Identification and characterization of a frequent genetic alteration toward the evolution of C2-photosynthesis in the genus *Moricandia*

Von der Naturwissenschaftlichen Fakultät
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades

Doktor der Naturwissenschaften
Dr. rer. nat.

genehmigte Dissertation
von

Waly Essam Farid Adwy, M.Sc.

[2018]

Prüfungsvorsitz:
Prof. Dr. Edgar Maiß

Prüferkollegium:
Prof. Dr. Jutta Papenbrock
Dr. Sascha Offermann
Prof. Dr. Edgar Maiß

Tag der Promotion: 29.03.2018

Erklärung kumulative Dissertation:

aus:

Gemeinsame Ordnung für die Promotion zum Doktor der Naturwissenschaften (Dr. rer. nat.) an der Gottfried Wilhelm Leibniz Universität Hannover (25.03.2013), § 8 Dissertation, (3):

A:

„[…][2] Es ist eine ausführliche Darstellung voranzustellen, die eine kritische Einordnung der Forschungsthemen und wichtigsten Erkenntnisse aus den Publikationen in den Kontext der wissenschaftlichen Literatur zum Thema vornimmt […]“

Die voranzustellende ausführliche Darstellung ist in dieser Arbeit aufgeteilt in die Kapitel 1 und 5.

B:

„[…] sowie die individuellen eigenen Beiträge und ggf. die Beiträge weiterer Autoren an den jeweiligen Publikationen darlegt.“

Publication (Chapter II)

Adwy W, Laxa M, Peterhansel C (2015) A simple mechanism for the establishment of C2-specific gene expression in *Brassicaceae*. Plant J 84: 1231–1238.

W.A. designed and conducted the experiments, analyzed data and wrote the manuscript; M.L. partially designed the experiments, provided the binary vector and chromatin material used in this work; C.P. designed, supervised this work and wrote the manuscript; all authors have reviewed this manuscript.

Manuscript I (Chapter III)

Adwy W, Schlüter U, Papenbrock J, Peterhänsel C, Offermann S Loss of the M-box from the glycine decarboxylase P-subunit promoter in C2 *Moricandia* species. Journal of Plant Research, submitted in April 2018.

The overall idea for the experiments was developed by W.A., and C.P; the more detailed design of the experiment was done by W.A.; U.S. provided plant material and transcript sequence data; the experiments were conducted by W.A. as well as the data analysis; the manuscript was written by W.A., J.P., U.S. and S.O., and reviewed by all authors.

Manuscript II (Chapter IV)

Adwy W, Peterhänsel C, Papenbrock J The M-box binding factor: a possible role for GATA transcription factors in regulating the mesophyll expression of the *AtGldp*1 promoter, in preparation

The overall idea for the experiments was developed by W.A. and C.P.; the more detailed design of the experiment was done by W.A.; the experiments were conducted by W.A. as well as the data analysis; the manuscript was written by W.A. and J.P., and reviewed by all authors.

Summary

During the process of photosynthesis plants fix atmospheric CO_2 into organic carbohydrates. The photosynthetic assimilation rates resulting from this reaction differ among different photosynthetic variants, in which C3 photosynthesis represents the most simple and ancestral photosynthetic variant, whereas C4 is the more advanced variant and more correlated with increased biomass. C4 photosynthesis evolved more than 60 times independently via intermediate C3-C4 variants that are referred to as C2 photosynthesis and are considered as a bridge between C3 and C4 photosynthesis. Understanding of the molecular basis of the evolution of C4 holds the potential of improving crop productivity. This thesis is mainly concerned with one of the earliest steps of development of the C4-trait, the restriction of glycine decarboxylase (GDC) complex to the bundle sheath (BS) of C2 plants. By using promoter deletion experiments this work shows that restriction of the GDC to the BS of C3 plants is possible. The *cis*-element responsible for this spatial expression was isolated from the promoter of *AtGldp*1, one of the two genes encoding the GDC-P subunit in *Arabidopsis thaliana*. The identified element is referred to as the M-box. Using bioinformatic analysis this work generalizes the presence of the M-box in the promoters of most C3 *Gldp*1 genes from *Brassicaceae* and suggests its loss from promoters of most C2 *Gldp* genes. Subsequently, this offers a possible molecular mechanism for restriction of GDC to BS in at least in genus *Moricandia*. The bioinformatic analysis was further validated experimentally by isolating additional C3 and C2 promoters from other species of the *Moricandia* that were not characterized or isolated before by restriction genome walking. We show that the M-box is found in the promoter of a C3 *Moricandia* species and absent in three *Gldp* promoters for three different *Moricandia* C2 species. The isolated promoters were further studied in *Arabidopsis*. Furthermore the M-box binding factor was defined to be a member of GATA transcription factor family via a yeast-one-hybrid screening coupled with several co-expression analyses and the processing of previous genome-wide footprint sequencing data. Collective results from this study identify and characterize a simple mechanism for establishing C2-specific gene expression in a C3 plant like *Arabidopsis*.

Keywords: *Arabidopsis thaliana*, *Brassicaceae*, *Moricandia*, bundle sheath, GATA transcription factor, M-box, *Gldp*1, photosynthesis.

Zusammenfassung

In der Photosynthese fixieren die Pflanzen anorganisches CO_2 aus der gasförmigen Form in der Atmosphäre und synthetisieren organische Kohlenhydrate. Die photosynthetischen Assimilationsraten, die sich aus dieser Reaktion ergeben, unterscheiden sich zwischen verschiedenen photosynthesevarianten, wobei C3 die einfachste und die ursprüngliche photosynthetische Variante ist und C4 eine abgeleitete Variante ist und mehr mit erhöhter Biomasse korreliert. Die C4-Photosynthese entwickelte sich mehr als 60 Mal unabhängig voneinander. Man nimmt an, dass intermediäre C3-C4-Varianten, auch als C2-Photosynthese bezeichnet, eine Brücke zwischen C3- und C4-Photosynthese bilden. Ein tieferes Verständnis der molekularen Grundlagen der Evolution von C4 kann genutzt werden, um die Pflanzenproduktion zu verbessern. Die vorliegende Arbeit beschäftigt sich hauptsächlich mit einem der ersten Entwicklungsschritte dieses Merkmals, der Beschränkung der Biosynthese des Glycin-Decarboxylase (GDC)-Komplexes auf die Bündelscheidenzellen (BS) von C2-Pflanzen. Durch die Verwendung von Promotordeletions-Mutanten konnte gezeigt werden, dass die Beschränkung der GDC auf die BS von C3-Pflanzen möglich ist. Ein verantwortliches *cis*-Element im Promotor des *AtGldp*1-Gens für diese räumliche Expression in BS konnte in *Arabidopsis thaliana* identifiziert werden; die isolierte Box wird als M-Box bezeichnet. Bioinformatische Analysen zeigen das Vorhandensein der M-Box in den Promotoren der meisten C3 *Gldp*1-Gene in *Brassicaceae* und lassen einen Verlust der M-Box in Promotoren der meisten C2 *Gldp*-Gene vermuten. Die Ergebnisse deuten auf einen möglichen molekularen Mechanismus für die Restriktion von GDC auf BS zumindest in der Gattung *Moricandia* hin. Die bioinformatische Analyse wurde experimentell validiert, indem zusätzliche C3- und C2-Promotoren aus anderen *Moricandia*-Arten unter Verwendung von "restriction genome walking" isoliert wurden, die zuvor nicht charakterisiert oder isoliert worden waren. Wir zeigen, dass die M-box in drei *Gldp*-Promotoren aus drei verschiedenen *Moricandia* C2-Spezies vorkommt. Andererseits, ist die M-box in dem Promoter einer C3-Spezies gefunden worden. Darüber hinaus wurde die Regulation der M-Box untersucht und eine Regulierung durch GATA-Transkriptionsfaktoren postuliert. Weiterhin wurde der M-Box bindende Faktor als ein Mitglied der GATA-Transkriptionsfaktorfamilie auf Basis eines Yeast-One-Hybrid-Screens verbunden mit mehreren Co-Expressionsanalysen und der Verarbeitung von vorhandenen " genome-wide footprint"-Sequenzierungsdaten identifiziert.

Zusammengefasst identifizieren und charakterisieren die erzielten Ergebnisse dieser Arbeit einen einfachen Mechanismus zur Etablierung einer C2-spezifischen Genexpression in einer C3-Pflanze wie *Arabidopsis*.

Schlüsselwörter: *Arabidopsis thaliana*, *Brassicaceae*, *Moricandia*, Bündelscheidenzellen, GATA Transkriptionsfaktor, M-Box, *Gldp*1, Photosynthese.

Abbreviations

AlyGldp	Gene encoding the GDC-P from *Arabidopsis lyrata*
ASF-2	Activating-sequence factor-2
AtGdch	Gene encoding the GDC-H of *A. thaliana*
AtGdct	Gene encoding the GDC-T of *A. thaliana*
AtGldp	Gene encoding the GDC-P of *A. thaliana*
AtLPD	Gene encoding the GDC-D of *A. thaliana*
*Att*R	Attachment region
Bla	Beta-lactamase, resistance against β-lactam antibiotics
BnGldp	Gene encoding the GDC-P from *Brassica napus*
BraGldp	Gene encoding the GDC-P from *Brassica rapa*
BS	Bundle sheath
BstGldp	Gene encoding the GDC-P from *Boechera stricta*
Cab/Lhc	Gene encoding the light-harvesting chlorophyll a/b protein
CBB	Calvin-Benson-Bassham
CCMs	Carbon concentration mechanisms
CcdB	Cytotoxic protein encoding gene B
CclGldp	Gene encoding the GDC-P from *Citrus clementine*
CDS	Coding sequence
CgrGldp	Gene encoding the GDC-P from *Capsella grandiflora*
CrbGldp	Gene encoding the GDC-P from *Capsella rubella*
CsiGldp	Gene encoding the GDC-P from *Citrus sinensis*
cTP	Chloroplast transient peptide
EARS	Evolutionary analysis of regulatory sequences

EsaGldp	Gene encoding the GDC-P from *Eutrema salsugineum*
***FtGldp*A**	Gene encoding the GDC-P group-A from *Flaveria trinervia*
GAPDH	Gene encoding the glycerinaldehyde-3-phosphate-dehydrogenase
GDC	Glycine decarboxylase
GDC-H	The H-subunit of the glycine decarboxylase
GDC-L	The L-subunit of the glycine decarboxylase
GDC-P	The P-subunit of the glycine decarboxylase
GDC-T	The T-subunit of the glycine decarboxylase
GEO	Gene expression omnibus
GLK	Gene encoding the *GOLDEN2-LIKE* transcription factor
*Gldp*1-P1	-113 bp of the promoter of the gene encoding the *Gldp*1 from *A. thaliana*
*Gldp*1-P2	-379 bp of the promoter of the gene encoding the *Gldp*1 from *A. thaliana*
*Gldp*1-P3	-576 bp of the promoter of the gene encoding the *Gldp*1 from *A. thaliana*
*Gldp*1-P4	-1270 bp of the promoter of the gene encoding the *Gldp*1 from *A. thaliana*
Gly	Glycine
GUS	Gene encoding β-glucuronidase
HIS	Histidin
H_{met}	H-subunit is methylaminated
H_{red}	H-subunit is reduced
LB	Left border of transfer DNA
LEU	Leucin
Loc	Location
M	Mesophyll

M::P2	M-box fused to the *Gldp*1-P2 promoter
MaGldp	Gene encoding the GDC-P from *Moricandia arvensis*
M-box	An upstream promoter element in the *Gldp*1 promoter important for mesophyll expression
MmGldp	Gene encoding the GDC-P from *Moricandia moricandioides*
MnGldp	Gene encoding the GDC-P from *Moricandia nitens*
mRNA	Messenger RNA
MS	Murashige & Skoog
MsGldp	Gene encoding the GDC-P from *Moricandia suffruticosa*
mTP	Mitochondrial transient peptide
MU35Smin	M-box fused (united) to the minimal 35S promoter (same as M::35Smin)
MUP2	M-box fused (united) to the P2 promoter of the *Gldp*1 (same as M::P2)
NAD^+	Nicotin amide dinucleotide (oxidized)
NADH	Nicotin amide dinucleotide (reduced)
NF-Y	Nuclear factor-Y
*npt*II	Gene encoding the neomycin phosphotransferase (kanamycin resistance gene)
Ori	Origin of replication (RK2 – for *A. tumefaciens* , ColE for *E. coli*)
P3Δ1	-517 bp of the promoter of the gene encoding the *Gldp*1 from *A. thaliana*
P3Δ2	-481 bp of the promoter of the gene encoding the *Gldp*1 from *A. thaliana*
P3Δ3	-421 bp of the promoter of the gene encoding the *Gldp*1 from *A. thaliana*
P3Δ4	-413 bp of the promoter of the gene encoding the *Gldp*1 from *A. thaliana*
pA35S	Termination poly-adenylation signal of the 35S from cauliflower mosaic virus

pAnos	Termination poly-adenylation signal of the nopaline synthase gene
PCR	Polymerase chain reaction
PEPC	Phospho*enol*pyruvate carboxylase
Pnos	Promoter of nopaline synthase
qRT-PCR	Quantitative real-time PCR
RB	Right border of transfer DNA
*rbc*S	Gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase
RC	Reliability class
RTL	Relative transcription level
RubisCO	Ribulose 1,5-bisphosphate carboxylase/oxygenase
SAR	Scaffold attachment region
SEM	Standard error of mean
SHMT	Serine hydroxymethyltransferase
T-DNA	Transfer DNA
TDO	Triple drop out medium
THF	Tetrahydrofolate
TRP	Tryptophan
V-box	An upstream promoter element in the *Gldp*1 promoter important for vascular expression
WT	Wild-type
Y1H	Yeast-one-hybrid
2-P1	-146 bp of the promoter of the gene encoding the *Gldp*2 from *A. thaliana*
2-P2	-470 bp of the promoter of the gene encoding the *Gldp*2 from *A. thaliana*
2-P3	-647 bp of the promoter of the gene encoding the *Gldp*2 from *A. thaliana*

2-P4	-1812 bp of the promoter of the gene encoding the *Gldp*2 from *A. thaliana*
2-PG	2-Phosphoglycolate
35S	Promoter of the 35S *cauliform mosaic* virus
35Smin	Minimal promoter of the 35S cauliflower mosaic virus
3-PGA	3-Phosphoglycerate
3AT	3-Aminotriazole

INDEX

Contents

Chapter I: General introduction

1.1 Assimilation of carbon dioxide in higher plants

Atmospheric CO_2 fixation in higher plants depends on the ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO). RubisCO catalyzes the fixation of CO_2 to ribulose-1,5-bisphosphate (RuBP) resulting in the production of the three carbon (C3) organic acid 3-phosphoglycerate (3-PGA). The efficiency of this reaction is decreased when oxygen reaches the active site of RubisCO (Zelitch, 1992) (Fig. 1A). RubisCO can catalyze the oxygenation of RuBP leading to the generation of one molecule 3-PGA and one molecule 2-phosphoglycolate (2-PG) (Bassham and Kirk, 1962). The generated 2-PG is toxic to the plant as it inhibits the enzymes of the Calvin-Benson-Bassham (CBB) cycle enzymes (Anderson, 1971). Some of the carbon atoms combined in 2-PG can be recycled back into 3-PGA via the photorespiratory pathway (Leegood et al., 1995) under consumption of additional energy (Bauwe et al., 2010; Raines, 2011). Hence, C3 plants lose part of their already captured carbon through oxygenation of RuBP. In a C3 leaf, not all cells contribute equally to the overall photosynthetic performance of the plant, for example the total number of bundle sheath cells (BS) account for 15% of the total number of cells in an *A. thaliana* first leaf, with an average of 22 chloroplasts per BS cell relative to 76 chloroplasts per mesophyll (M) cell, and hence BS have relatively less contribution to photosynthesis (Kinsman and Pyke, 1998).

C3 photosynthesis was probably the dominant CO_2 fixation mechanism in the past 4 billion years when atmospheric CO_2 levels were much higher than today's CO_2 level and oxygen levels were lower, known that RubisCO first appeared 3.4 billion years ago (Berner and Kothavala, 2001; Sage, 2004). Later during the carboniferous period (~ 280 million years ago) oxygen level was greater than the CO_2 level favoring high levels of photorespiration (Sage, 2004). These environmental changes might have acted as selection pressure for integrating other enzymes to a metabolism that concentrates CO_2 around RubisCO, leading to

the evolution of other photosynthetic variants such as C4 photosynthesis via an intermediate C3-C4 phenotype that probably acted as a bridge for the evolution of the C4 pathway (Sage, 2005). In C3-C4 intermediate plants, the glycine decarboxylase (GDC), a photorespiratory enzyme (for more details see chapter 1.3), is restricted to the BS of the leaf (Fig. 1B). For example, C3-C4 intermediate plant species like *Moricandia arvensis* lack the expression of the P-subunit of the GDC in their M compartment, while the other subunits of the GDC are still expressed. The lack of the P-subunit expression from the M prevents the assembly of a functional GDC in the M and effectively restricts glycine decarboxylation to the BS compartment only (Morgan et al., 1993). The restriction of GDC to the BS of intermediate plants resulted in a metabolic advantage relative to C3 plants. Glycine (Gly) accumulates in the mesophyll compartment of C2 leaves where the GDC is not functional, and is catabolized in the BS compartment where the GDC is expressed and active. Such pathway is beneficial to plants relative to C3 as it increases the CO_2 concentration in the BS. Therefore, CO_2 has a higher probability of being re-fixed by RubisCO expressed in the BS of C3-C4 intermediates (Keerberg et al., 2014). In C3-C4 intermediate plants, CO_2 is locked in the form of the C2 amino acid glycine; therefore this form of photosynthesis is also referred to as C2 photosynthesis and these plants are also called C2 plants. This metabolic pathway is suggested to be a necessary step toward C4 evolution (Monson and Rawthorne, 2000; Sage, 2005). There are also studies showing that C2 leaves are distinguished by differences in anatomy relative to a C3 leaf. These differences include more chloroplasts and mitochondria in BS cells, enlarged BS cells and narrower vein spacing (Rawsthorne, 1992). Such anatomy might probably not only contribute to increasing the CO_2 concentrations at the BS of C2 plants via higher chloroplast density but also provides a mechanistic support to hold the released CO_2 at the BS of C2 plants and reducing its efflux (Sage et al., 2014).

In C4 plants, other pre-existing metabolic enzymes were further distributed on different leaf compartments. CO_2 fixation takes place in the M cells by phospho*enol*pyruvate carboxylase

(PEPC) resulting in the synthesis of a C4 acid that is further processed in the BS cells (Fig. 1C) (Hatch, 1987). The C4 acid is specifically decarboxylated in the BS of C4 species providing a working biochemical CO_2 pump around RubisCO. Such spatial distribution of the pre-existing metabolic enzymes, PEPC in the M cells and RubisCO in the BS cells, re-order the photosynthetic tasks on different leaf tissues resulting in higher photosynthetic assimilation rates relative to C3 and C3-C4 intermediate plants. In addition, anatomical adaptations of C3-C4 intermediate plants become more pronounced in a C4 leaf that is known as "Kranz" anatomy (Sage et al., 2014). In such anatomy, frequency and size of BS cells is different compared to C3 and C2 plants. Furthermore, mesophyll (M) and bundle sheath (BS) cells are in 1:1 contact and in a wreath-like structure (Sage et al., 2014). Such C4-anatomical adaptations were previously linked to drought conditions, because C4 anatomy has evolved more than 60 independent times mostly in plants that occur in arid conditions (Sage, 2004).

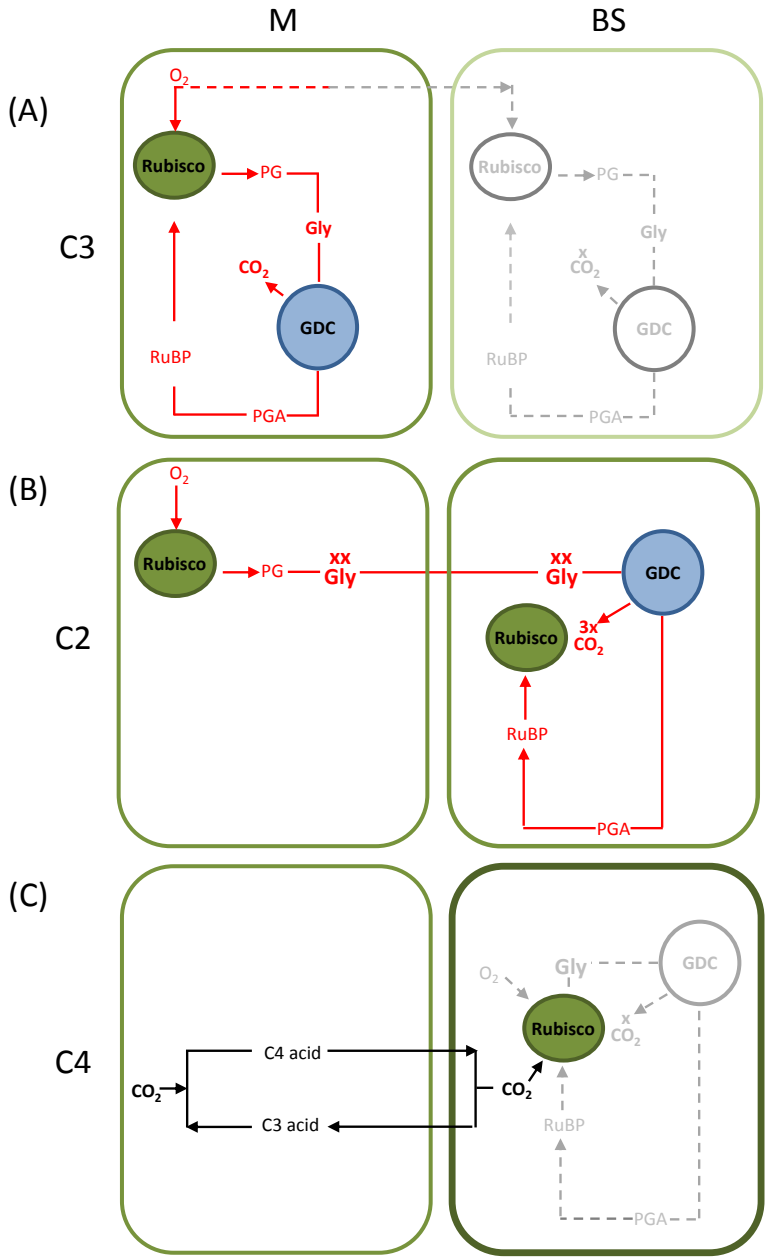


Figure 1. Schematic representation of photorespiration in C3, C2 (intermediate C3-C4), and C4 plants..

(A) In C3 plants O_2-fixation by RubisCO in mesophyll (M) cells results in the formation of 2-phosphoglycolate (PG). PG is converted back to 3-phosphoglycerate (PGA) by photorespiration and then further converted to ribulose-1,5-bisphosphate (RuBP) by the Calvin cycle. Photorespiration includes CO_2 release catalyzed by glycine decarboxylase (GDC) in mitochondria, which carries out decarboxylation of glycine (Gly), an intermediate product during recycling of PG. Photosynthesis and photorespiration also occur at the BS of C3 plants but at lower rates compared to M cells. Although the CO_2 released in the BS is probably trapped for longer time in the BS (xCO_2) by the above layer of M cells, it is probably less likely of being re-fixed.

(B) In C2 plants the expression of the GDC is restricted to the BS cells, this results in accumulation of Gly in the M compartment (xxGly) where the GDC is not active. Gly is specifically metabolized in the BS compartment where the GDC is active. The specific metabolism of glycine in the BS compartment enriches the BS with CO_2 ($3xCO_2$) (Keerberg et al., 2014) and increases the probability of its refixation. xxGly = accumulated glycine.

(C) In C4 plants, RubisCO is additionally restricted to BS cells and provided with CO_2 by a biochemical pump. By this, rates of O_2 fixation are further reduced and photorespiratory CO_2 release is insignificant.

1.2 Evolution of C4 photosynthesis

The evolution of the C4 photosynthetic pathway represents an example of how a complex genetic trait could develop in a plant. Steps of C4-evolution have been described before to necessitate genomic and anatomical pre-conditioning steps starting from a C3 ancestral plant (Fig. 2; Sage, 2004). Genomic pre-conditioning might include duplication events for necessary genes, such as genes encoding the P-subunit of the GDC (Schulze et al., 2016).

Gene duplication is an important molecular mechanism to provide multiple copies of important genes, and these copies probably serve as later substrates during evolution (Lynch and Conery, 2000). The presence of multiple copies of a gene would allow maintaining the function of one gene copy during gradual manipulation of the other (Lynch and Conery, 2000; Sage, 2004). The mechanism by which genes are manipulated could occur by losing a pre-existing gene copy (non-functionalization), changing or adjusting the function of one copy (neo-functionalization), or adjusting both copies for obtaining capacity of one copy (sub-functionalization) (Lynch and Conery, 2000).

Anatomical pre-conditioning includes reduction of the number of mesophyll cells and hence increasing vein density, this has been observed in number of *Flaveria* species where gradually the C4-Kranz anatomy was established (McKown and Dengler, 2007). There are indications that C4-anatomy already existed in some C3 taxa millions of years before the emergence of the physiological C4 pathway, and this is suggested to facilitate the development of such C4-syndrome (Christin et al., 2013). Another observed modification is the activation of BS cells, this step is characterized by the increase in size and organelle number of BS cells as observed in C3-C4 intermediate relative to C3 species and probably modified expression of genes encoding the P-subunit of GDC to have higher expression in the BS cells (Rawsthorne, 1992; Schulze et al., 2016).

In C3-C4 intermediate species, the GDC is restricted to the BS compartment providing a backbone for the C2 pathway (see 1.1; Fig. 1B). The mechanism of GDC restriction to the BS cells was explained in genus *Flaveria*, where a globally expressed gene copy encoding the P-subunit of the GDC was lost, and the BS-expressed version of this gene was retained in C3-C4 intermediate and C4 *Flaveria* species (Schulze et al., 2013).

Another important step during C4 evolution was the enhanced expression of the C4-enzymes . This probably has occurred as a result of the uneven nitrogen distribution as a result of GDC restriction to the BS. The biochemical adaptations for C3-C4 intermediate metabolism that rebalance nitrogen between M and BS compartments after GDC relocation are predicted to be important for C4 evolution (Mallmann et al., 2014). The enhancement of the PEPC activity in the M cells of C4 plants resulted in stable fixation of CO_2 to C4 acids. The CO_2 locked in C4 acids is then released nearby RubisCO specifically in the BS cells resulting in suppression of the oxygenase activity of RubisCO in these cells (Sage and Monson, 1999) (Fig. 1C).

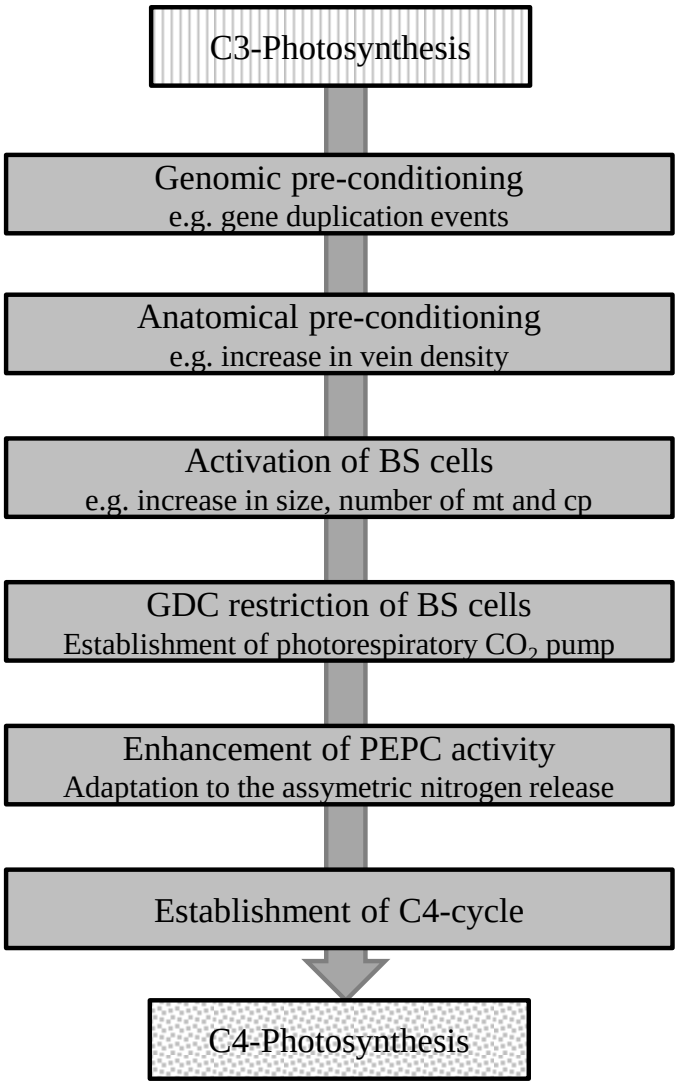


Figure 2. Sequence of events that are observed during evolution of C4-photosynthesis.

Model explaining the sequence of events that might have led to the evolution of the C4-syndrome. mt and cp: mitochondria and chloroplasts. BS: bundle sheath cells, PEPC: phospho*enol*pyruvate carboxylase, GDC: glycine decarboxylase. Figure adapted from Sage (2004).

1.3 The glycine decarboxylase complex

The glycine decarboxylase complex (GDC) is not only found in plants, but in all eukaryotes (Oliver, 1994). It is localized in the mitochondrial matrix (Oliver et al., 1990). In C3 plants, the GDC comprises about 30% of the mitochondrial matrix protein (Oliver et al., 1990). GDC has an essential function in C1-metabolism as it provides one-carbon compounds to several metabolic pathways such as nucleic acids synthesis, synthesis of lignin, betaines and alkaloids (Hanson et al., 2000; Oliver, 1994), and consequently, complete GDC knock-outs have been demonstrated to be lethal (Engel et al., 2007). In addition, it is a key enzyme in the photorespiratory metabolism (Bauwe, 2011).

