

**Evolution of the expression of the *GLDP* gene
family encoding the P subunit of glycine
decarboxylase in the genus *Flaveria***

Inaugural-Dissertation

Zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

Vorgelegt von

Stefanie Schulze

aus Krefeld

Düsseldorf, Mai 2013

Aus dem Institut für Entwicklungs-und Molekularbiologie der Pflanzen
Der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch- Naturwissenschaftlichen Fakultät der
Heinrich- Heine- Universität Düsseldorf

Referent: Prof. Dr. P. Westhoff

Koreferent: Prof. Dr. M. Feldbrügge

Tag der mündlichen Prüfung: 12.06.2013

Die hier vorliegende Dissertation habe ich eigenständig und ohne unerlaubte Hilfe angefertigt. Die Dissertation habe ich in der vorgelegten oder einer ähnlichen Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Düsseldorf, 3.5.2013

(Stefanie Schulze)

Contents

Abbreviations.....	III
I. Introduction.....	1
1. The C₄ syndrome and its biochemistry	1
1.1 Ribulose-1,5-bisphosphate carboxylase/ oxgenase (RubisCO) – a bifunctional enzyme	1
1.2. The C ₂ cycle – photorespiration – not only a wasteful process.....	2
1.3 C ₄ photosynthesis – a CO ₂ concentrating mechanism	4
2. The evolution of C₄ photosynthesis	6
2.1 C ₄ photosynthesis has a polyphyletic origin	6
2.2 When and why did C ₄ photosynthesis evolve?.....	7
2.3 C ₃ plants already contain all enzymes involved in C ₄ photosynthesis	7
2.4 The transition from C ₃ to C ₄ occurred stepwise	8
2.5 The genus <i>Flaveria</i> – a suitable model to study the evolution of C ₄ photosynthesis	11
3 Glycine decarboxylase	12
3.1 Enzymology of glycine decarboxylase and its function in the photorespiratory C ₂ cycle	12
3.2 Glycine decarboxylase in the genus <i>Flaveria</i>	14
4. Regulation of gene expression	16
4.1 The transcriptional process and its complexity.....	16
4.2 The connection between transcriptional and post-transcriptional regulation	17
4.3 Molecular changes in regulatory mechanisms of C ₄ genes leading to differential expression	19
4.3.1 The mesophyll expression module 1 (MEM1) of phosphoenolpyruvate carboxylase (PEPC) in the genus <i>Flaveria</i>	19
4.3.2 Relocation of RubisCO to the bundle sheath cells through post-transcriptional RNA regulation in <i>Flaveria</i>	21
4.3.3 The relocation of carbonic anhydrase (CA) from the chloroplast to the cytosol in the genus <i>Flaveria</i>	22
II Objectives.....	24
III. Theses.....	25
IV.A Summary.....	26

IV.B Zusammenfassung	27
V. Literature	29
VI. Chapters	39
1. Manuscript 1: Christian Wiludda, Stefanie Schulze, Udo Gowik, Sascha Engelmann Maria Koczor, Monika Streubel, Hermann Bauwe, Peter Westhoff (2012). Regulation of the photorespiratory GLDPA gene in C₄ Flaveria – an intricate interplay of transcriptional and post-transcriptional processes. <i>The Plant Cell</i> 24: 137-151	40
2. Manuscript 2: Stefanie Schulze, Julia Mallmann, Janet Burscheidt, Maria Koczor, Monika Streubel, Hermann Bauwe, Udo Gowik, Peter Westhoff (2013). Evolution of C₄ photosynthesis in the genus <i>Flaveria</i> – establishment of a photorespiratory CO₂ pump. Submitted for publication to <i>The Plant Cell</i>	57
3. Function and origin of the distal promoter (P_{R2}) of the <i>GLDP</i> genes in <i>Flaveria</i>.....	108
4. Identification of <i>cis</i>- regulatory elements for the functioning of the <i>GLDPA</i> 5' flanking region.....	118
5. The roles of splicing and upstream open reading frames on the expression of <i>GLDPA</i> of <i>F. trinervia</i>.....	128

Abbreviations

A	Adenine
<i>A. thaliana</i>	<i>Arabidopsis thaliana</i>
B	Bundle sheath
Bp	Base pairs
bZIP	Basic leucine zipper
°C	Degree Celsius
C	Cytosine
CA	Carbonic anhydrase
CAM	Crassulacean acid metabolism
CAT	Catalase
CO ₂	Carbon dioxide
DNA	Desoxyribonucleic acid
DR	Distal region of MEM1
<i>F.</i>	<i>Flaveria</i>
<i>Fa</i>	<i>Flaveria anomala</i>
<i>Fb</i>	<i>Flaveria bidentis</i>
<i>Fbr</i>	<i>Flaveria brownii</i>
<i>Fch</i>	<i>Flaveria chloraefolia</i>
<i>Fp</i>	<i>Flaveria pringlei</i>
<i>Fpa</i>	<i>Flaveria palmeri</i>
<i>Fpu</i>	<i>Flaveria pubescens</i>
<i>Fra</i>	<i>Flaveria ramossisima</i>
<i>Fro</i>	<i>Flaveria robusta</i>
<i>Ft</i>	<i>Flaveria trinervia</i>
<i>Fv</i>	<i>Flaveria vaginata</i>
G	Guanine
GDC	Glycine decarboxylase
GGT	Glutamate-glyoxylate aminotransferase
<i>GLDH</i>	Glycine decarboxylase H-protein gene
<i>GLDP</i>	Glycine decarboxylase P-protein gene
GOX	Glycolate oxidase
GUS	β-glucuronidase

GLYK	Glycerate kinase
h	Hour(s)
HCO ₃ ⁻	Bicarbonate
HPR	Hydroxypyruvate reductase
Inr	Initiator
kDA	Kilo Dalton
kb	Kilobases
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NH ₃	Ammonia
MCS	Multiple cloning site
MDH	Malate dehydrogenase
ME	Malic enzyme
MEM1	Mesophyll expression module 1
mRNA	Messenger ribonucleic acid
Mya	Million years ago
NMD	Nonsense mediated mRNA decay
nt	Nucleotides
O ₂	Molecular Oxygen
OAA	Oxaloacetate
ORF	Open reading frame
PEP	Phosphoenolpyruvate
PEPC	Phosphoenolpyruvate carboxylase
PEPCK	Phosphoenolpyruvate carboxylase kinase
PEPCK	Phosphoenolpyruvate carboxykinase
3-PGA	3-Phosphoglycerate
2-PG	2-Phosphoglycolate
PGP	Phosphoglycolate phosphatase
<i>ppcA</i>	Phosphoenolpyruvate carboxylase gene A of <i>Flaveria</i>
PPDK	Pyruvate, orthophosphate dikinase
PR	Proximal region of MEM1
Py	Pyrimidine
R	Purine (adenine or guanine)
5' RACE	Rapid amplification of 5' complementary DNA ends

RNA	Ribonucleic acid
RubisCO	Ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose 1,5-bisphosphate
SGT	Serine-glyoxylate aminotransferase
SHMT	Serine hydroxymethyltransferase
T	Thymine
TCA	Tricarboxylic acid
TF	Transcription factor
THF	Tetrahydrofolate
TP	Triose phosphate
TSS	Transcriptional start site
uORF	Upstream open reading frame
5' UTR	5' Untranslated region
UPF	Up-frameshift protein
Y	Pyrimidine (cytosine or thymine)

I. Introduction

1. The C₄ syndrome and its biochemistry

1.1 Ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) – a bifunctional enzyme

Photoautotrophic land plants conduct photosynthesis to make carbohydrates and O₂ out of CO₂ and light. About 85 % (Ehleringer et al., 1991) of the land plants perform C₃ photosynthesis. A minor part of the land plants perform specialized forms of photosynthesis such as C₄ photosynthesis and crassulacean acid metabolism (CAM) (Borland et al., 2011).

C₃ photosynthesis is initiated by the fixation of CO₂ via ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO), the most abundant protein in the world (Ellis, 1979). RubisCO is one of the slowest enzymes known and one of the largest enzymes with a molecular mass of 560 kDA (reviewed in Spreitzer and Salvucci, 2002). It is composed of a large subunit (55 kDA), and a small subunit (15 kDA) that together form a complex of eight large and eight small subunits (reviewed in Spreitzer and Salvucci, 2002).

RubisCO carboxylates its substrate ribulose-1,5-bisphosphate (RuBP) resulting in the C₃ intermediate 3-phosphoglycerate (3-PGA), thus starting the Calvin-Benson cycle (Calvin and Benson, 1948). As the name RubisCO implicates, the enzyme not only carboxylates, but also oxygenates, if O₂ instead of CO₂ is used as a substrate (Bowes et al., 1971; Ogren and Bowes, 1971; Lorimer, 1981). The oxygenation of RuBP produces only one molecule of 3-PGA and one molecule of 2-phosphoglycolate (2-PG) (Lorimer, 1981). 2-PG has to be recycled in a process called photorespiration (see 1.2), due to detrimental effects of 2-PG accumulation in leaves such as the inhibition of phosphofructokinase (Kelly and Latzko, 1976).

When RubisCO evolved more than 3 billion years ago the CO₂ concentration in the atmosphere was much higher than nowadays and almost no O₂ existed (Sage, 1999b; Sage, 2004). This led to the fact that the enzyme and the connecting metabolic pathways were already well established when the O₂ content of the atmosphere raised up to 20 % 0.6 billion years ago and RubisCO's oxygenase activity became a disadvantage (Sage, 1999b). Thus, RubisCO was already well established in photosynthetically organisms in the Pleistocene (2 to 0.01 million years ago), when the oxygenase activity reached levels up to 40 % of the

carboxylase activity at 30 °C, thereby leading to an increase of photorespiration (Sage, 1999b).

Alterations in the enzyme to increase the affinity to CO₂ or lower its affinity for O₂ come together with a slower enzymatic turnover rate, since the same catalytic centre recognizes both substrates (Whitney et al., 2011). Nonetheless some organisms such as red algae contain RubisCO with a higher affinity for CO₂ than average land plants (Andersson, 2008). In nature some organisms evolved mechanisms to avoid the oxygenase activity of RubisCO. One mechanism is C₄ photosynthesis, which is found only in higher plants, and the carbon concentrating mechanisms of algae and cyanobacteria (Giordano et al., 2005; Price et al., 2008). C₄ plants use phosphoenolpyruvate carboxylase (PEPC) to primary fix CO₂ in the form of bicarbonate (HCO₃⁻) and separate the primary fixation and RubisCO spatially. This concentrating mechanism of CO₂ around RubisCO further allowed these plants to evolve a faster but less specific RubisCO than C₃ plants (Kubien et al., 2008; Kapralov et al., 2011). In contrast, the CO₂ concentrating mechanisms of cyanobacteria and algae are biophysical pumps that are based on active transport of HCO₃⁻ and/or CO₂ from the surrounding area into the cell (Giordano et al., 2005). The actively acquired HCO₃⁻ and/or CO₂ then diffuses in the carboxysomes, in which carbonic anhydrase is exclusively located and thus CO₂ gets concentrated at the site of RubisCO, which is also located in the carboxysomes (Giordano et al., 2005).

1.2. The C₂ cycle – photorespiration – not only a wasteful process

The main function of the photorespiratory pathway is the recycling of 2-PG (Bowes et al., 1971; Ogren and Bowes, 1971). This pathway is therefore also called the photorespiratory C₂ cycle. It takes place in three organelles of the leaves, the chloroplast, the peroxisome and the mitochondrion (see Figure 1).

2-PG produced in the oxygenase reaction of RubisCO is dephosphorylated to glycolate by phosphoglycolate phosphatase. It is then transported from the chloroplast into the peroxisome. Glycolate oxidase catalyses the oxidation of glycolate to glyoxylate and H₂O₂. A catalase rapidly degrades the H₂O₂ and the glyoxylate is transaminated to glycine by glutamate-glyoxylate aminotransferase. Glycine is then transported into the mitochondrion, where it is decarboxylated and deaminated by glycine decarboxylase (GDC). CO₂, NH₃, NADH and the C₁ compound methylene tetrahydrofolate (CH₂-THF) are thereby produced. The CH₂-THF is used by serine hydroxymethyl transferase (SHMT), together with a second molecule of

be limited by nitrate supply. This indicates that photorespiratory nitrogen is needed in plants to support growth (Rachmilevitch et al., 2004).

Genes coding for the photorespiratory enzymes are found in all photosynthetic organisms even in cyanobacteria (Stabenau and Winkler, 2005; Hagemann et al., 2010). Although mechanisms to concentrate CO₂ around RubisCO were developed in many different ways, e.g. C₄ photosynthesis, it was shown that these plants still require an active C₂ cycle. Glycolate oxidase mutants of maize (*Zea mays*) plants were not able to survive in ambient air, but only at elevated CO₂. This clearly shows that even in maize, a C₄ plant, photorespiration cannot be completely abolished (Zelitch et al., 2009). It was also shown that organisms that perform C₄ photosynthesis do in fact accumulate 2-PG, even if they highly enhance the carboxylase activity of RubisCO and reduce its oxygenase activity, proving that they need photorespiration to recycle 2-PG (Yoshimura et al., 2004; Zelitch et al., 2009).

1.3 C₄ photosynthesis – a CO₂ concentrating mechanism

The vast majority of C₄ plants show a distinct leaf anatomy called “Kranz-anatomy” that was first described by Haberlandt (Haberlandt, 1881). C₄ plants have a specialized cell type directly surrounding the vascular bundle: the bundle sheath cells. One layer of mesophyll cells surrounds these cells. As a consequence each bundle sheath cell stays in direct contact with a mesophyll cell and vice versa (Hattersley, 1984; Dengler and Nelson, 1999). The two cell types differ in their biochemical properties (Figure 2).

The enzymes involved in C₄ photosynthesis are compartmentalized in the mesophyll and the bundle sheath cells. Comparative studies of the levels of RNA of separated mesophyll and bundle sheath cells of C₄ plants showed the differential expression of many of the enzymes involved in the C₄ pathway on the level of transcription (Sheen and Bogorad, 1987; Langdale et al., 1988; Sheen, 1999). The Calvin cycle, that in C₃ plants can be found in all photosynthetically active tissues, is restricted to the bundle sheath cells in C₄ plants (Hatch, 1987; Ehleringer and Monson, 1993) (see Figure 2 indicated in green). Analyses in *Sorghum bicolor* however revealed that only the regenerative part of the Calvin cycle is restricted to the bundle sheath cells, whereas the reducing phase is preferentially located in the mesophyll (Wyrich et al., 1998).

Although mesophyll and bundle sheath cells can be found in C_3 plants (Kinsman and Pyke, 1998), their function and biochemical properties had to change to fulfil their role in C_4 photosynthesis.

The first step in C_4 photosynthesis is the fixation of CO_2 , in the form of bicarbonate (HCO_3^-), in the mesophyll cells via PEPC together with the substrate phosphoenolpyruvate (PEP). The product of this step is the four-carbon acid oxaloacetate (OAA) and this type of photosynthesis is therefore called C_4 photosynthesis. Depending on the type of C_4 photosynthesis, OAA is either converted into malate or aspartate. Each of these four-carbon acids is able to diffuse into the bundle sheath cells.

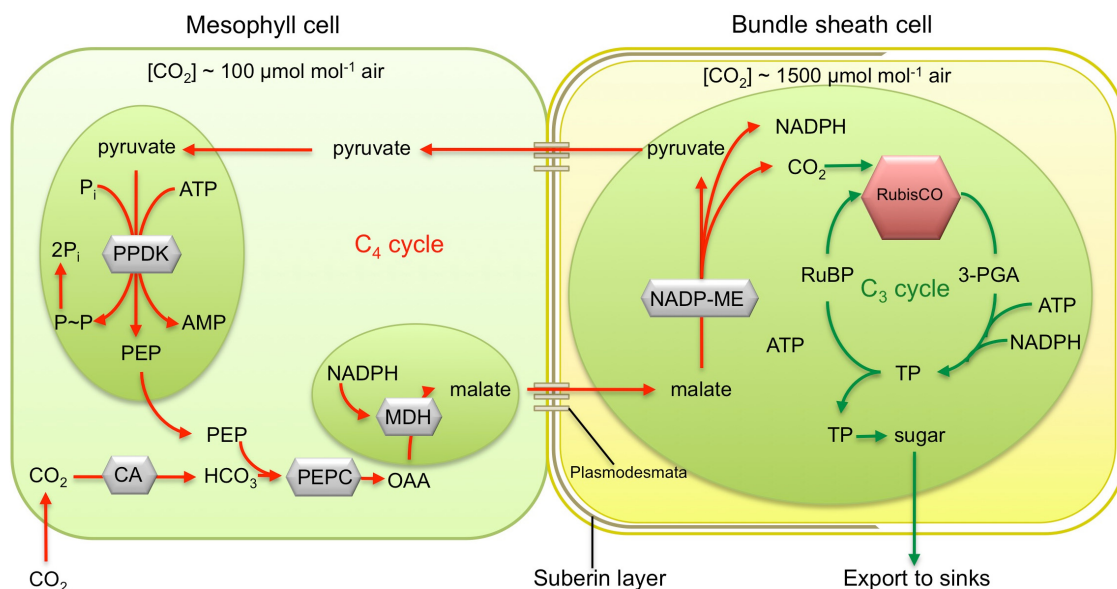


Figure 2 Schematic representation of the NADP-ME type of C_4 photosynthesis.

AMP, adenosine monophosphate; ATP, adenosine triphosphate; CA, carbonic anhydrase; MDH, malate dehydrogenase; NADPH, nicotinamide adenine dinucleotide phosphate; NADP-ME, NADP-malic enzyme; OAA, oxaloacetate; PEP, phosphoenolpyruvate; 3-PGA, 3-phosphoglycerate; PPDK, pyruvate-orthophosphate dikinase; RuBP, ribulose 1,5-bisphosphate; TP, triosephosphate. The C_3 cycle is indicated in green and the C_4 cycle in red. Adapted from Sage et al. (2012).

In the bundle sheath, cells malate or aspartate is decarboxylated using the enzymes NADP-malic enzyme (NADP-ME), NAD-malic enzyme (NAD-ME) or PEP carboxykinase (PEPCK) (Hatch et al., 1975). In the NADP-ME and the NAD-ME type, pyruvate is released in addition to CO_2 . In the NADP-ME type it is directly transported back to the mesophyll cells, and in the NAD-ME type it is first converted to alanine (NAD-ME type) and then transported to the mesophyll cells where it is reconverted to pyruvate (Kanai and Edwards,

1999). PEP is regenerated via pyruvate, orthophosphate dikinase (PPDK) to maintain the C₄ cycle (Figure 2). Plants with the PEPCK type of C₄ photosynthesis perform an additional step in the bundle sheath cells to convert the transport metabolite aspartate back to OAA (Kanai and Edwards, 1999). The following decarboxylation of OAA releases PEP in the bundle sheath cells, and diffusion leads to a regeneration of PEP in the mesophyll cells (Hatch, 1987; Kanai and Edwards, 1999; Wingler et al., 1999).

Although all C₄ plants are assigned to one of these three groups, some C₄ plants use a second decarboxylating enzyme at a lower activity level than the main decarboxylating enzyme (Furbank, 2011; Sage et al., 2012). For instance, the C₄ monocot plant *Zea mays* (maize) is assigned to the NADP-ME subtype. Nevertheless, it was shown that it also uses PEPCK to decarboxylate aspartate in addition to the decarboxylation of malate (Furumoto et al., 1999; Wingler et al., 1999; Furbank, 2011).

The release of CO₂ in the bundle sheath cells leads to a high concentration of CO₂ at the site of RubisCO (Figure 2), and thus to an almost saturated active site of RubisCO (von Caemmerer and Furbank, 2003).

2. The evolution of C₄ photosynthesis

2.1 C₄ photosynthesis has a polyphyletic origin

C₄ plants evolved at least 66 times independently among the angiosperms from C₃ ancestral plants (Sage et al., 2012). Today there are 22-24 C₄ lineages known within the grass family (Poaceae), which contain 4.500 C₄ species (Christin et al., 2008; Vicentini et al., 2008; Bouchenak-Khelladi et al., 2009; Christin and Besnard, 2009; Grass Phylogeny Working, 2012), six lineages in the sedges (Cyperaceae), containing 1.500 species (Besnard et al., 2009; Roalson et al., 2010), ten lineages in the Chenopodiaceae with around 500 C₄ species (Sage, 1999a; Kadereit and Freitag, 2011; Sage et al., 2011) and approximately 1.500 C₄ species in other eudicot families (Sage, 1999a). Even within one genus C₄ photosynthesis may have evolved independently more than once, for instance in the genus *Cleome* of the Cleomaceae family. The Cleomaceae constitute a sister clade to the Brassicaceae, which contain the model plant *Arabidopsis thaliana* (Sage et al., 2012). The fact that C₄ photosynthesis evolved at least

66 times independently in land plants makes it a remarkable example of convergent evolution (Sage et al., 2011).

2.2 When and why did C₄ photosynthesis evolve?

It was originally assumed that low CO₂ in the atmosphere was the major driving force for the evolution of C₄ photosynthesis (Ehleringer et al., 1991; Ehleringer et al., 1997), but today it is assumed that low CO₂ was merely a precondition that enabled other factors such as heat, aridity, high light, salinity and ecological disturbance to play crucial roles as additional driving forces (Ehleringer et al., 1991; Sage, 2004; Sage et al., 2012). All these factors favour increased rates of photorespiration in C₃ plants due to high leaf temperatures and decreased CO₂ levels in leaves. Thus, elevated levels of photorespiration are a common feature for the emergence of C₄ photosynthesis (Sage et al., 2012). Analyses of the biogeographical origins of C₄ photosynthesis indeed show that the habitats found are often characterised by hot, arid and saline environments (Sage et al., 2012). Phylogenetic analyses indicate that the grass subfamily Chloridoideae was the first lineage in which C₄ photosynthesis evolved during mid-Oligocene (30 Mya). In the dicot branch of the angiosperms the oldest lineage are the Caroxyloneae a subfamily of the Chenopodiaceae family that date back 22-25 Mya (Christin et al., 2011; Sage et al., 2012).

2.3 C₃ plants already contain all enzymes involved in C₄ photosynthesis

The polyphyletic origin of C₄ photosynthesis indicates that the changes required to transform a C₃ into a C₄ species must have been relatively easily achieved in genetic terms (reviewed in Westhoff and Gowik, 2010; Brown et al., 2011; Gowik and Westhoff, 2011). Since the biochemical core processes found in independent C₄ lineages are similar, it is assumed that metabolic processes operating in C₃ ancestors only needed to be reorganized in order to form the basis for C₄ evolution (West-Eberhard et al., 2011).

Indeed all enzymes involved in C₄ photosynthesis can be found in C₃ plants, in which they function in carbohydrate and nitrogen metabolism (Aubry et al., 2011; Brown et al., 2011). In many cases a small gene family encodes these enzymes and one of its members became altered to fulfil its new function in C₄ photosynthesis (Aubry et al., 2011). Often, this neo-functionalization was preceded by a duplication of the corresponding gene or the entire genome of the organism (Monson, 2003). Changes in the spatial distribution among the leaf

leading to a mesophyll or bundle sheath specific expression were necessary to establish a functional C_4 cycle. Transcriptional and post-transcriptional modifications regulating the amounts of the enzymes needed to be implemented, and the kinetic properties of the existing C_3 enzymes had to be adapted to their new role in the C_4 cycle (Sage, 2004; Sage et al., 2012).

The C_4 phosphoenolpyruvate carboxylase (PEPC) in the genus *Flaveria* provides a good example illustrating these evolutionary processes (see also 4.3.1). In C_3 plants PEPC has various functions that depend on the tissue and the developmental stage of the C_3 plant. It supplies carbon skeletons to the tricarboxylic acid cycle (Miyao and Fukayama, 2003) and to ammonium assimilation (Masumoto et al., 2010), operates in malate homeostasis during drought stress (González et al., 2003) and regulates stomatal conductance. In C_3 plants only low activities of PEPC are found that became up-regulated in the transition to C_4 and at the same time became restricted to the mesophyll (Stockhaus et al., 1997). The kinetic properties of PEPC changed together with the sensitivity to the inhibitor malate (Svensson et al., 1997).

2.4 The transition from C_3 to C_4 occurred stepwise

Sage (2004; 2012) proposes a model for the evolution of C_4 photosynthesis, which describes the main stages C_3 plants had to pass through in order to perform C_4 photosynthesis (Figure 3). The initial idea of a stepwise evolution from a C_3 species into a C_4 species was already discussed by Monson and Moore (1989). Not all C_3 clades seem to have the potential to evolve C_4 species, therefore the basal step is called the preconditioning for the evolution of C_4 photosynthesis. An essential precondition was the occurrence of extensive gene duplications and the enlargement of the genome size, since duplicated genes offer the possibility to alter one copy by neo-functionalization, while the other retains the original function (Monson, 2003).

A further feature of the preconditioning phase was the existence of close vein spacing, as can be seen in closely related C_3 members of many families that have C_4 members such as *Flaveria* or *Cleome* (Marshall et al., 2007; McKown and Dengler, 2007; Sage et al., 2012). Close veins in C_3 plants were found to lead to a higher photosynthetic capacity via a hydraulic benefit in hot climates (Brodribb et al., 2007; Brodribb and Feild, 2010). In contrast, a recent analysis of the PACMAD clade of the grasses, which in contrast to the BEP clade, evolved C_4 species, indicates that not only close veins but rather an increased bundle sheath to mesophyll ratio favoured the evolution of C_4 in this clade. In particular the interveinal distance in C_3 species of both the PACMAD and the BEP clade do not differ, but the bundle sheath to

mesophyll ratio does (Griffiths et al., 2013). In conclusion, it can be said that close veins in C_3 ancestral species of C_4 plants may have been preceded by an enhancement of the bundle sheath area (Griffiths et al., 2013).

A further aspect of the preconditioning phase was the acquisition of regulatory elements that can easily be altered to confer C_4 patterns of gene regulation. One example for this is the mesophyll expression module 1 (MEM1) in the promoter of the PEPC gene *ppcA* in the genus *Flaveria* (see also 4.3.1).

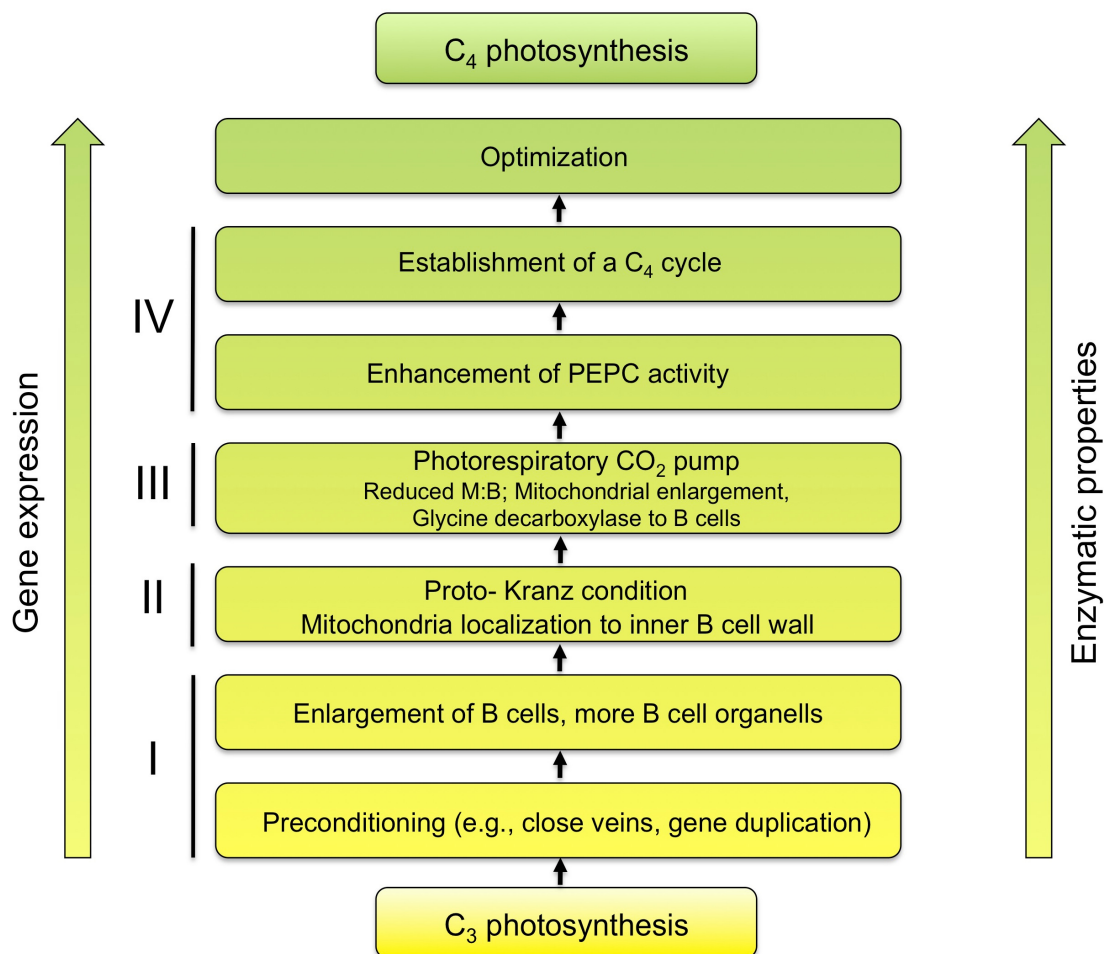


Figure 3. The evolutionary steps of C_4 photosynthesis.

Model-like representation of the main steps of the evolution from C_3 to C_4 photosynthesis. B, bundle sheath; M, mesophyll; PEPC, phosphoenolpyruvate carboxylase. Adapted from Gowik and Westhoff (2011) and Sage et al. (2012).

The second major phase was the establishment of a so-called “Proto-Kranz” condition (Muhaidat et al., 2011; Sage et al., 2012). Several C_3 species, within lineages that possess C_4 photosynthesis, exhibit enlarged bundle sheath cells and have increased organelle numbers compared to C_3 species of lineages without C_4 species (Sage et al., 2012). It can even be

observed that the mitochondria are already localized at the inner wall directing the vascular bundle. This is a feature commonly observed in C_3 - C_4 intermediate species. In some cases the bundle sheath cell chloroplast are facing the bundle sheath cell mitochondria. Muhaidat et al. (2011) hypothesize that this arrangement might lead to some form of single-celled glycine shuttle.

The third phase, the establishment of a photorespiratory CO_2 pump, is assumed to be a key step in the transition from C_3 to C_4 , and is directly influenced by the formation of the “Proto-Kranz” condition. Glycine decarboxylase (GDC), the CO_2 -releasing enzyme of photorespiration, became restricted to the bundle sheath cells (Monson et al., 1984; Sage, 2004; Bauwe, 2011; Sage et al., 2012)(Figure 4).

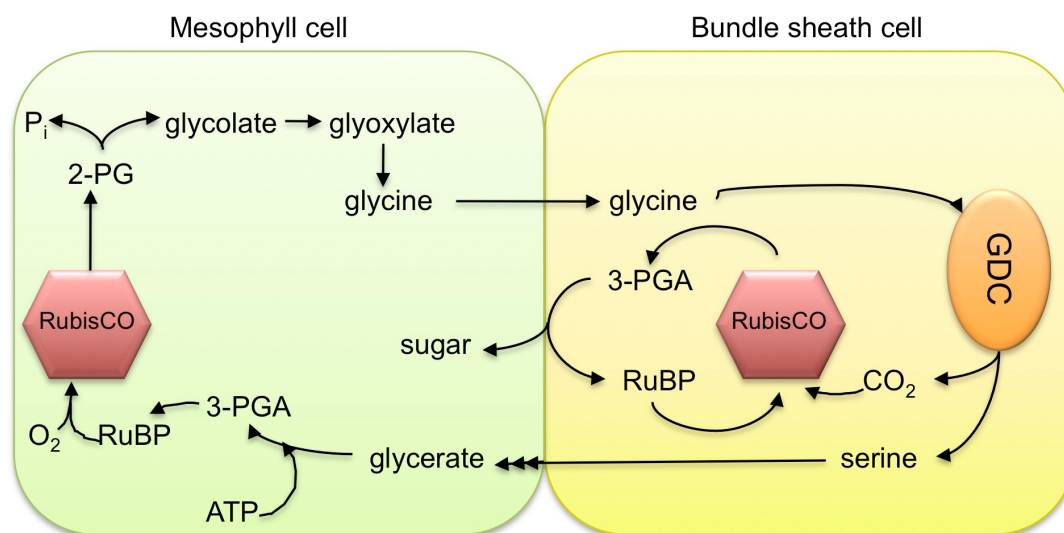


Figure 4. The photorespiratory glycine shuttle - a CO_2 pump.

Glycine produced in the mesophyll cells has to be shuttled in the bundle sheath cells where the glycine decarboxylase (GDC) is located. All photorespiratory CO_2 is released in the bundle sheath cells. A CO_2 concentrating mechanism thus suppresses the oxygenase activity of the bundle sheath cell ribulose-1,5-bisphosphate-carboxylase/oxygenase (RubisCO). ATP, adenosinetriphosphate; 2-PG, 2-phosphoglycolate; 3-PGA, 3-phosphoglycerate; RuBP, ribulose-1,5-bisphosphate. Adapted from Sage (2004).

This relocation of GDC to the bundle sheath cells changed the effect of photorespiration on the system drastically (Monson et al., 1984; Sage, 2004; Sage et al., 2012). Instead of being inhibitory, it now became a carbon source for the chloroplasts in the bundle sheath cells. Bauwe (2011) therefore called photorespiration “a bridge to C_4 photosynthesis”. This change favours the optimization of leaf anatomy and physiology to maximize this effect (Sage et al., 2012). The mitochondria of the bundle sheath cells became more numerous and larger, whereas the volume of the mesophyll cells was reduced. This stage can be observed in many

C₃-C₄ intermediate species of different genera such as *Flaveria* (Monson et al., 1984; Hylton et al., 1988) and *Heliotropium* (Muhaidat et al., 2011). Together with the previous steps, all required structural modifications for a functional C₄ cycle have been established at this step (Sage et al., 2012).

The final phase of C₄ evolution was the establishment of the C₄ cycle and the metabolic integration of mesophyll and bundle sheath cells to become one single metabolic system. Enzymes, such as PEPC and RubisCO, became compartmentalized into either mesophyll or bundle sheath cells. In the genus *Flaveria* two stages of intermediate species can be observed: those not having expressed PEPC, PDK and NADP-ME expressed in the C₄-like way, and those that already have these enzymes expressed in the C₄-like way (Ku et al., 1983; Ku et al., 1991). This indicates that the C₄ cycle became fully functional only after all other changes had occurred. To integrate the C₄ cycle, gene expression had to be changed in a large scale, mostly due to regulatory elements in the promoters that enhance expression and alter the spatial expression pattern (Sage et al., 2012). A fine-tuning in the interconnection of the C₄ cycle in the mesophyll cells and the C₃ cycle in the bundle sheath cells, such as the optimization of the kinetics of RubisCO, is the final step towards C₄ photosynthesis (Sage et al., 2012).

2.5 The genus *Flaveria* – a suitable model to study the evolution of C₄ photosynthesis

The genus *Flaveria* (Powell, 1978; McKown et al., 2005) is considered to be an ideal model to study the evolution of C₄ photosynthesis (Westhoff and Gowik, 2010; Gowik et al., 2011; Sage et al., 2012). It does not only contain true C₃ (e.g. *F. pringlei*) and true C₄ (e.g. *F. trinervia*) species, but also a broad range of intermediate species (Edwards and Ku, 1987; McKown et al., 2005; McKown and Dengler, 2007). These closely related species, expressing different stages of “C₄-ness”, make the genus *Flaveria* an excellent tool to study molecular mechanisms in the transition from C₃ plants to C₄ plants (e.g. (Bauwe and Chollet, 1986; Akyildiz et al., 2007; Furumoto et al., 2011)). Additionally the C₄ species *F. bidentis* can be stably transformed with *Agrobacterium tumefaciens*, thus facilitating transgenic studies directly within the genus (Chitty et al., 1994).

3 Glycine decarboxylase

3.1 Enzymology of glycine decarboxylase and its function in the photorespiratory C₂ cycle

All organisms possess glycine decarboxylase (GDC), since the enzyme is essential in connecting C₁, C₂ and C₃ metabolism (Oliver, 1994). Glycine decarboxylase consists of four subunits, namely P, T, L and H protein (Oliver et al., 1990) that together make up over one third of the soluble proteins of mitochondria of C₃ plants (Oliver and Raman, 1995). In plants, all four subunits are encoded in the nucleus and contain presequences that target them to the mitochondrial matrix (Oliver, 1994). Walker and Oliver (1986) also show that all genes are transcriptionally regulated in a light-dependent manner. The stoichiometric ratio of GDC subunits is assumed to be 4P : 27H : 9T : 2L, based on measurements of pea leaf mitochondria (Oliver, 1994). The P and the L proteins form dimers, the T protein is monomeric and the H protein can be mono- or dimeric (Oliver, 1994). While P, T and H proteins are unique to GDC, the L protein is also part of other enzyme complexes such as pyruvate dehydrogenase (Bourguignon et al., 1996). The P protein is the actual decarboxylating subunit of GDC and contains pyridoxal-5-phosphate (Walker and Oliver, 1986; Douce et al., 2001; Bauwe, 2011). The T protein is a tetrahydrofolate methylene transferase that shows similarities to formyltetrahydrofolate synthetases (Kopriva et al., 1995b). The L protein is a dihydrolipoamide dehydrogenase, and the H protein is a lipoamide-containing protein that does not catalyze a reaction itself, but works as a moveable substrate between the other subunits of GDC (Oliver et al., 1990; Douce et al., 2001).

Together with serine hydroxymethyltransferase (SHMT), GDC converts two molecules of glycine to one molecule of serine while consuming H₂O and NAD⁺, thereby releasing CO₂, NH₃, NADH and methyl tetrahydrofolate (CH₂-THF) (Oliver and Raman, 1995). Tetrahydrofolate (THF) is used as a carrier of the methyl group and is continuously cycling in the process (Figure 5) (Oliver and Raman, 1995), thus leading to the following net reaction:



(Douce et al., 2001)

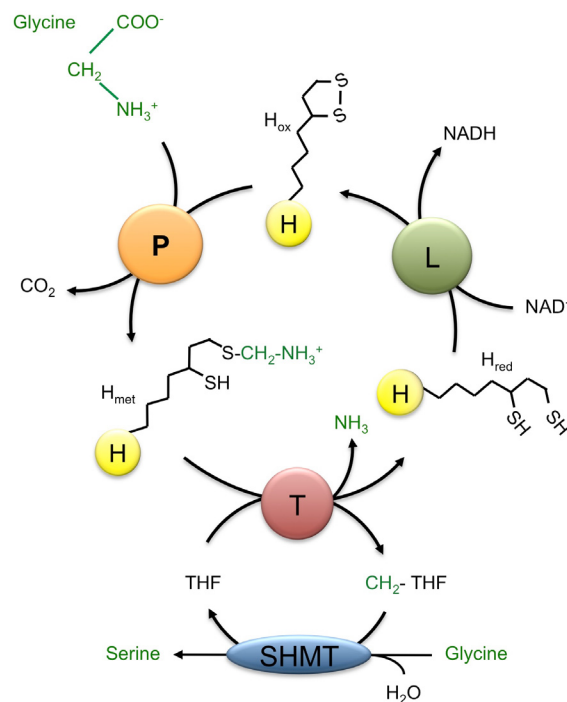


Figure 5. Schematic outlines of the reactions of glycine decarboxylase in mitochondria.

Glycine decarboxylase (GDC) contains four subunits P, L, T and H. The lipoamide-containing H protein undergoes a cycle of reductive methylamination that is catalysed by the P protein, methylamine transfer that is catalysed by the T protein and electron transfer that is catalysed by the L protein. Serine hydroxymethyltransferase (SHMT) takes part in the conversion of methyl tetrahydrofolate ($\text{CH}_2\text{-THF}$) to THF thereby producing serine while consuming a second molecule of glycine. NAD(H), nicotinamide adenine dinucleotide ox. (red.); H_{met} , H_{red} and H_{ox} , methylaminated, reduced and oxidized forms of the H protein. Adapted from Douce et al. (2001).

GDC in plants is not only involved in photorespiration but also in C_1 metabolism independent from photosynthesis, since serine and glycine are potential sources for C_1 compounds (Hanson and Roje, 2001). C_1 metabolism takes place in all tissues for the biosynthesis of purines, formylmethionyl-tRNA and thymidylate, as well as for the biosynthesis of several amino acids (Hanson and Roje, 2001). SHMT is not restricted to the mitochondria, since also a chloroplastic and a cytosolic variant exist (Besson et al., 1995). GDC – in contrast – is strictly localized to the mitochondria. Hence, all glycine formed in the cytosol has to be decarboxylated in this organelle (Mouillon et al., 1999). A serine-glycine shuttling cycle was therefore proposed to operate in all plant tissues. It connects the C_1 metabolism in the cytosol with the glycine metabolism performed by GDC in the mitochondria (Mouillon et al., 1999).

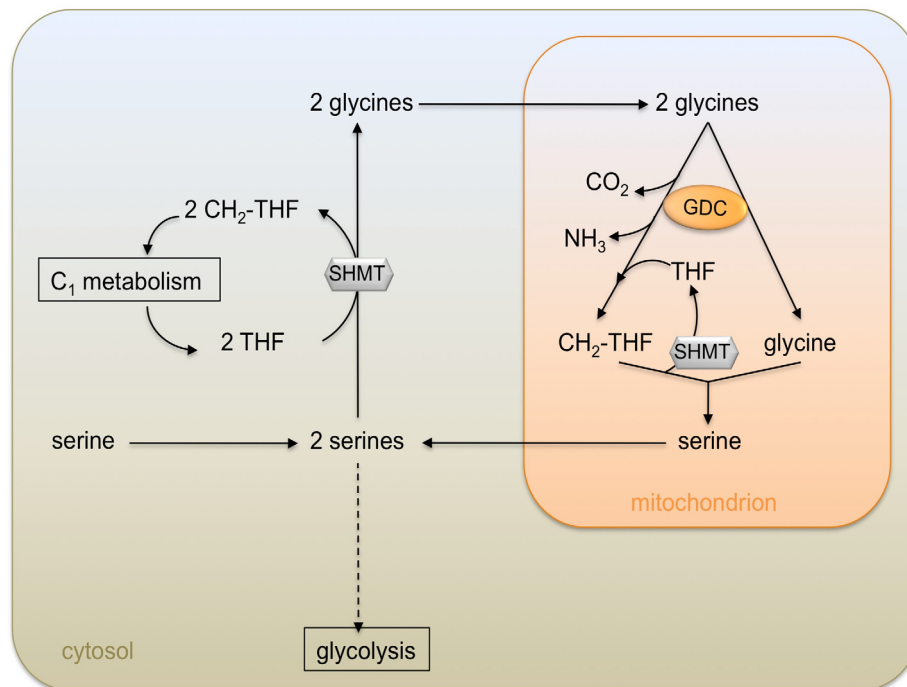


Figure 6. Diagram of the serine-glycine cycle between the cytosol and the mitochondrion connecting the C₁ metabolism in the cytosol with serine catabolism in the mitochondrion.

Cytosolic serine hydroxymethyltransferase (SHMT) converts serine into glycine in the cytosol. This glycine is shuttled into the mitochondrion where glycine decarboxylase (GDC) is located. The one-carbon compound is steadily recycled in the process and enters the C₁ metabolism in the cytosol. Adapted from Mouillon et al., (1999).

The fact that GDC is not only needed for photorespiration in plants was clearly shown by a double knock-out mutant of the C₃ plant *Arabidopsis thaliana* with no functional *GLDP* left. These plants were not viable even under elevated CO₂, thus non-photorespiratory conditions (Engel et al., 2007).

3.2 Glycine decarboxylase in the genus *Flaveria*

Glycine decarboxylase was intensively studied in the genus *Flaveria*. The P protein is encoded by a small gene family that contains three to five members (Kopriva and Bauwe, 1994; Bauwe et al., 1995; Bauwe and Kopriva, 1995; Chu, 1996) (see also chapter 2 manuscript 2). The H protein is also encoded by more than one gene in the C₃ species *F. pringlei* and *F. cronquistii*, whereas all other analysed species contained only one copy (Kopriva and Bauwe, 1995). The other two subunits, T and L, have not been further analysed in the genus *Flaveria*. It is assumed that the T protein is encoded by only one gene (Chu, 1996).

The expression of the genes encoding the four subunits is light-regulated in plants (Walker and Oliver, 1986). This up-regulation upon illumination was also directly demonstrated for the P, T and H protein in *F. anomala* and *F. pringlei* (Chu, 1996). Therefore, it is likely that all subunits of GDC in *Flaveria* are positively influenced by light.

In the C₄ species *F. trinervia*, all four subunits are located in the bundle sheath cells. This was proven with protein localisation studies, although minor amounts are still detectable in the mesophyll cells (Hylton et al., 1988; Morgan et al., 1993). These studies further showed that this predominant expression can already be found in the C₃-C₄ intermediate *F. linearis* (Morgan et al., 1993).

The bundle sheath localisation of the P protein in the C₄ species *F. trinervia* is dependent on the promoter of *GLDPA* gene that is sufficient to confer bundle sheath specific expression of a reporter gene both in transgenic *F. bidentis* (C₄) and *Arabidopsis thaliana* (C₃) plants (Engelmann et al., 2008). The second *GLDP* gene of *F. trinervia* is a pseudogene due to an insertion that interrupts the reading frame in the first exon (Cossu and Bauwe, 1998).

In *F. pringlei* all three genes encoding the H protein are active in leaves but show differences in their expression level in stems and particularly in roots (Kopriva and Bauwe, 1995). The C₄ species *F. trinervia* encodes only one H protein in its genome but organ specific splicing, a post-transcriptional regulatory mechanism, takes place (Kopriva et al., 1995a). One of these transcripts dominates in leaves and stems, while the other transcript accumulates predominately in roots (Kopriva et al., 1995a). Analyses of more *Flaveria* species ranging from C₃ through intermediates with different levels of “C₄-ness” up to C₄ species revealed that not only *F. trinervia* but also the other analysed advanced C₄ species contain both transcript variants (Kopriva et al., 1996). In contrast, none of the C₃ or intermediate species showed the alternative splicing observed in the C₄ species (Kopriva et al., 1996). Recent studies have shown that the alternative, C₄-specific variant of the H protein increases the activity of the P protein compared to the variant found in C₃, intermediate and C₄ species (Hasse et al., 2009).

Recent studies on the promoters of the T protein of the C₄ species *F. trinervia* and the C₃ species *F. pringlei* show that the gene of the C₄ species is expressed bundle sheath specifically, whereas the gene of the C₃ species is expressed ubiquitously in the leaf (Jan Emmerling, personal communication).

4. Regulation of gene expression

4.1 The transcriptional process and its complexity

The transcription of protein-coding genes by RNA polymerase II requires *cis*-regulatory sequences that are usually located upstream of the coding region of the genes and, besides RNA polymerase II, a set of general and gene-specific transcription factors and their auxiliary proteins (reviewed in Vedel and Scotti, 2011; Liu et al., 2013). The transcriptional controlling regions of protein-coding genes are complex and diverse (reviewed in Juven-Gershon et al., 2008). The sequences located 5' of the transcriptional start site (TSS) are called the promoter. Promoters are often separated into a core and a regulatory part (reviewed in Vedel and Scotti, 2011).

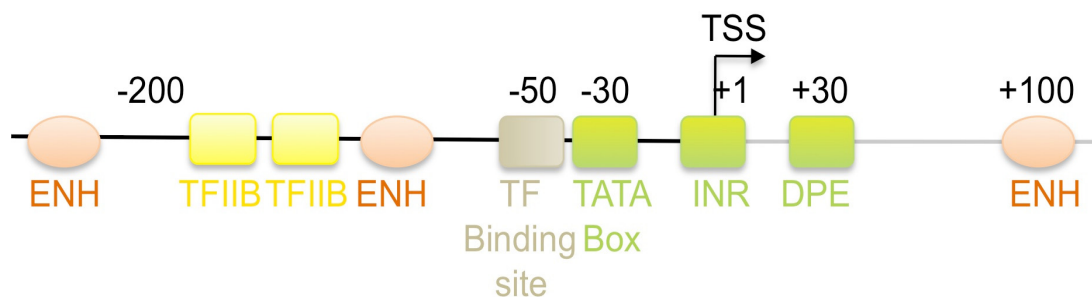


Figure 7. General structure of the transcriptional control region of protein-coding genes.

