Understanding the Bundle Sheath in C_4 evolution– Forward Genetic and Transcriptomic Approaches

Inaugural – Dissertation

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

vorgelegt von

Florian Döring

aus Haan

Düsseldorf, März 2017

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

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

Referent: Prof. Dr. Peter Westhoff
Koreferent: Prof. Dr. Maria von Korff Schmising

Tag der mündlichen Prüfung: 31.05.2017

Erklärung

Ich versichere an Eides Statt, dass die Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der "Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf" erstellt worden ist. Die Dissertation habe ich in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Düsseldorf, 20.03.2017

(Florian Döring)

Contents

Abbreviations

A	Adenine
ABA	Abscisic acid
AD Primer	Arbitrary degenerated primer
A. thaliana	*Arabidopsis thaliana*
ATP	Adenosine triphosphate
bp	Base pairs
BS	Bundle sheath
°C	Degree Celsius
C_1, C_2, C_3, C_4	One-, two-, three-, four-carbon molecule
C	Cytosine
CA	Carbonic anhydrase
CCM	CO_2-concentrating mechanism
CDS	Coding DNA sequence
CO_2	Carbon dioxide
DIT1+2	Dicarboxylate transporter 1+2
DNA	Deoxyribonucleic acid
EMS	Ethyl methanesulfonate
F. trinervia (Ft)	*Flaveria trinervia*
G	Guanine
GDC	Glycine decarboxylase complex
GFP	Green fluorescent protein
GGT	Glutamate-glyoxylate aminotransferase
$GLDPA_{Ft}$	Glycine decarboxylase PA gene of Flaveria trinervia
$GLDT_{Ft}$	Glycine decarboxylase T gene of Flaveria trinervia
GLYK	Glycerate kinase
GOX	Glycolate oxidase
h	Hour(s)
HCO_3^-	Bicarbonate (hydrogen carbonate)
HPR	Hydroxypyruvate reductase
Kb	Kilobases
kDA	Kilo Dalton
LB	T-DNA left border

LUC	Luciferase
MDH	Malate dehydrogenase
M	Mesophyll
M1, M2, M3, M4	1st, 2nd, 3rd or 4th mutant plant generation
ME	Malic enzyme
min	Minute(s)
mRNA	Messenger ribonucleic acid
N_2	Molecular nitrogen
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NH_3	Ammonia
nt	Nucleotides
O_2	Molecular oxygen
OAA	Oxaloacetate
ORF	Open reading frame
OsO_4	Osmium tetroxide
PCR	Polymerase chain reaction
PEP	Phospho*enol*pyruvate
PEPC	Phospho*enol*pyruvate carboxylase
PFD	Photon flux density
3-PGA	3-Phosphoglycerate
2-PG	2-Phosphoglycolate
PGLP	Phosphoglycolate phosphatase
PPDK	Pyruvate, orthophosphate dikinase
PPT	Phosphinothricin
RB	T-DNA right border
RNA	Ribonucleic acid
RubisCO	Ribulose 1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose 1,5-bisphosphate
s	Second(s)
SGT	Serine:glyoxylate aminotransferase
SHMT	Hydroxymethyltransferase
SNP	Single-nucleotide polymorphism
T	Thymine

TAIL-PCR	Thermal Asymmetric Interlaced-PCR
T-DNA	Transfer DNA
TEM	Transmission electron microscopy
TF	Transcription factor
THF	Tetrahydrofolate

I. Introduction

1. Photosynthesis and photorespiration

1.1 The two sites of 1,5-bisphosphate carboxylase/oxygenase

Life on earth strongly depends on energy from the sun. Photoautotrophic organisms, including higher plants, algae, and some bacteria, are capable of converting this light energy into chemical energy that manifests in carbohydrates. This process is called photosynthesis and it takes place in special compartments of photosynthetically active cells—the chloroplasts. Photosynthesis can be divided into two sections: the light-dependent and the light-independent reactions. In the former case, light energy and photosynthetic cleavage of H_2O is used to build up the energy-rich molecule ATP and the reducing equivalent NADPH. In the light-independent reactions of photosynthesis, both compounds are needed to assimilate CO_2 into carbohydrates. This happens in a cycle of reactions known as the Calvin-Benson cycle, which is named after its discoverers (Bassham et al., 1950). Three different mechanisms of CO_2 fixation exist in plants: C_3, C_4, and crassulaceaen acid metabolism (CAM). They differ in the time and spatial separation of the carboxylation reaction. The vast majority of land plants perform C_3 photosynthesis (Ehleringer et al., 1991), the photosynthetic pathway in which ribulose 1,5-bisphosphate (RuBP) carboxylase/oxygenase (RubisCO) catalyzes the transfer of CO_2 to RuBP to produce two molecules of the three-carbon compound 3-phosphoglycerate (3-PGA). 3-PGA is needed in the Calvin-Benson cycle to generate triosephosphates and to regenerate RuBP.

Besides the carboxylase activity, RubisCO also catalyzes the oxygenation of RuBP, especially under lower CO_2 concentrations (Ogren and Bowes, 1971; Bowes et al., 1971; Lorimer, 1981). In this case, only one molecule of 3-PGA is generated, along with one molecule of 2-phosphoglycolate (2-PG). Since 2-PG is useless and toxic to the plant, it has to be detoxified in a process called photorespiration (Bowes et al., 1971; Tolbert, 1971), which occurs in all plants, algae, and cyanobacteria. The extent of the oxygenase activity of RubisCO primarily depends on the CO_2 and O_2 concentrations in the atmosphere (Ehleringer et al., 1977). In general, a high $[O_2]/[CO_2]$ ratio promotes the oxygenase reaction. Additionally, the temperature has an impact on the carboxylase/oxygenase activity of RubisCO. The solubility of CO_2 decreases much faster at higher temperatures compared to O_2. This impedes transport of CO_2 within the cell and, consequently, lowers the CO_2 concentration around RubisCO, which favors RubisCO's oxygenase activity at leaf temperatures over 20–25 °C (Ehleringer et al., 1977; Jordan and Ogren, 1984; Brooks and Farquhar, 1985).

In general, RubisCO prefers CO_2 over O_2 (Jordan and Ogren, 1984), but today's atmosphere contains 20.95% O_2 and only 0.04% CO_2. Accordingly, each CO_2 molecule faces more than 500 O_2 molecules. Consequently, between 25% and 33% of RuBP is oxygenated and not carboxylated under the given conditions in C_3 plants (Sharkey, 1988; Ehleringer et al., 1991; Tcherkez, 2013). This has not always been the case. RubisCO arose approximately three billion years ago (Wacey, 2009) in a completely different environment and atmosphere. The CO_2 concentration was at least 100 times higher (Kasting and Howard, 2006) and there was nearly no O_2 present (Buick, 2008). Hence, the oxygenase activity of RubisCO was almost completely suppressed and there was hardly any evolutionary pressure on it. Only over the last 30 million years did the atmospheric composition change in such a way that the oxygenase activity of RubisCO became a severe problem for plants (Sage, 2004).

1.2 Photorespiration: An expensive but essential process in plants

Photorespiration can solve the problem of RubisCO oxygenase activity and, since there is no other known metabolism in plants that can deal with 2-PG, it is an indispensable process (Peterhansel et al., 2010). The photorespiratory cycle, also known as the C2 cycle, takes place in three different organelles—the chloroplast, the peroxisome, and the mitochondrion (Figure 1). More than 20 different enzymes and transporters are needed in this pathway to recover one molecule of 3-PGA from two molecules of 2-PG (Hagemann and Bauwe, 2016). It begins with the dephosphorylation of 2-PG into glycolate by 2-PG phosphatase (PGLP) in the chloroplast. A glycolate/glyoxylate transporter exports glycolate from the chloroplast into the cytosol, where it subsequently diffuses into the peroxisome (Pick et al., 2013). In the peroxisome, glycolate is irreversibly oxidized by glycolate oxidase (GOX), which produces glyoxylate and H_2O_2. A catalase immediately detoxifies the H_2O_2 and glyoxylate is transaminated to glycine by the action of glutamate:glyoxylate aminotransferase (GGT) and serine:glyoxylate aminotransferase (SGT). Glycine is then transported into the mitochondrion, where the mitochondrial glycine decarboxylase multienzyme system (GDC) and serine hydroxymethyltransferase (SHMT) catalyze the decarboxylation and deamination of two molecules of glycine to one molecule of serine, CO_2, and NH_3. GDC consists of four proteins, the P-, T-, L-, and H-proteins, which are all needed for the catalytic activity of the whole complex (Engel et al., 2007). Serine is then transported from the mitochondrion back to the peroxisome, where it is transaminated by SGT to hydroxypyruvate. Hydroxypyruvate reductase (HPR) subsequently reduces the hydroxypyruvate to glycerate. To complete the photorespiratory cycle, glycerate leaves the peroxisome and enters

the chloroplast to eventually become phosphorylated by glycerate kinase (GLYK) to 3-PGA, which directly enters the Calvin-Benson cycle again.

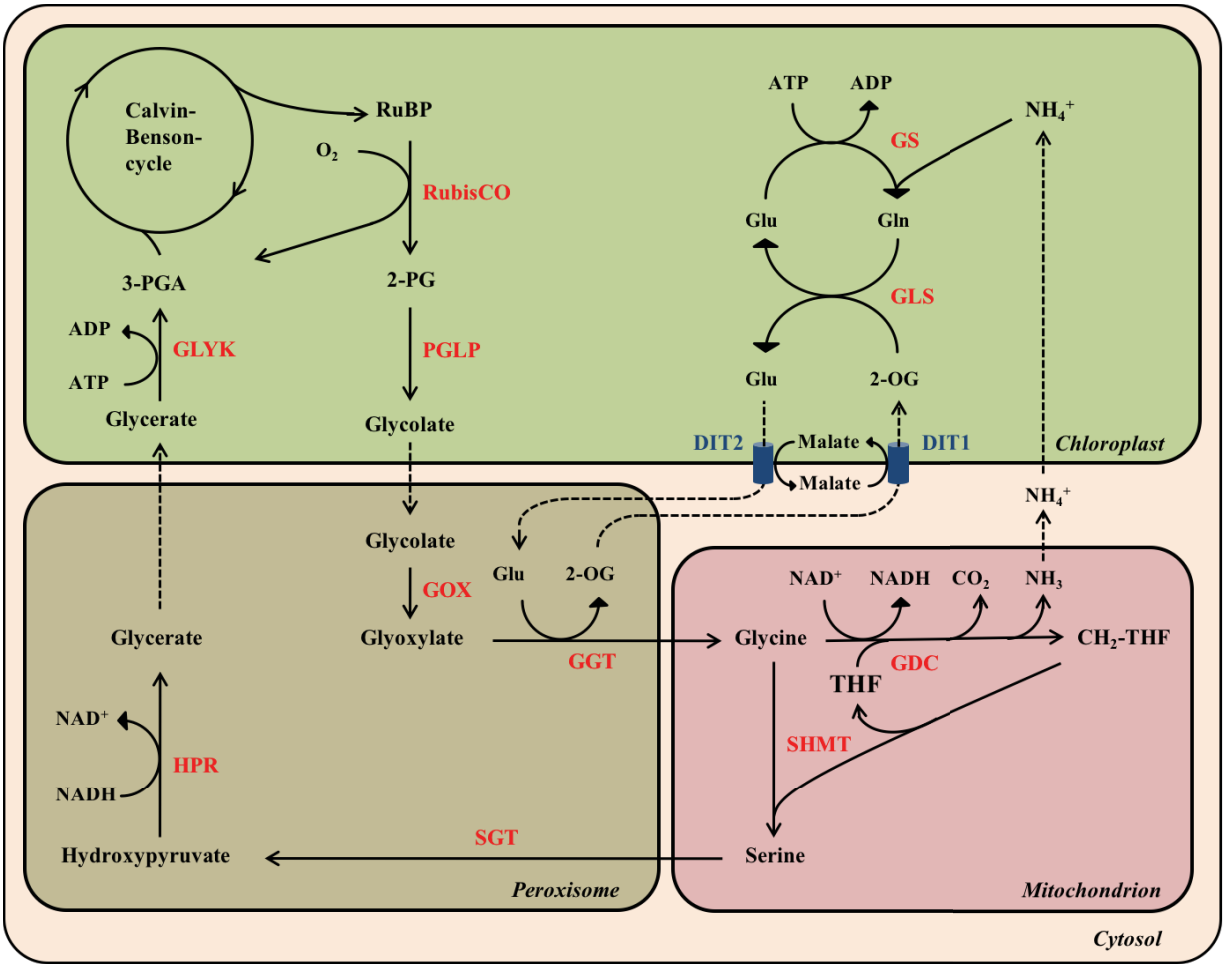


Figure 1. The reactions of the photorespiratory C_2 cycle.

Distribution of photorespiratory reactions between chloroplasts, peroxisomes, and mitochondria. Enzymes of the pathway are highlighted in red. 2-OG, 2-oxoglutarate; 3-PGA, 3-phosphoglycerate; DIT1+2, dicarboxylate transporter 1+2; GDC, glycine decarboxylase complex; GGT, glutamate glyoxylate aminotransferase; Gln, glutamine; GLS, glutamate synthase; Glu, glutamate; GLYK, glycerate kinase; GOX, glycolate oxidase; GS, glutamine synthetase; HPR, hydroxypyruvate reductase; PGLP, phosphoglycolate phosphatase; SGT, serine glyoxylate aminotransferase; SHMT, serine hydroxymethyltransferase; THF, tetrahydrofolate.

During photorespiration, not only CO_2 but also ammonia is released in the mitochondria. Consequently, it makes sense that the core C_2 cycle is tightly connected to the re-assimilation of ammonia (Keys, 2006). The recovery of ammonia requires a lot of energy and it is carried out by two enzymes in the chloroplast: glutamine synthetase (GS2) and ferredoxin-dependent glutamate synthase (GLS) (Figure 1). GS2 combines glutamate and ammonia to glutamine, which is used by GLS for the transamination of oxoglutarate. This results in two molecules of glutamate that are needed for a new cycle of ammonia refixation and for glycine synthesis in the peroxisomes during photorespiration.

In conclusion, the photorespiratory pathway successfully eliminates the useless and toxic intermediate 2-PG and converts it back to 3-PGA. But there is also a flip side to photorespiration. The process is highly expensive in terms of energy as well as the efficiency of carbon fixation. ATP and NADPH are needed to operate photorespiration and CO_2 is lost throughout the decarboxylation of glycine by GDC in the mitochondria (Figure 1). Only three of four carbon atoms that enter the photorespiratory cycle eventually become fixed. As much as 25% is lost as photorespiratory CO_2. For each RuBP that gets oxygenated, it takes 3.25 ATP and 2 NADPH to recover it in the photorespiratory cycle (Wingler et al., 2000). Especially the re-assimilation of lost ammonia is a wasteful process.

All in all, it is assumed that there is an extra energy cost of approximately 50% for photosynthesis caused by the oxygenase activity of RubisCO (Peterhansel et al., 2010), and photorespiration can reduce the efficiency of C_3 photosynthesis by up to 30% (Ogren, 1984; Bauwe et al., 2010; Fernie et al., 2013). The question then is why plants use such an expensive cycle in terms of energy and CO_2 lost. The answer to this is simple: It is the only way to get rid of 2-PG and keep the Calvin-Benson cycle operating under the current atmospheric conditions. As long as O_2 is present, it will always compete with CO_2 for the active site of RubisCO and only photorespiration allows the Calvin-Benson cycle to work in the presence of O_2.

1.3 Photorespiration as a primary target for crop improvement

Many studies provide clear evidence that the photorespiratory cycle is essential for plants by knocking out different genes of the pathway leading to so-called photorespiratory phenotypes. These are mutants that only grow under elevated CO_2 concentrations where the rate of photorespiration is very low. Although they do not grow in ambient atmosphere and develop chlorosis, they mostly can be rescued by shifting them back to higher CO_2 concentrations (Sharkey, 1988; Somerville, 2001; Bauwe et al., 2012). There are also photorespiratory mutants that cannot survive even in non-photorespiratory conditions, such as GDC mutants. This is because of the crucial role of GDC in one-carbon metabolism (Engel et al., 2007).

Photorespiration costs energy and freshly assimilated CO_2. Hence, many groups worldwide are looking for ways to minimize photorespiration in plants, especially in crops to gain more yield and to cover the rising demands for food worldwide in the coming years (Kebeish et al., 2007; Peterhansel et al., 2008; Evans, 2013; Long et al., 2015; Walker et al., 2016). There are different ideas on how to achieve this goal. One strategy could be to make a better RubisCO with more affinity for CO_2 (Marcus et al., 2011; Parry et al., 2013). Avoiding the high costs of ammonia refixation is also of interest (Carvalho et al., 2011). Besides these and other ideas, there is one

possibility that has not only been tried artificially, but which is already present and developed in nature: CO_2-concentrating mechanisms (CCMs). The idea is as simple as it sounds. If the CO_2 concentration around RubisCO increases, less oxygenase activity occurs and CO_2 becomes fixed more efficiently. Consequently, there would be less photorespiration in the leaves of the plant (Price et al., 2013). There is one very successful way of photosynthesis in which this mechanism is already optimized: C_4 photosynthesis.

2. The biochemistry and evolution of the C_4 syndrome

2.1 C_4 photosynthesis: A solution to photorespiration

The basic idea of C_4 photosynthesis and CCMs in general is the concentration of CO_2 around the carboxylating enzyme RubisCO. Consequently, the oxygenase reaction is suppressed and photorespiratory rates are reduced. This is achieved by anatomical and biochemical modifications to the original C_3 pathway and mostly by the division of labor between two photosynthetic cell types: bundle sheath cells and mesophyll cells (Hatch, 1987). To this end, C_4 plants have a very special leaf anatomy, which is referred to as the Kranz anatomy, first described by Haberlandt (1881). The vasculature of C_4 plants is surrounded by a cell layer of big bundle sheath cells forming a wreath-like structure, which, in turn, is encircled by, typically, only one cell layer of mesophyll cells (Figure 2).

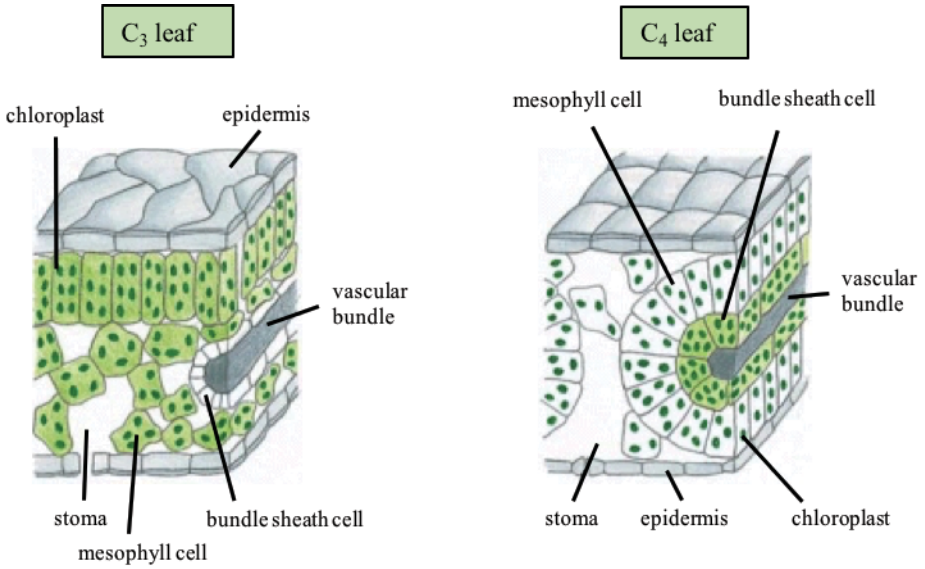


Figure 2. The typical leaf architecture of a C_3 and a C_4 leaf.

Leaves of C_4 plants largely differ in their anatomy from leaves of their C_3 relatives. They have a higher bundle sheath/mesophyll cell ratio with more contact between both cell types. Bundle sheath cells in general are larger in C_4 plants and contain more organelles. Source: Adapted from Alberts B, Johnson A (2002).

The interveinal distance and the leaf thickness is often reduced in C_4 plants (Dengler and Nelson, 1999; Leegood, 2002). This guarantees optimal contact between both cell types, which are also connected through many plasmodesmata to facilitate the transport and exchange of metabolites (Laetsch, 1974; Botha, 1992).

The C_4 cycle starts in the cytosol of the mesophyll, where CO_2 is converted to bicarbonate (HCO_3^-) by carbonic anhydrase (CA) and, subsequently, fixed into oxaloacetate (OAA) (Figure 3). Therefore, PEP carboxylase (PEPC) connects HCO_3^- and the acceptor molecule phospho*enol*pyruvate (PEP). Because the four-carbon acid oxaloacetate is the first stable product after the fixation of CO_2, this type of photosynthesis is referred to as C_4 photosynthesis. In the next step, OAA is either reduced to malate or transaminated to aspartate, depending on the type of C_4 photosynthesis. The C_4 acids will then diffuse into the bundle sheath cells and will become decarboxylated by NADP-malic enzyme (NADP-ME), NAD-malic enzyme (NAD-ME), or PEP carboxykinase. In some C_4 plants, such as maize, a combination of these three enzymes is also possible (Furbank, 2011; Wang et al., 2014). The released CO_2 is directly fixed by RubisCO and eventually enters the Calvin-Benson cycle. The remaining three-carbon compounds of the decarboxylation reaction move back to the mesophyll and—if NADP-ME or NAD-ME were used as decarboxylating enzymes—become converted to PEP again by pyruvate-orthophosphate dikinase (PPDK).

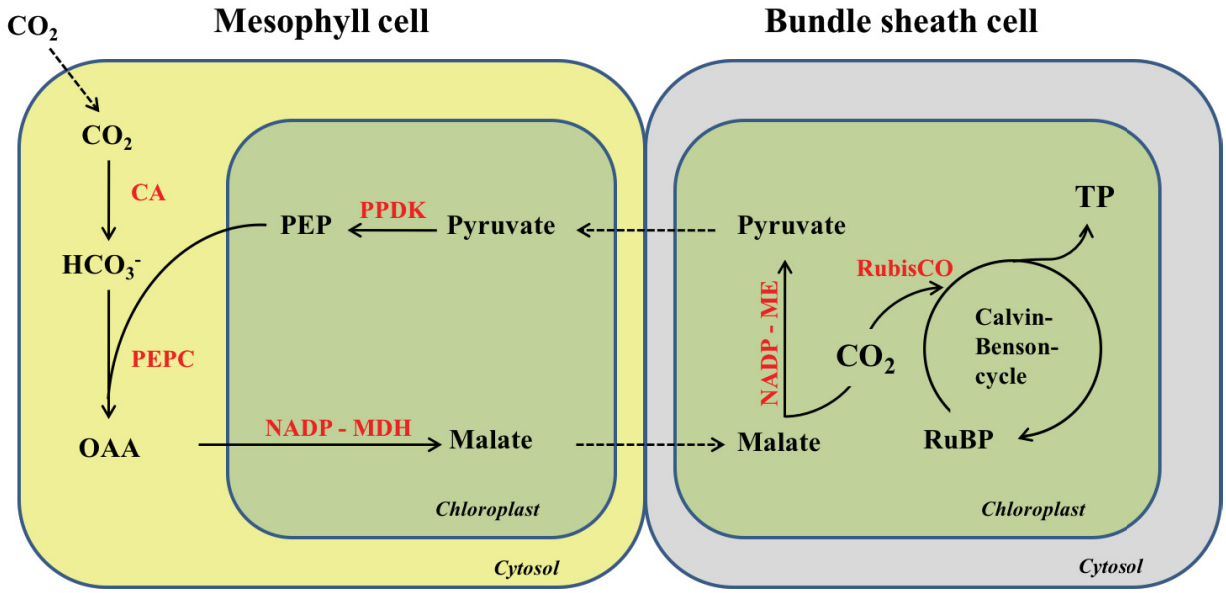


Figure 3. The reactions of the C_4 pathway using the NADP-ME.

The core C_4 cycle of the NADP-malic enzyme type and the distribution of labor between mesophyll and bundle sheath cells. CA, carbonic anhydrase; NADP-MDH, NADP-dependent malate dehydrogenase; NADP-ME, NADP-malic enzyme; OAA, oxaloacetate; PEP, phospho*enol*pyruvate; PEPC, phospho*enol*pyruvate carboxylase; PPDK, pyruvate-orthophosphate dikinase; RuBP, ribulose 1,5-bisphosphate; TP, triosephosphate.

The prefixation of CO_2 by PEPC in the mesophyll and its release in the bundle sheath chloroplasts results in a biochemical CO_2 pump that enables RubisCO to work under high CO_2 concentrations. Since the oxygenase activity of RubisCO mainly depends on the $[CO_2]/[O_2]$ ratio around it, this carbon-concentrating mechanism effectively represses photorespiration (Furbank, 2011). As a result, photorespiratory rates are significantly lower in C_4 plants as compared to C_3 plants. However, even C_4 plants have to run the photorespiratory cycle to some extent (Yoshimura et al. 2004). Due to the fact that RubisCO can work much more efficiently in C_4 plants, they require 50–80% less RubisCO for a given photosynthetic rate (Sage and Zhu, 2011). Consequently, C_4 plants have a better nitrogen-use efficiency, because RubisCO is by far the most abundant enzyme in plants (Oaks, 1994). Besides this, C_4 plants can keep their stomata closed for a longer time, which reduces water losses through transpiration, resulting in a better water-use efficiency (Sage, 2001a). For that reason, C_4 plants often grow in hot and arid environments with high light intensities—conditions that are not favored by C_3 plants (Ehleringer et al., 1997).

The presence of an organelle-rich bundle sheath is a feature that all C_4 plants with Kranz anatomy share. They pile large number of chloroplasts, mitochondria, and peroxisomes in their bundle sheath (Brown and Hattersley, 1989). However, the chloroplasts can differ at the ultrastructural level among the subtypes of C_4 photosynthesis (Woo et al., 1970). In the NADP-ME type of C_4 photosynthesis, the chloroplasts contain few grana and, therefore, only low photosystem II (PSII) activity, while there is high PSII activity in the bundle sheath of NAD-ME plants. Consequently, only photosystem I-mediated cyclic electron flow is possible in NADP-ME plants to generate ATP.

2.2 C_4 photosynthesis evolved many times independently

C_4 photosynthesis only occurs in the angiospermes (Ehleringer et al., 1997). In evolutionary terms, C_4 photosynthesis is a rather recent invention that evolved at least 66 times independently from the original C_3 pathway over the last 35 million years (Sage et al., 2011a; Aliscioni et al., 2012). Molecular phylogenies and genomic analysis of protein evolution allowed to estimate the number of distinct C_4 lineages and, therefore, the *de novo* origins of C_4 photosynthesis (Sage et al., 2012; Sage, 2016). It is assumed that there are 20–23 distinct lineages in the grasses (Poaceae) with 5,044 C_4 species in total (Aliscioni et al., 2012; Christin et al., 2012; Besnard et al., 2014; Osborne et al., 2014), six lineages in the sedges (Cyperaceae) with 1,322 C_4 species (Besnard et al., 2009; Roalson et al., 2010; Larridon et al., 2013; Sage, 2016), nine lineages in the Chenopodiaceae comprising 558 C_4 species (Kadereit et al., 2012; Sage, 2016), and, finally, more

than 1,750 other C_4 species in the eudicots with 34 *de novo* origins (Sage, 2016). In summary, the percentage of plants performing C_4 photosynthesis is relatively low. Approximately 8,000 species of more than 250,000 plant species worldwide perform this type of photosynthesis (Sage, 2016). Nevertheless, C_4 plants account for 23% of the terrestrial biomass production on earth (Still et al., 2003; Sage et al., 2012), which underlines their success and importance in agriculture.

With many independent origins of C_4 photosynthesis, there is also a lot of variation between different lineages. Each line is unique in its way and there are only two steps that they all have in common: the conversion of CO_2 to HCO_3^- and the carboxylation of PEP by PEPC (Sage et al., 1999). Thus, it may be more accurate to think of it as the "C_4 syndrome" rather than a fixed and defined pathway (Sage et al., 2012).

2.3 A conceptual model of the stepwise evolution of the C_4 trait

A low CO_2 concentration in the atmosphere has always been supposed to be the "driving force of C_4 evolution" (Ehleringer et al., 1991; Ehleringer et al., 1997), but these days it is assumed that low CO_2 is only a precondition for C_4 evolution that enables other factors to play a decisive role. Parameters such as heat, aridity, high light, salinity, and ecological disturbance can lower CO_2 concentrations in the leaves of C_3 plants and, alone or in combination, promote photorespiration (Moore, 1983; Sage, 2004; Osborne and Freckleton, 2009; Sage et al., 2012). Under these conditions, photorespiration can become a severe problem in terms of photorespiratory CO_2 loss. Thus, plants that can find a solution to relocate CO_2 to the chloroplasts for refixation would strongly benefit in these environments and outcompete other C_3 species. The restriction of GDC activity to an inner compartment, the bundle sheath, is the simple solution to this problem, invented and adopted by many plants (Sage et al., 2012). The spatial separation of RubisCO oxygenation and GDC activity resulting in a photorespiratory CO_2 pump is a very important intermediate step toward C_4 photosynthesis. This process is also referred to as C_2 photosynthesis (Vogan et al., 2007), since CO_2 is transported in the two-carbon compound glycine to the side of RubisCO (Sage et al., 2012; Bräutigam and Gowik, 2016). All current models of C_4 evolution propose that C_2 photosynthesis has to be established before a functional C_4 metabolic cycle can operate and replace the C_3 cycle in the mesophyll (Rawsthorne, 1992; Sage RF, Li M, 1999). One famous and well-known model of C_4 evolution has been proposed by Sage et al. (2004; 2012) and describes the gradual evolution of the C_4 trait, summarized in five major phases: (1) preconditioning, (2) evolution of proto-Kranz anatomy, (3) establishment of C_2 photosynthesis, (4) introduction of a C_4 metabolic cycle in the bundle sheath and mesophyll, and (5) optimization processes of the C_4 syndrome (Figure 4). Further models, like the biochemical

model from Heckmann et al. (2013), can give more evidence for a step-by-step evolution in one direction, a situation they refer to as the "Mount Fuji fitness landscape." It relies on the assumption that each step a plant takes toward C_4 has to increase fitness.