GDC is a multi-enzyme complex; it comprises four protein subunits, the P-, H-, T-, and L- subunits. Components of the GDC catalyze together catabolic decarboxylation and deamination of the amino acid glycine. In collaboration with serine hydroxymethyltransferase (SHMT), it catalyzes the conversion of two molecules of glycine to one molecule of serine. CO_2 and NH_3 are released as biproducts of the decarboxylation reaction (Fig. 3; Douce et al., 2001). This reaction is important in processing the intermediate glycine that is produced during recycling of photorespiratory 2-PG that occurs as a biproduct of RubisCO's oxygenation reaction.

Catabolism of glycine starts by the reaction of the co-factor of the P-subunit, pyridoxal phosphate, with the amino moiety of glycine. As a result, glycine is decarboxylated (Fig. 3). The remaining residue is carried by the lipoamide cofactor of the H-subunit; hence, the H-subunit is methylaminated (Fig. 3, H_{met}). The deamination of glycine is further carried out by the aminotransferase activity of the T-subunit in a reducing tetrahydrofolate-dependent reaction. Subsequently, this leaves the lipoamide cofactor of the H-subunit in a reduced form (H_{red}). The H_{red} is recovered back after the oxidation of its disulfide terminus and in parallel NAD^+ is reduced to NADH. This is catalyzed by the dihydrolipoamide dehydrogenase of the L-subunit.

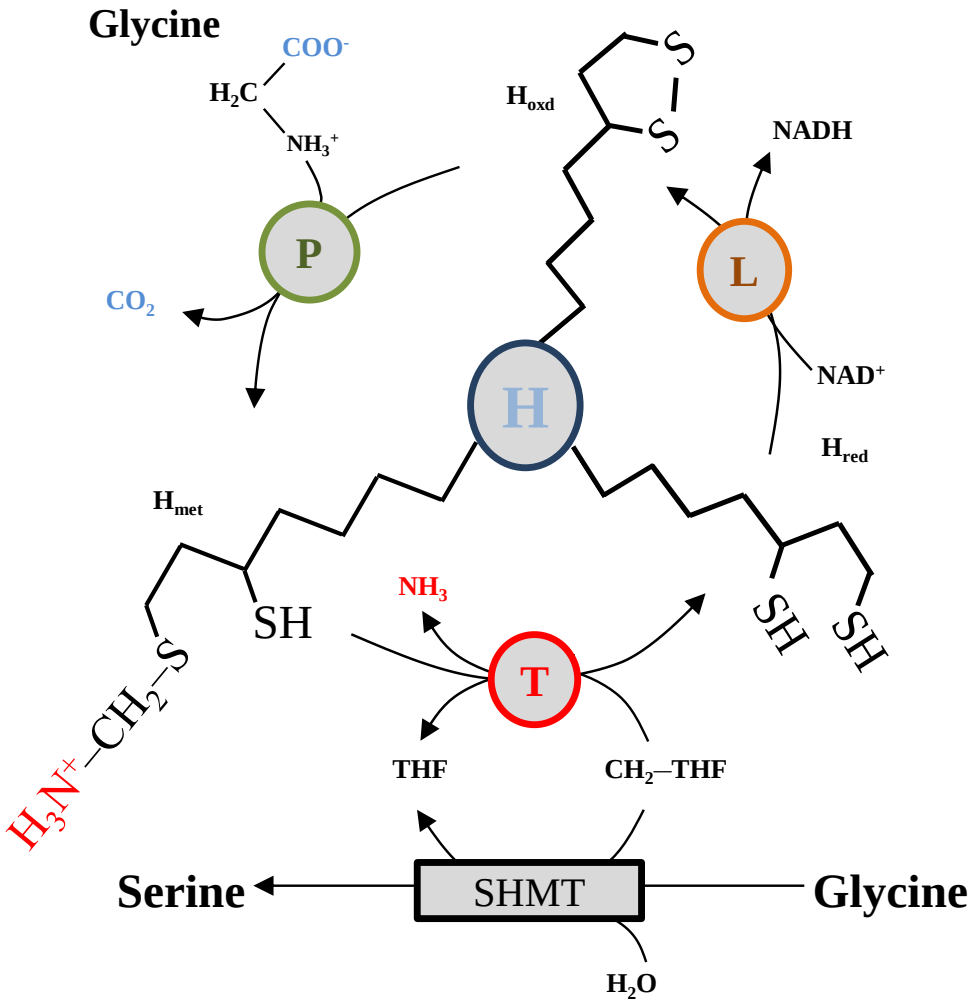


Figure 3. Glycine catabolism by the glycine decarboxylase complex.

The glycine decarboxylase (GDC) is composed of four protein subunits, the P-, H-, T-, and L- proteins. The P-subunit catalyzes decarboxylation of glycine. Methylamine and electron transfer are carried out by the H-, T-, and L- subunits. The biproducts of glycine decarboxylation are CO_2, NH_3 and a C1-tetrahydrofolate residue (CH2–THF). H_{met}, H_{red}, H_{ox} are methylaminated, reduced, and oxidized forms, respectively, of the H-protein. The amino acid serine is synthesized in collaboration with the serine hydroxymethyl transferase (SHMT).

1.3.1 Genes and regulation of the GDC complex

All subunits of the GDC are nuclear-encoded. The GDC-P subunit of *A. thaliana* is encoded by two genes (*AtGldp*1; AT4G33010 and *AtGldp*2; AT2G26080; Bauwe and Kolukisaoglu, 2003), the GDC-H subunit by three genes (*AtGdch*1; AT2G35370, *AtGdch*2; AT2G35120, and *AtGdch*3; AT1G32470; Srinivasan and Oliver, 1992), the GDC-T by one gene (*AtGdct*; AT1G11860; Bauwe and Kolukisaoglu, 2003), and the L-subunit by two genes (*AtLPD*1; AT3G17240 and *AtLPD*2; AT1G48030; Lutziger and Oliver, 2001). Protein products of all of these genes have mitochondrial targeting peptide signals (Bauwe and Kolukisaoglu, 2003).

Transcripts of genes encoding the GDC seem to accumulate in response to light (Vauclare et al., 1996; Douce et al., 2001) similar to the transcripts of *rbcS*. Light regulation was studied for some of the genes encoding GDC in more detail, for instance a promoter region for one of the H-subunit coding genes was found to positively regulate a fused reporter gene when plants are exposed to light (Srinivasan and Oliver, 1995). Other levels of regulations were also suggested for genes encoding the GDC complex, for example, post-transcriptional regulation was implied due to the lack of correlation between the increased amounts of transcripts of genes encoding the different subunits of the GDC and the amount of translated protein (Vauclare et al., 1998). Moreover, there is a negative correlation between the accumulation of serine and NADH and the translation of the different subunits of the GDC complex implying a post-translational regulation via metabolic feedback inhibition (Douce and Neuburger, 1999).

1.4 *Brassicaceae*

The *Brassicaceae* family belongs to the eudicotyledons which is a monophyletic clade of flowering plants or angiosperms. It is also referred to as cabbage, crucifers or mustards family. The *Brassicaceae* family has about 372 genera and 4060 species (www.theplantlist.org, updated 28.11.2017). Most species in this family perform C3 photosynthesis, except for eight species that are known to perform C2 photosynthesis (Sage et al., 2011; Sage et al., 2014). These *Brassicaceae* C2 species belong to either the genus *Moricandia* or *Diplotaxis*. The genus *Moricandia* includes five known C2 species; *M. nitens*, *M. suffruticosa*, *M. spinosa*, *M. arvensis* and *M. sinaica* (Sage et al., 2011). The genus *Diplotaxis* comprises of 35 accepted species (www.theplantlist.org), and have three reported case of C2 photosynthesis; *D. tenuifolia*, *D. erucoides*, and *D. muralis* (Sage et al., 2014).

A. thaliana is a model C3 plant belonging to the *Brassicaceae family*. There are no known C4 species in the *Brassicaceae family* (Sage et al., 2011). However, there is a close relationship between C4 *Cleome* species and *A. thaliana. A. thaliana* belongs to the *Brassicaceae* which *is a* sister group to *Cleomoideae* and both are merged into the *Capparaceae* (Brown et al., 2005). The tight relation between *Arabidopsis thaliana* and C4 *Cleome* species like *C. gynandropsis* makes the *Brassicaceae* family an interesting subject to study the evolution of C4 photosynthesis. This includes the availability of molecular phylogeny data, accessibility of different genome databases, an easy or applicable transformation system and the available range of mutants (Brown et al., 2005).

Genome sequencing data for *A. thaliana* are published in high accuracy and high quality. In addition, different microarray experiments are available for this model species, (Zimmermann et al., 2004). Moreover, a wide coverage of T-DNA insertion lines (Sessions et al., 2002) is available for this species. These advantages can accelerate future research objectives, such as installing a C4 pathway in *A. thaliana*. Developing C4 *A. thaliana* plant

would help in understanding the C4 pathway, and such information is of a high interest in the photosynthetic scientific community as it holds the potential of transferring the C4 trait to other C3 crops like wheat, potato and rice, therefore meeting the demand of the increasing world population (von Caemmerer et al., 2012; Hibberd and Furbank, 2016; von Caemmerer et al., 2017).

In this work the molecular mechanism of the restriction of GDC to the BS of C2 species within the *Brassicaceae* family was investigated. This was based on the hypothesis that a C2-spatial expression pattern might be controlled by pre-transcriptional regulation already existing in C3 plants. Objectives of this study were mainly tested in *A. thaliana* which is a model C3 species belonging to this family. Findings were further extended to different species within this family. Three main objectives were designed in order to test the aforementioned hypothesis. These are: (1) characterizing the upstream promoters of the genes encoding the GDC-P subunit, (2) analyzing the findings from the first objective in the context of phylogeny as well as carrying out bioinformatic analyses in order to answer the question; how this C2-expression pattern might have developed in the *Brassicaceae*?, (3) Implementing a molecular mechanism in C3 *Arabidopsis thaliana* in order to mimic the C2 pathway. The following experiments were carried out in order to fulfill the aforementioned objectives:

(1) The 5' flanking regions of genes encoding the GDC-P subunits were fused to a reporter *gusA* gene and transformed into *Arabidopsis thaliana*. Transcriptional and chromatin regulation were analyzed for the full promoters of each gene. Truncated versions for each promoter were tested in order to define promoter *cis* elements. Transgenic lines representing the different promoter versions were then analyzed for *gusA* expression patterns. Phylogenetic analysis was performed to characterize different GDC-P encoding genes in *Brassicaceae*. *Finally,* a bioinformatic analysis was performed to characterize the conserved motifs in promoters of *Gldp*1 genes from *Brassicaceae* (Publication 1; Adwy et. al., 2015).

(2) Based on the previous findings (Publication 1: Adwy et. al., 2015), a highly conserved promoter region between *Gldp*1 genes from *Brassicaceae* was studied in detail by deletion analysis. Therefore, truncated promoter versions were fused to the *gusA gene*, transformed into *A. thaliana* and the *gusA* expression pattern was studied. Promoters from C3 and C2 species from the genus *Moricandia* were isolated via restriction genome walking in order to check potential conservation of the identified *cis* elements between *Moricandia* species and *Arabidopsis thaliana*. In addition to bioinformatics analysis, the isolated promoters from *Moricandia* were then studied in *A. thaliana*. (Manuscript 1; Adwy et al., submitted).

(3) Based on findings from publication one and manuscript one, a yeast-one-hybrid screening coupled to experimental and bioinformatic co-expression analyses was used to identify binding sites and possible regulating factors for mesophyll specific expression of the *AtGldp*1 promoter (Manuscript 2; Adwy et al., in preparation).

1.5 Literature

Anderson L (1971) Chloroplast and cytoplasmic enzymes Pea leaf triose phosphate isomerases. Biochimica et Biophysica Acta **235:** 237–244.

Bassham J, Kirk M (1962) The effect of oxygen on the reduction of CO_2 to glycolic acid and other products during photosynthesis by chlorella. Biochemical and Biophysical Research Communications **9:** 376–380.

Bauwe H (2011) Photorespiration: The bridge to C4 photosynthesis. In Raghavendra AS, Sage RF, eds, C4 Photosynthesis and related CO_2 concentrating mechanisms. Springer Netherlands, Dordrecht, 81–108.

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

Bauwe H, Kolukisaoglu U (2003) Genetic manipulation of glycine decarboxylation. Journal of Experimental Botany **54:** 1523–1535.

Berner R, Kothavala Z (2001) GEOCARB III: a revised model of atmospheric CO_2 over phanerozonic time. American Journal of Science **301:** 182–204.

Brown NJ, Parsley K, Hibberd JM (2005) The future of C4 research--maize, *Flaveria* or *Cleome*? Trends in Plant Science **10:** 215–221.

Caemmerer S von, **Ghannoum O, Furbank RT** (2017) C4 photosynthesis: 50 years of discovery and innovation. Journal of Experimental Botany **68:** 97–102.

Caemmerer S von, **Quick WP, Furbank RT** (2012) The development of C4-rice: current progress and future challenges. Science **336:** 1671–1672.

Christin P-A, Osborne CP, Chatelet DS, Columbus JT, Besnard G, Hodkinson TR, Garrison LM, Vorontsova MS, Edwards EJ (2013) Anatomical enablers and the evolution of C4 photosynthesis in grasses. Proceedings of the National Academy of Sciences **110:** 1381–1386.

Douce R, Bourguingnon J, Neuburger M, Rebeille F (2001) The glycine decarboxylase system: a fascinating complex. Trends in Plant Science **6:** 167–176.

Douce R, Neuburger M (1999) Biochemical dissection of photorespiration. Current Opinion in Plant Biology **2:** 214-222.

Emanuelsson O, Nielsen H, Brunak S, Heijne G von (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of Molecular Biology **300:** 1005–1016.

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

Hanson AD, Gage DA, Shachar-Hill Y (2000) Plant one-carbon metabolism and its engineering. Trends in Plant Science **5:** 206–213.

Hatch M (2002) C4 photosynthesis: discovery and resolution. Photosynthesis Research **73:** 251–256.

Hibberd J, Furbank R (2016) Fifty years of C4 photosynthesis. Nature **538:** 177–179.

Keerberg O, Pärnik T, Ivanova H, Bassüner B, Bauwe H (2014) C2 photosynthesis generates about 3-fold elevated leaf CO_2 levels in the C3-C4 intermediate species *Flaveria pubescens*. Journal of Experimental Botany **65:** 3649–3656.

Kinsman EA, Pyke KA (1998) Bundle sheath cells and cell-specific plastid development in *Arabidopsis* leaves. Development **125:** 1815–1822.

Leegood R, Lea P, Adcock M, Häusler R (1995) The regulation and control of photorespiration. Journal of Experimental Botany **46:** 1397–1414.

Lutziger I, Oliver DJ (2001) Characterization of two cDNAs encoding mitochondrial lipoamide dehydrogenase from *Arabidopsis*. Plant Physiology **127:** 615–623.

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicated genes. Science **290:** 1151–1155.

Mallmann J, Heckmann D, Bräutigam A, Lercher JM, Weber PA, Westhoff P, Gowik U (2014) The role of photorespiration during the evolution of C4 photosynthesis in genus *Flaveria*. eLIFE **3:** e02478.

McKown AD, Dengler NG (2007) Key innovations in the evolution of kranz anatomy and C4 vein pattern in *Flaveria* (*Asteraceae*). American Journal of Botany **94:** 382–399.

Monson RK, Rawsthorne S (2000) CO_2 assimilation in C3-C4 intermediate plants. In RC Leegood, Sharkey TD, von Caemmerer S, eds, Photosynthesis: Physiology and metabolism. Springer Netherlands, Dordrecht, 533–550.

Morgan CL, Turner SR, 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.

Oliver D (1994) The glycine decarboxylase complex from plant mitochondria. Annual Review of Plant Physiology and Plant Molecular Biology **45:** 323–337.

Oliver DJ, Neuburger M, Bourguignon J, Douce R (1990) Interaction between the component enzymes of the glycine decarboxylase multienzyme complex. Plant Physiology 94: 833–839.

Raghavendra AS, Sage RF, eds, C4 Photosynthesis and related CO_2 concentrating mechanisms. Springer Netherlands, Dordrecht.

Raines CA (2011) Increasing photosynthetic carbon assimilation in C3 plants to improve crop yield: current and future strategies. Plant Physiology **155:** 36–42.

Rawsthorne S (1992) C3-C4 intermediate photosynthesis: linking physiology to gene expression. The Plant Cell **2:** 267–274.

Sage RF (2004) The evolution of C4 photosynthesis. New Phytologist **161:** 341–370.

Sage RF (2005) Atmospheric CO_2, environmental stress, and the evolution of C4 photosynthesis. In Baldwin IT, Caldwell MM, Heldmaier G, Jackson RB, Lange OL, Mooney HA, Schulze E-D, Sommer U, Ehleringer JR, Dearing M, Cerling TE, eds, a history of atmospheric CO_2 and its effects on plants, animals, and ecosystems. Springer, 185–213.

Sage RF, Khoshravesh R, Sage TL (2014) From proto-kranz to C4 kranz: building the bridge to C4 photosynthesis. Journal of Experimental Botany **65:** 3341–3356.

Sage R, Monson RK (1999) C4 plant biology. Academic Press, San Diego.

Schlüter U, Bräutigam A, Gowik U, Melzer M, Christin P-A, Kurz S, Mettler-Altmann T, Weber APM (2016) Photosynthesis in C3–C4 intermediate *Moricandia* species. Journal of Experimental Botany: **68:** 191-206.

Schulze S, Mallmann J, Burscheidt J, Koczor M, Streubel M, Bauwe H, Gowik U, Westhoff P (2013) Evolution of C4 photosynthesis in the genus *Flaveria*: establishment of a photorespiratory CO_2 pump. The Plant Cell **25:** 2522–2535.

Schulze S, Westhoff P, Gowik U (2016) Glycine decarboxylase in C3, C4 and C3–C4 intermediate species. Current Opinion in Plant Biology **31:** 29–35.

Sessions A, Burke E, Presting G, Aux G, McElver J, Patton D, Dietrich B, Ho P, Backwaden J, Ko C, Clarke J, Cotton D, Bullis D, Snell J, Miguel T, Hutchison D, Kimmerly B, Mitzel T, Katagiri F, Glazebrook J, Law M, Goff S (2002) A High-throughput *Arabidopsis* reverse genetics system. The Plant Cell **14:** 2985–2994.

Srinivasan R, Oliver DJ (1992) H-Protein of the glycine decarboxylase Multi-enzyme complex cDNA encoding the protein from *Arabidopsis thaliana*. Plant Physiology **98:** 1518–1519.

Srinivasan R, Oliver DJ (1995) Light-dependent and tissue-specific expression of the H-protein of the glycine decarboxylase complex. Plant Physiology **109:** 161–168.

Vauclare P, Diallo N, Bourguingnon J, Macherei D, Douce R (1996) Regulation of the expression of the glycine decarboxylase complex during pea leaf development. Plant Physiology **112:** 1523–1530.

Zelitch I (1992) Control of plant productivity by regulation of photorespiration. BioScience **42:** 510–516.

Zimmermann P, Hirsch-Hoffmann M, Henning L, Gruissem W (2004) GENEVESTIGATOR. Arabidopsis Microarray Database and Analysis Toolbox. Plant Physiology **136:** 2621–2632.

Chapter II: Publication 1

A simple mechanism for establishment of C2-specific gene expression in *Brassicaceae*

Waly Adwy[1,2], Miriam Laxa[1], Christoph Peterhänsel[1]

[1]Institute of Botany, Leibniz University Hannover, Herrenhäuser Str. 2, 30419 Hannover, Germany

[2]Department of Genetics, Faculty of Agriculture, Cairo University, Gamaa St. 13, 12619 Giza, Egypt

Author contribution:	W.A. designed and conducted the experiments, analyzed data and wrote the manuscript; M.L. partially designed the experiments, provided the binary vector and chromatin material used in this work; C.P. designed, supervised this work and wrote the manuscript; all authors have reviewed this manuscript.
Journal:	The Plant Journal
Date of Publication:	November 2015
DOI:	10.111/tpj.13084
PubMed-ID:	26603271

the plant journal

SEB Society for Experimental Biology

The Plant Journal (2015) 84, 1231–1238 doi: 10.1111/tpj.13084

A simple mechanism for the establishment of C_2-specific gene expression in Brassicaceae

Waly Adwy[1,2], Miriam Laxa[1] and Christoph Peterhansel[1,*]

[1]Institut für Botanik, Leibniz Universität Hannover, Herrenhäuserstrasse 2, 30419 Hannover, Germany, and
[2]Department of Genetics, Cairo University, 13 Gamaa Street, 12619 Giza, Egypt

Received 15 October 2015; accepted 9 November 2015; published online 25 November 2015.
*For correspondence (e-mail cp@botanik.uni-hannover.de).

SUMMARY

The transition of C3, via C2 towards C4 photosynthesis is an important example of stepwise evolution of a complex genetic trait. A common feature that was gradually emphasized during this trajectory is the evolution of a CO_2 concentration mechanism around Rubisco. In C2 plants, this mechanism is based on tissue-specific accumulation of glycine decarboxylase (GDC) in bundle sheath (BS) cells, relative to global expression in the cells of C3 leaves. This limits photorespiratory CO_2 release to BS cells. Because BS cells are surrounded by photosynthetically active mesophyll cells, this arrangement enhances the probability of re-fixation of CO_2. The restriction of GDC to BS cells was mainly achieved by confinement of its P-subunit (GLDP). Here, we provide a mechanism for the establishment of C2-type gene expression by studying the upstream sequences of C3 *Gldp* genes in *Arabidopsis thaliana*. Deletion of 59 bp in the upstream region of *AtGldp*1 restricted expression of a reporter gene to BS cells and the vasculature without affecting diurnal variation. This region was named the 'M box'. Similar results were obtained for the *AtGldp*2 gene. Fusion of the M box to endogenous or exogenous promoters supported mesophyll expression. Nucleosome densities at the M box were low, suggesting an open chromatin structure facilitating transcription factor binding. *In silico* analysis defined a possible consensus for the element that was conserved across the Brassicaceae, but not in *Moricandia nitens*, a C2 plant. Collective results provide evidence that a simple mutation is sufficient for establishment of C2-specific gene expression in a C3 plant.

Keywords: photorespiration, carbon concentration mechanism, evolution, promoter, nucleosome.

INTRODUCTION

C_3 plants lose part of their previously fixed carbon through oxygenation of ribulose-1,5-bisphosphate (RuBP) by Rubisco and subsequent photorespiration (Peterhansel et al., 2010). C_4 plants evolved a biochemical CO_2 pump to concentrate CO_2 around Rubisco by separation of primary and secondary carbon fixation in two tissues, mesophyll and bundle sheath (BS) cells (Langdale, 2011). Bundle sheath cells are not a novel invention of C_4 plants; they are also present in C_3 plants.

The function of BS cells in C_3 plants is still not fully understood, but the available data suggest that BS cells are involved in loading and unloading of assimilates to the vasculature or mechanical support (Kinsman and Pyke, 1998; Leegood, 2008). The chloroplasts of the BS cells might contribute to the re-fixation of CO_2 released from malate that is transported in the vasculature from the root to the leaf (Hibberd and Quick, 2002), but might also participate in nitrogen assimilation (Leegood, 2008). A recent characterization of ribosome-bound RNAs that were enriched in BS cells suggested additional functions in the assimilation and further metabolism of sulfur-containing compounds, the synthesis and export of branched-chain amino acids and glucosinolates or trehalose metabolism (Aubry et al., 2014). These functions are probably also important for mesophyll development (Lundquist et al., 2014).

The evolution of C_4 photosynthesis required multiple anatomical and physiological adaptations. An early step in the evolution of C_4 physiology was probably the development of a basal CO_2 concentration mechanism that was previously referred to as a C_3–C_4 intermediate state and that is now called C_2 photosynthesis (Sage et al., 2012; Williams et al., 2013; Mallmann et al., 2014). Some present-day species use a C_2 cycle for CO_2 fixation, having not established a full C_4 cycle. It functions by confining glycine decarboxylase (GDC), the CO_2-releasing enzyme of

photorespiration, to BS cells. This increases the probability of re-fixation of CO_2 released from photorespiration, because BS cells are surrounded by mesophyll cells that contain Rubisco and can fix released CO_2 before it diffuses from the leaf. Establishment of the C_2 cycle also required changes in leaf anatomy, such as enlargement of BS cells and reduction in vein spacing that might have preceded physiological adaptations (Christin et al., 2013).The C_2 cycle is seemingly sufficient at enhancing internal CO_2 concentration in the leaf three-fold and reducing photorespiratory losses significantly (Keerberg et al., 2014).

The mechanism by which GDC was shifted to BS cells has been studied in the model genus *Flaveria* that contains C_3, C_4 and C_2 plants. In C_2 and C_4 species, the P-subunit of the GDC complex is almost exclusively expressed in BS cells. Interestingly, C_3 species of *Flaveria* already contained one BS-specific isoform of the gene encoding the P-subunit (*Gldp*) and one additional isoform that was expressed in all photosynthetic tissues. The latter isoform was inactivated during evolution of C_2 photosynthesis (Schulze et al., 2013). An upstream promoter of the BS-specific isoform was employed to also provide some basal transcription in mesophyll cells that is probably required for one-carbon metabolism independent of photorespiration (Wiludda et al., 2012).

As well as the analyses in *Flaveria*, some molecular information is also available on *Gldp* genes in Brassicaceae. C_4 photosynthesis does not appear to have evolved in Brassicaceae (Sage et al., 2012), but C_2 plants have been described in the genera *Moricandia* and *Diplotaxis* (Apel et al., 1997).

A comparison of the *Gldp* promoter sequences of *Arabidopsis thaliana* (C_3), *Brassica napus* (C_3) and *Moricandia nitens* (C_2) revealed conserved regions in the C_3 species that were absent in the C_2 species (Zhang et al., 2004). Deletion analyses of the *B. napus* promoter identified a 135-bp element that controlled promoter strength in the leaf (Zhang et al., 2004). Homologous sequences to this element were cloned from different *Moricandia* and *Diplotaxis* species and putative conserved elements were suggested (Zhang et al., 2004).

In *A. thaliana*, the genes encoding the P-subunit of GDC are *Gldp*1 and *Gldp*2. According to mutant analyses, these two genes are functionally redundant and one of the two genes is sufficient for normal plant growth, at least under laboratory conditions (Engel et al., 2007). However, double knockouts are lethal even under high-CO_2 growth conditions where photorespiration is largely suppressed (Engel et al., 2007). Thus, GDC activity seems to be essential for plant growth independent of photorespiration.

In this study, we analyzed promoter deletion mutants of Arabidopsis *Gldp* genes. We show that deletion of a highly conserved sequence element was sufficient to restrict promoter activity to BS cells and veins. This element was not required for diurnal promoter regulation or promoter activity in non-photosynthetic tissues. Implications for the evolution of C_2 metabolism are discussed.

RESULTS

We wanted to analyze transcriptional regulation of the *Gldp*1 gene in *A. thaliana*. To this end, we created mutants containing GUS reporter constructs on which the *Gldp*1 upstream region had been deleted to varying extents and analyzed three important features: promoter strength, diurnal regulation and spatial expression pattern (Figure 1). A quantitative analysis of between 4 and 32 independent transformation events is also shown in Figure S1 in the Supporting Information.

The P4 construct contained all intergenic sequences (1270 bp) upstream of the *Gldp*1 transcribed sequence and the 5^0 untranslated region (UTR) of the gene. For this construct, we observed strong accumulation of GUS mRNA in the morning (1 h) and a four-fold reduction of GUS mRNA amounts in the afternoon (7 h) (Figure 1b). This pattern perfectly resembled diurnal variation in transcript abundance of the endogenous *Gldp*1 gene (Figure S2), and was also supported by the abundance of RNA polymerase II on the endogenous promoter at 1 and 7 h (Figure S2). The

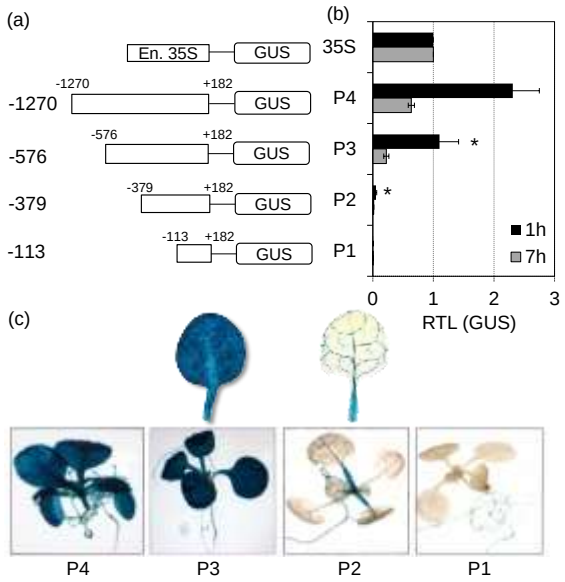


Figure 1. Analysis of *AtGldp*1 promoter deletions.

(a) Schematic representation of the enhanced 35S promoter::GUS construct and promoter deletion constructs.

(b) Relative transcript levels (RTL) of GUS in 2-week-old seedlings of stable upstream deletion mutants (n = 4 SEM) after 1 h and 7 h light. * $P < 0.05$ relative to next promoter deletion.

(c) Histochemical GUS staining of 2-week-old upstream deletion mutants (UDMs). A representative picture is shown for each construct. A summary of all tested events is given in Figure S1. An enlarged photograph of a single leaf is included for the P3 and P2 constructs.

promoter was active in all leaf cells, the leaf petiole and the roots (Figure 1c).