DPE, downstream promoter element; ENH, enhancer; INR, initiator motif; TF, transcription factor; TFIIB, transcription factor TFIIB binding site. The black line indicates the 5' flanking region of a gene and the grey line indicates the transcribed RNA. Adapted from Vedel and Scotti (2011).

Core promoters are diverse in eukaryotes. In general two types of promoters can be distinguished, those with a single or a few clearly defined TSSs, called focused promoters and those named dispersed promoters that possess no clearly defined TSS (reviewed in Juven-Gershon et al., 2008). In some promoters a TATA-box or a TATA-box derivative can be found that is located about 30 bp upstream of the TSS, but there are also TATA-less promoters known, in which the TSS is controlled by an initiator motif (Vedel and Scotti, 2011) (Figure 7). This initiator exhibits the consensus sequence PyPyAN [TA] PyPy with Py standing for pyrimidine (C or T) and N standing for any nucleotide (Yamamoto et al., 2007; Vedel and Scotti, 2011). Promoters with an initiator motif sometimes in addition contain a downstream promoter element that is located 30 basepairs upstream of the TSS and plays a

similar role as the TATA-box (reviewed in Vedel and Scotti, 2011) (Figure 7). Promoters with a TATA-box are more often focused than those without a TATA-box (Juven-Gershon et al., 2008).

The regulatory part of the promoter contains *cis*-regulatory sequences and their respective *trans*-acting factors (Figure 7). Transcription factors themselves are regulated in a temporal and spatial way, thus the occurrence or non-occurrence of a transcription factor might be essential for the transcription of one gene in a specific tissue (see e. g. Fuda et al., 2009; Ge et al., 2010; Ong and Corces, 2011; Yosef and Regev, 2011). In addition to promoters the correct transcription of protein-coding genes requires the action of enhancers that may be located (far) up- or downstream of the coding region of the gene or even in introns (reviewed in Vedel and Scotti, 2011) (Figure 7).

Further complexity can be found by the examination of the direction of the transcriptional process. It was found that many promoters are transcribed not only in the direction of the encoded gene but also in the opposite direction producing non-coding RNAs (reviewed in Jacquier, 2009; Wei et al., 2011). This bidirectional organisation is a common feature of coding genes and was observed in organisms ranging from bacteria to mammals (reviewed in Wei et al., 2011). Although RNA polymerase II seems to be able to start the transcription process in both directions equally, transcripts of the coding genes seem to be more abundant than those of the non-coding RNAs, thereby indicating a regulatory process after the initiation of transcription (reviewed in Wei et al., 2011).

Bidirectional transcription is not the only complexity of promoters. It is often the case that upstream of some genes not only one, but two promoters are located, either leading to alternative transcripts of the same gene or producing transcripts of an overlapping reading frame of a second gene. These promoters can be convergent, thus have overlapping transcripts, they can be oriented in tandem with one promoter upstream of the other, not always having overlapping transcripts (reviewed in Shearwin et al., 2005). The arrangement of coupled promoters interferes with the transcription of the corresponding transcripts by either blocking the binding of RNA polymerase II or by interrupting the transcriptional elongation (reviewed in Shearwin et al., 2005).

4.2 The connection between transcriptional and post-transcriptional regulation

During the transcription process, carried out by RNA polymerase II, a spliceosome complex binds to the pre-mRNA and removes any introns and connects the exons (reviewed in

Hoskins and Moore, 2012). During this process a so-called exon junction complex is deposited 20-24 nucleotides upstream of the exon-exon junction (Le Hir et al., 2000; Le Hir and Andersen, 2008). It was shown that this complex takes part in various processes during mRNA processing, including pre-mRNA splicing (Le Hir et al., 2000), mRNA export (reviewed in Reed, 2003) and localisation (reviewed in Tange et al., 2004), translation (Diem et al., 2007) and mRNAs stability (reviewed in Tange et al., 2004; Hwang and Kim, 2013).

After intron excision the ribosomes bind to the mRNA molecule and translate the codon triplets into amino acids, starting at the first AUG coding for a methionine and terminating at a stopcodon, mostly TAA in eukaryotes. Since translation is already initiated during the processing of mRNAs control mechanisms are necessary that regulate the accumulation of RNAs and the degradation of incorrect transcripts.

Degrading processes are initiated if mRNAs contain premature translation-termination codons that do not lead to functional proteins. These mRNAs are degraded through a set of different mechanisms (reviewed in Hwang and Kim, 2013). One mechanism that deals with the stability and degradation of mRNAs is the so-called “nonsense mediated mRNA decay” (NMD) (reviewed in Brogna and Wen, 2009; Hwang and Kim, 2013). The initiation of NMD is controlled in the very first round of translation (reviewed in Hwang and Kim, 2013). At this step the exon junction complexes are still associated to the mRNA and released when the ribosome reads through. In normal transcripts the termination codon is located downstream of the last exon junction complex, and thus, all complexes will be removed in this first round of translation. If a premature termination codon is located upstream of an exon junction complex, this complex will remain at the mRNA and thus, is an indicator for a false transcript (reviewed in Hwang and Kim, 2013). It was further shown that the existence of upstream open reading frames (uORFs) that are located in the 5' UTR (5' untranslated region), could cause NMD (Kertesz et al., 2006; Nyiko et al., 2009). These uORFs need to be of a minimal length of 50 nucleotides to interfere with the stability (Kertesz et al., 2006; Nyiko et al., 2009). Another fact that is important in NMD is the distance of the stop codon to the 3' UTR. Long distances between the stop codon and the polyadenylation site, independent if their origin is from uORFs or premature termination codons in the coding sequence, decrease the stability of mRNAs (Kertesz et al., 2006; Hori and Watanabe, 2007).

In some cases it was shown that mRNAs, even if they contain a premature termination codon, escape NMD. These mRNAs were found to undergo other mechanisms. One of them is called “nonsense mediated translational repression” that is dependent on the length of the 3' UTR and the existence of an intron downstream of the premature termination codon

(reviewed in Hwang and Kim, 2013). A second mechanism is called “nonsense associated alternative splicing” and leads to splice variants, in which the premature termination codon is skipped through the splicing process (reviewed in Hwang and Kim, 2013).

4.3 Molecular changes in regulatory mechanisms of C₄ genes leading to differential expression

The expression pattern of the genes involved in C₄ photosynthesis must have been changed during the transition from a C₃ to a C₄ species. The expression of many genes had to become either mesophyll or bundle sheath specific (see Figure 2). Different mechanisms are known to achieve this spatial distribution (reviewed in Hibberd and Covshoff, 2010). One mechanism is the exchange of *cis*-elements and/or the respective *trans*-acting factors. The transcription of the gene changed from ubiquitous to mesophyll or bundle sheath specific (see 4.3.1. for an example). A second mechanism is the establishment of differential accumulation of RNAs regulated at the post-transcriptional level, using cell specific RNA stability and/or cell specific degradation of RNAs (see 4.3.2. for an example). Further regulation can be found not only on the level of spatial expressions patterns, but also on the level of the intracellular localization of the resulting protein (see 4.3.3. for an example)

4.3.1 The mesophyll expression module 1 (MEM1) of phosphoenolpyruvate carboxylase (PEPC) in the genus *Flaveria*

The change of an already existent *cis*-element of a C₃ species in transition to a C₄ species that changes the expression from ubiquitously to mesophyll specifically is found in the 5' region of the *PEPC* gene in the genus *Flaveria*. *PEPC* is encoded by a small gene family that contains four distinct subclasses (*ppcA* to *ppcD*) in *Flaveria*, each of which show a different expression profile (Hermans and Westhoff, 1992). The *ppcA* gene was found to encode the C₄ isoform in the C₄ species *F. trinervia* (Hermans and Westhoff, 1992; Stockhaus et al., 1997). Promoter-reporter gene studies with the 5' flanking sequence of *ppcA* of the C₄ species *F. trinervia* and the C₃ species *F. pringlei* in transgenic *F. bidentis* (C₄) plants revealed that the promoter of *F. trinervia* drives expression of the reporter gene in the mesophyll cells exclusively. On the other hand the promoter of the C₃ plant *F. pringlei* drives the expression of the reporter gene to a much lower extent and firstly in cells associated with the vasculature and only later in the mesophyll (Stockhaus et al., 1997).

Yeast two-hybrid screens identified the CACT motif and its adjacent sequence as a possible binding site for transcription factors of the bZIP (basic leucine zipper) family (Akyildiz, 2007). These transcription factors do not seem to bind to the C₃ variant of sub-module B of the MEM1 (Akyildiz, 2007). Investigations of reporter gene constructs with the C₄ promoter lacking the CACT sequence led to the conclusion that this element alone is probably not the only *cis* regulatory element responsible for the mesophyll expression of the *ppcA* promoter (Gowik et al., 2004).

A further difference is located in the sub-module A (Figure 8). At the very first position of the sub-module a guanine (G) in the C₄ species and an adenine (A) in the C₃ species is found (Gowik et al., 2004). Additionally, only advanced C₄-like species inherit a G, whereas all other analysed intermediates inherit an A like the C₃ species (Gowik et al., 2004; Akyildiz et al., 2007).

The two sub-modules, A and B, are in direct connection in *F. trinervia* but in all other analysed *Flaveria* species about 100 nucleotides lay between the sub-modules (Gowik et al., 2004; Akyildiz et al., 2007). The sequence separating the sub-modules was analysed and no regulatory characteristics were found (Akyildiz et al., 2007). Both sub-modules need to be in the C₄ state to confer mesophyll specific expression of the reporter gene shown by the fact that transgenic lines with only one sub-module in the C₄ stage is not sufficient to assure mesophyll specific expression (Akyildiz et al., 2007).

4.3.2 Relocation of RubisCO to the bundle sheath cells through post-transcriptional RNA regulation in the genus *Flaveria*

RubisCO is exclusively found in the bundle sheath cells in C₄ plants and is absent from the mesophyll (see Figure 2). The mechanism of this restriction is an example of post-transcriptional regulation of cell specificity. Patel et al. (2006) show that in the C₄ species *Flaveria bidentis* the restriction of the small subunit of RubisCO to the bundle sheath cells is regulated by the 5' and the 3' untranslated regions (UTRs) of the gene (Patel et al., 2006). Both UTRs together were sufficient to drive bundle sheath specific reporter gene expression under a constitutive promoter in transgenic *F. bidentis* plants (Patel et al., 2006). RNA *in situ* hybridizations confirmed that the RNA of the reporter gene only accumulates in the bundle sheath cells although the constitutive promoter is used both in mesophyll and bundle sheath cells (Patel et al., 2006). It was concluded that the bundle sheath specific expression is

conferred by post-transcriptional degradation of the RNA in the mesophyll cells and that elements located in the UTRs lead to this degradation (Patel et al., 2006; Patel and Berry, 2008).

4.3.3 The relocation of carbonic anhydrase (CA) from the chloroplast to the cytosol in the genus *Flaveria*

Not only cell specificity but also relocation within the cell was necessary for some enzymes involved in C₄ photosynthesis, including carbonic anhydrase (CA). The C₄ plant *F. bidentis* was chosen for the analysis of the β -carbonic anhydrase (β -CA), which is the essential CA in photosynthesis (Cavallaro et al., 1994; Tetu et al., 2007). Three genes were found to encode isoforms of CA, named *CA1* (Cavallaro et al., 1994), *CA2* and *CA3* (Tetu et al., 2007). The expression of these three genes was determined on RNA level with quantitative RT-PCR and revealed that *CA3* transcripts showed the highest levels in leaves, thus CA3 was most likely the one CA important in photosynthesis (Tetu et al., 2007). A protein immunolocalization in leaf cross-sections showed that CA3 could be detected in the cytoplasm of the mesophyll cells exclusively (Tetu et al., 2007).

Other species of the genus *Flaveria* were also analysed for the presence of CA genes. The C₃ species *F. pringlei* was chosen for a detailed analysis (Tanz et al., 2009). In *F. pringlei* the same three β -CA genes as in *F. bidentis* were found, encoding *CA1*, *CA2* and *CA3* (Tanz et al., 2009). Phylogenetic analyses revealed that the CA genes of *F. bidentis* and *F. pringlei* cluster together and that *CA1* and *CA3* most likely share a common ancestral gene that was duplicated after the genus *Flaveria* diverged from other dicots but before *F. pringlei* and *F. bidentis* diverged from each other (Tanz et al., 2009). In contrast to the *F. bidentis* CA3, that of *F. pringlei* is located in the chloroplast and not in the cytoplasm (Tanz et al., 2009). Sequence comparisons between the *CA3* cDNAs from *F. bidentis* and *F. pringlei* showed that indeed the cDNA of the C₃ species contained a transit peptide to the chloroplast that was lost in the transition to C₄ photosynthesis and is thus no longer found in *F. bidentis* (Tanz et al., 2009). Studies of further *Flaveria* species showed that the transit peptide is lost only in the advanced C₄-like species *F. vaginata*, whereas all other analysed species, ranging from C₃ to C₃-C₄ intermediates to the C₄-like species *F. brownii*, still contained a transit peptide at the N-terminus of CA3 (Ludwig, 2011).

The *CA3* gene did not only need to be relocated from the chloroplast to the cytosol, but a mesophyll specific expression also had to be established. A MEM1-like structure was found in the 5' flanking region of the *CA3* gene in *F. bidentis* (Tanz et al., 2009; Ludwig, 2011). It has not yet been analysed if this MEM1-like element functions in a similar way as MEM1 of *ppcA*.

II Objectives

In plants, glycine decarboxylase (GDC) plays an essential role in photorespiration and the C₁ metabolism. While photorespiration is restricted to the bundle sheath cells in C₄ plants, C₁ metabolism is needed in all cells. In order to understand the restriction of GDC to the bundle sheath cells and the simultaneous maintenance of C₁ metabolism the P protein (*GLDP*) of GDC was analysed in the genus *Flaveria*.

(1) The 5' flanking region of the *GLDPA* gene of the C₄ species *Flaveria trinervia* was analysed in order to understand the underlying regulatory mechanism of a previously found bundle sheath specificity. "Rapid amplification of 5' complementary DNA ends" (5' RACE) was used to determine the transcriptional start site. RNAseq data were analysed to define the abundance of transcripts. Two splice variants deriving from the *GLDPA* promoter give rise to two versions of the GLPDA protein, containing either a full-length or a truncated mitochondrial targeting sequence. Promoter deletion and recombination *GUS* reporter gene constructs of the *GLDPA* 5' flanking region were analysed in transgenic *F. bidentis* (C₄) and *A. thaliana* (C₃) plants. (Manuscript 1: Wiludda et al., 2012).

(2) In order to understand the changes that *GLDP* underwent in the transition from C₃ to C₄ photosynthesis in the genus *Flaveria* and how the bundle sheath specificity was acquired, a phylogenetic analysis was conducted with different species, ranging from C₃ through intermediate species to C₄ species. The spatial expression behaviour of different members of the gene family was analysed with *GUS* reporter gene constructs both in transgenic *F. bidentis* (C₄) and *A. thaliana* (C₃). RNAseq data were consulted to gain insight into the expression levels and the abundance of both transcript variants in a total of nine members of the genus. 5' flanking regions of the C₃ species *F. pringlei* and *F. robusta* were compared to those of the orthologs of the C₄ species *F. trinervia*. It was analysed, whether the tandem promoter structure found for the 5' flanking sequence of *GLDPA* of *F. trinervia* is also applicable to orthologous genes in the other species. Consequently, the expression of both promoters were analysed in *GUS* reporter gene studies (Manuscript 2: Schulze et al., 2013).

III. Theses

(1) The regulatory mechanisms underlying the spatial expression pattern of the *GLDPA* promoter of the C₄ plant *F. trinervia* are quite complex. The 5' flanking sequence contains not only one, but two promoters oriented in tandem. Reporter gene studies show that both promoters exhibit a distinct spatial expression. The proximal promoter drives bundle sheath specific expression of the reporter gene, whereas the distal promoter drives reporter gene expression in all photosynthetic active leaf tissues. In uncoupled analyses the distal promoter shows higher reporter gene activity than the proximal promoter. Contrary to this, the output from the distal promoter is minimal in the context of the tandem promoter as RNAseq data show. Transcripts from the distal promoter undergo splicing in order to produce functional proteins. A post-transcriptional mechanism, probably involving “nonsense mediated mRNA decay”, regulates the amount of transcripts. The proximal promoter is most likely needed for photorespiratory GLDP, whereas the distal promoter is needed for the maintenance of GLDP for the C₁ metabolism (Manuscript 1: Wiludda et al., 2012)

(2) The establishment of a bundle sheath specific *GLDP* gene is analysed in the transition from C₃ to C₄ photosynthesis in the genus *Flaveria*. Phylogenetic analyses discovered a gene family with three distinct clusters. The *GLDP* genes of the C₃ species *F. pringlei* and *F. robusta*, clustering together with the bundle sheath specific *GLDPA* gene of *F. trinervia*, contain 5' flanking sequences that are sufficient to drive bundle sheath specific reporter gene expression. This shows that a bundle sheath specific *GLDP* was already present in C₃ species of *Flaveria*. In contrast, the second leaf-expressed group contains genes that are expressed ubiquitously in the case of C₃, but became a pseudogene in the C₄ species. This knock-off is a gradual process, shown by analyses of RNAseq data that comprised C₃-C₄ intermediates. In addition, the knock-down of the ubiquitously expressed *GLDP* in the intermediates leads to the establishment of a photorespiratory CO₂ pump. The tandem promoter structure of *GLDPA* of *F. trinervia* exists in orthologous genes in the C₃ species, but the distal promoter is cryptic and had to be activated in the transition to C₄. This activation is visible through RNAseq studies of C₃, intermediate and C₄ species. The activation is needed to maintain C₁ metabolism and had to be established prior of the deactivation of the ubiquitously expressed *GLDP* gene (Manuscript 2: Schulze et al., 2013).

IV.A Summary

C₄ photosynthesis is, in most cases, dependent of the spatial separation of the primary fixation of CO₂ in the form of bicarbonate by PEPC and the localisation of RubisCO. RubisCO as well as photorespiration is localised in the bundle sheath cells in C₄ plants. One of the key enzymes of photorespiration is glycine decarboxylase (GDC). GDC is further involved in C₁ metabolism that is essential in all cells of plants. A previous study of the gene for the P protein of GDC (*GLDPA*) of the C₄ plant *Flaveria trinervia* showed that the 5' flanking sequence is responsible for the bundle sheath specific expression of the *GLDPA* gene (Engelmann et al., 2008).

This study provides a detailed analysis of the *GLDPA* 5' flanking region with reporter gene studies and “rapid amplification of 5' complimentary DNA ends” (5' RACE). The analysis revealed that the 5' flanking region contains two promoters operating in tandem. The proximal promoter is responsible for the bundle sheath expression, whereas the distal promoter drives the expression of the reporter gene in all photosynthetic tissues. RNAseq experiments show that in the context of the complete promoter-composition the distal promoter produces only minor amounts of transcripts while the vast majority of transcripts derive from the proximal promoter. Regulation of transcript amounts is presumably regulated post-transcriptionally on the level of RNA stability. These results indicate that *GLDPA* is predominately expressed in the bundle sheath cells to fulfil photorespiration. Minor amounts of *GLDPA* are expressed in the other tissues to fulfil C₁ metabolism.

C₃ species of the genus *Flaveria* were chosen to elucidate the evolution of the bundle sheath specificity of GLDP. Phylogenetic analyses indicate that *GLDP* is encoded by a gene family, which clusters into three groups. Group I, II and III. Reporter gene studies of the 5' flanking sequences of group I genes (to which *GLDPA* of *F. trinervia* belongs) of two C₃ species show that these 5' flanking regions drive bundle sheath specific expression already. In contrast, group II genes drive reporter gene expression in all photosynthetic tissues of C₃ leaves, but in the C₄ species *F. trinervia*, the group II gene became a pseudogene. RNAseq analyses with *Flaveria* species, also containing intermediates, show that the knock-off was a gradual process and did not happen *ad hoc*. A further result is that the tandem promoter structure is already present in the group I *GLDP* genes of the genus, but no transcripts derive from this promoter, showing that this promoter is cryptic and is activated in the transition from C₃ to C₄ photosynthesis.

IV.B Zusammenfassung

In den meisten Fällen basiert C₄-Photosynthese auf der räumlichen Trennung der primären Fixierung von CO₂ in Form von Bicarbonat durch PEPC und der Lokalisierung der RubisCO. In C₄ Pflanzen ist RubisCO in den Bündelscheidenzellen lokalisiert, aus diesem Grund ist auch die Photorespiration auf die Bündelscheidenzellen reduziert. Eines der Hauptenzyme der Photorespiration ist die Glycindecarboxylase (GDC). GDC ist zusätzlich in den C₁ Stoffwechsel involviert, der in allen Zellen der Pflanze essentiell ist. Untersuchungen des P-Proteins der GDC in der C₄ Pflanze *Flaveria trinervia* (*GLDPA*) haben gezeigt, dass der 5'-flankierende Bereich für die Bündelscheiden-spezifische Expression dieses Gens verantwortlich ist (Engelmann et al., 2008).

In der vorliegenden Studie wurde eine detaillierte Analyse der 5'-flankierenden Region des *GLDPA* Gens der C₄ Pflanze *F. trinervia* durchgeführt. Hierzu wurden Verkürzungskonstrukte mit einem Reporter-Gen hergestellt und Transkripte mittels „rapid amplification of 5' complementary DNA ends“ (5' RACE) analysiert. Es konnte gezeigt werden, dass der 5'-flankierende Bereich nicht nur einen, sondern zwei Promotoren enthält, die in einer Tandem-Orientierung vorliegen. Der proximale Promoter ist für die Bündelscheiden-spezifische Expression des Reporter-Gens verantwortlich, der distale Promoter hingegen für eine Expression in allen photosynthetisch aktiven Geweben des Blattes. Mittels RNAseq-Experimenten konnte gezeigt werden, dass im Gesamtkontext des *GLDPA* Gens, wie er im Genom von *F. trinervia* vorliegt, nur ein Bruchteil der Transkripte vom distalen Promoter stammen und der Hauptteil der Transkripte vom proximalen Promoter ausgeht. Die Regulierung der Transkriptmenge des distalen Promoters wird höchstwahrscheinlich auf post-transkriptioneller Ebene durch die Stabilität der entstehenden RNAs reguliert. Die vorliegenden Resultate deuten darauf hin, dass *GLDPA* vorherrschend in den Bündelscheidenzellen exprimiert wird, in denen es Teil der Photorespiration ist und nur minimale Mengen in allen anderen Zelltypen exprimiert werden, um dort den C₁ Stoffwechsel aufrecht zu erhalten.

Um herauszufinden, wie sich die Bündelscheiden-spezifische Expression des *GLDP*-Gens im Laufe der Entwicklung der C₄-Photosynthese im Genus *Flaveria* entwickelt hat wurden C₃-Spezies hinzugezogen. Phylogenetische Analysen zeigten, dass es eine *GLDP*-Genfamilie gibt, die sich in drei distinkte Gruppen aufteilen lässt, diese wurden Gruppe I, II und III

genannt. Reporter-Gen Studien mit 5'-flankierenden Sequenzen der Gruppe I *GLDP* Gene, zu denen auch *GLDPA* von *F. trinervia*, gehört, zeigten, dass auch die 5'-flankierenden Regionen der C₃-Spezies Bündelscheiden-spezifische Expression des Reporter-Gens hervorrufen. Im Gegensatz dazu zeigten Konstrukte mit dem 5'-flankierenden Bereich des Gruppe II Gens der C₃-Spezies *F. pringlei* eine ubiquitäre Expression des Reporter-Gens. In der C₄-Spezies *F. trinervia* ist das Gruppe II Gen zu einem Pseudogen geworden. RNAseq Untersuchungen anhand von *Flaveria* Spezies, deren Photosynthesetyp von C₃- über intermediärer hin zu C₄-Photosynthese reichten, konnten zusätzlich zeigen, dass dieser Prozess graduell verlaufen und das Gruppe II Gen nicht abrupt abgeschaltet wurden ist. Die Tandem-Promoter-Struktur des *GLDPA*-Gens von *F. trinervia* innerhalb der Gruppe I-Gene ist konserviert und bereits in den C₃-Spezies zu finden. Allerdings wird der distale Promoter in den C₃-Spezies nicht abgelesen und ist somit kryptisch.

V. Literature

- Akyildiz, M.** (2007). Identification of *cis*- and *trans*-regulatory factors controlling the expression of the C₄ phosphoenolpyruvate carboxylase gene of the C₄ dicot *Flaveria trinervia*. [PhD thesis]. Heinrich Heine Universität, Düsseldorf.
- 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 C₄ dicot *Flaveria trinervia*. *Plant Cell* **19**, 3391-3402.
- Andersson, I.** (2008). Catalysis and regulation in Rubisco. *J Exp Bot* **59**, 1555-1568.
- Aubry, S., Brown, N.J., and Hibberd, J.M.** (2011). The role of proteins in C₃ plants prior to their recruitment into the C₄ pathway. *J Exp Bot* **62**, 3049-3059.
- Bauwe, H.** (2011). Photorespiration: The bridge to C₄ photosynthesis. In *C₄ Photosynthesis and Related CO₂ Concentrating Mechanisms*, A.S. Raghavendra and R.F. Sage, eds (Springer Netherlands), pp. 81-108
- Bauwe, H., and Chollet, R.** (1986). Kinetic properties of phosphoenolpyruvate carboxylase from C₃, C₄, and C₃-C₄ intermediate species of *Flaveria* (Asteraceae). *Plant Physiol* **82**, 695-699.
- Bauwe, H., and Kopriva, S.** (1995). The *gdcSP* gene from *Flaveria pringlei* (Asteraceae). *Plant Physiol* **107**, 655.
- Bauwe, H., Chu, C.C., Kopriva, S., and Nan, Q.** (1995). Structure and expression analysis of the *gdcSP* and *gdcPB* genes encoding two P-isoproteins of the glycine-cleavage system from *Flaveria pringlei*. *Eur J Biochem* **234**, 116-124.
- Besnard, G., Muasya, A.M., Russier, F., Roalson, E.H., Salamin, N., and Christin, P.-A.** (2009). Phylogenomics of C₄ photosynthesis in sedges (Cyperaceae): Multiple appearances and genetic convergence. *Mol Biol Evol* **26**, 1909-1919.
- Besson, V., Neuburger, M., Rebeille, F., and Douce, R.** (1995). Evidence for three serine hydroxymethyltransferases in green leaf cells. Purification and characterization of the mitochondrial and chloroplastic isoforms. *Plant Physiol Biochem* **33**, 665-673.
- Borland, A.M., Barrera Zambrano, V.A., Ceusters, J., and Shorrock, K.** (2011). The photosynthetic plasticity of crassulacean acid metabolism: an evolutionary innovation for sustainable productivity in a changing world. *New Phytol* **191**, 619-633.
- Bouchenak-Khelladi, Y., Anthony Verboom, G., Hodkinson, T.R., Salamin, N., Francois, O., Ní Chonghaile, G., and Savolainen, V.** (2009). The origins and diversification of C₄ grasses and savanna-adapted ungulates. *Glob Change Biol* **15**, 2397-2417.
- Bourguignon, J., Merand, V., Rawsthorne, S., Forest, E., and Douce, R.** (1996). Glycine decarboxylase and pyruvate dehydrogenase complexes share the same dihydrolipoamide dehydrogenase in pea leaf mitochondria: evidence from mass spectrometry and primary-structure analysis. *Biochem J* **313** (Pt 1), 229-234.

- Bowes, G., Ogren, W.L., and Hageman, R.H.** (1971). Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochem Biophys Res Commun* **45**, 716-722.
- Brodrribb, T.J., and Feild, T.S.** (2010). Leaf hydraulic evolution led a surge in leaf photosynthetic capacity during early angiosperm diversification. *Ecol Lett* **13**, 175-183.
- Brodrribb, T.J., Feild, T.S., and Jordan, G.J.** (2007). Leaf maximum photosynthetic rate and venation are linked by hydraulics. *Plant Physiol* **144**, 1890-1898.
- Brogna, S., and Wen, J.** (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat Struct Mol Biol* **16**, 107-113.
- Brown, N.J., Newell, C.A., Stanley, S., Chen, J.E., Perrin, A.J., Kajala, K., and Hibberd, J.M.** (2011). Independent and parallel recruitment of preexisting mechanisms underlying C₄ photosynthesis. *Science* **331**, 1436-1439.
- Calvin, M., and Benson, A.A.** (1948). The Path of carbon in photosynthesis. *Science* **107**, 476-480.
- Cavallaro, A., Ludwig, M., and Burnell, J.** (1994). The nucleotide sequence of a complementary DNA encoding *Flaveria bidentis* carbonic anhydrase. *FEBS Lett* **350**, 216-218.
- Chitty, J.A., Furbank, R.T., Marshall, J.S., Chen, Z., and Taylor, W.C.** (1994). Genetic transformation of the C₄ plant, *Flaveria bidentis*. *Plant J* **6**, 949-956.
- Christin, P.-A., and Besnard, G.** (2009). Two independent C₄ origins in Aristidoideae (Poaceae) revealed by the recruitment of distinct phosphoenolpyruvate carboxylase genes. *Am J Bot* **96**, 2234-2239.
- Christin, P.-A., Besnard, G., Samaritani, E., Duvall, M.R., Hodkinson, T.R., Savolainen, V., and Salamin, N.** (2008). Oligocene CO₂ decline promoted C₄ photosynthesis in grasses. *Curr Biol* **18**, 37-43.
- Christin, P.A., Osborne, C.P., Sage, R.F., Arakaki, M., and Edwards, E.J.** (2011). C₄ eudicots are not younger than C₄ monocots. *J Exp Bot* **62**, 3171-3181.
- Chu, C.** (1996). Molecular structure and expression patterns of glycine decarboxylase genes from *Flaveria pringlei* (C₃) and *Flaveria anomala* (C₃-C₄). [PhD thesis]. Martin-Luther University, Halle-Wittenberg.
- 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 C₄ plants. In C₄ plant biology, Sage R. F and R.K. Monson, eds (San Diego: Academic Press), pp. 133-172.
- Diem, M.D., Chan, C.C., Younis, I., and Dreyfuss, G.** (2007). PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. *Nat Struct Mol Biol* **14**, 1173-1179.

- Douce, R., Bourguignon, J., Neuburger, M., and Rébeillé, F.** (2001). The glycine decarboxylase system: a fascinating complex. *Trends Plant Sci* **6**, 167-176.
- Edwards, G.E., and Ku, M.** (1987). Biochemistry of C₃-C₄ intermediates. In *The biochemistry of plants*. Hatch M.D and Boardman N.K. , eds (San Diego: Academic Press), pp 275-325.
- Ehleringer, J.R., and Monson, R.K.** (1993). Evolutionary and ecological aspects of photosynthetic pathway variation. *Annu Rev Ecol Sys* **24**, 411-439.
- Ehleringer, J.R., Cerling, T.E., and Helliker, B.R.** (1997). C₄ photosynthesis, atmospheric CO₂, and climate. *Oecologia* **112**, 285-299.
- Ehleringer, J.R., Sage, R.F., Flanagan, L.B., and Pearcy, R.W.** (1991). Climate change and the evolution of C₄ photosynthesis. *Trends Ecol Evol* **6**, 95-99.
- Ellis, R.J.** (1979). The most abundant protein in the world. *Trends Biochem Sci* **4**, 241-244.
- 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 C₄ species *Flaveria trinervia*: analysis of transcriptional control in transgenic *Flaveria bidentis* (C₄) and *Arabidopsis* (C₃). *Plant Physiol* **146**, 1773-1785.
- Fuda, N.J., Ardehali, M.B., and Lis, J.T.** (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**, 186-192.
- Furbank, R.T.** (2011). Evolution of the C₄ photosynthetic mechanism: are there really three C₄ acid decarboxylation types? *J Exp Bot* **62**, 3103-3108.
- Furumoto, T., Hata, S., and Izui, K.** (1999). cDNA cloning and characterization of maize phosphoenolpyruvate carboxykinase, a bundle sheath cell-specific enzyme. *Plant Mol Biol* **41**, 301-311.
- Furumoto, T., Yamaguchi, T., Ohshima-Ichie, Y., Nakamura, M., Tsuchida-Iwata, Y., Shimamura, M., Ohnishi, J., Hata, S., Gowik, U., Westhoff, P., Brautigam, A., Weber, A.P., and Izui, K.** (2011). A plastidial sodium-dependent pyruvate transporter. *Nature* **476**, 472-475.
- Ge, X., Chang, F., and Ma, H.** (2010). Signaling and transcriptional control of reproductive development in *Arabidopsis*. *Curr Biol* **20**, R988-997.
- Giordano, M., Beardall, J., and Raven, J.A.** (2005). CO₂ concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. *Annu. Rev. Plant Biol* **56**, 99-131.

- González, M.-C., Sánchez, R., and Cejudo, F.** (2003). Abiotic stresses affecting water balance induce phosphoenolpyruvate carboxylase expression in roots of wheat seedlings. *Planta* **216**, 985-992.
- Gowik, U., and Westhoff, P.** (2011). The path from C₃ to C₄ photosynthesis. *Plant Physiol* **155**, 56-63.
- Gowik, U., Brautigam, A., Weber, K.L., Weber, A.P., and Westhoff, P.** (2011). Evolution of C₄ photosynthesis in the genus *Flaveria*: how many and which genes does it take to make C₄? *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 C₄ plant *Flaveria trinervia*, the promoter of the C₄ phosphoenolpyruvate carboxylase gene. *Plant Cell* **16**, 1077-1090.
- Grass Phylogeny Working, G., II.** (2012). New grass phylogeny resolves deep evolutionary relationships and discovers C₄ origins. *New Phytol* **193**, 304-312.
- Griffiths, H., Weller, G., Toy, L.F., and Dennis, R.J.** (2013). You're so vein: bundle sheath physiology, phylogeny and evolution in C₃ and C₄ plants. *Plant Cell Environ* **36**, 249-261.
- Haberlandt, G.** (1881). Vergleichende Anatomie des Gewebesystems bei Pflanzen. *Jahrbuch der Wissenschaftlichen Botanik* **13**, 74-188.
- Hagemann, M., Eisenhut, M., Hackenberg, C., and Bauwe, H.** (2010). Pathway and importance of photorespiratory 2-phosphoglycolate metabolism in cyanobacteria. *Adv Exp Med Biol* **675**, 91-108.
- Hanson, A.D., and Roje, S.** (2001). One-carbon metabolism in higher plants. *Annu Rev Plant Physiol Plant Mol Biol* **52**, 119-137.
- Hasse, D., Mikkat, S., Hagemann, M., and Bauwe, H.** (2009). Alternative splicing produces an H-protein with better substrate properties for the P-protein of glycine decarboxylase. *FEBS Journal* **276**, 6985-6991.
- Hatch, M., Kagawa, T., and Craig, S.** (1975). Subdivision of C₄-pathway species based on differing C₄ acid decarboxylating systems and ultrastructural features. *Aust J Plant Physiol* **2**, 111-128.
- Hatch, M.D.** (1987). C₄ photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochem Biophys Acta* **895**, 81-106.
- Hattersley, P.W.** (1984). Characterization of C₄ type leaf anatomy in grasses (Poaceae). Mesophyll: bundle sheath area ratios. *Ann Bot* **53**, 163-180.
- Hermans, J., and Westhoff, P.** (1992). Homologous genes for the C₄ isoform of phosphoenolpyruvate carboxylase in a C₃ and a C₄ *Flaveria* species. *Mol Gen Genet* **234**, 275-284.
- Hibberd, J.M., and Covshoff, S.** (2010). The regulation of gene expression required for C₄ photosynthesis. *Annu Rev Plant Biol* **61**, 181-207.

- 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.
- Hoskins, A.A., and Moore, M.J.** (2012). The spliceosome: a flexible, reversible macromolecular machine. *Trends Biochem Sci* **37**, 179-188.
- Hwang, J., and Kim, Y.K.** (2013). When a ribosome encounters a premature termination codon. *BMB Rep* **46**, 9-16.
- Hylton, C.M., Rawsthorne, S., Smith, A.M., Jones, D.A., and Woolhouse, H.W.** (1988). Glycine decarboxylase is confined to the bundle-sheath cells of leaves of C₃-C₄ intermediate species. *Planta* **175**, 452-459.
- Igamberdiev, A.U., and Gardeström, P.** (2003). Regulation of NAD- and NADP-dependent isocitrate dehydrogenases by reduction levels of pyridine nucleotides in mitochondria and cytosol of pea leaves. *Biochim Biophys Acta - Bioenergetics* **1606**, 117-125.
- Jacquier, A.** (2009). The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat Rev Genet* **10**, 833-844.
- Juven-Gershon, T., Hsu, J.-Y., Theisen, J.W.M., and Kadonaga, J.T.** (2008). The RNA polymerase II core promoter — the gateway to transcription. *Curr Opin Cell Biol* **20**, 253-259.
- Kadereit, G., and Freitag, H.** (2011). Molecular phylogeny of *Camphorosmeae* (Camphorosmoideae, Chenopodiaceae): Implications for biogeography, evolution of C₄-photosynthesis and taxonomy. *Taxon* **60**, 51-78.
- Kanai, R., and Edwards, G.E.** (1999). The biochemistry of C₄ photosynthesis. In *C₄ plant biology* Sage R.F. and Monson R.K. (eds). (San Diego:Academic press) 49-87.
- Kapralov, M.V., Kubien, D.S., Andersson, I., and Filatov, D.A.** (2011). Changes in Rubisco kinetics during the evolution of C₄ photosynthesis in *Flaveria* (Asteraceae) are associated with positive selection on genes encoding the enzyme. *Mol Biol Evol* **28**, 1491-1503.
- Kelly, G.J., and Latzko, E.** (1976). Inhibition of spinach-leaf phosphofructokinase by 2-phosphoglycollate. *FEBS Lett* **68**, 55-58.
- 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.
- Kinsman, E.A., and Pyke, K.A.** (1998). Bundle sheath cells and cell-specific plastid development in *Arabidopsis* leaves. *Development* **125**, 1815-1822.
- Kopriva, S., and Bauwe, H.** (1994). P-protein of glycine decarboxylase from *Flaveria pringlei*. *Plant Physiol* **104**, 1077-1078.
- Kopriva, S., and Bauwe, H.** (1995). H-protein of glycine decarboxylase is encoded by multigene families in *Flaveria pringlei* and *F. cronquistii* (Asteraceae). *Mol Gen Genet* **249**, 111-116.

- Kopriva, S., Cossu, R., and Bauwe, H.** (1995a). Alternative splicing results in two different transcripts for H-protein of the glycine cleavage system in the C₄ species *Flaveria trinervia*. *Plant J* **8**, 435-441.
- Kopriva, S., Chu, C.-c., and Bauwe, H.** (1996). H-protein of the glycine cleavage system in *Flaveria*: alternative splicing of the pre-mRNA occurs exclusively in advanced C₄ species of the genus. *Plant J* **10**, 369-373.
- Kopriva, S., Turner, S.R., Rawsthorne, S., and Bauwe, H.** (1995b). T-protein of the glycine decarboxylase multienzyme complex: evidence for partial similarity to formyltetrahydrofolate synthetase. *Plant Mol Biol* **27**, 1215-1220.
- Ku, M.S., Monson, R.K., Littlejohn, R.O., Nakamoto, H., Fisher, D.B., and Edwards, G.E.** (1983). Photosynthetic characteristics of C₃-C₄ intermediate *Flaveria* species : I. Leaf anatomy, photosynthetic responses to O₂ and CO₂, and activities of key enzymes in the C₃ and C₄ pathways. *Plant Physiol* **71**, 944-948.
- Ku, M.S., Wu, J., Dai, Z., Scott, R.A., Chu, C., and Edwards, G.E.** (1991). Photosynthetic and photorespiratory characteristics of *Flaveria* species. *Plant Physiol* **96**, 518-528.
- Kubien, D.S., Whitney, S.M., Moore, P.V., and Jesson, L.K.** (2008). The biochemistry of Rubisco in *Flaveria*. *J Exp Bot* **59**, 1767-1777.
- Langdale, J.A., Rothermel, B.A., and Nelson, T.** (1988). Cellular pattern of photosynthetic gene expression in developing maize leaves. *Genes Dev* **2**, 106-115.
- Le Hir, H., and Andersen, G.R.** (2008). Structural insights into the exon junction complex. *Curr Opin Struct Biol* **18**, 112-119.
- Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J.** (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J* **19**, 6860-6869.
- Liu, X., Bushnell, D.A., and Kornberg, R.D.** (2013). RNA polymerase II transcription: structure and mechanism. *Biochim Biophys Acta* **1829**, 2-8.
- Lorimer, G.H.** (1981). The carboxylation and oxygenation of ribulose 1,5-bisphosphate: The primary events in photosynthesis and photorespiration. *Annu Rev Plant Physiol* **32**, 349-382.
- Ludwig, M.** (2011). The molecular evolution of beta-carbonic anhydrase in *Flaveria*. *J Exp Bot* **62**, 3071-3081.
- Ludwig, M., and Burnell, J.** (1995). Molecular comparison of carbonic anhydrase from *Flaveria* species demonstrating different photosynthetic pathways. *Plant Mol Biol* **29**, 353-365.
- 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 C₃ to C₄ photosynthesis. *Plant J* **51**, 886-896.

- Masumoto, C., Miyazawa, S., Ohkawa, H., Fukuda, T., Taniguchi, Y., Murayama, S., Kusano, M., Saito, K., Fukayama, H., and Miyao, M.** (2010). Phosphoenolpyruvate carboxylase intrinsically located in the chloroplast of rice plays a crucial role in ammonium assimilation. *Proc Natl Acad Sci USA* **107**, 5226-5231.
- McKown, A.D., and Dengler, N.G.** (2007). Key innovations in the evolution of Kranz anatomy and C₄ vein pattern in *Flaveria* (Asteraceae). *Am J Bot* **94**, 382-399.
- McKown, A.D., Moncalvo, J.M., and Dengler, N.G.** (2005). Phylogeny of *Flaveria* (Asteraceae) and inference of C₄ photosynthesis evolution. *Am J Bot* **92**, 1911-1928.
- Miyao, M., and Fukayama, H.** (2003). Metabolic consequences of overproduction of phosphoenolpyruvate carboxylase in C₃ plants. *Arch Biochem Biophys* **414**, 197-203.
- Monson, R.K.** (2003). Gene duplication, neofunctionalization, and the evolution of C₄ photosynthesis. *Int J Plant Sci* **164**, S43-S54.
- Monson, R.K., and Moore, B.d.** (1989). On the significance of C₃-C₄ intermediate photosynthesis to the evolution of C₄ photosynthesis. *Plant Cell Environ* **12**, 689-699.
- Monson, R.K., Edwards, G.E., and Ku, M.S.B.** (1984). C₃-C₄ intermediate photosynthesis in plants. *BioScience*, 563-574.
- 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 C₃-C₄ intermediate species from different genera. *Planta* **190**, 468-473.
- Mouillon, J.-M., Aubert, S., Bourguignon, J., Gout, E., Douce, R., and Rébeillé, F.** (1999). Glycine and serine catabolism in non-photosynthetic higher plant cells: their role in C₁ metabolism. *Plant J* **20**, 197-205.
- Muhaidat, R., Sage, T.L., Frohlich, M.W., Dengler, N.G., and Sage, R.F.** (2011). Characterization of C₃-C₄ intermediate species in the genus *Heliotropium* L. (Boraginaceae): anatomy, ultrastructure and enzyme activity. *Plant Cell Environ* **34**, 1723-1736.
- Nyiko, T., Sonkoly, B., Merai, Z., Benkovics, A.H., and Silhavy, D.** (2009). Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. *Plant Mol Biol* **71**, 367-378.
- Ogren, W.L., and Bowes, G.** (1971). Ribulose diphosphate carboxylase regulates soybean photorespiration. *Nat New Biol* **230**, 159-160.
- Oliver, D.J.** (1994). The glycine decarboxylase complex from plant mitochondria. *Annu Rev Plant Physiol Plant Mol Biol* **45**, 323-337.
- 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.

- Oliver, D.J., Neuburger, M., Bourguignon, J., and Douce, R.** (1990). Interaction between the component enzymes of the glycine decarboxylase multienzyme complex. *Plant Physiol* **94**, 833-839.
- Ong, C.T., and Corces, V.G.** (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet* **12**, 283-293.
- Patel, M., and Berry, J.O.** (2008). Rubisco gene expression in C₄ plants. *J Exp Bot* **59**, 1625-1634.
- Patel, M., Siegel, A.J., and Berry, J.O.** (2006). Untranslated regions of *FbRbcS1* mRNA mediate bundle sheath cell-specific gene expression in leaves of a C₄ plant. *J Biol Chem* **281**, 25485-25491.
- Powell, A.** (1978). Systematics of *Flaveria* (Flaveriinae-Asteraceae). *Ann Mo Bot Gard* **65**, 590-636.
- Price, G.D., Badger, M.R., Woodger, F.J., and Long, B.M.** (2008). Advances in understanding the cyanobacterial CO₂-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. *J Exp Bot* **59**, 1441-1461.
- Rachmilevitch, S., Cousins, A.B., and Bloom, A.J.** (2004). Nitrate assimilation in plant shoots depends on photorespiration. *Proc Natl Acad Sci USA* **101**, 11506-11510.
- Rawsthorne, S., Hylton, C.M., Smith, A.M., and Woolhouse, H.W.** (1988). Photorespiratory metabolism and immunogold localization of photorespiratory enzymes in leaves of C₃ and C₃-C₄ intermediate species of *Moricandia*. *Planta* **173**, 298-308.
- Reed, R.** (2003). Coupling transcription, splicing and mRNA export. *Curr Opin Cell Biol* **15**, 326-331.
- Roalson, E.H., Hinchliff, C.E., Trevisan, R., and da Silva, C.R.M.** (2010). Phylogenetic relationships in *Eleocharis* (Cyperaceae): C₄ photosynthesis origins and patterns of diversification in the spikerushes. *Syst Bot* **35**, 257-271.
- Sage, R.F.** (1999a). The taxonomic distribution of C₄ photosynthesis. In *C₄ plant biology* Sage R.F. and Monson R.K. (eds). (San Diego:Academic press) 551-584.
- Sage, R.F.** (1999b). Why C₄ photosynthesis? .In *C₄ plant biology* Sage R.F. and Monson R.K. (eds). (San Diego:Academic press) 3-16.
- Sage, R.F.** (2004). The evolution of C₄ photosynthesis. *New Phytol* **161**, 341-370.
- Sage, R.F., Christin, P.A., and Edwards, E.J.** (2011). The C₄ 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 C₄ photosynthesis. *Annu Rev Plant Biol* **63**, 19-47.
- Shearwin, K.E., Callen, B.P., and Egan, J.B.** (2005). Transcriptional interference--a crash course. *Trends Genet* **21**, 339-345.