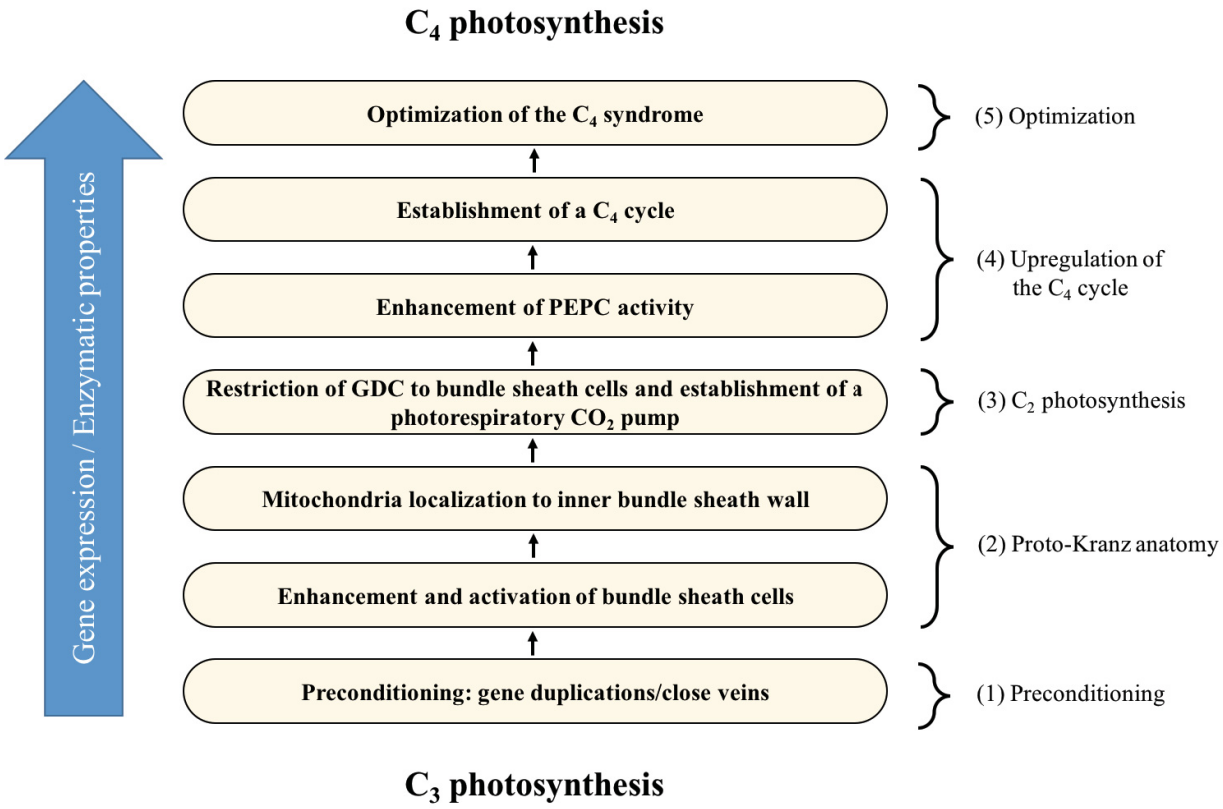


Figure 4. Evolution of C_4 photosynthesis.

The gradual evolution of the C_4 trait in five major phases. GDC, glycine decarboxylase complex; PEPC, phospho*enol*pyruvate carboxylase. Source: Adapted from Gowik and Westhoff (2011), and Sage et al. (2012).

In their current model, Sage et al. (2012) assume that C_4 evolution has to start with a kind of preconditioning that raises the potential for C_3 plants to evolve C_4 photosynthesis. Most C_3 families do not have C_4 plants at all, while many C_4 lineages cluster, which indicates that it was easier for some plants to evolve C_4 than for others (Sage, 2004; Sage et al., 2011a). One of these preconditions is the appearance of gene duplications and a large genome size. More copies of a gene allow plants to experiment with one copy that could be neo-functionalized through mutations without disturbing the original function (Monson, 2003; Wang et al., 2009). Another trait facilitating the start of C_4 evolution is high vein density. This can easily be observed in C_3 plants that are closely related to members of C_4 clades, like in Cleome, Heliotropium, or Flaveria (McKown and Dengler, 2007; Marshall et al., 2007; Christin et al., 2009; Sage et al., 2011*b*; Muhaidat et al., 2011). It is assumed that closer veins are an adaptation to hot and dry conditions that lead to reduced evaporation within leaves (Scoffoni et al., 2011; Osborne and Sack, 2012;

Sage et al., 2012). Additionally, more veins increase the overall stability of plant leaves, especially in windy environments (Sage, 2004).

Once preconditions are met, it is more likely for a plant to continue with the next step in C_4 evolution—the establishment of proto-Kranz anatomy, including the enhancement and activation of the bundle sheath. This is a very important step since it only takes a few key changes to set the basis for all subsequent steps toward C_4 (Westhoff and Gowik, 2010; Sage et al., 2012). In some C_3 species that are closely related to the C_2 species, bundle sheath cells are enlarged and contain more organelles than other C_3 species (Sage et al., 2012). Most mitochondria in the bundle sheath cells are already positioned at the inner centripetal wall and some photosynthetically active chloroplasts are located right beside them. This specific positioning of organelles could establish a kind of single-cell glycine shuttle in the bundle sheath (Muhaidat et al., 2011). Consequently, photorespiratory glycine has to move to the mitochondria at the inner bundle sheath periphery to continue the photorespiratory pathway. Due to larger vacuoles in the bundle sheath, the released CO_2 would be trapped and easily re-assimilated by adjacent chloroplasts (Sage et al., 2014). This arrangement can produce a glycine sink in which excessive glycine, produced in the mesophyll under high rates of photorespiration, moves to the bundle sheath to get decarboxylated. This would lead to higher CO_2 concentrations around RubisCO and increase its efficiency. Under these conditions, plants are ready to enter the third phase of C_4 evolution: the establishment of a photorespiratory CO_2 pump.

The formation of the proto-Kranz anatomy can already improve RubisCO efficiency by reducing the CO_2 compensation point (Sage et al., 2012). As more and more photorespiratory glycine gets decarboxylated in the bundle sheath, a loss of the mesophyll GDC function is not critical to the plant any longer, whereas in true C_3 plants this would create strong photorespiratory phenotypes. Now that GDC activity is restricted to the bundle sheath, C_2 photosynthesis is introduced with the necessity of a tight interplay between bundle sheath and mesophyll cells. In this way, CO_2 lost through photorespiration becomes a new resource that can be efficiently refixated in the bundle sheath by RubisCO (Bauwe, 2010). Consequently, the negative effect of photorespiration is significantly lowered (Monson et al., 1984). Once the basic CO_2 shuttle is established, further anatomical changes ensure the optimization of this process, which includes a reduction in mesophyll cell size and an increase in bundle sheath organelle number and size (Brown et al., 1983; Brown and Hattersley, 1989; Muhaidat et al., 2011; Sage et al., 2012). The outcome of this is a tight connection between mesophyll and bundle sheath cells, and, thus, a first version of the Kranz anatomy. The relocation of GDC expression from the mesophyll to the bundle sheath and the initiation of C_2 photosynthesis is considered a key step in the evolution of

C_4 photosynthesis and proceeded gradually (Monson and Rawsthorne, 2000; Schulze et al., 2013; Mallmann et al., 2014). The photorespiratory CO_2 pump does not only accumulate CO_2 in the bundle sheath, but it also creates a nitrogen imbalance between bundle sheath and mesophyll cells that needs to be fixed. Mallmann et al. (2014) hypothesize that the initial steps of C_4 photosynthesis evolved to recirculate photorespiratory ammonia from the bundle sheath to the mesophyll cells to prevent ammonia accumulation in the bundle sheath. As a side effect, important C_4 components were established and could be used to evolve a functional C_4 cycle, which is described in the fourth phase: the up-regulation of the C_4 cycle.

This phase mainly includes the compartmentalization of most C_4 cycle enzymes into either bundle sheath or mesophyll, which is mostly achieved by differential gene expression (Wyrich et al., 1998; Westhoff and Gowik, 2010). To establish an operating C_4 photosynthesis, the spatial separation of the two carboxylating enzymes PEPC and RubisCO is of key importance (Gowik and Westhoff, 2011). PEPC has to be expressed in the mesophyll and RubisCO in the bundle sheath. The mechanism of differentially expressed genes can be explained by changes in their cis-regulatory elements or by alterations of trans-acting factors, such as transcription factors (Doebley and Lukens, 1998; Brown et al., 2011). It has been shown that even single changes within one small cis-regulatory element can lead to mesophyll-specific expression of PEPC (Akyildiz et al., 2007). The establishment of a glycine shuttle generates a gradient for CO_2 efflux driven by high CO_2 concentrations in the bundle sheath cell (Caemmerer, 2000). PEPC activity can now rise in the mesophyll to refix some of the CO_2, which escapes the bundle sheath. With increasing PEPC activity, the expression of other C_4 cycle enzymes needed for regeneration of PEP can also be elevated (Sage, 2004).

A loss of RubisCO in the mesophyll reduces the selection pressure for high expression of photorespiratory genes, because there is no longer oxygenase activity in the mesophyll, while in the bundle sheath it is largely suppressed due to a more efficient RubisCO (Bräutigam and Gowik, 2016). Consequently, most photorespiratory genes become bundle sheath-specific in C_4 plants as there is no need for photorespiration in the mesophyll. This could be shown in highly optimized C_4 grasses like *Zea mays*, *Sorghum bicolor,* or *Setaria italica*, all of which contain a fully developed C_4 cycle (Li et al., 2010; John et al., 2014; Döring et al., 2016). Once a fully functional C_4 cycle is introduced, only optimization is needed in phase five to tap the full potential of C_4 photosynthesis. This includes fine tuning of kinetics and regulation of many enzymes to optimize photosynthetic efficiency (Sage, 2004).

2.4 The need for more C_4 plants to feed the world

Around 7.5 billion people live on earth in the year 2016. The number is constantly rising and it is estimated to reach the 10 billion mark around 2060 (United Nations, Department of Economic and Social Affairs, 2015). Already, poverty and malnutrition are big issues worldwide, and demands on food will increase further in the future. Consequently, agricultural yields have to be boosted and it is estimated that cereal production alone must increase by 50% by 2030 to meet projected demands on food (Covshoff and Hibberd, 2012). This can only be achieved if photosynthesis in plants is substantially improved to use solar energy more efficiently. As shown by nature, C_4 plants can be the solution due to their high radiation-use efficiency (RUE). C_4 photosynthesis outperforms C_3 photosynthesis in environments where CO_2 is a limiting factor, but it is only used by about 3% of all vascular plant species (Edwards et al., 2010). Thus, most crops are still using the C_3 pathway and waste a lot of potential by doing so. Theoretical models predict that the RUE of C_3 plants could be improved by up to 50% by using the C_4 pathway (Mitchell and Sheehy, 2006; Zhu et al., 2010). Thus, converting C_3 crops into C_4 crops can be a solution and many groups worldwide focus their work on how to artificially speed up evolution of C_4 photosynthesis in C_3 plants (Reynolds et al., 2011).

The integration of a fully functional C_4 cycle requires many small and large changes. Fortunately, all enzymes involved in C_4 photosynthesis are already present in C_3 plants, where they function in carbohydrate and nitrogen metabolism (Aubry et al., 2011; Brown et al., 2011). Conversion from C_3 to C_4, therefore, does not need the introduction of completely new players, but rather changes in regulation, kinetics, and tissue-specificity of existing enzymes and their corresponding genes (Doebley and Lukens, 1998; Sage, 2004; Schuler et al., 2016). Working groups all over the world try to contribute small parts of the big puzzle, thus creating a large toolbox that will eventually help to assemble all pieces to a functional C_4 pathway. Activation of the bundle sheath, which mainly comprises larger bundle sheath cells and more organelles inside, is only one, but an extremely important, piece of this box.

3. Forward genetic screens in the model plant *Arabidopsis thaliana* to identify bundle sheath mutants

3.1 The rise of the model species A. thaliana

Arabidopsis thaliana (*A. thaliana*) is a small dicotyledonous plant that belongs to the Brassicaceae and is known as the most important model system in plant biology. Arabidopsis is

not grown for food, feed, or fiber. Its flowers are small and inconspicuous, with little attraction. So how can it be that a weed with no commercial value is in the focus of so many studies worldwide? To answer this question, it helps to go back in time and study the situation of the emerging field of molecular biology in plants in the 1970s. At that time, many plants were considered as model genetic systems. These included maize, tomato, pew, rice, and barley (Meinke et al., 2002). For a long time, no plant was chosen for studying general processes that are relevant to all plants, thus leading to duplication of effort and a lack of shared resources (Meinke et al., 2002). In the late 1970s and early 1980s, the researchers David Meinke, Maarten Koorneef, and Chris Summerville, amongst others, essentially discovered the huge potential of the small angiosperm *A. thaliana* as a model species in plant biology (Figure 5). This species was first chosen as a genetic model organism by Laibach in the first half of the 20th century (Laibach, 1943). The combination of many traits and features makes Arabidopsis the perfect organism for studying the general processes of plants in Arabidopsis first, before applying the knowledge to other species.

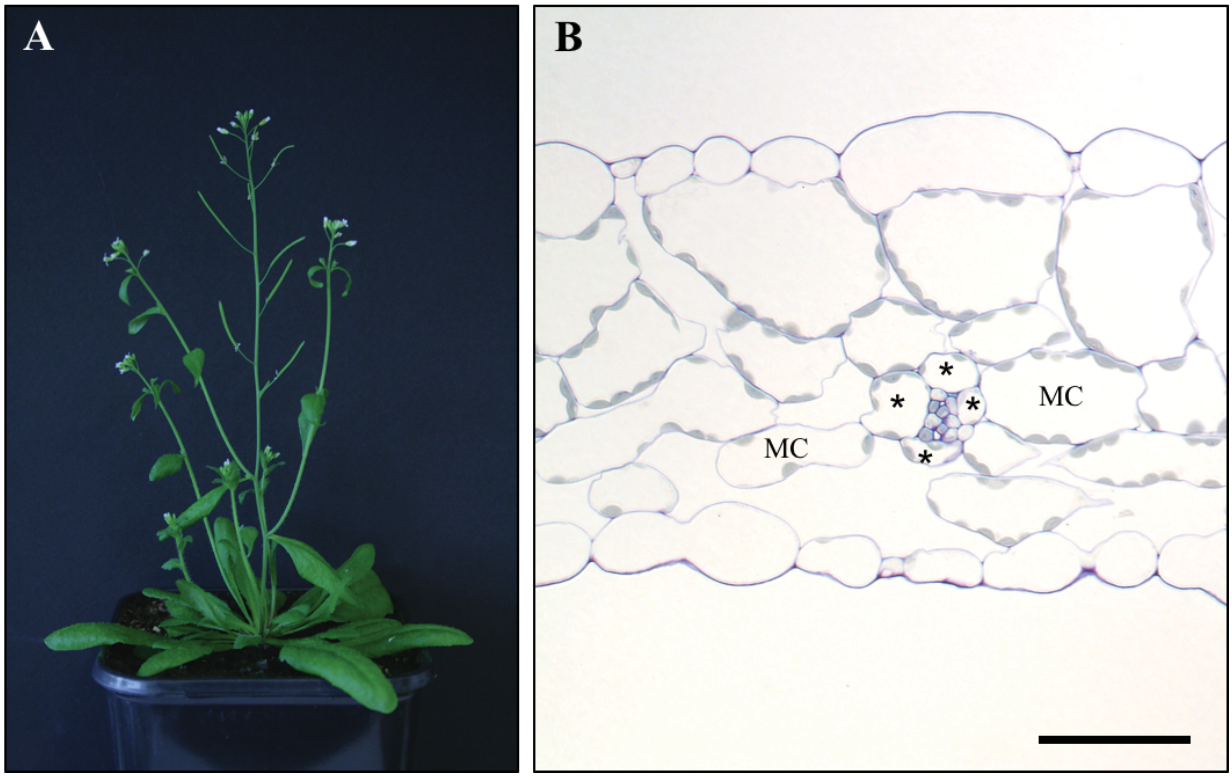


Figure 5. The model plant *A. thaliana.*

(A) A mature seven-week-old *A. thaliana* wild type (Ecotype Columbia) with developed siliques and flowers. (B) Cross section of an Arabidopsis leaf. Bundle sheath cells are indicated with an asterisk. MC, mesophyll cell. Scale bar = 50µm.

Firstly, Arabidopsis has a small size and does not need much place to grow, which enables large-scale studies with thousands of plants. The generation time is remarkably short, with only five to six weeks needed under optimal conditions. Arabidopsis can easily grow on soil, on media in petri dishes, or in liquid media (Siedlecka and Krupa, 2002). In terms of reproduction, some features make Arabidopsis very popular among biologists, such as its high fecundity (up to 10,000 seeds per plant) and the possibility of self-fertilization as well as creating outcrosses with other ecotypes. Arabidopsis is genetically simple and accessible. It has a small genome size of 125 Mbp and only 27,500 genes, which is moderate compared to many crops. Maize, for example, has a genome size of approximately 240 Mbp, which is mostly due to whole genome duplication events that did not occur in Arabidopsis. One of the most outstanding qualities of Arabidopsis is the simplicity of generating transgenic plants via *Agrobacterium*-mediated gene transfer. By simply dipping floral parts into an *Agrobacterium* solution containing the transgene, genetic material can be reliably transferred. In the year 2000, the genome of Arabidopsis was completely sequenced and the list of available mutant lines for almost all genes is large and still being extended (The Arabidopsis Genome Initiative, 2000). To sum this up, the genetic resources available are huge and of great importance for nearly every process studied in Arabidopsis. Whenever something interesting is revealed in the model plant, the new insights can help to understand equivalent situations in other plants of interest, such as crops that are economically more important.

3.2 A. thaliana contains rudimentary bundle sheath cells

Large and pronounced bundle sheath cells are a key feature of C_4 plants. They are tightly packed with chloroplasts and mitochondria, ensuring high photosynthetic rates. In short, the bundle sheath is of essential value for C_4 plants. But what do they look like in C_3 plants and what is their function in a non-C_4 plant? For a long time, C_3 bundle sheath cells received only a little attention, while their C_4 counterparts have been well-characterized (Dengler et al., 1985; Langdale et al., 1989). In the late 1990s, Kinsman and Pyke took a closer look at the bundle sheath of *A. thaliana* and revealed that bundle sheath cells contribute around 15% of the chloroplast-containing cells in the leaves. They are smaller than mesophyll cells and contain significantly fewer chloroplasts. While the average number of chloroplasts per cell is 76 in the mesophyll, it is only 22 in the bundle sheath (Kinsman and Pyke, 1998). This clearly shows the big difference in photosynthetic capacity in this cell type between C_3 and C_4. Despite its moderate contribution to leaf photosynthesis, the bundle sheath must have other functions in Arabidopsis. Since it directly links the vasculature and the mesophyll, it has an important role in transport of water and

photoassimilates, and it might also contribute to the mechanical strength within the leaf (Van Bel, 1993; Kinsman and Pyke, 1998). Most bundle sheath cells in Arabidopsis have parallel and elongated sidewalls that are in close contact to the veins. Longitudinal sections of developing vascular strands can reveal that bundle sheath cells divide periclinally and anticlinally to the adjacent vein. Thus, it appears that cell differentiation is determined by position rather than by a distinct cell lineage (Kinsman and Pyke, 1998).

In a nutshell, the present bundle sheath of Arabidopsis is completely C_3 and shows no characteristics of the C_4 equivalent. Nevertheless, experiments using bundle sheath promoters from the Asteracean C_4 species *Flaveria trinervia* showed that the expression specificities are maintained also in C_3 Arabidopsis (Engelmann et al., 2008). This indicates that there are already trans-acting factors present in the C_3 bundle sheath of Arabidopsis that are able to correctly recognize heterologous C_4-characteristic cis-regulatory elements. This provides further evidence that evolution from C_3 to C_4 must have been relatively easy in genetic terms. Unfortunately, very little is known so far about the specific genes that control size and organelle number of the bundle sheath, genes that could be manipulated in order to get a more C_4-like bundle sheath in C_3 plants. This lack of information requires a simple method to identify new genes and to link them to the desired process. Forward genetic approaches with large-scale screens are powerful tools for gathering new information by using the model plant *A. thaliana*.

3.3 EMS and activation tagging screens are powerful tools for identifying new genes in A. thaliana

Since the genome of Arabidopsis is completely sequenced, genetic screens are of great importance to assign specific functions to the ~25,000 genes. Reverse genetic approaches are of great help to identify the function of particular candidate genes, but they are biased and do not take all genes into account, whereas forward genetics allows the directed and unbiased analysis of any given process of interest. The principle of a forward genetic screen is to randomly mutate genes in the whole genome of the plant, which often creates large numbers of mutations that may or may not lead to an aberrant phenotype. In most cases, thousands of mutants are analyzed for the phenotype of interest and, once an interesting mutant is found, the phenotype needs to be linked to a gene, which is called “mapping of the gene” (Page and Grossniklaus, 2002).

There are three ways to induce a large number of mutations in the plant genome: by (1) chemical agents such as ethyl methanesulfonate (EMS), (2) radiation-based methods like *y*-ray and fast neuron bombardment, or (3) biological agents such as transposons and T-DNA (Sikora et al., 2011; Serrat et al., 2014). Chemical agents are often used since they have a high

mutation frequency, are reliable, and are easy to use. EMS is by far the most popular among them. It alkylates guanine bases, forming O^6-ethylguanine, which can pair with thymine residues but not with cytosine residues during DNA replication. Thus, a single-nucleotide polymorphism (SNP) is generated by the transition of a G/C base pair into an A/T base pair (Kim et al., 2006). Since EMS-induced SNPs are completely random among all G/C base pairs in the whole genome, the result of it can be manifold. If a substitution takes place in the coding region of a gene, it may lead to a non-synonymous mutation and change the amino acid at this position or even induce a premature stop codon. It may also result in a synonymous mutation with mostly no effect at all. In non-coding regions, mutations may still change regulatory sequences such as gene promoters, which can affect gene transcription in a positive or negative manner. It might also interfere with splice sites, resulting in aberrant splicing or altered mRNA stability (Sikora et al., 2011). To conclude, EMS mutagenesis is a powerful tool in forward genetics because of its diversity in changing the activity of a gene. While it is comparatively easy to generate numerous interesting mutant lines, the drawback of EMS is that genetic mapping of the responsible mutated gene can be quite tedious. In the 1990s, the total effort of map-based cloning was three to five years because there was no physical map available in Arabidopsis. The sequenced Arabidopsis genome in the early 2000s greatly reduced the time required, but it still took up to one year to isolate a gene until recently, when the combination of bulk segregant analysis and whole-genome resequencing has been shown to speed up this process to a large extent (Jander et al., 2002; Schneeberger et al., 2009). Decreasing costs of genome sequencing in general will favor this approach even more.

Besides the mutagenic compound EMS, small DNA fragments such as T-DNA or transposable elements can also serve as mutagens by simply being inserted at a random position in the genome with the chance to land within the coding region of a gene, which, in most cases, disrupts the gene function resulting in a complete knockout (Bhatt et al., 1996). However, loss-of-function screens often fail to identify redundant genes and those genes that are indispensable for the developing plant and whose knockout results in early embryonic or in gametophytic lethality (Weigel et al., 2000). An activation tagging screen, as a modified version of a classical T-DNA insertion mutant screen, can solve this problem by combining loss-of-function with gain-of-function mutations. It is based on a strong promoter sequence on the T-DNA itself that can activate and overexpress nearby genes. By the use of cell-specific promoters, it can also be limited to separate compartments of interest.

While the mutation frequency is rather high in EMS-based genetic screens and the mapping of genes is still laborious, the reverse situation can be seen in insertional mutagenesis screens.

Depending on the EMS concentration, mutation frequencies ranging from 1/300 to 1/30,000 per locus have been reported (Page and Grossniklaus, 2002), whereas T-DNA insertion results in only one or a few insertions per plant (Does et al., 1991). With a median gene length of 2.1 kb in Arabidopsis, it requires ~180,000 T-DNA inserts to have a 95% chance of mutating a specific gene (Bhatt et al., 1996). On the other hand, the insertion point of the T-DNA can be easily mapped using PCR-based methods such as thermal asymmetric interlaced PCR (TAIL-PCR) or inverse PCR (iPCR) (Liu and Whittier, 1995; McKinney et al., 1995). Thus, a combination of EMS mutagenesis and activation tagging could combine the advantages of both screens to uncover more bundle sheath developmental genes in Arabidopsis.

II. Scientific aims

C_4 photosynthesis leads to a concentration of CO_2 around the main carboxylating enzyme RubisCO, which results in an effective suppression of its oxygenase activity. Thus, photorespiration is largely reduced in C_4 plants and most enzymes are shifted to the bundle sheath. Since the C_4 pathway is more efficient than C_3 photosynthesis under high light, high temperature, and the present CO_2 concentration in the atmosphere, it is an ambitious goal to introduce this superior type of photosynthesis into existing C_3 plants. One of the key points in the early steps of C_4 evolution is the activation of the bundle sheath. So far, there is little information about enzymes and their genes that can be manipulated in order to design a more C_4-like bundle sheath in C_3 plants. Forward genetic screens might be the key to identify genes involved in the development and activation of the bundle sheath.

(1) Kranz anatomy and the division of labor between two different cell types, the bundle sheath and the mesophyll, are the hallmarks of C_4 photosynthesis. The enzymes of the core C_4 pathway are strictly divided between both cell types. Although photorespiration is largely reduced in C_4 plants, even there it is still an essential process. However, the photorespiratory cycle is only needed in the bundle sheath since in C_4 plants RubisCO is missing from the mesophyll cells. It is interesting to see to what extent C_4 plants shifted the expression of genes related to photorespiration to the bundle sheath after the establishment of true C_4 photosynthesis. In this study, the cell-specific expression of photosynthetic and photorespiratory genes was examined in *Sorghum bicolor*, a highly optimized plant species with regard to the C_4 pathway. This was achieved by RNA *in situ* hybridization, quantitative real-time PCR, and transcriptome analysis of isolated mesophyll and bundle sheath cells (Manuscript 1: Döring et al., 2016).

(2) Larger bundle sheath cells and more organelles inside are needed in most C_3 plants before a C_4 cycle can be established. In order to identify genes that are involved in these processes, two large-scale forward genetic screens were established in the model plant *Arabidopsis thaliana*. Since bundle sheath cells are rather small and inconspicuous in the C_3 Brassicaceaen species, a reporter gene construct was introduced to specifically express luciferase (LUC) or the green fluorescent protein (GFP) in the bundle sheath of transgenic plants. This was accomplished by the use of the promoter of the *GLDPA* gene from the C_4 Asteraceaen species *Flaveria trinervia*. This promoter is highly active in the bundle sheath and vascular tissue of *A. thaliana* leaves. With EMS mutagenesis and activation tagging, numerous mutants were generated and screened for

differences in reporter gene expression. After verification of the aberrant phenotype in subsequent plant generations, the most promising mutant lines were analyzed in more detail with high-resolution microscopy (Manuscript 2+3: Döring et al., 2017, unpublished).

III.A Summary

In the presence of high CO_2 concentrations, the oxygenase activity of RubisCO is effectively suppressed, and hence, photorespiration is strongly reduced in C_4 plants. Nevertheless, small amounts of 2-PG are produced in the bundle sheath of C_4 plants that need to be detoxified by the photorespiratory pathway.

In this study, it was analyzed whether the expression of genes related to photorespiration becomes bundle sheath-specific in the fully optimized C_4 species *Sorghum bicolor*. By the use of transcriptome analysis, qPCR, and *in situ* hybridization, it was shown that all genes of the core photorespiratory cycle are at least preferentially, if not specifically, expressed in the bundle sheath, except for GLYK, which is expressed to a much higher degree in the mesophyll than in the bundle sheath. These results support the assumption that the photorespiratory pathway is shifted almost completely to the bundle sheath in plants after they established true C_4 photosynthesis. The reduction and exclusion of most photorespiratory genes from the mesophyll represents an optimization and enhances the nitrogen-use efficiency of the C_4 grass *Sorghum bicolor*.

The evolution of C_4 photosynthesis proceeds step by step and each small change contributes to the general fitness of the plant. Large and organelle-rich bundle sheath cells are a requirement before a CO_2 shuttle toward the bundle sheath can develop. However, the genetic basis of these changes remains unknown so far.

In this study, two forward genetic screens (EMS and activation tagging) were established in *Arabidopsis thaliana* to identify genes involved in bundle sheath ontogeny and maintenance (*BSOM* genes). Two reporter gene lines that express either luciferase or chloroplast-targeted GFP in the bundle sheath and vasculature were created and used as a genetic background in both mutant screens. Differences in the reporter gene signal in Arabidopsis leaves were used as a proxy to isolate mutants with potentially altered bundle sheath anatomy. It was shown that both screens resulted in numerous primary mutants with increased or decreased reporter gene activity. High-resolution light microscopy and transmission electron microscopy could reveal that the bundle sheath anatomy is affected in four EMS-generated mutant lines and activation tagging mutant line AT-2. Furthermore, *BSOM2,* a gene that encodes an ABA efflux transporter, was mapped from activation tagging line AT-2. It was shown that both the overexpression and knockout of *BSOM2* results in more bundle sheath cells and an enlarged vasculature.

III.B Zusammenfassung

Unter hohen CO_2-Konzentrationen kann die Oxygenaseaktivität von RubisCO weitestgehend unterdrückt werden. Dies hat zur Folge, dass die Photorespiration in C_4-Pflanzen stark reduziert wird. Dennoch entstehen auch in den Bündelscheidenzellen der C_4-Pflanzen geringe Mengen an 2-PG, welche daraufhin über den photorespiratorischen Kreislauf entgiftet werden müssen.

In dieser Arbeit wurde untersucht, ob die Gene, die in Verbindung mit der Photorespiration stehen, in der voll-optimierten C_4-Pflanze *Sorghum bicolor* ein bündelscheidenspezifisches Expressionsmuster aufweisen. Mit Hilfe von Transkriptomanalysen, qPCR und *in situ*-Hybridisierung konnte gezeigt werden, dass alle Gene des eigentlichen Photorespirationszyklus zumindest präferentiell, wenn nicht sogar spezifisch in den Bündelscheidenzellen exprimiert werden, mit Ausnahme von GLYK, welches wesentlich stärker im Mesophyll als in der Bündelscheide exprimiert wird. Diese Ergebnisse bekräftigen die Annahme, dass der Prozess der Photorespiration in Pflanzen, die bereits ausschließlich C_4-Photosynthese betreiben, fast vollständig in die Bündelscheidenzellen verlagert wird. Die Drosselung oder Deaktivierung vieler photorespiratorischer Gene im Mesophyll ist Teil der Optimierung der C_4-Photosynthese in *Sorghum bicolor* und verbessert unter anderem die Stickstoffnutzungseffizienz.

Die Evolution der C_4-Photosynthese verläuft schrittweise, wobei jede kleine Veränderung der Pflanze einen Vorteil bringt. Große und organellenreiche Bündelscheidenzellen sind eine Grundvoraussetzung für die Etablierung eines CO_2-Konzentrierungsmechanismus in Richtung Bündelscheidenzelle. Allerdings ist die genetische Grundlage hierfür noch weitestgehend unbekannt.