In the P3 construct, 694 bp of the upstream sequence were deleted (Figure 1a). We observed a two-fold reduction in maximum accumulation of mRNA in the morning, but diurnal transcript variation remained mainly unaffected (Figure 1b). Moreover, the spatial expression pattern was identical to that of the P4 construct (Figure 1c).

In the P2 construct, a further 197 bp were deleted. This resulted in a strong drop of maximum RNA accumulation by more than 10-fold. However, the amounts of mRNA were still lower at 7 h than at 1 h, indicating that diurnal variation was unaffected by this deletion. Interestingly, the spatial expression pattern was drastically altered (Figure 1c). We observed GUS staining along the leaf veins but no longer in leaf mesophyll cells. Petioles and roots were still stained. Because deletion of this region abolished expression in mesophyll cells we provisionally defined it as the 'M box'.

A further deletion down to 113 bp of the upstream sequence (P1 construct) reduced mRNA accumulation to less than 1% of the full promoter. Consequently, diurnal variation could not be measured (Figure 1b) and we did not observe significant GUS staining in leaves. However, the promoter was still active in roots, albeit at reduced levels (Figure 1c). Because deletion of this region resulted in loss of expression in veins, it was named the 'V box'.

The spatial expression pattern observed for the P2 construct resembled the expression of *Gldp* genes in C_2 plants, i.e. restriction of *Gldp* expression to the vasculature (see Introduction). We were therefore interested in further studying the M box as defined by this deletion construct.

First, we created subdeletions to narrow down the sequences required for promoter activity in mesophyll cells. Subdeletion constructs are shown in Figure 2(a). Deletion of the upstream 59 bp of the DNA sequence contained in the P3 construct (P3D1) reduced promoter activity down to the level observed for the P2 construct. Diurnal regulation was again unaffected by this deletion (Figure 2b). Spatial distribution of the GUS signal in the P3D1 mutant resembled the pattern observed for the mutant containing the P2 construct, i.e. labeling of leaf veins, petioles, the hypocotyl and roots, but loss of GUS signal from mesophyll cells (Figure 2c). Further deletions did not change this pattern (Figure 2c). This pattern was consistent in 11 independent transformation events (Figure S3a) and comparable to *FtGLDPA*-PR7 constructs containing a promoter construct from the *Flaveria trinervia GLDPA* promoter (Figure s3b) that had been shown to be active in veins and BS cells in a previous study (Wiludda et al., 2012). In transverse sections, mutants having the P3 construct showed GUS staining in all leaf cells, whereas those having the P3D1 construct had staining in veins, but also adjacent BS cells (Figure 2d). GUS staining of BS cells

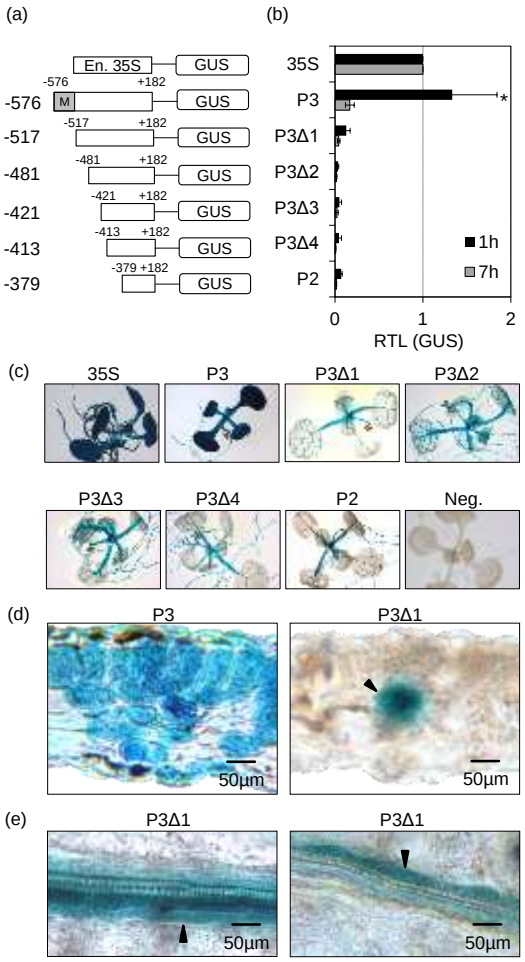


Figure 2. The M box of the *AtGldp*1 promoter.

(a) Schematic representation of subdeletions in the P3 upstream deletion mutant.

(b) Relative transcript level (RTL) of GUS in 2-week-old seedlings (n = 4 SEM). *P < 0.05 relative to the P3D1 construct.

(c) Histochemical GUS staining of 2-week-old seedlings of P3-subdeletion lines and controls. A representative image is shown for each construct. A summary of all tested events is given in Figure S1. Genotypes are indicated in the figure.

(d) Transverse sections of histochemical GUS staining of 2-week-old seedlings overexpressing the P3 or the P3D1 construct.

(e) Longitudinal sections of histochemical GUS staining of 2-week-old seedlings overexpressing the P3D1 construct. Arrowheads point to bundle sheath cells.

was also apparent in longitudinal sections of leaves from mutants transformed with the P3D1 construct (Figure 2e). These analyses allowed the M box to be pinned down to 59 bp ranging from 576 bp to 517 bp of the *AtGldp*1 upstream region. Promoter::reporter constructs without the M box were inactive in mesophyll cells but they were still active in BS cells.

Second, we analyzed nucleosome density on the *AtGldp*1 promoter (Figure 3a). We expected that regions important for promoter activity and/or bound by transcription factors would show reduced nucleosome density (Lee et al., 2004). We observed a low nucleosome density in comparison with adjacent regions between 500 and 600 bp where the M box was located. This drop in nucleosome density was independent of diurnal regulation and occurred in chromatin isolated from plants after both 1 and 7 h of light.

Third, we tested whether the M-box region was conserved on orthologous *Gldp*1 promoters from other Brassicaceae. We used the EARS algorithm that was specifically designed for comparison of promoter sequences where conserved elements might strongly vary in distance to the transcription initiation site (TIS) (Picot et al., 2010). We compared 2000 bp upstream of the TIS of *Gldp*1 homologues from the sequenced genomes of Brassicaceae species (Goodstein et al., 2012). A phylogenetic tree of Gldp sequences from the different species is shown in Figure S4. Beside *Gldp*1 homologues, we included the *AtGldp*2

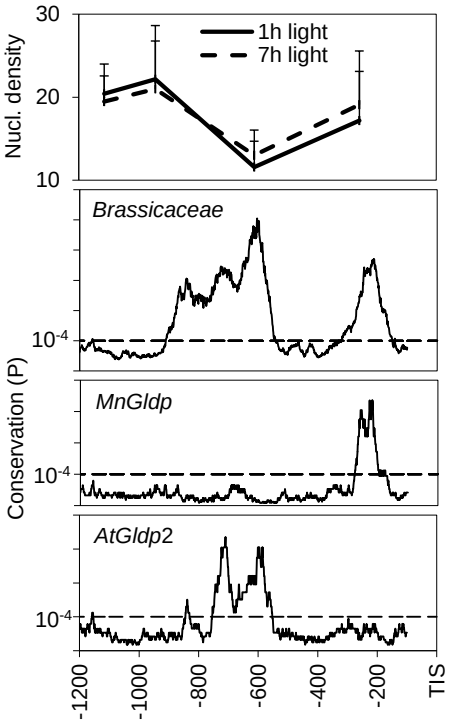


Figure 3. DNA accessibility at the M box and sequence conservation.
(a) Nucleosome densities at different regions of the *AtGldp*1 promoter. Nucleosome (Nucl.) densities were recorded by chromatin immunoprecipitation at 1 h and 7 h after illumination (n = 3 SEM).
(b) Conservation profile of *Gldp* promoter sequences. The upper panel shows the conservation among all available Brassicaceae sequences (see Figure S5). The middle panel shows conservation profile for the *Moricandia nitens* homologue. The lower panel shows conservation profile for *AtGldp*2. The y-axis is the alignment score expressed as P-values. The default significance threshold was P = 0.0001. Alignment was computed using the EARS algorithm (Picot et al., 2010). TIS, transcription initiation site.

promoter sequence as well as the previously isolated *B. napus* and *M. nitens* sequences. We observed two promoter regions that were highly conserved in this multispecies comparison (Figure 3b, Brassicaceae).

The first conserved region was located between 900 bp and 500 bp relative to the TIS on the *AtGldp*1 gene. Conservation in this region peaked between 650 bp and 500 bp. This peak overlapped with the experimentally validated M box (see Figure 2). A sequence alignment of the highest conserved region is shown in Figure S5(a). Within the upstream sequences from Brassicaceae, the M-box homology region was very differently positioned. Interestingly, we observed a fully conserved CCAAT binding site for nuclear factor Y (NF-Y) transcription factors (Mantovani, 1999) in this region. However, conservation in this region was not detected in a pair-wise comparison of the *A. thaliana* and *M. nitens* (Figure 3b, *MnGldp*) promoters, suggesting that the M box was not present on the *Moricandia Gldp* promoter.

The second conserved region according to EARS was located between 300 bp and 100 bp relative to the TIS on the *AtGldp*1 promoter. This region was conserved in all Brassicaceae *Gldp*1 genes including *M. nitens*, but not on the *AtGldp*2 promoter (Figure 3b, AtGldp2). The conserved region overlapped with the V box on the *AtGldp*1 promoter (see Figure 1). The corresponding sequence alignment (Figure s5b) again indicated variable positioning of this conserved region on the analyzed promoters. However, we observed a highly conserved AT-rich region on all promoter sequences. Thus, sequence analysis indicated high conservation of the M-box and V-box regions on *Gldp*1 homologues in the Brassicaceae, but absence of an M box on the *M. nitens* promoter. On the *AtGldp*2 gene, the M box could be detected but not the V box.

Fourth, we experimentally tested whether a functional M-box region also existed on the *AtGldp*2 gene, as suggested by the upstream sequence conservation profile (Figure 3). Deletion experiments with the upstream sequence are shown in Figure 4. Deletion of a 177-bp sequence containing the M-box homology region (2-P2 construct) resulted in a drastic reduction in promoter activity in seed-lings. In histochemical GUS stainings, plants expressing the 2-P2 construct showed staining of the roots and the vasculature. Compared with the P2 construct of *AtGldp*1, we observed more variation between individual events, for example more than 20% of all events still showed some GUS staining in M cells (Figure S1).

Our results up to this point indicated that C_3 Brassicaceae *Gldp* upstream sequences contained an element required for expression in M cells (59-bp M box). To analyze whether this box was also sufficient to induce expression of promoters in the mesophyll, we fused the M box to the P2 deletion construct of the *AtGldp*1 promoter that was not active in mesophyll cells (M::P2 construct, Figure 5a).

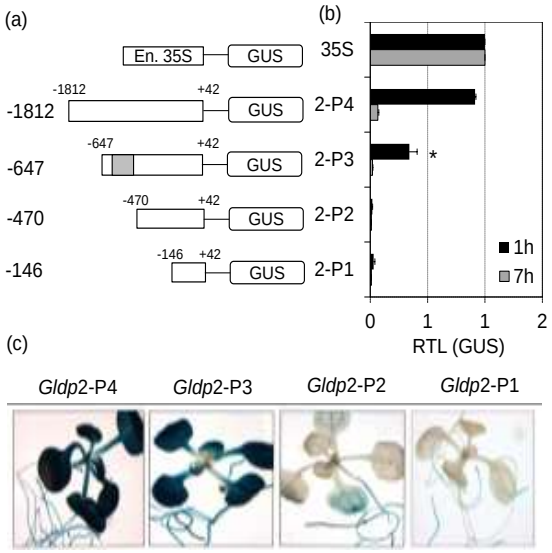


Figure 4. Analysis of *AtGldp2* promoter deletions.

(a) Schematic representation of the enhanced 35S promoter::GUS construct and upstream deletion mutants of *AtGldp2*. The gray box indicates the putative M-box region.

(b) Relative transcript levels (RTL) of GUS in 2-week-old seedlings (n = 4 SEM). *P < 0.05 relative to the 2-P2 construct.

(c) Histochemical GUS staining of 1-week-old and 2-week-old seedlings. A typical image is shown for each construct. A summary of all tested events is given in Figure S1.

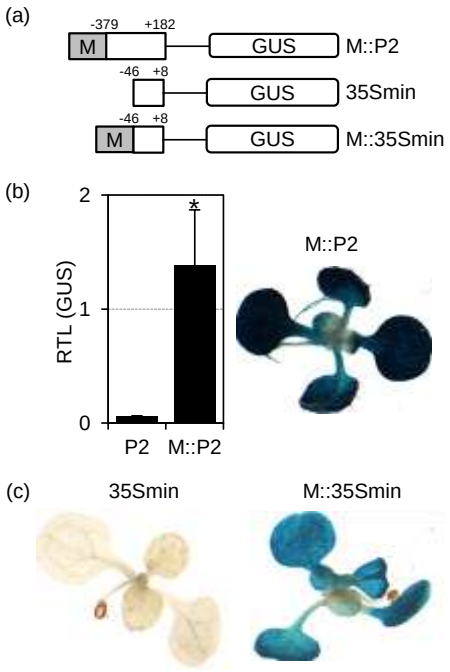


Figure 5. The M box is sufficient for controlling mesophyll expression.

(a) Schematic representation of the construct.

(b) Complementation of the P2 upstream deletion mutant of *AtGldp1*. Relative transcript levels (RTL) of GUS in 2-week-old seedlings (n = 4 SEM; *P < 0.05) and histochemical GUS staining of a 2-week-old seedling.

(c) Enhancement of the truncated 35S minimal (35Smin) promoter. Histochemical GUS staining of a 2-week-old seedling. Genotypes are indicated in the figure. A typical image is shown for each construct. A summary of all tested events is given in Figure S1.

We found that GUS mRNA accumulation was more than 10-fold stronger in leaves from M::P2 plants compared with P2 plants and was comparable to P3 plants (Figure 5b, cf. Figure 1). This coincided with a reconstitution of GUS expression in mesophyll cells (M::P2, Figure 5b). We also added the M box to the 35S minimal promoter (Benfey et al., 1990) (M::35Smin construct, Figure 5a). Basal activity of the 35S minimal promoter was undetectable by GUS staining in our hands (35Smin, Figure 5c). Addition of the M box strongly enhanced promoter activity in leaves but not in roots (M::35Smin, Figure 5c). These data indicate that the M box can act as a tissue-specific enhancer.

Together, a simple mutation of a highly conserved promoter element of *Gldp* genes was sufficient to restrict promoter activity to BS and vasculature cells. This is reminiscent of the Gldp expression pattern observed in C_2 plants.

DISCUSSION

Using a combination of promoter-deletion studies and bioinformatic analyses, we identified separate DNA motifs (boxes) required for the expression of Gldp genes in mesophyll cells and the vasculature. The vasculature element was also active in BS cells (Figure 2). This observation is supported by results showing that the Scarecrow (SCR) transcription factor binds the *AtGldp1* promoter in vivo (Cui et al., 2014). Scarecrow is part of the SCR/shortroot (SHR) transcription factor system that is important for the control of BS-specific gene expression in the leaf (Slewinski et al., 2012; Fouracre et al., 2014). The vasculature element was conserved in the upstream sequences of *Gldp1* homologues from all tested Brassicaceae, suggesting that these plants had the potential to develop C2-type expression of GLDP by simple mutation of the M box. Support for this scenario is provided by the lack of a detectable M box and the presence of the V box on the *Gldp* promoter from *M. nitens*, the only currently available promoter sequence from a C2 plant in the Brassicaceae (Zhang et al., 2004). As C2 metabolism was probably an intermediate state during the development of C4 metabolism (Sage, 2004), it will be interesting to test whether an M box is present in the *Gldp* promoters of *Gynandropsis gynandra* (*Cleomaceae*), the nearest C4 relative to Arabidopsis (Brown et al., 2005).

Deletion of the M box selectively abolished *Gldp* expression in mesophyll cells, but did not affect promoter activity in leaf petioles or roots, for example (Figure 2). Conversely, the M box was sufficient to induce expression of the weak 35S minimal promoter or a vasculature/BS-specific version of the *Gldp1* promoter in mesophyll cells

(Figure 5), although we did not test whether this induction can be also observed in other tissues. Thus, this promoter element is probably recognized at least in mesophyll cells by a transcription factor. Binding of a transcription factor is also consistent with the low nucleosome density in this region (Figure 3a). On the M-box sequence, we observed strong conservation of a CCAAT-binding motif for NF-Y (Petroni et al., 2012). Interestingly, in rice cell-type-specific gene expression in mesophyll cells and veins, respectively, has been associated with the presence of CCAAT boxes and specific NF-Y proteins (Jiao et al., 2009). Nuclear factor Y itself is made up from three subunits, but requires a nearby binding site for another transcription factor in order to function (Mantovani, 1999). A second highly conserved DNA sequence that was observed adjacent to the CCAAT motif is the CTTTCA motif. This motif, as well as the CCAAT box, was enriched among *cis* elements of genes that were preferentially expressed in mesophyll cells of maize (Wang et al., 2014), suggesting that they might be also involved in mesophyll expression in other plants. The CCAAT and CTTTCA motifs are separated by 20 bp, corresponding to exactly two helical turns on a nucleosome (Gottesfeld, 1987). This is important because the rotational positioning on the DNA helix is seemingly influencing the interaction of NF-Y factors and other adjacent transcription factors (Mantovani, 1999). Thus, both sites may well interact to control mesophyll expression of *Gldp* genes.

Even the shortest promoter versions of *AtGldp*1 and *AtGldp*2 (P1) were still active in roots, and diurnal regulation was unaffected by the deletions (Figures 1 and 4). We therefore assume that elements for root expression and diurnal control are located in the core promoter downstream of 113 bp (*AtGldp*1) and 146 bp (*AtGldp*2), respectively. The abundance of RNA polymerase II on the *AtGldp*1 promoter was also high in the morning and low in the evening (Figure S2), indicating that diurnal regulation is controlled by promoter activity. Consistent with our data, diurnal regulation has also been associated with elements on the core promoter, for example on the Constans gene (Ito et al., 2012) or, more generally, genes that encode organellar proteins in Arabidopsis (Giraud et al., 2010). As we included the endogenous 5′ UTRs in our promoter–reporter constructs, we cannot exclude an impact of RNA stability on diurnal regulation or root expression.

The importance of the restriction of GLDP to BS cells in the evolution of C_2 species was highlighted in a study of *M. nitens* (C_2) and *B. napus* (C_3) hybrids (Rawsthorne et al., 1998). In these hybrids, BS cell anatomy was similar to the C_3 plant and did not show the typical adaptations such as enlargement and higher organelle content of the BS cells of C_2 species. However, GLDP was enriched in BS cells over mesophyll cells, and this already resulted in a reduction in the CO_2 compensation point, a typical physiological property of C_2 plants. Very similar observations including enrichment of BS cells in GLDP and reduced compensation points were also described for hybrids of *Diplotaxis tenuifolia* (C_2) and *Raphanus sativa* (C_3) (Ueno, 2003). Thus, limited enrichment of GLDP in BS cells might already provide an evolutionary advantage towards the development of C_4 characteristics (Sage, 2004). This resembles the situation in C_3 Flaveria where one *Gldp* gene was ubiquitously expressed and the other only in BS cells. This set-up facilitated the development of C_2-specific gene expression by pseudogenization of the ubiquitously expressed *Gldp* gene (Schulze et al., 2013). In Arabidopsis, the two Gldp genes are both expressed in leaves and roots, and are apparently functionally redundant (Engel et al., 2007). A recent study of the BS translatome in Arabidopsis also revealed that both genes are expressed in BS cells (Aubry et al., 2014). We observed a conserved M-box region on both promoters, but a conserved V-box region only on the *Gldp*1 promoter (Figure 3). The latter is consistent with the less defined spatial expression pattern of *Gldp*2 M-box deletion mutants (Figure 4), although it remains to be shown whether *Gldp*2 reporter constructs are active in BS cells. Thus, the Arabidopsis *Gldp*1 promoter already contained all information required for C_2-type expression and simple mutation or deletion of an upstream *cis*-acting element (M box) was sufficient to establish C_2-type expression without affecting other important promoter features such as promoter strength, diurnal regulation or expression in non-photosynthetic tissues. This scenario would be comparable to the *Flaveria phosphoenolpyruvate carboxylase* (*Pepc*) promoter that acquired tissue specificity by mutation of an upstream promoter element (Akyildiz et al., 2007). Even if this mutation initially only occurred on one of the two *Gldp* genes, the hybrid studies suggest that this might have provided an advantage. Alternatively, mutation of the trans-acting element controlling mesophyll expression of *Gldp* genes through interaction with the M box would probably simultaneously abolish expression of both *Gldp* genes in mesophyll cells. The absence of the M box on the *M. nitens Gldp* promoter (Zhang et al., 2004) currently favors the former hypothesis.

EXPERIMENTAL PROCEDURES

Plant growth and transformation

Arabidopsis thaliana Col-0 plants were grown under short-day conditions for 14 days at 21°C/8-h light (120 µE m^{-2} sec^{-1}) and at 18°C/16-h darkness. Inflorescences of 8-week-old Arabidopsis plants (Col-0) were transformed by the *Agrobacterium*-mediated floral dip method (Clough and Bent, 1998).

DNA extraction and PCR

Genomic DNA was extracted from Col-0 as described in Edwards et al. (1991). Promoter fragments were amplified using Phusion Hot start II DNA polymerase (Thermo Scientific, http://www.ther-moscientific.com/). Standard parameters were used (denaturation

at 98°C for 10 sec, annealing at 60°C for 30 sec and extension at 72°C for 1 min). Chimeric constructs were amplified as 5′ end mutations. Primers are listed in Table S1.

Cloning

For directional cloning of PCR generated fragments we used pENTR™/D-TOPO according to the manufacturer's protocol (Invit-rogen, http://www.invitrogen.com/). Sequenced fragments were subcloned into a binary vector for plant transformation, pSAG (Figure S6), via homologous recombination using LR clonase (Invitrogen).

Quantitative RT-PCR

The RNA was extracted using TRIzol. Extracted RNA was used for first-strand cDNA synthesis using M-MLV Reverse Transcriptase (Promega, http://www.promega.com/) following the manufacturer's protocol. Quantitative RT-PCR analysis was carried out using Platinum SYBR Green qPCR SuperMix (Invitrogen). A PCR reaction without template DNA was included to monitor DNA contamination. Endogenous *AtGldp*1 and *AtGldp*2 transcripts were standardized to the level of GAPDH in each sample, and GUS transcripts were standardized to the level of *npt*II transcripts driven by the selection marker on each construct. The primers used for PCR amplification are listed in Table S1.

Leaf sections and GUS stains

The GUS activity was analyzed histochemically by transferring plant samples to a GUS staining buffer (Jefferson et al., 1987). Intact seedlings were stained and at 37°C, overnight (o/n), and de-stained in 96% ethanol at room temperature (RT; 21LC), o/n. Leaf sections were obtained by embedding non-stained leaves from 2-week-old plants in 6% (w/v) agarose. Agarose-leaf molds were attached to a titanium vibratome plate using superglue and leaf sections (thickness 45–65 lm) were obtained in a matrix of sodium phosphate buffer (50 mM Na_2HPO_4 pH 7.0) using a vibratome (Leica VT 1000S, Leica Microsystems, http://www.leica-microsystems.com/). Leaf sections were transferred to GUS staining buffer, stained at 37°C for 1 h and de-stained in 96% ethanol at RT for 1 h. Sections were mounted for microscopy using a painting brush that had been previously soaked with 10% glycerol. Sections were always repeated from leaves of three independent transformation events.

Conservation analysis

Evolutionary analysis of regulatory sequences (EARS) (Picot et al., 2010) was used for alignment of promoter sequences. The window size was 100 bp. Probabilities were plotted using Microsoft Excel 2010.

Chromatin immunoprecipitation

Chromatin fragments were precipitated as described in Jaskiewicz et al. (2011) using an antibody directed to an invariant domain of histone H3 (ab1791, Abcam, http://www.abcam.com/) or directed to RNA polymerase II (ab817, Abcam). The presence of DNA in the precipitate was quantified by qPCR relative to chromatin input. Primers are listed in Table S1.

ACCESSION NUMBERS

Arabidopsis *Gldp* genes (At4g33010 and At2g26080) are published in the Arabidopsis Information resource. Sequence data for other Brassicaceae species are available at Phytozome: *Arabidopsis lyrata*, *AlyGldp*1 (sc7:3493441), *AlyGldp*2 (sc4:481465); *Boechera stricta*, *BstGldp*1 (Bostr. 7867s1152.1); *Capsella grandiflora*, *CgrGldp*1 (Cagra.4093s 0031.1), *CgrGldp*2 (Cagra.3997s0002.1); *Capsella rubella*, *CrbGldp*1 (Carubv10004049m); *Brassica rapa*, *BraGldp*1 (Brara.A00490.1); *Eutrema salsugineum*, *EsaGldp* (Thhalv 10001891m); *Citrus clementine*, *CclGldp* (Ciclev10007310m) and *Citrus sinensis*, *CsiGldp* (orange1.1g001531m). *Brassica napus BnGldp* and *M. nitens MnGldp* sequences are available at GenBank under accession numbers AY54471 and AY544772.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Deutsche Forschungsgemeinschaft to CP (PE819/7-1) and from the Deutscher Akademischer Austauschdienst to WA. We would like to thank Natalija Hohnjec for providing help with the vibratome and Sascha Offermann for critical reading of the manuscript.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Summary of the frequency of different expression patterns in independent transformation events.

Figure S2. Analysis of expression control by *AtGldp* promoters.

Figure S3. Histochemical GUS staining of 2-week-old seedlings.

Figure S4. Molecular phylogenetic analysis of *Gldp* genes from the Brassicaceae.

Figure S5. DNA sequence alignment of promoter upstream sequences.

Figure S6. Map of the binary plasmid used for GUS assays.

Table S1. Primers used in this study.

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 C4 dicot *Flaveria trinervia*. *Plant Cell*, **19**, 3391–3402.

Apel, P., Horstmann, M. and Pfeffer, M. (1997) The Moricandia syndrome in species of the Brassicaceae - evolutionary aspects. *Photosynthetica*, **33**, 205–215.

Aubry, S., Smith-Unna, R.D., Boursnell, C.M., Kopriva, S. and Hibberd, J.M. (2014) Transcript residency on ribosomes reveals a key role for the *Arabidopsis thaliana* bundle sheath in sulfur and glucosinolate metabolism. *Plant J.* **78**, 659–673.

Benfey, P.N., Ren, L. and Chua, N.-H. (1990) Tissue-specific expression from CaMV 35S enhancer subdomains in early stages of plant development. *EMBO J.* 9, 1677–1684.

Brown, N.J., Parsley, K. and Hibberd, J.M. (2005) The future of C_4 research– maize, *Flaveria* or *Cleome*? *Trends Plant Sci.* **10**, 215–221.

Christin, P.-A., Osborne, C.P., Chatelet, D.S., Columbus, J.T., Besnard, G., Hodkinson, T.R., Garrison, L.M., Vorontsova, M.S. and Edwards, E.J. (2013) Anatomical enablers and the evolution of C4 photosynthesis in grasses. *Proc. Natl Acad. Sci. USA*, **110**, 1381–1386.

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.

Cui, H., Kong, D., Liu, X. and Hao, Y. (2014) SCARECROW, SCR-LIKE 23 and SHORT-ROOT control bundle sheath cell fate and function in *Arabidopsis thaliana*. *Plant J.* **78**, 319–327.

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.

Engel, N., van den Daele, K., Kolukisaoglu, U., 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.

Fouracre, J.P., Ando, S. and Langdale, J.A. (2014) Cracking the Kranz enigma with systems biology. J. Exp. Bot. **65**, 3327–3339.

Giraud, E., Ng, S., Carrie, C., Duncan, O., Low, J., Lee, C.P., van Aken, O., Millar, A.H., Murcha, M. and Whelan, J. (2010) TCP transcription factors link the regulation of genes encoding mitochondrial proteins with the circadian clock in *Arabidopsis thaliana.* Plant Cell, **22**, 3921–3934.

Goodstein, D.M., Shu, S., Howson, R. et al. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178– D1186.

Gottesfeld, J.M. (1987) DNA sequence-directed nucleosome reconstitution on 5S RNA Gene of *Xenopus laevis*. *Mol. Cell. Biol*. **7**, 1612–1622.

Hibberd, J.M. and Quick, W.P. (2002) Characteristics of C_4 photosynthesis in stems and petioles of C_3 flowering plants. *Nature*, **412**, 451–454.

Ito, S., Song, Y.H., Josephson-Day, A.R., Miller, R.J., Breton, G., Olmstead, R.G. and Imaizumi, T. (2012) FLOWERING BHLH transcriptional activators control expression of the photoperiodic flowering regulator CONSTANS in Arabidopsis. *Proc. Natl Acad. Sci.* USA, **109**, 3582–3587.

Jaskiewicz, M., Peterhansel, C. and Conrath, U. (2011) Detection of histone modifications in plant leaves. *J. Vis. Exp.* **55**, e3096.

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

Jiao, Y., Tausta, S.L., Gandotra, N. et al. (2009) A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nat. Genet.* **41**, 258–263.

Keerberg, O., Parnik, T., Ivanova, H., Bassuner, B. and Bauwe, H. (2014) C_2 photosynthesis generates about 3-fold elevated leaf CO_2 levels in the C_3-C_4 intermediate species *Flaveria pubescens*. *J. Exp. Bot.* **65**, 3649–3656.

Kinsman, E.A. and Pyke, K.A. (1998) Bundle sheath cells and cell-specific plastid development in Arabidopsis leaves. *Development*, **125**, 1815–1822.