- Sheen, J.** (1999). C₄ gene expression. *Annu Rev Plant Physiol Plant Mol Biol* **50**, 187-217.
- Sheen, J.Y., and Bogorad, L.** (1987). Differential expression of C₄ pathway genes in mesophyll and bundle sheath cells of greening maize leaves. *J Biol Chem* **262**, 11726-11730.
- Spreitzer, R.J., and Salvucci, M.E.** (2002). Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu Rev Plant Biol* **53**, 449-475.
- Stabenau, H., and Winkler, U.** (2005). Glycolate metabolism in green algae. *Physiol Plant* **123**, 235-245.
- Stockhaus, J., Schlue, U., Koczor, M., Chitty, J.A., Taylor, W.C., and Westhoff, P.** (1997). The promoter of the gene encoding the C₄ form of phosphoenolpyruvate carboxylase directs mesophyll-specific expression in transgenic C₄ *Flaveria* spp. *Plant Cell* **9**, 479-489.
- Svensson, P., Bläsing, O.E., and Westhoff, P.** (1997). Evolution of the enzymatic characteristics of C₄ phosphoenol pyruvate carboxylase. *Euro J Biochem* **246**, 452-460.
- Tange, T.Ø., Nott, A., and Moore, M.J.** (2004). The ever-increasing complexities of the exon junction complex. *Curr Opin Cell Biol* **16**, 279-284.
- Tanz, S.K., Tetu, S.G., Vella, N.G., and Ludwig, M.** (2009). Loss of the transit peptide and an increase in gene expression of an ancestral chloroplastic carbonic anhydrase were instrumental in the evolution of the cytosolic C₄ carbonic anhydrase in *Flaveria*. *Plant Physiol* **150**, 1515-1529.
- Tetu, S.G., Tanz, S.K., Vella, N., Burnell, J.N., and Ludwig, M.** (2007). The *Flaveria bidentis* β-carbonic anhydrase gene family encodes cytosolic and chloroplastic isoforms demonstrating distinct organ-specific expression patterns. *Plant Physiol* **144**, 1316-1327.
- Vedel, V., and Scotti, I.** (2011). Promoting the promoter. *Plant Science* **180**, 182-189.
- Vicentini, A., Barber, J.C., Aliscioni, S.S., Giussani, L.M., and Kellogg, E.A.** (2008). The age of the grasses and clusters of origins of C₄ photosynthesis. *Glob Change Biol* **14**, 2963-2977.
- von Caemmerer, S., and Furbank, R.T.** (2003). The C₄ pathway: an efficient CO₂ pump. *Photosynth Res* **77**, 191-207.
- Walker, J.L., and Oliver, D.J.** (1986). Glycine decarboxylase multienzyme complex. Purification and partial characterization from pea leaf mitochondria. *J Biol Chem* **261**, 2214-2221.
- Wei, W., Pelechano, V., Jarvelin, A.I., and Steinmetz, L.M.** (2011). Functional consequences of bidirectional promoters. *Trends Genet* **27**, 267-276.
- West-Eberhard, M.J., Smith, J.A.C., and Winter, K.** (2011). Photosynthesis, reorganized. *Science* **332**, 311-312.

- Westhoff, P., and Gowik, U.** (2010). Evolution of C₄ photosynthesis--looking for the master switch. *Plant Physiol* **154**, 598-601.
- Whitney, S.M., Houtz, R.L., and Alonso, H.** (2011). Advancing our understanding and capacity to engineer nature's CO₂-sequestering enzyme, Rubisco. *Plant Physiol* **155**, 27-35.
- Windhövel, A., Hein, I., Dabrowa, R., and Stockhaus, J.** (2001). Characterization of a novel class of plant homeodomain proteins that bind to the C₄ phosphoenolpyruvate carboxylase gene of *Flaveria trinervia*. *Plant Mol Biol* **45**, 201-214.
- Wingler, A., Walker, R.P., Chen, Z.H., and Leegood, R.C.** (1999). Phosphoenolpyruvate carboxykinase is involved in the decarboxylation of aspartate in the bundle sheath of maize. *Plant Physiol* **120**, 539-546.
- Wyrich, R., Dressen, U., Brockmann, S., Streubel, M., Chang, C., Qiang, D., Paterson, A., and Westhoff, P.** (1998). The molecular basis of C₄ photosynthesis in sorghum: isolation, characterization and RFLP mapping of mesophyll- and bundle-sheath-specific cDNAs obtained by differential screening. *Plant Mol Biol* **37**, 319 - 335.
- Yamamoto, Y.Y., Ichida, H., Matsui, M., Obokata, J., Sakurai, T., Satou, M., Seki, M., Shinozaki, K., and Abe, T.** (2007). Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genomics* **8**, 67.
- Yosef, N., and Regev, A.** (2011). Impulse control: temporal dynamics in gene transcription. *Cell* **144**, 886-896.
- Yoshimura, Y., Kubota, F., and Ueno, O.** (2004). Structural and biochemical bases of photorespiration in C₄ plants: quantification of organelles and glycine decarboxylase. *Planta* **220**, 307-317.
- Zelitch, I., Schultes, N.P., Peterson, R.B., Brown, P., and Brutnell, T.P.** (2009). High glycolate oxidase activity is required for survival of maize in normal air. *Plant Physiol* **149**, 195-204.

VI. Chapters

1. Manuscript 1: Christian Wiludda, Stefanie Schulze, Udo Gowik, Sascha Engelmann, Maria Koczor, Monika Streubel, Hermann Bauwe, Peter Westhoff (2012). **Regulation of the photorespiratory *GLDPA* gene in *C₄ Flaveria* – an intricate interplay of transcriptional and post- transcriptional processes.** *The Plant Cell* 24: 137-151
2. Manuscript 2: Stefanie Schulze, Julia Mallmann, Janet Burscheidt, Maria Koczor, Monika Streubel, Hermann Bauwe, Udo Gowik, Peter Westhoff (2013). **Evolution of *C₄* photosynthesis in the genus *Flaveria* – establishment of a photorespiratory *CO₂* pump.** Submitted for publication to *The Plant Cell*.
3. **Function and origin of the distal promoter (P_{R2}) of the *GLDP* genes in *Flaveria*.**
4. **Identification of *cis*- regulatory elements for the functioning of the *GLDPA* 5' flanking region.**
5. **The roles of splicing and upstream open reading frames on the expression of *GLDPA* of *F. trinervia*.**

Chapter 1

Manuscript 1

Regulation of the photorespiratory *GLDPA* gene in C_4 *Flaveria* – an intricate interplay of transcriptional and post- transcriptional processes

The Plant Cell, Vol. 24: 137–151, January 2012, www.plantcell.org © 2012 American Society of Plant Biologists. All rights reserved.

Regulation of the Photorespiratory *GLDPA* Gene in *C₄ Flaveria*: An Intricate Interplay of Transcriptional and Posttranscriptional Processes ^W

Christian Wiludda,^a Stefanie Schulze,^a Udo Gowik,^a Sascha Engelmann,^a Maria Koczor,^a Monika Streubel,^a Hermann Bauwe,^b and Peter Westhoff^{a,1}

^aHeinrich-Heine-Universität Düsseldorf, Institut für Entwicklungs- und Molekularbiologie der Pflanzen, 40225 Duesseldorf, Germany

^bUniversität Rostock, Abteilung Pflanzenphysiologie, 18059 Rostock, Germany

The mitochondrial Gly decarboxylase complex (GDC) is a key component of the photorespiratory pathway that occurs in all photosynthetically active tissues of *C₃* plants but is restricted to bundle sheath cells in *C₄* species. GDC is also required for general cellular *C₁* metabolism. In the Asteracean *C₄* species *Flaveria trinervia*, a single functional *GLDP* gene, *GLDPA*, encodes the P-subunit of GDC, a decarboxylating Gly dehydrogenase. *GLDPA* promoter reporter gene fusion studies revealed that this promoter is active in bundle sheath cells and the vasculature of transgenic *Flaveria bidentis* (*C₄*) and the Brassicacean *C₃* species *Arabidopsis thaliana*, suggesting the existence of an evolutionarily conserved gene regulatory system in the bundle sheath. Here, we demonstrate that *GLDPA* gene regulation is achieved by an intricate interplay of transcriptional and posttranscriptional mechanisms. The *GLDPA* promoter is composed of two tandem promoters, *P_{R2}* and *P_{R7}*, that together ensure a strong bundle sheath expression. While the proximal promoter (*P_{R7}*) is active in the bundle sheath and vasculature, the distal promoter (*P_{R2}*) drives uniform expression in all leaf chlorenchyma cells and the vasculature. An intron in the 5' untranslated leader of *P_{R2}*-derived transcripts is inefficiently spliced and apparently suppresses the output of *P_{R2}* by eliciting RNA decay.

INTRODUCTION

C₄ photosynthesis is based on the division of labor between two distinct photosynthetically active cell types: mesophyll and bundle sheath cells. After conversion to HCO_3^- , CO_2 is initially fixed in mesophyll cells by phosphoenolpyruvate carboxylase in the form of either malate or Asp and then transported into bundle sheath cells. There CO_2 is released, refixed by ribulose-1,5-bis-phosphate carboxylase/oxygenase (Rubisco), and finally enters the Calvin-Benson cycle as it occurs in *C₃* plants. As a bifunctional enzyme, Rubisco is able to catalyze the carboxylation as well as the oxygenation of its substrate ribulose 1,5-bisphosphate. The fixation of O_2 leads to accumulation of phosphoglycolate, which is toxic for plant cells. To regenerate phosphoglycerate from phosphoglycolate, photorespiration is essential. However, this metabolic pathway leads to the loss of previously fixed CO_2 and thus decreases the efficiency of photosynthesis. The high concentration of CO_2 in the bundle sheath cells, caused by the *C₄* cycle, suppresses the oxygenase reaction of Rubisco and, thereby, photorespiration effectively (Anderson, 1971; Ogren, 1984; Hatch, 1987; Leegood et al., 1995; Foyer et al., 2009).

Nevertheless, photorespiratory processes are not completely repressed in *C₄* species (Yoshimura et al., 2004; Zelitch et al., 2009; Bauwe et al., 2010).

C₄ plants have evolved at least 62 times independently from *C₃* ancestors (Sage et al., 2011), indicating that the conversion from *C₃* toward *C₄* photosynthesis did not require drastic alterations but could have been implemented rather easily in genetic terms (Sage, 2004; Brown et al., 2011; Gowik and Westhoff, 2011). It is assumed that a crucial step toward *C₄* photosynthesis was the establishment of a functional photorespiratory CO_2 pump, which required the restriction of the Gly decarboxylase complex (GDC) to the bundle sheath cells (Bauwe and Kolukisaoglu, 2003; Sage, 2004). In *C₃* plants, GDC accumulates in all photosynthetically active cells. By contrast, in *C₄* plants, GDC occurs only in bundle sheath but not in mesophyll cells, and consequently photorespiratory activity of *C₄* plants is restricted to the bundle sheath. In *C₃-C₄* intermediate species, which are considered an evolutionary link in the transition from *C₃* to *C₄* plants, GDC activity is restricted to the bundle sheath cells (Ohnishi and Kanai, 1983; Rawsthorne et al., 1988; Morgan et al., 1993; Yoshimura et al., 2004). The genus *Flaveria* includes *C₄*, *C₃*, and several *C₃-C₄* intermediate species (Powell, 1978; McKown et al., 2005) and therefore represents a well-studied model system to study molecular mechanisms of the evolutionary transition from *C₃* to *C₄* photosynthesis (Westhoff and Gowik, 2004; Brown et al., 2005). GDC is located in the mitochondria and consists of four subunits, the L-, H-, P-, and T-proteins. Together, the four proteins cleave Gly, resulting in the release of CO_2 , NH_3 , and a

¹Address 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).

^WOnline version contains Web-only data.

www.plantcell.org/cgi/doi/10.1105/tpc.111.093872

tetrahydrofolate-bound C₁ residue (Oliver, 1994; Douce et al., 2001). Aside from its involvement in the photorespiratory pathway, GDC also contributes to the C₁ metabolism in all biosynthetic tissues that is essential for the synthesis of proteins, nucleic acids, pantothenates, and methylated molecules (Mouillon et al., 1999; Hanson and Roje, 2001).

The *GLDPA* gene encodes the P-subunit of GDC in the C₄ plant *Flaveria trinervia* (Cossu and Bauwe, 1998). In situ hybridization studies showed that *GLDPA* transcripts accumulate only in bundle sheath cells, suggesting that *GLDPA* is specifically transcribed in this tissue (Engelmann et al., 2008). In agreement with this finding, the *GLDPA* 5'-flanking region from -1 to -1571 (with regard to the translational start site at +1; here referred to as the *GLDPA* promoter) directs expression of reporter genes in the bundle sheath but not in the mesophyll cells. In *Flaveria bidentis* (C₄), the promoter is also active, although to a varying degree, in the vasculature of leaves and roots, stomata, and in the pericycle cells of roots (Figure 1). Surprisingly, the *GLDPA* promoter exhibited a similar activity in *Arabidopsis thaliana* (C₃) (Engelmann et al., 2008; Figure 1), suggesting that the regulatory networks controlling bundle sheath gene expression are similar in the Brassicacean C₃ species *Arabidopsis* and the Asteracean C₄ species *F. bidentis*. Promoter deletion and recombination experiments identified a distal region that enhanced promoter activity and an intermediate segment that appeared to contain mesophyll-repressing sequences (Engelmann et al., 2008).

In this study, we demonstrate that the functional architecture of the *GLDPA* promoter is much more complex than previously thought. We show that the *GLDPA* promoter is in fact composed of two subpromoters acting in tandem to ensure strong bundle sheath and sparse mesophyll expression of the gene. We discuss why the bundle sheath-exclusive expression of a single leaf-specific *GLDP* gene, as predicted and requested by the functional model of C₄ photosynthesis, must allow limited expression in the mesophyll.

RESULTS

Analysis of 5' Ends of Transcripts of the *GLDPA* Gene of *F. trinervia* Revealed Two Independent Transcription Start Sites

The 1571-bp *GLDPA* 5'-flanking region was subdivided into seven segments with region 1 being most distal and region 7 being most proximal to the translational start site (Engelmann et al., 2008; Figure 2A). Since the transcription start site (TSS) of the *GLDPA* gene of *F. trinervia* had not been determined experimentally, rapid amplification of 5' cDNA ends (5'-RACE; Frohman et al., 1988) was used for mapping 5' ends of *GLDPA* transcripts as present in total leaf extracts. The 5'-RACE analysis revealed two RNA 5' end classes with one starting in the most proximal region 7 predominantly at

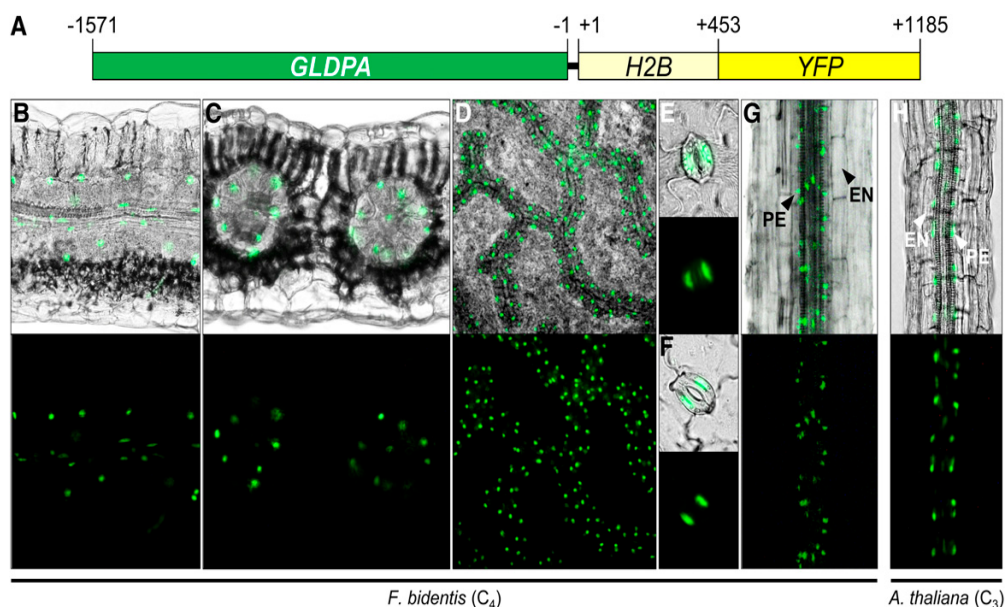


Figure 1. Fluorescence Microscopy Analysis of *GLDPA* Promoter Activity in Transgenic *F. bidentis* and *Arabidopsis*.

(A) Schematic presentation of the *GLDPA*-*H2B*-*YFP* construct that was transformed into *F. bidentis* or *Arabidopsis* to express YFP fused to histone 2B (*H2B*) under the control of the *GLDPA* promoter. The *H2B*-*YFP* fusion protein is retained in the nucleus of the expressing cell, which prevents any diffusion of the reporter protein into adjacent cells (Boisnard-Lorig et al., 2001).

(B) to (H) The localization of YFP was examined by fluorescence microscopy in longitudinal (B) and cross sections (C) of *F. bidentis* leaves, whole *F. bidentis* leaf blades in top view (D), guard cells of both the upper (E) and lower epidermis (F) of *F. bidentis*, and in roots of *F. bidentis* (G) and *Arabidopsis* (H). The fluorescence image is displayed underneath, and the corresponding merge of the fluorescent signal and the bright-field picture above. In the root, single endodermis (EN) and pericycle (PE) cells are indicated by an arrowhead.

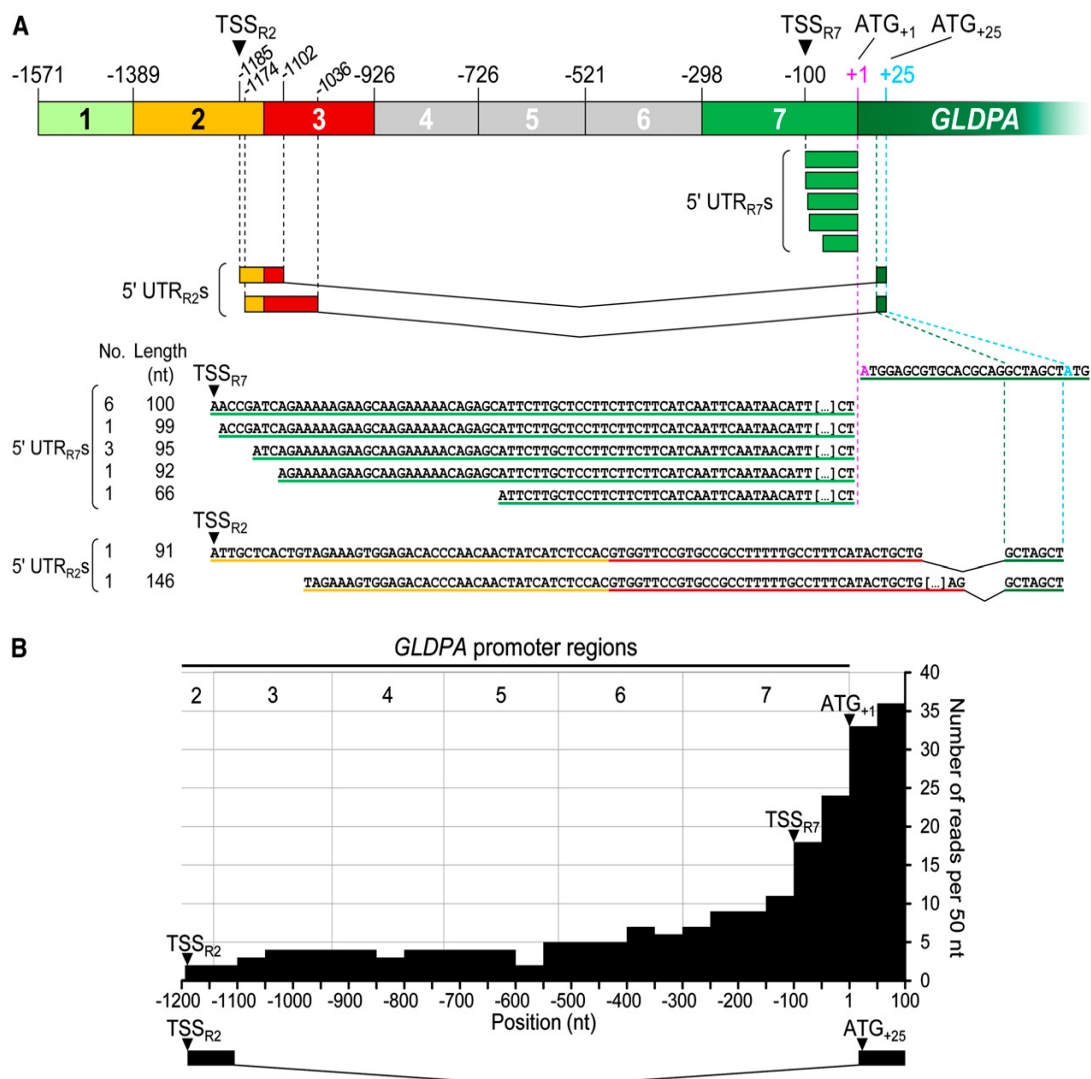


Figure 2. Analysis of Transcript 5' Ends of the Endogenous *GLDPA* Gene of *F. trinervia*.

(A) The *GLDPA* promoter and its transcriptional output based on 5'-RACE. The dissection of the *GLDPA* promoter into seven regions has been described by Engelmann et al. (2008). The schematic structure of the 5' UTRs of the two types of RNAs originating from region 2 (5' UTR_{R2s}) or region 7 (5' UTR_{R7s}) and their corresponding cDNA sequences are depicted below the schematic drawing of the *GLDPA* promoter. The TSSs within region 2 (TSS_{R2}) and 7 (TSS_{R7}) are indicated as well as the start codons used when transcription starts from region 2 (ATG₊₂₅) or region 7 (ATG₊₁) and the number (No.) of the 5' UTRs detected for each 5' UTR variant whose length is stated in nucleotides (nt).

(B) Diagram showing the read coverage of the *GLDPA* contig derived from 454 sequencing reads (Gowik et al., 2011). The coverage upstream of the translational start site (ATG₊₁) up to 100 nucleotides downstream was analyzed in 50-nucleotide windows. A contig corresponding to the 91-nucleotide spliced variant starting from TSS_{R2} that was detected by 5'-RACE **(A)** was represented by only two 454 reads. The TSSs (TSS_{R2} and TSS_{R7}) and the start codons (ATG₊₁ and ATG₊₂₅) are marked by arrowheads. The different *GLDPA* promoter regions 2 to 7 shown as columns are allocated to their respective positions.

nucleotide -100 upstream of the predicted translational start codon at +1 (ATG₊₁) and the other starting in the distal region 2 between nucleotides -1185 and -1174 (Figure 2A). Half of the analyzed transcripts starting from region 7 contained a 5' untranslated region (5' UTR_{R7}) of 100 nucleotides. The remaining 5' UTR_{R7s} were slightly shorter with a length between 66 and 99 nucleotides. The UTRs of the two detected 5' ends of RNAs transcribed from region 2 (5' UTR_{R2s}) included parts of region 2

and 3 but lacked regions 4, 5, 6, and 7 as well as the first 17 nucleotides of the predicted *GLDPA* open reading frame (Figure 2A). Fourteen individual and randomly selected 5'-RACE products were sequenced. Twelve of them started in region 7 and only two in region 2 (Figure 2A), indicating that the dominant TSS is that one located in region 7.

The comparison of the 5' UTR_{R2s} with the DNA sequence of the *GLDPA* promoter identified the signatures of a spliceosomal

intron with two putative GT splice donor sites at -1103 and -1037 within region 3 and a shared AG splice acceptor site at $+16$ within the open reading frame. If splicing occurs, regardless of which donor site is used, the next available putative start codon at position $+25$ (ATG₊₂₅; Figure 2A) could be used, resulting in the shortening of the mitochondrial GLDPA presequence by eight amino acids.

The analysis of the leaf transcriptome of *F. trinervia* by 454 pyrosequencing confirmed the 5'-RACE data (Figure 2B; Gowik et al., 2011). The most distal reads detected for the *GLDPA* gene started exactly at position -1185 within region 2. Additionally, the 91-nucleotide 5' UTR_{R2} splicing variant starting at -1185 (Figure 2A) was also found twice by 454 sequencing. In contrast with the low abundance of transcripts in the range from -1185 to -100 , the frequency of mRNAs increased at or downstream of position -100 within region 7. This shows that region 7 is transcriptionally more active than region 2, which is consistent with the results obtained by 5'-RACE. We conclude that the *GLDPA* promoter contains two putative TSSs with the major and proximal TSS located in region 7 predominantly at position -100 (TSS_{R7}) and the distal and minor TSS in region 2 around position -1185 (TSS_{R2}).

The Proximal and Distal TSSs of the *F. trinervia* *GLDPA* Gene Are Functional in Transgenic *F. bidentis* and *Arabidopsis*

To test whether the two putative TSSs are used in a transgenic promoter-reporter gene context, 5'-RACE experiments were performed with transgenic *F. bidentis* and *Arabidopsis* both containing the *GLDPA*-Ft: β -glucuronidase (*GUS*) chimeric gene (see Supplemental Figure 1 online; Engelmann et al., 2008). In both *F. bidentis* and *Arabidopsis*, all RNA 5' ends started between position -90 and -100 (i.e., in region 7). Transcripts that originated from region 2 at position -1185 were detected, despite their very low abundance. The 5' UTRs of these mRNAs were not spliced. This is to be expected because the splice acceptor site that occurs in the *GLDPA* reading frame is not available due to the substitution of the *GLDPA* reading frame by the *GUS* sequence. Taken together, these findings demonstrate that regions 2 and 7 of the *GLDPA* promoter function as separate promoters in transgenic plants of both *Arabidopsis* and *F. bidentis*.

Both *GLDPA* Transit Peptide Variants Ensure Mitochondrial Import

The P-subunit of Gly decarboxylase is located in the mitochondria; hence, the *GLDPA* precursor protein should contain a mitochondrial targeting sequence (presequence) at its N terminus (Tanudji et al., 1999; Huang et al., 2009). Analysis of the *GLDPA* coding sequence by UniProtKB predicts a transit peptide of 63 amino acids, which is equivalent to 189 nucleotides of the nucleotide sequence (Cossu and Bauwe, 1998). The use of the distal promoter and the removal of the intron shift the putative translational start site to position $+25$ nucleotides. This would result in a presequence truncated by eight residues at its N terminus. Therefore, we investigated whether the full-size transit peptide of 63 amino acids can target the green fluorescent protein (GFP) to mitochondria and, if so, whether a deletion of the eight N-terminal

residues would interfere with mitochondrial targeting. The two different *GLDPA* presequence variants were fused with the *GFP* reporter gene, and the various constructs were transiently expressed in leaf protoplasts of *Nicotiana benthamiana* under the control of the constitutive cauliflower mosaic virus (*CaMV*) 35S promoter (Odell et al., 1985). The distribution of GFP throughout the cell was analyzed by confocal microscopy (Figure 3).

When the 35S:*GLDPA*_{mt}-Ft-*mgfp6* construct containing the full-length *GLDPA* presequence (*GLDPA*_{mt}-Ft) fused to *GFP* was analyzed, GFP fluorescence was exclusively detected in the mitochondrial network. As expected, in the absence of any mitochondrial targeting peptide (construct 35S:*mgfp6*), GFP was evenly distributed throughout the cytoplasm with no visible association to any cellular organelle. When the truncated transit peptide was investigated (35S:*GLDPA*_{mt Δ 24}-Ft-*mgfp6*), the cellular pattern of GFP fluorescence was indistinguishable from that obtained with the 35S:*GLDPA*_{mt}-Ft-*mgfp6* construct. Therefore, the absence of the first eight amino acids from the *GLDPA* presequence did not affect the mitochondrial targeting of the passenger protein. We conclude that both transit peptide variants are capable of targeting the *GLDPA* protein into mitochondria.

Region 7 of the *GLDPA* Promoter Directs Bundle Sheath- and Vasculature-Specific Gene Expression in Both *F. bidentis* and *Arabidopsis*

The presence of the putative TSS identified in region 7 predominantly at position -100 (TSS_{R7}) and in region 2 at position -1185 (TSS_{R2}), respectively, raises the question whether regions 2 and 7 function as promoters that initiate transcription at these positions. To test the promoter function of region 7, the corresponding segment was fused to *GUS* (construct *GLDPA*-Ft-7; Figure 4A), and the expression pattern of the chimeric gene was analyzed in leaves of both transgenic *F. bidentis* and *Arabidopsis* (Figure 4).

Transgenic *GLDPA*-Ft-7 plants of *F. bidentis* exhibited *GUS* activity only in bundle sheath cells and vascular bundles (Figures 4B to 4D). This expression pattern was indistinguishable from that of the full-length *GLDPA* promoter (Figure 1; Engelmann et al., 2008). However, the promoter activity of *GLDPA*-Ft-7 was much lower than that of the full-length promoter (Figure 4H; Engelmann et al., 2008), indicating that other regions of the *GLDPA* promoter enhance the activity of region 7.

An almost identical expression pattern was observed in leaves of transgenic *GLDPA*-Ft-7 plants of *Arabidopsis* (Figures 4E to 4G). The two species differed only in the extent of *GUS* staining within the vasculature, which was less in the C₄ plant compared with the C₃ plant. As in *F. bidentis*, the *GLDPA*-Ft-7 promoter activity in *Arabidopsis* was very low (Figure 4H).

These findings demonstrate that region 7 is a functional promoter (P_{R7}) that can initiate transcription on its own. Furthermore, P_{R7} directs gene expression specifically in bundle sheath cells and the vascular bundles like the full-length *GLDPA* promoter.

The 5' UTR in P_{R7}-Derived Transcripts Does Not Contribute to Gene Expression Specificity

RNAs transcribed from P_{R7} at TSS_{R7} contain a 100-nucleotide-long 5' UTR (5' UTR_{R7}100) that is part of the subpromoter P_{R7} as

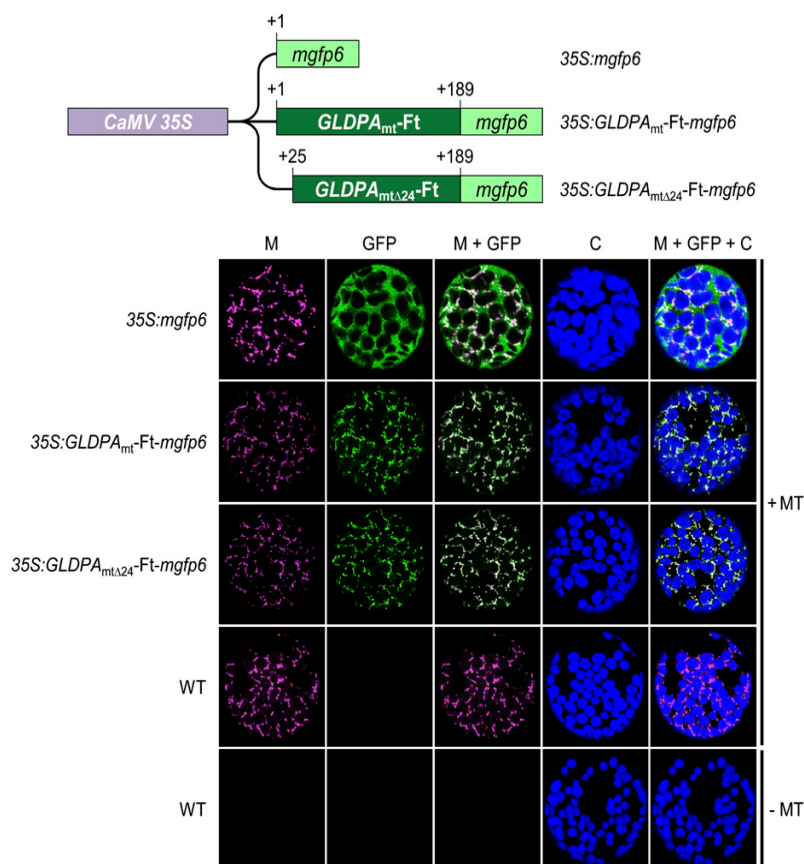


Figure 3. Localization Study of the Two Different Transit Peptide Variants of GLDPA.

The structures of the three constructs used for transient expression in leaves of *N. benthamiana* are diagrammed at the top. 35S:GLDPA_{mt}-Ft-mgfp6 contains the full-length GLDPA sequence encoding the predicted presequence for mitochondrial targeting (GLDPA_{mt}-Ft). In 35S:GLDPA_{mtΔ24}-Ft-mgfp6, the transit peptide lacks the eight amino terminal residues. 35S:mgfp6 is devoid of any transit peptide sequence and served as a control. For visualizing mitochondria, MitoTracker staining was performed (+MT) or omitted as negative control (-MT). Three different channels were used to separate the fluorescence signals of MitoTracker-labeled mitochondria (magenta color), GFP (green color), and chlorophyll of chloroplasts (blue color) from each other. When merging MitoTracker and GFP fluorescence (M + GFP), white color indicates overlapping of both signals. All three fluorescence signals are merged in the last column (M + GFP + CP). C, chloroplasts; M, mitochondria; MT, MitoTracker; WT, wild type.

defined above. It has been reported that the 5' UTRs of transcripts from *C₄* genes can be responsible for the bundle sheath-specific accumulation of the corresponding RNAs (Patel et al., 2006). To analyze whether the 5' UTR_{R7}100 of the GLDPA gene contributes to or may be even responsible for the observed bundle sheath specificity of GLDPA expression, the 5' UTR_{R7}100 was fused to the *GUS* coding sequence, and the transgene was stably expressed in *Arabidopsis* driven by the CaMV 35S promoter (35S:GLDPA-Ft-5'UTR_{R7}100-*GUS*). Transgenic *Arabidopsis* plants containing a 35S:*GUS* gene served as controls (Figure 5A).

Independently of whether the 5' UTR_{R7}100 was inserted between the 35S promoter and the *GUS* gene or not, the *GUS* gene was expressed in all inner tissues of mature rosette leaves as well as in cotyledons, roots, and partially in hypocotyls of young seedlings (Figures 5B to 5I). Thus, the 5' UTR_{R7}100 did not alter the expression pattern of the reporter gene. Transgenic *F. bidentis* and *Arabidopsis* plants harboring the construct GLDPA-Ft-7 expressed the *GUS* reporter gene specifically in the bundle sheath

and the vasculature (Figure 4). Since the construct GLDPA-Ft-7 contains the proximal promoter and the 5' UTR_{R7}100, we conclude that the 5' UTR_{R7}100 most likely does not contribute to bundle sheath-specific expression in either *Arabidopsis* or *F. bidentis*. Transgenic *Arabidopsis* plants showed a sevenfold higher *GUS* activity in total leaf extracts when the 5' UTR_{R7}100 was present compared with those plants expressing the 5' UTR_{R7}100-less *GUS* variant (Figure 5J). In addition, young seedlings harboring 35S:GLDPA-Ft-5'UTR_{R7}100-*GUS* exhibited much stronger *GUS* staining within the whole primary root than did 35S:*GUS* seedlings (Figures 5B and 5F).

These findings indicate that the 5' UTR_{R7}100 is not involved in the bundle sheath specificity of transcript accumulation. It neither destabilizes transcript accumulation in the mesophyll cells nor enhances transcript accumulation in the bundle sheath cells and the vasculature. We conclude that the bundle sheath- and vasculature-specific expression of genes driven by P_{R7} is regulated transcriptionally.

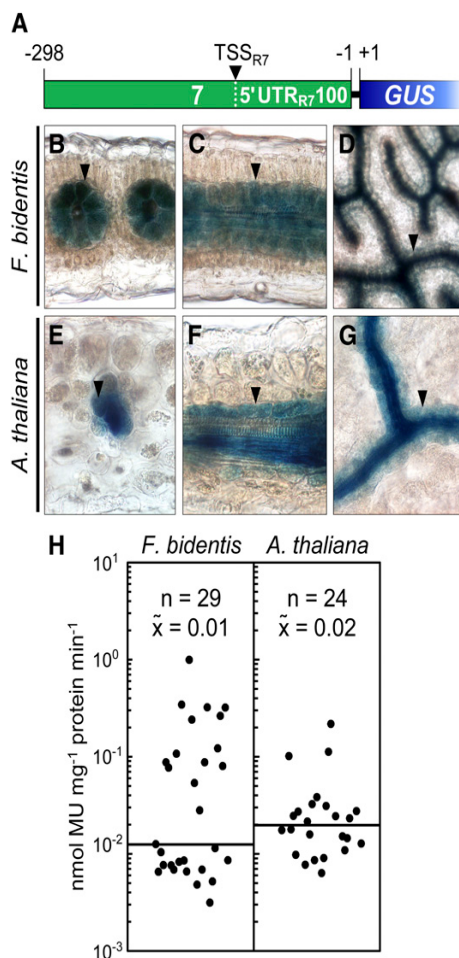


Figure 4. Functional Analysis of Region 7 of the *GLDPA* Promoter in Leaves of Transgenic *F. bidentis* and *Arabidopsis*.

(A) Schematic presentation of the *GLDPA*-Ft-7 construct. 5' UTR_{R7}100, 100-bp 5' untranslated region of *GLDPA* region 7.

(B) to (G) Histochemical localization of GUS activity in cross sections [(B), (C), (E), and (F)] and leaf blades in top view [(D) and (G)] of leaves of transgenic *F. bidentis* [(B) to (D)] and *Arabidopsis* [(E) to (G)]. Single bundle sheath cells are indicated by arrowheads. Incubation times for the GUS staining were 17 h [(E) and (F)], 29 h (G), 43 h (B), 66 h (C), and 70.5 h (D).

(H) Fluorometrical quantification of GUS activities of transgenic *F. bidentis* and *Arabidopsis* plants transformed with the *GLDPA*-Ft-7 construct. Each single dot represents one independent transgenic line. The number of lines examined (*n*) is indicated above as well as the median of all values (\bar{x}), also displayed as a black line in the diagram. MU, 4-methylumbelliferone.

Region 2 Activates Gene Expression in the Leaf Chlorenchyma and Vascular Tissues in Both *F. bidentis* and *Arabidopsis*

To investigate the promoter activity contained in region 2 of the *GLDPA* promoter, the sequence of region 2 was fused to the *GUS* reporter gene and the expression pattern and strength of the resulting *GLDPA*-Ft-2:*GUS* gene (construct *GLDPA*-Ft-2;

Figure 6A) was analyzed in transgenic *F. bidentis* and *Arabidopsis* (Figure 6). Transgenic *GLDPA*-Ft-2 plants of both *F. bidentis* and *Arabidopsis* showed an indistinguishable *GUS* expression pattern. Uniform *GUS* staining was detectable in all inner leaf tissues, namely, the chlorenchyma (mesophyll and bundle

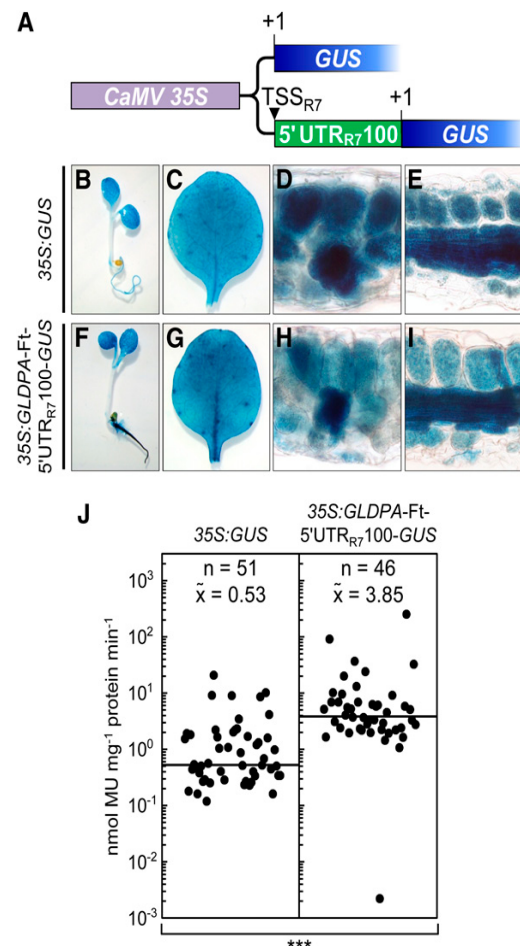


Figure 5. Analysis of the Gene Regulatory Properties of the 100-bp 5' Untranslated Region of *GLDPA* Region 7 (5' UTR_{R7}100) in Transgenic *Arabidopsis*.

(A) Schematic structure of the two constructs used for transformation. 35S:*GUS* consists of the *CaMV* 35S promoter and the *GUS* reporter gene, while 35S:*GLDPA*-Ft-5'UTR_{R7}100-*GUS* additionally contains the 100-bp long 5' UTR_{R7}.

(B) to (I) Histochemical GUS staining of *Arabidopsis* transformed with 35S:*GUS* or 35S:*GLDPA*-Ft-5'UTR_{R7}100-*GUS* in seedlings [(B) and (F)], young leaf blades [(C) and (G)], and cross sections of mature rosette leaves [(D), (E), (H), and (I)]. Staining was for 1 h [(H) and (I)], 3 h [(C) and (G)], 4 h [(B) and (F)], or 16 h [(D) and (E)].

(J) Quantitative measurements of expression strength by analyzing GUS activities in leaf extracts of transgenic *Arabidopsis* plants transformed with the 35S:*GUS* or 35S:*GLDPA*-Ft-5'UTR_{R7}100-*GUS* construct. Each single dot represents one independent transgenic *Arabidopsis* line. The number of lines examined (*n*) is indicated above as well as the median of all values (\bar{x}), which is additionally charted as a black line in the diagram. The Mann-Whitney U test was used to test for significantly different GUS activities (***) $P < 0.001$. MU, 4-methylumbelliferone.

sheath cells) and the vascular bundles. No tissue was stained preferentially (Figures 6B to 6E). Therefore, region 2 of the *GLDPA* promoter is essentially a general leaf promoter (P_{R2}) that functions in both *F. bidentis* and *Arabidopsis* with no obvious cell or tissue preference. The promoter activity of region 2 was much stronger than that of region 7 alone in both *F. bidentis* (~400-fold) and *Arabidopsis* (~1000-fold) (Figures 4H and 6F), reaching almost the promoter activity of the full-length *GLDPA* promoter at least in *Arabidopsis* (Engelmann et al., 2008).

Region 1 Enhances the Promoter Activities of Regions 2 and 7 of the *GLDPA* Promoter

Regions 1 and 2 together were previously suggested to act as a general transcriptional enhancing module of the *GLDPA* promoter (Engelmann et al., 2008). However, our findings revealed that region 2 alone is a strong autonomous promoter. To investigate whether region 1 alone enhances transcriptional activity, it was combined with either the proximal promoter, region 7 (P_{R7}), or the distal promoter, region 2 (P_{R2}), fused to *GUS*, and analyzed in transgenic *F. bidentis* and *Arabidopsis* plants. The activity of P_{R7} or P_{R2} in the presence of region 1 (constructs *GLDPA*-Ft-1-7 and *GLDPA*-Ft-1-2) was then compared with that of the constructs *GLDPA*-Ft-7 and *GLDPA*-Ft-2, which lack region 1 (see Supplemental Figure 2 online).

In both *F. bidentis* and *Arabidopsis*, the addition of region 1 to the P_{R2} and P_{R7} promoter segments caused similar effects. In combination with P_{R7} , region 1 enhanced promoter activity 15- to 18-fold, while the enhancing effect of region 1 on P_{R2} was small to moderate (approximately twofold). In both cases, the addition of region 1 did not alter the spatial expression patterns of the attached promoters. Therefore, region 1 exhibits only a quantitative enhancing effect but contains no cell or tissue specificity component.

Region 7 Represses the Promoter Activity of Region 2 of the *GLDPA* Promoter Stably in *F. bidentis* but Only Partially in *Arabidopsis*

Our data showed that the full-length *GLDPA* promoter functions essentially as a bundle sheath- and vasculature-specific promoter with minute amounts of transcripts derived from the nonspecific distal subpromoter P_{R2} (region 2). The question arose as to how this expression pattern could be achieved in view of the fact that the nonspecific subpromoter P_{R2} alone is about two to three magnitudes stronger than the specific proximal subpromoter P_{R7} (region 7) alone. Previous experiments had shown that a recombined promoter consisting of regions 1, 2, 3, and 7 in the order given (*GLDPA*-Ft-1-2-3-7) directed an expression pattern that was indistinguishable from that of the full-length promoter (Engelmann et al., 2008). A plausible hypothesis is that the activity of P_{R2} within the *GLDPA* promoter is repressed by the proximal promoter P_{R7} and/or region 3. To identify the component in the *GLDPA* promoter that suppresses its activity in the mesophyll tissue, various combinations of regions 1, 2, 3, and 7 were analyzed with regard to their expression specificities in both *F. bidentis* and *Arabidopsis*.

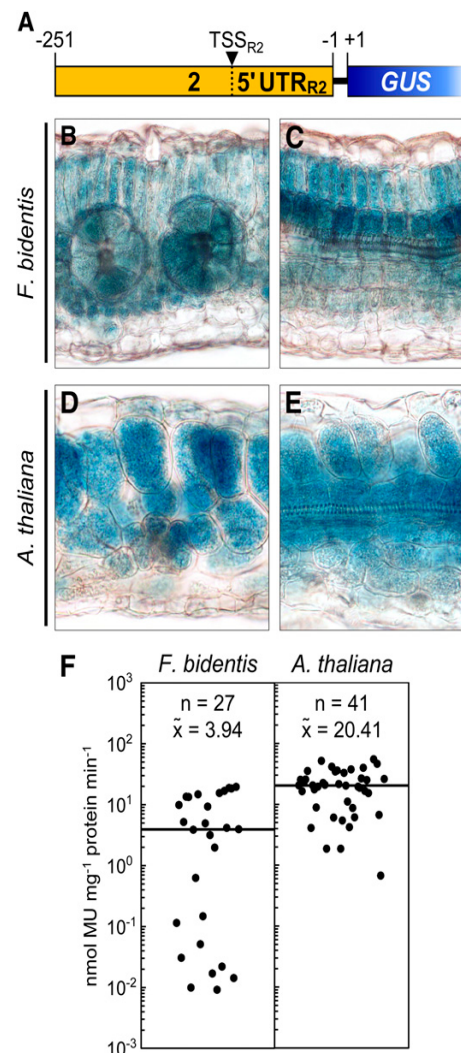


Figure 6. Functional Analysis of Region 2 of the *GLDPA* Promoter in Leaves of Transgenic *F. bidentis* and *Arabidopsis*.

(A) Schematic structure of the *GLDPA*-Ft-2 construct. 5' UTR_{R2}, 5' untranslated region of *GLDPA* region 2.

(B) to (E) Histochemical GUS staining in cross sections of leaves of transgenic *F. bidentis* or *Arabidopsis* harboring the *GLDPA*-Ft-2 construct. Incubation times for the GUS staining procedure were 1.5 h (E) and 2 h (B) to (D).

(F) Fluorometrical quantification of GUS activities of transgenic *F. bidentis* and *Arabidopsis* plants transformed with the *GLDPA*-Ft-2 construct. Each single dot represents one independent transgenic line. The number of lines examined (n) is indicated above as well as the median of all values (\bar{x}), which is also displayed as a black line in the diagram. MU, 4-methylumbelliferone.

Transgenic *F. bidentis* plants expressing *GLDPA*-Ft-1-2-7: *GUS* (*GLDPA*-Ft-1-2-7; Figure 7A) retained the expression specificity in bundle sheath cells (Figure 7B), suggesting that, in *F. bidentis*, region 3 is not required for promoter specificity. As expected, omission of region 1 (*GLDPA*-Ft-2-7; Figure 7A) did not affect the spatial GUS staining pattern (Figure 7C). Therefore,

in *F. bidentis*, the presence of P_{R7} alone suffices to repress the activity of P_{R2} .

By contrast, in *Arabidopsis*, the absence of region 3 in the promoter constructs caused a loss of bundle sheath and vasculature specificity. All transgenic *Arabidopsis* plants harboring *GLDPA-Ft-1-2-7* showed *GUS* expression in all inner leaf tissues (Engelmann et al., 2008). Thus, the expression pattern resembles that of P_{R2} alone (Figure 6). In contrast with the stable *GUS* expression pattern of transgenic *GLDPA-Ft-1-2-7* plants, *Arabidopsis* lines containing the *GLDPA-Ft-2-7* construct varied in their *GUS* expression patterns between two extremes: bundle sheath/vasculature-specific *GUS* staining to an expression in all inner leaf tissues (Figure 8). In the presence of region 3 (*GLDPA-Ft-2-3-7*; Figure 9A), all transgenic *Arabidopsis* plants exhibited *GUS* expression in the bundle sheath and the vasculature (Figures 9C and 9G). In *Arabidopsis*, therefore, P_{R7} is not sufficient to suppress P_{R2} but needs region 3 for stable repression. With regard to *GLDPA-Ft-1-2-7 Arabidopsis* plants, the presence of region 1 enhances the activity of P_{R2} , and the partially repressive function of P_{R7} is overcome.