Im Rahmen dieser Arbeit wurden zwei vorwärts gerichtete genetische Verfahren (EMS-Mutagenese und „*activation tagging*“) in der C_3-Pflanze *Arabidopsis thaliana* etabliert, um Gene für die Ontogenie und Erhaltung der Bündelscheide (*BSOM:* „*bundle sheath ontogony and maintenance“*) zu identifizieren. Es wurden zwei Reportergenlinien etabliert, welche entweder Luciferase oder Chloroplasten-lokalisiertes GFP in den Bündelscheidenzellen und dem Leitgewebe der Blätter exprimieren. Diese Linien dienten als genetischer Hintergrund für beide Mutageneseansätze. Mutanten mit potentiell veränderter Anatomie der Bündelscheidenzellen wurden anhand von Unterschieden in der Signalstärke der Reportergene in den untersuchten Mutanten selektiert. In beiden vorwärts gerichteten genetischen Verfahren konnten zahlreiche Mutanten mit verstärkter oder abgeschwächter Reportergenaktivität identifiziert werden. Durch hochauflösende Licht- und Transmissionselektronenmikroskopie konnte gezeigt werden, dass die Anatomie der Bündelscheide und des Leitgewebes in vier EMS-generierten Mutantenlinien und

in der „*activation tagging*“-Linie AT-2 verändert ist. Des Weiteren wurde *BSOM2*, ein Gen, welches für einen ABA-Effluxtransporter kodiert, aus der Mutante AT-2 als verantwortliches Gen identifiziert. Es konnte experimentell nachgewiesen werden, dass sowohl die Überexpression als auch der komplette Verlust von *BSOM2* jeweils zu mehr Bündelscheidenzellen und mehr Leitgewebe in den Blättern der Pflanzen führt.

IV. Literature

Akyildiz M, Gowik U, Engelmann S, Koczor M, Streubel M, 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**: 3391–402

Alberts B, Johnson A LJ (2002) Molecular Biology of the Cell. 4th edition. Garland Science, New York

Aliscioni S, Bell HL, Besnard G, Christin PA, Columbus JT, Duvall MR, Edwards EJ, Giussani L, Hasenstab-Lehman K, Hilu KW, et al (2012) New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. New Phytol **193**: 304–312

Aubry S, Brown NJ, Hibberd JM (2011) The role of proteins in C3 plants prior to their recruitment into the C4 pathway. J Exp Bot **62**: 3049–3059

Bassham, AJ, Benson, AA, Calvin M (1950) The path of carbon in photosynthsis. J Biol Chem **185**: 781–7

Bauwe H (2010) Chapter 6 Photorespiration: The bridge to C4 photosynthesis. C4 Photosynth. Relat. CO2 Conc. Mech. Springer Netherlands, pp 81–108

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

Bauwe H, Hagemann M, Kern R, Timm S (2012) Photorespiration has a dual origin and manifold links to central metabolism. Curr Opin Plant Biol **15**: 269–275

Van Bel AJE (1993) Strategies of phloem loading. Annu Rev Plant Physiol Plant Mol Biol **44**: 253–281

Besnard G, Christin P-A, Malé P-JG, Lhuillier E, Lauzeral C, Coissac E, Vorontsova MS (2014) From museums to genomics: old herbarium specimens shed light on a C3 to C4 transition. J Exp Bot **65**: 6711–21

Besnard G, Muasya AM, Russier F, Roalson EH, Salamin N, Christin PA (2009) Phylogenomics of C4 photosynthesis in sedges (Cyperaceae): Multiple appearances and genetic convergence. Mol Biol Evol **26**: 1909–1919

Bhatt AM, Page T, Lawson EJR, Lister C, Dean C (1996) Use of Ac as insertional mutagen in Arabidopsis. Plant J **9**: 935–945

Botha CEJ (1992) Plasmodesmatal distribution, structure and frequency in relation to assimilation in C3 and C4 grasses in southern Africa. Planta **187**: 348–58

Bowes G, Ogren WL, Hageman RH (1971) Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. Biochem Biophys Res Commun **45**: 716–722

Bräutigam A, Gowik U (2016) Photorespiration connects C3 and C4 photosynthesis. J Exp Bot **67**: 2953–2962

Brooks A, Farquhar GD (1985) Effect of temperature on the CO2/O2 specificity of ribulose-1,5-bisphosphate carboxylase/oxygenase and the rate of respiration in the light - Estimates from gas-exchange measurements on spinach. Planta **165**: 397–406

Brown NJ, Newell CA, Stanley S, Chen JE, Perrin AJ, Kajala K, Hibberd JM (2011) Independent and parallel recruitment of preexisting mechanisms underlying C4 photosynthesis. Science **331**: 1436–1439

Brown RH, Bouton JH, Rigsby L, Rigler M (1983) Photosynthesis of grass species differing in carbon dioxide fixation pathways : VIII. Ultrastructural characteristics of Panicum species in the Laxa Group. Plant Physiol **71**: 425–31

Brown RH, Hattersley PW (1989) Leaf anatomy of C3-C4 species as related to evolution of C4 photosynthesis. Plant Physiol **91**: 1543–1550

Buick R (2008) When did oxygenic photosynthesis evolve? Philos Trans R Soc B Biol Sci **363**: 2731–2743

Caemmerer S Von (2000) Biochemical models of leaf photosynthesis. Tech Plant Sci **53**: 1689–1699

Carvalho J de F, Madgwick PJ, Powers SJ, Keys AJ, Lea PJ, Parry MA, Keys A, Bird I, Cornelius M, Lea P, et al (2011) An engineered pathway for glyoxylate metabolism in tobacco plants aimed to avoid the release of ammonia in photorespiration. BMC Biotechnol **11**: 111

Christin P-A, Salamin N, Kellogg EA, Vicentini A, Besnard G (2009) Integrating phylogeny into studies of C4 variation in the grasses. Plant Physiol **149**: 82–87

Christin P-A, Wallace MJ, Clayton H, Edwards EJ, Furbank RT, Hattersley PW, Sage RF, Macfarlane TD, Ludwig M (2012) Multiple photosynthetic transitions, polyploidy, and lateral gene transfer in the grass subtribe Neurachninae. J Exp Bot **63**: 6297–308

Covshoff S, Hibberd JM (2012) Integrating C4 photosynthesis into C3 crops to increase yield potential. Curr Opin Biotechnol **23**: 209–214

Dengler NG, Dengler RE, Hattersley PW (1985) Differing ontogenetic origins of PCR ("Kranz") sheaths in leaf blades of C4 Grasses (Poaceae). Am J Bot **72**: 284–302

Dengler NG, Nelson T (1999) Leaf structure and development in C4 plants. C4 Plant Biol. pp 133–172

Doebley J, Lukens L (1998) Transcriptional regulators and the evolution of plant form. Plant Cell **10**: 1075–82

Does MP, Dekker BMM, de Groot MJA, Offringa R (1991) A quick method to estimate the T-DNA copy number in transgenic plants at an early stage after transformation, using inverse PCR. Plant Mol Biol **17**: 151–153

Döring F, Streubel M, Bräutigam A, Gowik U (2016) Most photorespiratory genes are preferentially expressed in the bundle sheath cells of the C4 grass Sorghum bicolor. J Exp Bot **67**: 3053–3064

Edwards EJ, Osborne CP, Stromberg CAE, Smith SA, Bond WJ, Christin PA, Cousins AB, Duvall MR, Fox DL, Freckleton RP, et al (2010) The origins of C4 grasslands: Integrating evolutionary and ecosystem science. Science **328**: 587–591

Ehleringer J, Björkman O, Bjorkman O (1977) Quantum yields for CO2 uptake in C3 and C4 plants. Plant Physiol **59**: 86–90

Ehleringer JR, Cerling TE, Helliker BR (1997) C4 photosynthesis, atmospheric CO2, and climate. Oecologia **112**: 285–299

Ehleringer JR, Sage RF, Flanagan LB, Pearcy RW (1991) Climate change and the evolution of C4 photosynthesis. Trends Ecol Evol **6**: 95–9

Engel N, van den Daele K, Kolukisaoglu U, Morgenthal K, Weckwerth W, Pärnik T, Keerberg O, Bauwe H, Parnik T, Keerberg O, et al (2007) Deletion of glycine decarboxylase in Arabidopsis is lethal under nonphotorespiratory conditions. Plant Physiol **144**: 1328–35

Engelmann S, Wiludda C, Burscheidt J, Gowik U, Schlue U, Koczor M, Streubel M, Cossu R, Bauwe H, Westhoff P (2008) The gene for the P-subunit of glycine decarboxylase from the C4 species Flaveria trinervia: analysis of transcriptional control in transgenic Flaveria bidentis (C4) and Arabidopsis (C3). Plant Physiol **146**: 1773–1785

Evans JR (2013) Improving photosynthesis. Plant Physiol **162**: 1780–93

Fernie AR, Bauwe H, Eisenhut M, Florian A, Hanson DT, Hagemann M, Keech O, Mielewczik M, Nikoloski Z, Peterhänsel C, et al (2013) Perspectives on plant photorespiratory metabolism. Plant Biol **15**: 748–753

Furbank RT (2011) Evolution of the C4 photosynthetic mechanism: Are there really three C4 acid decarboxylation types? J Exp Bot **62**: 3103–3108

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

Hagemann M, Bauwe H (2016) Photorespiration and the potential to improve photosynthesis. Curr Opin Chem Biol **35**: 109–116

Hatch M (1987) C4 photosynthesis: a unique blend of modified biochemistry, anatomy abd ultrastructure. Biochim Biophys Acta **895**: 81–106

Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber APM, Lercher MJ (2013) Predicting C4 photosynthesis evolution: Modular, individually adaptive steps on a mount fuji fitness landscape. Cell **153**: 1579–88

Hibberd JM, Sheehy JE, Langdale JA (2008) Using C4 photosynthesis to increase the yield of rice-rationale and feasibility. Curr Opin Plant Biol **11**: 228–231

Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL, Llc CG, Street S (2002) Arabidopsis map-based cloning in the post-genome era. Plant Physiol **129**: 440–450

John CR, Smith-Unna RD, Woodfield H, Covshoff S, Hibberd JM (2014) Evolutionary convergence of cell-specific gene expression in independent lineages of C4 grasses. Plant Physiol **165**: 62–75

Jordan DB, Ogren WL (1984) The CO2/O2 specificity of ribulose 1,5-bisphosphate carboxylase/oxygenase - Dependence on ribulosebisphosphate concentration, pH and temperature. Planta **161**: 308–313

Kadereit G, Ackerly D, Pirie MD (2012) A broader model for C_4 photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae s.s.). Proceedings Biol Sci **279**: 3304–11

Kasting JF, Howard MT (2006) Atmospheric composition and climate on the early earth. Philos Trans R Soc B Biol Sci **361**: 1733–1742

Kebeish R, Niessen M, Thiruveedhi K, Bari R, Hirsch H-J, Rosenkranz R, Stäbler N, Schönfeld B, Kreuzaler F, Peterhänsel C (2007) Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in Arabidopsis thaliana. Nat Biotechnol **25**: 593–599

Keys AJ (2006) The re-assimilation of ammonia produced by photorespiration and the nitrogen economy of C3 higher plants. Photosynth Res **87**: 165–175

Kim Y, Schumaker KS, Zhu J-K (2006) EMS mutagenesis of Arabidopsis. Methods Mol Biol **323**: 101–103

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

Laetsch WM (1974) The C4 syndrome: A structural analysis. Annu Rev Plant Physiol **25**: 27–52

Laibach F (1943) Arabidopsis thaliana (L.) Heynh. Als Objekt für genetische und entwicklungsphysiologische Untersuchungen. Bot Arch **44**: 439–455

Langdale JA, Lane B, Freeling M, Nelson T (1989) Cell lineage analysis of maize bundle sheath and mesophyll cells. Dev Biol **133**: 128–39

Larridon I, Bauters K, Reynders M, Huygh W, Muasya AM, Simpson DA, Goetghebeur P (2013) Towards a new classification of the giant paraphyletic genus Cyperus (Cyperaceae): Phylogenetic relationships and generic delimitation in C4 Cyperus. Bot J Linn Soc **172**: 106–126

Leegood RC (2002) C4 photosynthesis: principles of CO2 concentration and prospects for its introduction into C3 plants. J Exp Bot **53**: 581–590

Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR, et al (2010) The developmental dynamics of the maize leaf transcriptome. Nat Genet **42**: 1060–1067

Liu Y-GG, Whittier RF (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics **25**: 674–681

Long SP, Marshall-Colon A, Zhu XG (2015) Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. Cell **161**: 56–66

Lorimer GH (1981) The carboxylation and oxygenation of ribulose 1,5-biphoshate: The primary events in photosynthesis and photorespiration. Annu Rev Plant Physiol Plant Mol Biol **32**: 349–383

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

Marcus Y, Altman-Gueta H, Wolff Y, Gurevitz M (2011) Rubisco mutagenesis provides new insight into limitations on photosynthesis and growth in Synechocystis PCC6803. J Exp Bot **62**: 4173–4182

Marshall DM, Muhaidat R, Brown NJ, Liu Z, Stanley S, Griffiths H, Sage RF, Hibberd JM (2007) Cleome, a genus closely related to Arabidopsis, contains species spanning a developmental progression from C3 to C4 photosynthesis. Plant J **51**: 886–896

McKinney EC, Ali N, Traut A, Feldmann KA, Belostotsky DA, McDowell JM, Meagher RB (1995) Sequence-based identification of T-DNA insertion mutations in Arabidopsis: actin mutants act2-1 and act4-1. Plant J **8**: 613–622

McKown AD, Dengler NG (2007) Key innovations in the evolution of Kranz anatomy and C4 vein pattern in Flaveria (Asteraceae). Am J Bot **94**: 382–99

Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M (2002) Arabidopsis thaliana: A model plant for genome analysis. Science **282**: 662–682

Mitchell PL, Sheehy JE (2006) Supercharging rice photosynthesis to increase yield. New Phytol **171**: 688–93

Monson RK (2003) Gene duplication, neofunctionalization, and the evolution of C4 photosynthesis. Int J Plant Sci Int J Plant Sci **1643**: 43–54

Monson RK, Edwards GE, Ku MSB (1984) C3-C4 Intermediate photosynthesis in plants. Bioscience **34**: 563-566-574

Monson RK, Rawsthorne S (2000) CO2 assimilation in C3-C4 intermediate plants. Photosynth. Physiol. Metab. Springer Netherlands, pp 85–113

Moore PD (1983) Plants and the palaeoatmosphere. J Geol Soc London **140**: 13–25

Muhaidat R, Sage TL, Frohlich MW, Dengler NG, Sage RF (2011) Characterization of C3-C4 intermediate species in the genus Heliotropium L. (Boraginaceae): Anatomy, ultrastructure and enzyme activity. Plant, Cell Environ **34**: 1723–1736

Oaks A (1994) Efficiency of nitrogen utilization in C3 and C4 cereals. Plant Physiol **106**: 407–414

Ogren WL (1984) Photorespiration: Pathways, regulation, and modification. Annu Rev Plant Biol **35**: 415–442

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

Osborne CP, Freckleton RP (2009) Ecological selection pressures for C4 photosynthesis in the grasses. Proceedings Biol Sci **276**: 1753–60

Osborne CP, Sack L (2012) Evolution of C4 plants: a new hypothesis for an interaction of CO2 and water relations mediated by plant hydraulics. Philos Trans R Soc Lond B Biol Sci **367**: 583–600

Osborne CP, Salomaa A, Kluyver TA, Visser V, Kellogg EA, Morrone O, Vorontsova MS, Clayton WD, Simpson DA (2014) A global database of C4 photosynthesis in grasses. New Phytol **204**: 441–446

Page DR, Grossniklaus U (2002) The art and design of genetic screens: Arabidopsis Thaliana. Nat Publ Gr **3**: 124–136

Parry MAJ, Andralojc PJ, Scales JC, Salvucci ME, Carmo-Silva AE, Alonso H, Whitney SM (2013) Rubisco activity and regulation as targets for crop improvement. J Exp Bot **64**: 717–730

Peterhansel C, Horst I, Niessen M, Blume C, Kebeish R, Kürkcüoglu S, Kreuzaler F (2010) Photorespiration. Arab B 1–24

Peterhansel C, Niessen M, Kebeish RM (2008) Metabolic engineering towards the enhancement of photosynthesis. Photochem Photobiol **84**: 1317–1323

Pick TR, Bräutigam A, Schulz M a, Obata T, Fernie AR, Weber APM (2013) PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a unique class of metabolite transporters. Proc Natl Acad Sci U S A **110**: 3185–90

Price GD, Pengelly JJL, Forster B, Du J, Whitney SM, Von Caemmerer S, Badger MR, Howitt SM, Evans JR (2013) The cyanobacterial CCM as a source of genes for improving photosynthetic CO2 fixation in crop species. J Exp Bot **64**: 753–768

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

Reynolds M, Bonnett D, Chapman SC, Furbank RT, Manès Y, Mather DE, Parry MAJ (2011) Raising yield potential of wheat. I. Overview of a consortium approach and breeding strategies. J Exp Bot **62**: 439–52

Roalson EH, Hinchliff CE, Trevisan R, da Silva CRM (2010) Phylogenetic relationships in Eleocharis (Cyperaceae): C4 photosynthesis origins and patterns of diversification in the Spikerushes. Syst Bot **35**: 257–271

Sage RF (2001) Environmental and evolutionary preconditions for the origin and diversification of the C4 photosynthetic syndrome. Plant Biol **3**: 202–213

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

Sage RF (2016) A portrait of the C4 photosynthetic family on the 50th anniversary of its discovery: Species number, evolutionary lineages, and Hall of Fame. J Exp Bot **67**: 4039–4056

Sage RF, Christin PA, Edwards EJ (2011a) The C4 plant lineages of planet earth. J Exp Bot **62**: 3155–3169

Sage RF, Khoshravesh R, Sage TL (2014) From proto-Kranz to C4 Kranz: building the bridge to C4 photosynthesis. J Exp Bot **65**: 3341–56

Sage RF, Monson RK, Kellogg E a. (1999) Phylogenetic aspects of the evolution of C4 photosynthesis. C4 Plant Biol. pp 411–444

Sage RF, Sage TL, Kocacinar F (2012) Photorespiration and the evolution of C4 photosynthesis. Annu Rev Plant Biol **63**: 19–47

Sage RF, Zhu X-G (2011) Exploiting the engine of C4 photosynthesis. J Exp Bot **62**: 2989–3000

Sage TL, Sage RF, Vogan PJ, Rahman B, Johnson DC, Oakley JC, Heckel MA (2011b) The occurrence of C2 photosynthesis in Euphorbia subgenus Chamaesyce (Euphorbiaceae). J Exp Bot **62**: 3183–3195

Sage RF, Li M MR (1999) The taxonomic distribution of C4 photosynthesis. C4 Plant Biol 551–585

Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jørgensen J-E, Weigel D, Andersen SU (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat Methods **6**: 550–551

Schuler ML, Mantegazza O, Weber APM (2016) Engineering C4 photosynthesis into C3 chassis in the synthetic biology age. Plant J **87**: 51–65

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 CO2 pump. Plant Cell **25**: 2522–35

Scoffoni C, Rawls M, McKown A, Cochard H, Sack L (2011) Decline of leaf hydraulic conductance with dehydration: relationship to leaf size and venation architecture. Plant Physiol **156**: 832–43

Serrat X, Esteban R, Guibourt N, Moysset L, Nogués S, Lalanne E, Shimamoto K, Kyozuka J, Feng Q, Zhang Y, et al (2014) EMS mutagenesis in mature seed-derived rice calli as a new method for rapidly obtaining TILLING mutant populations. Plant Methods **10**: 5

Sharkey TD (1988) Estimating the rate of photorespiration in leaves. Physiol Plant **73**: 146–152

Siedlecka A, Krupa Z (2002) Simple method of Arabidopsis thaliana cultivation in liquid nutrient medium. Acta Physiol Plant **24**: 163–166

Sikora P, Chawade A, Larsson M, Olsson J, Olsson O, Sikora P, Chawade A, Larsson M, Olsson J, Olsson O (2011) Mutagenesis as a tool in plant genetics, functional genomics, and breeding. Int J Plant Genomics **2011**: 314829

Somerville CR (2001) An early Arabidopsis demonstration resolving a few issues concerning photorespiration. Plant Physiol **127**: 3

Still CJ, Berry JA, Collatz GJ, DeFries RS (2003) Global distribution of C3 and C4 vegetation: Carbon cycle implications. Global Biogeochem Cycles **17**: 6-1-6–14

Tcherkez G (2013) Is the recovery of (photo) respiratory CO2 and intermediates minimal? New Phytol **198**: 334–338

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408**: 796–815

Tolbert NE (1971) Microbodies-Peroxisomes and Glyoxysomes. Annu Rev Plant Physiol **22**: 45–74

United Nations, Department of Economic and Social Affairs PD 2015 (2015) World_Population_2015. https://esa.un.org/unpd/wpp/publications/Files/World_Population_2015_Wallchart.pdf 1–2

Vogan PJ, Frohlich MW, Sage RF (2007) The functional significance of C3-C4 intermediate traits in Heliotropium L. (Boraginaceae): gas exchange perspectives. Plant Cell Environ **30**: 1337–45

Wacey D (2009) Early life on earth. **31**: 47–53

Walker BJ, VanLoocke A, Bernacchi CJ, Ort DR (2016) The costs of photorespiration to food production now and in the future. Annu Rev Plant Biol **67**: 107–129

Wang X, Gowik U, Tang H, Bowers JE, Westhoff P, Paterson AH (2009) Comparative genomic analysis of C4 photosynthetic pathway evolution in grasses. Genome Biol **10**: R68

Wang Y, Bräutigam A, Weber APM, Zhu XG (2014) Three distinct biochemical subtypes of C4 photosynthesis? A modelling analysis. J Exp Bot **65**: 3567–3578

Weigel D, Ahn JH, Blazquez MA, Borevitz JO, Christensen SK, Fankhauser C, Ferrandiz C, Kardailsky I, Malancharuvil EJ, Neff MM, et al (2000) Activation tagging in Arabidopsis. Plant Physiol **122**: 1003–1013

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

Wingler A, Lea PJ, Quick WP, Leegood RC (2000) Photorespiration: metabolic pathways and their role in stress protection. Philos Trans R Soc Lond B Biol Sci **355**: 1517–29

Woo KC, Anderson JM, Boardman NK, Downton WJ, Osmond CB, Thorne SW (1970) Deficient photosystem II in agranal bundle sheath chloroplasts of C4 plants. Proc Natl Acad Sci U S A **67**: 18–25

Wyrich R, Dressen U, Brockmann S, Streubel M, Chang C, Qiang D, Paterson a H, Westhoff P (1998) The molecular basis of C4 photosynthesis in sorghum: isolation, characterization and RFLP mapping of mesophyll- and bundle-sheath-specific cDNAs obtained by differential screening. Plant Mol Biol **37**: 319–335

Yoshimura Y, Kubota F, Ueno O (2004) Structural and biochemical bases of photorespiration in C4 plants: Quantification of organelles and glycine decarboxylase. Planta **220**: 307–317

Zhu X-G, Long SP, Ort DR (2010) Improving photosynthetic efficiency for greater yield. Annu Rev Plant Biol **61**: 235–261

V. Manuscripts

1. Florian Döring, Monika Streubel, Andrea Bräutigam, Udo Gowik (2016). **Most photorespiratory genes are preferentially expressed in the bundle sheath cells of the C_4 grass *Sorghum bicolor***. Journal of Experimental Botany, **67**, 3053-3064

2. Florian Döring, Kumari Billakurthi, Udo Gowik, Stefanie Sultmanis, Roxana Khoshravesh, Shipon Das Gupta, Tammy Sage, Peter Westhoff (2017). **An EMS-based genetic screen to identify bundle sheath ontogeny and maintenance (*BSOM*) genes in *Arabidopsis thaliana*.** Publication in preparation (The Plant Journal).

3. **Altered expression of *BSOM2*, a gene coding for an ABA efflux transporter, results in increased bundle sheath and vascular tissue in the leaf of *Arabidopsis thaliana***. Unpublished work.

Manuscript 1

Most photorespiratory genes are preferentially expressed in the bundle sheath cells of the C_4 grass *Sorghum bicolor*

Journal of Experimental Botany, Vol. 67, No. 10 pp. 3053–3064, 2016
doi:10.1093/jxb/erw041 Advance Access publication 14 March 2016
This paper is available online free of all access charges (see http://jxb.oxfordjournals.org/open_access.html for further details)

RESEARCH PAPER

Most photorespiratory genes are preferentially expressed in the bundle sheath cells of the C_4 grass *Sorghum bicolor*

Florian Döring[1], Monika Streubel[1], Andrea Bräutigam[2,3,*] and Udo Gowik[1,†]

[1] Institute of Plant Molecular and Developmental Biology, Universitätsstrasse 1, Heinrich-Heine-University, D-40225 Düsseldorf, Germany
[2] Institute of Plant Biochemistry, Universitätsstrasse 1, Heinrich-Heine-University, D-40225 Düsseldorf, Germany
[3] Cluster of Excellence on Plant Sciences (CEPLAS) 'From Complex Traits towards Synthetic Modules', D-40225 Düsseldorf, Germany

* Present address: Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstraße 3, D-06466 Stadt Seeland, Germany.
† Correspondence: gowik@uni-duesseldorf.de

Received 2 December 2015; Accepted 21 January 2016

Editor: Martin Hagemann, University Rostock

Abstract

One of the hallmarks of C_4 plants is the division of labor between two different photosynthetic cell types, the mesophyll and the bundle sheath cells. C_4 plants are of polyphyletic origin and, during the evolution of C_4 photosynthesis, the expression of thousands of genes was altered and many genes acquired a cell type-specific or preferential expression pattern. Several lines of evidence, including computational modeling and physiological and phylogenetic analyses, indicate that alterations in the expression of a key photorespiration-related gene, encoding the glycine decarboxylase P subunit, was an early and important step during C_4 evolution. Restricting the expression of this gene to the bundle sheath led to the establishment of a photorespiratory CO_2 pump. We were interested in whether the expression of genes related to photorespiration remains bundle sheath specific in a fully optimized C_4 species. Therefore we analyzed the expression of photorespiratory and C_4 cycle genes using RNA *in situ* hybridization and transcriptome analysis of isolated mesophyll and bundle sheath cells in the C_4 grass *Sorghum bicolor*. It turns out that the C_4 metabolism of *Sorghum* is based solely on the NADP-dependent malic enzyme pathway. The majority of photorespiratory gene expression, with some important exceptions, is restricted to the bundle sheath.

Key words: C_4 photosynthesis, CO_2 fixation, differential gene expression, evolution, photorespiration, *Sorghum bicolor*.

Introduction

C_4 plants evolved multiple times from C_3 ancestors. The C_4 photosynthetic pathway leads to concentration of CO_2 around the main carboxylating enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO). This is achieved by a set of anatomical and biochemical modifications to the original C_3 pathway (Hatch, 1987). In the presence of high CO_2 concentrations, the oxygenase activity of RubisCO, which always competes with the carboxylation reaction, is effectively suppressed and hence photorespiration is strongly reduced in C_4 plants (Hatch, 1987). Photorespiration occurs when O_2 is used by RubisCO, which leads to the production of 2-phosphoglycolate (2-PG), a compound which is toxic for the plant cell and which needs to be detoxified (Anderson, 1971). Photorespiration takes place in chloroplasts, peroxisomes, and mitochondria. Throughout the regeneration of phosphoglycerate from phosphoglycolate, previously fixed

CO_2 is lost and additional energy and reduction equivalents are needed. Hence photorespiration can reduce the efficiency of photosynthesis in C_3 species by up to 30% (Ogren, 1984; Bauwe *et al.*, 2010; Raines, 2011; Fernie *et al.*, 2013). Therefore, C_4 photosynthesis can be of great advantage in conditions that promote photorespiration, such as hot, arid, and saline environments, in which plants have to close their stomata in order to avoid water loss through transpiration but which in consequence hinders the uptake of CO_2 (Sage, 2004). C_4 plants can keep their stomata closed for a longer time, because the CO_2 pump facilitates high rates of photosynthesis even under low CO_2 concentrations in the intercellular air space of the leaf and therefore minimizes water loss.

Leaves of C_4 plants show anatomical differences compared with those of C_3 plants. The vascular bundles are surrounded by organelle-rich bundle sheath cells, which, in turn, are surrounded by mostly one layer of mesophyll cells. This leads to a wreath-like appearance, which is termed Kranz anatomy (Haberlandt, 1904; Laetsch, 1974). In C_4 leaves, bundle sheath cells are enlarged and the interveinal distance is reduced (Dengler and Nelson, 1999). To allow the efficient interchange of metabolites between mesophyll and bundle sheath cells, both cell types are connected through numerous plasmodesmata (Botha, 1992).

In most species, C_4 photosynthesis largely depends on the division of labor between mesophyll and bundle sheath cells, in which the CO_2 assimilatory enzymes are compartmentalized. The C_4 pathway begins with the conversion of CO_2 to bicarbonate by carbonic anhydrase (CA) in the cytosol of mesophyll cells and the subsequent fixation into the C_4 acid oxaloacetate by phosphoenolpyruvate carboxylase (PEPC) with the 3-carbon compound phosphoenolpyruvate (PEP) as CO_2 acceptor. Afterwards, oxaloacetate is either reduced to malate or transaminated to aspartate, which is transported to the bundle sheath cells. There, CO_2 is released by decarboxylation of the C_4 compounds through a decarboxylating enzyme, either an NADP-dependent malic enzyme (NADP-ME), an NAD-dependent malic enzyme (NAD-ME), a PEP-carboxykinase (PEP-CK), or, as shown recently, a combination of these (Furbank, 2011; Y. Wang *et al.*, 2014). The released CO_2 is immediately refixed by RubisCO and enters the Calvin–Benson cycle. Less RubisCO is needed compared with C_3 plants as it works more efficiently under these conditions (Long, 1999). This results in a better nitrogen use efficiency of C_4 plants, since RubisCO is by far the most abundant protein in the leaves of higher plants (Long, 1999). Pyruvate, the other product of the decarboxylation, is transferred to the mesophyll cells where PEP is regenerated by pyruvate phosphate dikinase (PPDK).

C_4 photosynthesis has evolved at least 66 times independently from the original C_3 pathway (Sage *et al.*, 2011, 2012). To better understand the changes underlying the evolution of C_4 on the gene level, in recent years several studies aimed at creating transcriptome atlases of total leaf RNA of various pairs of closely related C_4 and C_3 species (Bräutigam *et al.*, 2011, 2014; Gowik *et al.*, 2011; Mallmann *et al.*, 2014). The development of C_3 and C_4 leaves was studied by analyzing the gene expression in different developmental stages of dicot leaves and the developmental gradients found in the leaves of C_3 and C_4 grasses (Li *et al.*, 2010; Pick *et al.*, 2011; Kulahoglu *et al.*, 2014; L. Wang *et al.*, 2014; Ding *et al.*, 2015). The coordination of the two different cell types was analyzed using mesophyll and bundle sheath transcriptomes of the C_4 grasses maize and *Setaria viridis* (Li *et al.*, 2010; Chang *et al.*, 2012; John *et al.*, 2014; Tausta *et al.*, 2014). It turned out that C_4 photosynthesis is a complex trait and its evolution involved changes in the expression of thousands of genes. Genes encoding the enzymes and transporters of the C_4 pathway had to be up-regulated and acquired tissue-specific expression. In addition, several other metabolic pathways must also have been regulated differentially in mesophyll and bundle sheath cells to enable this efficient type of photosynthesis including high nitrogen and water use efficiency attributed to C_4 plants.