Langdale, J.A. (2011) C_4 cycles: past, present, and future research on C_4 photosynthesis. *Plant Cell*, **23**, 3879–3892.

Lee, C.-K., Shibata, Y., Rao, B., Strahl, B.D. and Lieb, J.D. (2004) Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat. Genet.* **36**, 900–905.

Leegood, R.C. (2008) Roles of the bundle sheath cells in leaves of C3 plants. *J. Exp. Bot.* **59**, 1663–1673.

Lundquist, P.K., Rosar, C., Brautigam, A. and Weber, A.P.M. (2014) Plastid signals and the bundle sheath: mesophyll development in reticulate mutants. *Mol. Plant*, **7**, 14–29.

Mallmann, J., Heckmann, D., Brautigam, A., Lercher, J.M., Weber, P.A., Westhoff, P. and Gowik, U. (2014) The role of photorespiration during the evolution of C_4 photosynthesis in genus *Flaveria*. *ELIFE*, **3**, e02478.

Mantovani, R. (1999) The molecular biology of the CCAAT-binding factor NF-Y. *Gene*, **239**, 15–27.

Peterhansel, C., Horst, I., Niessen, M., Blume, C., Kebeish, R.,Kurkcuoglu, S. and Kreuzaler, F. (2010) Photorespiration. *Arabidopsis Book*, **8**, e0130.

Petroni, K., Kumimoto, R.W., Gnesutta, N., Calvenzani, V., Fornari, M., Tonelli, C., Holt, B.F. and Mantovani, R. (2012) The promiscuous life of plant NUCLEAR FACTOR Y transcription factors. *Plant Cell*, **24**, 4777– 4792.

Picot, E., Krusche, P., Tiskin, A., Carre, I. and Ott, S. (2010) Evolutionary analysis of regulatory sequences (EARS) in plants. *Plant J.* **64**, 165–176.

Rawsthorne, S., Morgan, C.L., O'Neill, C.M., Hylton, C.M., Jones, D.A. and Frean, M.L. (1998) Cellular expression pattern of the glycine decarboxylase P protein in leaves of an intergeneric hybrid between the C_3-C_4 intermediate species *Moricandia nitens* and the C_3 species *Brassica napus*. *Theor. Appl. Genet.* **96**, 922–927.

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

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

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

Slewinski, T.L., Anderson, A.A., Zhang, C. and Turgeon, R. (2012) Scarecrow plays a role in establishing Kranz anatomy in maize leaves. *Plant Cell Physiol.* **53**, 2030–2037.

Ueno, O. (2003) Structural and biochemical dissection of photorespiration in hybrids differing in genome constitution between *Diplotaxis tenuifolia* (C_3-C_4) and *Radish* (C_3). *Plant Physiol.* **132**, 1550–1559.

Wang, L., Czedik-Eysenberg, A., Mertz, R.A. et al. (2014) Comparative analyses of C_4 and C_{23} photosynthesis in developing leaves of maize and rice. *Nat. Biotechnol.* **32**, 1158–1165.

Williams, B.P., Johnston, L.G., Covshoff, S. and Hibberd, J.M. (2013). Phenotypic landscape inference reveals multiple evolutionary paths to C4 photosynthesis. *ELIFE*, **2**, e0096.

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

Zhang, C., Xu, G., Huang, R., Chen, C. and Meng, J. (2004) A dominant gdcP-specific marker derived from *Moricandia nitens* used for introducing the C_3-C_4 character from *M. nitens* into *Brassica* crops. Plant Breeding, 123, 438–443.

SUPPORTING INFORMATION

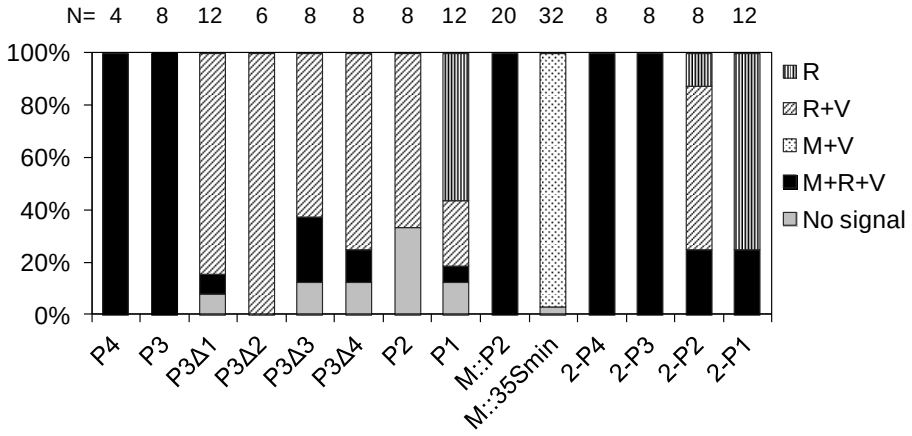


Figure S1: Summary of the frequency of different expression patterns in independent transformation events. N: number of independent events, R: root, V: vasculature, M: mesophyll.

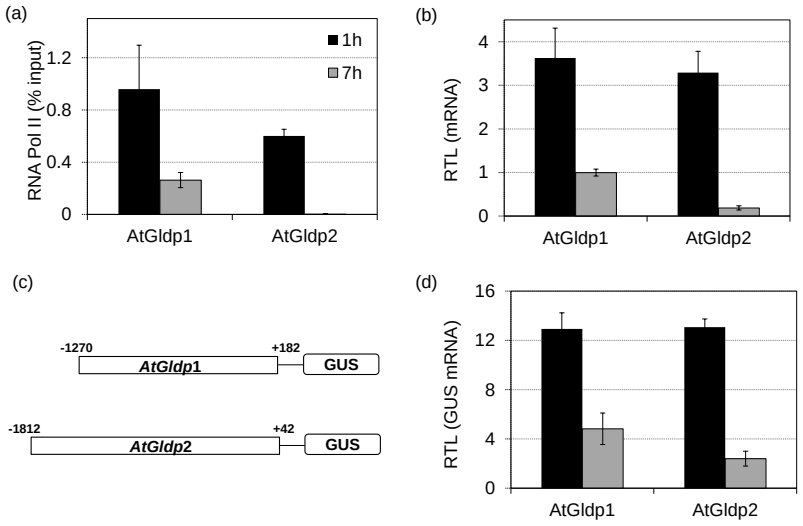


Figure S2: Analysis of expression control by *AtGldp* promoters. (a) Relative enrichment of RNA polymerase II on positions +400bp and +495bp relative to the transcription initiation site (TIS) of *AtGldp*1 and *AtGldp*2, respectively. Chromatin immunoprecipitates (ChIPs) from 1h and 7h after onset of light were analysed. (n=3±SEM). (b) Relative transcript level (RTL) in 2-week-old wild type seedlings 1h and 7h after onset of light. (n=3±SEM). (c) Schematic representation of the promoter::reporter GUS constructs. (d) RTL of GUS mRNA in 2-week-old mutant seedlings. RNA accumulation was analysed at 1h and 7h after onset of light (n=3±SEM).

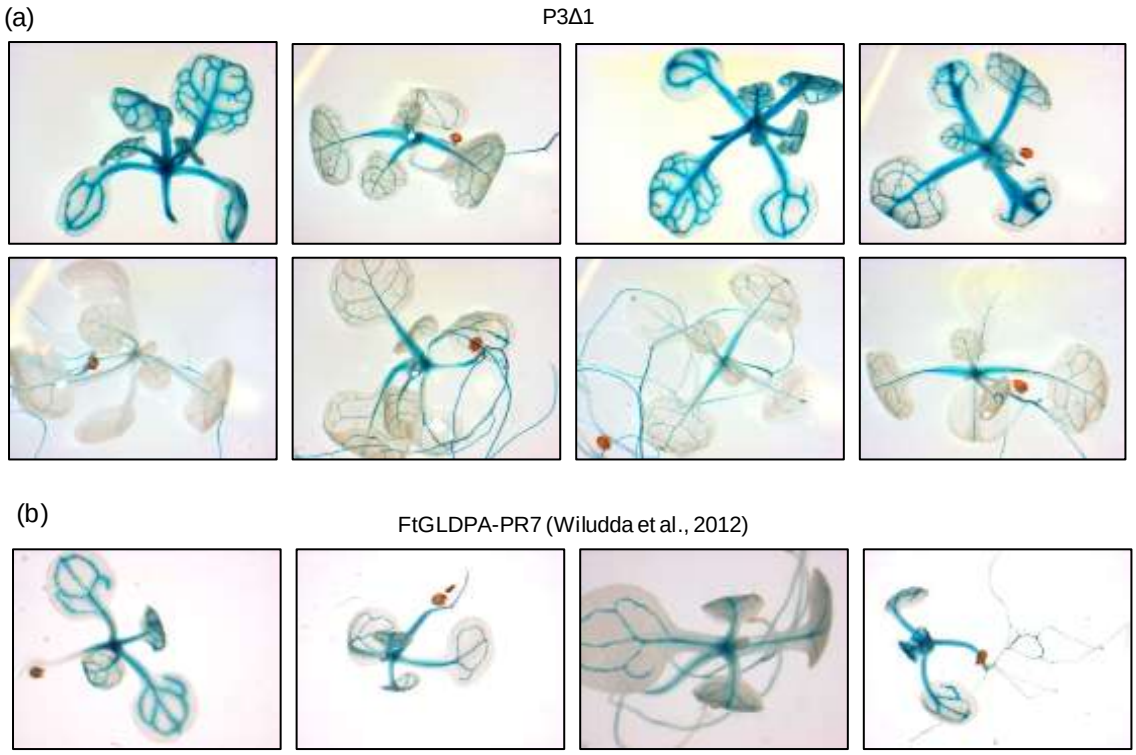


Figure S3: Histochemical GUS staining of 2-week-old seedlings. (a) 8 independent events transformed with the P3Δ1 construct. (b) 4 independent events transformed with the *FtGLDPA*-PR7 construct that had been shown to control BS-specific expression in a previous study (Wiludda et al., 2012).

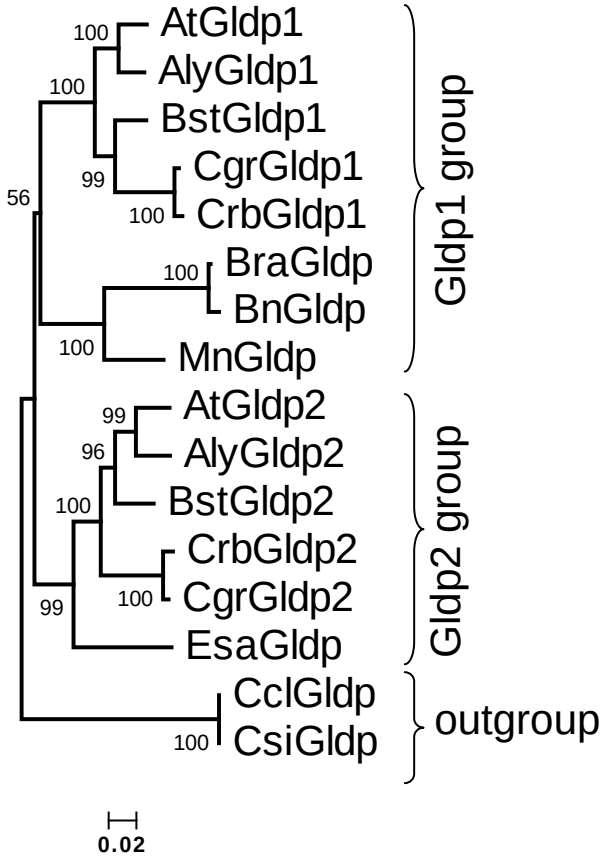


Figure S4: Phylogenetic analysis of *Gldp* genes from the *Brassicaceae*. The tree was constructed using the Neighbor-joining method in MEGA6 (Tamura *et al.*, 2013) with 1000 bootstrap replicates (Felsenstein, 1985) and is based on the first 1.2 kb of the coding sequence of each gene. GLDP genes from *Citrus sinensis* (Csi) and *Citrus clementine* (Ccl) were used as the outgroup. Ath: *Arabidopsis thaliana*, Aly: *Arabidopsis lyrata*, Bst: *Boechera stricta*, Cgr: *Capsella grandiflora* , Crb: *Capsella rubella*, Bra: *Brassica rapa*, Bn: *Brassica napus*, Esa: *Eutrema salsugineum* (Previously *Thellungiella halophila*), Mn: *Moricandia nitens*.

References:

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

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.*, **30**, 2725–2729.

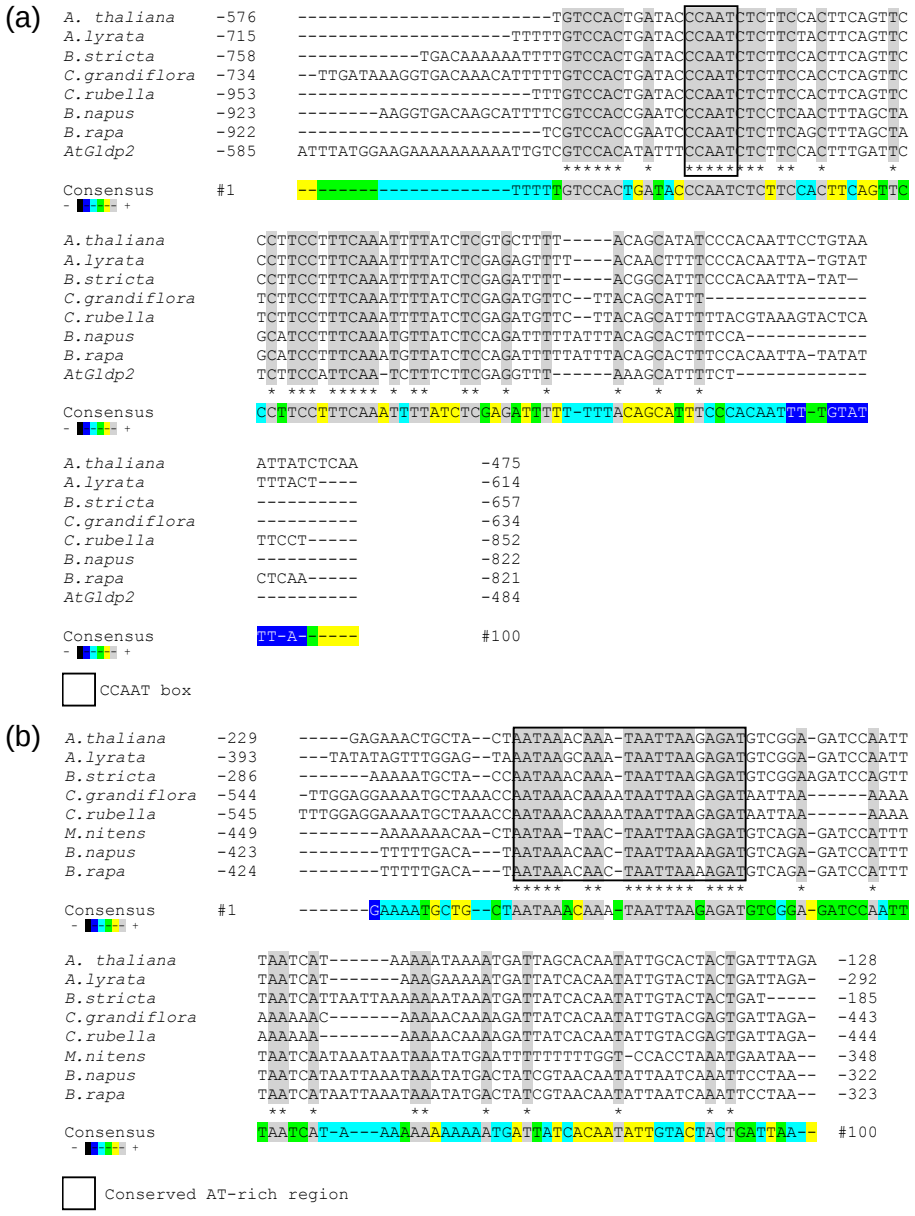


Figure S5: DNA sequence alignment of upstream sequences. Alignment of the top scoring window overlapping with the (a) M box, and (b) V box. Colors correspond to the degree of sequence conservation. Alignment was computed with CLUSTALW (Larkin *et al.*, 2007).

References:

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.

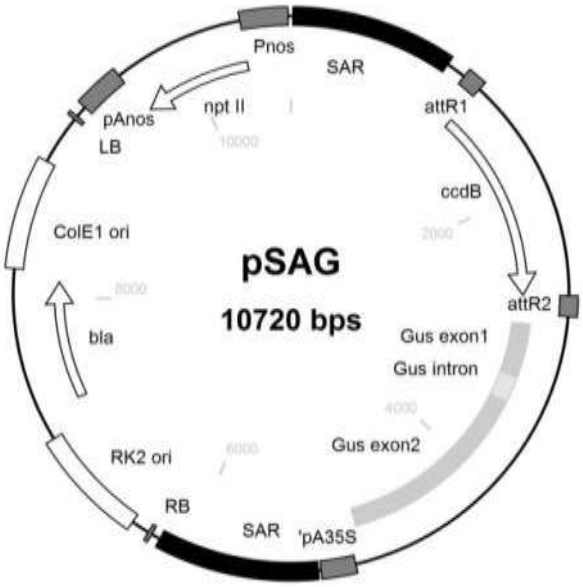


Figure S6: Map of the binary plasmid used for GUS assays. The vector contains scaffold attachment regions (SAR) that reduce variation caused by genome integration sites, recombination-based cloning of promoters, and a GUS gene with a plant intron that is inactive in bacteria. The intron also allows discrimination of pre-mRNAs synthesized from the promoter and mRNAs that might accumulate.

Table S1: Primers used in this study.

Primer name	Sequence (5' to 3')
4768_*Gldp*1_Promoter_rv	TGGGAAAAAAGGTTGCAGTC
6917_*Gldp*2_Promoter_rv	AACCAAACCAAAAAAAAACAAATGC
6930_*Gldp*1_P1_fw	CACCCTGCAACTTTTCACCAACCA
6931_*Gldp*1_P2_fw	CACCTATGTCCCATTAGAGGGGAA
6932_*Gldp*1_P3_fw	CACCTGTCCACTGATACCCAATCT
4767_*Gldp*1_P4_fw	CACCTCACTTTTCATTATTACTATTTG
6933_*Gldp*2_P1_fw	CACCGCTTATTGAGTTGAGTGGTTG
6934_*Gldp*2_P2_fw	CACCAGAGATGTGTATCCAGTTGCA
6935_*Gldp*2_P3_fw	CACCGTGAACCATAATAATCATAAGCTC
6937_*Gldp*2_P4_fw	CACCCCCACAACATCAAACTTCATG
7849_*Gldp*1-P3Δ1	CACCGCTTTTACAGCATATCCCACAA
7850_*Gldp*1-P3Δ2	CACCTCAAGGATTTGGTGCATAAATTC
7852_*Gldp*1-P3Δ3	CACCGAACCAAAAAAACTTAACATTG
7853_*Gldp*1-P3Δ4	CACCAAAACTTAACATTGATTAAACTT
8112_*Gldp*1_(-1142)_F	GGTATGCCTATCACTCTATTTGTTT
8113_*Gldp*1_(-1142)_R	TGCATACTTGCTTGTAAAAACTAATTG
8114_*Gldp*1_(-611)_F	GCATTCCGCCACCTAATTTATCA
8115_*Gldp*1_(-611)_R	AATTTGAAAGGAAGGGAACTGAAG T
8120_*Gldp*1_(+400)_F	GTCTCTTCCCTTTCCCCTTTC
8121_*Gldp*1_(+400)_R	GAGGGAAAGTATCGCTGGGTT
8172_*Gldp*1_(-888)_Fw	TTTTCTTGACGCCTTTTCGAGC
8173_*Gldp*1_(-888)_Rv	CTTAAAGATCGTGGAAAAGATTACT
8174_*Gldp*1_(-250)_Fw	GAAATTTGCGAACTAAAGACACTAAA
8175_*Gldp*1_(-250)_Rv	ATTTGTTTATTAGTAGCAGTTTCTCC
8176_*Gldp*1_(+147)_Fw	GGTCCGTGAGTGTATATTGTATAT
8177_*Gldp*1_(+147)_Rv	AAAACGAAGAGAGTAAGAGAGATG
8190_*Gldp*2_(-495)_Fw	GGAACAGAGGGAGGGAGTAT
8191_*Gldp*2_(-495)_RV	CATGAAACATTTGTATGTGACCTC
8687_M_59::P2_Mut.A	TTCAAATTTTATCTCGTTATGTCCCATTAGAGGGGAA
8735_M_59::P2_Mut.B	CCACTTCAGTTCCCTTCCTTTCAAATTTTA
8736_M_59::P2_Mut.C	ATACCCAATCTCTTCCACTTCAGTTCCCTTCCTT
8737_M_59::P2_Mut.D	CACCTGTCCACTGATACCCAATCTCTTCCACTTCAG
8800_35Smin_Template	GCAAGACCCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGACACGCTG
8802_35Smini_R	CAGCGTGTCCTCTCCAAATG
8803_35Smini_F	CACCGCAAGACCCTTCCTCTATA
8804_M_59:35Smin _Mut.1	ATTTTATCTCGTGCAAGACCCTTCCTCTATATAA
8805_M_59:35Smin _Mut.2	CCCTTCCTTTCAAATTTTATCTCGTGCAAGACCCTTCCTCTATATAA
4249_GAPDH_qPCR_FW	TTGGTGACAACAGGTCAAGCA
4250_GAPDH_qPCR_RV	AAACTTGTCGCTCAATGCAATC
4746_GLDP1_mRNA_FW	CATGCAATTGCTGATGCAGCTT

4748_GLDP1/2_mRNA_RV	CATCATCCAAGGTGGTTGTTTC
4747_GLDP2_mRNA_FW	GAGCCCTTATCTGACCCATCC
8112_*Gldp*1_-8112_Fw	GGTATGCCTATCACTCTATTTGTTT
8113_*Gldp*1_-8112_Rv	TGCATACTTGCTTGTAAAAACTAATTG
8114_*Gldp*1_-611_Fw	GCATTCCGCCACCTAATTTATCA
8115_*Gldp*1_-611_Rv	AATTTGAAAGGAAGGGAACTGAAGT
8174_*Gldp*1_-258_Fw	GAAATTTGCGAACTAAAGACACTAAA
8175_*Gldp*1_-258_Rv	ATTTGTTTATTAGTAGCAGTTTCTCC
8176_*Gldp*1_+90_Fw	GGTCCGTGAGTGTATATTGTATAT
8176_*Gldp*1_+90_Rv	AAAACGAAGAGAGTAAGAGAGATG
8172_*Gldp*1_-888_Fw	TTTTCTTGACGCCTTTTCGAGC
8173_*Gldp*1_-888_Rv	CTTAAAGATCGTGGAAAAGATTACT
8120_*Gldp*1_+400_Fw	GTCTCTTCCCTTTCCCCTTTC
8121_*Gldp*1_+400_Rv	GAGGGAAAGTATCGCTGGGTT
8196_*Gldp*2_+495_Fw	GCAACACCGGATGAACAAGC
8197_*Gldp*2_+495_Fw	TTAATCCTTCATCAAATATCCCGG
5112_GUSint_mRNA_Fw	GAAGCCGATGTCACGCCG
5113_GUSint_mRNA_Rv	TTGCCGTTTTCGTCGGTAATC
5115_Kanamycin_Std_Fw	CTGTCCGGTGCCCTGAATG
5116_Kanamycin_Std_Rv	CAACGTCGAGCACAGCTGC

Chapter III: Manuscript 1

Loss of the M-box from the glycine decarboxylase P-subunit promoter in C2 *Moricandia* species

Waly Adwy[1,3], Urte Schlüter[2], Jutta Papenbrock[1], Christoph Peterhänsel[1] and Sascha Offermann[1]

[1]Institute of Botany, Leibniz University Hannover, Herrenhäuser Str. 2, 30419 Hannover, Germany

[2]Institute of Plant Biochemistry, Heinrich Heine University, Universitätsstr. 1, 40225 Düsseldorf, Germany

[3]Department of Genetics, Faculty of Agriculture, Cairo University, Gamaa St. 13, 12619 Giza, Egypt

Journal:	Journal of Plant Research
Date of submission:	April 2018
Author contribution :	The overall idea for the experiments was developed by W.A., and C.P; the more detailed design of the experiment was done by W.A.; U.S. provided plant material and transcript sequence data; the experiments were conducted by W.A. as well as the data analysis; the manuscript was written by W.A., J.P., U.S. and S.O., and reviewed by all authors.
GenBank accessions:	KY386899, KY386900, and KY386901

Title:

Loss of the M-box from the glycine decarboxylase P-subunit promoter is frequent in C2 *Moricandia* species.

Authors

Waly Adwy[1,3], Urte Schlüter[2], Jutta Papenbrock[1], Christoph Peterhansel[1] and Sascha Offermann[1]

[1]Institute of Botany, Leibniz University Hannover, Herrenhäuser Str. 2, 30419 Hannover, Germany

[2]Institute of Plant Biochemistry, Heinrich Heine University, Universitätsstr. 1, 40225 Düsseldorf, Germany

[3]Department of Genetics, Faculty of Agriculture, Cairo University, Gamaa St. 13, 12619 Giza, Egypt

ABSTRACT

C2 photosynthesis operates by shuttling photorespiratory glycine (C2) from mesophyll (M) to bundle sheath (BS) cells, followed by decarboxylation and release of CO_2 around RubisCO. C2 plants are characterized by low apparent photorespiration and enhanced refixation of photorespiratory CO_2 and the C2 pathway is thought to represent an intermediate step for the evolution from C3 to C4 photosynthesis. Restriction of glycine decarboxylation to the BS cells is considered to be a prerequisite for C2 photosynthesis. In the C3 plant species *Arabidopsis thaliana*, a *cis*-element required for expression of the P-subunit of glycine decarboxylase (GDC-P) in M cells (termed the M-box) was previously identified in the promoter of *A. thaliana* glycine decarboxylase P-subunit 1 (*AtGldp*1). Consequently, the loss of this element restricted *Gldp*1 expression to the BS cells. To investigate conservation, *Gldp* promoter sequences from another C3 and two additional C2 *Moricandia* species were isolated by genome walking. In comparison to *AtGldp*1, the M-box was conserved in promoter from C3 *Moricandia moricandioides*, but was not found in the promoters of *M. nitens, M. suffruticosa, and M. arvensis*, indicating the loss of the M-box from several C2 *Moricandia* species. The *AtGldp*1 M-box was further analyzed in detail using promoter::GUS fusions. Results show that interaction between two promoter regions containing predicted CAAT and GATA elements are required for expression of the GUS reporter in M cells and these elements including their spacing are conserved in the promoters of different members of the *Brassicaceae.*

INTRODUCTION

Ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) plays a key role in photosynthetic carbon fixation. It catalyzes the fixation of CO_2 to the acceptor molecule ribulose-1,5-bisphosphate (RuBP) resulting in the production of the three carbon (C3) organic acid 3-phosphoglycerate (3-PGA). The efficiency of the carboxylation reaction is reduced when oxygen competes with CO_2 at the active site of RubisCO (Peterhänsel et al., 2010; Zelitch, 1992). RubisCO can then catalyze the oxygenation of RuBP resulting in the generation of one molecule 3-PGA and one molecule 2-phosphoglycolate (2-PG) (Bassham and Kirk, 1962). The accumulation of 2-PG is considered to be toxic due to inhibition of Calvin-Benson-Bassham (CBB) cycle enzymes (Anderson, 1970). 2-PG is detoxified via the photorespiratory pathway which also recycles most of the carbon locked in 2-PG back into 3-PGA for re-use in the CBB cycle (Leegood et al., 1995) under investment of additional energy (Bauwe et al., 2010; Raines, 2011).

One of the key reactions of the photorespiratory pathway is the decarboxylation and deamination of two molecules of glycine to one molecule of serine under release of CO_2 and NH_3. The reaction is catalyzed by mitochondrial glycine decarboxylase (GDC) which consists of four different subunits, named the P-, H-, L-, and the T-protein (Douce et al., 2001; Douce and Neuburger, 1989). In C3 plants, photosynthesis and the photorespiratory pathway take place within the same cell type (mesophyll, M cells). However, some plant species restrict glycine decarboxylation to bundle sheath (BS) cells (Rawsthorne, 1992). In these species, photorespiratory glycine (a C2 amino acid) is shuttled from the M to the neighboring BS cells where the decarboxylation reaction by the GDC complex takes place. This results in an increased CO_2 concentration in the BS which favors the carboxylation of RuBP by RubisCO. Consequently, C2 species have a lower CO_2 compensation point relative to C3 species (Schlüter et al., 2016; Monson and Rawthorne, 2000).

About 43 plant species distributed over 21 lineages carry out C2 photosynthesis (Sage et al., 2011). Restriction of GDC-P to BS cells is probably a common feature of C2 species but some might have lost additionally other members of the photorespiratory pathway such as GDC-H, GDC-T and the serine hydroxymethyltransferase (SHMT) from the M cells (Schulze et al., 2016; Morgen et al. 1993).

Also, the sequence of evolutionary events leading to the restriction of glycine decarboxylation to the BS cells was likely not uniform. For example, C3 *Flaveria* species have two copies of GDC-P, one copy that is exclusively expressed in BS cells and one copy that is expressed in all cell and tissue types. During evolution towards C2 and C4, the BS expressed copy became dominant whereas the globally expressed copy was deactivated (Schulze et al., 2013). In contrast, there are many C3 plants including monocots (Khoshravesh et al., 2016) and other C3 dicots such as the *Brassica* branch of the *Brassicaceae* with only a single GDC-P copy (Schlüter et al., 2016). Accordingly C2 and C4 evolved in these families without having multiple GDC-P copies.