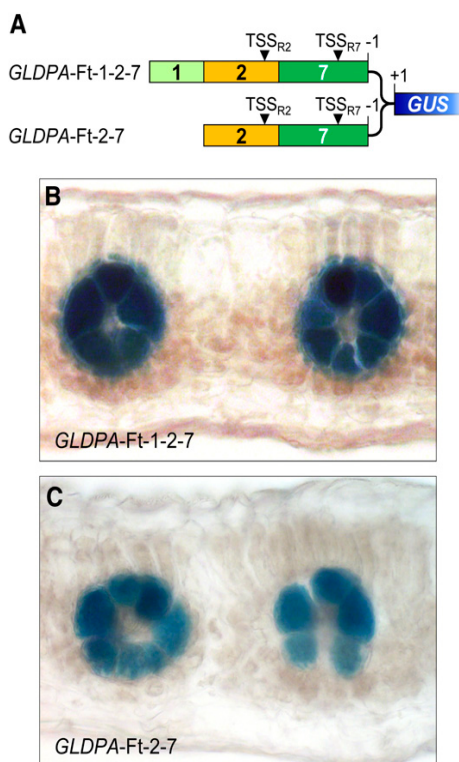


Figure 7. Functional Analysis of the Interactions of Regions 2 and 7 of the *GLDPA* Promoter in Transgenic *F. bidentis*.

(A) Schematic structure of the constructs *GLDPA-Ft-1-2-7* and *GLDPA-Ft-2-7*.

(B) and **(C)** Histochemical *GUS* staining in leaf cross sections of transgenic *F. bidentis* transformed with either *GLDPA-Ft-1-2-7* or *GLDPA-Ft-2-7*. Incubation times for the *GUS* staining procedure were 2 h **(B)** and 6 h **(C)**.

In *Arabidopsis*, Region 3 Cannot Maintain Bundle Sheath Specificity on Its Own and Needs the Presence of P_{R7}

Since region 3 of the *GLDPA* promoter is absolutely required for the suppression of P_{R2} activity in *Arabidopsis*, the question arose whether in *Arabidopsis* P_{R7} also is required for region 3 to be functional. In the natural context of the *GLDPA* promoter, region 3 is located 3' to the subpromoter P_{R2} followed further downstream by P_{R7} . Therefore, we wanted to know whether the position of region 3 with respect to P_{R2} is important for its suppressing activity and whether region 3 can also influence the activity of P_{R7} .

To investigate whether region 3 alone could repress P_{R2} activity, it was fused downstream (*GLDPA-Ft-2-3*) as well as upstream (*GLDPA-Ft-3-2*) of P_{R2} (Figure 9A). In both cases, all transgenic *Arabidopsis* plants exhibited the same uniform *GUS* expression pattern in the chlorenchyma and vasculature as detected for P_{R2} alone (Figures 9D, 9E, 9H, and 9I). By contrast, the combination of P_{R2} , region 3, and P_{R7} (*GLDPA-Ft-2-3-7*; Figure 9A) caused specific *GUS* expression in the bundle sheath and the vasculature (Figures 9C and 9G). Therefore, region 3 alone is not sufficient to suppress P_{R2} , independent of its location down- or upstream of P_{R2} , but the presence of P_{R7} , in addition, is necessary.

To examine whether region 3 could affect also P_{R7} activity, the corresponding transgene (*GLDPA-Ft-3-7*; Figure 9A) was constructed and analyzed in transgenic *Arabidopsis*. As expected, the combination of region 3 and P_{R7} led to the same bundle sheath- and vasculature-specific expression pattern as P_{R7} alone (Figures 9B and 9F), indicating that region 3 has no influence on the spatial activity of P_{R7} .

However, the *GLDPA-Ft-3-7* construct was 20-fold more active than *GLDPA-Ft-7* (Figures 4H and 9J). Interestingly, region 3 was also able to enhance P_{R2} activity (cf. constructs *GLDPA-Ft-3-2* [Figure 9J] and *GLDPA-Ft-2* [Figure 6F]), although to a much less degree than observed for P_{R7} (twofold versus 20-fold). Thus, region 3 can enhance transcription of P_{R7} and P_{R2} with a comparable strength as detected for region 1.

DISCUSSION

The commonly believed evolutionary scenario of C_4 photosynthesis predicts that relatively early along the path toward C_4 photosynthesis, a photorespiratory CO_2 pump was established by compartmentalization of Gly decarboxylase activity in the bundle sheath (Sage, 2004; Bauwe, 2011). All available experimental data confirm the final outcome of this evolutionary process, namely, that in present C_4 species, Gly decarboxylase accumulates predominantly in the bundle sheath (Li et al., 2010; Majeran et al., 2010). Along these lines, the activity of the *GLDPA* promoter of the Asteracean C_4 species *F. trinervia* was found to be restricted to the bundle sheath cells and vasculature in leaves of both transgenic *F. bidentis* (C_4) and *Arabidopsis* (C_3) (Engelmann et al., 2008). This suggested that the bundle sheath-specific accumulation of *GLDPA* mRNAs should be essentially regulated by transcription (Engelmann et al., 2008). Data presented in this article show that *GLDPA* gene regulation is much more complex than previously thought. We discuss and provide evidence that the mRNA output of this gene is determined by a combination of transcriptional and posttranscriptional means.

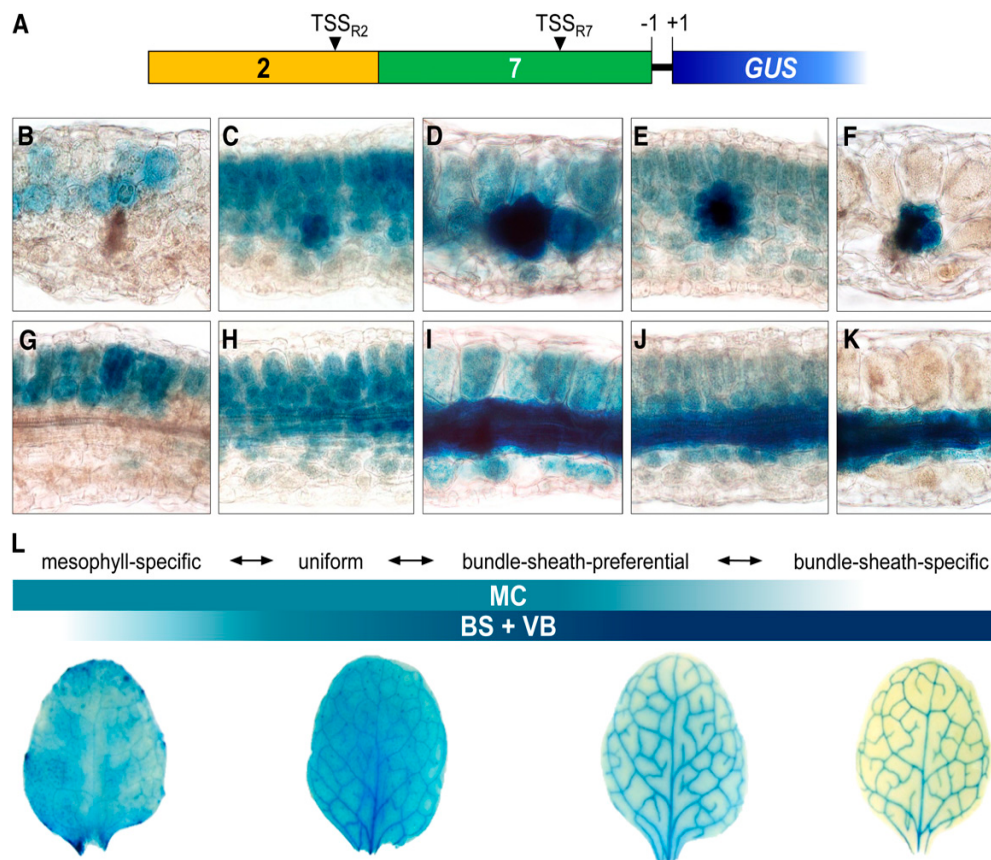


Figure 8. Functional Analysis of the Interactions of Regions 2 and 7 of the *GLDPA* Promoter in Transgenic *Arabidopsis*.

(A) Schematic structure of the promoter-reporter gene construct *GLDPA*-Ft-2-7.

(B) to (K) Analysis of GUS staining patterns in leaf cross sections of five independent transgenic *Arabidopsis* lines ([B]/[G], [C]/[H], [D]/[I], [E]/[J], and [F]/[K]) carrying the *GLDPA*-Ft-2-7 construct. Incubation times for the GUS staining procedure were 3.5 h ([C], [E], [H], and [J]), 4 h ([B], [D], [F], [I], and [K]), and 5 h (G).

(L) GUS staining in leaf blades of four different *GLDPA*-Ft-2-7 *Arabidopsis* lines representative of the smooth transition of the various expression patterns detected. This transition is schematically depicted as blue bars representing the varying intensity of GUS staining of the mesophyll (MC) and the bundle sheath cells, including the vascular bundles (BS + VB). Incubation times for the GUS staining procedure were 6, 3, 3.5, and 6 h (from left to right).

The *GLDPA* Promoter Consists of Two Promoters in Tandem

The *GLDPA* promoter is composed of tandem promoters, P_{R2} and P_{R7} , that together ensure a strong bundle sheath expression. The two subpromoters are not easily recognized by inspection of their corresponding nucleotide sequences. No reliable candidates for TATA boxes can be detected in the predicted distance of 25 to 40 bp upstream of the two transcriptional initiation sites (Joshi, 1987; Bernard et al., 2010; Zuo and Li, 2011). This is not surprising since, for instance in *Arabidopsis*, only 20 to 30% of all promoters contain a TATA box/variant (Molina and Grotewold, 2005; Yamamoto et al., 2009; Bernard et al., 2010). Recently, TC elements have been proposed as a novel class of regulatory elements that control transcription in plants (Bernard et al., 2010). Indeed, several TC elements are located around TSS_{R2} but predominantly around TSS_{R7} within the predicted range of 50 bp up- and/or downstream of the corresponding TSS (see Supplemental Figure 3 online; Zuo and Li, 2011). The possible importance of TC elements for transcriptional regulation of P_{R7} is supported by the fact that the motifs

CCCTTT, CCTTCT, and TCTTCT are unique to region 7 within the *GLDPA* promoter and that TCTTCT is one of the three TC element types most frequently observed (Bernard et al., 2010). TSS_{R2} is flanked by a sequence repeat that is very similar to the predicted Initiator (Inr) motif shown to be essential for the light-dependent activity of the *psaDb* promoter from *Nicotiana sylvestris* (Nakamura et al., 2002; see Supplemental Figure 3 online). According to the YR rule (YR, Y = C or T, R = A or G, TSS underlined), most of the *Arabidopsis* promoters contain the CA dimer sequence around their TSSs (Yamamoto et al., 2007), which is also true for TSS_{R2} . Yamamoto et al. (2007) consider this YR rule to represent a less stringent form of Inr. This indicates that Inr elements might be crucial for transcriptional activity of P_{R2} .

The *GLDPA* Subpromoters Diverge in Their Specificities

When analyzed separately, the two subpromoters diverge in their spatial expression profiles. The proximal promoter (P_{R7}) alone,

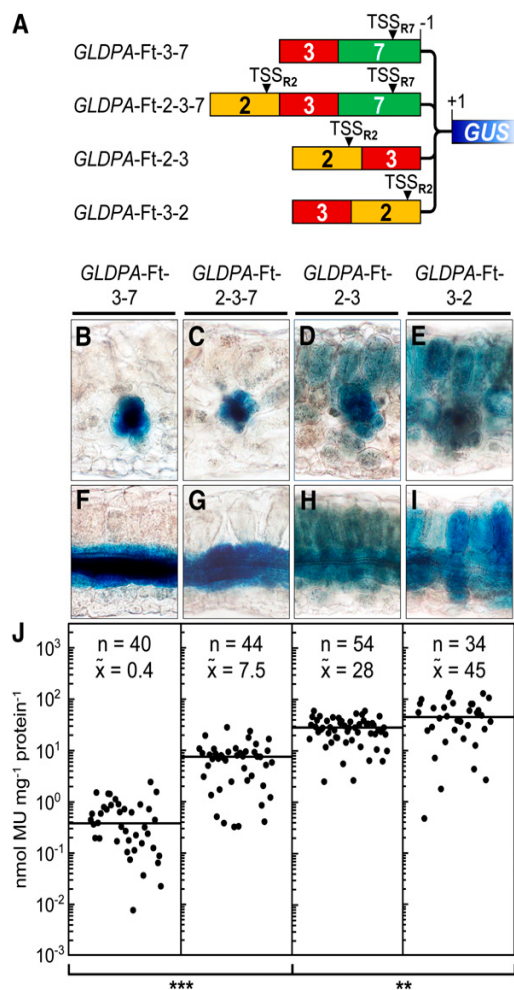


Figure 9. Functional Analysis of Region 3 of the *GLDPA* Promoter in Transgenic *Arabidopsis*.

(A) Schematic structure of the constructs *GLDPA-Ft-3-7*, *GLDPA-Ft-2-3-7*, *GLDPA-Ft-2-3*, or *GLDPA-Ft-3-2*.

(B) to (I) Histochemical GUS staining in cross sections of leaves of transgenic *Arabidopsis* transformed with *GLDPA-Ft-3-7*, *GLDPA-Ft-2-3-7*, *GLDPA-Ft-2-3*, or *GLDPA-Ft-3-2*. Incubation times for GUS staining were 0.5 h **(D)** and **(H)**, 1 h **(E)**, 2.5 h **(I)**, 3.5 h **(C)** and **(G)**, 5 h **(B)**, or 6 h **(F)**. **(J)** Fluorometrical measurement of GUS activities in transgenic *Arabidopsis* transformed with *GLDPA-Ft-3-7*, *GLDPA-Ft-2-3-7*, *GLDPA-Ft-2-3*, or *GLDPA-Ft-3-2*. Each single dot represents one independent transgenic line. The number of plants analyzed (*n*) is indicated at the top of each diagram as well as the median values (\bar{x}), which are also added as black lines in the diagrams. The Mann-Whitney U test was used to test for significantly different GUS activities (****P* < 0.001; ***P* < 0.01). MU, 4-methylumbelliferone.

defined by region 7, is relatively weak, and, as observed for the full-length *GLDPA* promoter, specifically active in the bundle sheath and the vasculature (Figure 4). The distal promoter (P_{R2}) alone, defined by region 2, is strong and drives expression in all inner leaf cells, including the mesophyll (Figure 6). By contrast, in the context of the full-length *GLDPA* promoter, which exhibits strong specific activity, the final RNA output from both pro-

motors, as measured by 5'-RACE and RNA sequencing experiments, is just the reverse. RNAs transcribed from the proximal promoter dominate the *GLDPA* transcript population, and RNAs derived from P_{R2} are in the minority (Figure 2). We could show that the activity of P_{R7} is enhanced in the presence of regions 1 and 3 of the *GLDPA* promoter, which might explain the high accumulation of P_{R7} -derived transcripts. When the two subpromoters (i.e., regions 2 and 7) are combined and fused to the *GUS* reporter gene, the readout of this chimeric gene in *F. bidentis* is similar to that of the complete *GLDPA* promoter (cf. Figures 1 and 7). This finding suggests that the proximal bundle sheath-specific promoter P_{R7} turns off the activity of the unspecific distal promoter P_{R2} . How could a downstream promoter interfere with the activity of an upstream promoter and even disable it?

That a strong upstream promoter can shut off the activity of a downstream promoter is well documented (Mazo et al., 2007). This phenomenon is called transcriptional interference and may be defined as "the in cis suppression of one transcriptional process by another" (Palmer et al., 2011). Transcriptional interference has been documented as a general regulatory process affecting the transcription from adjacent convergent or tandem promoters (Palmer et al., 2011). Promoters can impede or even block one another by different ways. One possibility is the transcription from a strong regulatory promoter that might impair the recruitment of the transcription initiation complex or the transcriptional elongation of a neighboring target gene (Mazo et al., 2007; Palmer et al., 2011). Transcriptional interference has been reported for plants. The strong 35S promoter of a T-DNA inserted upstream of the *Arabidopsis Riba1* gene results in large transcripts that run over the *Riba1* promoter and thereby inhibit *Riba1* transcription (Hedtke and Grimm, 2009).

In all known cases of transcriptional interference occurring with tandem promoters, the upstream promoter blocks the activity of the downstream promoter. However, with respect to the *GLDPA* promoter, just the opposite is true: The downstream promoter P_{R7} inhibits the output from the upstream promoter P_{R2} . Could a roadblock mechanism explain the transcriptional interference between the *GLDPA* subpromoters (i.e., a pausing RNA polymerase II) (Levine, 2011) or a scaffold of general transcription factors (Yudkovsky et al., 2000) residing at the downstream promoter inhibit the progress of RNA polymerase II from the upstream promoter? There is increasing evidence that RNA polymerase II pauses quite often after having initiated transcription and having produced a nascent transcript of ~30 to 50 nucleotides (Levine, 2011). Therefore, it is conceivable that an RNA polymerase II pausing at P_{R7} might represent a roadblock that impedes the elongation of transcripts originating from P_{R2} and is involved in the suppression of the P_{R2} output.

The *GLDPA* 5' Flanking Region: A Player in Posttranscriptional Control

One hallmark of the *GLDPA* gene of *F. trinervia* is an intron located within the 5' UTR of P_{R2} -derived transcripts. Depending on the splice donor site used, the intron commences 84 or 139 nucleotides behind the respective TSS of the distal

promoter P_{R2} within region 3 and ends 17 nucleotides behind the first nucleotide of the *GLDPA* reading frame (Figure 2). RNA sequencing experiments using the 454 technology (Gowik et al., 2011) revealed that this 5' intron is present in *GLDPA* transcripts, and the sequence reads cover the intron region uniformly (see Supplemental Figure 4 online). By contrast, sequence reads from the gene-internal introns are not detectable. The accumulation of unspliced P_{R2} -derived transcripts with respect to their 5' UTR demonstrates that the splicing efficiency of the 5' intron is drastically lower than that of the gene-internal ones.

To prevent the accumulation of aberrant mRNAs (i.e., mRNAs that are erroneously or not completely spliced), eukaryotes have developed various quality control systems (Egecioglu and Chanfreau, 2011). Spliceosomal DEXD/H box ATPases provide the first layer of defense. They act as kinetic proofreading systems and limit the escape of unspliced or erroneously spliced RNAs from the spliceosome (Egecioglu and Chanfreau, 2011). Despite the accuracy of these proofreading activities, unspliced mRNAs may escape detection; therefore, external quality control systems have been built up. The nuclear exosome takes part in the degradation of unspliced RNAs (Houseley et al., 2006; Fasken and Corbett, 2009), although the molecular mechanisms by which unspliced RNAs are recognized are not yet clear. Exon-junction complexes that are deposited on spliced RNAs might be involved in the recognition mechanism (Egecioglu and Chanfreau, 2011). These protein complexes also play a prominent role in nonsense-mediated mRNA decay. Nonsense-mediated mRNA decay is a eukaryotic mRNA

surveillance mechanism that detects and degrades mRNAs containing premature termination codons (Chang et al., 2007; Brogna and Wen, 2009). In plants, long 3' UTRs, introns that are located in 3' UTRs, and upstream open reading frames (uORFs) within the 5' UTR can trigger nonsense-mediated mRNA decay (Kertész et al., 2006; Hori and Watanabe, 2007; Nyikó et al., 2009).

The 5' intron sequence that is present in unspliced transcripts derived from P_{R2} contains several uORFs. The one that starts directly upstream of TSS_{R7} encodes more than 35 amino acids (see Supplemental Figure 5 online). uORFs are considered to have the potential to elicit the nonsense-mediated mRNA decay response when they give rise to proteins that are larger than the critical threshold of 35 amino acids (Nyikó et al., 2009). When P_{R2} is combined with P_{R7} (*GLDPA*-Ft-2-7), a 135-nucleotide uORF commencing directly upstream of TSS_{R7} might encode a protein of 45 amino acids that should promote nonsense-mediated mRNA decay (see Supplemental Figure 6 online). This 135-nucleotide uORF is present only in transcripts originating from P_{R2} . The 5'-RACE analyses of transgenic *GLDPA*-Ft-2-7 plants of *Arabidopsis* and *F. bidentis* showed that transcription started from either TSS_{R2} (P_{R2}) or TSS_{R7} (P_{R7}). While P_{R7} -derived RNAs exhibited more or less the same length, P_{R2} -derived transcripts appeared to be destabilized because many RNAs started randomly between TSS_{R2} and TSS_{R7} , indicating RNA degradation (see Supplemental Figure 6 online). Taken together, the presented data demonstrate that the mRNA output of the *GLDPA* gene of *F. trinervia* is regulated by an intricate interplay of transcriptional and posttranscriptional mechanisms.

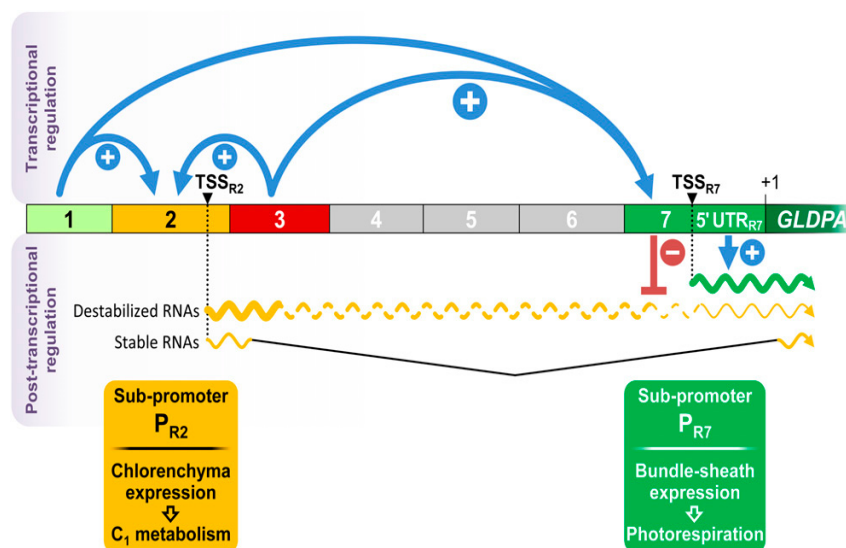


Figure 10. The Expression of the *GLDPA* Gene of *F. trinervia* Is Regulated by an Intricate Interplay of Transcriptional and Posttranscriptional Mechanisms.

The proximal promoter P_{R7} is sufficient to confer expression specifically in bundle sheath cells and the vascular bundles of leaves but can be effectively enhanced by regions 1 and 3. Transcripts generated at TSS_{R7} are presumably stabilized by their 5' UTR (5' UTR $_{R7}$), finally resulting in the accumulation of *GLDPA* protein in the distinct cell types to contribute to photorespiration. The activity of the distal promoter P_{R2} in all inner leaf tissues is also enhanced by regions 1 and 3. Transcripts from TSS_{R2} are supposed to be destabilized when they contain the sequence of region 7, which impedes RNA accumulation. The problem of RNA instability can be overcome by splicing out impairing elements, assuring at least small amounts of stable *GLDPA* transcript, and, thus, *GLDPA* protein, additionally in the mesophyll cells to serve the C_1 metabolic pathway.

Evolution of the *GLDPA* Promoter: The Necessity of Being Bundle Sheath Specific, but Not Completely

The presence of the 5' intron in P_{R2} -derived transcripts and of an alternative ATG codon 25 nucleotides behind the major translational start site results in a *GLDPA* protein variant whose mitochondrial targeting peptide is truncated by eight amino acids. Our import experiments showed that, nevertheless, the truncated transit peptide is capable of directing an attached passenger protein to the mitochondria (Figure 3). This shows that P_{R2} - and P_{R7} -derived mRNAs yield a *GLDPA* protein that accumulates in the mitochondria. One wonders why the *GLDPA* promoter contains one subpromoter with the desired specificity in the bundle sheath and a second subpromoter that is active in all internal leaf cells, including the mesophyll. Moreover, the second, nonspecific promoter is not allowed to express its full potential but is almost, even though not completely, switched off by a combination of transcriptional and posttranscriptional means.

Mutational analysis with *Arabidopsis* revealed that a *GLDP* double mutant in which both of the two *GLDP* genes were knocked out is lethal, even under nonphotorespiratory conditions (Engel et al., 2007). This indicates that the activity of GDC is indispensable and that all biosynthetically active cells need Gly decarboxylase activity for one-carbon metabolism (Hanson and Roje, 2001). Although *GLDPA* transcripts could not be observed in mesophyll cells by in situ hybridization (Engelmann et al., 2008), more sensitive methods like laser microdissection followed by RNA sequencing or immunogold labeling allow the detection of traces of *GLDP* transcripts and *GLDP* protein in mesophyll cells of C_4 and C_3 - C_4 intermediate plants (Rawsthorne et al., 1988; Yoshimura et al., 2004; Li et al., 2010). According to 454 pyrosequencing data, in C_4 *Flaveria* species, the *GLDPA* gene is the only active leaf *GLDP* gene (Gowik et al., 2011). Therefore, we hypothesize that the *GLDPA* gene of *F. trinervia* (C_4) must fulfill two purposes: First, it has to serve the requirements of the photorespiratory pathway, and its activity should therefore be restricted to the bundle sheath cells; second, small amounts of GDC activity are needed in all biosynthetically active cells, and in the mesophyll cells; therefore, the regulatory system of the *GLDPA* gene has to be somewhat leaky. This is achieved by a complex interplay of transcriptional and posttranscriptional mechanisms (Figure 10). The *GLDPA* promoter is in fact composed of two subpromoters acting in tandem. The proximal promoter P_{R7} is weak but suffices for bundle sheath- and vasculature-specific gene expression. To ensure strong transcriptional activity of P_{R7} , it is enhanced by regions 1 and 3. The strong distal promoter P_{R2} is active in the vasculature and all chlorenchyma tissues of the leaf, including the mesophyll. Its activity is also raised in the presence of regions 1 and 3. However, P_{R7} suppresses the activity of P_{R2} apparently posttranscriptionally by destabilizing P_{R2} -derived transcripts when they contain an intron, including the sequence of region 7, within their 5' UTRs. Thus, the removal of this intron is essential for generating *GLDPA* protein. Inefficient splicing in the 5' UTR drastically reduces the amounts of *GLDPA* transcripts starting from P_{R2} .

We provided conclusive evidence that an intricate combination of transcriptional and posttranscriptional regulation ensures small amounts of *GLDPA* mRNAs accumulate in the mesophyll cells of C_4 *Flaveria* species. It remains to be investigated how this

pattern of gene expression regulation evolved in the genus *Flaveria*. These studies are underway and should elucidate the adaptive changes in gene expression that are a central component of C_4 evolution.

METHODS

Generation of Chimeric Promoters and Cloning of Promoter-Reporter Gene Constructs

The DNA amplification and cloning procedures were accomplished according to Sambrook and Russell (2001). The dissection of the *GLDPA* promoter from *Flaveria trinervia* into seven regions and the cloning of the constructs *GLDPA*-Ft-7 (previously referred to as *GLDPA*-Ft- $\Delta 6$), *GLDPA*-Ft-1-2-7, and *GLDPA*-Ft:H2B-yellow fluorescent protein (*YFP*) (previously referred to as *GLDPA*-Ft:H2B:YFP) have been described by Engelmann et al. (2008). All *GLDPA* promoter regions were amplified by PCR by means of the Phusion High-Fidelity DNA Polymerase (New England Biolabs) or the *Pfu* DNA Polymerase (Stratagene) using the *GLDPA*-Ft construct (Cossu, 1997; Engelmann et al., 2008) as template and the corresponding oligonucleotides containing the respective restriction sites (see Supplemental Tables 1 and 2 online). For *GLDPA*-Ft-2-3-7, region 2-3 as *Xba*I-*Bcl*I and 7 as a *Bcl*I-*Xma*I fragment were cloned together into the *Xba*I-*Xma*I-digested binary plant transformation vector pBI121 (Clontech Laboratories; Jefferson et al., 1987; Chen et al., 2003) lacking the *CaMV* 35S promoter. *GLDPA*-Ft-2-7 and *GLDPA*-Ft-3-7 were constructed by exchanging region 2-3 of *GLDPA*-Ft-2-3-7 with regions 2 and 3 as *Xba*I-*Bcl*I fragments, respectively. For cloning of *GLDPA*-Ft-2 and *GLDPA*-Ft-2-3, regions 2 and 2-3 were ligated as *Xba*I-*Xma*I fragments into the *Xba*I-*Xma*I-digested pBI121 vector lacking the 35S promoter. Region 3 was cloned as an *Xba*I-*Xba*I fragment into *GLDPA*-Ft-2 previously cut with *Xba*I to generate *GLDPA*-Ft-3-2. The correct orientation of region 3 was confirmed by DNA sequencing. 35S:*GUS* was constructed by inserting the 35S promoter amplified from pBI121-35S:H2B:YFP (Boisnard-Lorig et al., 2001) as a *Hind*III-*Xba*I-*Xma*I fragment into the *Hind*III-*Xma*I-digested pBI121 vector, whereas for 35S:*GLDPA*-Ft-5' UTR_{R7}100-*GUS*, the 35S promoter was directly excised from pBI121-35S:H2B:YFP as a *Hind*III-*Xba*I fragment to ligate it together with the *Xba*I-*Xma*I-digested 100-bp 5' UTR of region 7 (5' UTR_{R7}100) into the *Hind*III-*Xma*I-restricted pBI121 vector. For 35S:*mgfp6*, 35S:*GLDPA*_{mt}-Ft-*mgfp6*, and 35S:*GLDPA*_{mt Δ 24}-Ft-*mgfp6*, Gateway Technology (Invitrogen) was applied starting with the generation of Gateway-compatible recombination fragments by PCR. Regarding 35S:*mgfp6*, the primers ATG-5'-*attB1* and ATG-3'-*attB2* (see Supplemental Table 1 online) were used for simple primer dimer formation, elongation, and amplification generating a start codon (ATG) with flanking *attB* sites, whereas for 35S:*GLDPA*_{mt}-Ft-*mgfp6* and 35S:*GLDPA*_{mt Δ 24}-Ft-*mgfp6*, *F. trinervia* cDNA was used for the amplification of the full-length N-terminal presequence of the *GLDPA* gene (*GLDPA*_{mt}, 189 bp; primers: *GLDPA*_{mt}-5'-*attB1*/*GLDPA*_{mt}-3'-*attB2* and *attB1*/*attB2* adapter) as annotated by GenBank (accession number Z99767) or a shorter version lacking the first 24 bp (*GLDPA*_{mt Δ 24}, 165 bp; primers: *GLDPA*_{mt Δ 24}-5'-*attB1*/*GLDPA*_{mt}-3'-*attB2* and *attB1*/*attB2* adapter). The *attB*-flanked PCR products were recombined into pDONRTM221 (Invitrogen) and afterwards into pMDC83 (Curtis and Grossniklaus, 2003). All generated constructs were verified by DNA sequencing.

Transformation of *Arabidopsis thaliana* and *Flaveria bidentis*

All chimeric promoter-reporter gene constructs were transformed into either the *Agrobacterium tumefaciens* strain AGL1 (Lazo et al., 1991) or GV3101 (pMP90) (Koncz and Schell, 1986) via electroporation. Transgenic *F. bidentis* was generated according to Chitty et al. (1994) by means

of *Agrobacterium* AGL1 containing the respective construct. *Arabidopsis* (ecotype Columbia) was transformed by the floral dip method (Clough and Bent, 1998) modified according to Logemann et al. (2006) using the *Agrobacterium* strain GV3101 harboring the appropriate construct. The presence of the respective transgene within the genome of each single independent *F. bidentis* T0 and *Arabidopsis* T1 line was verified by PCR after DNA isolation as described by Edwards et al. (1991).

In Situ Analysis of GUS and Detection of Its Activity

The fifth leaf from the top of 40- to 50-cm-tall transgenic *F. bidentis* T0 plants or three mature rosette leaves of 3- to 4-week-old transgenic *Arabidopsis* T1 plants prior to flowering were used for the fluorometrical quantification of GUS activity according to Jefferson et al. (1987) and Kosugi et al. (1990). The statistical significance of the difference between two data sets was analyzed by means of the Mann-Whitney U test (<http://elegans.som.vcu.edu/~leon/stats/utest.html>). Histochemical GUS analyses were performed as described by Engelmann et al. (2008). The fifth leaf from the top of transgenic *F. bidentis* T0 plants (40 to 50 cm) and single rosette leaves as whole blades or manually cut cross sections as well as young seedlings of transgenic T1 *Arabidopsis* plants were used for the histochemical GUS analysis in situ, respectively.

Transient Gene Expression in Leaves of *Nicotiana benthamiana* and Isolation of Protoplasts

The *Agrobacterium*-mediated infiltration of leaves of *N. benthamiana* was performed according to Waadt and Kudla (2008) by means of *Agrobacterium* GV3101 (pMP90) containing the respective construct and the *Agrobacterium* strain p19 for suppression of gene silencing (Voinnet et al., 2003). After 4 d, two infiltrated leaves per plant were harvested for protoplast isolation. Four leaf pieces (~0.7 × 0.7 cm each) per blade were cut out with a razor blade, transferred into 5 mL of enzyme solution (Yoo et al., 2007), vacuum-infiltrated three times for 30 s, and then incubated at room temperature for 2 h. After light shaking to release protoplasts, remaining leaf pieces were removed, and MitoTracker Orange CMTMRos (Invitrogen) was added to a final concentration of 150 nM to the suspension of protoplasts for labeling mitochondria. After incubation for 15 min at 37°C, protoplasts were centrifuged at 500g for 1 min, the supernatant was removed, and the sedimented leaf cells were resuspended in 100 µL W5 solution (Yoo et al., 2007) for analysis by confocal microscopy.

Confocal Laser Scanning Microscopy and Fluorescence Microscopy

The analysis of protoplasts by confocal laser scanning microscopy was performed with the LSM 510 (Carl Zeiss). Protoplasts were excited at 488 nm (for detection of GFP and chlorophyll fluorescence) and at 561 nm (for detection of MitoTracker fluorescence), respectively. To visualize specifically GFP fluorescence, a 505- to 550-nm band-pass emission filter was used. The fluorescence of MitoTracker Orange CMTMRos-labeled mitochondria was observed using a 575- to 615-nm band-pass filter, and the autofluorescence of chlorophyll was recorded by means of a 650-nm long-pass filter.

Leaf cross sections, complete leaf blades, and roots of transgenic *F. bidentis* as well as roots of transgenic *Arabidopsis* carrying the *GLDPA*-Ft:H2B-YFP chimeric gene were analyzed with the aid of an Axiophot fluorescence microscope (Carl Zeiss) that was equipped with an integrated HBO-UV lamp (Carl Zeiss) and a DP50-CU camera (Olympus Optical) using the filter set F41-028 (excitation, HQ 500/20; beam splitter, Q 515 LP; emission, HQ 535/30; AHF Analysentechnik). Bright-field and fluorescence pictures were merged with Adobe Photoshop 7.0 (Adobe

Systems). Prior to fluorescence microscopy, leaf blades were extracted with 95% ethanol according to Zhou et al. (2005). Leaf blades were harvested from 30-cm-tall *F. bidentis* grown in the greenhouse. Roots of young seedlings of *Arabidopsis* and *F. bidentis* were taken from plants cultivated on agar medium in a climate chamber.

Analysis of mRNA 5' Ends by 5'-RACE and 454 Sequencing

Total RNA from leaves of *F. trinervia* was isolated according to Westhoff et al. (1991). After enrichment by the Oligotex mRNA Midi Kit (Qiagen) 0.5 µg of poly(A⁺) mRNA was used for cDNA first-strand synthesis performed with the SMART RACE cDNA amplification kit (Clontech Laboratories) and PowerScript reverse transcriptase (Clontech Laboratories) according to the manufacturers' protocols. For PCR amplification of 5' UTRs with the Phusion High-Fidelity DNA polymerase (New England Biolabs), the gene-specific 3' oligonucleotide *GLDPA*-RACE4 (5'-GAGATCTTG-GACTTGACTGTGTC-3') and the SMART-II-A-Primer (5'-AAGCAGTGG-TATCAACGCAGAGT-3') were used. The PCR fragment was cloned subsequently using the CloneJET PCR cloning kit (Fermentas). Sixty independent clones were analyzed by colony PCR using the SMART-II-A-Primer and the gene-specific *GLDPA*-RACE6 oligonucleotide (5'-ACACCGTACATAGCAGCCATG-3'). These PCR products were verified by restriction endonuclease analyses, leading to the identification of 51 potentially correct clones. Plasmid DNA was isolated from 14 of them for DNA sequencing.

The generation of 454 reads was described by Gowik et al. (2011). The *F. trinervia* reads were mapped on the *GLDPA* gene sequence with the CLC Genomic Workbench (version 4.8). CAP3 (Huang and Madan, 1999) was used for the de novo assembly of the *F. trinervia* 454 reads. The *GLDPA* contig was identified with BLAST (Altschul et al., 1997), and the read coverage was determined in 50-nucleotide windows.

Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession number Z99767 (*F. trinervia* *GLDPA*).

Supplemental Data

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

Supplemental Figure 1. Analysis of mRNA 5' Ends in Leaves of Transgenic *Arabidopsis* and *F. bidentis* Harboring the *GLDPA*-Ft:GUS Transgene.

Supplemental Figure 2. Functional Analysis of Region 1 of the *GLDPA* Promoter in Transgenic *Arabidopsis* and *F. bidentis*.

Supplemental Figure 3. Distribution of TC-Rich Elements and Initiator-Like Motifs within Region 2 and Region 7 of the *GLDPA* Promoter.

Supplemental Figure 4. Splicing Pattern of the *GLDPA* Transcript Analyzed by 454 Sequencing.

Supplemental Figure 5. Distribution of Upstream Open Reading Frames in P_{R2}-Derived Transcripts.

Supplemental Figure 6. Analysis of mRNA 5' Ends in Leaves of Transgenic *Arabidopsis* and *F. bidentis* Harboring *GLDPA*-Ft-2-7.

Supplemental Table 1. Oligonucleotides Used for the Amplification of the Different *GLDPA*-Ft Promoter Regions or the *GLDPA*-Ft Presequence.

Supplemental Table 2. Oligonucleotide Combinations for the Amplification of the Appropriate Promoter Parts or Gateway-Compatible Fragments.

ACKNOWLEDGMENTS

This work was supported by the "Sonderforschungsbereich 590-Inhärente und adaptive Differenzierungsprozesse" and the "Forschergruppe 1186" funded by the Deutsche Forschungsgemeinschaft.

AUTHOR CONTRIBUTIONS

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

Received November 16, 2011; revised December 23, 2011; accepted January 12, 2012; published January 31, 2012.

REFERENCES

- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- Anderson, L.E. (1971). Chloroplast and cytoplasmic enzymes. II. Pea leaf triose phosphate isomerases. *Biochim. Biophys. Acta* **235**: 237–244.
- Bauwe, H. (2011). Photorespiration: The bridge to C₄ photosynthesis. In *C₄ Photosynthesis and Related CO₂ Concentrating Mechanisms*, Advances in Photosynthesis and Respiration Series, Vol. 32, A.S. Raghavendra and R.F. Sage, eds (Dordrecht, The Netherlands: Springer), pp. 81–108.
- Bauwe, H., Hagemann, M., and Fernie, A.R. (2010). Photorespiration: Players, partners and origin. *Trends Plant Sci.* **15**: 330–336.
- Bauwe, H., and Kolukisaoglu, U. (2003). Genetic manipulation of glycine decarboxylation. *J. Exp. Bot.* **54**: 1523–1535.
- Bernard, V., Brunaud, V., and Lecharny, A. (2010). TC-motifs at the TATA-box expected position in plant genes: A novel class of motifs involved in the transcription regulation. *BMC Genomics* **11**: 166.
- Boisnard-Lorig, C., Colon-Carmona, A., Bauch, M., Hodge, S., Doerner, P., Bancharrel, E., Dumas, C., Haseloff, J., and Berger, F. (2001). Dynamic analyses of the expression of the HISTONE:YFP fusion protein in *Arabidopsis* show that syncytial endosperm is divided in mitotic domains. *Plant Cell* **13**: 495–509.
- Brogna, S., and Wen, J. (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat. Struct. Mol. Biol.* **16**: 107–113.
- Brown, N.J., Newell, C.A., Stanley, S., Chen, J.E., Perrin, A.J., Kajala, K., and Hibberd, J.M. (2011). Independent and parallel recruitment of preexisting mechanisms underlying C₄ photosynthesis. *Science* **331**: 1436–1439.
- Brown, N.J., Parsley, K., and Hibberd, J.M. (2005). The future of C₄ research—Maize, *Flaveria* or *Cleome*? *Trends Plant Sci.* **10**: 215–221.
- Chang, Y.F., Imam, J.S., and Wilkinson, M.F. (2007). The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* **76**: 51–74.
- 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 C₄ plant, *Flaveria bidentis*. *Plant J.* **6**: 949–956.
- 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. (1997). Charakterisierung der Glycinecarboxylase-Gene von *Flaveria trinervia* (C₄) und ihre Expression in transgenen *Nicotiana tabacum*, *Flaveria pubescens* und *Solanum tuberosum*. PhD dissertation (Hannover, Germany: Universität Hannover).
- Cossu, R., and Bauwe, H. (1998). Two genes of the GDCSP gene family from the C₄ plant *Flaveria trinervia*: GDCSPA encoding P-protein and GDCSPB, a pseudo-gene. *Plant Physiol.* **116**: 445–446.
- Curtis, M.D., and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* **133**: 462–469.
- Douce, R., Bourguignon, J., Neuburger, M., and Rébeillé, F. (2001). The glycine decarboxylase system: A fascinating complex. *Trends Plant Sci.* **6**: 167–176.
- Edwards, K., Johnstone, C., and Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res.* **19**: 1349.
- Egecioglu, D.E., and Chanfreau, G. (2011). Proofreading and spell-checking: A two-tier strategy for pre-mRNA splicing quality control. *RNA* **17**: 383–389.
- Engel, N., van den Daele, K., Kolukisaoglu, Ü., Morgenthal, K., Weckwerth, W., Pärnik, 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 C₄ species *Flaveria trinervia*: Analysis of transcriptional control in transgenic *Flaveria bidentis* (C₄) and *Arabidopsis* (C₃). *Plant Physiol.* **146**: 1773–1785.
- Fasken, M.B., and Corbett, A.H. (2009). Mechanisms of nuclear mRNA quality control. *RNA Biol.* **6**: 237–241.
- Foyer, C.H., Bloom, A.J., Queval, G., and Noctor, G. (2009). Photorespiratory metabolism: Genes, mutants, energetics, and redox signaling. *Annu. Rev. Plant Biol.* **60**: 455–484.
- Frohman, M.A., Dush, M.K., and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**: 8998–9002.
- Gowik, U., Bräutigam, A., Weber, K.L., Weber, A.P.M., and Westhoff, P. (2011). Evolution of C₄ photosynthesis in the genus *Flaveria*: How many and which genes does it take to make C₄? *Plant Cell* **23**: 2087–2105.
- Gowik, U., and Westhoff, P. (2011). The path from C₃ to C₄ photosynthesis. *Plant Physiol.* **155**: 56–63.
- 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.D. (1987). C₄ photosynthesis: A unique blend of modified biochemistry, anatomy and ultrastructure. *Biochim. Biophys. Acta* **895**: 81–106.
- Hedtke, B., and Grimm, B. (2009). Silencing of a plant gene by transcriptional interference. *Nucleic Acids Res.* **37**: 3739–3746.
- 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.
- Houseley, J., LaCava, J., and Tollervey, D. (2006). RNA-quality control by the exosome. *Nat. Rev. Mol. Cell Biol.* **7**: 529–539.
- Huang, S., Taylor, N.L., Whelan, J., and Millar, A.H. (2009). Refining the definition of plant mitochondrial presequences through analysis of sorting signals, N-terminal modifications, and cleavage motifs. *Plant Physiol.* **150**: 1272–1285.
- Huang, X., and Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Res.* **9**: 868–877.
- Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: β -Glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**: 3901–3907.

- Joshi, C.P.** (1987). An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Res.* **15**: 6643–6653.
- Kertész, S., Kerényi, Z., Mérai, Z., Bartos, I., Pálfi, 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 the T_L-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.
- 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.
- Lazo, G.R., Stein, P.A., and Ludwig, R.A.** (1991). A DNA transformation-competent Arabidopsis genomic library in *Agrobacterium*. *Biotechnology (N Y)* **9**: 963–967.
- Leegood, R.C., Lea, P.J., Adcock, M.D., and Häusler, R.E.** (1995). The regulation and control of photorespiration. *J. Exp. Bot.* **46**: 1397–1414.
- 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., Ülker, B., and Somssich, I.E.** (2006). An improved method for preparing *Agrobacterium* cells that simplifies the Arabidopsis transformation protocol. *Plant Methods* **2**: 16.
- Majeran, W., Friso, G., Ponnala, L., Connolly, B., Huang, M., Reidel, E., Zhang, C., Asakura, Y., Bhuiyan, N.H., Sun, Q., Turgeon, R., and van Wijk, K.J.** (2010). Structural and metabolic transitions of C₄ leaf development and differentiation defined by microscopy and quantitative proteomics in maize. *Plant Cell* **22**: 3509–3542.
- Mazo, A., Hodgson, J.W., Petruk, S., Sedkov, Y., and Brock, H.W.** (2007). Transcriptional interference: An unexpected layer of complexity in gene regulation. *J. Cell Sci.* **120**: 2755–2761.
- McKown, A.D., Moncalvo, J.M., and Dengler, N.G.** (2005). Phylogeny of *Flaveria* (Asteraceae) and inference of C₄ photosynthesis evolution. *Am. J. Bot.* **92**: 1911–1928.
- Molina, C., and Grotewold, E.** (2005). Genome wide analysis of Arabidopsis core promoters. *BMC Genomics* **6**: 25.
- Morgan, C.L., Turner, S.R., and Rawsthorne, S.** (1993). Coordination of the cell-specific distribution of the four subunits of glycine decarboxylase and serine hydroxymethyltransferase in leaves of C₃-C₄ intermediate species from different genera. *Planta* **190**: 468–473.
- Mouillon, J.M., Aubert, S., Bourguignon, J., Gout, E., Douce, R., and Rébeillé, F.** (1999). Glycine and serine catabolism in non-photosynthetic higher plant cells: Their role in C₁ metabolism. *Plant J.* **20**: 197–205.
- Nakamura, M., Tsunoda, T., and Obokata, J.** (2002). Photosynthesis nuclear genes generally lack TATA-boxes: A tobacco photosystem I gene responds to light through an initiator. *Plant J.* **29**: 1–10.
- Nyikó, T., Sonkoly, B., Mérai, Z., Benkovics, A.H., and Silhavy, D.** (2009). Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. *Plant Mol. Biol.* **71**: 367–378.
- Odell, J.T., Nagy, F., and Chua, N.H.** (1985). Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature* **313**: 810–812.
- Ogren, W.L.** (1984). Photorespiration: Pathways, regulation, and modification. *Annu. Rev. Plant Physiol.* **35**: 415–442.
- Ohnishi, J., and Kanai, R.** (1983). Differentiation of photorespiratory activity between mesophyll and bundle sheath cells of C₄ plants. I. Glycine oxidation by mitochondria. *Plant Cell Physiol.* **24**: 1411–1420.
- Oliver, D.J.** (1994). The glycine decarboxylase complex from plant mitochondria. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **45**: 323–327.
- 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.
- Patel, M., Siegel, A.J., and Berry, J.O.** (2006). Untranslated regions of *FbRbcS1* mRNA mediate bundle sheath cell-specific gene expression in leaves of a C₄ plant. *J. Biol. Chem.* **281**: 25485–25491.
- Powell, A.M.** (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 C₃ and C₃-C₄ intermediate species of *Moricandia*. *Planta* **173**: 298–308.
- Sage, R.F.** (2004). The evolution of C₄ photosynthesis. *New Phytol.* **161**: 341–370.
- Sage, R.F., Christin, P.A., and Edwards, E.J.** (2011). The C₄ plant lineages of planet Earth. *J. Exp. Bot.* **62**: 3155–3169.
- Sambrook, J., and Russell, D.W.** (2001). *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
- Tanudji, M., Sjöling, S., Glaser, E., and Whelan, J.** (1999). Signals required for the import and processing of the alternative oxidase into mitochondria. *J. Biol. Chem.* **274**: 1286–1293.
- Voinnet, O., Rivas, S., Mestre, P., and Baulcombe, D.** (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* **33**: 949–956.
- Waadt, R., and Kudla, J.** (April 1, 2008). *In planta* visualization of protein interactions using bimolecular fluorescence complementation (BiFC). *CSH Protoc.* <http://dx.doi.org/10.1101/pdb.prot4995>.
- Westhoff, P., and Gowik, U.** (2004). Evolution of c4 phosphoenolpyruvate carboxylase. *Genes and proteins: A case study with the genus Flaveria*. *Ann. Bot. (Lond.)* **93**: 13–23.
- 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 C₄ plants. *Planta* **184**: 377–388.
- Yamamoto, Y.Y., Ichida, H., Matsui, M., Obokata, J., Sakurai, T., Satou, M., Seki, M., Shinozaki, K., and Abe, T.** (2007). Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genomics* **8**: 67.
- Yamamoto, Y.Y., Yoshitsugu, T., Sakurai, T., Seki, M., Shinozaki, K., and Obokata, J.** (2009). Heterogeneity of Arabidopsis core promoters revealed by high-density TSS analysis. *Plant J.* **60**: 350–362.
- Yoo, S.D., Cho, Y.H., and Sheen, J.** (2007). Arabidopsis mesophyll protoplasts: A versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**: 1565–1572.
- Yoshimura, Y., Kubota, F., and Ueno, O.** (2004). Structural and biochemical bases of photorespiration in C₄ plants: Quantification of organelles and glycine decarboxylase. *Planta* **220**: 307–317.
- Yudkovsky, N., Ranish, J.A., and Hahn, S.** (2000). A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**: 225–229.
- Zelitch, I., Schultes, N.P., Peterson, R.B., Brown, P., and Brutnell, T.P.** (2009). High glycolate oxidase activity is required for survival of maize in normal air. *Plant Physiol.* **149**: 195–204.
- Zhou, X., Carranco, R., Vitha, S., and Hall, T.C.** (2005). The dark side of green fluorescent protein. *New Phytol.* **168**: 313–322.
- Zuo, Y.C., and Li, Q.Z.** (2011). Identification of TATA and TATA-less promoters in plant genomes by integrating diversity measure, GC-Skew and DNA geometric flexibility. *Genomics* **97**: 112–120.