It is widely accepted that the development of a photorespiratory CO_2 pump, often termed C_2 photosynthesis, was an important intermediate step during the evolution of the C_4 pathway (Bauwe, 2011; Sage *et al.*, 2012; Heckmann *et al.*, 2013; Williams *et al.*, 2013). The photorespiratory pump is based on the restriction of one of the key photorespiratory enzyme complexes, the glycine decarboxylase complex (GDC), to the bundle sheath cells (Rawsthorne *et al.*, 1988a). Photorespiratory glycine has to move to the bundle sheath for decarboxylation, and CO_2 is released mainly in this compartment, leading to increased CO_2 concentrations and allowing RubisCO to work more efficiently (Bauwe, 2011; Heckmann *et al.*, 2013). The photorespiratory pump can lead to a 3-fold enrichment of CO_2 in the bundle sheath cells (Keerberg *et al.*, 2014). The analysis of C_3–C_4 intermediate *Flaveria* species implied that the effect of the photorespiratory pump on C_4 evolution might be quite direct and provided a mechanistic explanation for how the photorespiratory pump and C_4 photosynthesis interact (Mallmann *et al.*, 2014). The glycine shuttle induces a nitrogen imbalance between mesophyll and bundle sheath cells, and the introduction of important components of the C_4 pathway, as well as the C_4 pathway itself, are highly efficient ways to correct this imbalance. This implies that C_4 evolution is a metabolic exaptation as the C_4 pathway developed in the first place to transport nitrogen and was not directly related to improving photosynthetic efficiency (Mallmann *et al.*, 2014). Hence, photorespiration and the cell-specific expression of photorespiratory genes in the mesophyll and bundle sheath cells of C_3–C_4 intermediates were of key importance for the evolution of C_4 photosynthesis.

In the present study, we examined how the expression of photorespiratory genes changed after the transition to true C_4 photosynthesis. Therefore we analyzed the expression of photosynthetic and photorespiratory genes in the C_4 grass *Sorghum bicolor* by RNA *in situ* hybridization and transcriptome analysis of isolated mesophyll and bundle sheath fractions. *Sorghum bicolor* is a highly optimized plant species with regard to the C_4 pathway. Methods for the isolation of mesophyll and bundle sheath cells are available (Wyrich *et al.*, 1998) and its genome is fully sequenced (Paterson *et al.*, 2009), allowing transcriptome analysis with plain high-throughput sequencing as well as with a serial analysis of gene expression (SAGE) approach since the short sequence reads could be

directly mapped to the genome or the derived transcriptome sequence (Bräutigam and Gowik, 2010). We determined transcript abundances within our mesophyll and bundle sheath RNA preparations by Illumina sequencing and additionally by SuperSage (Matsumura *et al.*, 2003), a combination of SAGE with next-generation sequencing methods.

We hypothesized that the distribution of photorespiratory gene expression is similar to the enzyme distributions determined previously (Ohnishi and Kanai, 1983; Gardeström *et al.*, 1985; Ohnishi *et al.*, 1985) and that it is comparable in specificity with the distribution of genes related to the C_4 pathway.

Materials and methods

Plant material, RNA isolation, and cDNA synthesis

Sorghum bicolor L. Tx430 (Pioneer Hi-Bred, Plainview, TX, USA) was grown on soil (Floraton 1, Floragard, Oldenburg, Germany) in the greenhouse of the Heinrich-Heine University (Düsseldorf, Germany) with supplementary light for 14 h per day (~300 µmol m^{-2} s^{-1}). For the *in situ* analysis, we harvested the middle thirds of the second leaf from 3-week-old plants and took 2 × 5 mm sections from it. For isolation of mesophyll and bundle sheath RNA, we harvested the upper two-thirds of the second leaf from 10-day-old seedlings. For generation of the cell-specific mRNAs, we separated the bundle sheath and vascular bundles enzymatically from the mesophyll and epidermal cells as described in Wyrich *et al.* (1998). We isolated 15 independent mesophyll and 19 independent bundle sheath samples. Cross-contaminations of the RNA preparations were controlled by dot blot analysis following standard procedures. Five independent mesophyll and bundle sheath preparations were pooled for the SuperSage analysis. For cDNA synthesis and Illumina sequencing, we pooled five other preparations for each tissue. Total RNA from intact *Sorghum* leaves was isolated according to Westhoff *et al.* (1991). Poly(A)$^+$ RNA was enriched by two consecutive rounds of oligo(dT) purification with the Oligotex mRNA Midi Kit (Qiagen, Hilden, Germany). cDNA libraries for Illumina sequencing were prepared with the SMARTer PCR cDNA Synthesis Kit (Clontech-Takara Bio Company, Otsu, Japan), with 300 ng of poly(A)$^+$ RNA as starting material. The purity and integrity of total RNA, poly(A)+ RNA, and cDNA were verified spectroscopically with a NanoDrop ND-1000, with the Agilent 2100 Bioanalyzer and by agarose gel electrophoresis.

SuperSage/Illumina sequencing

The SuperSage analysis was performed by GenXPro Inc. (Frankfurt, Germany) (Matsumura *et al.*, 2003). The mesophyll, bundle sheath, and total cDNA libraries were sequenced each in one lane of an Illumina flow cell with an Illumina Genome Analyser II by GATC Biotech AG (Konstanz, Germany) following standard protocols. The read length was 40 bp. The cDNAs were prepared from pooled total RNAs.

Mapping/statistics

The SuperSage tags as well as the Illumina reads were mapped on the *S. bicolor* transcriptome [version 1.4 (Sbicolor_79_transcript_primaryTranscriptOnly.fa) in the case of the SuperSage tags, and version 3.1 (Sbicolor_313_v3.1.transcript_primaryTranscriptOnly.fa) in the case of the Illumina reads (http://phytozome.jgi.doe.gov)]. The SuperSage tags were mapped with BLAST (Altschul *et al.*, 1990) by GenXPro Inc. Two mismatches were allowed and only tags that were found at least twice were counted. Tag counts were transformed to tags per million (tpm). For the mapping of the Illumina reads, we used BOWTIE (Langmead *et al.*, 2009). The best hit for each Illumina read was retained, and hit counts were then transformed to reads per kilobase and million (RPKM) to normalize for the number of reads available for each cDNA library.

Log2 ratios were calculated and differentially expressed transcripts were called using the R package DEGseq (Wang *et al.*, 2010) on the non-normalized read counts followed by a Bonferroni correction to account for the accumulation of alpha-type errors when conducting multiple pairwise comparisons.

qRT-PCR

Quantitative real-time PCR (qRT-PCR) followed standard procedures and was performed with an ABI7500 fast Real Time PCR system. The primers were designed to target photorespiratory genes of *S. bicolor* and to generate amplicons of 170 bp. The specificity of PCRs was verified by melting curve analysis and agarose gel electrophoresis. To estimate the efficiency of the PCRs, four consecutive 5-fold dilutions of the cDNAs were tested with each primer pair. Only reactions with efficiencies >90% were considered for further analysis. As template we used total RNAs pooled from five independent mesophyll and bundle sheath preparations each, not used for SuperSAGE or Illumina sequencing.

RNA in situ *hybridization*

The tissue was fixed for 16 h in a mixture of 3.7% formaldehyde, 50% ethanol, and 5% acetic acid at 4 °C. Dehydration and embedding was done in the Tissue Processor Leica ASP300S using the following program: 1 h in 50% ethanol, 1 h in 70% ethanol, 1 h in 95% ethanol, 3 × 1 h in 100% ethanol, 2 × 1 h in 100% xylene, 1 h in 100% xylene (37 °C), 2 × 10 min in histowax (62 °C), and 20 min in histowax (62 °C). Subsequently the samples were embedded in paraffin and cut into 12 µm sections with a microtome.

Probe labeling: for the generation of hybridization probes, the respective cDNAs were amplified by PCR and cloned into pJET1.2/blunt plasmid (Thermo Scientific, St. Leon-Rot, Germany). After linearization of the vector with appropriate restriction enzymes, T7 RNA polymerase was used to generate both sense and antisense probes, which were labeled with digoxigenin (DIG)-labeled UTP using the DIG RNA Labeling kit (Roche, Mannheim, Germany). Subsequently the probes were hydrolyzed to a size of ~150–200 bases.

Pre-hybridization, hybridization, and post-hybridization steps were based on the protocol described by Simon (2002). Only deviations from this protocol are mentioned below. First the sections were dewaxed in Roti®-Histol for 10 min and rehydrated in a decreasing ethanol concentration series: 2 × 1 min in 100% ethanol, 1 min in 95% ethanol, 1 min in 85% ethanol, 1 min in 50% ethanol, 1 min in 30% ethanol, and 1 min in ddH_2O. Afterwards the sections were treated with 10 µg ml^{-1} proteinase K for 30 min at 37 °C, post-fixed and acetylated as described by Simon (2002), and finally dehydrated in a reverse order of the ethanol concentration series used before. For the hybridization, 150 ng of probe was used for each slide. The sections were incubated for 16 h at 50 °C in a humid chamber.

After hybridization, the sections were washed three times in washing buffer (2× SSC, 50% formamide) for 30 min at 50 °C and twice in NTE buffer (500 mM NaCl, 10 mM Tris, 1 mM EDTA, pH 8.0) for 5 min at 37 °C. After RNase A treatment, the sections were washed again twice in NTE at room temperature for 5 min and in washing buffer for 1 h at 50 °C.

For immunological detection, all steps were performed on a shaking platform. First the sections were washed in buffer 1 (100 mM Tris-HCl pH 7.5, 150 mM NaCl) for 5 min, before they were incubated in buffer 2 (buffer 1 containing 0.5% blocking reagent; Roche) for 40 min. Subsequently they were incubated in buffer 3 (buffer 1 containing 0.3% Triton X-100, 1% normal sheep serum, and sheep anti-DIG–alkaline phosphatase at a dilution of 1:2000) for 2 h, after which they were washed four times

in buffer 1 containing 0.3% Triton X-100 for 15 min. Then the sections were washed in buffer 1 for 5 min, incubated in buffer 4 (0.1 M Tris-HCl pH 9.5, 0.1 M NaCl, and 50 mM $MgCl_2$) for 5 min, and finally stained in buffer 5 [buffer 4 containing 10% polyvinyl alcohol, 0.16 mM nitroblue tetrazolium (NBT), and 0.15 mM BZIP] in a humid chamber for 12–16 h. The reaction was stopped by washing the sections twice in distilled water, after which they were mounted with Entellan® (Merck Millipore, Darmstadt, Germany).

Results

Mesophyll and bundle sheath RNAs

Mesophyll and bundle sheath cells of *S. bicolor* for RNA preparations were separated by enzymatic digestion of leaf cell walls as described in Wyrich *et al.* (1998). It has to be considered that the mesophyll fraction also contains epidermis cells whereas the bundle sheath fraction contains all vascular tissues. The cross-contamination of mesophyll and bundle sheath preparations was analyzed by dot blot analysis using a PEPC and an NADP-ME cDNA as hybridization probes (Fig. 1). PEPC is thought to be mesophyll specific in *Sorghum* whereas NADP-ME was shown to be exclusively expressed in bundle sheath cells (Wyrich *et al.*, 1998). Since no signals indicating cross-contamination were visible, it can be assumed that the RNA preparations are pure and that the cross-contamination of mesophyll and bundle sheath RNAs is <5% (Fig. 1).

Transcriptome analysis via SuperSage and RNA-Seq

To create transcriptome atlases of *Sorghum* bundle sheath and mesophyll tissue, we performed plain Illumina sequencing and a SuperSage analysis. With the SuperSage method, we obtained $>6.8\times10^6$ tags (total leaf, 1 098 800; mesophyll, 3 349 814; bundle sheath, 2 421 27) that could be assigned to >12 000 (12 937) of the 34 211 predicted *Sorghum* genes, whereas 2327 genes exhibited a significantly different expression between mesophyll and bundle sheath cells ($P<0.01$) (Table 1). With plain Illumina sequencing we produced $>36\times10^6$ reads (total leaf, 17 704 772; mesophyll, 10 420 446; bundle sheath, 8 695 328) which could be mapped to 23 244 *Sorghum* genes. With this method, we identified 1705 genes as being expressed significantly differentially between mesophyll and bundle sheath cells ($P<0.01$) (Table 1; Supplementary Table S1 at *JXB* online). With the SuperSage approach, we detected far fewer transcripts compared with the plain Illumina sequencing. This is most probably due to limitations of the SuperSage method. A transcript will not be recognized if the cleavage site of the anchoring enzyme, which is needed to produce the DNA fragments used as tags, is not present in the transcript (Matsumura *et al.*, 2003).

In total, we were able to detect 12 154 transcripts expressed within the *Sorghum* leaf with at least one read in both experiments, which corresponds to 35% of the total number of transcripts predicted from the *Sorghum* genome sequence (Paterson *et al.*, 2009). A total of 455 (3.7%) of them were more abundant in mesophyll cells and 401 (3.2%) in the bundle sheath in both experiments.

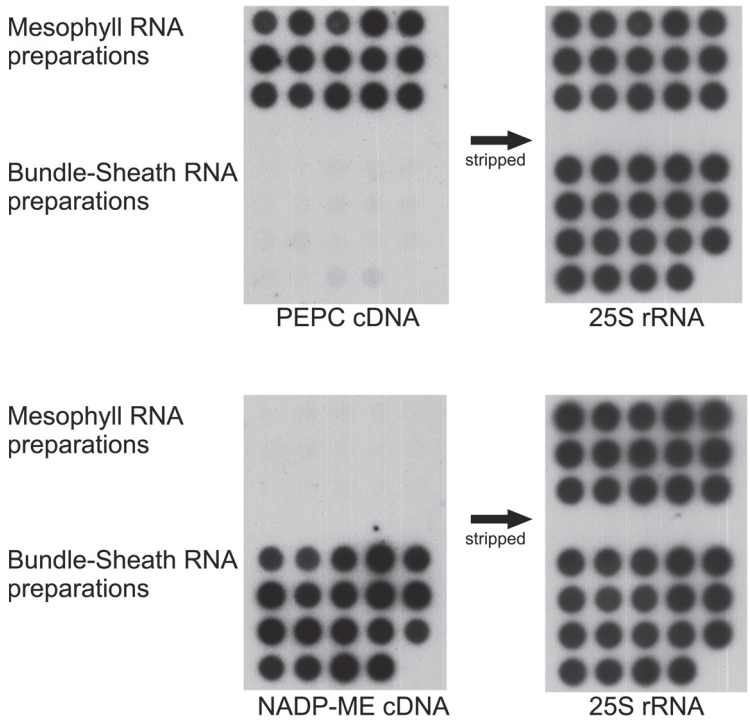


Fig. 1. Dot-blot analysis of independent mesophyll and bundle sheath RNA preparations. *Sorghum bicolor* PEPC cDNA, NADP-ME cDNA, and 25S rRNA were used as probes.

The enzymatic separation of mesophyll and bundle sheath cells influences gene expression

During the separation of mesophyll and bundle sheath cells by enzymatic digest, the tissue is incubated for up to 2.5h at 25 °C. It is known that this treatment stresses the plant cells and leads to the expression of stress-related genes (Sawers *et al.*, 2007). To account for this problem, we isolated RNA from complete, unstressed *Sorghum* leaves. We assumed that mesophyll and bundle sheath RNA accounts for a comparable fraction of the whole leaf RNA. Based on this premise, we identified 3697 genes within the SuperSage experiment and 3724 genes within the RNA-Seq experiment that were up-regulated >3-fold apparently due to the enzymatic treatment. To test this assumption, we analyzed the representation of Gene Ontology (GO) terms for the up-regulated genes. Indeed, we found an over-representation of GO terms related to stress response among these 3-fold up-regulated genes in the SuperSage as well as in the RNA-Seq experiment (Tables 2, 3). The genes found to be >3-fold up- or down-regulated after enzyme treatment were tagged.

The photorespiratory cycle mainly takes place in the bundle sheath in S. bicolor

It was assumed earlier that in C_4 plants the photorespiratory pathway is mainly located in the bundle sheath cells since in C_4 plants, RubisCO, the entry enzyme of photorespiration, is restricted to this cell type (Bauwe, 2011). One exception is glycerate kinase (GLYK), which catalyzes the regeneration of 3-phosphoglycerate (3-PG) and was found to be restricted to the mesophyll cells (Usuda and Edwards, 1980). The present transcriptome analysis largely supports these expectations (Fig. 2; Supplementary Table S2), as do the *in situ* hybridizations (Fig. 2; Supplementary Fig. S1). We detected a strong signal in the bundle sheath for most transcripts of the core photorespiratory pathway with genes that show virtually no expression in the mesophyll and can be seen as bundle sheath specific, such as phosphoglycolate phosphatase (PGLP), glycolate oxidase (GOX), serine hydroxymethyl transferase (SHM), and the H, P, and T subunit of the GDC (Fig. 2; Supplementary Fig. S1). However, there are also genes such as glycine 2-oxoglutarate aminotransferase (GGT) and the GDC L subunit that, although preferentially expressed in the bundle sheath, still seem to be expressed to a certain extent in the mesophyll (Fig. 2; Supplementary Fig. S1). Taken together, this implies that all genes of the core photorespiratory pathway are at least preferentially if not specifically expressed in the bundle sheath, except for GLYK that is expressed to a much higher level in the mesophyll than in the bundle sheath (Fig. 2; Supplementary Table S2). We did not obtain any *in situ* hybridization signal for GLYK. This may be caused by the low absolute expression of the gene observed even in the mesophyll (Supplementary Table S2).

Table 1. *Overview of the SuperSage and RNA-Seq results*

	SuperSage	RNA-Seq
Total reads:	6 870 541	36 820 546
Genes detected (*S. bicolor* 34 211 genes):	12 937	23 244
Percentage:	37	67
Differentially expressed:	2327	1705
Percentage:	6.8	4.9

The transcriptome analysis reveals detailed insight into the C_4 pathway of S. bicolor

Sorghum bicolor belongs to the NADP-ME type of C_4 plants. The genes encoding PEPC, malate dehydrogenase (MDH), or PPDK are expected to be expressed specifically or at least strongly preferentially in the mesophyll in these plants, whereas the genes encoding NADP-ME or RubisCO are bundle sheath specific. The results of our transcriptome analyses are essentially in line with these expectations (Fig. 3; Supplementary Table S3). Although PEPC

Table 2. *GO term over-representation analysis of genes up-regulated >3-fold in mesophyll or bundle sheath RNAs compared with total leaf RNA within the Illumina RNA-Seq experiment*

The 10 most strongly over-represented GO terms are shown. Analysis was performed using the Gene Ontology Consortium database (http://geneontology.org).

GO term	GO name	*P*-value
GO:0050896	Response to stimulus	9.40E-17
GO:1901701	Response to oxygen-containing compound	1.57E-13
GO:0042221	Response to chemical	5.65E-12
GO:0001101	Response to acid chemical	5.65E-12
GO:0006950	Response to stress	2.83E-11
GO:0044699	Single-organism process	4.35E-11
GO:0071704	Single-organism cellular process	5.77E-11
GO:0009719	Response to endogenous stimulus	1.20E-10
GO:0071229	Cellular response to acid chemical	2.57E-10
GO:0010033	Response to organic substance	2.79E-10

P-values are corrected by the Bonferroni method.

Table 3. *GO term over-representation analysis of genes up-regulated >3-fold in mesophyll or bundle sheath RNAs compared with total leaf RNA within the SuperSage experiment*

The 10 most strongly over-represented GO terms are shown. Analysis was performed using the Gene Ontology Consortium database (http://geneontology.org).

GO term	GO name	*P*-value
GO:0042221	Response to chemical	3.88E-17
GO:1901700	Response to oxygen-containing compound	1.24E-16
GO:0050896	Response to stimulus	2.02E-16
GO:0009987	Cellular process	8.90E-16
GO:0044237	Cellular metabolic process	1.98E-15
GO:0044699	Single-organism process	1.73E-14
GO:0009628	Response to abiotic stimulus	1.83E-14
GO:0044710	Single-organism metabolic process	1.92E-14
GO:0006950	Response to stress	4.25E-14
GO:0010033	Response to organic substance	4.74E-14

P-values are corrected by the Bonferroni method.

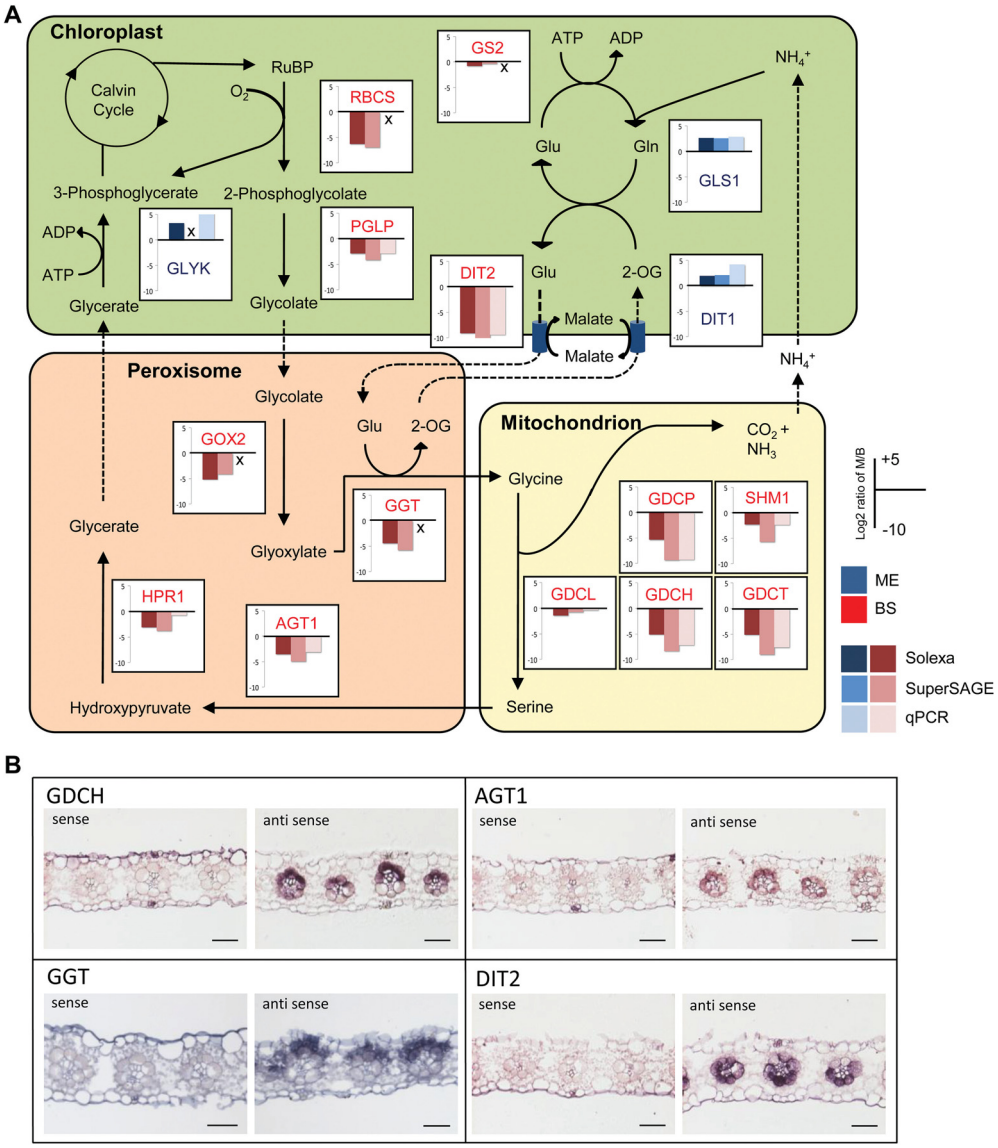


Fig. 2. (A) Distribution of photorespiratory genes between mesophyll and bundle sheath cells. Preferential gene expression in the mesophyll and bundle sheath is indicated by blue or red color, respectively. AGT, serine glyoxylate aminotransferase; DIT1+2, dicarboxylate transporter 1+2; GDCP/GDCL/GDCH/GDCT, glycine decarboxylase H, L, P, and T subunit; GGT, glutamate glyoxylate aminotransferase; GLS, glutamate synthase; GLYK, glycerate kinase; GOX2, glycolate oxidase 2; GS, glutamine synthetase; HPR, hydroxypyruvate reductases; PGLP, phosphoglycolate phosphatase; SHM, serine hydroxymethyltransferase; RBCS, ribulose bisphosphate carboxylase/oxygenase small subunit. (B) RNA *in situ* hybridization of *Sorghum bicolor* leaves with probes for transcripts related to photorespiration. Scale bars=50 µm.

was found to be expressed preferentially in the mesophyll, as expected, the absolute transcript levels as estimated by the Illumina sequencing appear to be quite low compared with NADP-ME or PPDK. In contrast, PEPC transcript levels turned out to be quite high when determined by the SuperSage method (Supplementary Table S3). If and how we selected against detecting high levels of the PEPC during the Illumina analysis is unclear. We detected virtually no expression of bundle sheath genes such as NADP-ME or RubisCO in the mesophyll, indicating that our mesophyll RNA preparations were not cross-contaminated with bundle sheath RNA (Supplementary Tables S2, S3). The fact that we detected some expression of typical mesophyll genes such as PEPC in the bundle sheath indicates some contamination of our bundle sheath RNA preparation with mesophyll RNA in the range of ~5% (Supplementary Table S3).

Recent results indicate that the classification of the different types of the C_4 pathway is not as clear-cut as previously

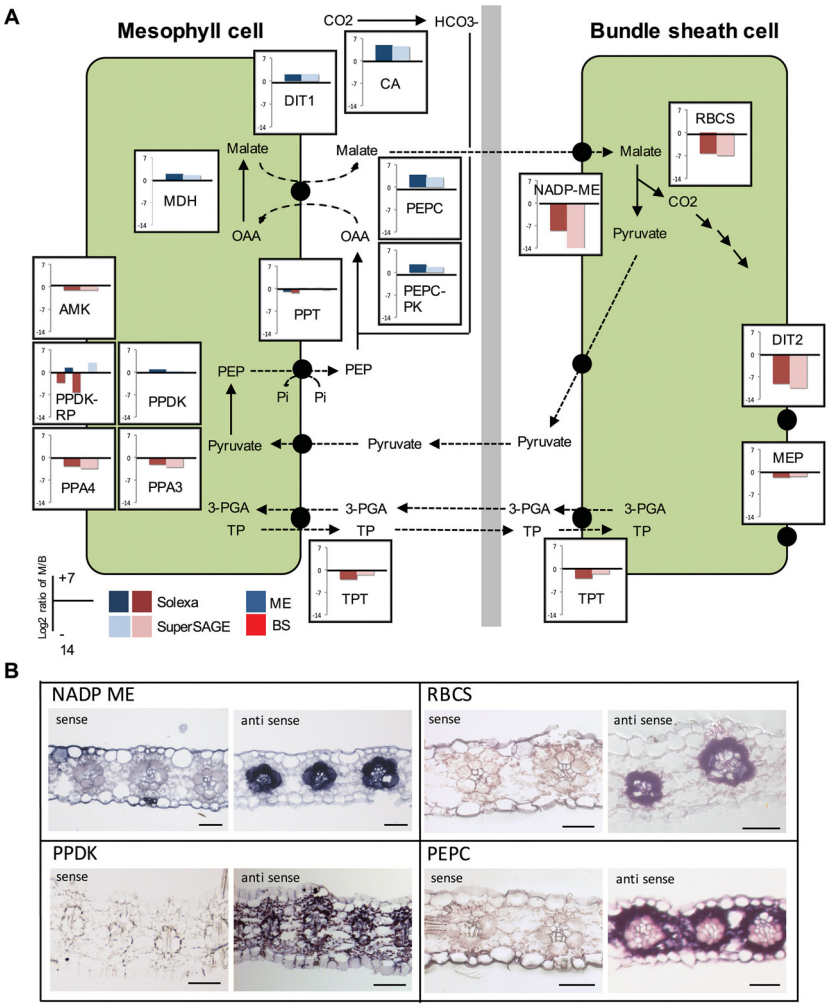


Fig. 3. (A) Distribution of C_4 cycle genes between mesophyll and bundle sheath cells. Preferential gene expression in the mesophyll and bundle sheath is indicated by blue or red color, respectively. AMK, AMP kinase; CA, carbonic anhydrase; DIT1+2, dicarboxylate transporter 1+2; MDH, NADP-dependent malate dehydrogenase; MEP, mesophyll envelope protein; NADP-ME, NADP-dependent malic enzyme; PEPC, phosphoenolpyruvate carboxylase; PEPC-PK, phosphoenolpyruvate carboxylase protein kinase; PPA3+4, pyrophosphorylase 3+4; PPDK, pyruvate phosphate dikinase; PPT, phosphoenolpyruvate phosphate translocator; PPDK-RP, PPDK regulatory protein; RubisCO, ribulose bisphosphate carboxylase/oxygenase; TPT, triosephosphate phosphate translocator. (B) RNA *in situ* hybridization of *Sorghum bicolor* leaves with probes for transcripts related to the C_4 pathway. Scale bars=50 µm.

thought (Furbank, 2011; Pick *et al.*, 2011; Y. Wang *et al.*, 2014). Maize, which was assumed to be an archetypal NADP-ME-type C_4 plant, uses in parallel the PEP-CK type pathway to a considerable extent (Wingler *et al.*, 1999; Pick *et al.*, 2011). Interestingly this does not apply for *Sorghum*, although maize and *Sorghum* possess a common C_4 origin. We did not find a highly expressed PEP-CK gene in bundle sheath cells and no abundantly expressed NAD-ME genes could be detected (Supplementary Table S3). It follows that *Sorghum* instead of maize should be considered as the NADP-ME C_4 archetype. It was shown earlier that, in contrast to the common textbook models of this pathway, some NADP-ME species use alanine and aspartate as transport metabolites in parallel to malate and pyruvate (Meister *et al.*, 1996; Gowik *et al.*, 2011). We were interested in whether the same is true for *Sorghum*, but the results are inconclusive. While we could identify a highly expressed aspartate aminotransferase (AspAT) gene in mesophyll as well as in bundle sheath cells, we have not found an alanine aminotransferase (AlaAT) that is highly expressed in both cell types. The most highly abundant AlaAT transcript, which belongs to the most abundant transcripts identified in this study, is clearly mesophyll specific. The function of this highly abundant AlaAT in the mesophyll remains unknown. We found another AlaAT gene which was significantly more highly expressed in the bundle sheath compared with the mesophyll (Supplementary Table S3) but, since its overall abundance is much lower, it is unclear if the overall AlaAT transcript abundance in the bundle sheath

allows the considerable usage of alanine and aspartate as transport metabolites. The up-regulated AspAT, on the other hand, is predicted to be localized in the chloroplast (TargetP score: 0.968). This in line with other NADP-ME species that synthesize and decarboxylate aspartate in the chloroplasts of mesophyll and bundle sheath cells (Meister *et al.*, 1996; Gowik *et al.*, 2011).