A. thaliana is a model C3 plant in the *Brassicaceae* family. It has two mitochondria-localized GDC-P copies encoded by nuclear *AtGldp*1 and *AtGldp*2 but the relative contribution of the two *AtGldp* homologs to the total GDC-P protein pool is currently not clear. Since only the double knock-out of *AtGldp*1 and *AtGldp*2 is lethal, they are considered to be functionally redundant. Surprisingly, the double knock-out was lethal even under non-photorespiratory conditions emphasizing the importance of the GDC-P for basic C1 metabolism (Engel et al., 2007).

We have previously shown that the loss of a promoter element known as the M-box from the *AtGldp*1 promoter results in C2-like expression of a downstream fused *gusA* reporter gene. Furthermore, the M-box was found to be conserved in promoters of *Gldp* genes from C3 *Brassicaceae* species, but it was not identified in the promoter of *Moricandia nitens* (Adwy et al., 2015) which is one out of five C2 species in the genus *Moricandia* (Schlüter et al., 2016). In this study, we show that loss of the M-box is observed in several C2 *Moricandia* species. Furthermore we analyzed in detail the *cis*-elements in the M-box that are responsible for expression in M cells.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

A. thaliana Col-0 plants used for stable transformation were grown on one-half strength Murashige and Skoog (MS) agar medium supplemented with vitamins (Duchefa, RV Haarlem, Netherlands). *A. thaliana* plants (Col-0) used for transformation were grown for 8 weeks. Mutant seedlings were grown for two weeks on MS agar plates supplemented with vitamins and the antibiotic kanamycin at a concentration of (25 µg/ml). Plants were grown under controlled short day conditions (8 h light, ~150 µE m^{-2} s^{-1}/16 h dark, 22°C/21°C) in a Percival growth chamber (CLF Plant Climatics, Wertingen, Germany) at ambient CO_2 levels.

PCR and Cloning

A. thaliana DNA was extracted as described in Edwards et al. (1991). DNA from different *Moricandia* species was extracted using the CTAB method as described in Doyle (1991). Promoter::*gusA* fusions were generated as 5'-end mutations and directionally cloned into a pENTR™/D-TOPO® vector. Sequence-confirmed fragments were then transferred to a plant binary vector (pSAG; Adwy et al., 2015) using LR clonase® (Invitrogen, Waltham, USA, Germany). Standard PCR conditions were used for amplification of promoter constructs (denaturation at 98°C for 10 s, annealing at 60°C for 30 s, and extension at 72°C for 1 min). All cloned inserts were sequenced and compared to the TAIR database (https://www.arabidopsis.org/). All primers are listed in Table S1.

Plant Transformation and Screening

A. thaliana plants with 8-week-old inflorescences were used for *Agrobacterium*-mediated floral dip transformation as described in Clough and Bent (1998). Positively transformed plants were identified after seedling screening on MS-agar medium supplemented with vitamins and kanamycin (25 µg/ml).

GUS Staining

Histochemical GUS staining was analyzed in 2-week-old mutants by transferring seedlings to GUS staining solution and incubation overnight at 37°C. GUS-stained plants were de-stained by incubating

in 96% ethanol for another day before being mounted for microscopy. The GUS staining buffer recipe was: 50 mM NaCl, 100 mM, Tris-HCl (pH 7.0), 1 mM potassium ferricyanide ($K_3[Fe(CN)_6]$), 1 mM potassium ferrocyanide ($K_4[Fe(CN)_6]$), 0.1% (v/v) Triton X-100, 5% (v/v) 5-bromo-4-chloro-3-indolyl *β*-D-glucuronide (Jefferson et al., 1987).

Quantitative RT-PCR

RNA was extracted using the TRIzol method after Chomczynski and Sacchi (1987) from 2-week-old plants and then used for first strand cDNA synthesis using M-MLV reverse transcriptase (Promega, Madison, USA) according to the manufacturer's protocol. Quantitative RT-PCR was carried out using Platinum® SYBR® Green qPCR SuperMix (Invitrogen, Germany). A non-template PCR control was included to control for contamination with genomic DNA. Transcription level of *gusA* was normalized to the expression of the *npt*II gene. All primers are listed in Table S1.

Restriction Genome Walking

Moricandia moricandioides (line 04-0393-10-00, Botanical Garden Osnabrück), *M. arvensis* (line MOR1, IPK Gatersleben), *M. suffruticosa* (line 0105433, Royal Kew Gardens), and *M. nitens* (line 0209858, Royal Kew Gardens) were used to isolate promoters of *Gldp* genes (more information are available in Schlüter et al., 2016). Restriction libraries (RL) were generated using DNA from each species after being digested using different blunt end cutters (*Dra*I, *Eco*RV, *Pvu*II and *Stu*I; Promega, USA). Cleaned RL were further ligated to adapters using T4-ligase at 16°C overnight (adapters sequences are listed in Table S1). Primary PCR using gene specific primers (GSP) and adapter-based primers was performed using each ligation reaction as a template. Nested PCR was then carried out to obtain more specific products. DNA fragments amplified by PCR were cleaned and sequenced using GSP.

Conservation Analysis

Sequences for the promoters of *MmGldp*, *MaGldp* and *MsGldp* were determined by restriction genome walking (Suppl. Fig. 1). The *MnGldp* promoter sequence was identified previously by Zhang et al. (2004). Promoter alignments were performed by testing the probability of conservation of each

promoter relative to the *AtGldp*1 promoter. The EARS software (Picot et al., 2010) was used for alignments.

Data Analyses

Data were analyzed using R software (http://www.R-project.org) for statistical analysis. Student's *t*-test was based on a two-tailed hypothesis and non-paired observations.

ACCESSION NUMBERS

Accession numbers for promoter sequences of *Gldp* genes are as follows: *Arabidopsis lyrata* (*AlyGldp*1: sc7:3493441), *Arabidopsis thaliana* (*AtGldp*1:At4g33010, *AtGldp*2: At2g26080), *Boechera stricta* (*BstGldp*1: Bostr.7867s1152.1), *Brassica napus* (*BnGldp*: AY54471), *Brassica rapa* (*BraGldp*1: Brara.A00490.1), *Capsella grandiflora* (*CgrGldp*1: Cagra.4093s0031.1), *Capsella rubella* (*CrbGldp*1: Carubv10004049m), *Moricandia nitens* (*MnGldp*: AY544772).

Sequences for *Gldp* promoters from *Moricandia moricandioides* (*MmGldp*), *Moricandia arvensis* (*MaGldp*), and *Moricandia suffruticosa* (*MsGldp*) were identified by genome walking and were deposited in GenBank according to the following accession numbers KY386901, KY386899, and KY386900.

RESULTS

Lack of the M-box is a common feature in C2 *Moricandia Gldp* promoters

Previously, a 59 nucleotide long *cis*-element (M-box) required for mesophyll (M) specific expression was identified in the promoter of *A. thaliana Gldp*1. Sequence analysis indicated that this M-box is generally present in C3 *Brassicaceae* but it was not found in the C2 *Brassicaceae* species *M. nitens* (Adwy et al., 2015).

In order to test if the absence of the M-box is a common feature for C2 *Moricandia* species, additional promoter sequences of *Gldp* genes from the two C2 *Moricandia* species *M. suffruticosa* and *M. arvensis*, and for comparison from the C3 *Moricandia* species *M. moricandioides* were isolated by restriction genome walking (Suppl. Fig. 1). Conservation analysis indicated two highly conserved regions in the *M. moricandioides Gldp* promoter (Fig. 1) that overlap with the previously tested M-box and the element required for expression in BS and the vascular tissue (V-box) from the *AtGldp*1 promoter (Adwy et al., 2015). In contrast, the three C2 *Moricandia* species *M. nitens*, *M. suffruticosa*, and *M. arvensis* missed any predictions for the M-box region and only the V-box was conserved.

To correlate the computer predictions with *in vivo* function, we tested exemplarily the tissue expression pattern of a C3 and a C2 *Moricandia* promoter. *A. thaliana* plants transformed with the *Gldp* promoter from the C2 species *M. arvensis* showed no GUS expression in the M, thereby resembling the expression pattern observed for the *AtGldp*1 promoter lacking the M-box (P3Δ1) (Fig. 2A). In contrast, both the full length *AtGldp*1 promoter as well as the *Gldp* promoter from the C3 *Moricandia* species *M. moricandioides* showed GUS staining in all tissues.

Quantitatively, *gusA* mRNA accumulation in the *A. thaliana* lines carrying the C3 *M. moricandioides Gldp* promoter was similar to the endogenous full length *AtGldp*1 promoter, whereas mRNA accumulation in the line carrying the C2 promoter from *M. arvensis* was approximately four times lower (Fig. 2B). This is again similar to the *A. thaliana* mutant lacking the M-box (P3Δ1) and can probably be attributed to restriction of expression to the BS and vascular tissue only.

In summary, the conservation analysis indicates that the lack of the M-box is a common feature for the C2 *Moricandia Gldp* promoters tested. Furthermore, the presence of the M-box in the C3 *Moricandia Gldp* promoter correlates with expression in all leaf tissue types whereas M cell expression was not

observed when using the *Gldp* promoter from the C2 *Moricandia* species without the M-box.

Short motifs in both, the 5' and 3' region of the M-box, are required for expression of *AtGldp*1 in the mesophyll

In order to further define the sequences responsible for M-specific expression in the M-box, promoter deletions mutants were created and tested for their tissue-specific expression pattern. The M-box was initially divided into a 5' (M-5') and a 3' (M-3') segment (Fig. 3A). Fusing either of the two segments to the P2 minimal promoter was not sufficient to induce expression in M cells (Fig. 3B, C). This could either indicate that both, the 5' and 3' regions, are required for M cell expression or alternatively that a critical element was disrupted. To discriminate between these possibilities, another deletion mutant encompassing the cut site and an additional two nucleotides to the left and five nucleotides to the right (M-5'/3'Δn) was created. M cell expression was not disturbed in this mutant indicating that indeed elements in both fragments are required for M expression.

To further narrow down the exact sequences required for M cell expression, several deletions from the 5' end were created. Deleting the first four (M-5'Δ1::P2), eight (M-5'Δ2::P2) or 12 nucleotides (M-5'Δ3::P2) from the 5' end reduced slightly the expression strength in M cells but did not abolish it. In contrast, removing the next six nucleotides resulted in complete loss of M cell expression (M-5'Δ4::P2). Removing the following five nucleotides did not change the expression pattern any further (M-5'Δ5::P2). At the 3' end of the M-box, deleting the last 25 nucleotides (M-3' Δ1::P2) or the last 10 nucleotides (M-3' Δ2::P2) resulted in a complete loss of M cell expression.

The identified sequences in the 5' (M-5'Δ4) and 3' (M-3'Δ2) region of the M-box required for M cell expression were analyzed for conservation in other C3 *Brassicaceae* species and for the presence of potential transcription factor binding sites (Fig. 4). Alignment of the *AtGldp*1 M-box with the homologous sequences of eight other C3 *Brassicaceae* species indicated four conserved regions including a fully conserved CAAT box and a GATA box in reverse orientation exactly in the functionally important M-5'Δ4 and M-3'Δ2 regions. The sequence between the predicted CAAT and the GATA boxes was not fully conserved between the different species. However, the spacing was exactly 33 nucleotides in all cases.

In summary, the data shows that sequence elements in both the 5' and 3' regions are required for M cell expression but neither one is sufficient. Furthermore, CAAT and GATA transcription factor binding sites are predicted in the sequence elements required for M cell expression and they are conserved between different C3 *Brassicaceae* species.

DISCUSSION

The M-box has been characterized previously in *A. thaliana* as an element required for expression of the *Gldp*1 gene in the M-cells. Consequently, deleting the M-box resulted in a C2-like expression pattern of the fused reporter in BS cells and the vasculature only (Adwy et al., 2015). Here we show that the M-box is also present in the promoters of the C3 *Moricandia* species *M. moricandioides*, but not in any of the three tested promoters of the C2 *Moricandia* species *M. nitens*, *M. arvensis* and *M. suffruticosa* (Fig. 1). This poses the question, if lack of the M-box is a common feature for C2 *Moricandia* species and, possibly, even for all C2 *Brassicaceae* species.

The *Brassicaceae* family comprises more than 4000 species with the majority being C3 species and only five appearances of C2 photosynthesis in the genus *Moricandia* and three C2 species in the genus *Diplotaxis* (Apel et al., 1997; Sage et al., 2014). Conserved regions in the 5'-upstream flanking regions of *Gldp* genes from the C3 *Brassicaceae* species *B. napus* and *A. thaliana* have been reported before (Zhang et al., 2004). Reporter-GUS-fusions later identified a 135 bp 5'-upstream flanking region required for mesophyll expression in the promoter of *B. napus Gldp* (Zhang et al., 2011). Interestingly, this region encompasses a sequence homolog to the M-box sequences described in our previous study (Adwy et al., 2015) and also including the CAAT and GATA boxes described in this study (Fig. 4), thereby providing further evidence for the importance of the M-box for expression in the mesophyll. However, the same study also observed the M-box in the promoters of *Gldp* genes from the two C2 species *Moricandia spinosa* and *Diplotaxis tenuifolia* which were not included in this

study. This could indicate that either other unknown *cis*-elements are missing in the promoters of these species or, alternatively, not all C2 members of the *Brassicaceae* restrict GDC expression to the BS cells and vasculature via loss of the M-box. While GDC restriction to BS cells is commonly achieved by the loss of the GDC-P expression from the mesophyll compartment (Rawsthorne, 1992), it could theoretically also be achieved by restricting the expression of other subunits of the GDC complex such as the GDC-H, GDC-T, GDC-L subunits or even SHMT (Morgen et al., 1993).

In summary, there is currently no clear evidence that the loss of the M-box is a strictly conserved mechanism for achieving a C2-like spatial expression pattern of *Gldp* in all C2 *Brassicaceae* species, however, it is observed in several *Moricandia* species. Conversely, the M-box was found to be present in all C3 *Brassicaceae* species tested, including C3 *M. moricandioides* (Fig. 4). It needs to be mentioned that in the study by Zhang et al. (2011) also a 5'-upstream flanking regions of the *Gldp* gene from *M. moricandioides* was identified via restriction genome walking. In contrast to our results, this sequence does not seem to contain the M-box and the putative CAAT and GATA transcription factor binding sites. However, the reported sequence by Zhang et al. (2011) encompasses only the region between -900 bp and -705 bp. In our study, we were able to extend this up to the -1039 bp position and the M-box including the CAAT and GATA transcription factor binding sites were clearly identified in this region. It would be interesting to make a more detailed comparison of all promoter features of C2 and C3 *Brassicaceae* GLDP promoters identified in this and the previous studies. However, besides the regions homolog to the 135 bp region required for mesophyll expression in *B. napus*, sequence information was not made publically available by Zhang et al. (2011) which prevents further comparisons, for example by promoter conservation analysis using the EARS algorithm (Fig.1). In conclusion, there is evidence that the M-box is present in all C3 *Brassicaceae* species analyzed so far, including *M. moricandioides*.

As we have demonstrated, loss of expression of *Gldp* in the M cells can easily be achieved by mutation or loss of small defined *cis*-elements containing predictions for CAAT and GATA boxes (Fig. 3). In plants, complex formation of different CAAT box binding factors (nuclear factors Y, NF-Ys) has been associated with diverse functions including embryo development, flowering time control, drought stress and nodule development (Laloum et al., 2012). Promoter regulation driven by CAAT and GATA boxes in plants is also not unique to *AtGldp*1. The phytochrome response of the light-harvesting chlorophyll a/b protein (*Cab*/*Lhc*) encoding gene in *Lemna gibba* can be disturbed by mutation of either two promoter regions containing CAAT and GATA motifs (Kehoe et al., 1994), and these motifs are conserved in *Cab* promoters from wheat (Nagy et al., 1987), *A. thaliana* (Leutwiler et al., 1986; Karlin-Neumann et al., 1988; Sun et al., 1993), pea (Cashmore, 1984; Simpson et al. 1986; White et al., 1992) and maize (Sullivan et al., 1989). In other studies, deletion of CAAT and GATA sequences have resulted in general reduction of transcript levels implying that CAAT and GATA regulation is not only limited in conferring phytochrome responses (Gidoni et al., 1989). It will be interesting to see what factors exactly bind at the predicted CAAT and GATA motifs of the *AtGldp*1 promoter and if their presence is conserved in other promotors regulating gene expression of genes involved in photorespiration as well**.**

Acknowledgements:

This work was funded by the Deutsche Forschungsgemeinschaft (PE819/7-1).

REFERENCES

Adwy W, Laxa M, Peterhansel C. A simple mechanism for the establishment of C2-specific gene expression in *Brassicaceae*. Plant J. 2015;84: 1231–1238.

Anderson L. Chloroplast and cytoplasmic enzymes. II. Pea leaf triose phosphate isomerases. Biochim Biophys Acta. 1970;235**:** 237–244.

Apel P, Horstmann M, Pfeffer M. The Moricandia syndrome in species of the Brassicaceae - evolutionary aspects. Photosynthetica. 1997;33: 205–215.

Bassham J, Kirk M. The effect of oxygen on the reduction of CO_2 to glycolic acid and other products during photosynthesis by *chlorella*. Biochem Biophys Res Commun. 1962;9: 376-380.

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

Cashmore A. Structure and expression of a pea nuclear gene encoding a chlorophyll a/b-binding polypeptide. PNAS. 1984;81: 2960–2964.

Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem.. 1987;162: 156–159.

Clough SJ, Bent AF. Floral dip: a simplified method for *Agrobacterium* mediated transformation of *Arabidopsis thaliana*. Plant J. 1998;16**:** 735–743.

Douce R, Neuburger M. The uniqueness of plant mitochondria. Annu Rev Plant Phys. 1989;40**:** 371–414.

Douce R, Bourguingnon J, Neuburger M, Rebeille F. The glycine decarboxylase system: a fascinating complex. Trends Plant Sci. 2001;6: 167–176.

Doyle J. DNA Protocols for Plants. In Hewitt GM, Johnston AWB, Young JPW, editors. Molecular techniques in taxonomy. Springer, Berlin Heidelberg. 1991; pp. 283–293.

297 Edwards K, Johnstone C, Thompson C. A simple and rapid method for the preparation of plant
298 genomic DNA for PCR analysis. Nucleic Acids Res. 1991;19: 1349.

299 Engel N, van den Daele K, Kolukisaoglu U, Morgenthal K, Weckwerth W, Pärnik T, Keerberg O,
300 Bauwe H. Deletion of glycine decarboxylase in *Arabidopsis* is lethal under non-photorespiratory
301 conditions. Plant Physiol. 2007;144: 1328–1335.

302 Gidoni D, Brosio P, Bond-Nutter D, Bedbrook J, Dunsmuir P. Novel *cis*-acting elements in petunia
303 *Cab* gene promoters. Mol Gen Genet. 1989;215: 337–344.

304 Jefferson RA, Kavanagh TA, Bevan MW. GUS fusions: *ß*-glucuronidase as a sensitive and versatile
305 gene fusion marker in higher plants. EMBO J. 1987;6: 3901–3907.

306 Karlin-Neumann GA, Sun, L, Tobin EM. Expression of light-harvesting chlorophyll a/b-protein genes
307 is phytochrome-regulated in etiolated *Arabidopsis thaliana* seedlings. Plant Physiol. 1988;88: 1323-
308 1331.

309 Kehoe DM, Degenhardt J, Winicov I, Tobin EM. Two 10-bp regions are critical for phytochrome
310 regulation of a *Lemna gibba Lhcb* gene promoter. Plant Cell. 1994;6: 1123-1134.

311 Khoshravesh R, Stinson CR, Stata M, Busch FA, Sage RF, Ludwig M, Sage TL. C3-C4 intermediacy
312 in grasses: organelle enrichment and distribution, glycine decarboxylase expression, and the rise of C2
313 photosynthesis. J Exp Bot. 2016;67: 3065–3078.

314 Laloum T, Mita S, Gamas P, Baudin M, Niebel A. CCAAT-box binding transcription factors in plants:
315 Y so many? Trends Plant Sci. 2013;18: 157–166.

316 Leegood R, Lea P, Adcock M, Häusler R. The regulation and control of photorespiration. J Exp Bot.
317 1995;46: 1397–1414.

318 Leutwiler LS, Meyerowitz EM, Tobin EM. Structure and expression of three light-harvesting
319 chlorophyll a/b-binding protein genes in *Arabidopsis thaliana*. Nucleic Acids Res. 1986;14: 4051–
320 4064.

Monson RK, Rawsthorne S. CO_2 Assimilation in C3-C4 intermediate plants. In Leegood RC, Sharkey TD, Caemmerer S, editors. Photosynthesis: Physiology and Metabolism. Springer Netherlands, Dordrecht. 2000; pp. 533–550.

Morgan CL, Turner SR, Rawsthorne S. 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. 1993;190: 468–473.

Nagy F, Boutry M, Hsu MY, Wong M, Chua N-H. The 5' proximal region of the wheat *Cab*-7 gene contains a 268-bp enhancer-like sequence for phytochrome response. EMBO J. 1987;6: 2537-2542.

Orgen WL. Photorespiration: Pathways, Regulation, and Modification. Plant Physiol. 1984;35: 415–442.

Peterhansel C, Horst I, Niessen M, Blume C, Kebeish R, Kürkcüoglu S, Kreuzaler F. Photorespiration. Arabidopsis Book. 2010;8: e0130.

Picot E, Krusche P, Tiskin A, Carré I, Ott S. Evolutionary analysis of regulatory sequences (EARS) in plants. Plant J. 2010;64: 165–176.

Raines CA. Increasing photosynthetic carbon assimilation in C3 plants to improve crop yield: current and future strategies. Plant Physiol. 2011;155: 36–42.

Rawsthorne S. C3-C4 intermediate photosynthesis: linking physiology to gene expression. Plant Cell. 1992;2: 267–274.

Sage RF, Christin PA, Edwards EJ. The C4 plant lineages of planet Earth. J Exp Bot. 2011;62: 3155–3169.

Sage RF, Khoshravesh R, Sage TL. From proto-Kranz to C4 Kranz: building the bridge to C4 photosynthesis. J Exp Bot. 2014;65: 3341–3356.

Schlüter U, Bräutigam A, Gowik U, Melzer M, Christin PA, Kurz S, Mettler-Altmann T, Weber AP. Photosynthesis in C3–C4 intermediate *Moricandia* species. J Exp Bot. 2016;68: 191-206.

Schulze S, Mallmann J, Burscheidt J, Koczor M, Streubel M, Bauwe H, Gowik U, Westhoff P. Evolution of C4 photosynthesis in the genus *Flaveria*: establishment of a photorespiratory CO_2 pump. Plant Cell. 2013;25: 2522–2535.

Schulze S, Westhoff P, Gowik U. Glycine decarboxylase in C3, C4 and C3–C4 intermediate species. Curr Opin Plant Biol. 2016;31: 29–35.

Simpson J, Montagu MV, Herrera-Estrella L. Photosynthesis-associated gene families: differences in response to tissue-specific and environmental factors. Science. 1986;233: 34–38.

Sullivan TD, Christensen AH, Quail PH. Isolation and characterization of a maize chlorophyll a/b binding protein gene that produces high levels of mRNA in the dark. Mol Gen Genet. 1989;215: 431–440.

Sun L, Doxsee RA, Harel E, Tobin EM. CA-1, a novel phosphoprotein, interacts with the promoter of the *cab*140 gene in *Arabidopsis* and is undetectable in det1 mutant seedlings. Plant Cell. 1993;5: 109–121.

White MJ, Fristensky BW, Falconet D, Childs LC, Watson JC, Alexander L, Roe BA, Thompson WF. Expression of the chlorophyll-a/b-protein multigene family in pea (*Pisum sativum* L.). Planta. 1992;188: 190-198.

Zhang C, Xu G, Huang R, Chen C, Meng J. A dominant *gdcP*-specific marker derived from *Moricandia nitens* used for introducing the C3-C4 character from *M. nitens* into *Brassica* crops. Plant Breeding. 2004;123: 438–443.

Zhang C, Park IA, Xu F, Li M, Lim YP, Meng J. Transcriptionally and phylogenetically analyzing the P protein gene of glycine decarboxylase for understanding the evolution of C3–C4 species in Brassicaceae. Hortic Environ Biotechnol. 2011;52: 427–434.

Zelitch I. Control of plant productivity by regulation of photorespiration. BioScience. 1992;42: 510–516.

FIGURES

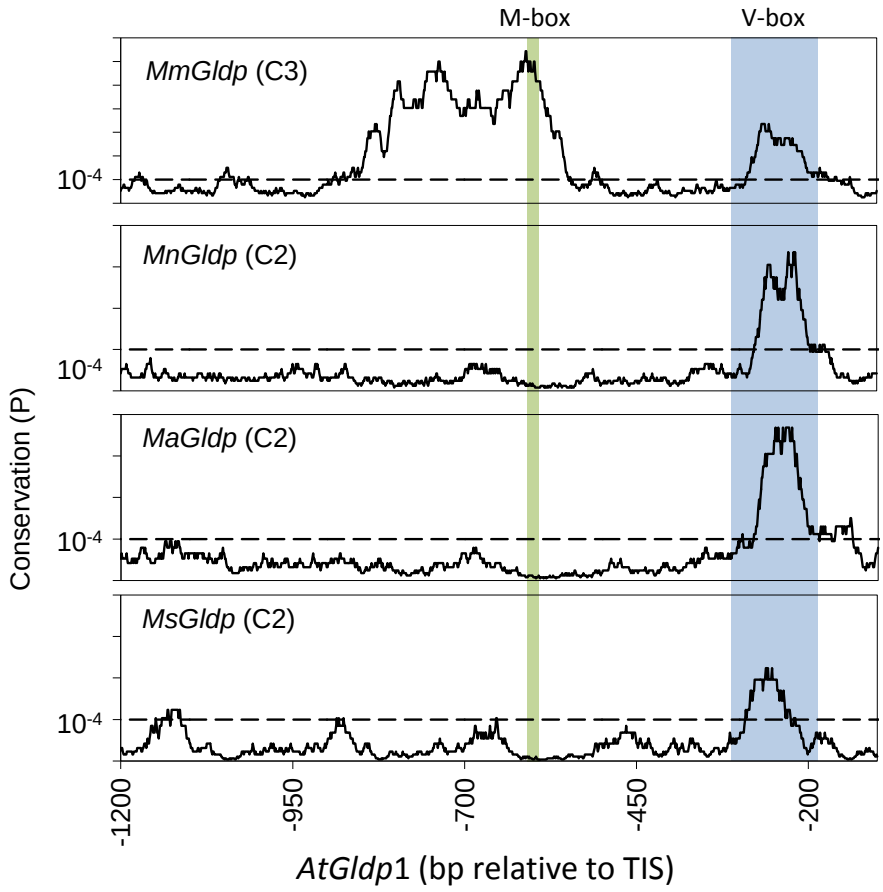


Fig. 1. Conservation analysis for *Gldp* promoters of the C3 *Moricandia* species *Moricandia moricandioides* (*MmGldp*) and the three C2 *Moricandia* species *Moricandia nitens* (*MnGldp*), *Moricandia arvensis* (*MaGldp*), and *Moricandia suffruticosa* (*MsGldp*). The *Gldp*1 promoter from *A. thaliana* was used as the reference sequence. Promoter regions that have been associated with expression in M (M-box), BS and the vasculature (V-box) as defined in Adwy et al. (2015) are indicated.

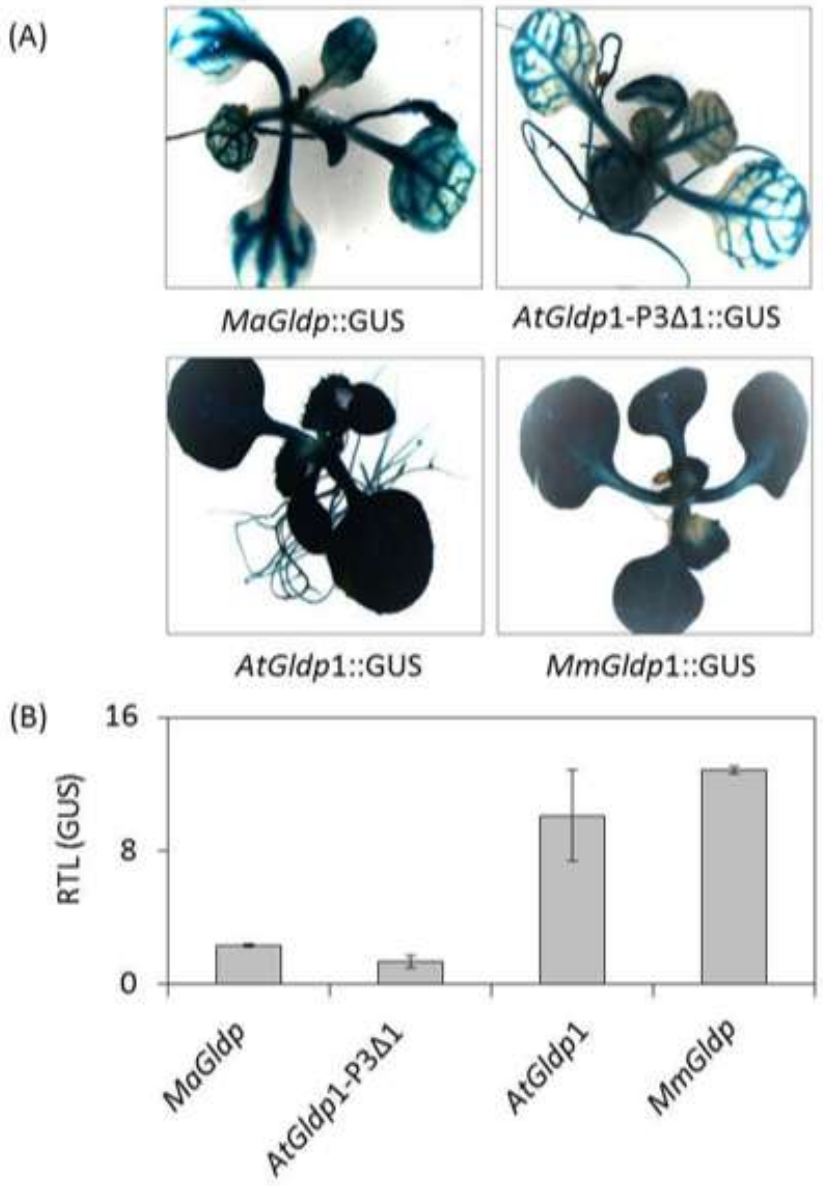


Fig. 2. (A) Histochemical GUS staining in 2-week-old Col-0 mutant seedlings stably transformed with promoter::*gusA* constructs from *Gldp* genes of *Moricandia arvensis* (Ma), the *Gldp*1 promoter fragment from *A. thaliana* lacking the M-box (*AtGldp*1-P3Δ1), *A. thaliana* (At) and *Moricandia moricandioides* (Mm). Pictures are representative for at least four independent transformation events. **(B)** Relative transcription levels (RTL) of promoter::*gusA* fusions are shown in Fig. 1A (n=4 ± SEM).