The authors' contributions

CW wrote this manuscript and performed all experiments except those listed below.

SS performed the fluorescence microscopic analysis of roots of transgenic *Flaveria bidentis* plants harboring the *GLDPA-Ft:H2B-YFP* construct (Figure 1G), analysed the histochemical GUS staining in leaves of transgenic *F. bidentis* plants harboring *GLDPA-Ft-2-7* construct (Figure 7C), carried out the bioinformatic analysis of *GLDPA* transcripts obtained by 454 pyrosequencing (Supplemental Figure 4), and detected mRNA ends by 5' RACE in transgenic *GLDPA-Ft* plants of *F. bidentis* and *Arabidopsis thaliana* (Supplemental Figure 1 [the 5' UTR_{R7S} in [B] were detected by CW] as well as in transgenic *GLDPA-Ft-2-7* plants of *F. bidentis* (Supplemental Figure 6B).

UG performed the 454 pyrosequencing of transcripts of *Flaveria trinervia* (Figure 2B and Supplemental Figure 4) and provided total RNA of leaves of *F. trinervia*.

SE generated the *GLDPA-Ft-1-2-7* construct and provided transgenic seeds of *A. thaliana* plants which had been transformed with the construct *GLDPA-Ft-7* by him.

MK and **MS** performed the transformation of *F. bidentis* plants.

HB provided the *GLDPA-Ft* construct.

PW and **UG** participated in drafting of the manuscript.

Manuscript 1 was published in *The Plant Cell* (Impact Factor: 8.987)

Chapter 2

Manuscript 2

Evolution of C₄ photosynthesis in the genus *Flaveria* – establishment of the photorespiratory
CO₂ pump

Evolution of C₄ Photosynthesis in the Genus *Flaveria* – Establishment of a Photorespiratory CO₂ Pump

Running Head

Evolution of *GLDP* expression in *Flaveria*

Authors:

Stefanie Schulze^{1,2}, Julia Mallmann¹, Janet Burscheidt¹, Maria Koczor¹, Monika Streubel¹, Hermann Bauwe³, Udo Gowik^{1,2}, Peter Westhoff^{1,2,a}

¹ Heinrich-Heine-Universität, Department Biologie, 40225 Düsseldorf, Germany

² Cluster of Excellence on Plant Sciences (CEPLAS) "From Complex Traits towards Synthetic Modules", 40225 Düsseldorf, Germany

³ Universität Rostock, Abteilung Pflanzenphysiologie, Albert-Einstein-Straße 3, 18059 Rostock, Germany

^a To whom correspondence should be addressed: Peter Westhoff, ¹ Department Biologie/Entwicklungs- und Molekularbiologie der Pflanzen, Universitätsstrasse 1, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany. Tel: 49 (0) 211 8112338. Fax: 49 (0) 211 8114871. E-mail: 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).

Estimated length of the article: 14.2 pages

ABSTRACT

C₄ photosynthesis is nature's most efficient answer to the dual activity of ribulose-1,5-bisphosphate carboxylase/oxygenase and the resulting loss of CO₂ by photorespiration. Glycine decarboxylase (GDC) is the key component of photorespiratory CO₂ release in plants and is active in all photosynthetic tissues of C₃ plants, but only in the bundle sheath cells of C₄ plants. The restriction of GDC to the bundle sheath is assumed to be an essential and early step in the evolution of C₄ photosynthesis leading to a photorespiratory CO₂ concentrating mechanism. In this study we analysed how the P protein of GDC (GLDP) became restricted to the bundle sheath during the transition from C₃ to C₄ photosynthesis in the genus *Flaveria*. We found that C₃ *Flaveria* species already contain a bundle sheath expressed *GLDP* gene in addition to a ubiquitously expressed second gene, which became a pseudogene in C₄ *Flaveria* species. Analyses of C₃-C₄ intermediate *Flaveria* species revealed that the photorespiratory CO₂ pump was not established in one single step, but gradually. The knowledge gained by this study sheds new light on the early steps in C₄ evolution.

INTRODUCTION

Ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO), the key enzyme of CO₂ fixation in plants, is a bi-specific 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, CO₂ is released leading to a net loss of photoassimilated CO₂. The loss of CO₂ becomes problematic especially under hot and arid conditions when stomata have to close to avoid water losses (reviewed in Sage, 2004). Under these conditions, CO₂ uptake is drastically reduced, and the relation between photosynthetic CO₂ fixation and photorespiratory CO₂ release becomes unfavourable.

The release of CO₂ is catalysed by glycine decarboxylase (GDC) in the mitochondria of plant cells (Oliver and Raman, 1995). GDC is a multi-protein 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 C₁ 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 knock-out mutant of *Arabidopsis thaliana* that possesses no active GDC

and cannot survive under elevated CO₂, i.e. non-photorespiratory conditions (Engel et al., 2007).

C₄ photosynthesis is one of nature's answers to cope with the oxygenase activity of RubisCO. It is essentially a CO₂ pump that concentrates CO₂ at the site of RubisCO. In the vast majority of all C₄ species, the CO₂-concentrating mechanism requires the close metabolic interaction of two different cells, mesophyll and bundle sheath cells. The bundle sheath cells form a wreath-like layer around the vasculature and harbour RubisCO and the Calvin-Benson cycle. Bundle sheath cells are surrounded by the mesophyll cells, which are devoid of RubisCO but contain phosphoenolpyruvate carboxylase, an oxygen-insensitive carboxylase (Hattersley, 1984; Dengler and Nelson, 1999). The atmospheric CO₂, after conversion to bicarbonate, is initially fixed by phosphoenolpyruvate carboxylase in the mesophyll, resulting in a four-carbon compound, malate and/or aspartate, after which this photosynthetic pathway is named C₄ photosynthesis. The C₄ compound diffuses along its concentration gradient via the plasmodesmata into the bundle sheath cells where it becomes decarboxylated by NADP/NAD malic enzymes or phosphoenolpyruvate carboxykinase (Hatch et al., 1975). The released CO₂ is finally channelled through RubisCO into the Calvin-Benson cycle. Due to the elevated CO₂ concentration at the site of RubisCO, its oxygenase reaction is largely abolished, and photorespiration is drastically reduced in C₄ 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).

C₄ photosynthesis has evolved up to 66 times independently within the angiosperms (Sage et al., 2012). This polyphyletic origin of C₄ photosynthesis suggests that the evolution of a C₃ into a C₄ species must have been relatively easy in genetic terms. The genus *Flaveria* (Powell, 1978) is an attractive model in which to study the transition from C₃ to C₄ photosynthesis. The genus includes not only true C₃ and C₄ species but also a large number of C₃-C₄ intermediate species with a differing degree of "C₄-ness" (Edwards and Ku, 1987; McKown et al., 2005). The evolutionary analysis of the kinetic and regulatory characteristics of C₄ phosphoenolpyruvate 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 C₃ to C₄ 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 absence in mesophyll cells is assumed to constitute a very early and essential step in the evolution towards C_4 photosynthesis (Sage, 2004; Bauwe, 2011; Sage et al., 2012). Restriction of GDC to the bundle sheath cells implicates that the photorespired CO_2 will only be released in the bundle sheath. This results in the establishment of a C_2 photorespiratory CO_2 pump creating a CO_2 -enriched environment for the RubisCO of the bundle sheath, but not of the mesophyll cells. Immunolocalisation 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, P-protein (Rawsthorne et al., 1988; Morgan et al., 1993).

The present study seeks to answer the question how the photorespiratory CO_2 pump was established during the evolution towards C_4 photosynthesis in the genus *Flaveria*. From previous work, we knew that C_4 *Flaveria* species contain 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 having 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 could show that the regulation of *GLDP* expression is complex involving the transcription from two promoters oriented in tandem and most probably also post-transcriptional 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 additionally *de novo* assembled sequences derived from RNAseq of different *Flaveria* species using 454 (Gowik et al., 2011) or Illumina sequencing (Mallmann, in preparation), respectively. The sequences were aligned by Clustal X (see supplementary data 1 online) and gene trees were constructed by using the maximum likelihood method as described in Materials and Methods. Figure 1A shows that the *GLDP*

genes of the various *Flaveria* species group into three clusters, I, II and III. All nine analysed *Flaveria* species contained one gene of each group with the exception of *F. pringlei* in which two group I and two group II genes were found (Figure 1A). It is known that *F. pringlei* is a tetraploid (Cameron et al., 1989), probably arisen by allopolyploidisation with the C₃-C₄ intermediate species *F. angustifolia* (Kopriva et al., 1996; McKown et al., 2005).

The phylogenetic analysis revealed that group I and II *GLDP* genes are more closely related to each other than to group III genes. Group I *GLDP* genes contain *GLDPA* of *F. trinervia* (C₄; formerly *gdcSPA* (Cossu and Bauwe, 1998), accession number Z99767) and *GLDPA* and *GLDPB* of *F. pringlei* (C₃; 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 characterised by sequencing of genomic clones. Group II *GLDP* genes are exemplified by the *GLDPE** pseudogene of *F. trinervia* (C₄; formerly *gdcSPB* (Cossu and Bauwe, 1998), accession number Z99768) and the *GLDPD* and *GLDPE* genes of *F. pringlei* (C₃; formerly *gdcSPD* and *gdcSPE* (Chu, 1996), accession numbers KC545950 and KC545951). The *GLDP* group III contains the *GLDPC* gene of *F. pringlei* (C₃, formerly *gdcSPC* (Chu, 1996), accession number KC545949).

To determine spatial expression patterns of the different *GLDP* genes, we used datasets from Illumina RNAseq experiments of roots, stems and leaves of *F. bidentis* (C₄) and *F. robusta* (C₃) (Mallmann, in preparation). Figure 1B shows that transcripts of group III *GLDP* genes could be detected predominantly in roots and stems in both species, while the *GLDP* group I and group II genes accumulated preferentially in leaves. Accordingly, group I and group II *GLDP* genes are mainly relevant for photorespiration, whereas group III *GLDP* genes most likely are not involved in photorespiration but in the maintenance of basal C₁ metabolism in stems and roots (Hanson and Roje, 2001).

The group I *GLDPA* genes of the C₃ *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 C₄) 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* (including the 5' untranslated region upstream of the AUG codon) were fused to the β -glucuronidase (GUS) reporter gene and transformed into *F. bidentis* (C₄), reporter gene activity was only observed in the bundle sheath cells and to a small degree in the vascular bundle, but not in the

mesophyll tissue (Engelmann et al., 2008). To get an insight how bundle sheath specific *GLDP* expression evolved in the genus *Flaveria*, we analysed the expression specificity of the *GLDPA* genes of *F. pringlei* (C₃) and *F. robusta* (C₃) and additionally that of the *GLDPB* gene of *F. pringlei* (C₃) in both the C₄ plant *F. bidentis* and the C₃ species *A. thaliana*.

The 5' flanking sequences of the *GLDPA* genes of *F. pringlei* (2217 bp; (Bauwe et al., 1995)) and *F. robusta* (1154 bp; accession number KC545947) and of the *GLDPB* gene of *F. pringlei* (2040 bp; (Bauwe et al., 1995)) were fused to the GUS reporter gene and transformed into *F. bidentis* plants. Figure 2A demonstrates that the 5' flanking sequences of the *GLDPA* genes of *F. pringlei* and *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 (cf. Engelmann et al., 2008; Wiludda et al., 2012). No GUS activity could be detected in the mesophyll cells (Figure 2A upper 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* (cf. Engelmann et al., 2008). The 5' flanking sequence of the *GLDPB* gene of *F. pringlei* exhibited a similar expression behaviour in *F. bidentis* as the two *GLDPA* genes (see Supplemental Figure 1A upper panel online).

Transformation systems for C₃ *Flaveria* species are not available. Since the Brassicacean C₃ species *A. thaliana* faithfully recapitulates the expression profile of the *GLDPA* 5' flanking region of the C₄ 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 *GLDPA* 5' flanking region of the C₃ *Flaveria* species *F. pringlei* in transgenic *Arabidopsis* resembles that observed for the C₄ 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 those in *F. bidentis* (Figure 2B). Thus, the expression profile of the *GLDPA* promoter of the C₃ *Flaveria* species *F. pringlei* in *Arabidopsis* is indistinguishable from that of the *GLDPA* promoter of *F. trinervia* in this C₃ species. The 5' flanking regions of the *GLDPA* gene of *F. robusta* and of the *GLDPB* gene of *F. pringlei* function also essentially as bundle sheath/vasculature specific promoters in *Arabidopsis*, however, they exhibit a faint activity in the mesophyll tissue (Figure 2A and Supplemental Figure 1A online).

Taken together, the expression specificities and quantities encoded by the 5' flanking sequences of the group I *GLDP* genes of the two C₃ *Flaveria* species are almost indistinguishable from that of the 5' flanking region of the *GLDPA* gene of the C₄ species *F.*

trinervia. This suggests that the last common ancestor, leading to present C₃ and C₄ *Flaveria* species, contained already a bundle sheath specific *GLDP* gene.

The group II *GLDPD* gene of the C₃ species *F. pringlei* is active in all photosynthetic tissues of both *F. bidentis* (C₄) and *A. thaliana* (C₃)

The group II gene *GLDPE** of *F. trinervia* (C₄) is known to be a pseudogene due to an insertion into the first exon leading to an interruption of the *GLDPE* reading frame (Cossu and Bauwe, 1998). Moreover, the 5' flanking region (1981 bp) of the *GLDPE** gene did not show any promoter activity when analysed in Arabidopsis (see Supplemental Figure 2 online). In order to assess the expression specificity of the 5' flanking sequences of group II *GLDP* genes of C₃ *Flaveria* species, 2733 bp of the 5' flanking region of the *GLDPD* gene of *F. pringlei* were fused to the GUS reporter gene and analysed in transgenic *F. bidentis* (C₄) and *A. thaliana* (C₃). 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 C₃ *Flaveria* species contain at least two *GLDP* genes from groups I and II each that differ in their expression patterns in leaves. While group I *GLDP* genes are expressed bundle sheath specifically/preferentially, group II *GLDP* genes are active in all photosynthetic tissues. During evolution towards C₄, the ubiquitously expressed group II *GLDP* gene was converted into a pseudogene, and hence *GLDP* expression in leaves became bundle sheath specific.

The 5' flanking sequences of both *GLDP* genes of *A. thaliana* do not direct any tissue specificity in leaves

To investigate whether a bundle sheath specific *GLDP* gene might be a common feature of C₃ plants, we analysed the promoters of the two *GLDP* genes of *A. thaliana*. We fused the 1852 bp and 1451 bp 5' flanking sequences of both genes, *GLDP1* (AT4G33010) and *GLDP2* (AT2G26080), to the GUS reporter gene and transformed these constructs into *A. thaliana*. 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 C₃ plants.

Group II *GLDP* genes were progressively inactivated during C₄ evolution, while group I *GLDP* genes show a maximum of expression in C₃-C₄ intermediate species

The genus *Flaveria* with its large number of C₃-C₄ intermediate species offers the unique opportunity to study the steps taken during evolution from C₃ to C₄ photosynthesis. To obtain insight into these changes, the expression of group I and group II *GLDP* genes was compared, by RNAseq, in the leaves of different *Flaveria* species, ranging from C₃ [*F. pringlei* (Fp) and *F. robusta* (Fro)], via C₃-C₄ intermediate species with varying degree of “C₄-ness” [*F. chloraefolia* (Fch), *F. pubescens* (Fpu), *F. anomala* (Fa) and *F. ramossisima* (Fra)], C₄-like [*F. brownii* (Fbr)] to fully fledged C₄ photosynthesis [*F. bidentis* (Fb) and *F. trinervia* (Ft)].

Figure 5 illustrates that the amounts of group II *GLDP* transcripts declined during C₄ evolution. Compared to the C₃ reference, they reached their highest level of reduction in the C₄-like species *F. brownii* and could not be detected anymore in the two C₄ *Flaveria* species. In contrast, group I *GLDP* transcripts increased continuously from C₃ over C₃-C₄ intermediate to the more C₄-like species, but, as to be expected, dropped to lower levels in the true C₄ species.

The tandem promoter structure of the *GLDPA* gene of *F. trinervia* (C₄) is evolutionary conserved

The 5' flanking sequence of the *GLDPA* gene of *F. trinervia* (C₄) 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 tissues (Wiludda et al., 2012). We wanted to know whether this promoter organisation applies also to the 5' flanking regions of *GLDPA* genes from C₃ *Flaveria* species.

The comparison of the 5' flanking sequences of the *GLDPA* genes from the C₃ species *F. pringlei* and *F. robusta* with those from the C₄ species *F. trinervia* and *F. bidentis* (accession number KC545946; Figure 6) revealed a high degree of conservation in regions 2 and 7 which define the distal and proximal promoters, respectively. High sequence similarities 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 *GLDPA* genes of *F. pringlei* and *F. robusta* function as promoters, they were fused to the GUS

reporter gene, and the constructs were transformed into *F. bidentis* (*GLDPA* from *F. pringlei*) and *A. thaliana* (*GLDPA* from *F. pringlei* and *F. robusta*).

Figure 7 illustrates that region 7 of the *GLDPA* 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 lighted up when the two regions from the *GLDPA* gene of *F. pringlei* (Figure 7) or *F. robusta* (see Supplemental Figure 3 online) were analysed in transgenic *A. thaliana*. It follows that the two regions of the 5' flanking sequence of the *GLDPA* genes of *F. pringlei* and of *F. robusta* possess the same promoter activity and specificity as the respective regions of the *GLDPA* 5' flanking sequence of *F. trinervia*. Moreover, it has to be inferred that the promoter structures of the *GLDPA* 5' flanking regions were functionally conserved during C₄ evolution.

The amounts of transcripts derived from the distal sub-promoter of the *GLDPA* gene of C₃ *Flaveria* species are negligible

When assayed by promoter-reporter gene fusions, the distal *GLDPA* sub-promoter of *F. trinervia* (defined by region 2) turned out to be much stronger than the proximal sub-promoter (defined by region 7) (Wiludda et al., 2012). In contrast, the RNA output from the two promoters was just the 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 C₃ and C₃-C₄ intermediate *Flaveria* species, we exploited the available RNAseq data available for C₃, C₃-C₄ and C₄ *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* (C₄), but also of *F. bidentis* (C₄), originated from the proximal transcriptional start site (TSS_{R7}) and that only small but clearly detectable amounts arose from the distal site (TSS_{R2}; Figure 8A and B) In 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 C₃ *Flaveria* species (Figure 8C and D). In C₃-C₄ intermediate species, the amounts of transcripts derived from the distal transcription start site were in between those of the C₄ and C₃ *Flaveria* species (see Supplemental Figure 4A-E online). Taken together, the data indicate that RNA output from the distal promoter of the *GLDPA* 5' flanking region is negligible in the C₃ species, but that it rose to small but clearly detectable amounts during C₄ evolution.

Splicing of transcripts derived from the distal transcriptional start site changed during C₄ evolution

Transcripts originating from the distal transcriptional start site of the *GLDPA* gene of *F. trinervia* contain a large intron of about 1.000 nucleotides that has to be spliced in order to generate functional *GLDPA* mRNAs (Wiludda et al., 2012). These spliced mRNAs 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). We proposed further that the low amounts of spliced mRNAs, which arise from the distal, ubiquitously expressed promoter, function in maintaining a basal level of C₁ metabolism in the mesophyll (Wiludda et al., 2012). Since RNA output from the distal promoter was barely detectable in the C₃ species (Figure 8C and D), but in small amounts in the C₃-C₄ intermediates (see Supplemental Figure 4A-E), we wanted to know when during C₄ evolution the 5' intron and its splicing was established.

Inspection of the RNAseq data did not detect any spliced *GLDPA* RNAs in the two C₃ species and the majority of the C₃-C₄ intermediates. In contrast, in the C₄-like C₃-C₄ intermediate species *F. brownii*, very small amounts of spliced *GLDPA* RNA accumulate (see Supplemental Figure 5 online). As to be expected, splicing of the 5' intron occurs also in the C₄ species *F. bidentis* (see Supplemental Figure 5 online). A functional spliced 5' intron was therefore not only observed in the C₄ *Flaveria* species *F. bidentis* and *F. trinervia*, but also in the C₄-like C₃-C₄ intermediate species *F. brownii* indicating that its presence is a typical feature of C₄ and C₄-like *Flaveria* species.

Interestingly, the splice acceptor sites differ in the two C₄ species (see Supplemental Figure 5 online). While the splice acceptor site of the *GLDPA* gene of *F. trinervia* is located about 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 about 15 nucleotides upstream (see Supplemental Figure 5 online). We conclude that the splicing of the 5' intron evolved independently during C₄ speciation in *Flaveria*.

DISCUSSION

The photorespiratory CO₂ pump is considered to be an essential and early step in the evolutionary trajectory towards C₄ photosynthesis (Sage, 2004; Bauwe, 2011; Sage et al.,

2012). The establishment of this pump requires that the glycine-decarboxylating step of the photorespiratory pathway, which is carried out by GDC, becomes restricted to the bundle sheath cells. The confinement of GDC to the bundle sheath cells implies that RubisCO in the bundle sheath cells operates at higher CO₂ levels than RubisCO in the mesophyll compartment (Bauwe, 2011). At the end of C₄ evolution, both RubisCO and the core photorespiratory pathway are entirely compartmentalized in the bundle sheath cells. Early studies with C₃-C₄ intermediate species (Morgan et al., 1993) indicated that the P-protein of GDC (GLDP), the actual decarboxylase, is responsible for this re-allocation of glycine decarboxylase activity. In all cases of C₃-C₄ intermediates investigated, GLDP is exclusively found in the bundle sheath cells (Hylton et al., 1988; Rawsthorne et al., 1988; Morgan et al., 1993; Turner et al., 1993; Muhaidat et al., 2011). These findings led to the conclusion that the - complete - restriction of *GLDP* expression to the bundle sheath cells was necessary for the establishment of the photorespiratory CO₂ pump and that this compartmentation step occurred rather early during C₄ evolution (Monson et al., 1984; 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, by using the genus *Flaveria* as an evolutionary model, provides conclusive answers to both questions.

C₃ *Flaveria* species contain a bundle sheath cell specific *GLDP* gene

Two scenarios can be imagined how to evolve a bundle sheath specific expression of *GLDP* during the transition from C₃ to C₄. Firstly, a ubiquitously expressed *GLDP* gene changed its expression behaviour to become bundle sheath specific. Alternatively, a *GLDP* gene with the requested bundle sheath specificity of expression was already present in C₃ species, and an additional, ubiquitously expressed *GLDP* gene became inactivated during C₄ evolution. The first scenario is best exemplified by the evolution of mesophyll expression specificity in case of the *ppcA* phosphoenolpyruvate carboxylase gene of the genus *Flaveria*. The *ppcA* gene is ubiquitously expressed in all leaf tissues of the C₃ *Flaveria* species, while its orthologue of the C₄ species is expressed only in the mesophyll cells. To generate a C₄ mode of expression, a *cis*-regulatory module of 41 base pairs located in the distal 5' flanking region was converted by two nucleotide changes into a bundle sheath repressing module of gene expression, and as a consequence, *ppcA* transcription became confined to mesophyll cells (Stockhaus et al., 1997; Gowik et al., 2004; Akyildiz et al., 2007).

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

The occurrence of bundle sheath specific *GLDP* genes is not a common feature of C_3 plants. The C_3 species *A. thaliana*, for instance, contains two *GLDP* genes both of which are expressed similarly in all chlorenchyma of the leaf 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 pre-conditioning syndrome that distinguishes C_3 taxa that evolved C_4 photosynthesis from others that did not (Sage, 2004; Sage et al., 2011; Sage et al., 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 ubiquitously expressed *GLDP* genes, in a C_3 species facilitated the evolution of C_3 - C_4 intermediate photosynthesis. This is because a knock-out or the drastic down-regulation 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).

The grasses are probably the oldest angiosperm lineage in which C_4 species evolved (Edwards et al., 2010) and therefore illustrate how fully optimised C_4 species finally look like with respect to metabolic organization and the underlying transcriptional regulation. While the genome of the C_3 grass *Oryza sativa* (rice) contains two *GLDP* genes (Goff et al., 2002) (LOC_Os01g51410; LOC_OS06g40940), the genomes of the C_4 grasses *Zea mays* (Schnable et al., 2009) (GRMZM2G104310), *Sorghum bicolor* (Paterson et al., 2009) (Sb08g003440) and *Setaria italica* (Bennetzen et al., 2012) (Si000068m) harbour 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 *Flaveria* species ranging from C₃ through C₃-C₄ intermediate to C₄ photosynthesis. Our investigations demonstrated that glycine decarboxylase disappeared from the mesophyll cells not abruptly, but gradually. We showed that the ubiquitously expressed group II *GLDP* genes were progressively inactivated in C₃-C₄ intermediates and C₄-like species and completely shut down in C₄ *Flaveria* species (Figure 5). In contrast, the bundle sheath specific group I *GLDP* genes were found to stay active, and their expression was even enhanced in the C₃-C₄ intermediates.

Is it plausible that the photorespiratory CO₂ pump was not established abruptly (Sage, 2004), but step by step? One can imagine that the capacities to efficiently decarboxylate large amounts of glycine and recapture the correspondingly large amounts of photorespiratory CO₂ were not *ab initio* present in the bundle sheaths of C₃ ancestors of contemporary C₄ plants. Indeed, bundle sheath cells of present C₃ species with “Proto-Kranz” anatomy (Muhaidat et al., 2011; Sage et al., 2012) are still relatively poor in chloroplasts and mitochondria (Muhaidat et al., 2011; Sage et al., 2012). If such a C₃ species would abruptly lose all its glycine decarboxylation activity in the mesophyll, it would most probably not be viable anymore. A gradual reduction of glycine decarboxylation in the mesophyll cells could initiate a series of steps organised in a positive feedback loop (Bauwe, 2011; Muhaidat et al., 2011; Sage et al., 2012). Glycine had to diffuse to the bundle sheath for decarboxylation thereby creating a higher CO₂ concentration around RubisCO in the bundle sheath. The RubisCO in the bundle sheath would become more active in CO₂ 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 glycine decarboxylation activity could then be shifted to the bundle sheath cells, and concomitantly the number of mitochondria would increase. The final outcome of this positive feedback loop would be a higher CO₂ re-fixation capacity than before. Of course, other evolutionary adaptations had to proceed in parallel, for instance the optimisation of organelle positioning in the bundle sheath cells, i.e. centripetal mitochondria and centrifugal chloroplasts, the up-regulation of inter- and intra-cellular glycine and serine transport, and/or changes in overall leaf anatomy.

The tandem promoter structure of *GLDPA* is conserved in both C₃ and C₄ *Flaveria* species, but the RNA output is not

The 5' flanking sequences of the group I *GLDPA* genes of C_3 *Flaveria* species are very similar to those of their counterparts in the C_4 species. The 5' introns are found at corresponding positions, and the distal and proximal *GLDPA* promoter regions are highly conserved (Figure 6). In fact, both regions from 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* (cf. Wiludda et al., 2012; Figure 7 and Supplemental Figure 3 online). In 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 *Flaverias* (Figure 8).

One may conclude from these findings that the distal *GLDPA* promoter is silent in the context of the authentic 5' flanking region in the C_3 *Flaverias*, whereas its counterpart in the C_4 *Flaverias* 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). On the other side, the different RNA output from the distal promoter in C_3 vs. C_4 *Flaveria* species may not be regulated transcriptionally but post-transcriptionally at the level of transcript stability. Indeed, tentative evidence indicates that the 5' intron is involved in controlling 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 the 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 able of synthesizing C_1 compounds regardless whether it photorespires or not (Hanson and Roje, 2001; Bauwe, 2011). A complete shut-down of glycine decarboxylase in the mesophyll cells of C_4 plants would thus be fatal (Bauwe, 2011). The promoter could be tight in C_3 and C_3 - C_4 intermediate species, because these plants possess a group II *GLDP* gene that is active in all leaf chlorenchyma cells (Figure 5) and consequently no selective pressure would favour 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 C₄ evolution (Christin et al., 2011a), and the large numbers of C₃-C₄ intermediate species (Edwards and Ku, 1987; McKown et al., 2005) suggest that evolution towards C₄ is still ongoing in *Flaveria*. We have used *Flaveria* to analyse the evolutionary trajectory towards the establishment of a photorespiratory CO₂ pump and its further integration into the C₄ pathway. The model derived from these studies shows that the photorespiratory CO₂ pump evolved step by step and that this gradual evolution was eased by the presence of duplicated *GLDP* genes differing in expression specificity (see Figure 9).

We do not know whether this evolutionary scenario is unique for *Flaveria* or whether it represents a general model for C₃-to-C₄ transitions. It would be worthwhile, therefore, to study other genera that are phylogenetically different, contain both C₃ and C₄ species, and ideally also C₃-C₄ intermediates, as for instance *Mollugo* (Christin et al., 2011b), *Cleome* (Marshall et al., 2007) or *Heliotropium* (Muhaidat et al., 2011). An alternative approach would pursue synthetic experimental evolution (Morange, 2009) by using C₃ model plants such as *Arabidopsis* that can easily be manipulated by genetic engineering and are very convenient for multiple rounds of mutation and selection due to their fast life cycles. The optimisation of photosynthesis by placing a C₄ pathway into current C₃ species is on the agenda of crop biologists (Hibberd et al., 2008; von Caemmerer et al., 2012), setting up a photorespiratory CO₂ pump by synthetic experimental evolution is an important and necessary component of this endeavour.

METHODS

Sequence alignments and phylogenetic analyses

Sequences used for construction of the phylogenetic tree of the *GLDP* genes of the *Flaveria* species were obtained from either known sequences (*GLDPA*-Ft (Cossu and Bauwe, 1998); *GLDPE**-Ft (Cossu and Bauwe, 1998); *GLDPA*-Fp (Bauwe et al., 1995); *GLDPB*-Fp (Bauwe et al., 1995); *GLDPC*-Fp (Chu, 1996); *GLDPD*-Fp (Chu, 1996) and *GLDPE*-Fp (Chu, 1996)) or from contigs assembled from either 454 (Gowik et al., 2011) or Illumina (Mallmann, in preparation) sequencing. Alignments of sequences for phylogenetic analyses were carried out 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 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 *F. pringlei*, *F. robusta*, *F. chloraefolia*, *F. pubescens*, *F. anomala*, *F. ramossisima*, *F. brownii*, *F. bidentis* and *F. trinervia* (Mallmann, in preparation) were mapped against the sequences of the *GLDPA* gene of *F. trinervia* and the *GLDPE** gene of *F. trinervia* to obtain an overall distribution of reads along the 5' flanking and coding regions of the genes. The 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*, *GLDPC* and *GLDPD* transcripts in roots, stems and leaves of *F. bidentis* and *F. robusta*, the coding sequences of *GLDPA*-Ft, *GLDPC*-Fro and *GLDPD*-Fro (the latter two obtained from full-length contigs from assembly of Illumina reads (Mallmann, in preparation) 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. The sequences for *GLDPC*-Fro and *GLDPD*-Fro are available as supplemental data 2 online.

Isolation of *GLDP* 5' flanking regions by vectorette PCR

5' flanking regions of *GLDP* genes from *F. robusta*, *F. bidentis* and *F. pubescens* 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 (Clontech, Mountain View, US). DNA for library construction was isolated from leaves of the respective *Flaveria* species with the Qiagen DNeasy Plant Mini Kit (Qiagen, Hilden, GER).

Cloning of promoter-reporter gene constructs

The 5' regions of *GLDPA*-Fp and *GLDPD*-Fp genes were fused to the β -glucuronidase (*uidA*, *GUS*) 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). Detailed sequences of the used oligonucleotides can be found in supplemental table 1 online.

Transformation of *Arabidopsis thaliana* and *Flaveria bidentis*

Transformation of *Arabidopsis thaliana* was carried out following the “Floral Dip” protocol (Clough and Bent, 1998) as modified by Logeman et al. (Logemann et al., 2006). Strain GV3101 of *Agrobacterium tumefaciens* (Holsters et al., 1980; Koncz and Schell, 1986) provided the helper plasmid for the transformations. *Flaveria bidentis* was transformed as described by Chitty et al. (Chitty et al., 1994) using *Agrobacterium tumefaciens* strain AGL1 (Hood et al., 1986; Lazo et al., 1991).

***In situ* detection of GUS and fluorimetric activity measurements**

Fluorimetric measurements of β -glucuronidase activity were carried out according to Jefferson et al. (Jefferson et al., 1987) and Kosugi et al. (Kosugi et al., 1990). In the case of *Flaveria bidentis*, the 5th leaf of a 40 to 50 cm tall T₀ plant was harvested for the analysis, in the case of *Arabidopsis thaliana*, three rosette leaves 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 was performed as described in Engelmann et al. (Engelmann et al., 2008).

Accession Numbers

Sequence data from this article are available in the EMBL/GenBank data libraries under the accession numbers Z36879 (*F. pringlei* GLDPB; *gdc*PA), Z54239 (*F. pringlei* GLDPA; *gdc*PB), KC545949 (*F. pringlei* GLDPC; *gdc*PC), KC545950 (*F. pringlei* GLDPD; *gdc*PD), KC545951 (*F. pringlei* GLDPE, *gdc*PE) Z99767 (*F. trinervia* GLDPA, *gdc*PA) and Z99768 (*F. trinervia* GLDPE*; *gdc*PB pseudogene), KC545946 (5' flanking sequence of *F. bidentis* GLDPA), KC545947 (5' flanking sequence of *F. robusta* GLDPA). The *A. thaliana* genes GLDP1 (AT4G33010) and GLDP2 (AT2G26080) can be found at TAIR (www.arabidopsis.org).

Supplemental Data

Supplemental Figure 1. Functional analysis of the 5' flanking sequence of the *GLDPB* gene of *F. pringlei*.

Supplemental Figure 2. Functional analysis of the 5' flanking sequence of *GLDPE** gene of *F. trinervia*.

Supplemental Figure 3. Functional analysis of the proximal (P_{R7}) and distal (P_{R2}) sub-promoters of the *GLDPA* gene of the C₃ species *F. robusta*.

Supplemental Figure 4. Transcript coverage of the *GLDPA* genes of the C₃-C₄ intermediates *F. chloraefolia*, *F. pubescens*, *F. anomala*, *F. ramossisima* and *F. brownii*.

Supplemental Figure 5. Splice variants of *GLDPA* transcripts derived from the distal transcriptional start site.

Supplemental Data 1. ClustalX alignment of the *GLDP* sequences used for the phylogenetic analysis.

Supplemental Data 2. Sequences of the *GLDPC* and *GLDPD* genes of *F. robusta*.

Supplemental Table 1. List of oligonucleotides used for amplification of promoter sequences for GUS constructs and for the amplification of the 5' splice site of transcripts derived from distal transcriptional start site of *F. bidentis*.

Acknowledgements

We thank Dr. Borjana Arsova for carefully reading this manuscript and giving usefull comments on this manuscript.

This work was funded by the Deutsche Forschungsgemeinschaft (DFG) through the Sonderforschungsbereich SFB 590 and the Forschergruppe FOR1186 (Promics).

Authors' 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. analysed the data.

S.S., H.B., U.G. and P.W. wrote the manuscript.

FIGURE LEGENDS

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 based on 203 nucleotide positions, starting at the ATG, which were aligned by using ClustalX 2.0.8

(Higgins and Sharp, 1988, 1989; Thompson et al., 1997; Larkin et al., 2007). Bootstrap values (1000 replicates) are shown next to the branches (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. **(B)** Abundance of *GLDPA* (group I), *GLDPC* (group III) and *GLDPD* (group II) transcripts in stems, roots and leaves of *F. bidentis* (Fb; C₄) and *F. robusta* (Fro; C₃) as measured by mapping RNAseq Illumina reads on the respective cDNAs.

Figure 2. Functional analysis of the 5' flanking sequences of the group I *GLDPA* genes of the C₃ species *F. pringlei* and *F. robusta*.

The 5' flanking sequences of *GLDPA*-Fp (2217 bp) and *GLDPA*-Fro (1154 bp) were fused to the GUS reporter gene and analysed in transgenic *F. bidentis* (C₄) and transgenic *A. thaliana* (C₃). **(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.

Figure 3. Functional analysis of the 5' flanking sequence of the group II *GLDPD* gene of the C₃ species *F. pringlei*.

The 5' flanking sequence of *GLDPD*-Fp (2733 bp) as a representative of group II *GLDP* genes of C₃ *Flaverias* was fused to the GUS reporter gene and analysed in transgenic *F. bidentis* (C₄) and transgenic *A. thaliana* (C₃). **(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.

Figure 4. Functional analysis of the 5' flanking sequences of *GLDP1* and *GLDP2* genes of *A. thaliana*.

The 5' flanking sequences of *GLDP1* (1852 bp) and *GLDP2* (1451 bp) of *A. thaliana* (C₃) were fused to the GUS reporter gene and analysed in transgenic *A. thaliana* (C₃). **(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.

Figure 5. Transcript abundance of group I and group II *GLDP* genes in C₃, C₃-C₄ intermediate and C₄ *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 I *GLDP* genes are displayed in green and transcripts abundances of group II 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*.

Figure 6. Schematic comparison of the 5' flanking sequences of group I *GLDP* genes from C₄ and C₃ *Flaveria* species.

The 5' flanking sequence of the *GLDPA* gene of *F. trinervia* was divided into seven functionally characterised regions (Engelmann et al., 2008). Regions 2 (orange) and 7 (dark green) contain the two sub-promoters, P_{R2} and P_{R7}. Region 1 (light green) enhances the activities of both sub-promoters. Region 3 (red) is required to suppress the mesophyll activity of the sub-promoter P_{R2}, but only in Arabidopsis. Regions 4 to 6 are not required for promoter activity. The 5' flanking sequences of the *GLDPA* genes from *F. bidentis* (C₄), *F. pringlei* (C₃) and *F. robusta* (C₃) are colour-labelled 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 sub-promoters P_{R7} and P_{R2} of the *GLDPA* gene of *F. pringlei*.

Regions 1-2 (*GLDPA*-Fp1-2; 641 bp) and 7 (*GLDPA*-Fp7; 318 bp) of the 5' flanking region of the *GLDPA* gene of *F. pringlei* (C₃) were fused to the GUS reporter gene and analysed in transgenic *F. bidentis* (C₄) and transgenic *A. thaliana* (C₃). (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.

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 (A) *F. trinervia* (C₄), (B) *F. bidentis* (C₄), (C) *F. pringlei* (C₃) and (D) *F. robusta* (C₃) were mapped on the sequence of the *GLDPA* gene of *F. trinervia* including its 5' flanking sequence. The numbers of reads covering each position of the gene sequence were counted.

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. Sub-functionalization remodelled the expression of the group I *GLDP* genes to become bundle sheath specific and led to an ancestral C_3 species with the same spatial *GLDP* expression pattern as of today's C_3 species. During transition to C_3 - C_4 intermediate photosynthesis, the expression of group II *GLDP* genes level was reduced and the distal *GLDPA* sub-promoter became activated. In the end, group II *GLDP* genes were inactivated by pseudogenization. Green: group I *GLDP* spatial expression; yellow: group II *GLDP* spatial expression.

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). Chapter 6 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 (Springer Netherlands), pp. 81-108.
- Bauwe, H., and Kopriva, S. (1995). The *gdcsPA* gene from *Flaveria pringlei* (Asteraceae). *Plant Physiol* **107**, 655.
- 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.
- Bennetzen, J.L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A.C., Estep, M., Feng, L., Vaughn, J.N., Grimwood, J., Jenkins, J., Barry, K., Lindquist, E., Hellsten, U., Deshpande, S., Wang, X., Wu, X., Mitros, T., Triplett, J., Yang, X., Ye, C.-Y., Mauro-Herrera, M., Wang, L., Li, P., Sharma, M., Sharma, R., Ronald, P.C., Panaud, O., Kellogg, E.A., Brutnell, T.P., Doust, A.N., Tuskan, G.A., Rokhsar, D., and Devos, K.M. (2012). Reference genome sequence of the model plant *Setaria*. *Nat Biotech* **30**, 555-561.

- Bevan, M.** (1984). Binary *Agrobacterium* vectors for plant transformation. *Nucleic Acids Res* **12**, 8711-8721.
- Brognia, 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 C₄ 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. *Molecular Breeding* **11**, 287-293.
- Chitty, J.A., Furbank, R.T., Marshall, J.S., Chen, Z., and Taylor, W.C.** (1994). Genetic transformation of the C₄ plant, *Flaveria bidentis*. *Plant J* **6**, 949-956.
- Christin, P.A., Osborne, C.P., Sage, R.F., Arakaki, M., and Edwards, E.J.** (2011a). C₄ eudicots are not younger than C₄ 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 C₃-C₄ 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* C₃ and *Flaveria anomala* C₃-C₄. [PhD thesis] (Martin-Luther University, Halle-Wittenberg).
- 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 C₄ plants. In *C₄ plant biology*, S.R. F and R.K. Monson, eds (San Diego: Academic Press), pp. 133-172.
- Edwards, E.J., Osborne, C.P., Stromberg, C.A., Smith, S.A., Bond, W.J., Christin, P.A., Cousins, A.B., Duvall, M.R., Fox, D.L., Freckleton, R.P., Ghannoum, O., Hartwell, J., Huang, Y., Janis, C.M., Keeley, J.E., Kellogg, E.A., Knapp, A.K., Leakey, A.D., Nelson, D.M., Saarela, J.M., Sage, R.F., Sala, O.E., Salamin, N., Still, C.J., and Tipple, B.** (2010). The origins of C₄ grasslands: integrating evolutionary and ecosystem science. *Science* **328**, 587-591.

- Edwards, G., and Ku, M.** (1987). Biochemistry of C₃-C₄ 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 C₄ species *Flaveria trinervia*: analysis of transcriptional control in transgenic *Flaveria bidentis* (C₄) and *Arabidopsis* (C₃). *Plant Physiol* **146**, 1773-1785.
- Felsenstein, J.** (1985). Confidence limits on phylogenies : an approach using the bootstrap. *Evolution* **39**, 783-791.
- Goff, S.A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.-l., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., and Briggs, S.** (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 C₄ photosynthesis in the genus *Flaveria*: how many and which genes does it take to make C₄? *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 C₄ plant *Flaveria trinervia*, the promoter of the C₄ phosphoenolpyruvate 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 C₄-pathway species based on differing C₄ acid decarboxylating systems and ultrastructural features. *Aust J Plant Physiol* **2**, 111-128.
- Hatch, M.D.** (1987). C₄ photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochem Biophys Acta* **895**, 81-106.
- Hattersley, P.W.** (1984). Characterization of C₄ type leaf anatomy in grasses (Poaceae). Mesophyll: bundle sheath area ratios. *Ann Bot* **53**, 163-180.
- Hibberd, J.M., Sheehy, J.E., and Langdale, J.A.** (2008). Using C₄ 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., Silva, B., Van Vliet, F., Genetello, C., De Block, M., Dhaese, P., Depicker, A., Inze, D., Engler, G., Villarroel, R., and 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.
- Hylton, C.M., Rawsthorne, S., Smith, A.M., Jones, D.A., and Woolhouse, H.W.** (1988). Glycine decarboxylase is confined to the bundle-sheath cells of leaves of C₃-C₄ intermediate species. *Planta* **175**, 452-459.
- 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 Science* **70**, 133-140.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G.** (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* **9**, 963-967.
- Levine, M.** (2011). Paused RNA polymerase II as a developmental checkpoint. *Cell* **145**, 502-511.
- Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S.L., Kebrom, T.H., Provar, N., Patel, R., Myers, C.R., Reidel, E.J., Turgeon, R., Liu, P., Sun, Q., Nelson, T., and Brutnell, T.P.** (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 C₃ to C₄ photosynthesis. *Plant J* **51**, 886-896.
- McKown, A.D., Moncalvo, J.M., and Dengler, N.G.** (2005). Phylogeny of *Flaveria* (Asteraceae) and inference of C₄ photosynthesis evolution. *Am J Bot* **92**, 1911-1928.
- Monson, R.K., Edwards, G.E., and Ku, M.S.B.** (1984). C₃-C₄ intermediate photosynthesis in plants. *BioScience*, 563-574.
- Morange, M.** (2009). Synthetic biology: a bridge between functional and evolutionary biology. *Biological 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 C₃-C₄ 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 C₃-C₄ 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., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Ollilar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Mehboob ur, R., Ware, D., Westhoff,**

- P., Mayer, K.F., Messing, J., and Rokhsar, D.S.** (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 C₃ and C₃-C₄ intermediate species of *Moricandia*. *Planta* **173**, 298-308.
- Sage, R.F.** (2004). The evolution of C₄ photosynthesis. *New Phytologist* **161**, 341-370.
- Sage, R.F., Christin, P.A., and Edwards, E.J.** (2011). The C₄ 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 C₄ photosynthesis. *Annu Rev Plant Biol* **63**, 19-47.
- Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L., Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E., Du, F., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., Van Buren, P., Vaughn, M.W., Ying, K., Yeh, C.T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.M., Deragon, J.M., Estill, J.C., Fu, Y., Jeddelloh, J.A., Han, Y., Lee, H., Li, P., Lisch, D.R., Liu, S., Liu, Z., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C.,**

- Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A., and Wilson, R.K.** (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 Research* **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 C₄ form of phosphoenolpyruvate carboxylase directs mesophyll-specific expression in transgenic C₄ *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.
- Turner, S.R., Hellens, R., Ireland, R., Ellis, N., and Rawsthorne, S.** (1993). The organisation and expression of the genes encoding the mitochondrial glycine decarboxylase complex and serine hydroxymethyltransferase in pea (*Pisum sativum*). *Mol Gen Genet* **236**, 402-408.
- von Caemmerer, S., Quick, W.P., and Furbank, R.T.** (2012). The development of C₄ rice: current progress and future challenges. *Science* **336**, 1671-1672.
- 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 C₄ *Flaveria*: an intricate interplay of transcriptional and posttranscriptional processes. *Plant Cell* **24**, 137-151.