Transcripts related to most of the known transporters thought to be directly involved in the NADP-ME C_4 pathway, such as the triosephosphate phosphate translocator (TPT), the PEP phosphate translocator (PPT), the dicarboxylate transporter (DIT/DCT/OMT), or the inner chloroplast envelope transporter MEP (Weber and von Caemmerer, 2010) could be identified, and most of them showed high abundance in agreement with their probable role in the C_4 pathway (Fig. 3; Supplementary Table S3). However, it has to be considered that they did not always show the expected distribution in the two cell types (e.g. the PPT was expected to be mesophyll specific but we also found high amounts of PPT transcripts in the bundle sheath). We could not detect high expression for the BASS2 and the NHD transporter that were shown to catalyze pyruvate transport across the chloroplast membrane in the C_4 *Flaveria* species (Furumoto *et al.*, 2011). This is in line with earlier results indicating that *Sorghum* uses a proton-dependent pyruvate transporter (Aoki *et al.*, 1992) instead of BASS, which was shown to be a pyruvate–sodium symporter (Furumoto *et al.*, 2011).

While most of the core C_4 genes are expressed either mesophyll or bundle sheath specifically, as expected, we found that PPDK transcripts are not only highly abundant in the mesophyll, but were also present in respectable amounts in the bundle sheath, with a mesophyll to bundle sheath ratio of only ~1 to 2 (Fig. 3). Along with that, we also found that transcripts related to the PPDK reaction such as pyrophosphatases, AMP kinase, or the PPT exhibit high levels in the bundle sheath cells and are partly even preferentially expressed in the bundle sheath (Supplementary Table S3).

To verify the tissue distribution of selected transcripts, we performed *in situ* hybridizations for typical C_4 genes such as PEPC, NADP-ME, PPDK, and RBCS (RubisCO small subunit). The obtained results largely support the outcome of the transcriptome analysis using SuperSage or RNA-Seq (compare Fig. 3A and B). *In situ* hybridization confirmed bundle sheath-specific expression for RBCS and NADP-ME, mesophyll-specific expression for the PEPC gene, and the preferential expression in the mesophyll cells of PPDK, with high PPDK transcript levels also in the bundle sheath.

Expression patterns of genes associated with photorespiration are variable

During photorespiration not only CO_2, but also nitrogen is released in the mitochondria in the form of NH_3 that becomes reassimilated in the chloroplasts. In contrast to the core photorespiratory pathway, the genes for nitrogen assimilation and the dedicated transporters do not show a tissue-specific expression pattern. Glutamine synthetase as well as glutamate synthase genes are expressed in mesophyll and bundle sheath cells, but glutamine synthetase is more highly expressed in the bundle sheath, and a ferredoxin-dependent glutamine oxoglutarate aminotransferase (Fd-GOGAT) shows higher transcript abundance in the mesophyll (Fig. 2; Supplementary Table D2).

Only a few transporters involved in the intracellular transport of photorespiratory metabolites are known to date. We could identify two transcripts corresponding to the plastid glycolate glycerate transporter (Pick *et al.*, 2013). Whereas one of the genes appears not to be expressed at all in the *Sorghum* leaf, the other one exhibits high amounts of transcripts in both cell types, but the expression in the bundle sheath is higher than in the mesophyll (Fig. 2; Supplementary Table S2). The mitochondrial transporter BOU, known to be needed for functional photorespiration in *Arabidopsis thaliana* (Eisenhut *et al.*, 2013), appears to be expressed only at a low level in the leaves of the C_4 plant *Sorghum* and does not show a strong tissue preference (Supplementary Table S2). *Sorghum* contains five genes encoding dicarboxylate transporters (DITs); four of these transporters are classified as DIT2 and one is classified as a DIT1 gene. The DIT1 gene is expressed to moderate levels and clearly is expressed preferentially in the mesophyll. One of the DIT2 genes is highly expressed in the bundle sheath (Fig. 2; Supplementary Table S2). The two transporters are thought to interact in the glutamate–oxoglutarate exchange across the chloroplast membrane during NH_3 reassimilation (Renne *et al.*, 2003; Bauwe *et al.*, 2010). Additionally the DIT proteins might be involved in the C_4 cycle of NADP-ME C_4 species and facilitate the exchange of malate and/or aspartate across the chloroplast membrane (Gowik *et al.*, 2011; Kinoshita *et al.*, 2011), which may explain the highly tissue-preferential expression of these genes in *Sorghum*.

Discussion

C_4 photosynthesis mainly evolved to enhance photosynthetic efficiency by avoiding photorespiration. It is widely accepted that an important initial step towards the evolution of C_4 was the establishment of a photorespiratory CO_2 pump (Bauwe, 2011; Sage *et al.*, 2012). This was achieved by restricting the activity of a central photorespiratory protein complex, the GDC, to the bundle sheath cells, allowing the release of photorespiratory CO_2 exclusively in this cell type (Hylton *et al.*, 1988; Rawsthorne *et al.*, 1988*b*). Finally that was realized by restricting the expression of either single GDC subunit genes or all GDC and SHM genes to the bundle sheath (Morgan *et al.*, 1993). Nevertheless, photorespiration is still essential in C_4 plants (Zelitch *et al.*, 2008) and we were interested in the tissue-specific expression of photorespiratory genes in the mesophyll and bundle sheath cells of a widely optimized C_4 species. Therefore we analyzed gene expression in leaves of *S. bicolor* using RNA-Seq on isolated mesophyll and bundle sheath transcripts and RNA *in situ* hybridization.

Photorespiration is largely confined to the bundle sheath cells in Sorghum

In C_4 plants, photorespiration is reduced to low levels compared with C_3 plants as a result of concentrating CO_2 around RubisCO (Hatch, 1987). Using RNA-Seq and SuperSage,

we were able to detect the transcripts of all core photorespiratory genes as well as of the genes encoding transporters known to be involved in photorespiration. The vast majority of the core photorespiratory genes are expressed preferentially in the bundle sheath. The only noticeable exceptions are GLYK, which is expressed preferentially in the mesophyll, and the two genes encoding the L subunit of the GDC complex (GDCL), which are nearly equally expressed in both cell types. This largely reflects earlier results from the analysis of mesophyll and bundle sheath transcriptomes and proteomes of the C_4 grass maize (Li *et al.*, 2010; Majeran *et al.*, 2010; Chang *et al.*, 2012) and studies on the enzyme activities in different C_4 species (Usuda and Edwards, 1980; Ohnishi and Kanai, 1983; Ohnishi *et al.*, 1985). Since in C_4 plants RubisCO is missing from the mesophyll cells, no 2-PG can be produced there and 2-PG detoxification in this cell type is no longer necessary. Consequently, the expression of photorespiratory genes was switched off in the mesophyll during C_4 evolution. The photorespiratory enzymes belong to the most highly abundant proteins in the leaves of C_3 species (Osborne and Freckleton, 2009; Bauwe, 2011). Accordingly, the decrease in these proteins adds to the reduction of RubisCO in C_4 plants and contributes to the better nitrogen use efficiency found for C_4 species (Oaks, 1994; Osborne and Freckleton, 2009).

GDCL is not only part of the GDC but is also connected to other multienzyme complexes such as the pyruvate dehydrogenase complex, the 2-oxoglutarate dehydrogenase complex, and the branched-chained 2-oxoacid dehydrogenase that are not involved in photorespiration and have important functions in general cell metabolism (Millar *et al.*, 1999; Marrott *et al.*, 2014). This explains why the genes encoding GDCL have to stay active in the mesophyll of C_4 plants. An explanation for the preferential expression of GLYK in the mesophyll is less obvious. In advanced C_4 species using the NADP-ME pathway, such as maize or *Sorghum*, the activity of photosystem II is greatly reduced in the bundle sheath (Woo *et al.*, 1970; Oswald *et al.*, 1990). This requires the reductive phase of the Calvin–Benson cycle to take place in the mesophyll cells, due to a lack of reducing equivalents in the bundle sheath, and is achieved by a phosphoglycerate–triose phosphate shuttle (Weber and von Caemmerer, 2010). It appears to be more efficient to transfer the photorespiratory glycerate directly to the mesophyll chloroplasts to regenerate 3-PG instead of importing it into the bundle sheath chloroplast for regeneration.

The genes involved in photorespiratory ammonia refixation, glutamine synthetase and glutamate synthase, show different expression patterns in mesophyll and bundle sheath cells. While two glutamine synthetase genes are expressed in both cell types with a bundle sheath preference, Fd-GOGAT is preferentially expressed in the mesophyll. This makes sense in the light of lacking reducing equivalents in the bundle sheath and one can assume that the released ammonia is fixed by glutamine synthetase and the resulting glutamine is partially transferred to the mesophyll to generate glutamate.

The plastidic glycolate glycerate transporter PLGG1 (Pick *et al.*, 2013) is expressed in both cell types. This might be due to the fact that glycolate has to be exported from bundle sheath chloroplasts and glycerate must be imported into the chloroplasts in the mesophyll. It is known that the mitochondrial transporter BOU is essential for photorespiration in *A. thaliana* (Eisenhut *et al.*, 2013). Like PLGG, BOU is expressed in both cell types, but the overall transcript abundance is much lower. Since the specific substrate for the BOU transporter is not known (Eisenhut *et al.*, 2013), one can only speculate about possible functions beside photorespiration.

Specificity of photorespiratory genes is as variable as that of C_4 genes

With the transcriptome analysis, we confirmed that *S. bicolor* belongs to the NADP-ME type of C_4 plants since all participating C_4 genes (Wang *et al.*, 2009) are expressed in a tissue-preferential manner as expected for the NADP-ME archetype. Recent studies in maize revealed that not only the NADP-ME pathway is operating, but a respectable level of PEP-CK activity, up to 25% of the NADP-ME activity, was also found (Pick *et al.*, 2011). In the leaf transcriptome of *S. bicolor* we could find neither any highly expressed PEP-CK gene nor any significantly expressed NAD-ME gene in the bundle sheath. Taken together, these results indicate that *Sorghum* relies solely on the NADP-ME pathway.

As expected, we found PPDK to be one of the most highly expressed genes in the *Sorghum* leaf. Surprisingly, the transcriptome analysis indicated that PPDK transcripts are not restricted to the mesophyll but are also found in high amounts in the bundle sheath, with a mesophyll to bundle sheath ratio of only ~1 to 2 (Fig. 3). We confirmed that the analysis detects only the gene encoding the chloroplast-targeted PPDK isoform and indeed the gene encoding the cytosolic isoform showed quite low expression in *Sorghum* leaves. Also the RNA *in situ* analysis indicates high amounts of PPDK transcripts in the bundle sheath cells (Fig. 3B). Since this analysis is strictly independent of the transcriptome analysis, it must be considered that *Sorghum* contains considerable amounts of PPDK in its bundle sheath cells. This is in contrast to the analysis of mesophyll and bundle sheath cells of maize or *S. viridis* where PPDK transcripts were found to be five and 20 times more abundant in the mesophyll than in the bundle sheath, respectively (Chang *et al.*, 2012; John *et al.*, 2014). Very similar patterns were also found for the transcripts of genes that functionally interact with PPDK such as the PPDK regulatory proteins, plastid-localized pyrophosphatases, an AMP kinase, and the plastid PEP translocator PPT (Fig. 3; Supplementary Table S3). For all these genes, we found considerable amounts of transcripts in the bundle sheath preparations that were often even higher than in the mesophyll. The most parsimonious explanation is that *Sorghum* is capable of regenerating substantial amounts of PEP in the bundle sheath cells. The existence of plants using extensively the PEP-CK type of the C_4 pathway shows that PEP can serve as a transport metabolite in the C_4 cycle. Due to up-regulation of photosystem I and cyclic electron transport in the bundle sheath chloroplasts (Supplementary Table S1; Kubicki *et al.*, 1994, 1996), *Sorghum* potentially produces high amounts of ATP in this compartment that can be used

for PEP regeneration. By regenerating PEP in the bundle sheath chloroplasts, the number of transport processes would be reduced since PEP can be exported by PPT and diffuse into the mesophyll where it could be carboxylated by PEPC in the cytosol.

All in all, it appears that the degree of cell specificity is quite comparable for photorespiratory and C_4 cycle genes. While most of the genes encoding core pathway enzymes are expressed in a highly cell type-specific manner, exceptions are the PPDK in the case of the C_4 cycle and GDCL in the case of photorespiration. This is notable since tissue specificity for C_4 enzymes such as PEPC or NADP-ME is necessary to avoid futile cycles and ensure the efficiency of the pathway, whereas tissue-specific expression of most photorespiratory genes has to be seen as optimization that saves nitrogen. The expression of auxiliary genes of both pathways was found to be not very tissue specific. This might be due to additional roles of the encoded protein in other important pathways as can be envisaged for the genes involved in primary nitrogen and amino acid metabolism.

Evolutionary aspects of restricting photorespiration to the bundle sheath

As discussed above, photorespiration was important for the evolution of C_4 photosynthesis in different ways. The avoidance of photorespiration was one of the driving forces towards C_4 photosynthesis, and the establishment of a photorespiratory pump was an important intermediate step during C_4 evolution (Bauwe, 2011; Sage *et al.*, 2012). The reduction and exclusion of the majority of photorespiratory reactions from the mesophyll represents an optimization and enhances the nitrogen use efficiency. This optimization could only happen after the implementation of a fully functional C_4 pathway and the complete down-regulation of RubisCO in the mesophyll since the oxygenase reaction of RubisCO would be fatal without PGLP and GOX activity present in the same compartment. This has a further implication for C_4 evolution: once PGLP and GOX are switched off in the mesophyll, the reintroduction of RubisCO into this compartment would be detrimental. Once these photorespiratory reactions are gone from the mesophyll due to optimization, a reversal from C_4 to C_3 photosynthesis becomes impossible.

Supplementary data

Supplementary data are available at *JXB* online.

Table S1. Excel worksheet providing quantitative information for all reads and all SuperSage tags mapped onto the reference transcriptome from *Sorghum bicolor*.

Table S2. Transcript abundance of genes related to photorespiration

Table S3. Transcript abundance of C_4 cycle genes and C_4-related transporters.

Table S4. Gene-specific primers used for qPCR and RNA *in situ* analysis.

Figure S1, RNA *in situ* hybridization of *Sorghum bicolor* leaves with probes for transcripts related to photorespiration

Acknowledgements

This work was supported by the Bill and Melinda Gates Foundation through the C4 Rice Project, the Deutsche Forschungsgemeinschaft through the Research Group FOR1186, and the Excellence Cluster EXC 1028 (From Complex Traits towards Synthetic Modules).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology **215,** 403–410.

Anderson LE. 1971. Chloroplast and cytoplasmic enzymes. II. Pea leaf triose phosphate isomerases. Biochimica et Biophysica Acta **235,** 237–244.

Aoki N, Ohnishi J, Kanai R. 1992. Two different mechanisms for transport of pyruvate into mesophyll chloroplasts of C_4 plants—a comparative study. Plant and Cell Physiology **33,** 805–809.

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

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

Botha CE. 1992. Plasmodesmatal distribution, structure and frequency in relation to assimilation in C3 and C4 grasses in southern Africa. Planta **187,** 348–358.

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

Bräutigam A, Kajala K, Wullenweber J, *et al*. 2011. An mRNA blueprint for C4 photosynthesis derived from comparative transcriptomics of closely related C3 and C4 species. Plant Physiology **155,** 142–156.

Bräutigam A, Schliesky S, Külahoglu C, Osborne CP, Weber APM. 2014. Towards an integrative model of C4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C4 species. Journal of Experimental Botany **65,** 3579–3593.

Chang YM, Liu WY, Shih AC, *et al*. 2012. Characterizing regulatory and functional differentiation between maize mesophyll and bundle sheath cells by transcriptomic analysis. Plant Physiology **160,** 165–177.

Dengler NG, Nelson T. 1999. Leaf structure and development in C4 plants. In: Rowan FS, Russell KM, eds. C4 plant biology . San Diego: Academic Press, 133–172.

Ding Z, Weissmann S, Wang M, *et al*. 2015. Identification of photosynthesis-associated C4 candidate genes through comparative leaf gradient transcriptome in multiple lineages of C3 and C4 species. PLoS One **10,** e0140629.

Eisenhut M, Planchais S, Cabassa C, *et al*. 2013. Arabidopsis A BOUT DE SOUFFLE is a putative mitochondrial transporter involved in photorespiratory metabolism and is required for meristem growth at ambient CO2 levels. The Plant Journal **73,** 836–849.

Fernie AR, Bauwe H, Eisenhut M, *et al*. 2013. Perspectives on plant photorespiratory metabolism. Plant Biology **15,** 748–753.

Furbank RT. 2011. Evolution of the C4 photosynthetic mechanism: are there really three C4 acid decarboxylation types? Journal of Experimental Botany **62,** 3103–3108.

Furumoto T, Yamaguchi T, Ohshima-Ichie Y, *et al*. 2011. A plastidial sodium-dependent pyruvate transporter. Nature **476,** 472–475.

Gardeström P, Edwards GE, Henricson D, Ericson I. 1985. The localization of serine hydroxymethyltransferase in leaves of C3 and C4 species. Physiologia Plantarum **64,** 29–33.

Gowik U, Bräutigam A, Weber KL, Weber APM, Westhoff P. 2011. Evolution of C4 photosynthesis in the genus Flaveria: how many and which genes does it take to make C4? The Plant Cell **23,** 2087–2105.

Haberlandt G. 1904. Physiologische Pflanzenanatomie . Leipzig: Verlag von Wilhelm Engelmann.

Hatch MD. 1987. C4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. Biochimica et Biophysica Acta **895,** 81–106.

Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber APM, Lercher MJ. 2013. Predicting C4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. Cell **153,** 1579–1588.

Hylton C, Rawsthorne S, Smith A, Jones DA, Woolhouse H. 1988. Glycine decarboxylase is confined to the bundle-sheath cells of leaves of C3–C4 intermediate species. Planta **175,** 452–459.

John CR, Smith-Unna RD, Woodfield H, Covshoff S, Hibberd JM. 2014. Evolutionary convergence of cell-specific gene expression in independent lineages of C4 grasses. Plant Physiology **165,** 62–75.

Keerberg O, Parnik T, Ivanova H, Bassuner B, Bauwe H. 2014. C2 photosynthesis generates about 3-fold elevated leaf CO2 levels in the C3–C4 intermediate species Flaveria pubescens Journal of Experimental Botany **65,** 3649–3656.

Kinoshita H, Nagasaki J, Yoshikawa N, Yamamoto A, Takito S, Kawasaki M, Sugiyama T, Miyake H, Weber AP, Taniguchi M. 2011. The chloroplastic 2-oxoglutarate/malate transporter has dual function as the malate valve and in carbon/nitrogen metabolism. The Plant Journal **65,** 15–26.

Kubicki A, Steinmuller K, Westhoff P. 1994. Differential transcription of plastome-encoded genes in the mesophyll and bundle-sheath chloroplasts of the monocotyledonous NADP-malic enzyme-type C4 plants maize and Sorghum. Plant Molecular Biology **25,** 669–679.

Kubicki A, Funk E, Westhoff P, Steinmüller K. 1996. Differential expression of plastome-encoded ndh genes in mesophyll and bundle-sheath chloroplasts of the C4 plant Sorghum bicolor indicates that the complex I-homologous NAD(P)H-plastoquinone oxidoreductase is involved in cyclic electron transport. Planta **199,** 276–281.

Kulahoglu C, Denton AK, Sommer M, *et al*. 2014. Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae C3 and C4 plant species. The Plant Cell **26,** 3243–3260.

Laetsch WM. 1974. C4 syndrome—structural-analysis. Annual Review of Plant Physiology and Plant Molecular Biology **25,** 27–52.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology **10,** R25.

Li P, Ponnala L, Gandotra N, *et al*. 2010. The developmental dynamics of the maize leaf transcriptome. Nature Genetics **42,** 1060–1067.

Long SP. 1999. Environmental responses. In: Sage RF, Monson RK, eds. C4 plant biology . San Diego: Academic Press, 215–249.

Majeran W, Friso G, Ponnala L, *et al*. 2010. Structural and metabolic transitions of C4 leaf development and differentiation defined by microscopy and quantitative proteomics in maize. The Plant Cell **22,** 3509–3542.

Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber AP, Westhoff P, Gowik U. 2014. The role of photorespiration during the evolution of C4 photosynthesis in the genus Flaveria. Elife **3,** e02478.

Marrott NL, Marshall JJ, Svergun DI, Crennell SJ, Hough DW, van den Elsen JM, Danson MJ. 2014. Why are the 2-oxoacid dehydrogenase complexes so large? Generation of an active trimeric complex. Biochemical Journal **463,** 405–412.

Matsumura H, Reich S, Ito A, Saitoh H, Kamoun S, Winter P, Kahl G, Reuter M, Kruger DH, Terauchi R. 2003. Gene expression analysis of plant host–pathogen interactions by SuperSAGE. Proceedings of the National Academy of Sciences, USA **100,** 15718–15723.

Meister M, Agostino A, Hatch MD. 1996. The roles of malate and aspartate in C4 photosynthetic metabolism of Flaveria bidentis (L.). Planta **199,** 262–269.

Millar A, Hill S, Leaver C. 1999. Plant mitochondrial 2-oxoglutarate dehydrogenase complex: purification and characterization in potato. Biochemical Journal **343,** 327–334.

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.

Oaks A. 1994. Efficiency of nitrogen utilization in C3 and C4 cereals. Plant Physiology **106,** 407–414.

Ogren WL. 1984. Photorespiration: pathways, regulation, and modification. Annual Review of Plant Physiology **35,** 415–442.

Ohnishi J-i, Kanai R. 1983. Differentiation of photorespiratory activity between mesophyll and bundle sheath cells of C4 plants I. Glycine oxidation by mitochondria. Plant and Cell Physiology **24,** 1411–1420.

Ohnishi J-i, Yamazaki M, Kanai R. 1985. Differentiation of photorespiratory activity between mesophyll and bundle sheath cells of C4 plants II. Peroxisomes of Panicum miliaceum L. Plant and Cell Physiology **26,** 797–803.

Osborne CP, Freckleton RP. 2009. Ecological selection pressures for C4 photosynthesis in the grasses. Proceedings of the Royal Society B: Biological Sciences **276,** 1753–1760.

Oswald A, Streubel M, Ljungberg U, Hermans J, Eskins K, Westhoff P. 1990. Differential biogenesis of photosystem-II in mesophyll and bundle-sheath cells of 'malic' enzyme NADP(+)-type C4 plants. A comparative protein and RNA analysis. European Journal of Biochemistry **190,** 185–194.

Paterson AH, Bowers JE, Bruggmann R, *et al*. 2009. The Sorghum bicolor genome and the diversification of grasses. Nature **457,** 551–556.

Pick TR, Bräutigam A, Schlüter U, *et al*. 2011. Systems analysis of a maize leaf developmental gradient redefines the current C4 model and provides candidates for regulation. The Plant Cell **23,** 4208–4220.

Pick TR, Bräutigam A, Schulz MA, Obata T, Fernie AR, Weber AP. 2013. PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a unique class of metabolite transporters. Proceedings of the National Academy of Sciences, USA **110,** 3185–3190.

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, Hylton CM, Smith AM, Woolhouse HW. 1988*a*. Photorespiratory metabolism and immunogold localization of photorespiratory enzymes in leaves of C3 and C3–C4 intermediate species of Moricandia. Planta **173,** 298–308.

Rawsthorne S, Hylton CM, Smith AM, Woolhouse HW. 1988*b*. Distribution of photorespiratory enzymes between bundle-sheath and mesophyll cells in leaves of the C3–C 4 intermediate species Moricandia arvensis (L.) DC. Planta **176,** 527–532.

Renne P, Dressen U, Hebbeker U, Hille D, Flugge UI, Westhoff P, Weber AP. 2003. The Arabidopsis mutant dct is deficient in the plastidic glutamate/malate translocator DiT2. The Plant Journal **35,** 316–331.

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

Sage RF, Christin P-A, Edwards EJ. 2011. The C4 plant lineages of planet Earth. Journal of Experimental Botany **62,** 3155–3169.

Sage RF, Sage TL, Kocacinar F. 2012. Photorespiration and the evolution of C4 photosynthesis. Annual Review of Plant Biology **63,** 19–47.

Sawers RJ, Liu P, Anufrikova K, Hwang JT, Brutnell TP. 2007. A multi-treatment experimental system to examine photosynthetic differentiation in the maize leaf. BMC Genomics **8,** 12.

Simon R. 2002. Molecular and biochemical analysis of Arabidopsis. DIG application manual for nonradioactive in situ hybridization . Mannheim, Germany: Roche Applied Science, 197–207.

Tausta SL, Li P, Si Y, Gandotra N, Liu P, Sun Q, Brutnell TP, Nelson T. 2014. Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C4-related processes. Journal of Experimental Botany **65,** 3543–3555.

Usuda H, Edwards GE. 1980. Localization of glycerate kinase and some enzymes for sucrose synthesis in C3 and C4 plants. Plant Physiology **65,** 1017–1022.

Wang L, Feng Z, Wang X, Wang X, Zhang X. 2010. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics **26,** 136–138.

Wang L, Czedik-Eysenberg A, Mertz RA, *et al*. 2014. Comparative analyses of C4 and C3 photosynthesis in developing leaves of maize and rice. Nature Biotechnology **32,** 1158–1165.

Wang X, Gowik U, Tang H, Bowers JE, Westhoff P, Paterson AH. 2009. Comparative genomic analysis of C4 photosynthetic pathway evolution in grasses. Genome Biology **10,** R68.

Wang Y, Bräutigam A, Weber AP, Zhu XG. 2014. Three distinct biochemical subtypes of C4 photosynthesis? A modelling analysis. Journal of Experimental Botany **65,** 3567–3578.

Weber AP, von Caemmerer S. 2010. Plastid transport and metabolism of C3 and C4 plants—comparative analysis and possible biotechnological exploitation. Current Opinion Plant in Biology **13,** 257–265.

Westhoff P, Offermann-Steinhard K, Höfer M, Eskins K, Oswald A, Streubel M. 1991. Differential accumulation of plastid transcripts encoding photosystem II components in the mesophyll and bundle-sheath cells of monocotyledonous NADP-malic enzyme-type C4 plants. Planta **184,** 377–388.

Williams BP, Johnston IG, Covshoff S, Hibberd JM. 2013. Phenotypic landscape inference reveals multiple evolutionary paths to C4 photosynthesis. Elife **2,** e00961.

Wingler A, Walker RP, Chen ZH, Leegood RC. 1999. Phosphoenolpyruvate carboxykinase is involved in the decarboxylation of aspartate in the bundle sheath of maize. Plant Physiology **120,** 539–546.

Woo KC, Anderson JM, Boardman NK, Downton WJ, Osmond CB, Thorne SW. 1970. Deficient photosystem II in agranal bundle sheath chloroplasts of C4 plants. Proceedings of the National Academy of Sciences, USA **67,** 18–25.

Wyrich R, Dressen U, Brockmann S, Streubel M, Chang C, Qiang D, Paterson AH, Westhoff P. 1998. The molecular basis of C4 photosynthesis in sorghum: isolation, characterization and RFLP mapping of mesophyll- and bundle-sheath-specific cDNAs obtained by differential screening. Plant Molecular Biology **37,** 319–335.

Zelitch I, Schultes NP, Peterson RB, Brown P, Brutnell TP. 2008. High glycolate oxidase activity is required for survival of maize in normal air. Plant Physiology **149,** 195–204.

Supplementary Data

Supplementary Table 1. Excel worksheet providing quantitative information for all reads and all SuperSage tags mapped onto the reference transcriptome from *Sorghum bicolor*. (not included) Available online:

https://academic.oup.com/jxb/article/67/10/3053/1749097/Most-photorespiratory-genes-are-preferentially#35993780

Supplementary Table 2. Transcript abundance of genes related to photorespiration.

Supplementary Table 3. Transcript abundance of C_4 cycle genes and C_4-related transporters.

Supplementary Table 4. Gene-specific primers used for qPCR and RNA *in situ* analysis.

Supplementary Figure 1. RNA *in situ* hybridization of *Sorghum bicolor* leaves with probes for transcripts related to photorespiration.