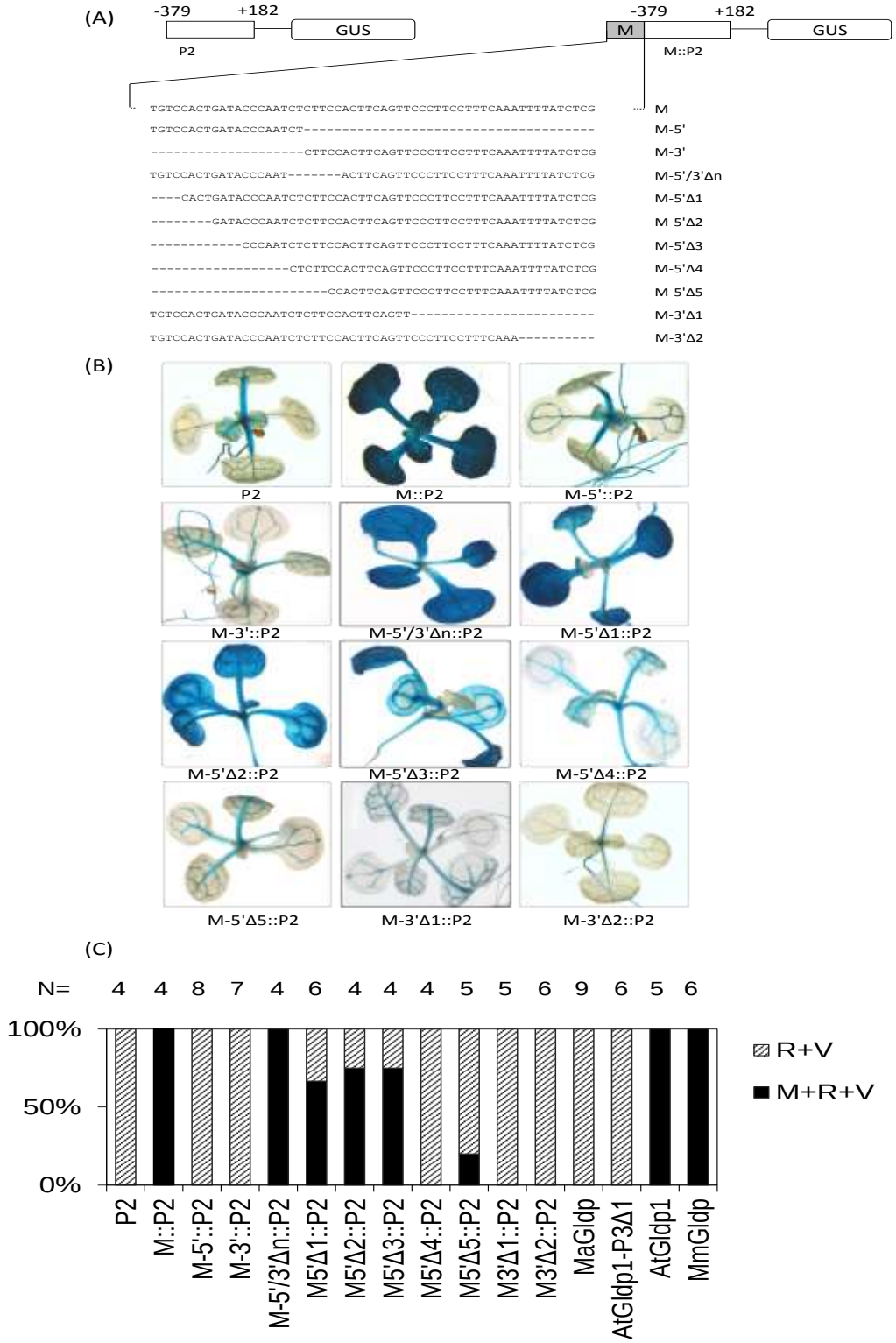


Fig. 3. Analysis of sequence elements required for M expression in the M-box. **(A)** Schematic illustrating of various deletion mutants. P2: minimal promoter that drives GUS expression in the vasculature and in BS cells but not in M cells (Adwy et al., 2015); M::P2: M-box fused to the P2 promoter; numbers indicate position relative to the transcription initiation site (TIS). **(B)** Histochemical GUS staining of 2-week-old stable mutants. Pictures are representative for at least three independent transformation events. **(C)** Frequency of spatial GUS

expression patterns in different independent transformation events. N: number of GUS stained plants from independent transformation events, R: root, V: vasculature, M: mesophyll.

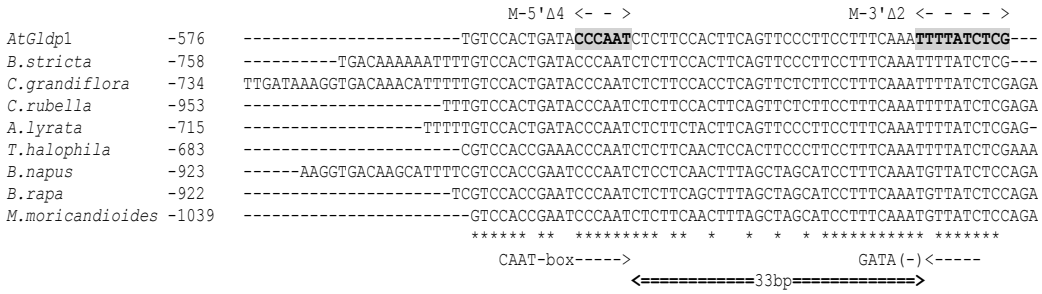
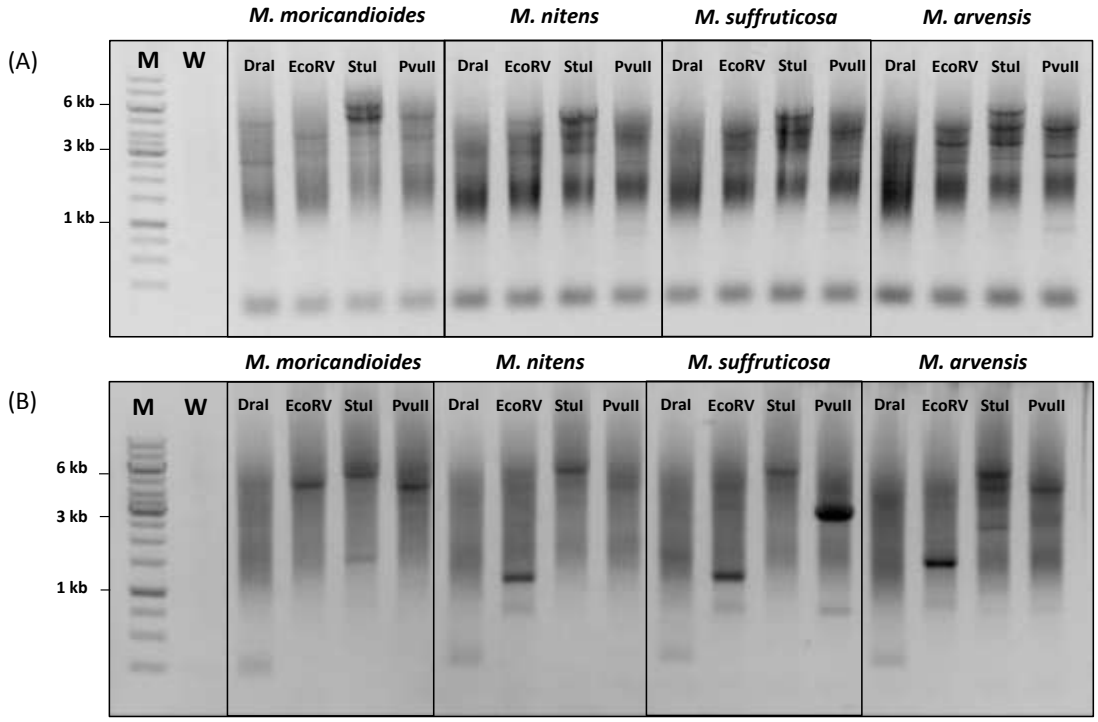


Fig. 4 DNA sequence alignment for M-box promoter regions of different C3 *Brassicaceae* species. Predicted CAAT- and GATA-binding sites are indicated. Arrows indicate the orientation relative to the coding strand. The conserved 33 nt spacing between the CAAT- and GATA-boxes is indicated with a double pointed bold arrow. Asterisks indicate nucleotides conserved in all species. Numbers indicate position of the sequence relative to the TIS.

SUPPLEMENTAL INFORMATION



Suppl. Fig. 1. Restriction genome walking (RGW) for *Gldp* promoters from different *Moricandia* species. **(A)** PCR using different RGW-libraries from DNA isolated from each species as a template M: 1Kb DNA ladder, W: water control. Primers were adapter-based (9661) and gene-specific primer (GSP1) (9607). **(B)** Nested PCR reaction using adapter-based primer (9662) and GSP2 (10061). All primers are listed in Suppl. Table 1. All GSPs are universal to the tested species and based on conservation of the first exon.

Supplemental Table 1. Primers and oligonucleotides used in this study.

Primer code	**Primer sequence (5' ==> 3')**
Primers for M-box mutations:	
9529-Mbox1	TCTTTATGTCCCATTAGAGGGGAA
9530-Mbox2	CCAATCTCTTTATGTCCCATTAGA
9531-Mbox3	ACTGATACCCAATCTCTTTATGTCCCATT
9532-Mbox4	CACCTGTCCACTGATACCCAATCTCTTTATGTC
9533-Mbox5	ATACCCAATCCACTTCAGTTCCCTTCCTT
9534-Mbox6	CACCTGTCCACTGATACCCAATCCACTTCA
9141-Mbox7	CACCCCACTTCAGTTCCCTTCCTTTCAAATTTTA
9675-Mbox8	GTTCCCTTCCTTTCAATATGTCCCATTA
9676- Mbox9	CCACTTCAGTTCCCTTCCTTTCAATATGTCCCATTA
9677- Mbox10	AATCTCTTCCACTTCAGTTCCCTTCCTTTCAATATGTCCCATTA
9678- Mbox11	GATACCCAATCTCTTCCACTTCAGTTCC
9679- Mbox12	CACCTGTCCACTGATACCCAATCTCTTCCACTTC
9306_M(-572)_F	CACCCACTGATACCCAATCTCTTC
9307_M(-568)_F	CACCGATACCCAATCTCTTCCACTT
9308_M(-558)_F	CACCCCCAATCTCTTCCACTTCAGT
9309_M(-553)_F	CACCCTCTTCCACTTCAGTTCCCTT
9310_M(-547)_F	CACCCCACTTCAGTTCCCTTCCT
Primers for qPCR	
5112_GUSint_mRNA_fw	GAAGCCGATGTCACGCCG
5113_GUSint_mRNA_rev	TTGCCGTTTTCGTCGGTAATC
5115_Kanamycin_fw	CTGTCCGGTGCCCTGAATG
5116_Kanamycin_rv	CAACGTCGAGCACAGCTGC
Primers for restriction genome walking	
9661_Ad1	GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGT
9662_Ad2	ACCAGCCC-NH2
9663_AP1-fw	GTAATACGACTCACTATAGGGC
9664_AP2-fw	ACTATAGGGCACGCGTGGT
10061_nPCR-univ_GSP2_Rv	TAAGCAAGCCTACGTGCGCGC
9607_GSP1_cons-Rv	CCATGAACGCGTTCGGCAATA
Primers for cloning the isolated promoters	
10093_MaGldp_F	CACCGGTGAGTTCTACCATTTTAGC
10081_MaGldp_R	TGGTTGGTGAAGAGGTTGTA
10069_MmGldp_F	CACCTTGGAGGACGATGGTTCCCA
10070_MmGldp_R	GGTGATGAGAGATTGGAAT

Chapter IV: Manuscript 2

The M-box binding factor: a possible role for GATA transcription factors in regulating the mesophyll expression of the *AtGldp*1 promoter

Waly Adwy[1,2], Christoph Peterhänsel[1], and Jutta Papenbrock[1]

[1]Institute of Botany, Leibniz University Hannover, Herrenhäuser Str. 2, 30419 Hannover, Germany

[2]Department of Genetics, Faculty of Agriculture, Cairo University, Gamaa St. 13, 12619 Giza, Egypt

Contribution

The overall idea for the experiments was developed by W.A. and C.P.; the more detailed design of the experiment was done by W.A.; the experiments were conducted by W.A. as well as the data analysis; the manuscript was written by W.A., and J.P., and reviewed by all authors

The M-box binding factor: a possible role for GATA transcription factors in regulating the mesophyll expression of the *AtGldp*1 promoter

ABSTRACT

Unlike C3 plants, C2 plants are able to utilize the photorespiratory glycine specifically in their bundle sheath (BS) cells. This gives C2 plants an advantage over C3 plants under photorespiratory conditions to lower the CO_2 losses resulting from photorespiration and enhancing the probability of its refixation. In *Brassicaceae*, restriction of glycine decarboxylase (GDC) in BS of C2 leaves has been achieved by the loss of a promoter *cis* element known as the M-box that is responsible for mesophyll (M) expression of the gene encoding the P-subunit of the GDC. This offers a mechanism for restricting the GDC to the BS of C2 plants. The restriction of GDC to BS is crucial for the C2 pathway. Theoretically, a knock-out mutation in an M-box binding factor could simulate one of the earliest steps of evolution of photosynthesis. In this study, a yeast-one-hybrid screen was carried out to identify factors that can bind to the M-box in yeast. Analysis points out to a transcription factor (TF) that could bind to the M-box region responsible for its M expression. The M-box binding factor belongs to the GATA-TF family and is commonly referred to as *Gata*5. A database search for available microarray data of this TF showed that *Gata*5 is preferentially expressed in shoots. Transcription of this factor increases by light, and this correlates with the light induction property that we tested in parallel for the M-box. A knock-out mutation in the *Gata5* locus was tested, and we show that it has no significant effect on the endogenous expression of *AtGldp*1. Available microarray data show preferential expression of three out of 30 GATA factors in shoot tissues suggesting possible redundancy in regulating the M expression of the M-box.

Keywords: GDC, GATA transcription factor, M-box, C2 photosynthesis.

INTRODUCTION

Assimilation of CO_2 in higher plants depends on the RubisCO enzyme, ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO), which can fix atmospheric CO_2 to sugar precursors. This reaction is highly influenced by the presence of oxygen in the vicinity of RubisCO, where oxygen competes with CO_2 at the active site of RubisCO (Zeltich, 1992). In a C3 photosynthetic pathway, oxygen uptake by RubisCO can occur instead of CO_2. The uptake of oxygen by RubisCO results in the formation of 2-phosphoglycolate (2-PG) (Orgen, 1984). This 2-PG compound is toxic to the plant and needs to be recycled by the pathway of photorespiration. Partial loss of the captured atmospheric CO_2 takes place during the recycling process of 2-PG, and this is not beneficial to plants (Anderson, 1971; Leegood et al., 1995).

There are many plant species that have developed other variants of photosynthesis in order to overcome the CO_2 losses from photorespiration. One of these variants is referred to as C2 photosynthesis. C2 photosynthesis is achieved on molecular level by restricting the expression of a photorespiratory enzyme known as the glycine decarboxylase (GDC) to the bundle sheath of C2 plants. C2 plants gain a metabolic advantage over those plants which perform C3 photosynthesis by molecular restriction of GDC to the BS, where CO_2 concentration takes place in the form of intermediate glycine (C2) and is utilized in the BS compartment of C2 plants. The released CO_2 byproduct from glycine utilization in the BS of C2 plants is trapped in the BS by the above existing mesophyll (M) tissue, subsequently have higher probability of being re-fixed by lasting for longer time in the vicinity of RubisCO expressed in the BS of these plants. In contrast, partial loss of the captured CO_2 would occur as a result of direct utilized of glycine in the M of C3 plants (Keeberg et al., 2014; Schulze et al., 2013).

We recently showed that such restriction of GDC to BS is possible in *Arabidopsis thaliana* which is a model C3 plant. This mechanism is simply by losing a promoter *cis* element which is necessary for M expression of the gene encoding the P-subunit of GDC referred to as the *AtGldp*1 (Adwy et al., 2015). Moreover, we showed that this M-box has two necessary *cis* elements that work together for conferring M expression, these are two short promoter stretches having CAAT and GATA boxes predictions (Adwy et al., submitted).

The sequence CAAT is generally found in most eukaryotic promoters and is required for general gene expression (Liberati et al., 1998a; Mantovani, 1999). It could be also found in promoters that have a regulated gene expression pattern at different developmental stages

or in different tissues (Ronchi et al., 1996). Out of 1500 genes encoding the transcription factors (TFs) regulating the expression of the genome of *A. thaliana* (Reichmann et al., 2000), there are 36 genes encoding a complex factor that can recognize the sequence CAAT, this factor is known as nuclear factor Y (NF-Y) (Siefers et al., 2009). This NF-Y complex has three subunits, NF-YA, NF-YB and NF-YC, and all subunits are required for DNA-binding (Sinha et al., 1995). Transcriptional activation domains are located on the N-terminal of the NF-YA and the C-terminal of the NF-YC (Li et al., 1992a). Moreover, the amino acid sequence of NF-Y factor has a histone fold motif (HFM) that enables specific binding to DNA on the level of histone via hydrophobic interaction, and probably regulates chromatin structure in a sequence specific manner (Arents and Moudrianakis, 1995; Luger et al., 1997).

The GATA sequence is found in many light-induced promoters. This includes the promoter of the light-harvesting chlorophyll a/b-protein genes (*Lhc* genes) (Kehoe et al., 1994), and the promoters of genes encoding the small subunit of RubisCO (*rbcS*) (Rolfe and Tobin, 1991). It has been shown that the light-perception property of the GATA *cis* element is dependent on the presence of a nearby CAAT box (Kehoe et al., 1994). Moreover, the GATA sequence in promoters seems also to regulate general expression strength regardless to light-perception or tissue-specific expression (Gidoni et al., 1989). Information about protein binding factors that could interact with the GATA sequence are limited in literature. A protein binding factor to a GATA motif that is found in the *Cab* promoter was described before (Lam and Chua, 1989). This GATA-binding factor was referred to as activating-sequence factor-2 (ASF-2). Later studies by Reyes et al. (2004) showed 30 genes encoding GATA-binding factors; however, the identity of ASF-2 is still not identified.

In this study, we screened an *A. thaliana* transcription factor cDNA library (Mitsuda et al., 2010) for a binding factor that could interact with the identified GATA element previously identified on the M-box, and shown to be regulating specific expression in the mesophyll (Adwy et al., 2015; Adwy et al., submitted). We addressed the light induction properties of the M-box as well as the binding factors. We studied the *Gldp*1 expression in a knock-out insertion line for the identified protein factor.

MATERIALS AND METHODS

Plant material

Arabidopsis thaliana Col-0 was used for light induction experiments and comparing light-induced expression patterns of different genes. Homozygous T-DNA insertion line SALK_134217-1-5 from the SALK collection (Sessions et al., 2002) was screened. Plants were grown under short day conditions 8 h light,~150 µE m^{-2} s^{-1}, 22°C), followed by 16 h dark (21°C) for two weeks. For light induction, 2-week-old plants were kept in dark conditions for 3 d (21°C) then exposed to light (~150 µE m^{-2} s^{-1}, 21°C).

Cloning and PCR

The vector pHIS2.1 was used to clone bait constructs for yeast-one-hybrid screening (Takara Bio Europe S.A.S. Saint-Germain-en-Laye, France), all inserts were designed as oligomers having *Eco*RI and *Spe*I sites on their 5' and 3' ends respectively (Eurofins, Hamburg, Germany). Inserts were double digested using *EcoR*I and *Spe*I (Thermo Scientific, Dreieich, Germany), and cloned using T4 ligase (16°C, overnight) after mixing with a previously linearized pHIS2.1 using *Eco*RI/*Spe*I digestion. Colony PCR screening was done using standard PCR protocol (denaturation at 98°C for 10 s, annealing at 60°C for 30 s, and extension at 72°C for 1 min), DreamTaq polymerase was used for amplification (Thermo Scientific). Primers used for pHIS2.1 screening are found in Table S3. Colony PCR in yeast was done using the same PCR protocol after incubating water suspended yeast cells with Lyticase (zymolyase) enzyme for 5 min at 37°C followed by 5 min 95°C for deactivation (Sigma, Taufkirchen, Germany).

Y1H screening

Bait plasmids were transformed to Y1HGold (Clontech) using the protocol described by Gietz and Schiestl (2007). Reporter yeast strains containing each bait construct were selected on synthetic drop-out medium including all amino acids except tryptophan (Clontech). Each strain was again grown at 30°C for 3 d in synthetic broth medium lacking tryptophan, and then co-transformed with a cDNA library representing all known transcription factors in *A. thaliana* provided by Mitsuda et al. (2010). Double transformed yeast was selected on synthetic drop-out medium that lacks tryptophan, leucine and histidine after adding 3-aminotriazole (3AT) at different concentrations (10, 15, 20, 25 and 30 mM) for identifying true positive interactions. Colony PCR in yeast was done using primers based on the prey vector (Table S1). Interactions that have been identified to be positive were based on

3AT selection and transformation of more than one prey vector (identified using yeast-colony PCR). Yeast growth pattern for each interaction and the sequenced prey vector are listed in.

RT-qPCR

RNA extraction was done using TRIzol after Chomczynski and Sacchi (1987) (38% water saturated phenol, 0.8 M guanidinium thiocyanate, 0.4 M ammonium thiocyanate, and 100 mM NaAc (pH 5.0), 5% glycerol). Extracted RNA was used for first strand cDNA synthesis using M-MLV reverse transcriptase (Promega, Mannheim, Germany) according to manufacturer protocol. Quantitative RT-PCR was carried out using platinum SYBR® Green qPCR SuperMix (Thermo Fisher Scientific, Schwerte, Germany). Water and non-template controls were run on each qPCR plate and contamination was controlled. All primers used during this study are listed in Table S1.

Co-expression analysis

Co-expression analysis was done by checking all available microarray probes published for the given AGI codes representing each gene on Genevestigator database (Zimmermann et al., 2004). Expression fold for each gene at each developmental stage, different anatomical regions, and under different experimentally tested perturbations were evaluated in a heat-map correlation. Overall correlation for each gene relative to *AtGldp*1 was calculated and used to interpret the degree of similar expression between the tested gene and *AtGldp*1. Genes like *AtGldp*2 and *SHMT* were used as positive controls for the correlation.

RESULTS

Yeast-one-hybrid screening and identification of the M-box binding factors

We previously showed that the M-box is a promoter *cis*-element that is conserved in C3 *Gldp*1 promoters from *Brassicaceae* but not found in the C2 promoters of *Gldp*1 group and its loss can result in a C2-like gene expression in C3 *A. thaliana* (Adwy et al., 2015). In order to identify which transcription factors are able to bind to the M-box, the full M-box was cloned upstream of a histidin nutrition marker (*HIS*3) gene and this plasmid was used as a bait and transformed to mutant yeast strains that are not able to grow on drop out medium lacking histidin or tryptophan (*TRP*1). Reporter yeast strains carrying the M-box controlling the expression of the *HIS*3 gene were transformed with a prey cDNA library representing transcription factors expressed in *A. thaliana*, known that prey plasmids carry a leucin nutrition marker for selection (Mitsuda et al., 2010). Reporter yeast strains double transformed with both the prey and the bait vectors were selected first on triple drop out medium (TDO) lacking leucin (LEU), tryptophan (TRP) and histidin (HIS). Colonies that appeared on plates were further transferred to a TDO-LEU-TRP-HIS medium after the addition of 3-aminotrizole (3AT) as a competitive inhibitor for *HIS*3 and therefore differentiating between false positives and true positives. 3AT was used at five different concentrations, 10, 15, 20, 25 and 30 mM. Yeast colonies that showed stable growth over increasing 3AT concentrations were defined as true positives. Interactions that have resulted due to co-transformation of more than one prey vector were defined by colony-PCR in yeast and were excluded. From 141 colonies, only 30 colonies were defined as true positive interactions, and the corresponding factors are listed after analysis of co-expression data with *AtGldp*1 using the available microarray data covering different developmental stages, expression in different anatomical regions, and response to different perturbations (Table S2, Zimmermann et al., 2004). As a control for the correlation analysis the *AtGldp*2 and serine hydroxymethyl transferase (SHMT) were integrated. Eight transcription factors showed an overall positive correlation to the *AtGldp*1, these factors are listed in Table 1. Hence, the yeast-one-hybrid screen coupled with co-expression analysis defines eight M-box binding factors.

Light-induced transcription conferred by the M-box

A genome-wide footprint experiment shows regulation of transcription factors occupancy around the M-box region in response to light (Fig. 1A, GSE53322; Sullivan et al., 2013), implying that light perception could be a common feature between the M-box and a possible M-box regulating factor. In order to define the latter suggested effect of light, transcription analysis was carried out under light-induced conditions. When the full *AtGldp*1 promoter was fused to *gusA* gene, *gusA* transcription level was 17-fold higher after 1h of light induction relative to dark conditions and this expression level increased 7-fold after 7h of light induction (Fig. 1B). A similar response was observed when fusing GUS reporter gene to the P3 promoter (-576 bp), but not for the P2 promoter (-379 bp) which has doesn't have the M-box, such that, *gusA* transcription level was 10-fold higher after 1h of light induction when *gusA* gene was fused to the P3 promoter, and this was increased 8-fold after 7h of light induction. When *gusA* was fused to the P2 promoter, that doesn't have the M-box sequence, transcription level of *gusA* didn't change after 1h of light induction and was increased only 2-fold after 7h of light induction (Fig. 1B). Light induced behavior of *gusA* transcripts was restored after fusing back the M-box to the P2 promoter (Fig. 1B, MUP2), where *gusA* transcripts increased 8-fold after 1h of light induction, and 31-fold after 7h of light induction. Similar response of transcripts was observed when fusing the M-box to the 35Smin promoter relative to only the 35Smin promoter controlling *gusA* expression, where *gusA* transcripts were increased by 3-fold following 1h and 7h of light induction (Fig. 1C).

Light induced transcription of the M-box binding factors

Up to here data indicate that light-response and previously described mesophyll expression of the *AtGldp*1 promoter are conferred by the M-box. In order to screen the list of M-box binding factors for an M-box regulator, light induced transcript of each factor was studied in correlation to the *AtGldp*1 transcripts. Transcripts of AT5G57660 and AT4G18890 were reduced by 28% and 24%, respectively, following 1h of light induction, and 20% and 26%, respectively, after 7h (Fig. 2A-B). Transcripts of AT5G66320, AT2G22200, AT5G47230 and AT4G26500 increased by 2.5-, 1.8-, 6.0-, 1.5-fold and 3.7-, 3.0-, 2.0-, 6.0-fold, respectively, following 1h and 7h of light induction (Fig. 2C-E). Transcripts of AT2G35700 and AT3G56220 didn't show any clear light induction after 1h of light, but rather increased 3.3-and 7.0-fold, respectively, following 7h of light induction and relative to 1h (Fig. 2G-H). Transcripts of *AtGldp*1 increased 4.0-fold after 1h and 30-fold after 7h of light induction, probably in response to one of these factors in the list (Fig. 2I). In correlation

with *AtGldp*1 expression, transcripts of an M-box regulating factor should be light-induced. Results show that out of eight M-box binding factors only six factors show light-induced transcription and the other two are obviously repressed by light. From the six light induced factors two factors show late light induction, and the other four have clear light induction from the first hour of light.

The M-box binding factor

In order to identify which of these factors is regulating the M-box, we repeated testing the binding of each factor to the M-box at higher 3AT concentration in a direct (1:1) yeast-one-hybrid. In addition, we conclusively tested the binding of each to a mutated M-box for the characterized *cis* element on the 3' end that has been shown previously to be important for M expression. Each prey vector representing each factor was tested for binding to each of three baits; bait vector having the full M-box cloned upstream of the *HIS*3 gene, bait vector having the M-box 3' end (M-box3'-*cis*) cloned upstream of the *HIS*3 gene, and a bait vector having the M-box-3'end after deleting the 10 nucleotides having the GATA prediction previously characterized to be important for M expression of the M-box (Fig. 3A). On testing the direct binding of each factor to the M-box, mutant yeast strains that have the full M-box bait controlling *HIS*3 gene were able to grow on medium lacking the three amino acids histidin, leucin, and tryptophan even at 50 mM 3AT (Fig. 3B, -TLH+3AT). This shows the stable binding of the eight factors to the M-box even after high competitive inhibition by adding the 3AT. On testing the binding of each to the M-box 3'-end, only three factors, AT5G66320, AT2G35700, and AT5G47230, were able to activate the expression of *HIS*3 gene when being controlled by the M-box 3'-end, and this enabled mutant yeast strains to grow on the medium lacking the three amino acids histidin, leucin, and tryptophan after high selective inhibition of 3AT (50 mM) (Fig. 3B, -TLH+3AT). The binding of five factors, AT5G57660, AT4G18890, AT2G22200, AT4G26500 and AT3G56220, was disturbed by losing the M-box-5' region, and these factors didn't activate the *HIS*3 gene. The lack of *HIS*3 gene activation was interpreted from the weak or no growth of yeast when these factors were co-transformed (Fig. 3B, Mbox-3'-*cis*). On deleting 10 nt on the M-box having the GATA prediction, the binding of the prey vector representing the AT5G66320 was disturbed, while bindings of the other M-box-3'-*cis* binding factors, AT2G35700, and AT5G47230, were not disturbed (Fig. 3B, 3'cisΔGATAA). These analyses end with the binding of only one transcription factor encoded by AT5G66320 to the M-box, and this binding was obviously disturbed after deleting the region from the M-box that is important for M expression.