FIGURES

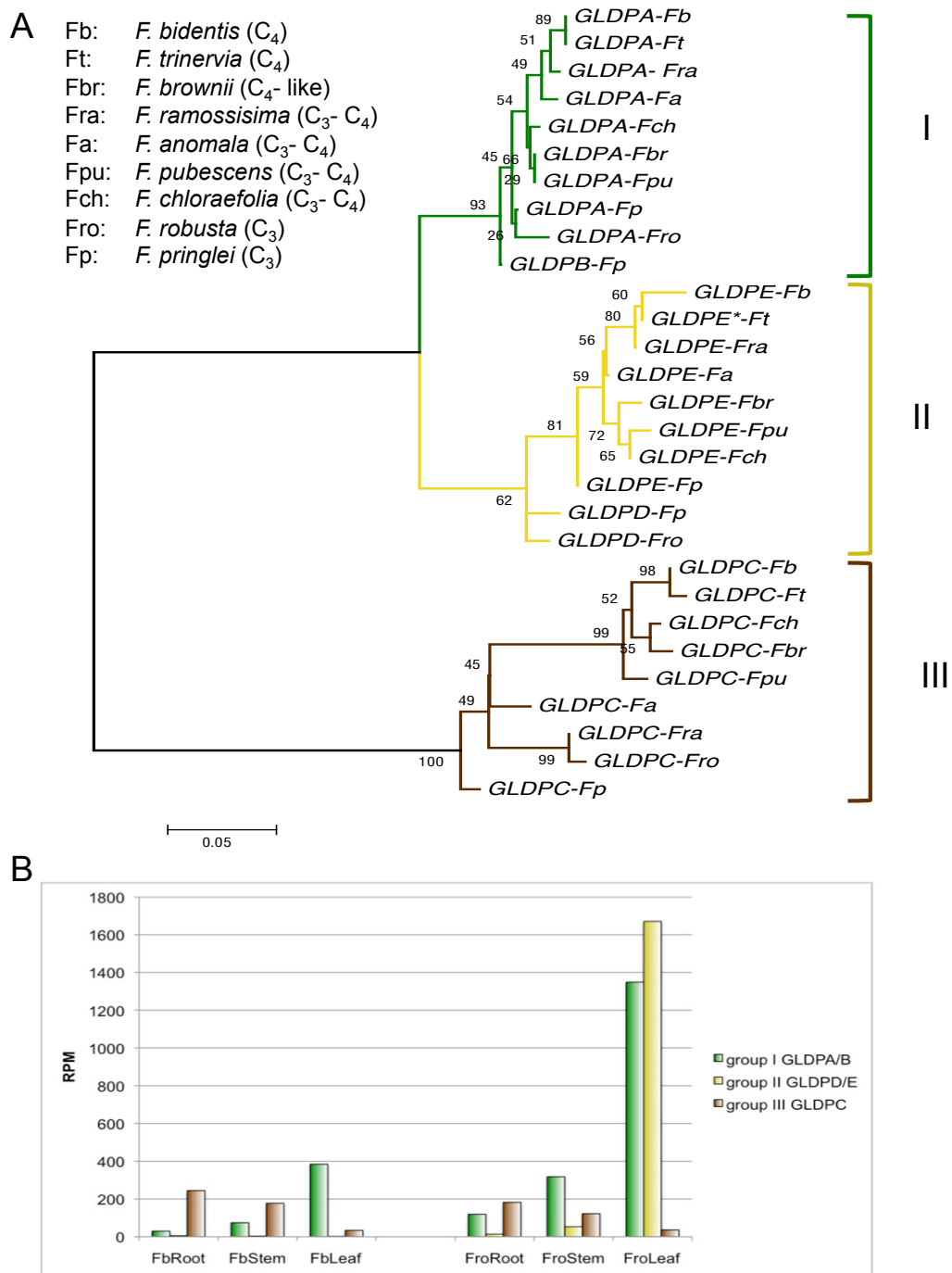


Figure 1

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 based on 203 nucleotide positions, starting at the ATG, which were aligned by using ClustalX 2.0.8 (Higgins and Sharp, 1988, 1989; Thompson et al., 1997; Larkin et al., 2007). Bootstrap values (1000 replicates) are shown next to the branches (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. **(B)** Abundance of *GLDPA* (group I), *GLDPC* (group III) and *GLDPD* (group II) transcripts in stems, roots and leaves of *F. bidentis* (Fb; C₄) and *F. robusta* (Fro; C₃) as measured by mapping RNAseq Illumina reads on the respective cDNAs.

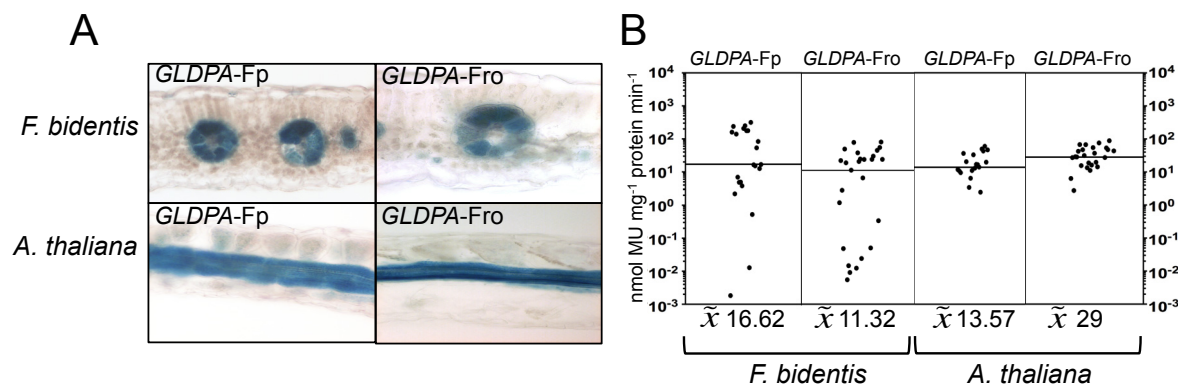


Figure 2

Figure 2. Functional analysis of the 5' flanking sequences of the group I *GLDPA* genes of the *C*₃ species *F. pringlei* and *F. robusta*.

The 5' flanking sequences of *GLDPA*-Fp (2217 bp) and *GLDPA*-Fro (1154 bp) were fused to the GUS reporter gene and analysed in transgenic *F. bidentis* (*C*₄) and transgenic *A. thaliana* (*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.

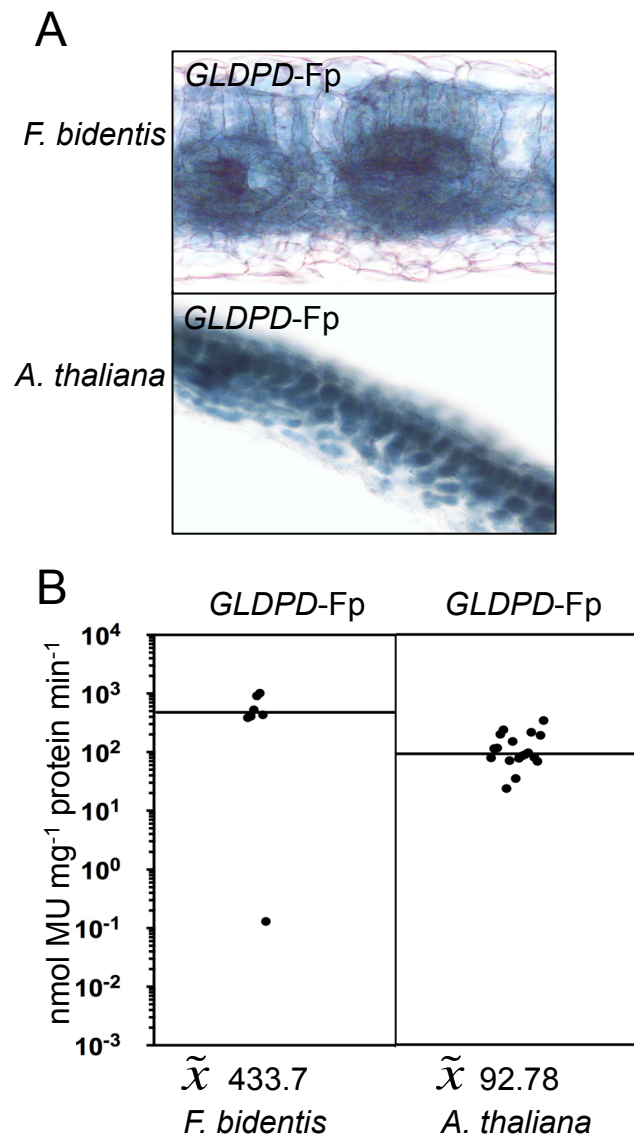


Figure 3

Figure 3. Functional analysis of the 5' flanking sequence of the group II *GLDPD* gene of the *C*₃ species *F. pringlei*.

The 5' flanking sequence of *GLDPD*-Fp (2733 bp) as a representative of group II *GLDP* genes of *C*₃ *Flaverias* was fused to the GUS reporter gene and analysed in transgenic *F. bidentis* (*C*₄) and transgenic *A. thaliana* (*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.

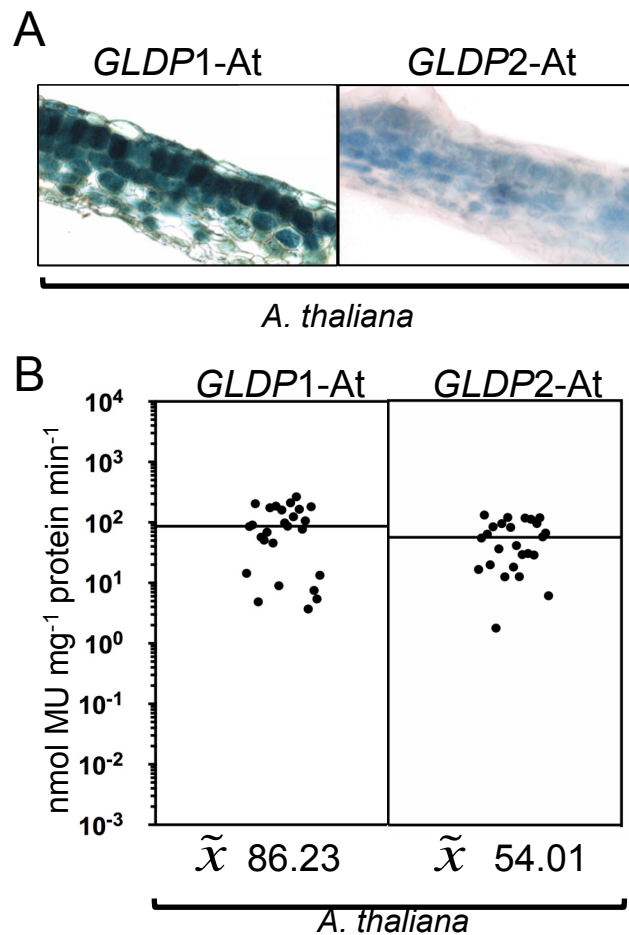


Figure 4

Figure 4. Functional analysis of the 5' flanking sequences of *GLDP1* and *GLDP2* genes of *A. thaliana*.

The 5' flanking sequences of *GLDP1* (1852 bp) and *GLDP2* (1451 bp) of *A. thaliana* (C₃) were fused to the GUS reporter gene and analysed in transgenic *A. thaliana* (C₃). (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.

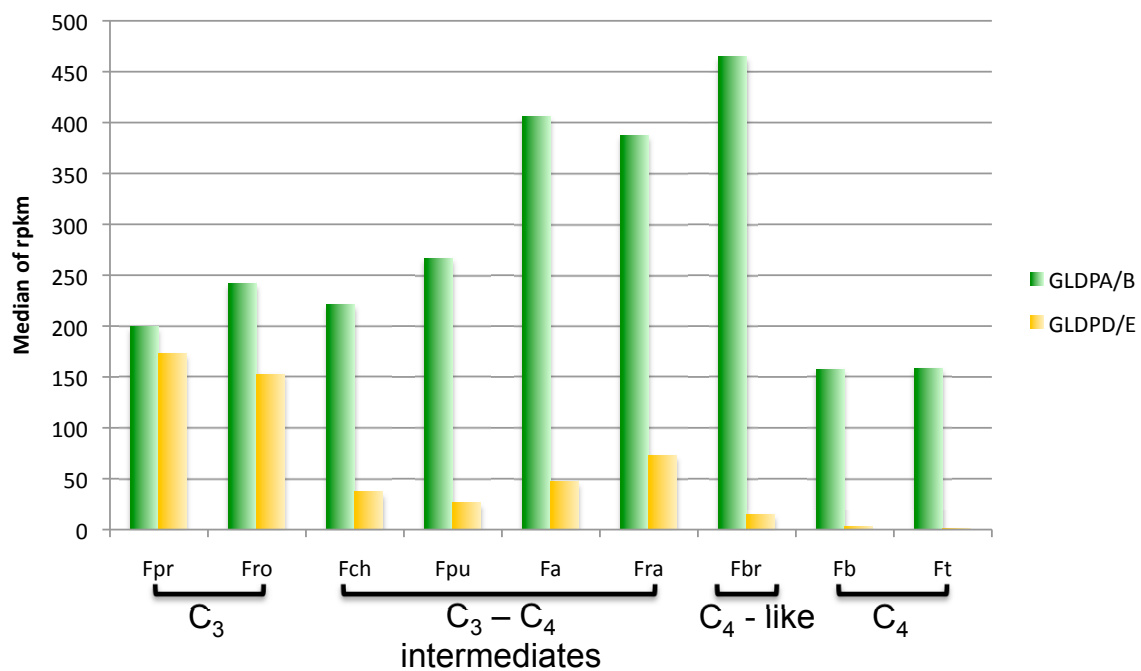


Figure 5

Figure 5. Transcript abundance of group I and group II *GLDP* genes in C₃, C₃-C₄ intermediate and C₄ *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 I *GLDP* genes are displayed in green and transcripts abundances of group II 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*.

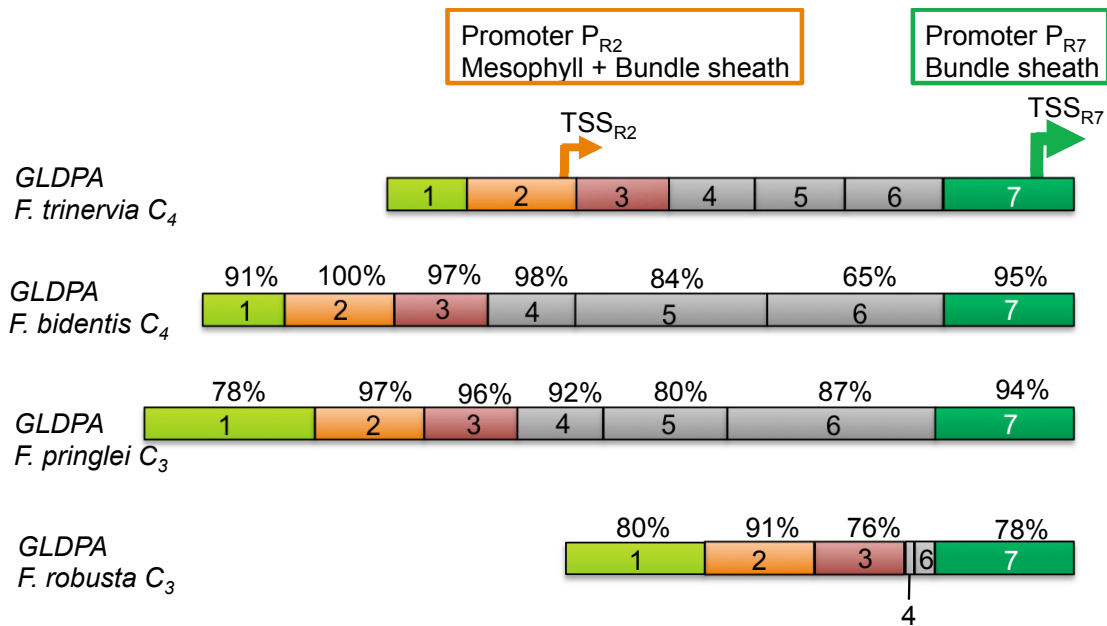


Figure 6

Figure 6. Schematic comparison of the 5' flanking sequences of group I *GLDP* genes from C₄ and C₃ *Flaveria* species.

The 5' flanking sequence of the *GLDPA* gene of *F. trinervia* was divided into seven functionally characterised regions (Engelmann et al., 2008). Region 2 (orange) and 7 (dark green) contain the two sub-promoters, P_{R2} and P_{R7}. Region 1 (light green) enhances the activities of both sub-promoters. Region 3 (red) is required to suppress the mesophyll activity of the sub-promoter P_{R2}, but only in *Arabidopsis*. Regions 4 to 6 are not required for promoter activity. The 5' flanking sequences of the *GLDPA* genes from *F. bidentis* (C₄), *F. pringlei* (C₃) and *F. robusta* (C₃) are colour-labelled 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.

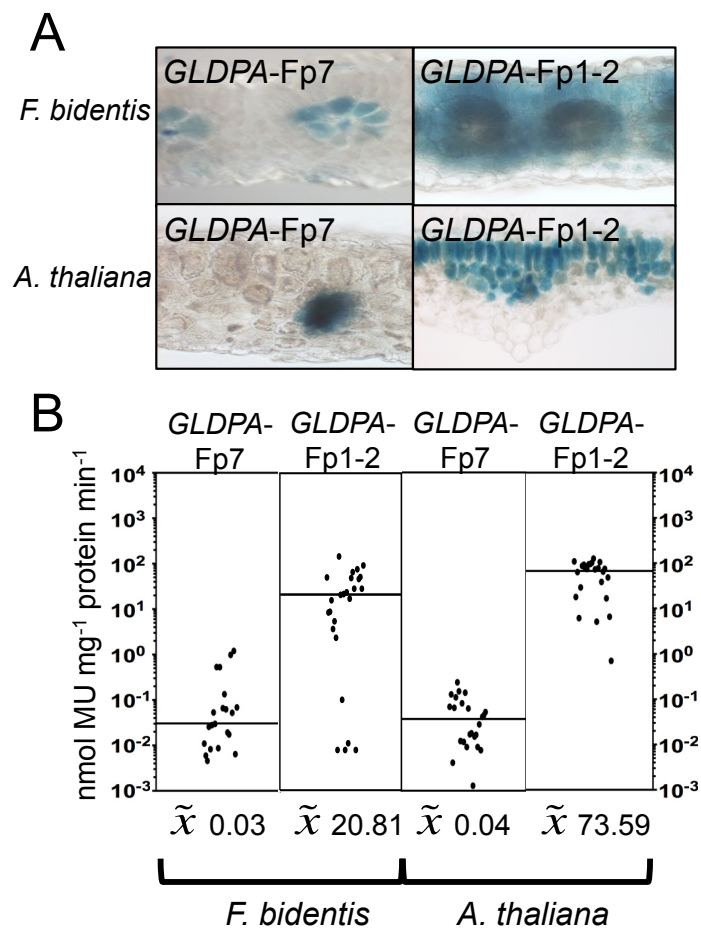


Figure 7

Figure 7. Functional analysis of the sub-promoters P_{R7} and P_{R2} of the *GLDPA* gene of *F. pringlei*.

Regions 1-2 (*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 analysed in transgenic *F. bidentis* (C_4) and transgenic *A. thaliana* (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.

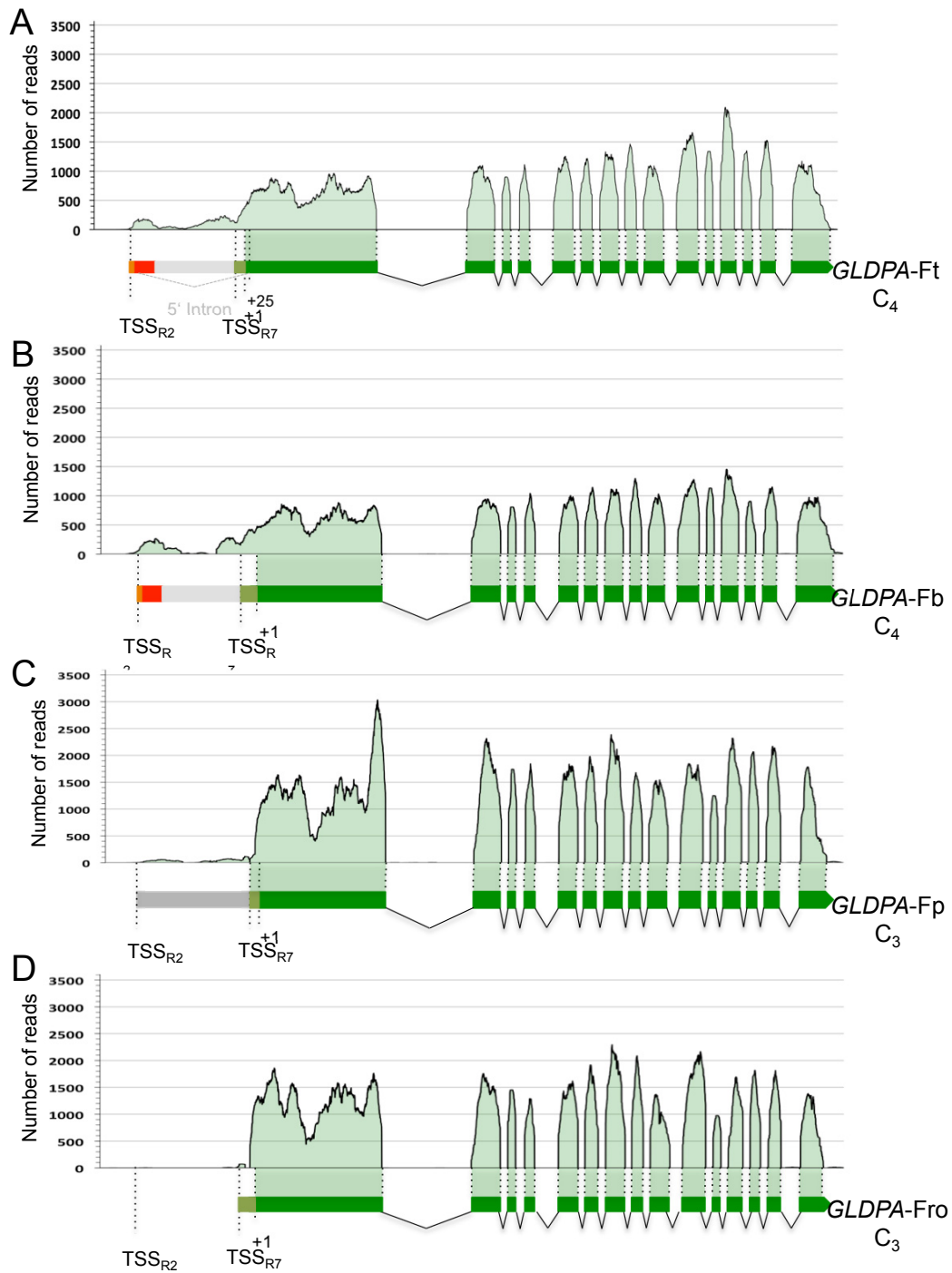


Figure 8

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 **(A)** *F. trinervia* (C₄), **(B)** *F. bidentis* (C₄), **(C)** *F. pringlei* (C₃) and **(D)** *F. robusta* (C₃) were mapped on the sequence of the *GLDPA* gene of *F. trinervia* including its 5' flanking sequence. The numbers of reads covering each position of the gene sequence were counted.

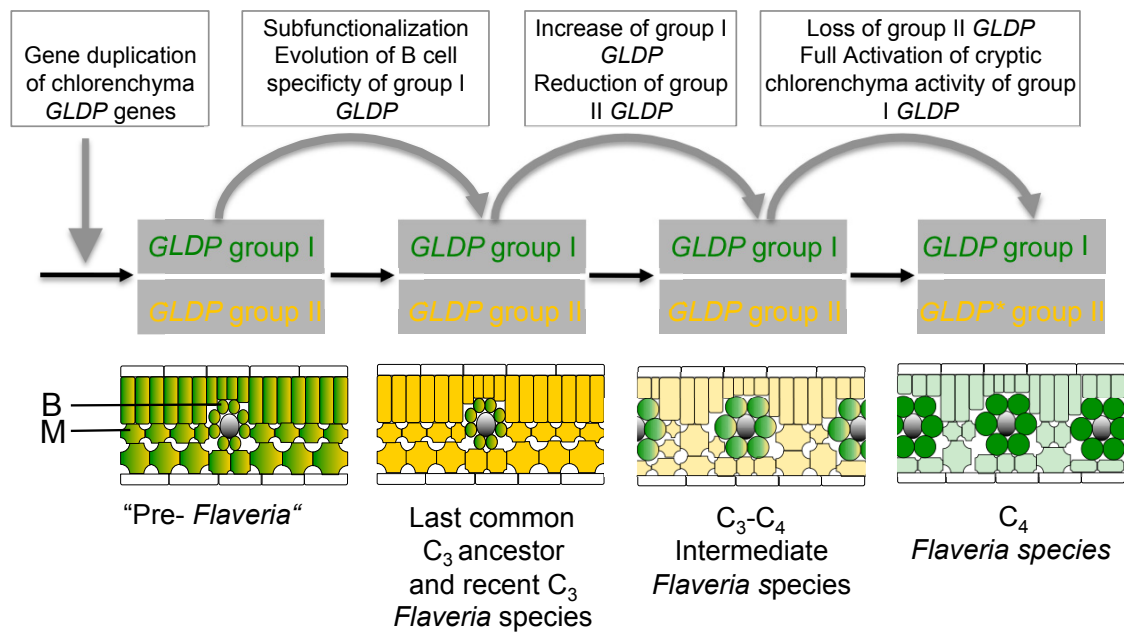
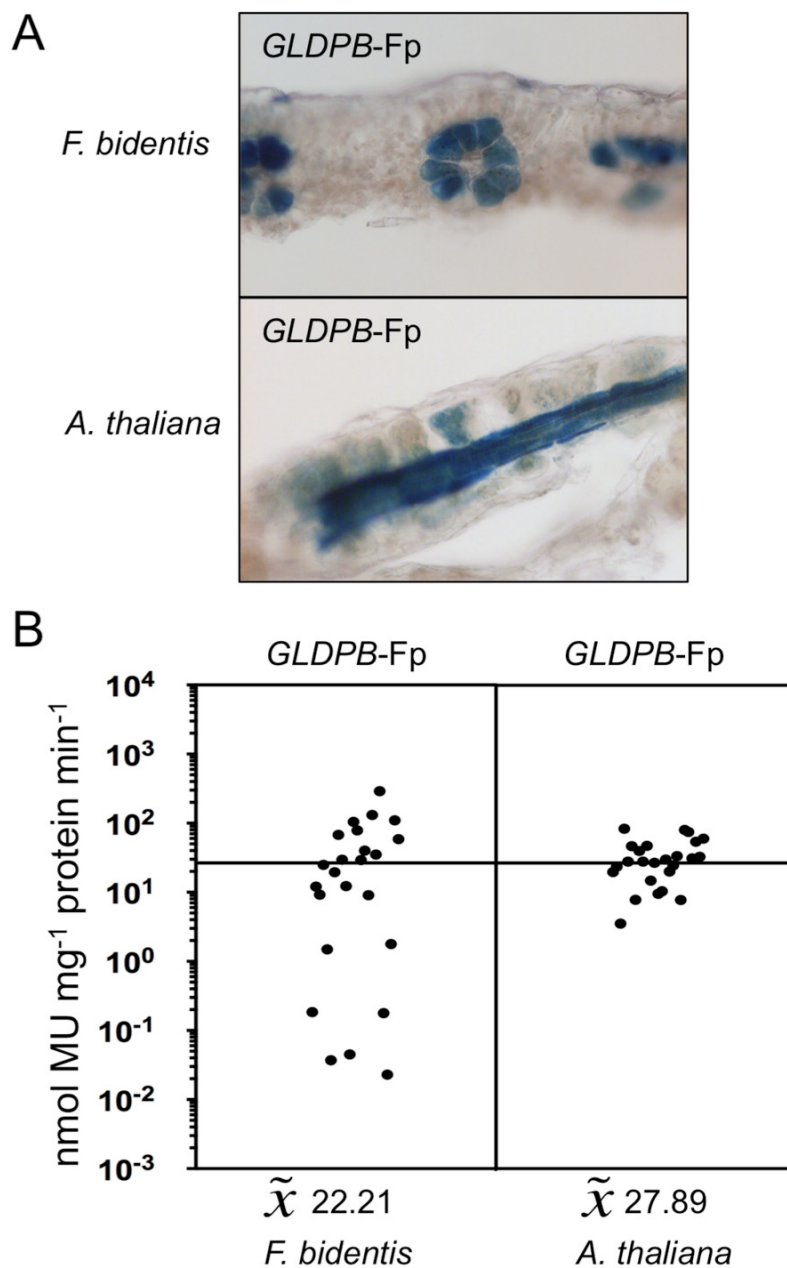


Figure 9

Figure 9. Model for the evolution of bundle sheath specific *GLDP* expression in the genus *Flaveria*.

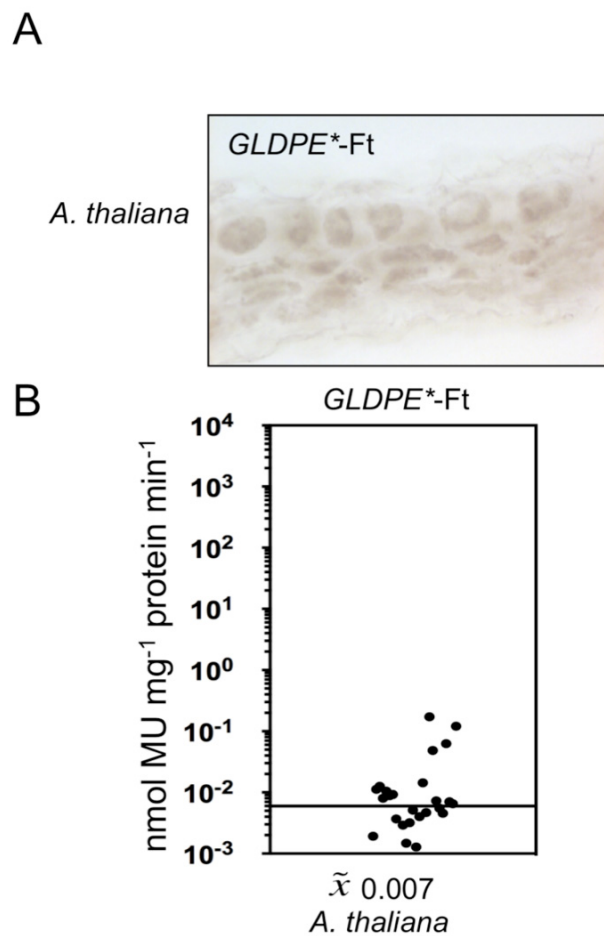
The duplication of a photorespiratory *GLDP* gene in early, ancestral C_3 *Flaveria* species led to two ubiquitously expressed *GLDP* genes with identical expression patterns in all chlorenchyma tissues (M, mesophyll; B, bundle sheath) of the leaf. Sub-functionalization remodelled the expression of the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

SUPPLEMENTAL FIGURES



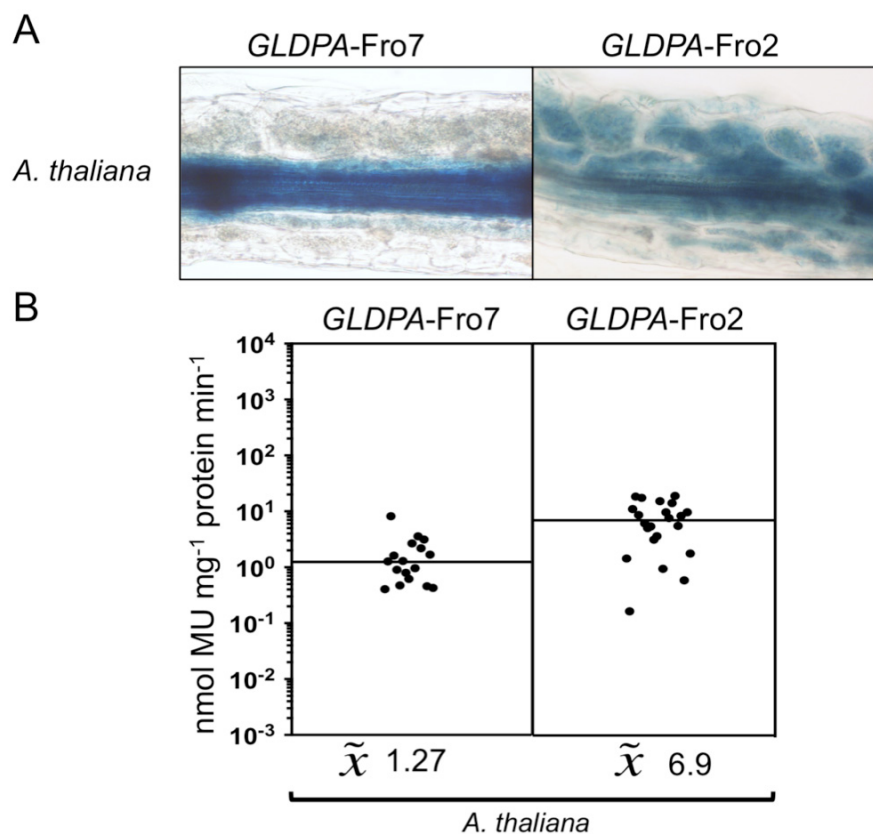
Supplemental Figure 1. Functional analysis of the 5' flanking sequence of the *GLDPB* gene of *F. pringlei*.

The 5' flanking sequence (2040 bp) of the *GLDPB* of *F. pringlei* was fused to the GUS reporter gene, inserted in the pBIN19 plant transformation vector (Bevan, 1984) and analysed in transgenic *F. bidentis* (C₄) and transgenic *A. thaliana* (C₃). **(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.



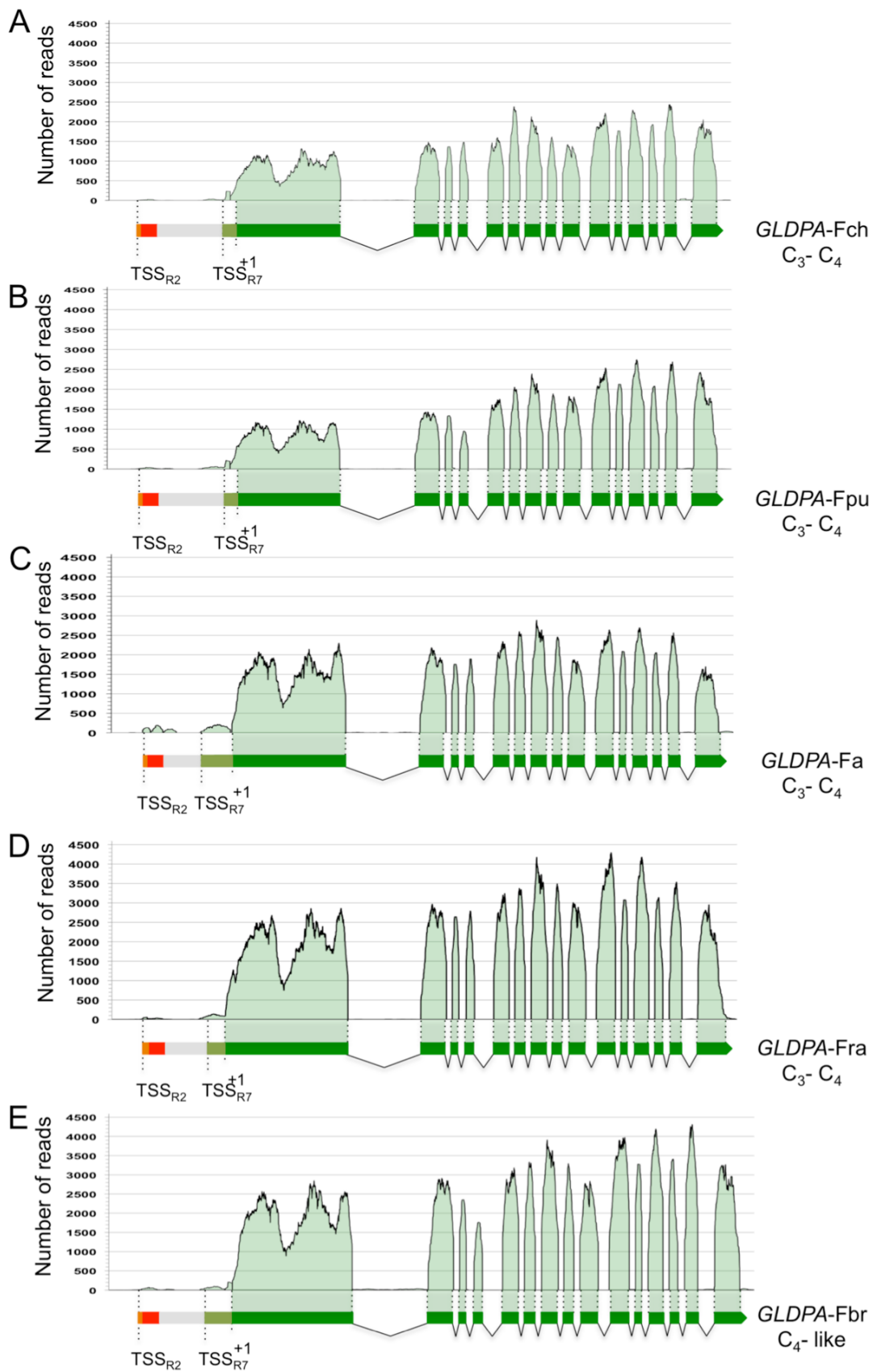
Supplemental Figure 2. Functional analysis of the 5' flanking sequence of *GLDPE gene of *F. trinervia*.**

The 5' flanking sequence (1881 bp) of the *GLDPE** gene of *F. trinervia* was fused to the GUS reporter gene in the plant expression vector pBI121 (Jefferson et al., 1987; Chen et al., 2003) and analysed in transgenic *A. thaliana*. **(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.



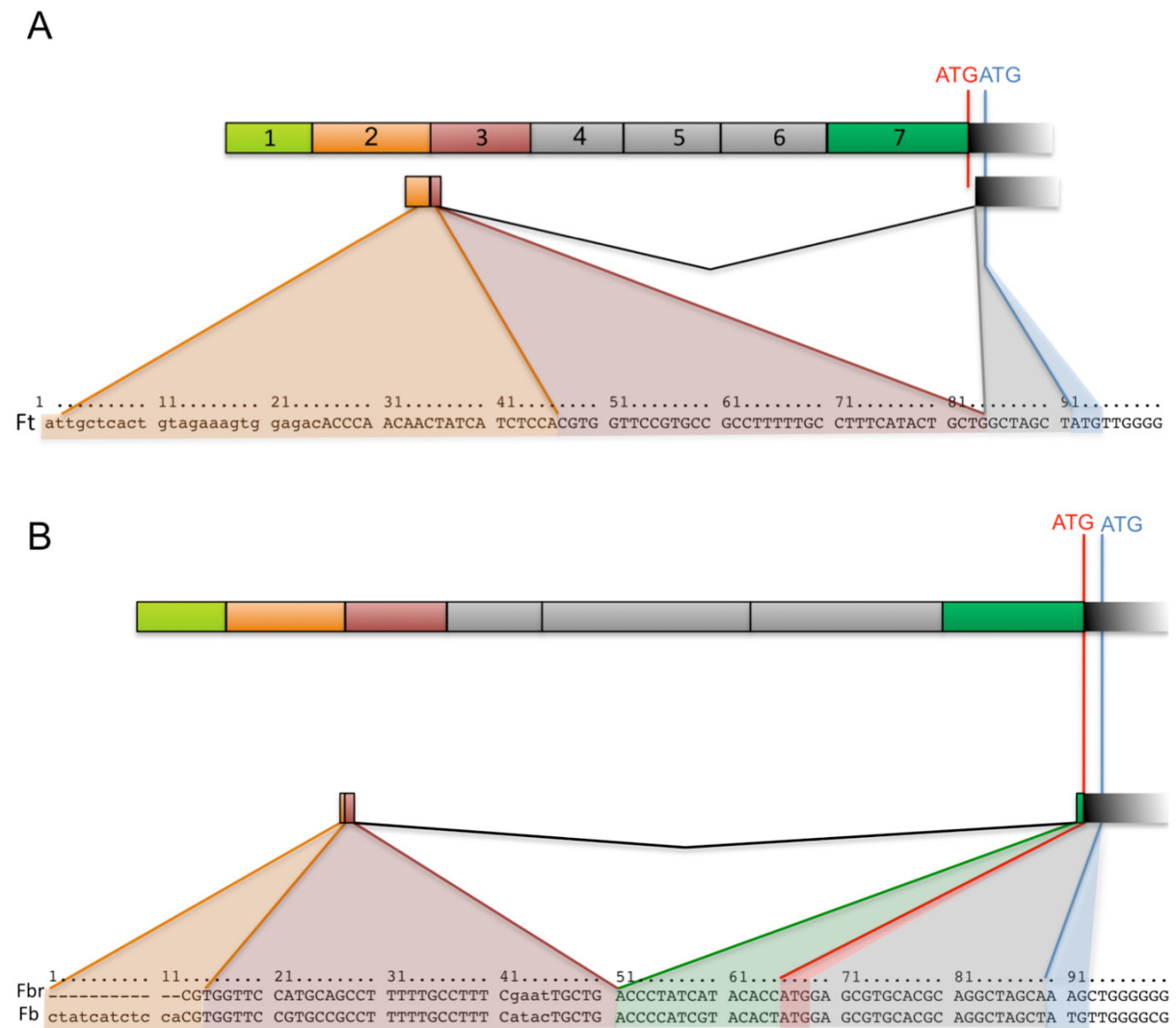
Supplemental Figure 3. Functional analysis of the proximal (P_{R7}) and distal (P_{R2}) sub-promoters of the *GLDPA* gene of the C_3 species *F. robusta*.

Regions 2 (568 bp) and 7 (318 bp) of the 5' flanking sequence of the *GLDPA* gene of *F. robusta* were fused to the GUS reporter gene in the plant expression vector pBI121 (Jefferson et al., 1987; Chen et al., 2003) and analysed in transgenic *A. thaliana* (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.



Supplemental Figure 4. Transcript coverage of the *GLDPA* genes of the C₃-C₄ intermediates *F. chloreaefolia*, *F. pubescens*, *F. anomala*, *F. ramossissima* and *F. brownii*.

Illumina reads obtained from sequencing the leaf transcriptomes of **(A)** *F. chloraefolia* (C₃), **(B)** *F. pubescens* (C₃-C₄), **(C)** *F. anomala* (C₃-C₄), **(D)** *F. ramosissima* (C₃-C₄) and **(E)** *F. brownii* (C₄-like) were mapped on the sequence of the *GLDPA* gene of *F. trinervia* including its 5' flanking sequence. The numbers of reads covering each position of the gene sequence were counted.



Supplemental Figure 5. Splice variants of *GLDPA* transcripts derived from the distal transcriptional start site.

Two different spliced forms of transcripts originating from the distal transcriptional start site TSS_{R2} were detected in the Illumina reads of the *C*₄ species *F. trinervia* (**A**) and *F. bidentis* (**B**). The splice variant of *F. bidentis* was originally found in a PCR reaction directed to transcripts derived from distal transcriptional start site. Mappings were performed against the functionally spliced transcripts. The *C*₄-like species *F. brownii* showed the same splicing behaviour as *F. bidentis* (**B**). In *F. trinervia*, the splice acceptor site is located in the coding region of *GLDPA* and therefore an alternative ATG has to be used (Wiludda et al., 2012). In contrast, both in *F. bidentis* and *F. brownii*, the splice acceptor sites are located upstream of the *GLDPA* open reading frame.