Supplementary Table 2. Transcript abundance of genes related to photorespiration

		Illumina Sequencing						SuperSage					
locusName	gene-symbol	rpmk total leaf	rpmk M	rpmk B	log2 M/B	P-value	3 fold change	tpm total leaf	tpm M	tpm B	log2 M/B	P-value	3 fold change
Chloroplast													
Sobic.005G042000	RBCS1A	4233,815	293,974	22898,380	-6,283	0	no	9570,651	137,152	17292,959	-6,978	0	no
Sobic.006G130300	PGLP1	25,734	33,785	253,103	-2,905	3,27023E-37	up	194,911	23,307	416,688	-4,160	2,0241E-88	no
Sobic.001G007000	ATPK5,PGLP2	52,506	11,795	33,562	-1,509	1	no	1,822	2,092	0,413	2,339	1	no
Sobic.009G004100	GLYK	27,647	42,375	4,273	3,310	6,27671E-05	no	0,000	0,000	0,000	NA	NA	no
Sobic.003G259200	GLYK	1,223	1,230	0,609	1,014	1	no	0,000	0,000	0,000	NA	NA	no
Peroxisome													
Sobic.001G065600	GOX	14,322	31,815	10,971	1,536	1	no	0,000	0,598	2,894	-2,276	1	up
Sobic.006G220600	GOX	57,505	14,972	4,153	1,850	1	down	1,822	5,080	6,201	-0,288	1	no
Sobic.006G220500	GOX	1,415	2,871	0,000	NA	NA	no	0,911	2,988	0,413	2,854	1	no
Sobic.002G036000	GOX	138,160	3,326	118,320	-5,153	1,83619E-24	no	376,159	57,968	1016,504	-4,132	1,3907E-218	no
Sobic.002G374650	GOX	0,132	0,000	0,296	NA	NA	no	0,000	0,000	0,000	NA	NA	no
Sobic.002G374700	GOX	3,102	0,698	2,032	-1,542	1	no	0,000	0,000	0,000	NA	NA	no
Sobic.002G006600	AOAT1,GGAT1,GGT1	287,781	59,842	1301,550	-4,443	4,6543E-281	no	183,981	11,355	634,953	-5,805	3,2558E-149	no
Sobic.007G213700	AGT,AGT1,SGAT	15,175	2,505	29,156	-3,541	0,017164041	no	325,154	34,363	1077,270	-4,970	3,9618E-250	no
Sobic.004G001300	HPR	730,797	128,514	1080,376	-3,072	2,8341E-177	no	900,778	95,319	1310,004	-3,781	1,9468E-267	no
Mitochondria													
Sobic.004G055700	GDCH	3,514	0,000	4,257	NA	NA	no	3,643	0,299	11,575	-5,276	1	no
Sobic.008G062500	GDCH	108,391	6,994	236,982	-5,082	1,2377E-51	no	147,549	0,896	293,087	-8,353	4,09372E-57	no
Sobic.006G220800	GDCT	9,135	1,117	38,370	-5,102	7,05627E-06	no	67,399	0,299	153,778	-9,007	2,55099E-26	no
Sobic.003G152900	GDCL,mtLPD1	21,060	26,198	69,466	-1,407	0,376737044	no	71,953	140,439	233,560	-0,734	0,010363406	no
Sobic.009G054600	GDCL,mtLPD1	6,972	2,338	10,766	-2,203	1	no	15,484	15,239	26,043	-0,773	1	no
Sobic.008G039900	GDCP,GLDP2	124,436	24,931	1013,739	-5,346	6,7786E-232	up	407,126	1,793	1186,816	-9,371	4,9493E-211	no
Sobic.001G097100	SHM1,SHMT1,STM	110,030	48,230	239,734	-2,313	1,61387E-26	no	92,901	4,482	235,627	-5,716	1,00358E-53	no
Sobic.008G144800	SHM4	17,807	1,419	35,486	-4,644	5,96822E-05	no	27,324	3,885	81,023	-4,383	2,22493E-15	no
Sobic.005G113300	SHM4	13,718	5,460	33,354	-2,611	0,118928169	no	79,239	27,191	103,345	-1,926	7,26477E-08	no

		Illumina Sequencing						SuperSage					
locusName	gene-symbol	rpmk total leaf	rpmk M	rpmk B	log2 M/B	P-value	3 fold change	tpm total leaf	tpm M	tpm B	log2 M/B	P-value	3 fold change
NH_3 fixation													
Sobic.001G451500	GLN1;1,GSR 1	26,844	52,994	87,376	-0,721	1	no	79,239	338,548	306,315	0,144	1	up
Sobic.004G247000	GLN1;1,GSR 1	72,527	39,107	68,080	-0,800	1	no	1203,162	151,495	729,617	-2,268	3,10968E-90	no
Sobic.006G249400	GLN2,GS2	72,433	106,444	182,593	-0,779	0,554868248	no	418,967	196,316	249,269	-0,345	1	no
Sobic.001G116400	GSR2	5,136	10,446	2,302	2,182	1	no	20,038	13,148	43,405	-1,723	0,800326939	no
Sobic.002G402700	FD-GOGAT,GLS1,GLU1,GLUS	256,269	389,295	62,028	2,650	6,66152E-56	no	751,407	441,934	73,169	2,595	1,23715E-58	no
Transport													
Sobic.001G288900	BOU	0,453	3,261	1,016	1,682	1	up	66,488	60,060	45,885	0,388	1	no
Sobic.007G226800	DCT,DIT2.1	3,018	0,290	5,829	-4,331	1	no	0,000	0,000	0,827	NA	NA	up
Sobic.004G035500	DCT,DIT2.1	79,690	1,917	1053,286	-9,102	1,5008E-190	up	481,811	0,598	849,498	-10,473	1,1487E-127	no
Sobic.002G233700	DCT,DIT2.1	28,701	21,046	10,628	0,986	1	no	10,930	2,988	0,413	2,854	1	no
Sobic.007G226700	DCT,DIT2.1	0,061	0,064	0,411	-2,687	1	up	0,000	0,000	0,000	NA	NA	no
Sobic.008G112300	DiT1	19,738	25,761	6,547	1,976	1	no	542,835	177,192	40,511	2,129	1,63389E-16	no
Sobic.001G283100	PLGG	529,288	420,125	940,388	-1,162	2,88768E-39	no	567,426	238,746	367,495	-0,622	0,000697181	no
Sobic.001G283000	PLGG	18,507	17,456	29,825	-0,773	1	no	0,000	0,000	0,000	NA	NA	no

Supplementary Table 3. Transcript abundance of C_4 cycle genes and C_4-related transporters.

		Illumina Sequencing						SuperSage					
locusName	gene-symbol	rpmk total leaf	rpmk M	rpmk B	log2 M/B	P-value	3 fold change	tpm total leaf	tpm M	tpm B	log2 M/B	P-value	3 fold change
C4 NADP-ME													
Sobic.003G234200	BCA4,CA4	317,888	606,438	20,134	4,913	9,2011E-139	no	1834,345	3579,400	157,912	4,503	9,2011E-139	no
Sobic.003G234400	BCA4,CA4	27,721	133,032	2,186	5,927	5,48218E-29	no	1799,734	1817,340	73,582	4,626	5,48218E-29	no
Sobic.003G234600	BCA4,CA4	0,671	0,654	0,301	1,121	1	no	0,000	0,000	0,000	NA	1	no
Sobic.003G234500	BETA CA2,CA18,CA2	8,218	34,696	2,554	3,764	0,000384212	no	0,000	2,092	0,000	NA	0,000384212	down
Sobic.010G160700	ATPPC2,PPC2	201,204	182,626	11,960	3,933	8,96566E-36	no	10943,221	9058,013	1062,802	3,091	8,96566E-36	no
Sobic.007G137600	MDH	72,086	387,252	102,225	1,922	4,06286E-37	up	67,399	138,945	47,539	1,547	4,06286E-37	no
Sobic.007G166300	MDH	165,925	368,467	19,380	4,249	1,92794E-78	no	690,384	547,712	20,256	4,757	1,92794E-78	no
Sobic.003G036200	NADP-ME4	1882,793	35,126	12952,277	-8,526	0	up	5182,433	0,299	11702,811	-15,257	0	no
Sobic.003G036000	NADP-ME4	251,720	19,124	1814,836	-6,568	0	up	0,000	0,000	0,000	NA	0	no
Sobic.009G108700	NADP-ME4	123,265	20,356	669,038	-5,039	3,0089E-150	no	0,000	0,000	0,000	NA	3,0089E-150	no
Sobic.009G132900	PPDK	5343,046	8768,331	4789,007	0,873	4,9659E-279	no	6289,051	5985,388	4767,108	0,328	4,9659E-279	no
Sobic.001G326900	PPDK	4,887	5,218	4,856	0,104	1	no	0,000	0,000	0,000	NA	1	no
Sobic.004G331700	AAT3,ASP5,ATAAT1	672,110	397,132	195,499	1,022	1,87001E-13	no	19,127	31,076	6,201	2,325	1,87001E-13	no
Sobic.002G375800	ALAAT2	27,885	9,611	49,477	-2,364	0,003286999	no	32,789	53,785	178,167	-1,728	0,003286999	down
Sobic.001G260800	ALAAT2	3048,522	5066,782	52,774	6,585	0	no	689,473	522,612	31,004	4,075	0	no
Sobic.001G260701	ALAAT2	0,541	0,000	0,520	NA	1	no	0,000	0,000	0,000	NA	1	no
C4 other													
Sobic.009G240700	mMDH1	19,529	5,444	18,960	-1,800	1	no	79,239	46,016	99,625	-1,114	1	no
Sobic.003G238500	mMDH2	30,559	0,953	8,624	-3,177	1	down	15,484	6,574	12,815	-0,963	1	no
Sobic.002G309400	NAD-ME1	21,993	34,710	32,878	0,078	1	no	0,000	0,000	0,000	NA	1	no
Sobic.001G201700	NAD-ME2	13,438	17,097	33,952	-0,990	1	no	6,376	25,996	12,402	1,068	1	no
Sobic.001G432800	PCK1,PEPCK	7,651	1,782	8,705	-2,288	1	no	0,000	0,000	0,000	NA	1	no
transport													
Sobic.003G002300	APE2,TPT	1082,659	523,077	3903,597	-2,900	0	no	4902,818	1623,414	4609,610	-1,506	0	no
Sobic.009G088200	APE2,TPT	2,637	1,168	0,358	1,706	1	down	0,000	0,000	0,000	NA	1	no

		Illumina Sequencing						SuperSage					
locusName	gene-symbol	rpmk total leaf	rpmk M	rpmk B	log2 M/B	P-value	3 fold change	tpm total leaf	tpm M	tpm B	log2 M/B	P-value	3 fold change
C4 NADP-ME													
Sobic.009G062500	CUE1,PPT	22,320	8,522	8,326	0,034	1	no	0,000	0,299	0,413	-0,468	1	down
Sobic.002G159900	CUE1,PPT	4,861	0,312	2,011	-2,687	1	down	3,643	5,080	5,374	-0,081	1	no
Sobic.003G050800	CUE1,PPT	33,603	35,119	67,603	-0,945	1	no	142,084	175,997	188,088	-0,096	1	no
Sobic.004G353100	CUE1,PPT	194,165	152,628	378,450	-1,310	1,01749E-17	no	986,393	880,283	695,720	0,339	1,01749E-17	no
Sobic.004G165000	BASS	16,480	9,040	21,089	-1,222	1	no	2,732	2,988	3,307	-0,146	1	no
Sobic.009G237100	BASS	61,350	66,195	16,752	1,982	0,000404007	no	61,934	46,315	9,094	2,348	0,000404007	no
Sobic.003G236800	BASS	3,711	1,192	0,560	1,091	1	down	0,000	0,000	0,000	NA	1	no
Sobic.002G141900	NHD1	15,979	3,477	12,055	-1,794	1	no	3,643	1,195	8,681	-2,861	1	no
Sobic.007G160200	DIC2	0,221	0,000	2,410	NA	1	up	363,408	158,367	304,662	-0,944	1	no
Sobic.003G431900	MEP	95,691	1,197	1053,861	-9,783	8,0009E-174	up	0,9108	0	2,4803	NA	NA	no
Sobic.009G124000	MEP	9,230	2,913	2,666	0,128	1	down	13,6619	2,0916	7,4409	-1,830870155	1	no
Sobic.001G000800	MEP	85,110	73,701	241,883	-1,715	5,85604E-17	no	143,9059	56,1755	153,7776	-1,452832403	5,43622E-08	no
regultation													
Sobic.004G219900	ATPPCK1,PPCK1	0,671	6,038	1,041	2,536	1	up	19,127	57,371	16,949	1,759	1	no
Sobic.004G338000	ATPPCK1,PPCK1	0,203	0,000	0,000	NA	1	no	2,732	3,586	2,894	0,309	1	no
Sobic.006G148300	ATPPCK1,PPCK1	0,211	2,960	0,068	5,453	1	up	10,930	2,988	18,602	-2,638	1	no
Sobic.002G324400	PPDK RP1	66,935	63,692	24,359	1,387	0,426463451	no	399,840	319,723	42,165	2,923	0,426463451	no
Sobic.002G324500	PPDK RP1	57,217	4,745	385,412	-6,344	1,94619E-86	up	0,000	0,000	0,000	NA	1,94619E-86	no
Sobic.002G324700	PPDK RP1	13,524	2,707	25,811	-3,253	0,172839961	no	0,000	0,000	0,000	NA	0,172839961	no
PPDK associated													
Sobic.007G113600	AMK	90,606	78,516	143,023	-0,865	0,929230866	no	40,986	39,741	58,287	-0,553	0,929230866	no
Sobic.007G009200	AMK	432,031	419,019	858,267	-1,034	9,65255E-29	no	2337,104	732,374	1501,812	-1,036	9,65255E-29	no
Sobic.001G252900	PPa1	0,262	0,000	0,440	NA	1	no	0,000	0,000	0,000	NA	1	no
Sobic.006G276700	PPa3	547,808	75,042	330,830	-2,140	6,76081E-34	no	71,953	16,434	132,282	-3,009	6,76081E-34	no
Sobic.003G367650	PPa3	1,557	31,622	1,903	4,055	0,000817402	up	0,000	0,000	0,000	NA	0,000817402	no

		Illumina Sequencing						SuperSage					
locusName	gene-symbol	rpmk total leaf	rpmk M	rpmk B	log2 M/B	P-value	3 fold change	tpm total leaf	tpm M	tpm B	log2 M/B	P-value	3 fold change
PPDK associated													
Sobic.009G152600	PPa3	1,626	5,503	0,503	3,453	1	no	0,000	0,000	0,000	NA	1	no
Sobic.004G268000	PPa4	249,840	108,758	699,846	-2,686	6,3416E-99	no	269,596	57,670	627,512	-3,444	6,3416E-99	no
Sobic.009G016000	PPa4	453,310	658,176	1936,839	-1,557	1,0496E-134	no	249,558	223,208	212,891	0,068	1,0496E-134	no
Sobic.004G311100	PPa6	93,319	103,226	203,877	-0,982	0,000738658	no	1515,566	1132,476	1333,567	-0,236	0,000738658	no

Supplementary Table 4. Gene-specific primers used for qPCR and RNA *in situ* analysis.

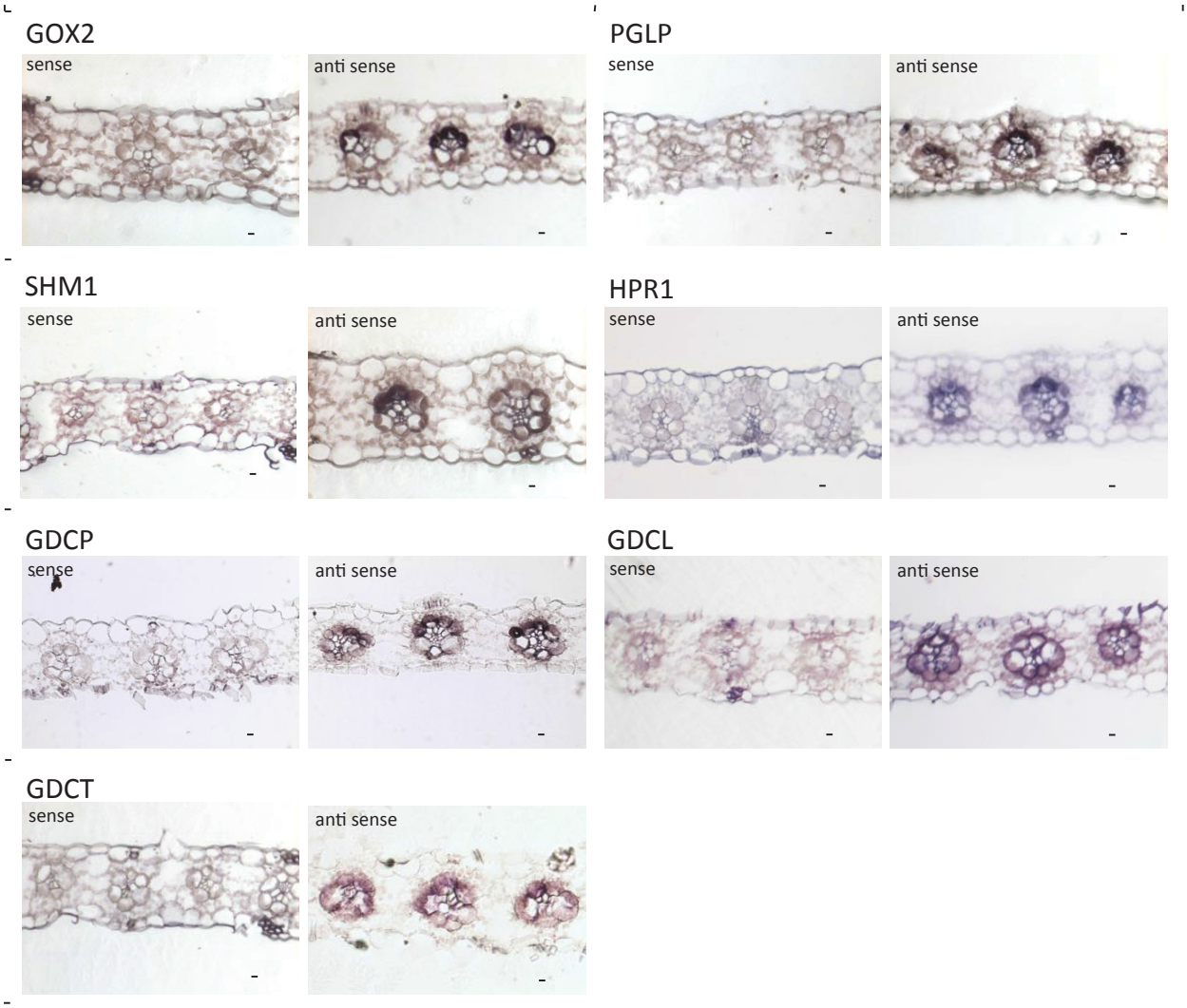
F: forward primer; R: reverse primer.

In Situ

Primer designation	Sequence	Orientation
IS-AGT1	TCGACTGGAAGGACTACCTCA	F
IS-AGT1	AATCTGGGCAGGGCTATACA	R
IS-DIT1	GTGTCTGCAGCAATTCTTGG	F
IS-DIT1	GAGAACGATCCCCAAGATCC	R
IS-DIT2	TGGGATACATTGGCTTGGTT	F
IS-DIT2	TAGAGATTGGTCGCGCTTTT	R
IS-GDCH	GCCTTTGCCTTGTCAATCTC	F
IS-GDCH	CATGGGGATTTTAAGAATCTGG	R
IS-GDCL	ACCACCTGCATCGAGAAGAG	F
IS-GDCL	ATGGTCTGCTCACCACCAG	R
IS-GDCP	CAAGCCCTGGTTTTATTGGA	F
IS-GDCP	CTGCAGTGTGCAGATGAGGT	R
IS-GDCT	CCACATCTACCTCGTCGTCA	F
IS-GDCT	CGAGGGCCTGTAGTACTTGG	R
IS-GGT	TGCCCTATTCCACAGTACCC	F
IS-GGT	ACAACAATCCCAGTGGCTTC	R
IS-GLS1	CAAACTGGACAGGTGCTTGA	F
IS-GLS1	TGGCACCATATCCAATCAGA	R
IS-GLYK	ATCGGCTACACCGATGAGAA	F
IS-GLYK	TCCAGGTTTTCCATCTGCTC	R
IS-GOX2	ACAGGAAAGTGGTGGAGCAG	F
IS-GOX2	ACTCGGTGATGATGTGGTTG	R
IS-GS2	GGTTGCTTTGCTGAAGGTCT	F
IS-GS2	CCTGTGCTTGACCCATCATA	R
IS-HPR1	GGTGACCCTCCCATCTACTTTAC	F
IS-HPR1	AGCAGCCTATATGCACATGTTTT	R
IS-ME	ATCAGTGGTCCTTGCAGGTT	F
IS-ME	GCATATGCCACCAGGTCTTT	R
IS-PEPC	AAGTACGGCGTGAAGCTGAC	F
IS-PEPC	GGGTGGTGATGTAGGGATTG	R
IS-PGLP	CCGTCGAGACCTTCATCTTC	F
IS-PGLP	AATCTGGCTGGATCGAGTTG	R
IS-PPDK	GAAGCTGTATGGCGAGTTCC	F
IS-PPDK	CATTCAGCGATAGCCACTCA	R
IS-RBCS	TTCCAGGGTCTCAAGTCCAC	F
IS-RBCS	ACCAGAGCAAACTCCAATGC	R
IS-SHM1	GCGACCTCCATCTACTTCCA	F
IS-SHM1	CACGGCTTGGTGAAGGTACT	R

qPCR

Primer designation	Sequence	Orientation
qPCR-ACTIN	CGTTTATTTGCATCGGACCT	F
qPCR-ACTIN	TAGCACCAGCACAATCCAAG	R
qPCR-AGT1	CGACTACGACGACGAATGG	F
qPCR-AGT1	GACCAACAGCTGAGTGTGGA	R
qPCR-DIT1	ACTATGGCATTGGCTCTGCT	F
qPCR-DIT1	TCGCAGTTAATGTGGTGCTC	R
qPCR-DIT2	TGGAGCTGGGTATCTTGACC	F
qPCR-DIT2	GCTGCTGTCATCACTTCCAA	R
qPCR-G6PI	ACCGGCCATCACTAAGTTTG	F
qPCR-G6PI	ATGAAGCCCTGAACTGCAAC	R
qPCR-GDCH	CTGATCAACACGAGCCCTTAC	F
qPCR-GDCH	ACTGCGTCATAAGCATGCAC	R
qPCR-GDCL	AAGGTGATCGCAGAGAAGGA	F
qPCR-GDCL	GAGGGCTTCGCTCACTGTAG	R
qPCR-GDCP	CGACTTGGGCAACCTGTTAT	F
qPCR-GDCP	TGCTGAGACAGGAATGATGC	R
qPCR-GDCT	CGCTATGGGCTACGTGAAAT	F
qPCR-GDCT	GCAACAAGAGAAATGCAAAGC	R
qPCR-GGT	GGTTCTGAGCCCTGATGTGT	F
qPCR-GGT	AATGTCAAATGGGCAGAAGC	R
qPCR-GLS1	GCCAGGACGTCCACACTAAT	F
qPCR-GLS1	CTTGGCAATCACACCTGCTA	R
qPCR-GLYK	GATCCACAGCTTGAGGTGGT	F
qPCR-GLYK	TCCAGGTTTTCCATCTGCTC	R
qPCR-GOX2	CGTTCCTTCTCTTCATCATTCAGA	F
qPCR-GOX2	CAACAAGGAACCAAGCCAGTTAC	R
qPCR-GS2	CACAGGTGAGTTGAATCCGC	F
qPCR-GS2	TCACCACGGAATGCAGCAG	R
qPCR-HPR1	CAAAGCTTTAGGAGCCAACG	F
qPCR-HPR1	TTATACCGGAACCCAAGCTG	R
qPCR-PFK	GGTGCTGGACAGGATCTCAT	F
qPCR-PFK	AGGTCGGGTCGATGTACTTG	R
qPCR-PGLP	GGCAAAGAAGTTCGGAATCA	F
qPCR-PGLP	AATCTGGCTGGATCGAGTTG	R
qPCR-PPDK	AGTTGAGAACTGCAACATAC	F
qPCR-PPDK	CTGGTTCTACCATCTTAATC	R
qPCR-RPN	CAATGCCTCAAGCTCAAACA	F
qPCR-RPN	TCAGTCACAGCTGCAGATCC	R
qPCR-SHM1	GAGCAAGGGCTACAAGTTGG	F
qPCR-SHM1	GGTGTCATTGCACTGCTGTC	R



Supplementary Figure 1. RNA *in situ* hybridization of *Sorghum bicolor* leaves with probes for transcripts related to photorespiration.

GDCL/GDCP/GDCT, glycine decarboxylase L, P, and T subunit; GOX2, glycolate oxidase; HPR1, hydroxypyruvate reductase; PGLP, phosphoglycolate phosphatase; SHM1, serine glyoxylate aminotransferase. Scale bars = 50µm.

The authors' contributions

FD and **UG** wrote this manuscript. **FD** performed all the experiments except those listed below.

UG performed the SuperSage and the Illumina transcriptome sequencing of *Sorghum bicolor* mesophyll and bundle sheath samples. He assembled the transcriptome data and did most of the mapping and statistics.

MS isolated mesophyll and bundle sheath RNA from *Sorghum bicolor*.

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

Manuscript 2

An EMS-based genetic screen to identify bundle sheath ontogeny and maintenance (*BSOM*) genes in *Arabidopsis thaliana*.

An EMS-based genetic screen to identify bundle sheath ontogeny and maintenance (*BSOM*) genes in *Arabidopsis thaliana*

Florian Döring[1]*, Kumari Billakurthi[1,2], Udo Gowik[1], Stefanie Sultmanis[3], Roxana Khoshravesh[3], Shipon Das Gupta[1], Tammy Sage[3] and Peter Westhoff[1,2]

[1]Institute of Plant Molecular and Developmental Biology, Universitätsstrasse 1, Heinrich-Heine-University, 40225 Duesseldorf, Germany

[2]Cluster of Excellence on Plant Sciences 'From Complex Traits towards Synthetic Modules', 40225 Duesseldorf and 50923 Cologne, Germany

[3]Department of Ecology and Evolutionary Biology, The University of Toronto, Toronto, ON, Canada M5S 3B2

*To whom correspondence should be addressed

Florian Döring
e-mail: florian.doering@hhu.de

Kumari Billakurthi
e-mail: kumari.billakurthi@hhu.de

Udo Gowik
e-mail: gowik@hhu.de

Stefanie Sultmanis
e-mail: stefanie.sultmanis@mail.utoronto.ca

Roxana Khoshravesh
e-mail: r.khoshravesh@utoronto.ca

Shipon Das Gupta
e-mail: shipan.gupta@hhu.de

Tammy Sage
e-mail: tammy.sage@utoronto.ca

Peter Westhoff
e-mail: west@hhu.de

Publication in preparation (The Plant Journal).

Abstract

The evolution of C_4 photosynthesis is a stepwise sequence of small changes leading from C_3 to C_4 plants in which each step on its own increases the fitness of the plant. One important precondition for the introduction of a functional C_4 cycle is the activation of the bundle sheath. This includes the enlargement of bundle sheath cells and an increase in organelle number inside of them. However, most C_3 plants possess only small bundle sheath cells that are sparsely populated with organelles. To engineer C_4 photosynthesis into existing C_3 crops, information about genes controlling bundle sheath cell size and organelle content is needed. Since very little is known so far about the specific genes that could be manipulated in order to create a more C_4-like bundle sheath, we established an EMS-based genetic screen in the Brassicaceaen C_3 species *Arabidopsis thaliana*. To this end, we created two reporter gene lines in which the bundle sheath cells of *A. thaliana* were labeled with easily detectable reporter genes. The activity of the reporter genes in the leaves was used as a proxy to identify mutants with altered bundle sheath anatomy. The mutant screen resulted in dozens of interesting mutants with alterations in the anatomy of bundle sheath cells and vascular tissue.

Introduction

C_4 plants are superior to C_3 plants with regard to their photosynthetic performance under many conditions, such as high light intensity, high temperature, and drought (Ehleringer et al., 1991). Since many years, researchers all over the world have been focusing on this special type of photosynthesis and try to understand the complex details of its biochemistry and evolution. It is still a major goal to introduce the C_4 type of photosynthesis into existing C_3 crops in the future to increase yield and meet increased food demands—in the face of a growing world population (Sheehy et al., 2007).

C_4 photosynthesis is characterized by the spatial separation of CO_2 uptake and fixation by ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO). This is guaranteed by the division of labor and the interaction between two different cell types, the bundle sheath and the mesophyll, in order to concentrate CO_2 in the bundle sheath around the RubisCO enzyme. In a typical C_4 plant, one layer of bundle sheath cells surrounds the vasculature in a wreath-like arrangement followed by one layer of mesophyll cells. This special order of cells in the leaves is termed Kranz anatomy (Haberlandt, 1904). C_4 photosynthesis occurs only in the angiosperms (Ehleringer et al., 1997) and evolved at least 66 times independently (Sage et al., 2011; Sage et al., 2012) indicating

that it is probably easy to evolve from C_3 to C_4 from a genetic point of view. The evolution of the C_4 trait proceeded step by step and each small change contributes to the general fitness of the plant (Gowik and Westhoff, 2011; Heckmann et al., 2013). The establishment of a functional photorespiratory CO_2 shuttle that shifts the photorespiratory CO_2 release to the bundle sheath cells is a key step during C_4 evolution and is also referred to as C_2 photosynthesis (Sage et al., 2014). C_2 photosynthesis serves as the basis upon which C_4 photosynthesis can develop and is therefore considered to be the evolutionary bridge between C_3 and C_4 photosynthesis (Sage et al., 2012; Mallmann et al., 2014; Bräutigam and Gowik, 2016). However, before a photorespiratory CO_2 pump can evolve, some essential morphological requirements at the level of the bundle sheath have to be met. This includes an increase in vein density, which is probably an adaption to hot and dry climates. Additionally, more veins raise photosynthetic activity in the bundle sheath due to a higher bundle sheath/mesophyll ratio in the leaf (Roth-Nebelsick et al., 2001; Lundgren et al., 2014). The combination of larger bundle sheath cells, enhanced organelle number, and the relocation of organelles within the bundle sheath cells may already create conditions in which photorespiratory CO_2, produced in the bundle sheath, is partly captured. This situation is termed proto-Kranz anatomy and, consequently, it could allow plants to survive a loss of glycine decarboxylase (GDC) expression in the mesophyll to establish C_2 photosynthesis (Sage et al., 2012; Sage et al., 2013).

Parenchymatous bundle sheath tissue also occurs in C_3 species; however, it is often not very well defined and contains only few chloroplasts, indicating that this tissue does not play a major role in leaf photosynthesis of C_3 plants (Kinsman and Pyke, 1998; Leegood, 2008). The exact physiological role of bundle sheath cells in C_3 plants is currently not understood. It is assumed that they function in phloem loading and unloading and may provide mechanical strength within the leaf (Van Bel, 1993; Kinsman and Pyke, 1998; Griffiths et al., 2013). However, studies with bundle sheath gene promoters (glycine decarboxylase P protein) from the Asteracean C_4 species *Flaveria trinervia* (Engelmann et al., 2008) showed that the expression specificities were maintained in the Brassicacean C_3 species *Arabidopsis thaliana*. This indicates that cryptic Kranz anatomy is present in C_3 species, and that the bundle sheath is an evolutionarily ancient invention in angiosperms (Westhoff and Gowik, 2010). The introduction of C_4 photosynthesis into C_3 species requires knowledge about the genes that are needed for the formation of a pronounced and photosynthetically active bundle sheath. Currently, very little is known of specific genes that control these processes and that might help for the construction of a large and organelle-rich bundle sheath in C_3. This knowledge is essential for the long-term goal of introducing C_4 photosynthesis into C_3 crops. Since the gene regulatory system operating in bundle sheath cells

of C_4 and C_3 monocots and dicots are at least partly conserved, we hypothesize that the C_3 model species *A. thaliana* can be used for the discovery of conserved genes that are required for the activation of the bundle sheath in angiosperms (Engelmann et al., 2008). When information on genes of a particular process is limited, the use of forward genetics is a powerful tool to identify genes related to the process of interest. In contrast to reverse genetics, it is completely unbiased and takes the whole genome into account to uncover players of specific biological phenomena. A successful forward genetic screen is defined by a high throughput and a reliable and robust screen for mutants in which thousands of plants have to be analyzed (Page and Grossniklaus, 2002). For this reason, Arabidopsis is best suited since it is small and simple to grow, it has a fully sequenced genome, short generation times, and it can be transformed by *Agrobacteria tumefaciens* with ease (Somerville and Koornneef, 2002).