We further attempted to study *AtGldp*1 expression in the corresponding T-DNA knock-out insertion lines for AT5G66320. We studied SALK_134217 that has been previously shown to have knocked out transcripts for AT5G66320 (Bi et al., 2005; Fig. 4). Amplification of T-DNA showed the expected product size in the tested SALK lines and not for WT (Fig. 4A). This shows successful T-DNA integration to the end of exon 2 and the beginning of the 3'-UTR of the AT5G66320 locus (Fig. 4B), where amplification of genomic DNA region using primers based on the AT5G66320 locus was only successful for WT but not for the mutants, showing that tested SALK_134217 lines are homozygous for the AT5G66320 T-DNA insertion. Expression of AT5G66320 was studied in SALK_134217, expression of AT5G66320 was almost none detectable compared to WT expression of AT5G66320 in Col-0 (Fig. 4C, AT5G66320). Expression of *AtGldp*1 in homozygous SALK_134217 lines was 30% less than WT (Fig. 4C, *AtGldp*1), however, the difference was not significant between replicates implying no change in expression of *AtGldp*1 in response to the knock-out mutation of AT5G66320 in SALK_134217. Data base search shows that the AT5G66320 belongs to GATA-transcription factor gene family (Reyes et al., 2004), and that only three GATA factors out of the 30 are preferentially expressed in shoots (Fig. S1; Zimmermann et al., 2004). Collective data suggest possible functional redundancy between the three GATA factors expressed in shoot tissues for conferring the M expression of the *Gldp*1 promoter.

DISCUSSION

The molecular restriction of the GDC into the bundle sheath compartment of C2 *Brassicaceae* plants is achieved by the loss of the M-box promoter *cis* element from the promoter of the *Gldp*1 homologs (Adwy et al., 2015). The presence of an M-box regulating factor would be interesting to identify. This is based on the hypothesis that a knock-out mutation in a gene encoding such factor might simulate steps of C2 evolution.

We analyzed the genome sequencing from a genome-wide- DNA foot printing experiment provided by Sullivian et al. (2013) in attempt to identify regions on the *AtGldp*1 with high occupancy with transcription factors (Fig. 1A). Surprisingly, photodynamic DNA-binding was found specifically on the upstream region that has the M-box, and transcripts of the *AtGldp*1 positively correlate with such temporal access to DNA, implying a positive regulation. Such photodynamic DNA accessibility is reported for less than 20% of *A. thaliana* genes (Sullivian et al., 2013). This analysis shows that there might be a correlation between light response and tissue specific expression conferred by the M-box, and subsequently suggests similar characteristics to the M-box regulating factor. Light-induction experiment shows that the M-box is also important for light perception of the *AtGldp*1, and we assume a co-expression correlation between the *AtGldp*1 expression and the M-box binding factor under light induction conditions (Fig. 1B). In summary, we define the following characteristics for an M-box regulating factor: binds to the M-box, have light-induced transcription, and expressed specifically in the mesophyll tissues.

A yeast-one-hybrid screen was carried out in order to identify such a factor (Table 1). In such screening, we have identified eight M-box binding factors that positively correlates with the expression of *AtGldp*1 mostly anatomically, and developmentally. Six of which show light induction in correlation to the *AtGldp*1, whereas two factors are clearly repressed with light (Fig. 2). This already shows that we can exclude the light-repressed factors as they don't show the same temporal expression pattern expected to regulate the *AtGldp*1. Only one factor, *Gata*5, has light induced transcripts and disturbed binding after mutating the M-box sequence responsible for its M-expression (Fig. 3). In summary, the *Gata*5 transcription factor is the only factor that showed disturbed binding after mutating the M-box, in addition to light induced transcript and preferential expression in the shoots of Arabidopsis (Fig. S1; Zimmermann et al., 2004). However, T-DNA knock-out in *Gata*5 locus has no influence on the *AtGldp*1 transcripts in the mutant (Fig. 4), implying possible functional redundancy between GATA-factors in regulating the M-box.

Previous gel shifting experiments point out to a GATA-factor that could bind specifically to the *Gata* sequence, and they describe that this factor is preferentially expressed in shoot tissues (Lam and Chua, 1989). It turned out later that there are 30 genes encoding different GATA-factors and they all have high homology in their zinc finger domain that enables binding to the DNA region with a *Gata* sequence motif (Reyes et al., 2004; Bi et al., 2005). Among the 30 factors the biological function of *Gata*21 was addressed by Bi et al. (2005), where they showed that a knock-out mutation in *Gata*21 appears to be deficient in chlorophyll synthesis, nitrate and carbon metabolism. GATA-transcription factors are not only specific to plants, but to all eukaryotes (Reichmann et al., 2000). It has been shown that outside of the plant kingdom GATA-factors could control gene expression at different developmental stages and in a tissue specific manner (Gidoni et al., 1989). The number of GATA transcription factors differs from species to another, for example there are 28 *Gata* genes in rice (Reyes et al., 2004), and six Gata factor sequences in humans with tissue specific expression pattern (Lentjes et al., 2016). Since we are looking for possibly redundant GATA factors in controlling the tissue specific expression of the *AtGldp*1, we looked into the microarray data published for all known *Gata* factors (Fig. S1). Data shows that there are three *Gata* factor genes with higher expression fold in shoots relative to roots, and therefore, might be redundantly regulating the *AtGldp*1 as well as other genes in shoots.

In conclusion, this study defines an M-box binding factor that belongs to the *Gata* transcription factor gene family and possibly regulating the M-box. The suggested redundancy between these factors would leave the loss of the M-box during C2 evolution to be the easiest possibility for achieving BS restriction of GDC.

REFERENCES

Adwy W, Laxa M, Peterhansel C (2015) A simple mechanism for the establishment of C2-specific gene expression in *Brassicaceae*. The Plant Journal **84:** 1231–1238.

Adwy W, Schlüter U, Papenbrock J, Peterhänsel C, Offermann S Loss of the M-box from the glycine decarboxylase P-subunit promoter in C2 *Moricandia* species. Journal of Plant Research, submitted in April 2018.

Anderson L (1971) Chloroplast and cytoplasmic enzymes pea leaf triose phosphate isomerases. Biochimica et Biophysica Acta **235:** 237–244.

Arents G, Moudrianakis E (1995) The histone fold: A ubiquitous architectural motif utilized in DNA compaction and protein dimerization. Proceeding of the National Academy of Sciences **92:** 11170–11174.

Bi Y-M, Zhang Y, Signorelli T, Zhao R, Zhu T, Rothstein S (2005) Genetic analysis of Arabidopsis GATA transcription factor gene family reveals a nitrate-inducible member important for chlorophyll synthesis and glucose sensitivity. The Plant Journal for Cell and Molecular Biology **44:** 680–692.

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry **162:** 156–159.

Gidoni D, Brosio P, Bond-Nutter D, Bedbrook J, Dunsmuir P (1989) Novel cis-acting elements in petunia *Cab* gene promoters. Molecular Genetics and Genomics **215:** 337–344.

Gietz RD, Schiestl RH (2007) Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Proceeding of the National Academy of Sciences **2:** 38–41.

Keerberg O, Pärnik T, Ivanova H, Bassüner B, Bauwe H (2014) C2 photosynthesis generates about 3-fold elevated leaf CO_2 levels in the C3-C4 intermediate species *Flaveria pubescens*. Journal of Experimental Botany **65:** 3649–3656.

Kehoe DM, Degenhardt J, Winicov I, Tobin EM (1994) Two 10-bp regions are critical for phytochrome regulation of a *Lemna gibba Lhcb* gene promoter. The Plant Cell **6:** 1123.

Lam E, Chua N-H (1989) ASF-2: A factor that binds to the *Cauliflower Mosaic* virus 35S promoter and a conserved GATA motif in *Cab* promoters. The Plant Cell **1:** 1147–1156.

Leegood R, Lea P, Adcock M, Häusler R (1995) The regulation and control of photorespiration. Journal of Experimental Botany **46:** 1397–1414.

Lentjes MHFM, Niessen HEC, Akiyama Y, Bruïne AP de, **Melotte V, van Engeland M** (2016) The emerging role of GATA transcription factors in development and disease. Expert Reviews in Molecular Medicine **18:** e3.

Li X-Y, Mantovani R, Huijsduijnen R, Andre I, Benoist C, Mathis D (1992) Evolutionary variation of the CCAAT-binding transcription factor NF-Y. Nucleic Acids Research **20:** 1087–1091.

Liberati C, Ronchi A, Lievens P, Ottolenghi S, Mantovani R (1998) NF-Y Organizes the γ-Globin CCAAT Boxes Region. Journal of Biological Chemistry **273:** 16880–16889.

Luger K, Mader A, Richmond R, Sargent D, Richmond T (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature **389:** 251–260.

Mantovani R (1999) The molecular biology of the CCAAT-binding factor NF-Y. Gene **239:** 15–27.

Mitsuda N, Ikeda M, Takada S, Takiguchi Y, Kondou Y, Yoshizumi T, Fujita M, Shinozaki K, Matsui M, Ohme-Takagi M (2010) Efficient yeast one-/two-hybrid screening using a library composed only of transcription factors in *Arabidopsis thaliana*. Plant and Cell Physiology **51:** 2145–2151.

Orgen WL (1984) Photorespiration: Pathways, regulation, and modification. Plant Physiology **35:** 415–442.

Reyes JC, Muro-Pastor MI, Florencio FJ (2004) The GATA family of transcription factors in *Arabidopsis* and rice. Plant Physiology **134:** 1718–1732.

Riechmann JL, Heard J, Martin G, Reuber L, -Z. C, Jiang, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang J, Ghandehari D, Sherman BK, -L. Yu G (2000) Arabidopsis Transcription Factors Genome-Wide Comparative Analysis Among Eukaryotes. Science **290:** 2105–2110.

Rolfe SA, Tobin EM (1991) Deletion analysis of a phytochrome-regulated monocot *rbcS* promoter in a transient assay system. Proceeding of the National Academy of Sciences **88:** 2683–2686.

Ronchi A, Berry M, Raguz S, Imam A, Yannoutsos N, Ottolenghi S, Grosveld F, Dillon N (1996) Role of the duplicated CCAAT box region in y-globin gene regulation and hereditary persistence of fetal haemoglobin. The European Molecular Biology Organization Journal **15:** 143–149.

Schulze S, Mallmann J, Burscheidt J, Koczor M, Streubel M, Bauwe H, Gowik U, Westhoff P (2013) Evolution of C4 photosynthesis in the genus *Flaveria*: establishment of a photorespiratory CO_2 pump. The Plant Cell **25:** 2522–2535.

Sessions A, Burke E, Presting G, Aux G, McElver J, Patton D, Dietrich B, Ho P, Backwaden J, Ko C, Clarke J, Cotton D, Bullis D, Snell J, Miguel T, Hutchison D, Kimmerly B, Mitzel T, Katagiri F, Glazebrook J, Law M, Goff S (2002) A High-throughput Arabidopsis reverse genetics system. The Plant Cell **14:** 2985–2994.

Siefers N, Dang KK, Kumimoto RW, Bynum WE, Tayrose G, Holt BF (2009) Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. Plant Physiology **149:** 625–641.

Sinha S, Maity SN, Lu J, de Crombrugghe B (1995) Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein-DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3. Proceeding of the National Academy of Sciences **92:** 1624–1628.

Sullivan AM, Arsovski AA, Lempe J, Bubb KL, Weirauch MT, Sabo PJ, Sandstrom R, Thurman RE, Neph S, Reynolds AP, Stergachis AB, Vernot B, Johnson AK, Haugen E, Sullivan ST, Thompson A, Neri FV, Weaver M, Diegel M, Mnaimneh S, Yang A, Hughes TR, Nemhauser JL, Queitsch C, Stamatoyannopoulos JA (2014) Mapping and

dynamics of regulatory DNA and transcription factor networks in *A. thaliana*. Cell Reports **8:** 2015–2030.

Zelitch I (1992) Control of plant productivity by regulation of photorespiration. BioScience **42:** 510–516.

Zimmermann P, Hirsch-Hoffmann M, Henning L, Gruissem W (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. Plant Physiology 136: 2621–2632.

FIGURES AND TABLES

Table 1: List of candidate M-box factors defined in a yeast-one-hybrid screen and listed after a heat-map correlation to the expression of *AtGldp*1 expression at different developmental stages, anatomical parts (shoots, and roots) and in response to all studied perturbations according to the published microarray probes retrieved from Genevestigator (Zimmermann et al., 2004). The *AtGldp*2 and *SHMT* are included as controls. The other genes are the M-box binding factors screened after the yeast-one-hybrid screen coupled with co-expression analyses.

	AGI code	Development	Anatomy	Perturbations	$\bar{x}r^2$
		***AtGldp* 1**			
		AT4G33010			
***AtGldp* 1**	**AT4G33010**				**1.00**
***AtGldp* 2**	**AT2G26080**				**0.80**
SHMT	**AT4G37930**				**0.86**
	AT5G57660				0.40
	AT5G66320				0.37
	AT2G22200				0.26
	AT4G26500				0.25
	AT2G35700				0.22
	AT4G18890				0.21
	AT5G47230				0.19
	AT3G56220	No data available			**?**

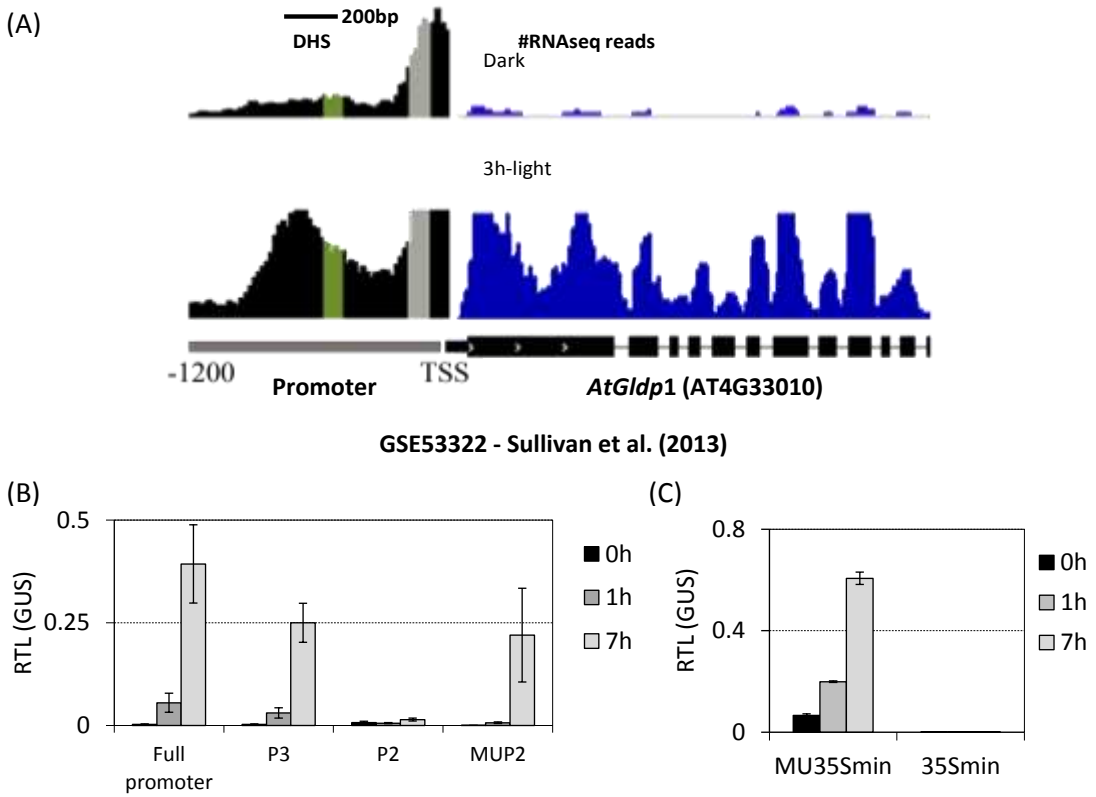


Figure 1. M-box light perception. (**A**) Transcription factors occupancy on *AtGldp*1 promoter. Left peaks; number of DNA reads corresponding to DNaseI hypersensitive sites (DHS), M-box (green), V-box (grey). Right peaks; RNA sequencing reads (dark vs. 3h of light). (Gene expression omnibus accession; GSE53322; https://www.ncbi.nlm.nih.gov/geo/; Sullivan et al. (2013)). (**B**) Relative transcription level of *gusA* (RTL), light induced *gusA* transcripts when being fused to full *AtGldp*1 promoter, P3, P2 and after fusing back the M-box to the P2 promoter (MUP2) (n=3±SEM). (**C**) RTL of *gusA* when being controlled by the M-box fused to the 35S min promoter (MU35Smin). (n=3±SEM).

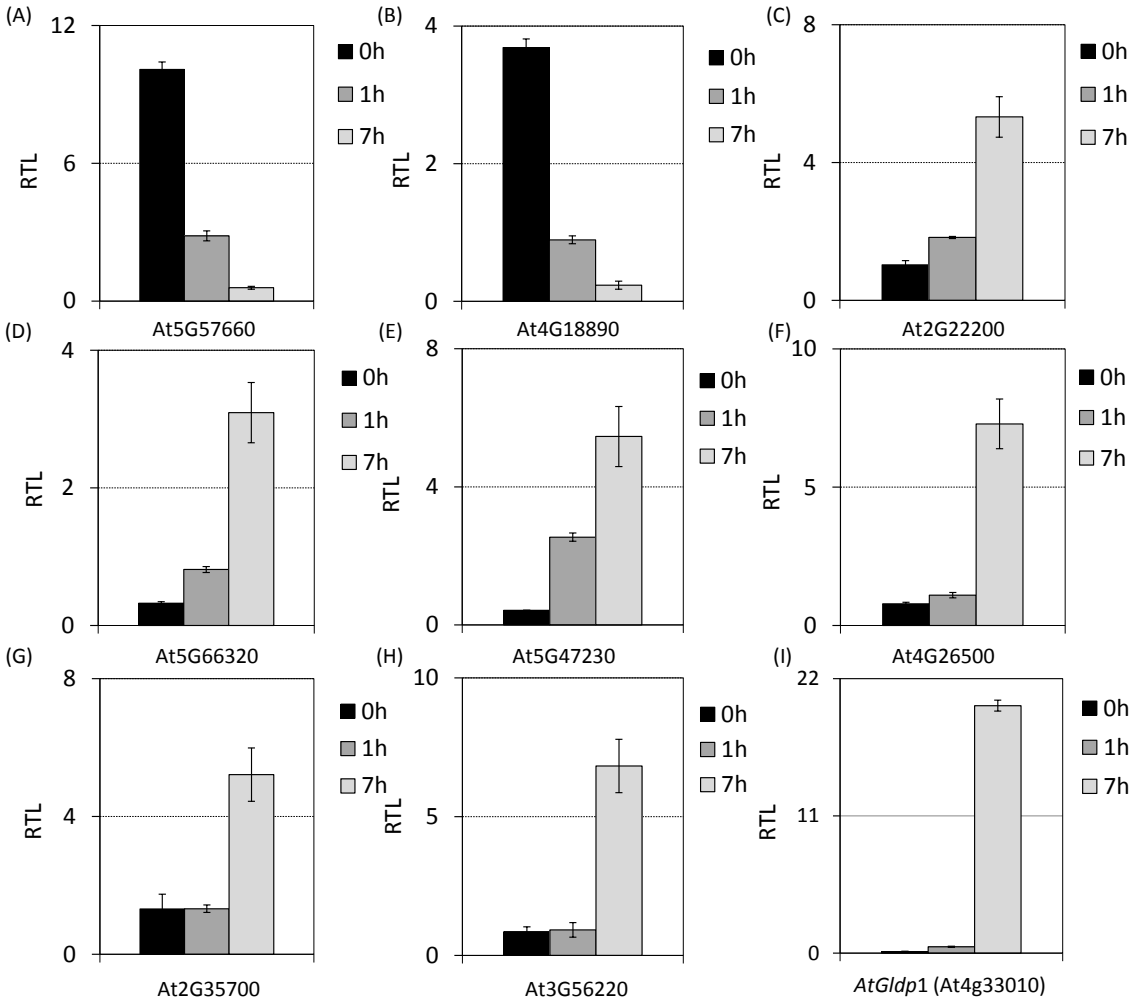


Figure 2. Light induction of genes encoding the M-box binding factors. Relative transcription level (RTL) for genes encoding M-box binding factors (**A-H**) and endogenous *AtGldp*1 (**I**) at dark conditions (0h), 1h and 7h after exposure to light, (n=3±SEM).

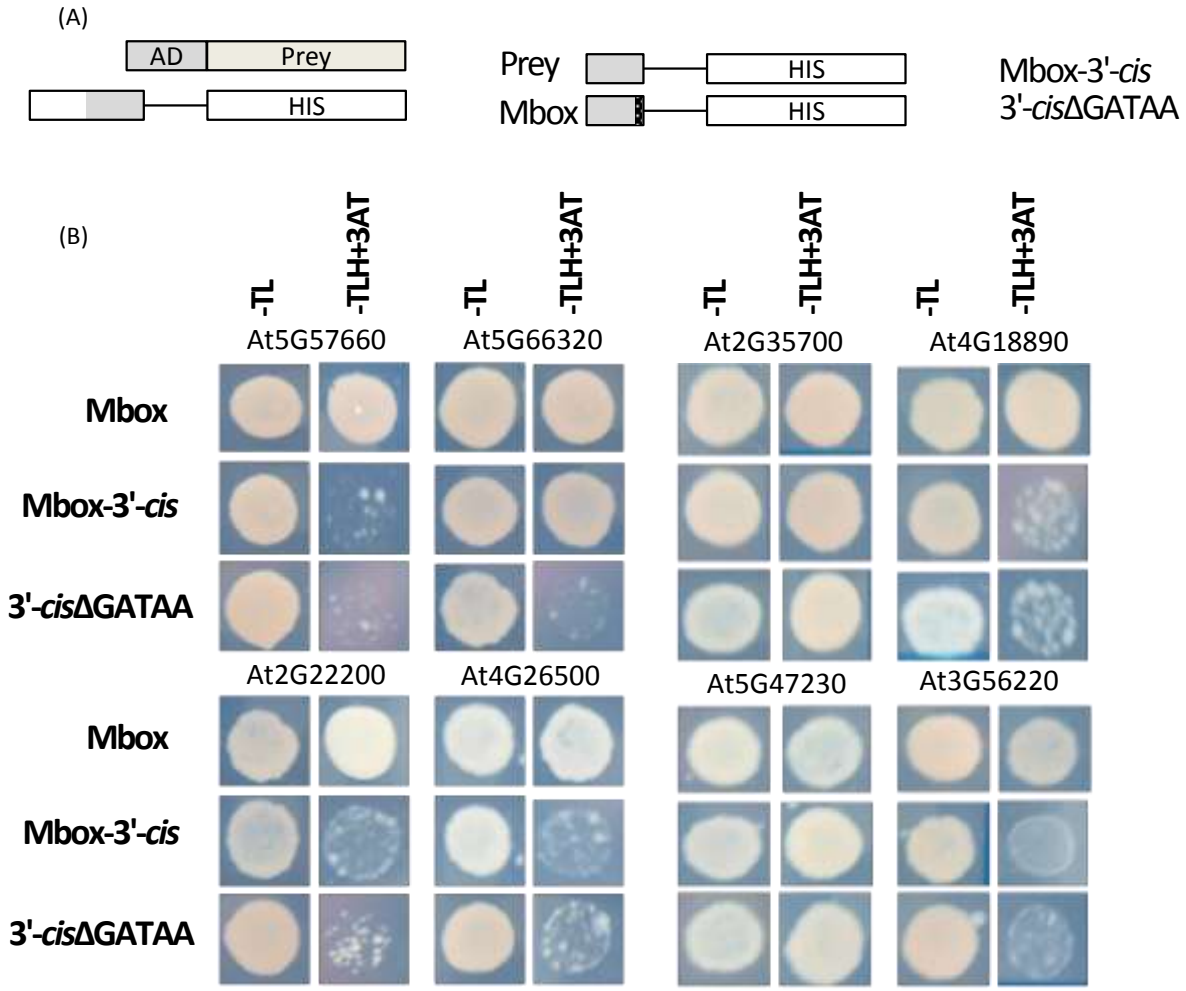


Figure 3. Interaction between M-box binding factors and different M-box baits in a direct yeast-one hybrid (Y1H). (**A**) Schematic representation of the construct, AD-Prey, full M-box, M-box-3'-cis elements, and M-box-3'-*cis*ΔGATA. (**B**) Direct Y1H to each bait; AD, GAL4-activating domain; -TL, synthetic complete medium lacking TRP and LEU; -TLH+3AT, synthetic complete medium lacking TRP, LEU, and HIS, 3-aminotriazole (3AT), competitive inhibitor for HIS3 enzyme. The -TLH+3AT medium contained 50 mM 3AT.

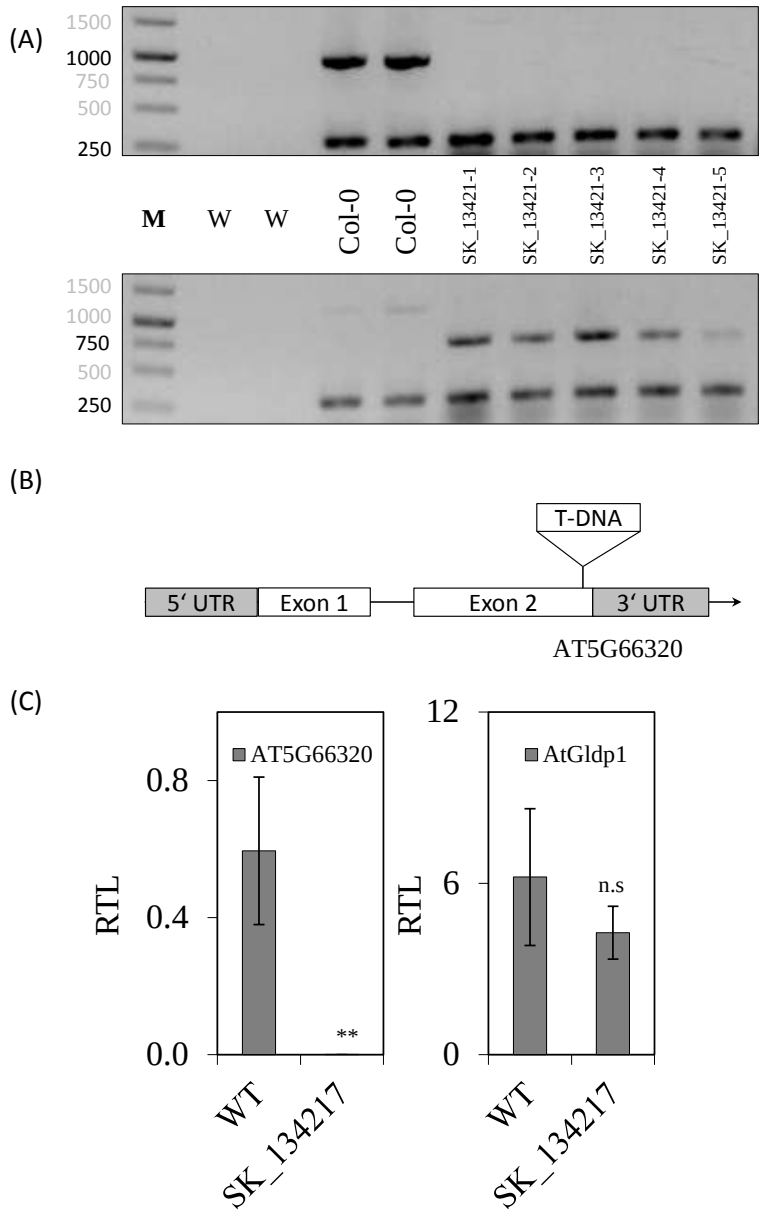


Figure 4. Characterization of *Arabidopsis* knockout line SALK_134217. (**A**) PCR with genomic DNA isolated from mutants vs. wild-type (WT) using gene-specific (upper gel) and T-DNA border (lower gel)-specific primers for AT5G66320 (SALK_134217-1-5). Control, positive 295 bp was amplified as a multiplex reaction in all samples (amplified: *AtGldp*1-P1 promoter). (**B**) Schematic representation of insertion in line SALK_134217 is illustrated. (**C**) Expression of AT5G66320 in SALK_134217 vs. Col-0 WT, and expression of *Gldp*1 in SALK_134217 vs. Col-0. Relative transcription level (RTL) in 2-week-old WT seedlings 1h after the onset of light, SK: SALK line, W: water, M: 1 kb DNA ladder. **= P<0.01 tested by student *t*-test (n=3±SEM).