SUPPLEMENTAL DATA

CLUSTAL 2.0.8 multiple sequence alignment

```

Fbr39904      ATGGAGCGTGCACGCAGGCTAGCCA-----TGCTGGGGCGTCTTGTGTGCACAAACC
Fpu42513      ATGGAGCGTGCACGCAGGCTAGCCA-----TGCTGGGGCGTCTTGTGTGCACAAACC
Fc21575       ATGGAGCGTGCACGCAGGCTAGCCA-----TGCTGGGGCGTCTTGTGTGCACAAACC
FbShoot17051  ATGGAGCGTGCACGCAGGCTAGCTA-----TGTTGGGGCGCCTTGTGTGCACAAACC
FtGLDPA       ATGGAGCGTGCACGCAGGCTAGCTA-----TGTTGGGGCGCCTTGTGTGCACAAACC
FraContig1584 ATGGAGCGTGCACGCAGGCTAGCTA-----TGCTGGGGCGCCTTGTGTGCACAAACC
Fa26629       ATGGAGCGTGCACGCAGGCTAGCTA-----TGCTGGGGCGCCTTGTGTGCACAAACC
FpGLDPA       ATGGAGCGTGCACGCAGGCTAGCCA-----TCCTGGGGCGCCTTGTGTGCACAAACC
Fro75076      ATGGAGCGTGCACGCAGGCTAGCCAACAAAGCCATCCTGGGGCGCCTGGTGTGCACAAACT
FpGLDPB       ATGGAGCGTGCACGCAGGCTAGCCAACAAAGCCATCCTGGGGCGCCTTGTGTGCACAAACC
FpGLDPD       ATGGAGCGTGCACGCAGGCTAGCAAAACAAAGCCATCCTCAGGCGCTTAGTTCACAAACC
Fro49132      ATGGAACGTGCTCGCAGGCTAGCAAAACAAAGCCATCCTCAGGCGCTTAGTTCACAAACC
FraContig10162 ATGGAACGTGCTCGCAGGCTAGCAAAACAAAGCCATCCTCAGGCGCTTAGTTCACAAACC
Fb40810       ATGGAACGTGCTCGCAGGCTAGCAAAACAAAGCCATCCTCAGGCGCTTAGTTCACAAACC
FtGLDPE*      ATGGAACGTGCTCGCAGGCTAGCAAAACAAAGCCATCCTCAGGCGCTTAGTTCACAAACC
Fa27221       ATGGAACGTGCTCGCAGGCTAGCAAAACAAAGCCATCCTCAGGCGCTTAGTTCACAAACC
Fc44326       ATGGAGCGTGCACGCAGGCTAGCAAAACAAAGCCATCCTCAGGCGCTTAGTTCACAAACC
Fpu36693      ATGGAGCGTGCACGCAGGCTAGCAAAACAAAGCCATCCTCAGGCGCTTAGTTCACAAACC
Fbr25606      ATGGAGCGTGCACGCAGGCTAGCAAAACAAAGCCATCCTCAGGCGCTTAGTTCACAAACC
FpGLDPE       ATGGAGCGTGCACGCAGGCTAGCAAAACAAAGCCATCCTCAGGCGCTTAGTTCACAAACC
FraContig36589 ATGGATCGTGTCTCGCAGGATAGCGAACAATAATCGATCCTCAAACGCCTAATCTCACAATCC
Fro71768      ATGGATCGTGTCTCGCAGGATAGCGAACAATAATCGATCCTCAAACGCCTAATCTCACAATCC
Fb17191       ATGGATCGAGCTCGAAGGCTAGCGAACAATAATCGATCCTGAAACGCCTGATTTTACAATCC
Ft12179       ATGGATCGAGCTCGAAGGCTAGCGAACAATAATCGATCCTGAAACGCCTGATTTTACAATCC
Fbr18208      ATGGATCGAGCTCGAAGGCTAGCGAACAATAATCGATCCTCAAACGCCTAATTTTACAATCC
Fc16238       ATGGATCGAGCTCGAAGGCTAGCGAACAATAATCGATCCTCAAACGCCTAATTTTACAATCC
Fpu8407       ATGGATCGAGCTCGAAGGCTAGCGAACAATAATCGATCCTCAAACGCCTAATTTTACAATCC
Fa17483       ATGGATCGAGCTCGTAGGCTAGCGAACAATAATCGATCCTCAAACGCCTAATTTTACAATCC
FpGLDPC       ATGGATCGTGTCTCGTAGGCTAGCGAACAATAATCGATCCTCAAACGCCTAATTTTACAATCC
*****

Fbr39904      AAA-CACAACCCATCAATATCATCCCCTGCTCTGTGTTACCTTCACGGTATGTTTCCTC
Fpu42513      AAA-CACAACCCATCAATATCATCCCCTGCTCTGTGTTACCTTCACGGTATGTTTCCTC
Fc21575       AAA-CACAACCCATCAATATCATCCCCTGCTCTGTGTTACCTTCACGGTATGTTTCCTC
FbShoot17051 AAA-CACAACCCATCAATATCATCCTCTGCTCTGTGTTACCTTCACGGTATGTTTCCTC
FtGLDPA       AAA-CACAACCCATCAATATCATCCCCTGCTCTGTGTTACCTTCACGGTATGTTTCCTC
FraContig1584 AAA-CACAACCCATCAATATCATCCCCTGCTCTGTGTTACCTTCACGGTATGTTTCCTC
Fa26629       AAA-CACAACCCATCAATATCATCCCCTGCTCTGTGTTACCTTCACGGTATGTTTCCTC
FpGLDPA       AAA-CACAACCCATCAATATCATCCCCTGCTCTGTGTTACCTTCACGGTATGTTTCCTC
Fro75076      AAA-CACAACCCATCAATATCATCCCCTGCTCTGTGTTACCTTCACGGTATGTTTCCTC
FpGLDPB       AAA-CACAACCCATCAATATCATCCCCTGCTCTGTGTTACCTTCACGGTATGTTTCCTC
FpGLDPD       AAA-CACAACCCATCAA-----CCCCTGCTCTATACTCACCTTCACGGTATGTTTCCTC
Fro49132      AAA-CACAACCCATCAA-----CCCCTGCTCTATACTCACCTTCACGGTATGTTTCCTC
FraContig10162 AAA-CACAACCCATCAA-----CCCCTGCTCTATACTCACCTTCACGGTATGTTTCCTC
Fb40810       AAA-CACAACCCATCAA-----CCC-TGCTCTATACTCACCTTCACGGTATGTTTCCTC
FtGLDPE*      AAA-CACAACCCATCAA-----CCCCTGCTCTATACTCACCTTCACGGTATGTTTCCTC
Fa27221       AAA-CACAACCCATCAA-----CCCCTGCTCTATACTCACCTTCACGGTATGTTTCCTC
Fc44326       AAA-CACAACCCATCAA-----CCCCTGCTCTATACTCACCTTCACGGTATGTTTCCTC
Fpu36693      AAA-CACAACCCATCAA-----CCCCTGCTCTATACTCACCTTCACGGTATGTTTCCTC
Fbr25606      AAA-CACAACCCATCAA-----CCCCTGCTCTATACTCACCTTCACGGTATGTTTCCTC
FpGLDPE       AAAACACAACCCATCA-----CCCCTGCTCTATACTCACCTTCACGGTATGTTTCCTC
FraContig36589 AAA-CAACACCCATCAT-----CCCCTGTTTTGTGTTTCATCATCAAGATTCGTTTCATC
Fro71768      AAA-CAACACCCATCAT-----CCCCTGTTTTGTGTTTCATCATCAAGATTCGTTTCATC
Fb17191       AAA-CAACACCCATCAT-----CCCCTGTTTTGTGTTTCATCATCAAGATTCGTTTTTC
Ft12179       AAA-CAACACCCGTCAT-----CCCCTGTTTTGTGTTTCATCATCAAGATTCGTTTTTC
Fbr18208      AAA-CAACACCCATCAT-----CCCCTGTTTTGTGTTTCATCATCAAGATACGTTCTTTC
Fc16238       AAA-CAACACCCATCAT-----CCCCTGTTTTGTGTTTCATCATCAAGATACGTTCTTTC
Fpu8407       AAA-CAACACCCATCAT-----CCCCTGTTTTGTGTTTCATCATCAAGATACGTTCTTTC
Fa17483       AAA-CAACACCCATCAT-----CCCCTGTTTTGTGTTTCATCTTCAAGATACGTTTCATC
FpGLDPC       AAA-CAACACCCATCAT-----CCCCTGTTTTGTGTTTCATCTTCAAGATACGTTTCATC
*** **

```

```

Fbr39904          TTTATCCCCTTATGTATGCAGTGGCACAACGTTAGATCAG
Fpu42513         TTTATCCCCTTATGTATGCAGTGGCACAACGTTAGATCAG
Fc21575          TTTATCCCCTTATGTATGCAGTGGCACAACGTTAGATCAG
FbShoot17051     TTTATCCCCTTATGTATGCAGTGGCACAACGTTAGATCAG
FtGLDPA          TTTATCCCCTTATGTATGCAGTGGCACAACGTTAGATCAG
FraContig1584    TTTATCCCCTTATGTATGCAGTGGCACAACGTTAGATCAG
Fa26629          TTTATCCCCTTATGTATGTGGTGGCACAACGTTAGATCAG
FpGLDPA          TTTATCCCCTTATGTATGTAGTGGCACAACGTTAGATCAG
Fro75076         TTTATCCCCTTATGTATGTAGTGGCACAACGTTAGATCAG
FpGLDPB         TTTATCTCCTTATGTATGTAGTGGCACAACGTTAGATCAG
FpGLDPD         TCTATCTCCTTACATATGCAGCAGCAGAAACGTTAGATTTG
Fro49132         TTTATCTCCTTACATTTGCAGTGGCAGAAACGTTAGATTTG
FraContig10162   TTTATCTCCTTACATATGCAGTGGCACAACGTTAGATTCG
Fb40810          TTTATCTCCTTACATATGCAGTGGCACAACGTTAAATTCG
FtGLDPE*         TTTATCTCCTTACATATGCAGTGGCACAACGTTAGATTCG
Fa27221         TTTATCTCCTTACATATGCAGTGGCACAACGTTAGATTTG
Fc44326         TTTATCTCCTTACATATGCAGTGGCACAACGTTAGATTTG
Fpu36693        TTTATCTCCTTACATATGCAGTGGTGCAGAAACGTTAGATTTG
Fbr25606        TTTATCTCCTTACATATGCAGTGGCACAACGTTAGATTTG
FpGLDPE         TTTATCTCCTTACATATGCAGTGGCACAACGTTAGATTTG
FraContig36589   ACTATCTTCAAACGTTTGGATCAACAGAAATGGTAGATCAG
Fro71768        ACTATCTTCAAACGTTTGGATCAACAGAAATGGTAGATCAG
Fb17191         ATCATCTTCGCACGTTTTGATCAGCAGAAATGGTGGATCAA
Ft12179         ATCATCTTCGCACGTTTTGATCAGCAGAAATGGTGGATCAA
Fbr18208        ATCATCTTCACACGTTTTAATCAGCAGAAATGGTGGATCAA
Fc16238         ATCATCTTCACACGTTTTGATCAGCAGAAATGGTGGATCAA
Fpu8407         ATCATCTTCACACGTTTTGATAAGCAGAAATGGTGGATCAA
Fa17483         ACTATCTTCTTACCTTTGATCAACAGAAATGGTGGATCAG
FpGLDPC         ATCATCTTCACACGTTTTGGATCAACAGAAATGGTAGATCAG
                *** * * * * *                ** * * **

```

Supplemental Data 1. ClustalX alignment of the *GLDP* sequences used for the phylogenetic analysis.

>FroGLDPC

```

ATGGATCGTGCTCGCAGGATAGCGAACAAATCGATCCTCAAACGCCTAATCTCACAAATCCAAACAACACCCATCA
TCCCCTGTTTTGTGTTTCATCATCAAGATTTCGTTTTCATCACTATCTTCAAACGTTTGGATCAACAGAAATGGTAGA
TCAGAGGGAAATATTTATGGGTTTTGGTCAAAAGTCGATTGATTTGATTGAAAGCTTTGAAACCCAGTGATACT
TTCGCTCGTCGCCATAACTCTGCAACCCCTGAAGAACAACGAAATGGCGGAATTTGTAGGGTTTTCTAGCCTT
GATTCGTTGATTGATGCCACTGTACCGAAATCAATTCGAATTTGGTCAGATGAAATTTGCCCAATTTGATCAAGGG
TTAACTGAATCTCAGATGATTGATCATATGCAAACATTAGCTGCTAAAAATAGGGTTTTTAAGTCTTTTATTGGT
ATGGGGTATTACAATACATTTGTTCTAGTGTTATTTTTGAGGAATATTATGGAGAACCCTGGTTGGTATACTCAG
TATACTCCTTATCAGGCTGAAATCGCGCAGGGGCGACTCGAATCTTTACTGAATTTTCAAACCATGGTTACGGAT
TTAACCGGTTTACCAATGTGCAATGCGTCGTTGCTAGACGAAGGGACTGCAGCTGCAGAAAGCTATGGCTATGTGT
AACAACTTCAAAGGGTAAAAAGAAGACATTTTTGATAGCAAGAACTGTCACCTCAAACGATTTGATATCTGC
AAGACTAGAGCAGATGGTTTTGATTTGAAAGTGGTGCAGATTGGATCTTAAAGATTTGATTTACTCGTCTGGTGT
GTTTTGTGGAGTGCTGGTGCAGTATCCTGGAAGTCAGAGTGGATTTGTTGGATTATGGTGGATTTGTGAACAATGCG
CACGCGAATGGGGTGAAGGTTGTGATGGCTAGTGATTTGTTGGCTTTGACTGTCTTGAAGCCCCCTGGCGAATTT
GGGGTTGATATTGTGGTGGGCTCGGCTCAGAGATTTGGGGTGCCAATGGGGTATGGAGGTCCCTCATGCTGCGTTT
CTTGCAACGTCTCAAGAATACAAGAGGATGATGCCTGGGAGAATTATTGGTGTAGTGTGATGCTTCTGGTAAA
CCTGCGTTGCGTATGGCGATGCAGACAAGAGAGCAGCATATTAGGAGGGATAAAGCGACTAGCAACATTTGCACA
GCTCAGGCATTGCTTCAAATATGGCTGCTATGTATGGTATATAACCATGGACCGAGGGCCTAAAAACAATCGCA
CAACGAGTCCATGGACTTGCTGCAACATTTGCTGCTGGATTAAGAAACTCGGGACAGTAGATGTCCAAGGTCTT
CCTTTTTTTGACACTGTTAAGATTAATGTACTGATTCAAGTGCAATATCTGAAGAAGCTCTTAAAAACAAGATG
AATCTACGTGTTGTTGATAAAAAACTATTACCGTCTCATTTGATGAAAACAACCACCATAGAAGATGTTGATACA
TTGTTTTAAAGTTTTCTCCTTTGGGAAACAGGTGACATTTACTGCTGCATCTCTTGCACCAGAGGTTGAAAATGTT
ATCCCTTCTGGGCTTGTAAAGGGACTCCATATATGACACATTTCAATTTTCAACTCTTTCACACAGAGCATGAG
CTACTCAGATATATCAGCAAACAGAGATGAAAGGATCTCTCATTGTGCCACAGTATGATTTCCCTGGGGTCTTGT
ACAATGAAGTCAATGCAACAACAGAGATGATGCCGGTGACATGGCCAGCATTTGCAGATGATCATCCATTTGCA
CCTATTGAACAGGCAGAAGGCTATCAAGAAATGTTTAAAGAAATTTGGGTGATATGTTGTGTACCATCACCAGTTT
GATTCCTTCTCTTTGCAACCTAATGCTGGTGTGCTGGAGAGTATGCCGGCTGATGGTTATCCGAGCATATCAT
ATGGCAAGAGGAGATCATCATAGAAATGTATGCATCATTTCTGTGTGATGACATGGAACAACATCCTGCAAGTGT
GCTATGTGTGGAATGAAAATCATTACTGTTGGAACAGATGCTAAAGGTAACATTAATATTTGAAGAGGTACGAAAG
GCTGCTGAAGCGAATAAAGACAATCTATCTGCTCTAATGGTAACATATCCTTCGACTCATGGAGTTTATGAAGAG
GGCATTGACGAGATTTGCAAAATTTATTCATGACAATGGAGGTCAGGTTTACATGGATGGAGCCAATATGAATGCA
CAGGTTGGTTTTAAACAAGTCTGGGTGGATTGGTGTGATGTTTGCCATCTGAACCTTCATAAGACTTTTTGCAAT
CCACATGGCGGAGGTGGGCCCCGAATGGGTCCAATCGGTGTGAAAAACACTTGGCGCCATACCTTGCCCTTCTCAT

```


CCTGTGGTTGGGACAGGAGGGTTACCAGCACCAGAAAAGGCCGAGCCACTTGGTACAATATCTGCAGCACCTTGG
 GGATCGGCACTTATCTTGGCCATATCATATACATACATCGCCATGATGGGATCTAAGGGACTTACGGATGCATCA
 AAGATAGCCATATTGAATGCAAACCTACATGGCAAAACGCCTTGAGAGTCGCTACCCCATCCTTTCCGCGGTGTC
 AATGGAACCGTGGCCCATGAATTCATTGTGCGATTTAAGACCCTTAAAACTACAGCTGGAATAGAGCCAGAAGAC
 GTTGCTAAACGTTTGTATAGATTACGGATTTTCATGGCCCAACAATGTCATGGCCGGTTTCTTGGAAACGCTAATGATT
 GAACCTACTGAAAGTAAAAGCAAAGCTGAGTTAGACAGGTTTTTGTGATGCTTTGATCTCCATTAGAGAAGAAATA
 GCTGAAATGAGAAAAGGAAACGCCGATATCAACAACAACGTCATCAAGGGAGCTCCTCATCTCTTCAAGTGCTC
 ATGGCTGATAAGTGGACTAAACCATACTCTCGAGAATATGCGCCCTACCCCTGCTTCATGGCTTCGGTCCGCCAAG
 TTCTGGCCTACCACATGCCGTGTGGATAATGTGTATGGCGATCGCAACCTCGTTTGCACCCCTTCAGTCGCCACAC
 GAGGAAGAAGAAAAGGCCGGCTGCAGCCACTGCTTAA

>FroGLDPD

ATGGAACGTGCTCGCAGGCTAGCAAACAAAGCCATCCTCAGGCGTCTAGTTTCAAAAACCAAACACAACCCATCA
 ACCCCTGCTCTATACTCACCTTCACGGTATGTTTCTCTTTATCTCCTTACATTTGCAGTGGCAGAAACGTTAGA
 TTTGATAGAAACCTTAATGGGTTTTGGTTCACAAGCTCGAACCATATCAGTTGAAGCATTGAAACCAAGTGACACT
 TTCCACGTCGCCATAACTCCGCTACACCAGAAGAACAGACCAAAAATGGCTGAATTCGTAGGGTACCCTAATCTT
 GATTCACTCATTGATGCCACTGTGCCAAATCAATCCGCTTGAATCGATGAAATACTCTAAGTTTGTATGAGGGG
 TTGACAGAATCCCAAATGATTGCTCATATGCAAGAATTAGCTTCAAAAATAAGATCTTTAAGTCTTTTATTTGGT
 ATGGGGTATTACAATACTTATGTTCTACTGTTATTTTGGAGAACATTATGGAGAACCCTGGTTGGTATACTCAG
 TACACACCGTACCAACGCGGAGATTGCGCAGGGGGCGGCTCGAGTCCCTGCTCAATTTCCAGACCATGGTTACTGAT
 CTGACCGGTTTTCCCATGTGCAATGCTTTCGTTACTAGATGAAGGAACTGCAGCTGCTGAGGCTATGGCTATGTGT
 AATAATATCCAAAAGGGTAAAAGAAAACCTTTATTATTGCTAGTAACTGCCACCCACAGACAATTGATATCTGT
 AAGACTAGGGCCGATGGGTTGATCTGAAAAGTGGTTACTTCGGATCTTAAAAATTTTGTACTCGTCTGGTGAT
 GTGTGTGGGGTGTGGTTCAGTATCCGGGGACTGAGGGTGAGGTGTTGGACTATGGTGAGTTCATTAAGAATGCA
 CATGCTAGTGGGGTGAAGGTGGTCATGGCTAGTGATTTGTTGGCTTTGACTATTTTGAAGCCACCTGGTGAACCT
 GGGGCTGATATTGTGGTTCGGTTCGGCTCAGCGGTTTTGGTGTCCAATGGGGTATGGTGGTCTCATGCAGCGTTT
 CTTGCGACGTCCCAAGAGTATAAGAGAATGATGCCGGGAAGAATTATTGGTGCTAGTGTGATTTCTTCTGGGAAA
 CCGGCTTTACGTATGGCGATGCAGACCAGAGAACAACATATCAGGAGAGACAAAAGCAACTAGTAACATCTGCACA
 GCTCAGGCATTGCTTGCGAACATGGCTGCTATGTTTGGTGTTTACCATGGACCAGAAGGCCTAAAAGACCATAGCC
 AAACGGGTCCACGGTCTCGCTGGCACATTTGCTGCAGGTTTGAAGAAGCTTGGGACGGTACAAGTTCAGGATCTT
 CCGTTTTTTGACACTGTGAAAATAACATGTGCTGATGCAAAATGCAATGCTGAAGAAGCTTACAAGCATAAGATG
 AATCTTCGTATTGTGGACAAAACACTATTACTGTGCTGCTCCTTTGATGAAACAACAACACTATTGAAGATGTTGATAC
 TTGTTCAAAGTGTGTTGCATTTGGGAAAACCTGTGACATTCCTGCTGCATCTATTGCACCCGAGGTTCAAGATGCC
 ATCCCTTCTGGGCTAGTGAGGGAGACTCCGTATTTGACTCATCCAATTTTCAACATGTATCATACGGCATGAG
 CTATTAAGATACCTTAGTAAGTTACAGTCAAGGATCTGTCAATTGTGTACAGTATGATTTCCCTTGGGCTCTTGT
 ACCATGAAGCTTAATGCAACTACAGAGATGATGCCTGTGACATGGCCAGCCTTTGCAGATATGCATCCTTTTCGCT
 CCGACCGAACAGGCAGAGGGTTATCAGGAAATGTTCAAGAATCTGGGTGACTTGTGTTGTACCATTACTGGATTT
 GATTTCTTCTCCTTGAACCCAATGCTGGTGTGCTGCGGGTGAATACGCTGGGCTGATGGTTATCCGAGCATATCAT
 AAGGCAAGAGGAGATCATCATAGGAATGTATGCATTATCCCTGTGTGTCAGCGCATGGAACGAATCCTGCTAGTGCT
 GCTATGTGTGGAATGAAAATTATCACTGTGGGAACCTGACTCCAAAGGTAACATTAATATTGAAGAGGTGCGTAAG
 GCTGCTGAGGCGAATAAAGAGAATCTATCTGCTCTTATGGTTACCTATCCTTCGACTCATGGAGTTTACGAAGAG
 GGTATAGATGAGATCTGCAAAATTATTCATGACAATGGAGGTGAGGTTTATATGGATGGAGCAAAACATGAATGCA
 CAGGTTGGCTTGACGAGTCCAGGGTGGATTGGTGTGATGTATGTCATCTAAACCTACACAAGACATTTTGCATT
 CCACATGGCGGTGGTGGTCCGGGAATGGGTCCCATTGGAGTGAAGAAAACACTTGGCACCTTACTTGCCTTCACAT
 CCTGTGGTGGCGACAGGCGGCATACCAGCCCCTGAACAAAGTCAGCCAGTGGGTACCATTTCTGCAGCCCCCTTGG
 GGTCTGCTCTTATCATGCCTATATCATAACACATACATTGCCATGATGGGATCTCAAGGAATAACAATGCATCA
 AAGATAGCTATCTTGAATGCAAACCTATATGGCAAAACGCTTGGAGAATCACTATCCAATTTCTTTCCGTGGTGTG
 AATGGAACAGTTGCCCATGAATTCATTGTTGATTTGAGACCTTTGAAGACTACTGCTGGTATAGAGCCAGAAGAT
 GTTGCTAAACGCTAATAGATTACGGGTTTTCATGGTCCAACAATGTCATGGCCAGTTCAGGAACACTCATGATT
 GAACCCACTGAAAGTAAAAGCAAGGCGGAGTTGGACCGATTCTGTGATGCTTTAATCTCCATTAGACAAGAAATA
 GCAGAGATCGAGAAAGGACAGTTGATCTCAACAACAATGTGATCAAGGGAGCCCCACCCGCCACAACACTACTC
 ATGGCCGATAAGTGGACAAAACCATACACTCGAGAATATGCAGCATAACCCCTGCTCCATGGCTTCGCGCTGCTAAG
 TTCTGGCCTACCACATGCCGTGTAGACAATGTGTATGGTGACCGCAACCTCATCTGCACCCCTTCAACCCGCCACAA
 GAGTATGAAGAGAAAAGCCGAAGCTACTGCTTAA

Supplemental Data 2. Sequences of the *GLDPC* and *GLDPD* genes of *F. robusta*.

The sequences were extracted from assembled contigs of RNAseq of leaves. The start (ATG) and stop (TAA) codons are marked in red.

Oligonucleotide	Sequence 5' → 3'
FroGLDPAfw	<i>aagctt</i> aaactttacttatctc
FroGLDPArv	cccgggggcgtatgatgaggtcta
AtGLDP1fw	<i>aagctt</i> gagctgggatgatgtctgctc
AtGLDP1rv	cccgggtgggaaaaaagggtgcagtea
AtGLDP2fw	<i>aggctt</i> aggcttcaaatcaacagtgacac
AtGLDP2rv	cccgggaaccaaaccataaaaaaaacaaatgc
FtGLDPE*fw	<i>aagctt</i> atcgctccacttccctac
FtGLDPE*rv	cccgggtgtgtgtgattgatgatgata
FpGLDPA_7fw	<i>aagctt</i> catttgatctgtaacaataa
FpGLDPA_7rv	cccgggggtgtatgatggggtctaattgg
FpGLDPA_1fw	<i>aagctt</i> aaacttaacttatatcaactt
FpGLDPA_2rv	cccgggggtggagatgatagttggtgggtgtct
FroGLDPA_7fw	<i>aagctt</i> gtgcaataagcctaaatg
FroGLDPA_7rv	cccgggggcgtatgatgaggtctaata
FroGLDPA_2fw	<i>aagctt</i> gaaataggatcagccacaccaa
FroGLDPA_2rv	cccgggggtggagatgatagttgtttg
Reg2fw	ctatcatctccacgtgggttc
GLDPArv	agagattcttgcctcttct

Supplemental Table 1. List of oligonucleotides used for amplification of promoter sequences for GUS constructs and for the amplification of the 5' splice site of transcripts derived from distal transcriptional start site of *F. bidentis*. Restriction sites (*HindIII* aagctt; *XmaI* cccggg) are labelled in italics.

The authors' contributions

SS wrote this manuscript and performed all experiments and analysed all data except those listed below.

JM performed Illumina transcriptome sequencing of the various *Flaveria* species and provided the raw data.

JB performed the qualitative and quantitative analysis of *GLDPD*-Fp in *F. bidentis* (Fig.3A upper picture and 3B left site).

MK and **MS** performed the transformation of *F. bidentis* plants.

HB provided the *GLDPA*-Fp, *GLDPB*-Fp and *GLDPD*-Fp constructs and the sequences of *GLDPC*, *GLDPD* and *GLDPE*.

UG assembled 454 and Illumina transcriptome data and provided contigs obtained from those and helped analysing the data.

HB, **UG** and **PW** participated in drafting of the manuscript.

Manuscript 2 was submitted for publication to *the Plant Cell* (Impact Factor: 8.987)

Chapter 3

Function and origin of the distal promoter (P_{R2}) of the *GLDP* genes in *Flaveria*

Function and origin of the distal promoter (P_{R2}) of the *GLDP* genes in *Flaveria*

Introduction

Glycine decarboxylase (GDC) is one of the key enzymes in plant photorespiration and is located in the bundle sheath cells of C_4 plants. The promoter of the gene encoding the PA protein (*GLDPA*) of GDC of the C_4 plant *F. trinervia* is a tandem promoter with bundle sheath specific expression of the proximal promoter and expression in all photosynthetically active tissues of the distal promoter (see chapter 1: manuscript 1). The 5' flanking sequences of the *GLDPA* genes of all analysed *Flaveria* species show high similarity to each other (see chapter 2: manuscript 2 Figure 6). In all cases the regions 1, 2, 3 and 7 are highly conserved and can be found in all analysed species. These regions carry a function in the promoter composition of the *GLDPA* genes (Engelmann et al. 2008 and chapter 1: manuscript 1) and both – region 2 and region 7 – contain a TSS preceded by an active promoter (chapter 1: manuscript 1). Both promoters of the C_3 species *F. pringlei* (chapter 2: manuscript 2 Figure 7) and *F. robusta* (chapter 2: manuscript 2 Supplemental Figure 4) are functional in their uncoupled form in GUS reporter gene studies. In contrast, RNAseq experiments show that in the C_3 species none or almost none transcripts derive from this promoter (chapter 2: manuscript 2 Figure 8C and 8D). In the C_3 - C_4 intermediate (chapter 2: manuscript 2 Supplemental Figure 4) and the C_4 (chapter 1: manuscript 1 and chapter 2: manuscript 2 Figure 8A and 8B) species transcripts from the distal promoter are detectable. So the question arises which differences can be found between the distal promoters of the *GLDPA* genes of these species. A second question comprises the origin of the distal promoter. To investigate this the 5' flanking region of the group II *GLDPD* gene of the C_3 species *F. pringlei* was consulted.

Results

The distal promoter is conserved in the group II *GLDPD* gene of *F. pringlei* (C_3)

The group II *GLDP* genes (see chapter 2: manuscript 2 Figure 1 for nomenclature) contain 5' flanking sequences that do not show such a conservation in their sequence, if compared to *GLDPA* of *F. trinervia* or between each other, as can be found in the *GLDPA* genes (see chapter 2: manuscript 2 Figure 6). Only the 5' untranslated region (5' UTR) corresponding to

that of the proximal promoter of the *GLDPA* genes of the transcribed genes shows similarities both to each other and to the 5' UTR of the *GLDPA* genes.

Interestingly, the 5' flanking sequence of the group II *GLDPD* gene of *F. pringlei* (C₃) contains a 390 bp long part that shows some similarity to regions 1, 2 and 3 of the group I *GLDP* genes (Figure 1A and B). Other parts of the 5' flanking sequence do not fit to any region of the group I *GLDP* genes. Region 2 of *GLDPA*-Ft is that part of the 5' flanking region that contains the distal promoter. Consequently, the corresponding sequence of *GLDPD*-Fp was tested for promoter activity in an isolated context in *GUS* reporter gene studies in transgenic *Arabidopsis thaliana* plants (Figure 1B and C).

Analyses of seedlings harbouring the construct *GLDPD*-Fp2-3 fused to the *GUS* reporter gene show that this part of the 5' flanking sequence of *GLDPD* of the C₃ species *F. pringlei* indeed functions as a promoter. *GUS* staining is detectable in the leaves of transgenic *A. thaliana* plants. The spatial expression pattern resembles the pattern of the distal promoter of the analysed group I *GLDP* genes. All photosynthetically active tissues in the leaves are stained (Figure 1C).

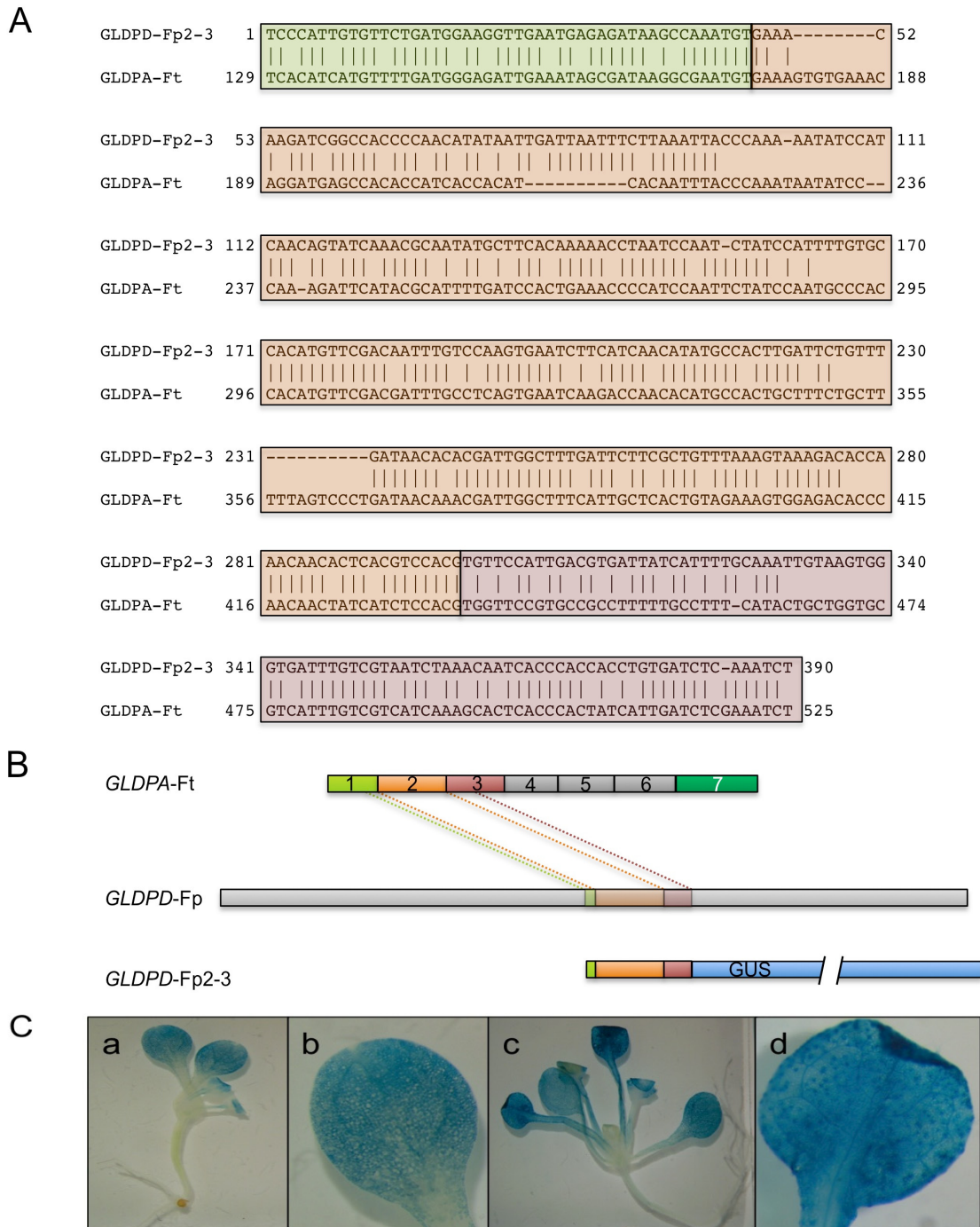


Figure 1. Functional analysis of the corresponding region of the distal promoter of *GLDPA* of *GLDPD* of *F. pringlei*.

(A) Sequence alignment (conducted with the BLAST alignment online tool) of the corresponding region of *GLDPD-Fp* with the 5' region of *GLDPA-Ft*. Colours indicate the respective region of the 5' region of *GLDPA-Ft*. Green, region 1; orange, region 2; red, region 3.

(B) Schematic presentation of the corresponding regions in the 5' flanking regions of *GLDPD-Fp* and *GLDPA-Ft* and the *GUS* reporter gene construct *GLDPD-Fp2-3*.

(C) Histochemical analysis of seedlings harbouring the *GLDPD-Fp2-3* construct. Seedlings were stained for 3h 20min (a and b) or 5h 30min (c and d). b shows a cotyledon leaf in detail, whereas d shows the first leaf in detail.

The transcriptional start site of the distal promoter shows variation on the nucleotide level

In order to analyse the distal promoter, the sequences upstream and downstream of the determined transcriptional start site (TSS) in region 2 as found with 5' RACE (see chapter 1: manuscript 1 Figure 2) of the *GLDPA* genes of *F. trinervia* (C₄), *F. bidentis* (C₄), *F. pubescens* (C₃-C₄), *F. pringlei* (C₃) and *F. robusta* (C₃) were aligned to investigate differences of the nucleotide level (Figure 2). Further the corresponding region of *GLDPD* of *F. pringlei* was included, since this region showed a relevant similarity and is able to act as a promoter in promoter-reporter gene studies (see Figure 1).

alignment position		201.....	211.....	221.....	231.....	241.....
FtGLDPA_2	137	GTGAATCAAG	ACCAACACAT	GCCACTGCTT	TCTGCTTTT	AGTCCCTGAT
FbGLDPA_2	137	GTGAATCAAG	ACCAACACAT	GCCACTGCTT	TCTGCTTTT	AGTCCCTGAT
FpuGLDPA_2	137	CTGAATCAAG	ACCAACACAT	GCCACTGCTT	TCTACTTTAG	AGTCGCTGAT
FpGLDPA_2	134	GTGAATCAAG	ACCAACACAT	GCCACTGCTT	TCTGCTTTT	AGTCGCTGAT
FroGLDPA_2	135	GTGAATCAAG	ACCAACACAT	GCCACTGCTT	TCTGCTTTAG	AGTCGCTGAT
FpGLDPD_2	198	GTGAATCttc	ATCAACATAT	GCCACTG GA T	TCTG g T----	-----TGAT
		*****	* ***** *	*****	* ** *	****
alignment position		251.....	261.....	271.....	281.....	291.....
FtGLDPA_2	187	AACAAACGAT	TGGCTTTC A T	<u>TGCTCACTGT</u>	AGAAAGTGGGA	GACACCCAAC
FbGLDPA_2	187	AACAAACGAT	TGGCTTTC A T	<u>TGCTCACTGT</u>	AGAAAGTGGGA	GACACCCAAC
FpuGLDPA_2	187	AACAAACGGT	TGGCTTTC A T	<u>TGCTCGCAGT</u>	AGAAAGTAGA	GACACCAAAC
FpGLDPA_2	184	AACAAACGAT	TGGCTTTC A T	<u>TGCTCACTGT</u>	TGAAAGTGGGA	GACACCCAAC
FroGLDPA_2	185	AACAAACGAT	TGGCTT GG T	<u>TGCTCGCAGT</u>	TGAAAGTAGA	GACACCAAAC
FpGLDPD_2	238	AACACACGAT	TGGCTT GA T	<u>TCTTCGCTGT</u>	TAAAGTAAA	GACACCAAAC
		**** *	*****	* * * * *	*****	* ***** *

Figure 2. The transcriptional start site in region 2 and its characteristics in different *Flaveria* species.

Sequence comparison between the transcriptional start site of the *GLDPA* genes of five *Flaveria* species, namely *F. trinervia* (C₄), *F. bidentis* (C₄), *F. pubescens* (C₃-C₄), *F. pringlei* (C₃) and *F. robusta* (C₃) and the *GLDPD* gene of *F. pringlei* (C₃) giving detailed information of the TSS in region 2. The CT-elements upstream of TSS_{R2} are marked in red. The sequence of the initiator motif is underlined and the transcription starting A, or G in the case of *F. robusta*, is marked in red and blue.

In case of the *GLDPA*-Ft tandem promoter an initiator (Inr) motif (Yamamoto et al., 2007) and CT-elements (Bernard et al., 2010) are located at the TSS in region 2 (chapter 1: manuscript 1). Upon analyses of the putative TSS in region 2 of additional four *GLDPA* 5' upstream regions, namely *F. bidentis* (C₄), *F. pubescens* (C₃-C₄ intermediate), *F. pringlei* (C₃) and *F. robusta* (C₃), it is evident that, although the CT-elements are mostly conserved (Figure 1) in the proposed range of -50 bp to +50 bp (Zuo and Li, 2011), the TSS itself shows, at least in the case of *F. robusta*, some divergence. In most cases transcription in plants starts with an A (Joshi, 1987) - Sawant et al. (2001) confirmed these findings and even stated that in 62% of

all highly expressed genes an A occurs at the TSS preceded by a C. All *GLDPA* TSSs in region 2 show this CA following the YR rule (Y stands for C or T and R for A or G) (Yamamoto et al., 2007), except for *GLDPA*-Fro where a GG can be found at this position (Figure 2).

The corresponding region of the group II gene *GLDPD* contains an A at this position but a G precedes this A, thus also not following the YR rule (Yamamoto et al., 2007). In addition, the CT-elements upstream of the TSS (Figure 2) are not conserved in the 5' flanking region of *GLDPD*. Thus, both the TSS of *GLDPA* of *F. robusta* and the putative TSS of *GLDPD* of *F. pringlei* show differences compared to those of the other analysed *GLDPA* genes.

Discussion

The finding that the distal TSS of *F. robusta* differs from the TSSs of the other *GLDPA* genes is highly consistent with the fact that *F. robusta* is the one plant that does not show any transcription from TSS_{R2} at all (chapter 2: manuscript 2 Figure 8D). It is further noticeable that the median activity measured for GUS activity of the distal promoter in *F. robusta* is low (6.9 nmol MU * mg⁻¹ protein * min⁻¹; see chapter 2: manuscript 2 Supplemental Figure 3) compared to the median activity of the distal promoter of *F. trinervia* (20.41 nmol MU * mg⁻¹ protein * min⁻¹; chapter 1: manuscript 1 Figure 6). The transcriptional start site might have changed in the transition from C₃ to the C₃-C₄ intermediate state, leading to a more effectively transcribed start site.

This leads to the assumption that either *F. robusta* is the more basal C₃ species in the genus *Flaveria* or that the *GLDPA* gene of *F. pringlei* originates indeed from hybridization with the intermediate species *F. angustifolia* (Kopriva et al. 1996; McKown et al. 2005) and thus must be considered as C₃-C₄ intermediate *GLDPA* gene. If the second alternative is true, the *GLDPB* gene of *F. pringlei* might be the more ancestral gene. The sequence of *GLDPB* of *F. pringlei* at TSS_{R2} it is not yet known, due to sequence limitations. Thus, it remains unclear, if *GLDPB*-Fp possesses CA or GG or even a GA as can be found in the putative TSS of *GLDPD* (Figure 2) at this position. It remains to be investigated if minor amounts of transcripts derive from the distal TSS of this gene, as it is the case for *GLDPA* of *F. pringlei* (see chapter 2: manuscript 2 Figure 8) or not.

The putative TSS of the corresponding promoter of *GLDPD* of *F. pringlei* is not experimentally confirmed yet. The fact that this part of the 5' flanking region of this gene functions as a promoter in promoter-reporter gene studies indicates that the transcription may

start there. The sequence comparison to the TSSs of the *GLDPA* genes illustrate a high diversity in the composition of the CT-elements that are located upstream of the TSS. Since a 5' RACE experiment with RNA isolated from leaves of *F. pringlei* did not show any indication for transcripts derived from a distal TSS in the *GLDPD* gene (data not shown), it can be assumed that this promoter may be cryptic like the distal promoter of *GLDPA* of *F. robusta* (see chapter 2: manuscript 2 Figure 8) in the context of the complete 5' flanking region. It is therefore possible that both the CA at the site of the TSS and the CT-elements are necessary for the activation of the promoter in the transition from C₃ to C₄ photosynthesis. The isolation of 5' flanking regions of more *GLDP* genes from different species as well as from *Sartwellia* spec., the potential ancestral species of *Flaveria*, will probably help to elucidate this question.

The fact that the structure of the distal promoter is also present in the group II *GLDPD* gene of *F. pringlei* gives two options, concerning the origin of this sequence in the 5' flanking sequence of the *GLDP* genes. The less probable one is that this promoter element originates from a transposable element and both 5' flanking regions gained this sequence independently from each other. It is more likely that the sequence was already part of the 5' flanking sequence before the gene duplicated (see chapter 2: manuscript 2 Figure 9). If the second case is true it is still possible that the sequence of the distal promoter originates from a transposable element, since transposable elements are able to activate transcription in a eukaryotic genome (Naito et al., 2009; Batut et al., 2012). A sequenced genome of a *Flaveria* species might help to settle this question. The genome might contain this sequence or identifiable derivatives of it in a higher number than a random sequence, if it originates from a transposable element.

Material and Methods

Standard molecular procedures such as polymerase chain reaction (PCR), reverse transcription, enzymatic restriction of DNA and gel electrophoresis were carried out after Sambrook et al. (Sambrook and Russell 2001).

Generation of promoter-reporter gene constructs

Transformation constructs for *Arabidopsis thaliana* plants were prepared in the pBI121 vector (Jefferson et al. 1987; Chen et al. 2003). Promoter DNA sequences were amplified with PCR and the correctness was verified with commercial sequencing (LGC Genomics (formerly Agowa), Berlin, GER or GATC Biotech, Konstanz, GER) Promoter constructs in the vector pBI121 were inserted using restriction sites available in the multiple cloning site (MCS) of the vector.

Sequence alignments

Sequence comparisons for similarity studies between the promoters were done with the Align sequence Nucleotide BLAST tool (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Altschul et al. 1990; States and Gish 1994; Zhang et al. 2000).

Transformation of *Arabidopsis thaliana*

Transformation of *Arabidopsis thaliana* plants was performed as described in chapter 2: manuscript 2.

In situ analysis of the β -glucuronidase and its fluorimetric measurement

The *in situ* analysis of transgenic plants carrying β -glucuronidase constructs and fluorimetric measurements of the activity were performed as described in chapter 2: manuscript 2.

Table 1: Oligonucleotides used for the construction of reporter gene constructs. Underlined italic fonts indicate restriction sites. *HindIII* (aagcct), *XmaI* (ccggg).

Oligonucleotide	Sequence 5' → 3'
FpGLDPD2analog5'fw_ <i>HindIII</i>	<u>aagcct</u> ttgtctcccattgtgtctga
FpGLDPD3analog3'rv_ <i>XmaI</i>	cccggg <u>ttt</u> caaacacccaaaagaaa

Literature

- Batut, P., Dobin, A., Plessy, C., Carninci, P., and Gingeras, T.R.** (2013). High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. *Genome Res* **23**, 169-180.
- Bernard, V., Brunaud, V., and Lecharny, A.** (2010). TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. *BMC Genomics* **11**, 166.
- 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 Breeding* **11**, 287-293.
- 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 C₄ species *Flaveria trinervia*: analysis of transcriptional control in transgenic *Flaveria bidentis* (C₄) and *Arabidopsis* (C₃). *Plant Physiol* **146**, 1773-1785.
- 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.
- Joshi, C.P.** (1987). An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Res* **15**, 6643-6653.
- 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.
- Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C.N., Richardson, A.O., Okumoto, Y., Tanisaka, T., and Wessler, S.R.** (2009). Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* **461**, 1130-1134.
- McKown, A.D., Moncalvo, J.M., and Dengler, N.G.** (2005). Phylogeny of *Flaveria* (Asteraceae) and inference of C₄ photosynthesis evolution. *Am J Bot* **92**, 1911-1928.
- Sambrook, J., and Russell, D.** (2001). *Molecular cloning: A laboratory manual*. (Cold Spring Harbor Laboratory Press).
- Sawant, S., Singh, P.K., Madanala, R., and Tuli, R.** (2001). Designing of an artificial expression cassette for the high-level expression of transgenes in plants. *TAG Theor Appl Genet* **102**, 635-644.
- States, D.J., and Gish, W.** (1994). Combined use of sequence similarity and codon bias for coding region identification. *J Comput Biol* **1**, 39-50.

- Yamamoto, Y.Y., Ichida, H., Matsui, M., Obokata, J., Sakurai, T., Satou, M., Seki, M., Shinozaki, K., and Abe, T.** (2007). Identification of plant promoter constituents by analysis of local distribution of short sequences. *BMC Genomics* **8**, 67.
- Zhang, Z., Schwartz, S., Wagner, L., and Miller, W.** (2000). A greedy algorithm for aligning DNA sequences. *J Comput Biol* **7**, 203-214.
- Zuo, Y.C., and Li, Q.Z.** (2011). Identification of TATA and TATA-less promoters in plant genomes by integrating diversity measure, GC-Skew and DNA geometric flexibility. *Genomics* **97**, 112-120.

Chapter 4

Identification of *cis*-regulatory elements for the functioning of the *GLDPA* 5' flanking region

Identification of *cis*-regulatory elements for the functioning of the *GLDPA* 5' flanking region

Introduction

In C_4 plants photorespiration is exclusively performed in the bundle sheath cells. Thus, the enzymes involved in this process had to be relocated to this cell type in the transition from C_3 to C_4 photosynthesis. One of the key enzymes of photorespiration is glycine decarboxylase (GDC). It was shown that the gene of the PA protein of GDC (*GLDPA*) of the C_4 plant *Flaveria trinervia* contains a 5' flanking region that is sufficient to confer this bundle sheath specific expression (Engelmann et al., 2008). Further this 5' flanking region contains two promoters oriented in tandem (chapter 1: manuscript 1).

The proximal promoter P_{R7} , defined by region 7, of the *GLDPA* gene of the C_4 plant *F. trinervia* drives *GUS* reporter gene expression exclusively in bundle sheath cells of leaves both in transgenic *F. bidentis* (C_4) and *A. thaliana* (C_3) plants (chapter 1: manuscript 1). This leads to the assumption that within this region some form of *cis*- regulatory element must be located that confers the bundle sheath specific expression. Further region 7 contains some element that suppresses the ubiquitous expression of the distal promoter, as the construct *GLDPA*-Ft1-2-3 is expressed ubiquitously, whereas *GLDPA*-Ft1-2-3-7 showed reporter gene expression restricted to the bundle sheath cells (Engelmann et al., 2008).

A further category of *cis*-regulatory elements enhances the expression in general if they are part of a promoter. Region 1 of the 5' flanking region of *GLDPA* of *F. trinervia* was found to function as such an enhancing element (chapter 1: manuscript 1).

Results

Two sub-regions of region 7 take part in maintaining bundle sheath specificity

In order to find *cis*-regulatory elements involved in the bundle sheath specificity in the 198 bp long fragment of region 7 up to the 5' untranslated region (UTR) of *GLDPA*-Ft region 7 was dissected in four overlapping parts (Figure 1).

The *GUS* reporter gene activity driven by region 7 alone is quite low (see chapter 1: manuscript 1 Figure 4), thus the assumption was that any deletion constructs might lead to an undetectable activity of the *GUS* reporter. One construct that showed the same spatial expression pattern but higher reporter gene activity in transgenic *A. thaliana* plants is

GLDPA-Ft1-2-3-7 (Engelmann et al. 2008). The combined existence of region 3 and region 7 effectively suppresses the expression, deriving from the distal promoter in region 2, of the *GUS* reporter gene in the mesophyll cells (Engelmann et al. 2008 and chapter 1: manuscript 1). Thus, the construct *GLDPA-Ft1-2-3-7* was selected as the basis of the deletion analysis of region 7 (Figure 1A).

The deletion analyses show that none of the four constructs (*GLDPA-Ft1-2-3-7_1-7UTR*, *GLDPA-Ft1-2-3-7_2-7UTR*, *GLDPA-Ft1-2-3-7_3-7UTR* and *GLDPA-Ft1-2-3-7_4-7UTR*) is able to completely drive the same spatial expression pattern of the *GUS* reporter gene as the construct *GLDPA-Ft1-2-3-7* does (cf. Engelmann et al. 2008). The insertions of the parts *7_2* and *7_3* are not able to suppress the mesophyll cell activity deriving from the distal promoter (Figure 1B). The parts *7_4* and even more so *7_1* on the other hand are able to partially suppress the mesophyll cell activity, but both of them do not lead to a complete bundle sheath restricted expression of the reporter gene (Figure 1B). Activity measurements of the four constructs show differences between them. Especially *GLDPA-Ft1-2-3-7_3-7UTR* led to a higher median activity, and thus show a significant difference ($p < 0.005$) to the other three constructs (Figure 1C).