Our main interest is the morphology and organelle content of the bundle sheath, which naturally is not very pronounced in Arabidopsis but rather small and not easily detectable. To this end, we modified the genetic background of Arabidopsis to facilitate the identification of the bundle sheath with reporter genes that were expressed under the control of the promoter of the glycine decarboxylase P gene of *Flaveria trinervia* (Engelmann et al., 2008). This promoter is highly active in the bundle sheath and vasculature in *A. thaliana*. In this study, we describe a forward genetic screen based on the chemical mutagen ethyl methanesulfonate (EMS) with bundle sheath-labeled reporter lines in *A. thaliana*. The level of reporter gene expression served as a proxy to collect mutants with altered bundle sheath anatomy. Microscopic analysis of selected mutants revealed anatomical changes in bundle sheath and vascular tissue.

Material and Methods

Plant material

A. thaliana (Ecotype Columbia) plants were used as genetic background for both reporter gene lines. The plants were grown under greenhouse conditions with supplementary light for 14 h per day at a photon flux density (PFD) of ~300 µmol m^{-2} s^{-1} or in climate chambers operating at 16 h light/8 h of darkness periods (~60 µmol m^{-2} s^{-1}) and a constant temperature of 21–22 °C. The seeds were surface-sterilized with a chloric solution containing 20% Dan Klorix (Colgate-Palmolive, Hamburg, Germany) and 0.02% Triton X-100 for 5 min and washed four times with sterile water. After sterilization, the seeds were incubated at 4 °C in the dark for at least 48 h before they were sown on either soil (Floraton 1, Floragard, Oldenburg, Germany) or petri dishes with ½ Murashige and Skoog (MS)-medium containing 0.6% agar and 1% sucrose.

Construction of reporter gene lines

The pGreen Gateway vector pGreen-*LUC68* containing the firefly luciferase 68 gene served as a backbone for the luciferase (*LUC*) reporter construct and was kindly provided by Franziska Turck (Adrian et al., 2010). The promoter region of the *GLDPA* gene from *Flaveria trinervia* was amplified by PCR by using the p*GLDPA*$_{Ft}$ construct (Engelmann et al., 2008) as a template and specific oligonucleotides that added *att*B1 and *att*B2 sites to the PCR product (Supplementary Table 1). The BP Clonase reaction (Gateway® BP Clonase® enzyme mix, ThermoFisher Scientific) between the PCR product and the Gateway entry vector pDONR221 was performed according to the Gateway manual and resulted in pENTRY221-p*GLDPA*$_{Ft}$, which was subsequently used for the LR Clonase reaction (Gateway® LR Clonase® enzyme mix, ThermoFisher Scientific) to introduce p*GLDPA*$_{Ft}$ into pGreen-*LUC68* (pGreen-p*GLDPA*$_{Ft}$::*LUC68*). We used the binary plant transformation vector pBI121 (Clonetech laboratories; Jefferson et al., 1987) to assemble the *GFP* reporter gene construct. The p*GLDPA*$_{Ft}$ promoter region (Engelmann et al., 2008), the RbcS transit peptide *RbcS.TP* from *A. thaliana* (Kim et al., 2010), and the s*GFP* gene were amplified by PCR by using the p*GLDPA*$_{Ft}$ construct (Engelmann et al., 2008), cDNA of Arabidopsis (Ecotype Columbia) and pGWB5 (Nakagawa et al., 2007) as templates, respectively. Appropriate restriction sites were added to all oligonucleotides (Supplementary Table 1). p*GLDPA*$_{Ft}$ was cloned as a *Hind*III-*Bam*HI fragment into a *HInd*III-*BamH*I-digested pBI121 (pBI121- p*GLDPA*$_{Ft}$), followed by the ligation of *RbcS-TP* as a *BamH*I-*EcoR*I fragment into *BamH*I-*EcoR*I-digested pBI121-p*GLDPA*$_{Ft}$ (pBI121-p*GLDPA*$_{Ft}$::*RbcS.TP*). Subsequently, the s*GFP* gene with *EcoR*I and *Sac*I restriction sites was introduced into *EcoR*I-*Sac*I-digested pBI121-p*GLDPA*$_{Ft}$::*RbcS.TP* resulting in the final reporter gene construct pBI121-p*GLDPA*$_{Ft}$::*RbcS.TP*-s*GFP*.

Transformation of *A. thaliana*

Both reporter gene constructs were transformed by electroporation into the *Agrobacterium tumefaciens* strain AGL1 (Lazo et al., 1991), and subsequently transformed into *A. thaliana* (Ecotype Columbia) by using the floral dip method (Zhang et al., 2006). T1 plants were analyzed by PCR to check for the presence of the reporter gene construct. Positive lines were carried to the T3 generation, and homozygous plants were selected for the mutant screens.

EMS mutagenesis

Approximately 40,000 seeds (~1.6 g) of the *LUC* reporter gene line and 120,000 (~4.8 g) seeds of the *GFP* reporter gene line were used for EMS mutagenesis. The seeds were initially washed with 0.1% TWEEN® 20 for 15 min, after which EMS (Sigma-Aldrich, St. Louis) was added to a final concentration of 0.25%. The mixture incubated for 16 h on a rotating platform. Subsequently, the seeds were washed four times with sterile water, incubated again for 1 h on a rotating platform, and washed one last time in sterile water. After three days at 4 °C, the seeds were sown evenly on soil, and M2 seeds were harvested in pools of about 30–50 plants. Each pool of M2 seeds was sown individually, and 14–17-day-old plants were used for the genetic screen.

Mutant screen

The first leaf pair of M2 plants was analyzed for aberrant reporter gene expression. In principle, plants with more, less, or diffused reporter gene signal were selected at this point. The screen for LUC activity was performed with the imaging system Night Owl LB983-NC100U (Berthold Technologies, Bad Wildbad, Germany) by using the in vivo imaging software indiGO (Berthold Technologies, Bad Wildbad, Germany). For this purpose, the leaves were incubated in a 1 mM luciferin solution for 5 min after which LUC activity was detected (exposure time: 120 s). The resulting signal in the bundle sheath of the EMS-mutagenized M2 populations was compared to the non-mutagenized reporter line. M2 plants with the GFP reporter gene were screened for aberrant *GFP* expression with the aid of a fluorescence binocular microscope (Nikon SMZ25, Düsseldorf, Germany). In the M3 generation all primary mutants were analyzed again to confirm the individual aberrant phenotype. Additionally, the signal intensity was measured for whole leaves and normalized to the leaf area with the software ImageJ (Schneider et al., 2012). Only mutant lines with at least 30% stronger or weaker signal intensities in the whole leaf were selected for further studies. The DNA was isolated from the mutant lines to check for point mutations in the reporter gene construct. The complete region (pBI121-p*GLDPA*$_{Ft}$::*RbcS.TP* or p*GLDPA*$_{Ft}$::*LUC68*) was amplified by PCR using the Phusion High-Fidelity DNA Polymerase (New England Biolabs), cloned into pJet1.2/blunt vector (ThermoFisher Scientific), and subsequently sequenced. Any mutant lines with point mutations within the reporter gene constructs were not used for further analysis.

Microscopic imaging

Sampling and preparation for light microscopy (LM) and transmission electron microscopy (TEM) was operated as described in Akhani and Khoshravesh (2013). This included (1) tissue fixation and dehydration, (2) resin infiltration and embedding, and (3) sectioning and staining. We took the fully expanded 2nd leaf pair of four-week-old Arabidopsis plants, removed the edges and the midvein of the leaf, and cut the remaining area into 1–2 mm^2 small sections that were immediately fixed with a fixative (1% glutaraldehyde, 1% paraformaldehyde, 0.1 M sodium cacodylate). These samples were post-fixed with OsO_4, followed by a series of ethanol dehydrations and Araldite resin infiltration and embedding. The resin-embedded tissue blocks were cut with a microtome (Microm HM 330 Microtome) to obtain sections of 1.5–2 micron thickness, which were subsequently stained with toluidine blue for LM. The same blocks were also used to create 50–70 nm thin sections for TEM by using a RMC MT-7 ultra-microtome (Boeckeler Instruments). Subsequently, the sections were stained with uranyl acetate and lead citrate.

Mapping by sequencing

Stable M4 mutant lines with intact reporter gene sequences were backcrossed with the corresponding non-mutagenized reporter gene line. F2 plants were scored for the individual aberrant phenotype, and genomic DNA was isolated from bulked leaf samples of 50–60 plants with the DNeasy Plant Maxi Kit (Qiagen, Hilden, Germany). DNA was eluted with 750 µl sterile water in two steps and concentrated to at least 50 ng/µl by vacuum infiltration. Sequencing libraries of the bulked mutant DNA as well as of the two original reporter gene lines were prepared as follows: 1 µg of each DNA sample was sheared with a Covaris S2x system (Covaris, Woburn, MA, USA) to a size of approximately 350 bp. The DNA library preparation was done with the TruSeq DNA PCR-Free LT Library Preparation Kit (Illumina) according to the manual. The library concentration was determined with the KAPA Library Quantification Kit Illumina® platforms (Kapabiosystems). The paired end sequencing (2 x 150 bp) was performed using an Illumina HiSeq3000 system by the "Genomics and Transcriptomics laboratory" of the Biologisch-Medizinisches Forschungszentrum at Düsseldorf University with 80 to 500 fold coverage. EMS induced mutations potentially responsible for the mutant phenotypes were identified by using SHOREmap v3.0 following the backcross procedure as described (http://bioinfo.mpipz.mpg.de/shoremap/guide.html; Sun and Schneeberger, 2015). The read mapping and SNP calling were performed by using SHORE v0.9.3 and Genomemapper v0.4.4.

Results

Construction of bundle sheath-labeled reporter gene lines

Since bundle sheath cells of C_3 plants are only detectable under microscopic examination, changes in their properties cannot be easily visualized non-destructively. We have therefore labeled the bundle sheath cells of *A. thaliana* by expressing firefly luciferase (*LUC*) or a chloroplast-located GFP (Sattarzadeh et al., 2010) under the control of the promoter of the glycine decarboxylase P protein gene (*GLDPA*) of the C_4 Asteraceaen species *F. trinervia* (Engelmann et al., 2008). The p*GLDPA*$_{Ft}$ promoter is active in bundle sheath cells and vascular tissue. Both reporter gene constructs were transformed into *A. thaliana* (Ecotype Columbia). Transgenic lines with high reporter gene activity were identified, and homozygous lines were obtained by selfing. Figure 1 shows the expression patterns of both reporter gene constructs in Arabidopsis leaves. In addition, leaf cross sections of the *GFP* reporter gene line allowed to identify single chloroplasts due to the RbcS transit peptide (Figure 1 B2,B3).

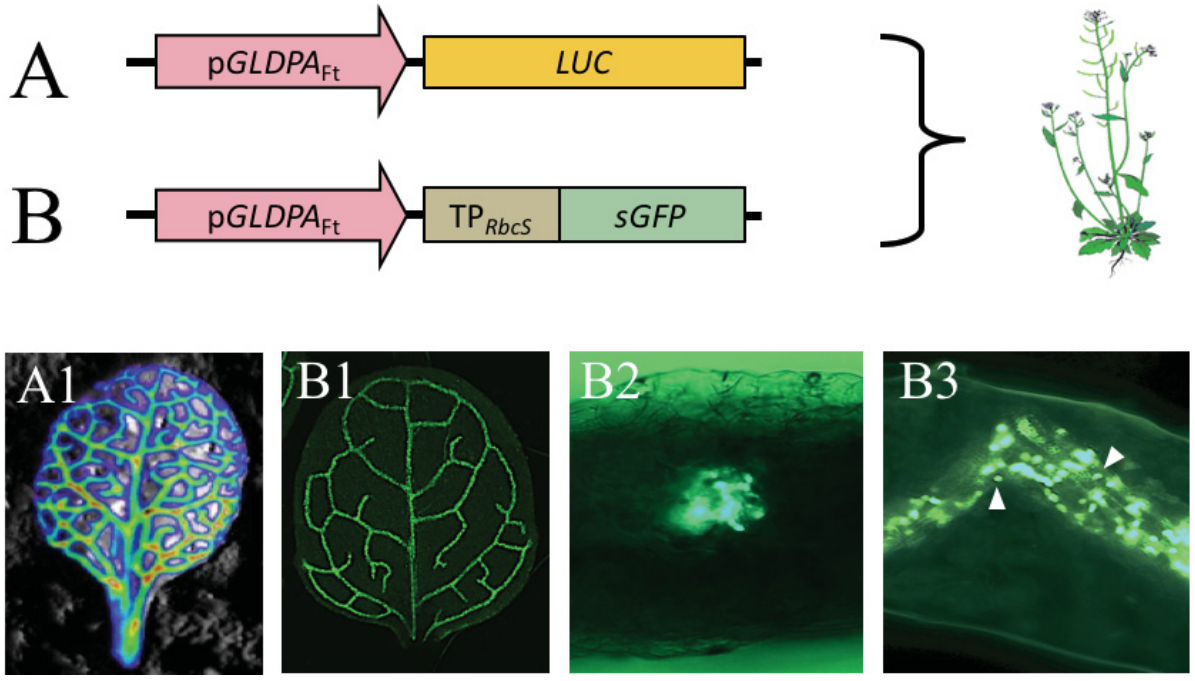


Figure 1. Labeling the bundle sheath in Arabidopsis leaves with reporter genes.

Construction of the *LUC*- (A) and *GFP* (B) reporter gene lines. Both reporter genes are under the control of the p*GLDP*A$_{Ft}$ promoter region of the C_4 dicot *Flaveria trinervia*. A chloroplast transit peptide of the small subunit of RubisCO was fused to the *GFP* gene. Luminescence (A1) and fluorescence imaging (B1) show the reporter gene expression patterns in young Arabidopsis leaves. Additionally, the GFP signal is restricted to chloroplasts in the bundle sheath and vasculature as can be seen in leaf cross sections (B2 + B3, arrowheads).

EMS-based genetic screen with bundle sheath-labeled reporter gene lines

160,000 seeds were mutagenized with EMS (40,000 seeds of the *LUC* reporter gene line, and 120,000 seeds of the *GFP* reporter gene line) and sowed on soil in large flats under greenhouse conditions. We observed a survival rate of 50% in the M1 generation and harvested seeds from the remaining 80,000 M1 plants in pools of 30–50 plants. Approximately 45,000 M1 plants were needed under the given EMS concentration to have a 95% chance of exploring a mutation in any given G:C base pair (Jander et al., 2003). Therefore, we have reached a saturating EMS screen by mutagenizing most G:C base pairs in the genome of *A. thaliana*. We expected the number of mutations per genome to be randomly distributed, thus following a Poisson distribution and calculated approximately one embryonic-lethal mutation per mutagenized genome (Pollock and Larkin, 2004). In addition, 2.2% of the plants in the M2 generation displayed a pale chlorophyll phenotype; therefore, the EMS treatment could be considered as a success (Kim et al., 2006). Each M2 pool was sown individually on large flats in the greenhouse, and single leaves or whole seedlings were screened for aberrant reporter gene expression (e.g. stronger or weaker reporter gene signal in the bundle sheath). 755 primary mutants were identified: 258 mutants with the *LUC* background and 497 mutants with the *GFP* background (Figure 2).

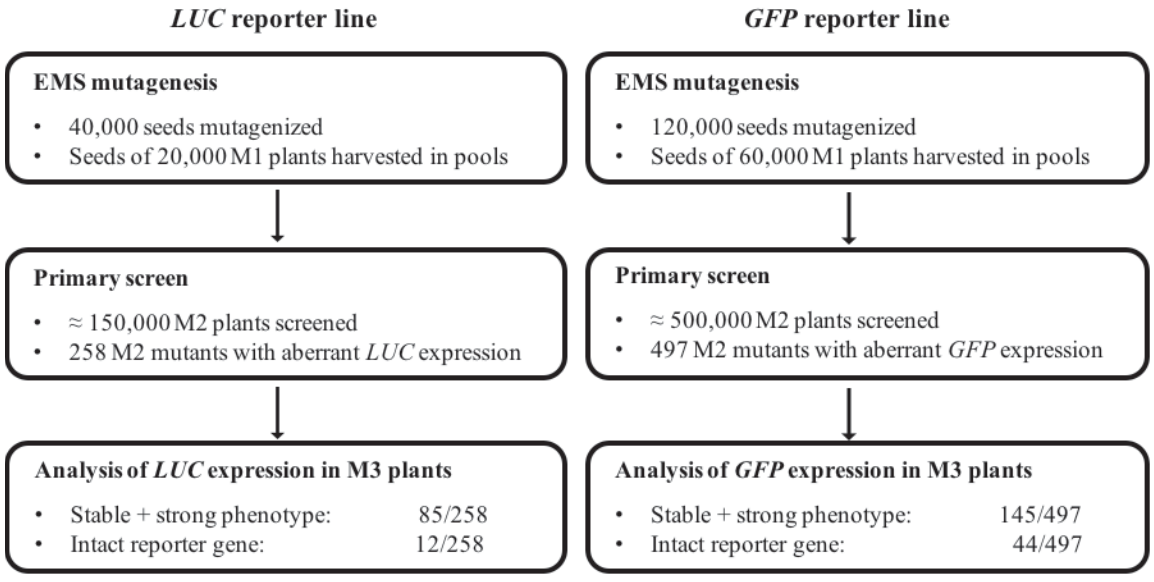


Figure 2. Workflow of the EMS-based genetic screens.

Workflow of both EMS screens (*LUC* and *GFP* reporter gene line). Altogether, seeds of about 80,000 M1 plants were harvested in pools of 30–50 plants. 650,000 M2 plants were screened for aberrant reporter gene activity, which resulted in more than 750 primary mutants. 56 mutant lines possessed stable aberrant phenotypes and intact reporter gene sequences.

We checked the phenotype of each mutant line in the M3 generation for its stability and quantified the reporter gene signal of each mutant line in single leaves using ImageJ software. A threshold of at least 30% difference in signal intensity between each mutant line and the reference line was chosen to exclude minor changes and to focus only on strong differences. Thereafter, 230 mutants remained (85 mutant lines with the *LUC* background and 145 mutant lines with the *GFP* background) (Figure 2). To exclude aberrant phenotypes that are based on point mutations in the appropriate reporter gene, we amplified and sequenced the whole reporter gene construct. Almost 75% of our mutant lines had to be discarded at that point due to mutations in this region. Nevertheless, 12 mutant lines with the *LUC* reporter gene and 44 mutant lines with the *GFP* reporter gene were both stable and contained intact reporter genes.

Analysis of EMS-generated mutant lines

Our primary screening criterium was based on the reporter gene signal intensity (Supplementary Figure 1). Additionally, we obtained several mutants in which the reporter gene signal was clearly detectable in the mesophyll. We considered some of these mutants for our further analysis as the loss in tissue-specificity of our reporter gene might be linked to altered bundle sheath or mesophyll development or to mutations in genes affecting the transcription and/or post-transcriptional regulation of the p*GLDPA*$_{Ft}$ promoter.

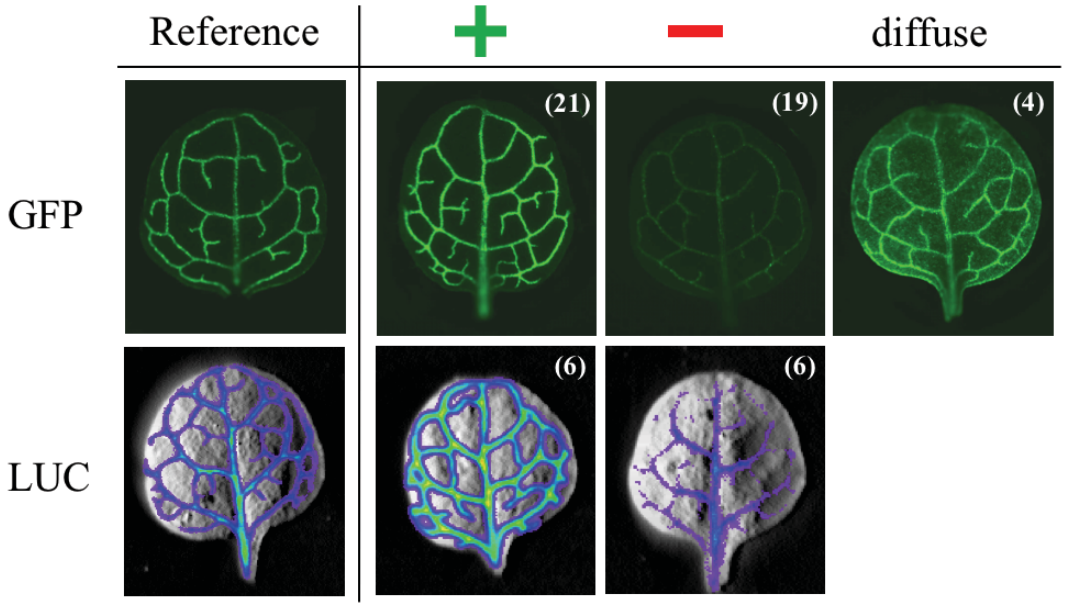


Figure 3. Categorization of 56 stable EMS mutant lines.

The selected mutant lines showed either more signal in the bundles (+), less signal in the bundles (-), or a diffuse signal that was also present in the mesophyll cells (diffuse). Numbers in brackets indicate the total quantity of mutant lines of each category. No mutants with a diffuse signal were obtained with the *LUC* reporter gene line.

In summary, the EMS-based mutant screen resulted in 21 and 6 mutant lines with increased reporter gene activity in the *GFP* and *LUC* background, respectively. Additionally, we obtained 19 and 6 mutants with less reporter gene activity in the *GFP* and *LUC* reporter lines, respectively (Figure 3). Four mutant lines possessed a diffuse GFP signal. Intriguingly, seven mutants with increased GFP signal intensity also contained larger bundle sheath strands (vasculature and bundle sheath) (Figure 4A). We assumed that this might be caused by either an increase in vasculature or bundle sheath, or a combination of both (Figure 4B).

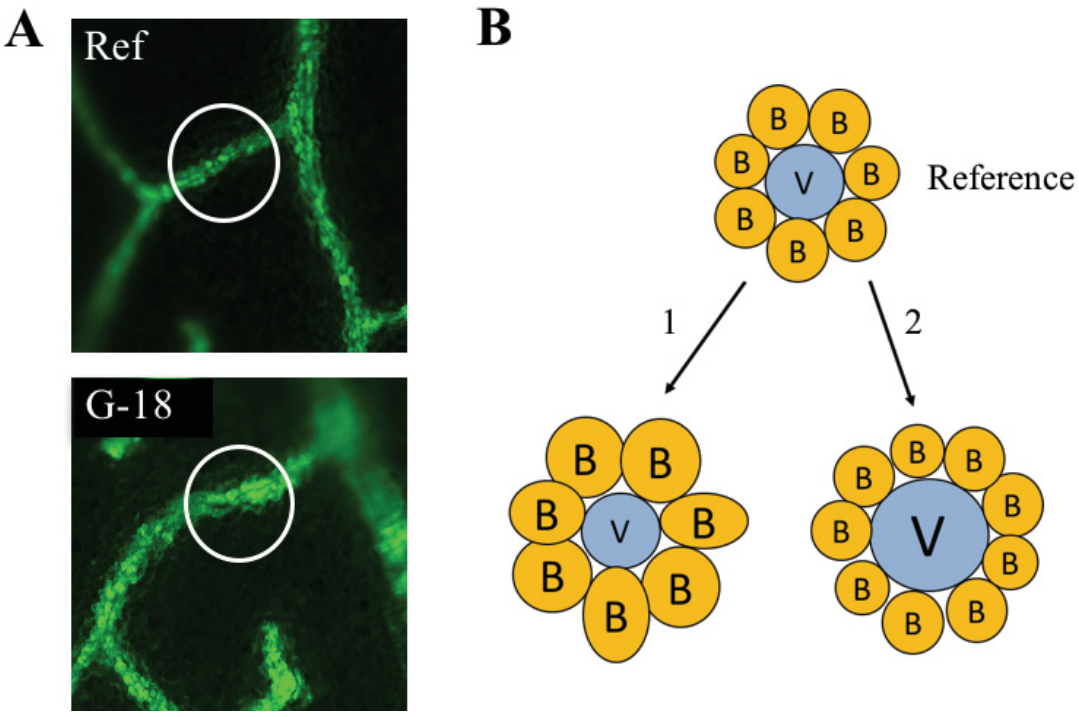


Figure 4. Larger bundle sheath strands in mutant lines with increased GFP signal.

(A) A close-up view of 3rd order veins of the reference line and mutant line G-18 as an example. In addition to an increased GFP signal intensity, we also observed larger bundle sheath strands in mutant line G-18 and six other mutant lines with the GFP background. (B) Possible explanations for larger bundle sheath strands: (1) Larger bundle sheath cells or (2) larger vasculature. B, bundle sheath cell; V, vasculature.

However, we could not clearly assign the change in reporter intensity to anatomical alterations in our mutants. To address this issue, we performed high-resolution microscopy and prepared samples for LM as well as TEM of a subset of 25 mutant lines (G01–G25). In general, we tried to select the strongest phenotypes in terms of signal intensity and width of the bundle sheath strands. Twenty mutant lines possessed an increased reporter gene signal whereas three mutant lines were chosen with less signal. Additionally, we included two mutant lines with a diffused reporter signal. We harvested three biological replicates for each mutant line and prepared leaf cross sections for LM of all samples. Subsequently, sections from each replicate were compared to sections of the reference line, and only higher-order veins were analyzed with respect to the

anatomy of the bundle sheath and vasculature. We took only those samples into account in which we could verify a phenotype in comparable veins in all three replicates. As a consequence, we could not draw a conclusion in eight mutant lines due to an insufficient sample size. In eleven mutant lines we could not find any differences in the anatomy of the bundle sheath cells and/or vascular tissue (Supplementary Table 2). However, we identified six mutants with clear differences in either bundle sheath cell number and/or vascular tissue size. Four mutants were defined by both an enlarged vasculature and more bundle sheath cells (mutant category 1), while two mutant lines possessed an enlarged vasculature but no obvious change to the bundle sheath cells while (mutant category 2). Mutant line G-19 and G-20 are shown in Figure 5 as an example for mutant category 1 and 2, respectively.

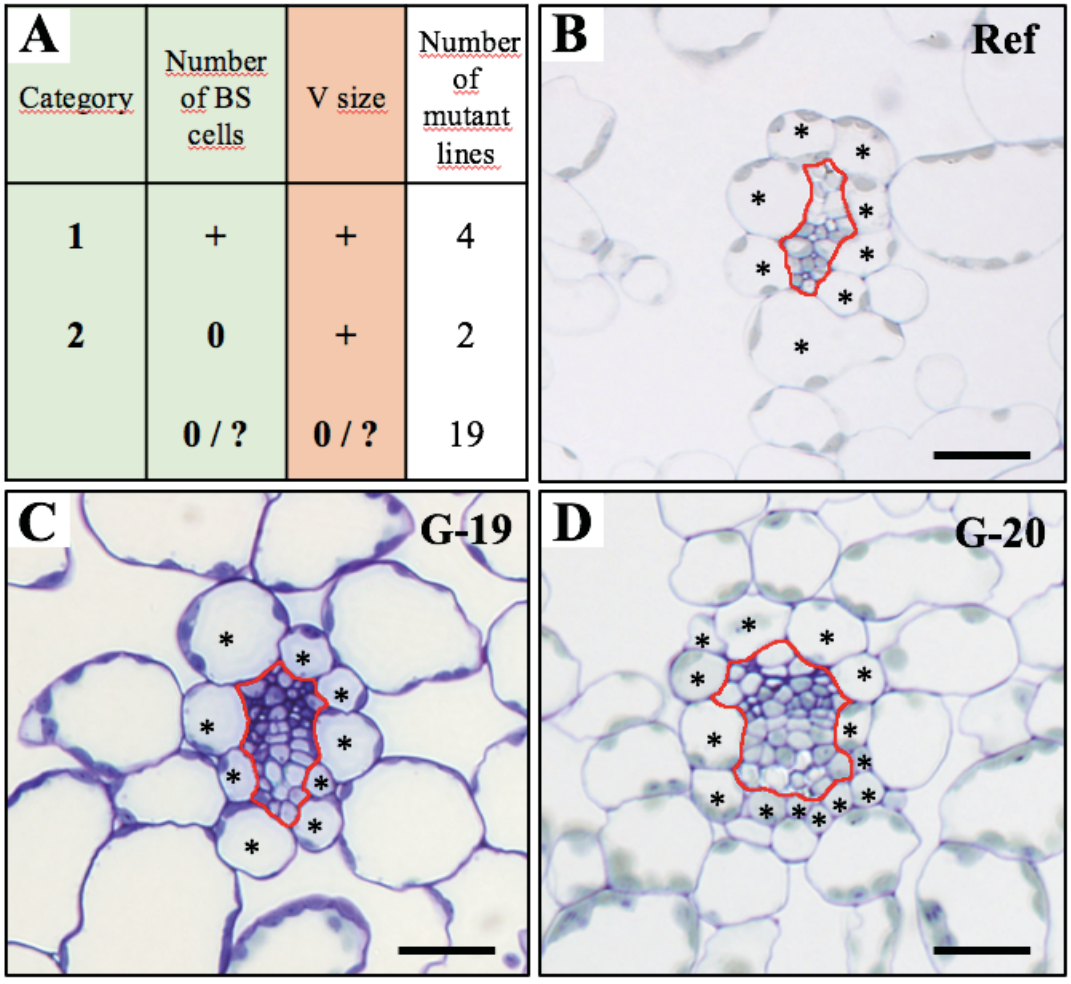


Figure 5. Light microscopic images of leaf cross sections from selected EMS mutant lines

(A) Categorization of the EMS mutant lines after LM. (B, C, and D) Leaf cross sections with representative 3° veins of (B) the reference line, (C) mutant line G-19 as an example for mutant category 2, and (D) mutant line G-20 as an example of mutant category 1. The vascular tissue in all sections is framed by a red line. Bundle sheath cells are indicated with an asterisk. +, increased number/size; 0, no change in number/size; ?, insufficient sample size to draw a conclusion; BS, bundle sheath V, vasculature. Scale bars = 25 µm.

Although bundle sheath cell number was increased in the mutants of category 2, bundle sheath cell size seemed to be decreased to some extent. The low sample size of independent bundles of the same vein order did not allow statistically relevant measurements with respect to bundle sheath cell size in leaf cross sections. However, it is clear that there was more variation in the size of the bundle sheath cells in mutant category 2, and many bundle sheath cells seemed to be smaller as compared to the bundle sheath cells of the reference line (Figure 5).