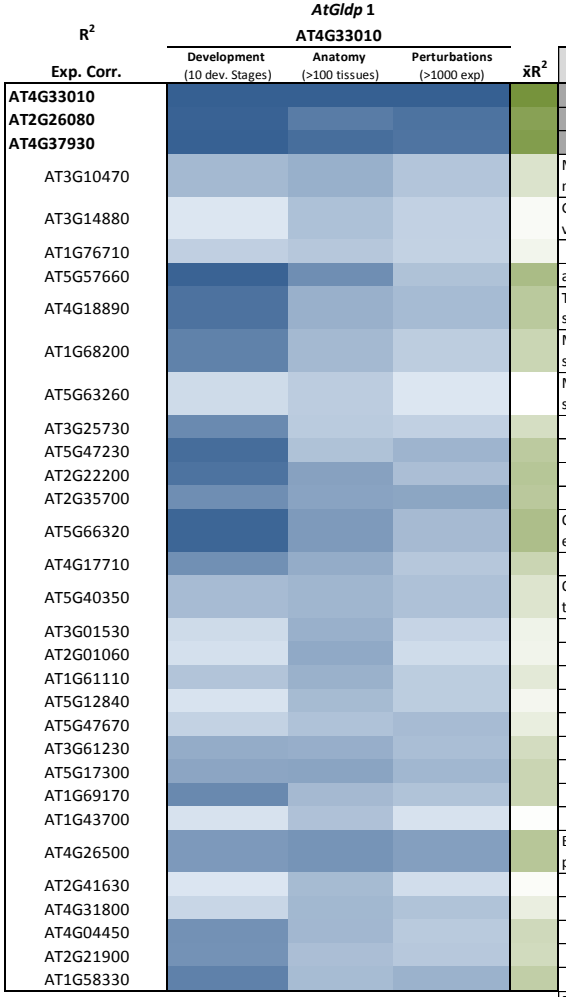
.

SUPPLEMENTARY INFORMATION

Table S1. Primers used in this study.

Primer name	Primer sequence (5'===>3')
Baits (Oligomers) added *EcoR*I and *Spe*I sites for cloning into pHIS2.1	
M-box-bait-F	GAATTCGAGCTCTGTCCACTGATACCCAATCTCTTCCACTTCAGTTCCCTTCCTTTCAAATTTTATCTCGTCACTAGT
M-box-bait-R	ACTAGTGACGAGATAAAATTTGAAAGGAAGGGAACTGAAGTGGAAGAGATTGGGTATCAGTGGACAGAGCTCGAATTC
M-box-3'-F	GAATTCGAGCTCACTTCAGTTCCCTTCCTTTCAAATTTTATCTCGTCACTAGT
M-box-3'-R	ACTAGTGACGAGATAAAATTTGAAAGGAAGGGAACTGAAGTGAGCTCGAATTC
M-boxΔGATA-F	GAATTCGAGCTCACTTCAGTTCCCTTCCTTTCAAATCACTAGT
M-boxΔGATA-R	ACTAGTGATTTGAAAGGAAGGGAACTGAAGTGAGCTCGAATTC
Primers for screening bait and prey vectors	
pHIS_F	TGGCAAGTGTAGCGGTCA
pHIS_R	TCGTTTATCTTGCCTGCTCAT
GAL4AD_F	ATACCACTACAATGGATG
GAL4AD_R	CGACGGATCCCTCTAGAC
Primers for genotyping *gata*5 mutant	
SALK_134217_F	ACATGTCGCGTGACCATGAAA
SALK_134217_R	CCGTAGCAAACGTAACCGCAA
LBa1	TGGTTCACGTAGTGGGCCATCG
qPCR primers	
BBx6- Fw	TCACTTGCAAAGCCGACGCC
BBx6-Rv	ACGGCGGTTTCGGCTGAGT
BEH3-Fw	GTCGCCAAAATCTCAGCTTCTTCG
BEH3-Rv	CCGATTTCGATTTTAGCCGGTTCG
EDF-Fw	CCGCCATTGAAGATTTCGCTTC
EDF-Rv	CCACGGTCTTTGTCTTACTCCT
ERF-AP2-Fw	TTCAGCTCCGTGGAGATATC
ERF-AP2-Rv	AACAGAGGAAGGGAGAGGATTC
ERF-38-Fw	GAGACAATGGGGAAAATGGG
ERF-38-Rv	TGGAGAAAGTACCGAGCCAGAT
ZW2-Fw	GTGGCCGGTGACGATATATATG
ZW2-Rv	TCCAGAGGATGAGTTTCTCG
GATA5_F	TTCGACCTCCTCCTCCTCTT
GATA5_R	TTCTTGGGAAACGGTGGCCT
SUFE1-F	ATCTGGAAGACCTGTCCCTG
SUFE1-R	TGCTGCAACCCGAGAAGAAC
bHLH-act-Fw	TGGCGTTTCTACTCATTGATT
bHLH-act-Rv	TCTTATATGTTTTAGCTTCTTCT

Table S2. M-box binding factors listed after co-expression analysis relative to the *AtGldp*1. Available data about each factor are listed. This includes function, DNA binding domain when defined, and member of which TF-family.



	Exp. Corr.	Function	Family members	DNA binding domain
***AtGldp* 1**	**AT4G33010**			
***AtGldp* 2**	**AT2G26080**			
SHMT	**AT4G37930**			
*	AT3G10470	Members might be involved in biological processes like RNA metabolism and chromatin-remodeling.	18 members	Zinc finger - C2H2-type
*	AT3G14880	GO cellular components (located in) flower pedicel, leaf vasculature, root vasculature, shoot epidermis	n.d	n.d
ASH1	AT1G76710			
BBX6	AT5G57660	act between circadian clock genes and genes controlling	17 members	Zinc finger - B box
BEH3	AT4G18890	Transcriptional repressor, expressed in collective leaf structure and roots	5 members	n.d
C3H15	AT1G68200	Most members are regulated by biotic and abiotic stresses- suggesting that they might have role during stress tolerance	2 members	Zinc finger-type CCCH
C3H67	AT5G63260	Most members are regulated by biotic and abiotic stresses- suggesting that they might have role during stress tolerance	9 members	Zinc finger-type CCCH
EDF 3	AT3G25730			
EDF 5	AT5G47230			
ERF / AP2	AT2G22200			
ERF 38	AT2G35700			
GATA 5	AT5G66320	C and N metabolims, light regulation and tissue specific expression (Expressed in Aerial parts)	30 members	Zinc finger -type IV
HDG4	AT4G17710			
MYB 24	AT5G40350	One of key genes regulating flower development - response to Jasmonate	> 100 member	MYB DNA bidning domain type R2R3
MYB 57	AT3G01530		> 100 member	MYB DNA bidning domain type R2R3
myb-like	AT2G01060			
NAC 25	AT1G61110			
NFY-A1	AT5G12840			
NFY-B6	AT5G47670			
PLIM2C	AT3G61230			
RVE 1	AT5G17300			
SPL6	AT1G69170			
SUE 3	AT1G43700			
SUFE 1	AT4G26500	Expressed in major tissues with higher expression in green parts, light dependent expression	3 members only 1 with TF activity	The only SUFE1 that has a BolA domain (DNA binding)
TFIIB	AT2G41630			
WRKY 18	AT4G31800			
WRKY 42	AT4G04450			
WRKY 59	AT2G21900			
ZW2	AT1G58330			
bHLH-act	AT3G56220	Protein expressed with actin interaction domain, strong expression in petal (mid-region) - cytoskelatal bending properties!	n.d	Helix-loop-helix DNA binding domain (Information by TAIR)

No microarray probes

* Abbreviation not defined

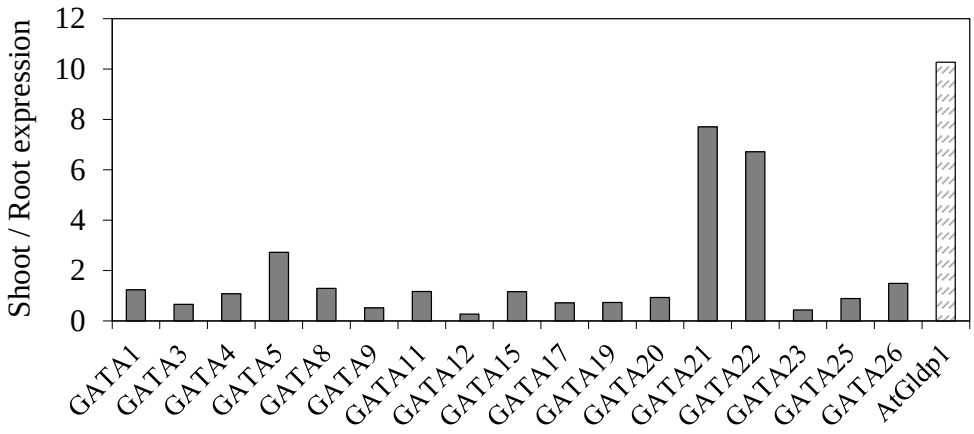


Figure S1. Expression of different *Gata* factors in shoots vs. roots. Microarray search was retrieved from Genevestigator database (Zimmermann et al., 2004). An expression fold of 1 indicates equal expression in both tissues. The *AtGldp*1 is included to the query. This query was updated on 15.09.2017; the missing *Gata* factors in this analysis are due to absence of microarray probes at the time of this analysis. Data for shoot expression was retrieved by Genevestigator as an average of 4580 from 224 experiments. Data for root expression was retrieved by Genevestigator as an average of 1081 samples from 52 experiments.

Chapter V: General discussion

The information required for C2 evolution is already present in C3 plants

C2 plants restrict expression of GDC to the BS compartment. This depends on the BS-specific localization of the P-subunit. The BS localization of the GDC-P subunit is the back-bone of the C2-photosynthetic pathway in many C2 plants (see chapter 1.1). The molecular mechanism of this BS-specific expression of the GDC-P was previously analyzed in the genus *Flaveria*. C3 *Flaveria* species already have a BS-specific and a globally expressed copy of genes encoding for the GDC-P subunit. Only the BS-specific copy of these genes was found to be actively expressed in C2 and C4 *Flaveria* species (see chapter 1.2; Schulze et al., 2013). It was suggested that *Brassicaceae* species lack a similar mechanism of evolution of photosynthesis as that described for *Flaveria* (Schulze et al., 2013) and it remains unclear how restriction of GDC to the BS was achieved in C2 *Brassicaceae* species.

The *AtGldp*1 is one of two genes encoding the GDC-P subunit in *A. thaliana*. The promoter of *AtGldp*1 has two *cis* elements called the M- and the V-boxes (Chapter II; Adwy et al., 2015). The M- and V-box control the spatial expression of a downstream fused *gusA* gene. The V-box is important for vascular and BS expression and the M-box is important for M expression. Accordingly, C2 *Brassicaceae* species might have lost expression of the GDC-P protein in the M by the loss of the M-box.

Comparison of *Gldp*1 promoters from different *Brassicaceae* species indicates conservation of the V-box in all C3 and C2 promoters. However, the M-box was only conserved in the promoters of C3 species and not found in the promoter of the C2 species *Moricandia nitens* implying an alteration in the promoter of C2 *Gldp* genes that might be responsible for restriction of the GDC to the BS compartment of C2 plants.

The *Brassicaceae* family comprises more than 4000 species, with frequent appearance of C2 species in the genus *Moricandia* (five C2 species) relative to genus *Diplotaxis* (three C2 species) (see chapter 1.4). Identifying different promoters for *Gldp* genes from *Moricandia*

showed the presence of the V-box in the promoters of both C3 and C2 species of *Moricandia* and the absence of the M-box only from the promoter of C2 *Moricandia* species like *M. nitens*, *M. suffruticosa*, and *M. arvensis* (Chapter III). Previously, a 135 bp M-box-like sequence was identified in the promoter of the *Gldp* gene from *B. napus* (C3). This 135 bp fragment was refered as being responsible for mesophyll specificity of that promoter (Zhang et al., 2011). Interestingly, this 135 bp M-box like sequence could be still identified in the promoter of *Gldp* genes from some C2 plants in *Brassicaceae* like *M. spinosa* and *D. tenuifolia* (Zhang et al., 2011). The latter finding of this M-box like element already suggests some exceptions for C2 members of *Brassicaceae* of having GDC restriction via the M-box loss, and suggests a different mechanism for GDC restriction rather than the M-box at least for these two C2 species. The GDC restriction to BS cells is commonly achieved by the loss of the GDC-P expression from the mesophyll compartment (Rawsthorne, 1992). Nevertheless, restriction of glycine decarboxylation to BS can be also achieved by restricting the expression of other subunits of the GDC to the BS this could be by restriction the GDC-H, GDC-T, GDC-L or even the SHMT (Morgen et al., 1993). In summary, the loss of the M-box from the promoters of C2 *Brassicaceae* is frequently observed in most C2 *Moricandia* species, however, this mechanism might possibly vary between different C2 species within the family of *Brassicaceae*.

Possible alterations toward C2 evolution in *Brassicaceae*

The GDC restriction to BS of C2 plants offer a weak pumping mechanism of CO_2 arround Rubisco in the BS of these plants (chapter 1). Anatomically, there are two possible routes for the released CO_2 from the GDC at the BS, either imported to the chloroplast or simply lost. In case of photorespiratory CO_2 that is imported into the chloroplasts of BS cells, the amount of the re-fixed CO_2 will not outperform a wild-type *A. thaliana* plant when the number of chloroplasts is the same as in the wild-type. Accordingly, other anatomical modifications might also be required for installing a C2 pathway in *A. thaliana*.

A recent study shows that the constitutive expression of a gene encoding the *GOLDEN2-LIKE* (GLK) transcription factor from C4 maize in C3 rice plant is associated with increased size of chloroplasts and mitochondria as well as higher plasmodesmata connections between M and BS cells, which mimics the proto-kranz anatomy (Wang et al., 2017). Interestingly, the induced proto-kranz modification in C3 rice was not sufficient to alter the photosynthetic parameters of the mutant rice plants relative to wild-type (Wang et al., 2017). Subsequently, restriction of GDC to BS as well as inducing a proto-kranz modification in a C3 plant could be interpreted to be both necessary for improving the photosynthetic performance of a C3 plant and none of the single modifications seems to be sufficient.

C2 metabolism involves glycine export from the mitochondria into the M cells, to be imported in the mitochondria of the BS cells (see chapter 1.1). This doesn't only necessitate increased plasmodesmatal connections between BS and M cells, but probably enhanced expression of a putative mitochondrial transporter that might be involved in glycine metabolism. A study by Eisenhut et al. (2013) identified a putative mitochondrial transporter that might be important for glycine metabolism, this transporter is encoded by the *A BOUT DE SOUFELE* (*BOU*) gene (Eisenhut et al., 2013). The *bau*-2 mutant accumulates significant amounts of glycine relative to wild-type at ambient CO_2 conditions (Eisenhut et al., 2013). Moreover, the expression of the *BOU* gene strongly correlates with the expression of both genes encoding the GDC-P subunit (Eisenhut et al., 2013). This might indicate another modification required for C2 metabolism that is linked with the molecular transport of glycine.

The *AtGldp*1 and *AtGldp*2 genes have a redundant function in keeping the plant alive, known that an *A. thaliana* mutant couldn't survive with a complete knock-out of both genes (Engel et al., 2007). However, the contribution of both *AtGldp*1 and *AtGldp*2 genes to

photosynthesis seems to vary, where only *gldp*1 mutants show a significant decrease in assimilation rates relative to wild-type (Manuscript 1; Adwy et al., submitted).

The first step of C2 evolution in *Brassicaceae* was probably the loss of *Gldp*2 coupled with some anatomical pre-conditioning toward C2 (Fig. 1; nonfunctional-adaptation). Genome wide expression data for members of the *Moricandia* lineage suggests the loss of *Gldp*2 homologs from C3 and C2 species of *Moricandia* (Schlüter et al., 2016). This might indicate that the loss of *Gldp*2 would bear no negative consequences but offers a step toward GDC restriction to the BS cells. However, this still remains to be shown for the genus *Diplotaxis*.

Anatomical changes were found to proceed physiological changes linked to evolution of photosynthesis (Christin et al., 2013). A similar scenario is also suggested for *Brassicaceae* such that anatomical modifications like increasing vein order, BS cell size, number and size of organelles might have started among C3 species of *Brassicaceae* as an anatomical pre-conditioning step (see chapter 1.2, Sage, 2004). The establishment of these anatomical modifications might have been the prerequisite for beneficial effects of the loss of GDC expression from the M.

A second step toward C2 development in *Brassicaceae* could be the restriction of GDC expression to the BS compartment, via the loss of the expression of the GDC-P from the M compartment (Morgan et al., 1993). This could be achieved by the loss of the M-box from the promoters of *Gldp* genes as shown for *M. nitens*, *M. suffruticosa*, and *M. arvensis* (chapter 3). It could also be different in other C2 species like *D. tenuifolia* and *M. spinosa* (Zhang et al., 2011). However, the restriction of GDC to the BS cells has to be achieved for enhancing CO_2 refixation in leaves of C2 plants (Keerberg et al., 2014). In parallel, it can be suggested that a more pronounced anatomical enablers for C2 photosynthesis have occurred in C2 plants (Fig. 1; nonfunctional adaptation). The latter suggested anatomical change is based on the

anatomical comparison that was previously done between leaves of C3 and C2 species in *Moricandia* indicates (Schlüter et al., 2016). C2 leaves show changes that differ from C3 and correlate with C2 evolution (Schlüter et al., 2016) such that *M. moricandioides* (C3) has smaller BS cells with less chloroplasts and reduced venation density. These features are enhanced in *M. suffruticosa* (C2) and even more enhanced in *M. arvensis* (C2) (Schlüter et al., 2016), suggesting a more pronounced anatomy as prerequisite for the C2 cycle. It is still not fully understood if C2 species developed some other molecular changes like enhanced expression of a mitochondrial glycine transporter for example the *BOU* (Eisenhut et al., 2013), or developed changes in the expression patterns of the *GLK* gene which has been shown before to regulate the development of chloroplasts (Wang et al., 2013).

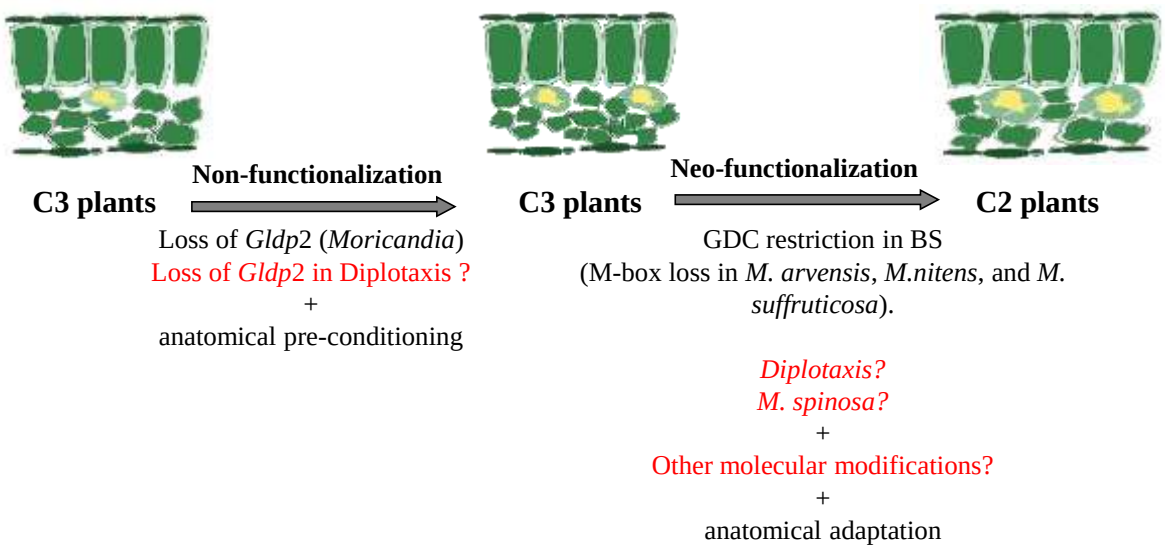


Figure 1. Proposed model for C2 evolution in *Brassicaceae*. Efficiency of CO_2-refixation was enhanced gradually during evolution. Step1: Nonfunctional adaptation: a redundant copy of *Gldp* was probably lost in addition to anatomical preconditioning. Step2: Neofunctional adaptation: GDC restriction to the bundle sheath (BS) cells of C2 plants, achieved in three C2 species (*M.nitens*, *M. suffruticosa*, *M. spinosa*) by the loss of the M-box from the promoter of *Gldp* genes, the mechanism could be different for other C2 species within *Brassicaceae*. Other molecular changes might be suggested, molecular changes related to plastid development and glycine transport.

Outlook

This work identifies and characterizes a simple mechanism for the establishment of C2-specific gene expression within the family of *Brassicaceae*. This mechanism seems to be frequently followed by a number of C2 species from genus *Moricandia*. It still remains to be shown how this is achieved for other members of C2 *Brassicaceae* like *D. tenuifolia* and *M. spinosa*. Restriction of GDC to the BS compartment of C2 plants might not be the only molecular alteration necessary to install a successful C2 pathway, but other molecular alterations could be also suggested that might be related to glycine transfort and proto-kranz development. It would be interesting to compare these changes between different C3 and C2 members of *Brassicaceae*, and to combine all of these alterations by means of genetic engineering for successful transition of a model C3 plant like *A. thaliana* into a C2 plant.

REFERENCES

Adwy W, Laxa M, Peterhansel C (2015) A simple mechanism for the establishment of C2 -specific gene expression in *Brassicaceae*. The Plant Journal 84: 1231–1238.

Adwy W, Schlüter U, Papenbrock J, Peterhänsel C, Offermann S Loss of the M-box from the glycine decarboxylase P-subunit promoter is conserved in C2 *Moricandia* species and is not detrimental for photosynthetic performance. Journal of Plant Research, submitted in April 2018

Christin P-A, Osborne CP, Chatelet DS, Columbus JT, Besnard G, Hodkinson TR, Garrison LM, Vorontsova MS, Edwards EJ (2013) Anatomical enablers and the evolution of C4 photosynthesis in grasses. Proceedings of the National Academy of Sciences 110: 1381–1386.

Eisenhut M, Planchais S, Cabassa C, Guivarc A, Justin A-M, Taconnat L, Renou J-P, Linka M, Gagneul D, Timm S, Bauwe H, Carol P, Weber APM (2013) Arabidopsis *A BOUT DE SOUFFLE* is a putative mitochondrial transporter involved in photorespiratory metabolism and is required for meristem growth at ambient CO_2 levels. The Plant Journal 73: 836–849.

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

Keerberg O, Pärnik T, Ivanova H, Bassüner B, Bauwe H (2014) C2 photosynthesis generates about 3-fold elevated leaf CO_2 levels in the C3-C4 intermediate species *Flaveria pubescens*. Journal of Experimental Botany **65:** 3649–3656.

Morgan CL, Turner SR, 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.

Rawsthorne S (1992) C3-C4 intermediate photosynthesis: linking physiology to gene expression. The Plant Cell **2:** 267–274.

Sage RF (2004) The evolution of C4 photosynthesis. New Phytologist 161: 341–370.

Schulze S, Mallmann J, Burscheidt J, Koczor M, Streubel M, Bauwe H, Gowik U, Westhoff P (2013) Evolution of C4 photosynthesis in the genus *Flaveria*: establishment of a photorespiratory CO_2 pump. The Plant Cell 25: 2522–2535.

Schulze S, Westhoff P, Gowik U (2016) Glycine decarboxylase in C3, C4 and C3–C4 intermediate species. Current Opinion in Plant Biology 31: 29–35.

Schlüter U, Bräutigam A, Gowik U, Melzer M, Christin P-A, Kurz S, Mettler-Altmann T, Weber APM (2016) Photosynthesis in C3–C4 intermediate *Moricandia* species. Journal of Experimental Botany 68: 191-206.

Wang P, Fouracre J, Kelly S, Karki S, Gowik U, Aubry S, Shaw MK, Westhoff P, Slamet-Loedin IH, Quick WP, Hibberd JM, Langdale JA (2013) Evolution of *GOLDEN2-LIKE* gene function in C_3 and C_4 plants. Planta 237: 481–495.

Wang P, Khoshravesh R, Karki S, Tapia R, Balahadia CP, Bandyopadhyay A, Quick WP, Furbank R, Sage TL, Langdale JA (2017) Re-creation of a key step in the evolutionary switch from C3 to C4 leaf anatomy. Current biology 27: 3278-3287.

Zhang C, Park I-a, Xu F, Li M, Lim YP, Meng J (2011) Transcriptionally and phylogenetically analyzing the P protein gene of glycine decarboxylase for understanding the evolution of C3–C4 species in Brassicaceae. Horticulture, Environment, and Biotechnology 52: 427–434.

Curriculum Vitae

Waly Adwy

Born on 12.03.1989 in Cairo, Egypt

e-mail 1: w.adwy@cu.edu.eg
e-mail 2: wallyessam@daad-alumni.de
e-mail 3: w.adwy@botanik.uni-hannover.de

EDUCATION AND TEACHING

03.2018 -10.2014	RESEARCH ASSISTANT - INSTITUTE OF BOTANY, GOTTFRIED WILHELM LEIBNIZ UNIVERSITY HANNOVER, GERMANY (PHD, CUM LAUDE – 1,0)
09.2014 -09.2012	MASTER OF SCIENCE INTERNATIONAL HORTICULTURE (MAJOR: PLANT BIOTECHNOLOGY), GOTTFRIED WILHELM LEIBNIZ UNIVERSITY HANNOVER, GERMANY (M.SC., SUMMA CUM LAUDE – 1,0)
09.2012 -09.2009	TEACHING ASSISTANT – DEPARTMENT OF GENETICS, FACULTY OF AGRICULTURE, CAIRO UNIVERSITY, EGYPT
09.2009 -09.2005	BACHELOR OF SCIENCE IN PLANT BIOTECHNOLOGY, CAIRO UNIVERSITY, FACULTY OF AGRICULTURE, DEPARTMENT OF GENETICS, EGYPT (B.SC., EXCELLENT WITH HONOUR)

List of Publications

Adwy W, **Laxa M, Peterhänsel C** (2015) A simple mechanism for the establishment of C2-specific gene expression in *Brassicaceae*. Plant J 84: 1231–1238.

Adwy W, Schlüter U, Papenbrock J, Peterhänsel C, Offermann S Loss of the M-box from the glycine decarboxylase P-subunit promoter in C2 *Moricandia* species. Journal of Plant Research, submitted in April 2018.

Contribution to conferences

(1) A simple mechanism for the establishment of C_2-specific gene expression in a C_3 plant, Photorespiration-Key for Better Crops, Rostock-Warnemünde, Germany, June 2015. (Oral presentation).

(2) Regulation of the *Gldp*1 promoter, C4 Photosynthesis, 50 years of discovery and innovation, QT Canberra, Australia, April 2016. (Poster presentation).

(3) The loss of the M-box is a frequent genetic alteration for C2-evolution in *Moricandia*, CO_2 Assimilation in Plants from Genome to Biome, Lucca (Barga), Italy, April 2017. (Poster presentation).

Acknowledgments

I was privileged when I had the chance to be a student of Christoph, Prof. Christoph Peterhänsel, he invested by all means in his students and therefore he earned their deep respect. He is a true scientist and was a great boss.

I would like to thank my colleagues in the institute of Botany, especially Dr. Christian Blume, from whom I learned about GCMS and gas-exchange measurements. Christian is an active scientist; he would be the one you can depend on if you need to collect samples during late night hours. Apart from science, and based on a number of barbecues, Christian is a maestro in grilling. I also want to thank Dr. Renke Perduns. I shared with Renke one office, and I learned from him about western-blots, Genome-seq analyses and Chromatin precipitation. Apart from that, I really enjoyed chatting with Renke. He is a fan of Hannover 96 football team by the way, and you might need to remember that when you meet him. I would also like to thank Dr. Denise Maibaum. I shared the office also with Denise, and we used to chat very often, this was really useful from time to time when you work for late hours of the day. Denise always had a spare tablet of Paracetamol when you have a headache, which always saved my day. I would like to express my thanks and appreciation for Dr. Miriam Laxa. I learned from Miriam about plant transformation and many other interesting things. Miriam is a bright scientist, and I wish her the very best in her scientific future. I would like to thank Dr. Ina Horst-Nießen. Ina is very organized and I enjoyed learning from her about chromatin in one of the practical courses that I visited, also if you have any troubleshoot with any device, and you want to learn how to solve it (not just getting it solved), I advise you to go to Ina. Birgit, Birgit, Birgit, “du bist wirklich lieb”, sie ist immer da, wenn du Hilfe brauchst. Yvonne Leye, Fabian Söffker, and I will never forget Lutz Krüger …. Lots of people deserve really kind words to express my thanks.

I would like to thank Prof. Jutta Papenbrock and Dr. Sascha Offermann as they supported all of the students from Christoph.

Prof. Edgar Maiß, I really enjoyed your lectures, asking, and chatting with you about every topic you taught us. When you attend a lecture for Prof. Maiß, you don't need to have a coffee before; it is interesting and interactive enough to keep you awake. I appreciate your help during my Masters and that you never hesitated to support during my PhD.

Only few times and I wish they were just many. I had a chance to chat with Roan and Tammy Sage during only two conferences. If you don't know who are Roan or Tammy, just google these names, you will understand why I wished many chances to meet them. Roan and Tammy are very inspiring not only to me but to most scientists working in this field.

It is self-evident that I owe you my deepest love and thanks to my wife "Dina". Apart from her studying with me at the same institute, she never failed to support me by her "all" and not her "few", I'd never be able to express my thanks to you. Never went-off from my heart or my mind my lovely Mother, Prof. Alia Hashem, I like to use here her academic title as it always kept me proud to have a mother scientist. I'd like to thank my Father, Dr. Essam Adwy for giving his best to us. I want to thank my brother Major Louay Adwy for being so kind and helpful whenever I needed him.

I had a blessing in my life; it is my daughter "Sedra". If you want to forget how hard your day was, just look at her smile.