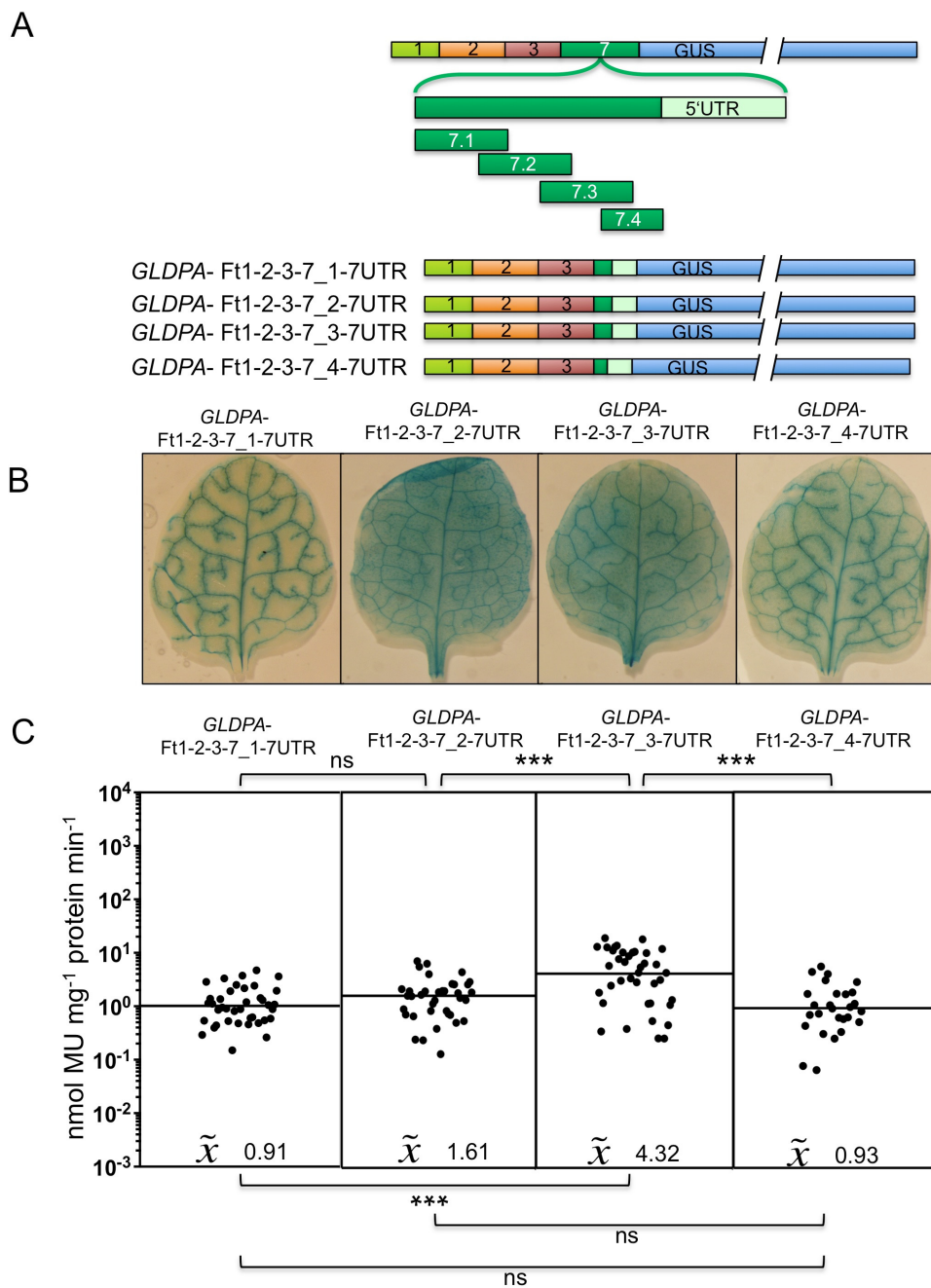


Figure 1. Functional analysis of deletion fragments of region 7 and its influence on the expression.

(A) Schematic structure of *GUS* reporter gene constructs harbouring the deletion fragments of region 7 of *GLDPA*-Ft.

(B) Histochemical *GUS* stainings of *A. thaliana* leaves harbouring *GLDPA*-Ft1-2-3-7_1-7UTR (left), *GLDPA*-Ft1-2-3-7_2-7UTR (second from left), *GLDPA*-Ft1-2-3-7_3-7UTR (second from right) and *GLDPA*-Ft1-2-3-7_4-7UTR (right). Incubation times were 3 h 30 min for all leaves.

(C) Fluorimetric measurements of *GUS* activities in leaves of *A. thaliana* plants with either *GLDPA*-Ft1-2-3-7_1-7UTR (left), *GLDPA*-Ft1-2-3-7_2-7UTR (second from left), *GLDPA*-Ft1-2-3-7_3-7UTR (second from right) and *GLDPA*-Ft1-2-3-7_4-7UTR (right). The black line indicates the median of the measured data from all independent lines. (***) $p < 0.005$, Mann – Whitney test). MU: 4-methylumbelliferone.

The region upstream of region 1 of *GLDPA* of *F. pringlei* increases the enhancing effect of region 1

The 5' flanking sequences of the *GLDPA* genes of *F. trinervia* and *F. pringlei* were originally isolated from fragmented genomic clones in the 1990s (Kopriva and Bauwe 1994; Bauwe et al. 1995; Chu 1996; Cossu and Bauwe 1998). This led to the fact that the 5' flanking sequences are not equal in these clones. *GLDPA* of *F. trinervia* is the gene, whose 5' flanking sequence was first analysed in detail with the construction of deletion constructs (Engelmann et al. 2008). Due to high sequence conservation, the regions that were applied to *GLDPA*-Ft can also be adapted for *GLDPA* genes of other *Flaveria* species (see chapter 2: manuscript 2 Figure 5). The originally isolated 5' flanking sequence of *GLDPA*-Fp reached farther upstream than the 5' flanking sequence of *GLDPA*-Ft, thus when the regions were adapted, a part of the 5' flanking sequence was excluded for better comparison. However, the construct containing the distal promoter P_{R2} of *GLDPA* of *F. pringlei* (*GLDPA*-Fp1-2) was constructed in two different designs: one starting at the originally isolated 5' sequence (*GLDPA*-Fp1*-2) end and one starting comparable to *GLDPA*-Ft (*GLDPA*-Fp1-2). Both constructs were transformed into *Arabidopsis thaliana* plants (Figure 2).

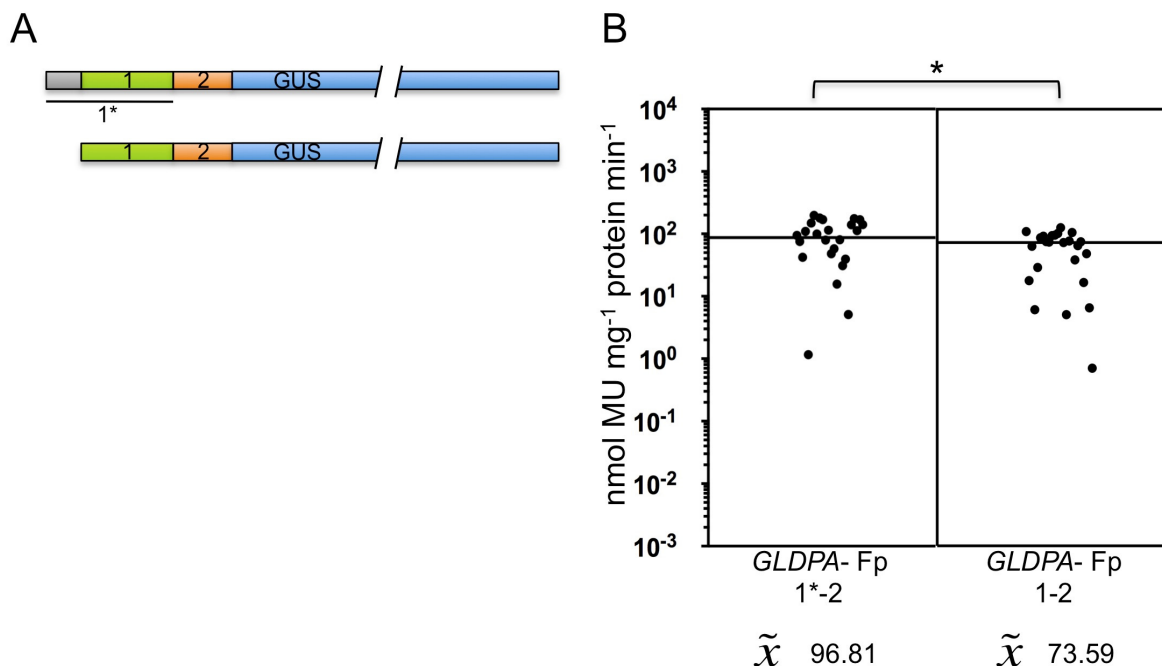


Figure 2. Influence of the sequence upstream of region 1 on the expression strength.

(A) Schematic structure of *GUS* reporter gene constructs harbouring *GLDPA*-Fp1-2 either in the original sequence length (upper) or in the comparable length to *GLDPA*-Ft (lower).

(B) Fluorimetric measurements of *GUS* activities in leaves of *A. thaliana* plants harbouring *GLDPA*-Fp1*-2 (left) or *GLDPA*-Fp1-2 (right). The black line indicates the median of the measured data from all independent lines. The measurements show a significant difference as indicated by the astericks ($p < 0.05$; Mann-Withney test). MU: 4-methylumbelliferone.

The analyses of the constructs *GLDPA*-Fp1*-2 and *GLDPA*-Fp1-2 show that both constructs drive the expression of the *GUS* reporter gene in the same spatial pattern (data not shown; cf chapter 2: manuscript 2 Figure 6). *GUS* staining is visible in all photosynthetically active tissues of the leaves. The activity measurements indicate that the attachment of the sequence upstream of region 1 leads to a significant increase ($p < 0.05$) of the activity of the promoter. Region 1 already has an enhancing effect (Engelmann et al. 2008 and chapter 1: manuscript 1). The comparison of the two constructs (Figure 1) indicates that the region upstream of region 1 enhances the activity even more.

The region upstream of region 1 is conserved in the *GLDPA* genes

Genome walking was performed to isolate the 5' flanking sequence of *GLDPA* of *F. bidentis* and the sequence upstream of region 1 of *GLDPA* of *F. trinervia*. Sequence alignments show that the conservation between the 5' flanking sequences of the *GLDPA* genes of *F. trinervia*, *F. bidentis* and *F. pringlei* reaches farther upstream than the annotated region 1 (Figure 3).

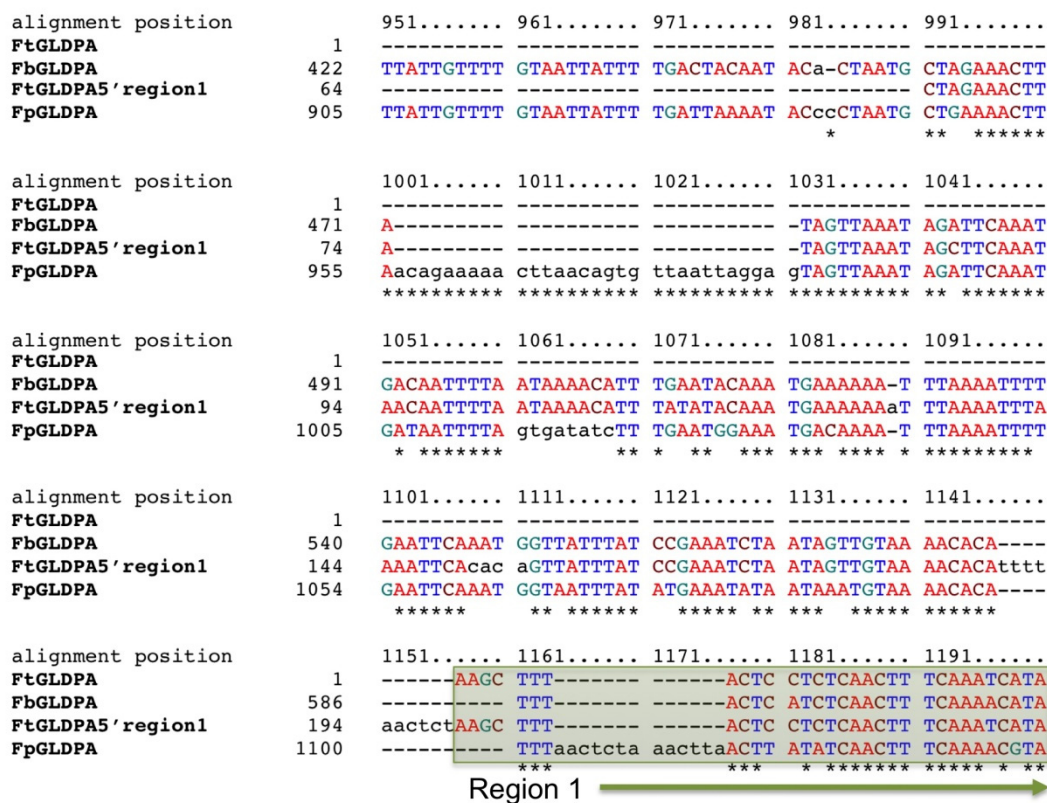


Figure 3. Alignment of the sequence upstream of region 1 of *F. trinervia*, *F. bidentis* and *F. pringlei*.

5' flanking sequences upstream of region 1 were aligned with the Genomatix DiAlign online alignment tool. Capitals indicate identity between sequences and asterisks indicate identity between all sequences. A green box indicates the start of region 1. Ft, *F. trinervia*; Fb, *F. bidentis*, Fp, *F. pringlei*.

The conservation of the sequence upstream of region 1 and the enhancing effect of this sequence in the analysed *GLDPA*-Fp1-2 construct (Figure 2) indicate that this sequence might be part of the upstream region of the *GLDPA* gene in *Flaveria* in general. Thus, the analysed GUS reporter gene constructs might not contain the full activity potential of the promoters of *GLDPA* genes of *Flaveria* species and the enhancing effect of region 1 and its upstream sequence might be even higher than originally found.

Discussion

The analyses of the deletion constructs of region 7 suggest that none of the selected parts contain all *cis*-regulatory elements that are necessary to confer the complete bundle sheath cell specificity that can be seen in the construct *GLDPA*-Ft1-2-3-7. It rather looks like as if the parts 7_1 and 7_4 are needed in combination. To test this possibility a construct that contains both parts fused together needs to be analysed. To eliminate the eventuality that a certain distance is needed between those two parts to allow the access of *trans*-acting factors, a second construct with a non-interfering spacer should be analysed in parallel.

The longest uORF that might interfere with transcript stability (see chapter 1: manuscript 1) is located in part 7_4 and thus might be the cause for a partial suppression of the mesophyll cell expression in the construct *GLDPA*-Ft1-2-3-7_4_7UTR. An analysis concerning this issue can be found in chapter 5. As for the function of part 7_1 it remains to be investigated if a *cis*-regulatory sequence is located within this part. Further deletions will allow narrowing down such a sequence that later on be used in, for instance, yeast-two hybrid screens to identify any binding partners.

The even more enhancing effect of the region upstream of region 1 of the *GLDPA* genes may be used for the analysis of the bundle sheath specificity of region 7 independent from region 2, thus independent from the distal promoter. A construct containing the prolonged region 1 and region 7 may show an activity high enough to analyse the deletions of region 7 directly. It further needs to be elucidated if this enhancing effect works in general, thus it has to be analysed in a heterologous background. Such a background could, for instance, be the promoter of the sulphate transporter 2;2 (Sultr 2;2) of *A. thaliana* that shows a bundle sheath specific expression but leads to very low activities in promoter-reporter gene studies (Prusko; 2010, and Sandra Kirschner personal communication).

Material and Methods

Standard molecular procedures such as polymerase chain reaction (PCR), reverse transcription, enzymatic restriction of DNA and gel electrophoresis were carried out after Sambrook et al. (Sambrook and Russell 2001).

Isolation of *GLDP* 5' flanking regions by vectorette PCR

Sequence upstream of the 5' flanking regions of *GLDPA* genes from *F. trinervia* 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 (Clontech, Mountain View, US). DNA for library construction was isolated from leaves of the respective *Flaveria* species with the Qiagen DNeasy Plant Mini Kit (Qiagen, Hilden, GER).

Generation of promoter-reporter gene constructs

Transformation constructs for *Arabidopsis thaliana* plants were prepared in the pBI121 vector (Jefferson et al. 1987; Chen et al. 2003). Promoter DNA sequences were either amplified with PCR and the correctness was verified with commercial sequencing (LGC Genomics (formerly Agowa), Berlin, GER or GATC Biotech, Konstanz, GER) or synthesized as complete fragments by GenScript (Piscataway, USA). Promoter constructs in the vector pBI121 were inserted using restriction sites available in the multiple cloning site (MCS) of the vector.

Sequence alignments

Sequence comparisons for similarity studies between the promoters were performed with the Genomatix DiAlign web interface (<http://www.genomatix.de/cgi-bin/dialign/dialign.pl>) (Morgenstern et al. 1996; Morgenstern et al. 1998; Morgenstern 1999).

Transformation of *Arabidopsis thaliana*

Transformation of *Arabidopsis thaliana* plants was performed as described in chapter 2: manuscript 2.

***In situ* analysis of the β -glucuronidase and its fluorimetric measurement**

The *in situ* analysis of transgenic plants carrying β -glucuronidase constructs and fluorimetric measurements of the activity were performed as described in chapter 2: manuscript 2.

Literature

- Bauwe, H., Chu, C.C., Kopriva, S., and Nan, Q.** (1995). Structure and expression analysis of the *gdcSP* and *gdcPB* genes encoding two P-isoproteins of the glycine-cleavage system from *Flaveria pringlei*. *Eur J Biochem* **234**, 116-124.
- Brogna, S., and Wen, J.** (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat Struct Mol Biol* **16**, 107-113.
- 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 Breeding* **11**, 287-293.
- Chu, C.** (1996). Molecular structure and expression patterns of glycine decarboxylase genes from *Flaveria pringlei* (C₃) and *Flaveria anomala* (C₃-C₄). [PhD thesis]. Martin-Luther University, Halle-Wittenberg.
- Cossu, R., and Bauwe, H.** (1998). The electronic Plant Gene Register. *Plant Physiol* **116**, 445-446.
- 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 C₄ species *Flaveria trinervia*: analysis of transcriptional control in transgenic *Flaveria bidentis* (C₄) and *Arabidopsis* (C₃). *Plant Physiol* **146**, 1773-1785.
- Hanson, A.D., and Roje, S.** (2001). One-carbon metabolism in higher plants. *Annu Rev Plant Physiol Plant Mol Biol* **52**, 119-137.
- 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.
- Kopriva, S., and Bauwe, H.** (1994). P-protein of glycine decarboxylase from *Flaveria pringlei*. *Plant Physiol* **104**, 1077-1078.
- Mazo, A., Hodgson, J.W., Petruk, S., Sedkov, Y., and Brock, H.W.** (2007). Transcriptional interference: an unexpected layer of complexity in gene regulation. *J Cell Sci* **120**, 2755-2761.
- 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 U S A* **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.
- 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.
- Prusko K. (2010).** Identifizierung *cis*-regulatorischer Elemente der Transkriptionskontrolle in photosynthetisch aktiven Blattzellen von *Arabidopsis thaliana*. [PhD thesis]. Heinrich Heine Universität, Düsseldorf.
- Sambrook, J., and Russell, D.** (2001). *Molecular cloning: A laboratory manual*. (Cold Spring Harbor Laboratory Press).
- Shearwin, K.E., Callen, B.P., and Egan, J.B.** (2005). Transcriptional interference--a crash course. *Trends Genet* **21**, 339-345.

Chapter 5

The roles of splicing and upstream open reading frames on the expression of *GLDPA* of *F. trinervia*

The roles of splicing and upstream open reading frames on the expression of *GLDPA* of *F. trinervia*

Introduction

The *GLDPA* gene, encoding the PA protein of glycine decarboxylase, of *Flaveria trinervia* (C₄) contains a tandem promoter in its 5' flanking sequence. The distal promoter produces a transcript that has to undergo splicing to be functional. The spliced transcript uses an alternative ATG, which is located downstream of the original ATG. This leads to a shorter, albeit still functional, mitochondrial-targeting sequence of the resulting protein (see chapter 1: manuscript 1).

The distal promoter, if detached from the rest of the 5' flanking sequence, drives expression of the *GUS* reporter gene in all photosynthetically active tissues of the leaf (see chapter 1: manuscript 1 Figure 5). *GUS* activity measurements indicate a high activity of this promoter. In contrast, RNAseq shows that transcripts originating from the distal promoter are quite rare in the context of the complete 5' flanking region (see chapter 1: manuscript 1 and chapter 2: manuscript 2 Figure 8A). It is assumed that upstream open reading frames (uORFs) negatively influence the RNA stability of these transcripts. These uORFs lead to a degradation of unspliced transcripts (see chapter 1: manuscript 1). All analysed promoter-reporter gene constructs lacked the splice acceptor site since it is located in the coding region of the *GLDPA* gene. The influence of the possibility of splicing is examined in constructs that have this splice acceptor site inserted upstream of the *GUS* reporter gene.

The longest uORFs that might influence the RNA stability is located directly upstream of the transcriptional start site of the proximal promoter. A construct containing only the uORF comprising part of region 7 and missing the others (*GLDPA*-Ft1-2-3-7_4-7UTR; see chapter 4 Figure 1) showed a bundle sheath preferential expression. To elucidate whether the uORF has any influence in this expressional behaviour a substitution of the ATG was performed.

Results

The possibility of splicing stabilizes transcripts from the distal promoter

Two promoter-reporter gene constructs were generated that lacked the proximal promoter but contained the complete rest of the 5' flanking region of the *GLDPA* gene of *F. trinervia*. One of them additionally contained the splice acceptor site of the coding region of the *GLDPA*

gene upstream of the *GUS* reporter gene (Figure 1A). The second ATG in this case is that of the *GUS* gene.

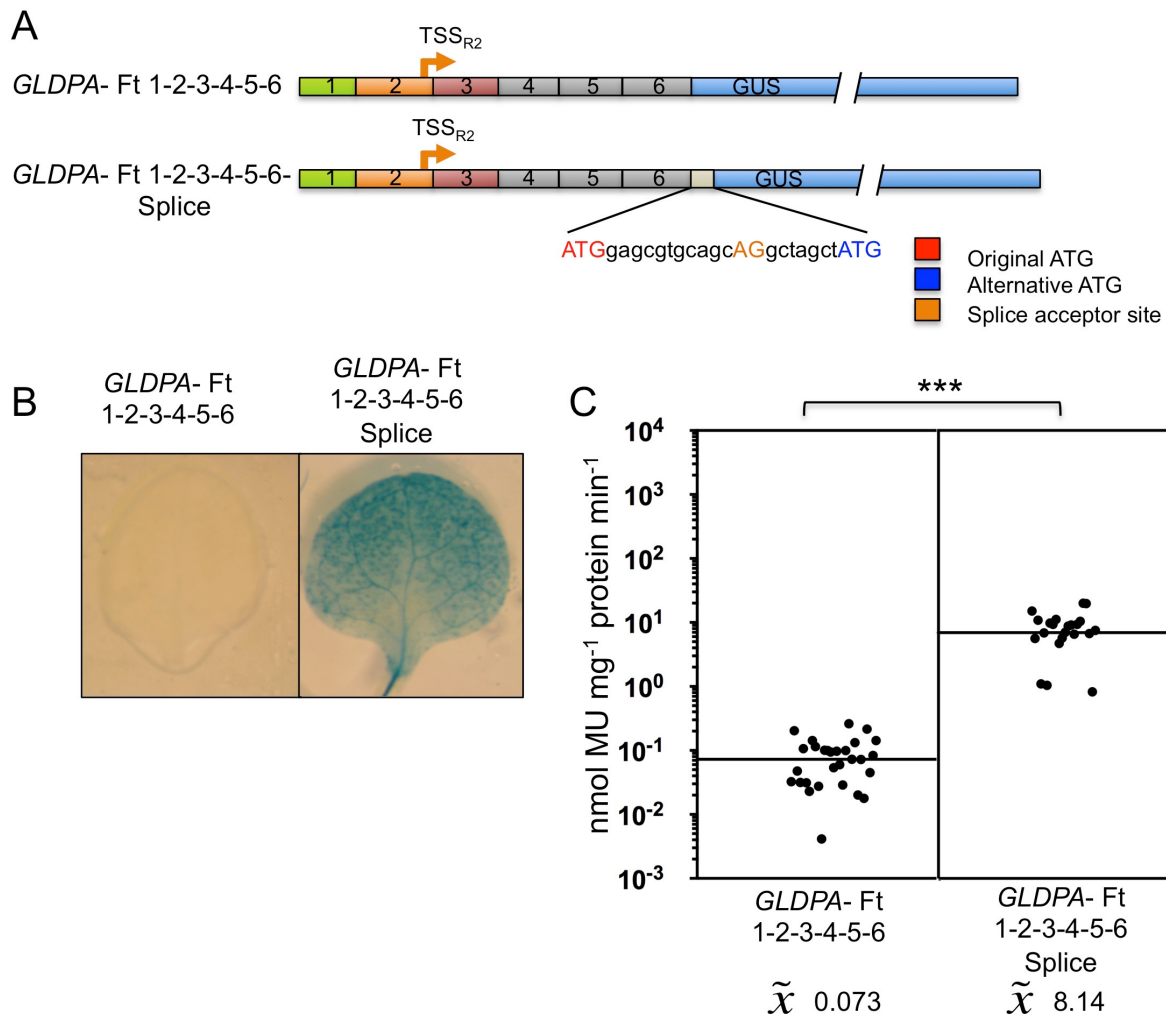


Figure 1. Functional analysis of the influence of the splice acceptor site on the distal promoter of *GLDPA* of the *C₄* species *F. trinervia*.

(A) Schematic structure of *GUS* reporter gene constructs harbouring the splice site of *GLDPA*- Ft and the regions 1-6. TSS: Transcription start site.

(B) Histochemical *GUS* stainings of *A. thaliana* leaves harbouring *GLDPA*- Ft1-2-3-4-5-6 (left, 1273 bp) or *GLDPA*- Ft1-2-3-4-5-6-Splice (right, 1273 bp plus 24 bp). Incubation times were 4 h for both leaves.

(C) Fluorimetric measurements of *GUS* activities in leaves of *A. thaliana* plants with either *GLDPA*- Ft1-2-3-4-5-6 or *GLDPA*-Ft1-2-3-4-5-6-Splice. The black line indicates the median of the measured data from all independent lines. (***) $p < 0.0001$, Mann – Whitney test). MU: 4-methylumbelliferone.

GUS constructs harbouring either *GLDPA*-Ft1-2-3-4-5-6 (1273 bp) or *GLDPA*-Ft1-2-3-4-5-6-Splice (1273 bp plus 24 bp) (Figure 1A) lead to a clear difference both in spatial expression patterns (Figure 1B), and in the activity of the reporter gene (Figure 1C). Whereas plants with the construct without the splice acceptor site (*GLDPA*-Ft1-2-3-4-5-6) show no *GUS* staining, in a comparable timeframe, and a low reporter gene activity, plants

transformed with the construct that contains the splice acceptor site (*GLDPA*-Ft1-2-3-4-5-6-Splice) show GUS staining throughout the leaf. The expression pattern of *GLDPA*-Ft1-2-3-4-5-6-Splice resembles the expression that is observed for the distal promoter *GLDPA*-Ft (*GLDPA*-Ft-2) (cf. chapter 1: manuscript 1 Figure 6).

Stabilized transcripts from the distal promoter negatively influence transcripts from the proximal promoter

The analysis of plants transformed with the complete 5' flanking sequence that contains both promoters and the splice acceptor site plus the second ATG (*GLDPA*-Ft-Splice; 1571 bp plus 24 bp) (Figure 2A), reveals that the spatial expression pattern of the reporter gene (Figure 2B) resembles that of the full length 5' flanking sequence without the splice acceptor site (cf. Engelmann et al. 2008 and chapter 1: manuscript 1 Figure 1).

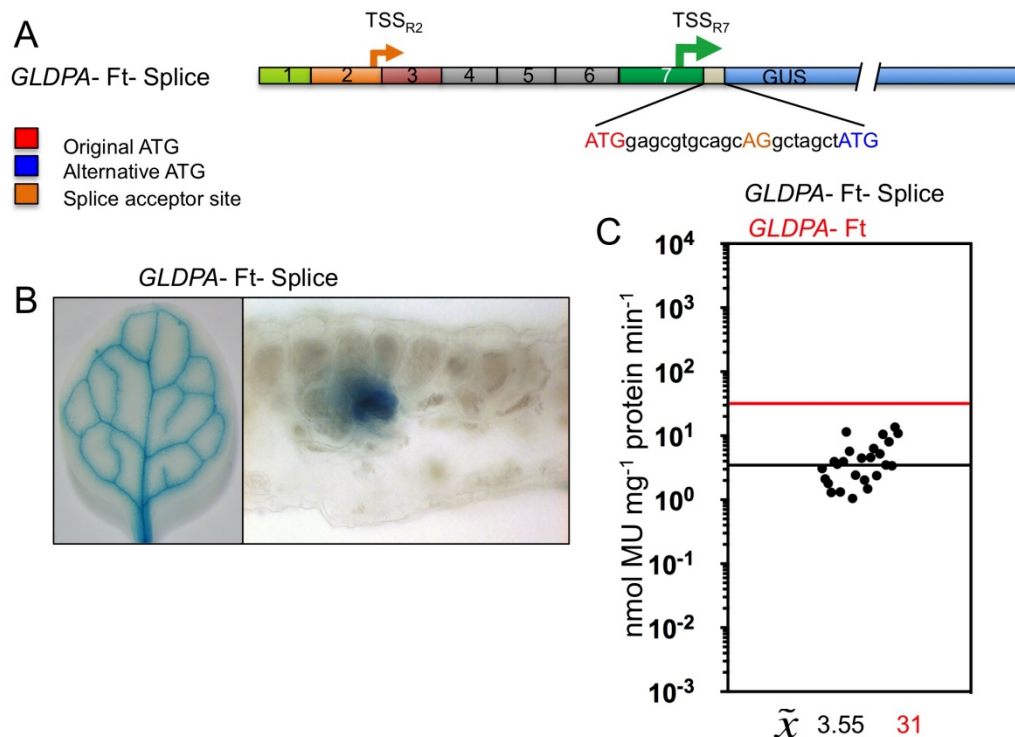


Figure 2. Functional analysis of the influence of the splice acceptor site on the distal promoter in the tandem promoter context of *GLDPA* of the *C*₄ species *F. trinervia*.

(A) Schematic structure of *GUS* reporter gene constructs harbouring the splice site of *GLDPA*-Ft and the composition of *GLDPA*-Ft. TSS: Transcriptional start site.

(B) Histochemical GUS staining of an *A. thaliana* leaf (left) and a leaf cross section (right) harbouring *GLDPA*-Ft-Splice (1571 bp plus 24 bp). Incubation times were 6 h 20 min and 1 h 10 min respectively.

(C) Fluorimetric measurements of GUS activities in leaves of *A. thaliana* plants harbouring *GLDPA*-Ft-Splice. The black line indicates the median of the measured data from all independent lines. The red line indicates the median of measurements for *GLDPA*-Ft performed by Sascha Engelmann (Engelmann et al., 2008). MU: 4-methylumbelliferone.

GUS staining is only detectable in the bundle sheath cells and the vasculature. The GUS activity, in contrast, shows a difference in comparison to the *GUS* reporter gene construct without the splice acceptor site. The median activity is tenfold lower than the median activity that is measured without the splice acceptor site (Figure 2C).

The longest upstream open reading frame in region 7 influences stability of transcripts from the distal promoter

To further investigate the possible influence of the long uORF in part 7_4 (see chapter 4 Figure 1), a construct was designed that consists of the same sequence as *GLDPA*-Ft1-2-3-7_4-7UTR, but at the same time had the T of the startcodon (ATG) of the uORF altered to a G leading to AGG instead of ATG (Figure 3A). *A. thaliana* plants were transformed with this reporter gene construct and seedlings were stained to detect the spatial expression pattern.

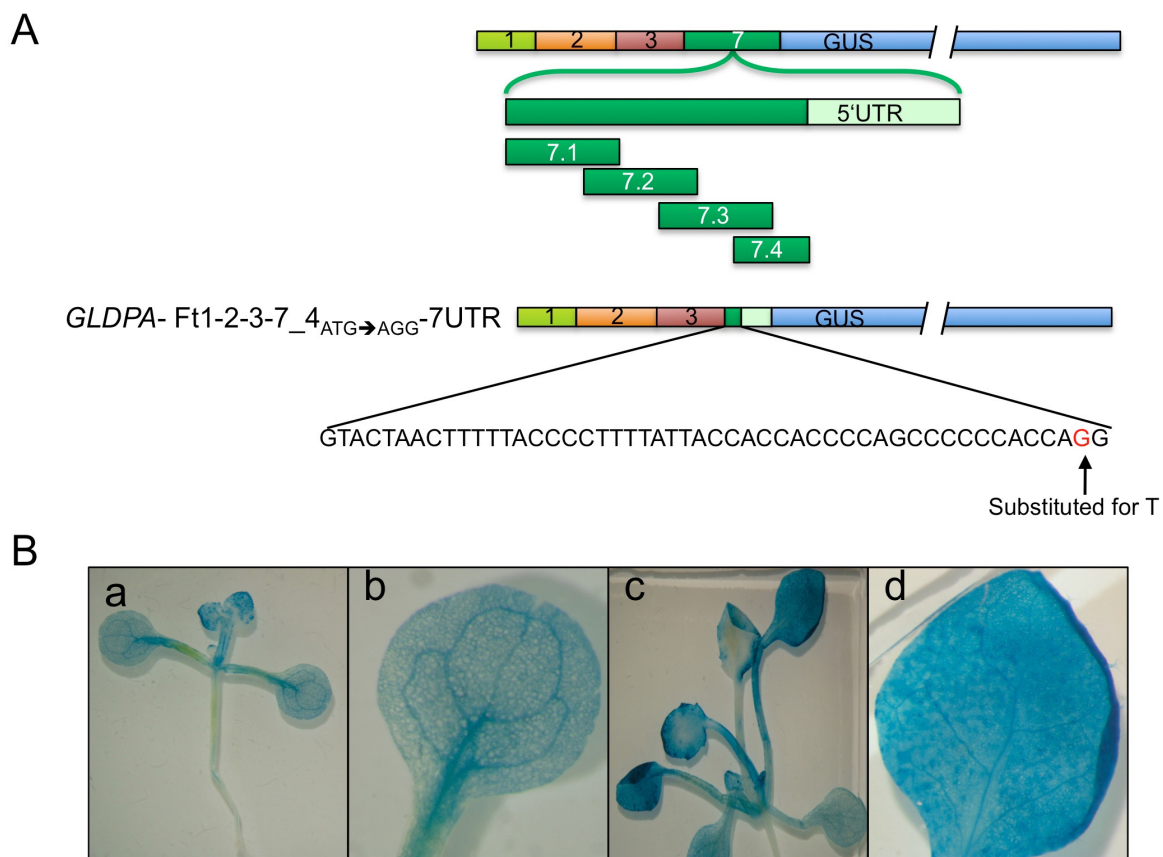


Figure 3. Functional analysis of the influence the upstream open reading frame on the suppression of the distal promoter of *GLDPA* of *F. trinervia*.

(A) Schematic structure of *GUS* reporter gene constructs harbouring the deletion fragment 7_4 with a substitution of the ATG of the uORF of region 7 of *GLDPA*-Ft to AGG.

(B) Histochemical GUS stainings of *A. thaliana* seedlings (a and c), cotyledon (b) and the first leaf (d) harbouring *GLDPA*-Ft1-2-3-7_4_{ATG→AGG}-7UTR. Incubation times were 3 h 20 min (a and b) and 3h (c and d), respectively.

Expression of the *GUS* reporter gene is maintained, but the spatial expression pattern changes. The leaves of the transgenic plants containing *GLDPA-Ft1-2-3-7_4_{ATG-AGG}-7UTR* show a ubiquitous staining (Figure 3B). All photosynthetically active tissues express the *GUS* reporter gene and no preference to the bundle sheath cells and the vasculature is detectable any more (cf. chapter 4 Figure 1B). The change in the spatial expression pattern might indicate that the exchange of ATG to AGG indeed no longer leads to a degradation of transcripts from TSS_{R2}. Thus, the ubiquitously expressed transcripts from the distal promoter might interfere with those transcripts from the proximal promoter and the bundle sheath specific expression of the proximal promoter is overlain by the ubiquitous expression of the distal promoter.

Discussion

Both promoters influence each other's transcription

The analyses of *GLDPA-Ft1-2-3-4-5-6* and *GLDPA-Ft1-2-3-4-5-6-Splice* indicate a stabilizing effect of the splice acceptor site on transcripts deriving from the transcriptional start site (TSS) in region 2 (TSS_{R2}). While *GLDPA-Ft1-2-3-4-5-6* can only lead to a transcript that contains all downstream regions in front of the translational start of the *GUS* gene, *GLDPA-Ft1-2-3-4-5-6-Splice* allows splicing, thus eliminating region 4 to 6. Any uORFs that might inhibit the RNA stability can be removed with the splicing step. This reduces the likelihood of the transcript to undergo “nonsense mediated mRNA decay” (NMD). This mechanism is effective whenever mRNAs contain reading frames that produce transcripts not leading to functional proteins (Kertesz et al. 2006; Brogna and Wen 2009; Nyiko et al. 2009). The fact that the median *GUS* activity of *GLDPA-Ft1-2-3-4-5-6* (8.14 nmol MU * mg⁻¹ protein * min⁻¹) is not even half as high as the median *GUS* activity of *GLDPA-Ft-2* (20.41 nmol MU * mg⁻¹ protein * min⁻¹; see chapter 1: manuscript 1 Figure 5) indicates that the occurrence of splicing is most likely not the only mechanism regulating the expression strength of the distal promoter in the context of the complete tandem promoter construction of *GLDPA* of *F. trinervia*.

The analyses of the full-length 5' flanking region with the additional splice acceptor site indicate that the insertion of the splice acceptor site, as it is found in the original context of the *GLDPA* gene, does not lead to an alteration in the spatial expression pattern. Transcripts from the distal promoter do not likely accumulate in high amounts; even so they would

probably be functional. This is consistent with the fact that a RNA *in situ* hybridization experiment, with a probe targeted against *GLDPA* mRNA, did not show any signal in the mesophyll cells of the leaves of *F. trinervia* (Engelmann et al., 2008). Thus, the amount of transcripts from the distal promoter is probably quite low. This is further supported by the fact that in RNAseq experiments, either with 454 or Illumina sequencing, only a minor percentage of transcripts originated from the TSS in the distal part of the tandem promoter (chapter 1: manuscript 1 and chapter 2: manuscript 2 Figure 8A).

The tenfold lower GUS activity (Figure 2C) indicates that the insertion of the splice acceptor site leads to a lower activity of the complete tandem-promoter. One explanation for this is that the possibility of splicing of transcripts, derived from the distal promoter, stabilizes mRNAs from TSS_{R2}, and thus, more transcripts can be produced. This might lead to an occupation of the TSS of the proximal promoter by the transcription machinery from TSS_{R2}. This might interfere with the transcription efficiency of TSS_{R7}, due to the preclusion of the occupation of the binding site for the RNA polymerase II and its associated factors. This mechanism is known and called “transcriptional interference” (Shearwin et al. 2005; Mazo et al. 2007; Palmer et al. 2011).

As discussed in chapter 1: manuscript 1, the existence of the proximal promoter interferes with the transcriptional strength of the distal promoter. This is shown by the fact that the activity of the distal promoter, when uncoupled from the tandem promoter context, drives high *GUS* reporter gene activity, whereas in the tandem composition, only a minor amount of transcripts from the distal promoter is found (cf. chapter 1: manuscript 1). This leads to the assumption that the existence of the two promoters causes interference. The distal promoter, when stabilized through splicing, inhibits the proximal one (Figure 2), while the existence of the proximal promoter leads to a suppression of the distal one (see chapter 1: manuscript 1).

uORFs play a role in transcript stability of the distal promoter

The elimination of the longest, and thus probably most interfering, uORF from sub-region 7_4 changes the spatial expression pattern of the construct *GLDPA*-Ft1-2-3-7_4-7UTR from bundle sheath preferentially to ubiquitously. This indicates that the existence of the uORF indeed plays a role in the suppression of transcripts derived from the distal TSS. “Nonsense-mediated mRNA decay” (NMD) occurs, if uORFs have a length of at least 50 nt (Kertesz et al., 2006; Brogna and Wen, 2009; Nyiko et al., 2009). In the case of the *GLDPA* gene the uORF is 246 nt long (see chapter 1: manuscript 1) and in the GUS constructs the uORF has a length of 141 nt. So in both situations an interference with transcript stability is likely.

To examine this influence directly the transformation of either *GLDPA*-Ft1-2-3-7_4-7UTR or *GLDPA*-Ft1-2-3-4-5-6 (Figure 1) into an *A. thaliana* plant that is deficient in one of the NMD factors, called UPF (up-frameshift protein), such as *upf 1-5*, that are no longer able to performed the degradation of transcripts via NMD (Kalyna et al., 2012) is a suitable way to analyse the influence of NMD. Transcripts from the distal TSS should be stabilized and detectable with, for instance, 5' RACE in a higher amount than in plants with the wild type *A. thaliana* background.

Material and Methods

Standard molecular procedures such as polymerase chain reaction (PCR), reverse transcription, enzymatic restriction of DNA and gel electrophoresis were carried out after Sambrook et al. (Sambrook and Russell 2001).

Generation of promoter-reporter gene constructs

Transformation constructs for *Arabidopsis thaliana* plants were prepared in the pBI121 vector (Jefferson et al. 1987; Chen et al. 2003) Promoter DNA sequences were either amplified with PCR and the correctness was verified with commercial sequencing (LGC Genomics (formerly Agowa), Berlin, GER or GATC Biotech, Konstanz, GER) or synthesized as complete fragments by GenScript (Piscataway, USA). Promoter constructs in the vector pBI121 were inserted using restriction sites available in the multiple cloning site (MCS) of the vector. For constructs containing the splice acceptor site and the second ATG of the *GLDPA* gene of *F. trinervia* in front of the reading frame of the GUS reporter gene a PCR fragment was inserted between the *XmaI* (5') of the MCS and a *SnaBI* (3') site in the GUS open reading frame. The 5' oligonucleotide contained in addition to the restriction site and 16 bp of the GUS coding sequence (including the ATG) the ATG and the sequence up to the second ATG of the open reading frame of *GLDPA* of *F. trinervia*. This PCR fragment was inserted in a *XmaI/SnaBI* digested pBI121 vector. Using the *HindIII* and *XmaI* restriction site of the MCS the resulting vector was also used to create the construct *GLDPA*-Ft-Splice.

Transformation of *Arabidopsis thaliana*

Transformation of *Arabidopsis thaliana* plants was performed as described in manuscript 2.

***In situ* analysis of the β -glucuronidase and its fluorimetric measurement**

The *in situ* analysis of transgenic plants carrying β -glucuronidase constructs and fluorimetric measurements of the activity were performed as described in manuscript 2.

Table 1: Oligonucleotides used for the construction of reporter gene constructs. Underlined italic fonts indicate restriction sites. *HindIII* (aagcct), *SnaBI* (tacgta) *XmaI* (cccggg).

Oligonucleotide	Sequence 5' →3'
FtSpliceATG5'fw_XmaI	<u>cccggg</u> atggagcgtgcacgcaggctagctatgttacgtcctgtag
FtSplice-3rv_GUS_SnaBI	<u>tacgtac</u> acttttcccggcaat
FtGLDPA-Reg6-5'rv_XmaI	<u>cccggg</u> tttgaacatcgtatcgtgc
FtGLDPA5'_HindIII	<u>aagcct</u> tactcctctcaactttcaaatc
FtGLDPA3'rv_XmaI	<u>cccggg</u> gagtgaagatgggggtgtaa

Literature

- Brogna, S., and Wen, J.** (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat Struct Mol Biol* **16**, 107-113.
- 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 Breeding* **11**, 287-293.
- 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.
- 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.
- Kalyna, M., Simpson, C.G., Syed, N.H., Lewandowska, D., Marquez, Y., Kusenda, B., Marshall, J., Fuller, J., Cardle, L., McNicol, J., Dinh, H.Q., Barta, A., and Brown, J.W.** (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*. *Nucleic Acids Res* **40**, 2454-2469.
- 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.

- Nyiko, T., Sonkoly, B., Merai, Z., Benkovics, A.H., and Silhavy, D.** (2009). Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. *Plant Mol Biol* **71**, 367-378.
- 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.
- Sambrook, J., and Russell, D.** (2001). *Molecular cloning: A laboratory manual*. (Cold Spring Harbor Laboratory Press).
- Shearwin, K.E., Callen, B.P., and Egan, J.B.** (2005). Transcriptional interference--a crash course. *Trends Genet* **21**, 339-345.

Danksagungen

Ich danke...

... **Prof. Dr. Peter Westhoff** dafür, dass ich die Möglichkeit von ihm bekommen habe in seinem Institut meine Doktorarbeit machen zu dürfen. Außerdem natürlich für die hilfreiche Betreuung während dieser Zeit und die vielen Möglichkeiten an interessanten Orten an interessanten Tagungen teilzunehmen.

... **Prof. Dr. Michael Feldbrügge** für die Übernahme des Koreferats.

...**Maria Koczor** und **Monika Streubel** dafür, dass sie die mühselige Transformation der Flaverien für mich durchgeführt haben und sich auch danach noch um diese gekümmert haben. Außerdem Monika für den gelungenen gemeinsamen Forschungsaufenthalt in Oxford und Maria für das Gefühl auch in der Uni ein bisschen bemuttert zu werden... meine Hosen darfst du trotzdem nicht flicken.

...**Udo Gowik** für das Betreuen der C₄ Gruppe, die große Hilfe beim Schreiben meines ersten Papers und den Aufbau einer Toleranz gegen viel zu starken Kaffee.

...dem **iGRAD Plant** Graduiertenprogramm.

...den **Gärtnern** des Dachgewächshauses der HHU für die Betreuung meiner Pflanzen.

...**Myles Levey** und **Julia Mallmann** für das beste Promics-Team, das es jemals gegeben hat und geben wird.

... **Julia Mallmann, Christian Wiludda und Christian Wever** für die gemeinsam verbrachten Stunden in den ersten Jahren meiner Doktorarbeitszeit.

...**Jan Emmerling** für die vielen wirklich hilfreichen Diskussion über die Komplexitäten der GDC in *Flaveria*, egal ob P oder T ;-). Natürlich auch größten Dank für alle Gespräche, die nicht fachlich relevant waren. Vielen Dank auch für das Korrekturlesen, war bestimmt nicht lustig. Und ein übergroßes Danke für alles, was nicht hier steht.

...**Daniela Lichtblau** für die zwar nur kurze aber nichts desto trotz gute gemeinsame Zeit, die wir mit mitnichten tiefsinnigen Gesprächen verbracht haben. Da mache ich gern den Scan-Hiwi ;)

...**Sandra Kirschner** für unsere Gespräche, die Hollandkurztrips, die leckere Pizza und für die künstlerische Vielfalt in unserem Labor.

...dem restlichen **Botanik IV** Institut für die angenehme Arbeitsumgebung und die Hilfsbereitschaft, die eigentlich immer von Allen für Alle zu finden war.

...**Jens Fleischhauer** für den Blick eines Sprachwissenschaftlers auf meine oft viel zu langen und verschlungenen englischen Sätze und das daraus resultierende Korrekturlesen.

... **Ryan Dux** for the insights of a native speaker into the depths of the english language. Thanks for reading this stuff, without even knowing me in person beforehand.

...**meiner Familie** dafür, dass sie immer an mich geglaubt hat und mich immer unterstützt hat. Besonders meinem großen Bruder ;-) Danny, der mich in der Zeit des Schreibens nicht hat vereinsamen lassen und durch gemeinsame Kochabende dafür gesorgt hat, dass ich zwischendurch mal was Leckeres zu Essen hatte.