoter of the gene encoding the C_4 form of phospho*enol*pyruvate carboxylase directs mesophyll-specific expression in transgenic C_4 *Flaveria* spp. Plant Cell **9:** 479–489.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28:** 2731–2739.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:** 4876–4882.

von Caemmerer, S., Quick, W.P., and Furbank, R.T. (2012). The development of C_4 rice: Current progress and future challenges. Science **336:** 1671–1672.

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.

Wiludda, C., Schulze, S., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., Bauwe, H., and Westhoff, P. (2012). Regulation of the photorespiratory *GLDPA* gene in C_4 *Flaveria*: An intricate interplay of transcriptional and posttranscriptional processes. Plant Cell **24:** 137–151.