Ultrastructural analysis of EMS mutant line G-19

The LM analysis of the 25 mutant lines primarily revealed changes in the bundle sheath cell number and the size of the vascular tissue. However, it failed to deliver insights with regard to cell ultrastructure such as the structure of the chloroplasts. In the reference line with the *GFP* reporter gene, the GFP is transported to the chloroplasts due to the *A. thaliana* RbcS transit peptide that was fused to the *GFP* gene. Consequently, changes in GFP signal in our mutant lines might be associated with chloroplast number, chloroplast size, or chloroplast anatomy. To this end, we viewed cross sections of leaves with TEM of ten mutant lines plus the reference line (G-10, G-13, G-14, G-16, G-18, G-19, G-20, G-22, G-23, and G-25). We only found comparable veins in all replicates in mutant line G-14, G-19, and G-25. A low vein density in most of our samples did not allow to analyze the remaining seven mutant lines, which will be sampled again in future work. While there were no obvious changes detectable in mutant lines G-14 and G-25 with respect to the ultrastructure of the cells, we observed differences in mutant line G-19. This mutant line was selected in the primary screen because of increased reporter gene signal in the leaf (+84%), and the LM analysis of G-19 showed that it also possessed a larger vasculature. By observation with TEM, we identified more sieve tube elements, phloem parenchyma cells, and companion cells in mutant line G-19 as compared to equivalent vein orders in the reference line (Figure 6). Additionally, we detected differences in chloroplast anatomy in all chloroplast-containing cells of the leaf. The plastids had varying degrees of thylakoid development, which is summarized in Figure 6. We observed many chloroplasts that were rather small and thin, and rarely contained any grana. However, there were also well-developed chloroplasts with many grana present in mutant line G-19. This variance in plastid anatomy was not only restricted to the cells of the bundle sheath and vasculature, but it could also be detected in the mesophyll.

With LM and TEM we could show that we got mutants from our EMS screen with altered anatomy of the bundle sheath and vasculature. Furthermore, we obtained one mutant line with

altered chloroplast anatomy. However, this is just a first insight and many more mutant lines still have to be investigated both at the level of LM and TEM.

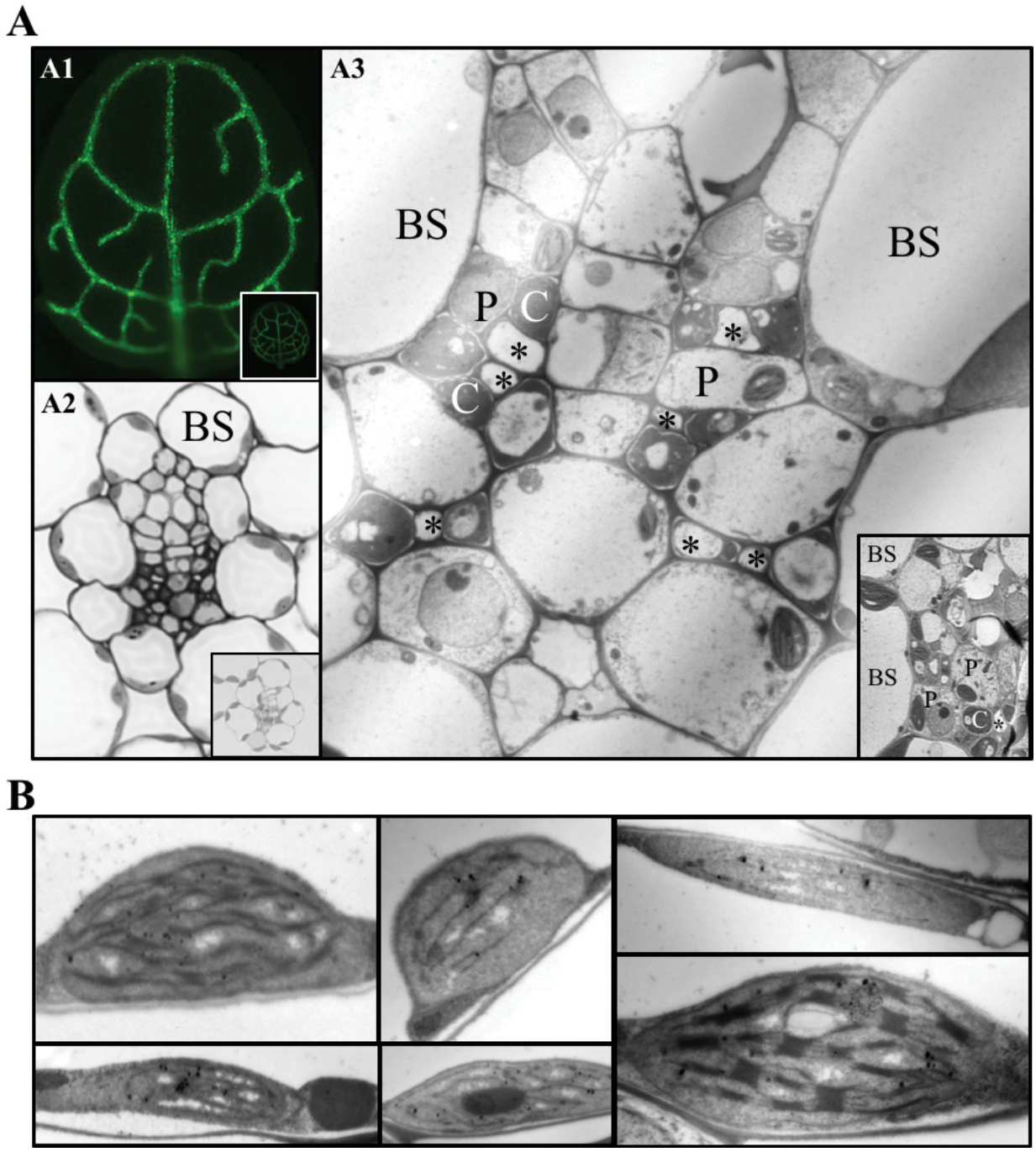


Figure 6. TEM images of mutant line G-19.

(A) Mutant line G-19 compared to the reference line (insets). The images show the GFP signal in the whole leaf (A1), a whole bundle (A2), and TEM images of mostly phloem and xylem tissue (A3). (B) TEM images of six different and representative chloroplasts of mutant line G-19. *, sieve tube element; BS, bundle sheath; C, companion cell; P, phloem parenchyma; inset: reference line. Scale bars in (A) = 5 µm, in (B) = 500 nm.

Mapping of the EMS induced mutations

As a proof of concept, we tried to map the genes for the first nine homozygous EMS-generated mutant lines that we obtained (two with the *LUC* and seven with the *GFP* reporter gene). The mutant lines were crossed with the appropriate non-mutagenized reference line. Since our genetic screen depends on a functional and reliable reporter gene expression in the bundle sheath, we could not perform outcrosses with a different ecotype. Nevertheless, it has been shown before that the use of backcross populations results in enough genetic diversity to identify the causative point mutation (Abe et al., 2012; James et al., 2013). Selfing of the F1 population resulted in an F2 generation in which the recessive mutant phenotype segregated according to Mendelian inheritance. We isolated, pooled, and sheared genomic DNA of about 50–60 F2 plants with the individual mutant phenotype. In this bulked segregate analysis, other EMS-induced point mutations that are close to the causable SNP as well as the causative mutation itself should be in linkage disequilibrium, and therefore, should not recombine. After library preparation, the samples were loaded and sequenced on four flow cell lanes by using the Illumina HighSeq3000 platform along with samples from the original reporter gene lines that served as a background in the further analyses.

A SHOREmap backcross analysis revealed a clear candidate region (Allele frequency > 0.9) in six out of the nine mutants (Supplementary Figures 3A, 3B, and 3C). For each of these six lines between two and ten mutations altering the coding region or splicing sites of protein encoding genes could be identified. The candidate genes will be analyzed further in the future. In one mutant (G-43), the analysis indicated two candidate regions on different chromosomes containing, in total, five candidates for the causative mutation. In one mutant (L-04), no clear candidate region based on the allele frequencies could be identified, and in one mutant (G-42), no candidate for a causative mutation could be identified although a region with high allele frequencies was present. These unexpected results are most likely due to problems in the selection of mutant F2 plants.

Discussion

C_4 plants exhibit a high photosynthetic capacity and an efficient use of nitrogen and water resources, and hence, they outcompete C_3 plants especially in hot and dry environments (Ehleringer et al., 1997). It is of major interest to implement this superior way of photosynthesis into existing C_3 crops, and there are already attempts under the way to integrate the C_4 pathway into rice (http://c4rice.irri.org). To achieve this ambitious goal, several biochemical as well as

anatomical modifications have to be introduced into C_3 plants that allow these plants to concentrate CO_2 around RubisCO. It is assumed that anatomical preconditions such as an enlarged and organelle-rich bundle sheath have to be established before a CO_2-concentrating mechanism can evolve (Sage et al., 2012). However, the genetic basis for these evolutionary changes remains mostly unknown.

EMS mutagenesis as a tool to identify novel genes involved in bundle sheath anatomy

In this study, we described a forward genetic mutant screen in the model plant *A. thaliana* based on the chemical mutagen EMS to identify genes that are involved in the development of bundle sheath cells with respect to bundle sheath cell size, bundle sheath cell number, and chloroplast development within the bundle sheath cells. EMS is one of the most commonly used mutagens for forward genetic screens because of its high efficiency to introduce mutations in the genome. Furthermore, it is reliable and easy to apply in Arabidopsis. Since EMS-induced SNPs are randomly distributed all over the plant genome, the results can be manifold. Compared to T-DNA insertional mutagenesis, which mostly results in complete gene knockouts, EMS mutagenesis can also lead to minor changes in protein activity (Kim et al., 2006; Sikora et al., 2011). Since we do not know whether or not the genes involved in bundle sheath anatomy are indispensable for plant viability, we decided to use an EMS-based genetic screen to also take non-lethal mutations of genes into account.

A reporter gene-labeled bundle sheath—the genetic background of our EMS screen

Since bundle sheath cells are rather small and inconspicuous in Arabidopsis, we designed a reporter gene line that labeled the bundle sheath in the leaf of *A. thaliana*. For this purpose, we used the promoter of the *GLDPA* gene of the C_4 Asteracean species *F. trinervia*. The p$GLDPA_{Ft}$ promoter is highly active in the bundle sheath and vasculature in the C_3 plant Arabidopsis. We used *LUC* and *sGFP* reporter genes to assemble two independent reporter gene lines for the EMS-based genetic screen (Figure 1). Both reporter genes were used in numerous forward genetic screens in *A. thaliana* before in order to label certain tissues, cells, or proteins (Chiu et al., 1996; Chinnusamy et al., 2002; Won et al., 2012; Zwiewka and Friml, 2012). In this study, we could show that the expression of both reporters clearly labeled the bundle sheath and vasculature in the leaves of transgenic plants (Figure 1). Thus, both lines could be used as a genetic background in the EMS-based mutant screens in which we used the reporter gene signal intensity as a proxy for alterations in bundle sheath or chloroplast anatomy. We started with both lines in parallel; however, we early shifted focus to the EMS screen using the *GFP* reporter gene, since it allowed

more spatial resolution and a higher throughput of the primary screen. Consequently, we obtained almost four times more homozygous and stable mutants with the *GFP* background (44) than mutants with the *LUC* background (12) (Figure 2).

Most mutant lines with altered reporter gene expression also showed alterations in the anatomy of the bundle sheath and the vasculature

In the primary screen we first selected for obvious changes in the reporter gene signal and collected plants with either increased, decreased, or diffused reporter gene activity in the leaves (Figure 3; Supplementary Figure 1). However, an effect on the bundle sheath anatomy remained unclear at this point since whole leaves and even hand-cut sections of leaves do not allow to reliably detect single bundle sheath cells. Nevertheless, in seven mutant lines we observed larger bundle sheath strands according to the reporter gene signal. We hypothesized that larger bundle sheath strands might be caused by either larger bundle sheath cells, increased vascular tissue with more surrounding bundle sheath cells, or a combination of both effects (Figure 4). In order to resolve this issue, it was mandatory to analyze the mutant lines in more detail with high-resolution LM. Since sampling and preparation for LM is quite cumbersome, we selected 25 mutant lines (out of 56) that covered the full scope of phenotypes that we obtained in the primary screen but focused on those with larger bundle sheath strands and more reporter gene activity (Supplementary Table 2). We sampled three biological replicates for each line and compared 3° veins among each other with respect to changes in bundle sheath and vascular tissue. We identified six mutants with alterations in the anatomy of the bundle sheath and/or the vasculature that could be classified into two categories: (1) Enlarged vasculature and (2) enlarged vasculature plus more bundle sheath cells (Figure 5). Eight mutant lines did not contain enough comparable veins in all replicates to draw a conclusion; hence, new samples need to be analyzed in future work. The remaining 11 mutant lines did not show any alterations in the anatomy of the bundle sheath or the vasculature (Supplementary Table 2).

Our selection of 25 mutant lines contained seven mutants (G14–20) that were primarily selected because of more GFP signal and enlarged bundle sheath strands in the leaves. It is not surprising that these mutants contained all four mutant lines of category 2 (enlarged vasculature plus more bundle sheath cells) as well as both mutant lines of category 1 (enlarged vasculature). In general, mutants of category 2 possessed the largest bundles among all analyzed mutant lines. Due to the low sample size of independent and comparable veins in the replicates of each mutant line, we could not perform statistically relevant measurements of bundle sheath and vasculature area (and number) but only got a qualitative impression of obvious changes. Nevertheless, we

obtained four mutant lines that seemed to have more bundle sheath cells. It is unclear whether the increase in bundle sheath cell number is a direct effect of the mutation or rather a secondary effect in response to increased vascular tissue. Additionally, bundle sheath cell size seemed to be reduced in these plants. In this respect, it would be worthwhile to analyze single-cell isolates in order to get quantitative data. Furthermore, this would allow quantification of plastid number and provide information on the photosynthetic capacity of the bundle sheath. However, many mutants (including G-19 and G-20 that are shown in Figure 5) showed a strong growth reduction (Supplementary Figure 2). We also observed slightly pale leaves in some mutant lines (G-10, G-16, and G-19), which certainly should negatively affect photosynthetic capacity. To get a closer look at chloroplast morphology, we prepared samples of ten mutant lines for TEM, of which we could analyze three lines so far (G-14, G-19, and G-25). It turned out that mutant line G-19 indeed contained alterations in chloroplast anatomy. We observed varying degrees of thylakoid development in all chloroplast-containing cells. This is in opposition to the cell patterning effect that only applied to the cells of the vasculature in this mutant. Chloroplasts were of varying sizes and shapes in G-19 and a considerable proportion was rather small and nearly grana-less, which should affect photosynthesis in the leaves. It would also be consistent with the slightly pale leaf structure and growth retardation of the mutant (Supplementary Figure 2). However, it remains puzzling how both individual phenotypes can be explained by a mutation of one single gene.

Mapping of EMS-induced mutations in our mutant lines resulted in a manageable number of candidate genes

Forward genetic screens result in many interesting mutant lines; however, it is tedious to map the mutated genes. As a proof of concept, we tried to isolate the genes for the first nine homozygous mutant lines that we obtained in the EMS screens. To map the EMS-induced causable SNPs, we had to create mapping populations. Since the detection of our mutant lines mainly depended on the activity of the reporter gene signal, we could not perform outcrosses with other Arabidopsis ecotypes as it is typically performed (Mokry et al., 2011; Hartwig et al., 2012). Nevertheless, there are reports of using backcross populations and thereby using EMS-induced SNPs as genetic markers to map the genes (Abe et al., 2012; James et al., 2013). We could show that in six out of nine mutant lines in which we have sequenced BCF2 backcross populations, a candidate region could be identified with 2–15 candidate genes each (Supplementary Figures 3A, 3B, and 3C). One mutant line resulted in two candidate regions, and in two lines we did not obtain any candidate genes. However, we demonstrated that using backcross populations to map

EMS-induced SNPs resulted in candidate regions in seven out of nine samples. Obviously, the specific mutated gene with the causative mutation still needs to be identified in each mutant line. To address this issue, one first could try to narrow down the list of candidate genes by predicted gene functions and use T-DNA knockout lines or CRISPR/Cas9 approaches to analyze the candidates. We recently started with the analysis of the genes for the seven mutant lines with clear candidate regions.

To conclude, we could show that our forward genetic EMS screen using Arabidopsis bundle sheath-labeled reporter gene lines allowed the identification of mutants with more bundle sheath cells and mutants with more vascular tissue. Additionally, we identified one mutant line with altered chloroplast anatomy. In how far the information on the genes that still have to be mapped can be used in the context of C_4 engineering remains unclear. Nevertheless, we established a forward genetic screen in *Arabidopsis thaliana*, which successfully identifies mutants with altered bundle sheath cells and chloroplasts.

Acknowledgements

This work was supported by the European Union through the 3to4 project—Converting C_3 to C_4 Photosynthesis for Sustainable Agriculture and the Deutsche Forschungsgemeinschaft through the Cluster of Excellence on Plant Sciences ‘From Complex Traits towards Synthetic Modules’. We thank the ‘Genomics and Transcriptomics laboratory’ of the ‘Biologisch-Medizinisches Forschungszentrum’ (BMFZ) at the Heinrich-Heine-University of Duesseldorf (Germany) for technical support and conducting the Illumina sequencing.

References

Abe A, Kosugi S, Yoshida KK, Natsume S, Takagi H, Kanzaki H, Matsumura H, Yoshida KK, Mitsuoka C, Tamiru M, et al (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol **30**: 174–178

Adrian J, Farrona S, Reimer JJ, Albani MC, Coupland G, Turck F (2010) cis-Regulatory elements and chromatin state coordinately control temporal and spatial expression of FLOWERING LOCUS T in Arabidopsis. Plant Cell **22**: 1425–1440

Akhani H, Khoshravesh R (2013) The relationship and different C4 Kranz anatomy of Bassia eriantha and Bassia eriophora, two often confused Irano-Turanian and Saharo-Sindian species. Phytotaxa **93**: 1–24

Van Bel AJE (1993) Strategies of phloem loading. Annu Rev Plant Physiol Plant Mol Biol **44**: 253–281

Bräutigam A, Gowik U (2016) Photorespiration connects C3 and C4 photosynthesis. J Exp Bot **67**: 2953–2962

Chinnusamy V, Stevenson B, Lee B, Zhu J-K (2002) Screening for gene regulation mutants by bioluminescence imaging. Sci STKE **2002**: pl10
Chiu W, Niwa Y, Zeng W, Hirano T, Kobayashi H, Sheen J (1996) Engineered GFP as a vital reporter in plants. Curr Biol **6**: 325–330
Ehleringer JR, Cerling TE, Helliker BR (1997) C4 photosynthesis, atmospheric CO2, and climate. Oecologia **112**: 285–299
Ehleringer JR, Sage RF, Flanagan LB, Pearcy RW (1991) Climate change and the evolution of C4 photosynthesis. Trends Ecol Evol **6**: 95–9
Engelmann S, Wiludda C, Burscheidt J, Gowik U, Schlue U, Koczor M, Streubel M, Cossu R, Bauwe H, Westhoff P (2008) The gene for the P-subunit of glycine decarboxylase from the C4 species Flaveria trinervia: analysis of transcriptional control in transgenic Flaveria bidentis (C4) and Arabidopsis (C3). Plant Physiol **146**: 1773–1785
Gowik U, Westhoff P (2011) The path from C3 to C4 photosynthesis. Plant Physiol **155**: 56–63
Griffiths H, Weller G, Toy LFM, Dennis RJ (2013) You're so vein: Bundle sheath physiology, phylogeny and evolution in C3 and C4 plants. Plant, Cell Environ **36**: 249–261
Haberlandt G (1904) Physiologische Pflanzenanatomie. Verlag von Wilhelm Engelmann, Leipzig, Germany
Hartwig B, James G V., Konrad K, Schneeberger K, Turck F (2012) Fast isogenic mapping-by-sequencing of ethyl methanesulfonate-induced mutant bulks. Plant Physiol **160**: 591–600
Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber APM, Lercher MJ (2013) Predicting C4 photosynthesis evolution: Modular, individually adaptive steps on a mount fuji fitness landscape. Cell **153**: 1579–88
James G, Patel V, Nordström KJ, Klasen JR, Salomé PA, Weigel D, Schneeberger K (2013) User guide for mapping-by-sequencing in Arabidopsis. Genome Biol **14**: R61
Jander G, Baerson SR, Hudak JA, Gonzalez KA, Gruys KJ, Last RL (2003) Ethylmethanesulfonate saturation mutagenesis in Arabidopsis to determine frequency of herbicide resistance. Plant Physiol **131**: 139–46
Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J **6**: 3901–7
Kim S, Lee D-S, Choi IS, Ahn S-J, Kim Y-H, Bae H-J (2010) Arabidopsis thaliana Rubisco small subunit transit peptide increases the accumulation of Thermotoga maritima endoglucanase Cel5A in chloroplasts of transgenic tobacco plants. Transgenic Res **19**: 489–97
Kim Y, Schumaker KS, Zhu J-K (2006) EMS mutagenesis of Arabidopsis. Methods Mol Biol **323**: 101–103
Kinsman EA a, Pyke KAA (1998) Bundle sheath cells and cell-specific plastid development in Arabidopsis leaves. Development **125**: 1815–1822
Lazo GR, Stein PA, Ludwig RA (1991) A DNA transformation-competent Arabidopsis genomic library in Agrobacterium. Biotechnol (Nature Publ Company) **9**: 963–7
Leegood RC (2008) Roles of the bundle sheath cells in leaves of C3 plants. J. Exp. Bot. pp 1663–1673
Lundgren MR, Osborne CP, Christin P-A (2014) Deconstructing Kranz anatomy to understand C4 evolution. J Exp Bot **65**: 3357–69
Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, Gowik U (2014) The role of photorespiration during the evolution of C4 photosynthesis in the genus Flaveria. Elife **2014**: e02478
Mokry M, Nijman IJ, van Dijken A, Benjamins R, Heidstra R, Scheres B, Cuppen E (2011) Identification of factors required for meristem function in Arabidopsis using a novel next generation sequencing fast forward genetics approach. BMC Genomics **12**: 256

Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, Niwa Y, Toyooka K, Matsuoka K, Jinbo T, Kimura T (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. J Biosci Bioeng **104**: 34–41

Page DR, Grossniklaus U (2002) The art and design of genetic screens: Arabidopsis Thaliana. Nat Publ Gr **3**: 124–136

Pollock DD, Larkin JC (2004) Estimating the degree of saturation in mutant screens. Genetics **168**: 489–502

Roth-Nebelsick A, Uhl D, Mosbrugger V, Kerp H (2001) Evolution and function of leaf venation architecture: A review. Ann Bot **87**: 553–566

Sage RF, Christin PA, Edwards EJ (2011) The C4 plant lineages of planet earth. J Exp Bot **62**: 3155–3169

Sage RF, Khoshravesh R, Sage TL (2014) From proto-Kranz to C4 Kranz: building the bridge to C4 photosynthesis. J Exp Bot **65**: 3341–56

Sage RF, Sage TL, Kocacinar F (2012) Photorespiration and the evolution of C4 photosynthesis. Annu Rev Plant Biol **63**: 19–47

Sage TL, Busch FA, Johnson DC, Friesen PC, Stinson CR, Stata M, Sultmanis S, Rahman BA, Rawsthorne S, Sage RF (2013) Initial events during the evolution of C4 photosynthesis in C3 species of Flaveria. Plant Physiol **163**: 1266–76

Sattarzadeh A, Fuller J, Moguel S, Wostrikoff K, Sato S, Covshoff S, Clemente T, Hanson M, Stern DB (2010) Transgenic maize lines with cell-type specific expression of fluorescent proteins in plastids. Plant Biotechnol J **8**: 112–25

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods **9**: 671–675

Schuler ML, Mantegazza O, Weber APM (2016) Engineering C4 photosynthesis into C3 chassis in the synthetic biology age. Plant J **87**: 51–65

Sheehy JE, Mitchell PL (Peter L., Hardy B (2007) Charting new pathways to C_4 rice. International Rice Research Institute

Sikora P, Chawade A, Larsson M, Olsson J, Olsson O, Sikora P, Chawade A, Larsson M, Olsson J, Olsson O (2011) Mutagenesis as a tool in plant genetics, functional genomics, and breeding. Int J Plant Genomics **2011**: 314829

Somerville C, Koornneef M (2002) A fortunate choice: the history of Arabidopsis as a model plant. Nat Rev Genet **3**: 883–889

Sun H, Schneeberger K (2015) SHOREmap v3.0: Fast and accurate identification of causal mutations from forward genetic screens. *In* JM Alonso, AN Stepanova, eds, Plant Funct. Genomics Methods Protoc. Springer New York, New York, NY, pp 381–395

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

Won SY, Li S, Zheng B, Zhao Y, Li D, Zhao X, Yi H, Gao L, Dinh TT, Chen X (2012) Development of a luciferase-based reporter of transcriptional gene silencing that enables bidirectional mutant screening in Arabidopsis thaliana. Silence **3**: 6

Zhang X, Henriques R, Lin S-S, Niu Q-W, Chua N-H (2006) Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. Nat Protoc **1**: 641–646

Zwiewka M, Friml J (2012) Fluorescence imaging-based forward genetic screens to identify trafficking regulators in plants. Front Plant Sci **3**: 97

Supplementary Data

Supplementary Table 1. Oligonucleotides used in this study.

Supplementary Table 2. Characterization of the 25 selected mutant lines after LM.

Supplementary Figure 1. Relative reporter gene signal intensity of all 56 EMS mutant lines.

Supplementary Figure 2. Growth of the ten mutants that were analyzed by TEM.

Supplementary Figure 3A. Allelic frequencies for mutant lines L-01, L-02, and L-03.

Supplementary Figure 3B. Allelic frequencies for mutant lines L-04, G-21, and G-32.

Supplementary Figure 3C. Allelic frequencies for mutant lines G-35, G-42, and G-43.

Supplementary Table 1. Oligonucleotides used in this study.

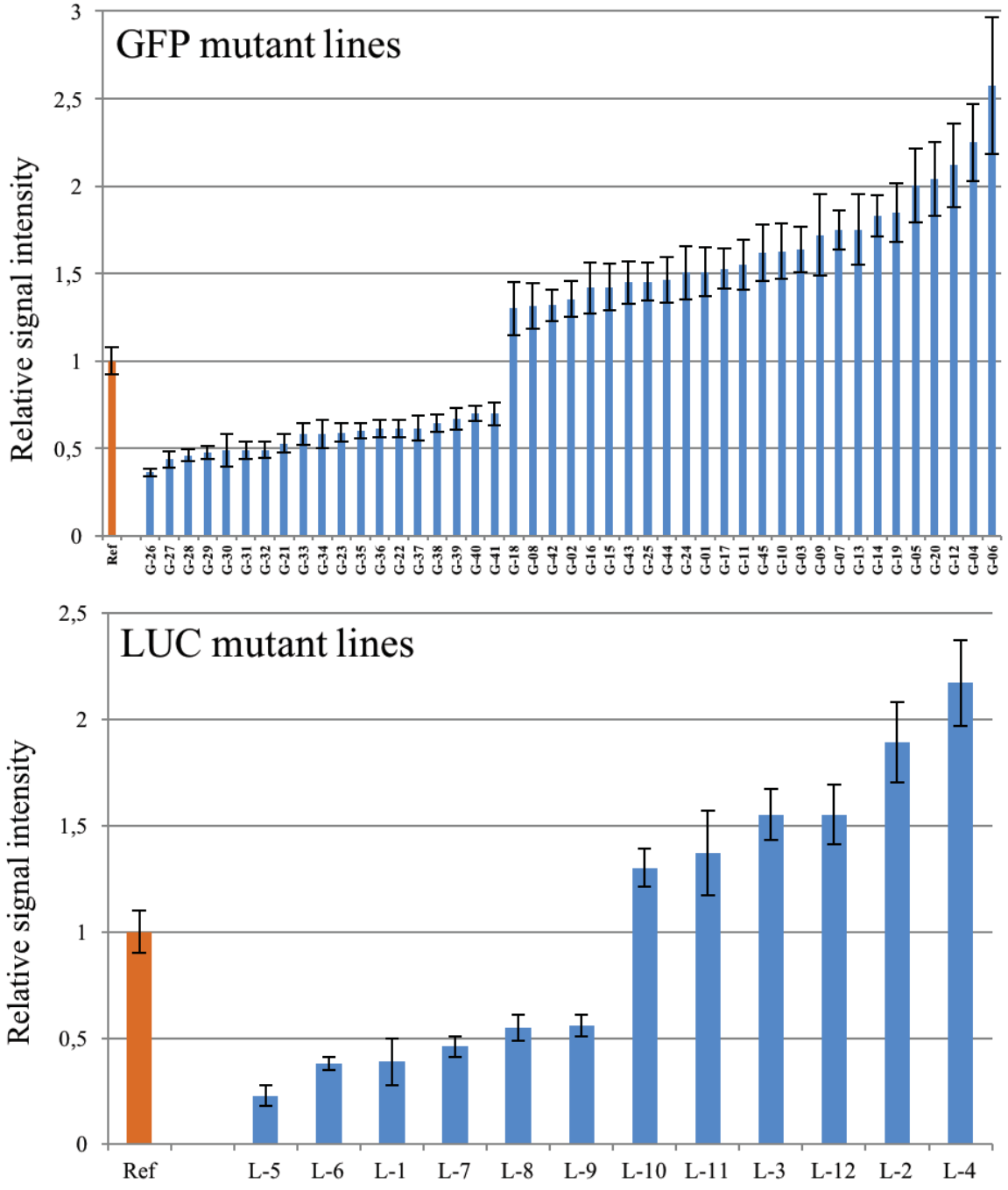
F, forward primer; R, reverse primer

Primer	Sequence (5'→ 3')	Orientation
p*GLDPA*-Ft-F	TACTCCTCTCAACTTTCAA	F
p*GLDPA*-Ft-R	AGTGTAAGATGGGGTCTAA	R
p*GLDPA*-Ft+*Att*B1	GGGGACAAGTTTGTACAAAAAAGCAGGCT	F
TACTCCTCTCAACTTTCAA		
p*GLDPA*-Ft+*Att*B2	GGGGACCACTTTGTACAAGAAAGCTGGGT	R
AGTGTAAGATGGGGTCTAA		
p*GLDPA*-Ft+*Hind*III	ATAAGCTTTACTCCTCTCAACTTTCAA	F
p*GLDPA*-Ft+*BamH*I	ATGGATCCGTGTAAGATGGGGTCTAA	R
RbcS.TP+*BamH*I	AAGGATCCATGGCTTCCTCTATGCTC	F
RbcS.TP+*EcoR*I	AAGAATTCTTCGGAATCGGTAAGGTC	R
sGFP+*EcoR*I	ATGAATTCATGGTGAGCAAGGGCGAG	F
sGFP+*Sac*I	ATGAGCTCTTACTTGTACAGCTCGTC	R
p*GLDT*-Ft+*Pvu*I	CGATCGTCGACCCGTAAATAGGTCAA	F
p*GLDT*-Ft+*Asc*I	GGCGCGCCGTGTGCTTTATTCTTTAGAAAC	R

Supplementary Table 2. Characterization of the 25 selected mutant lines after LM.

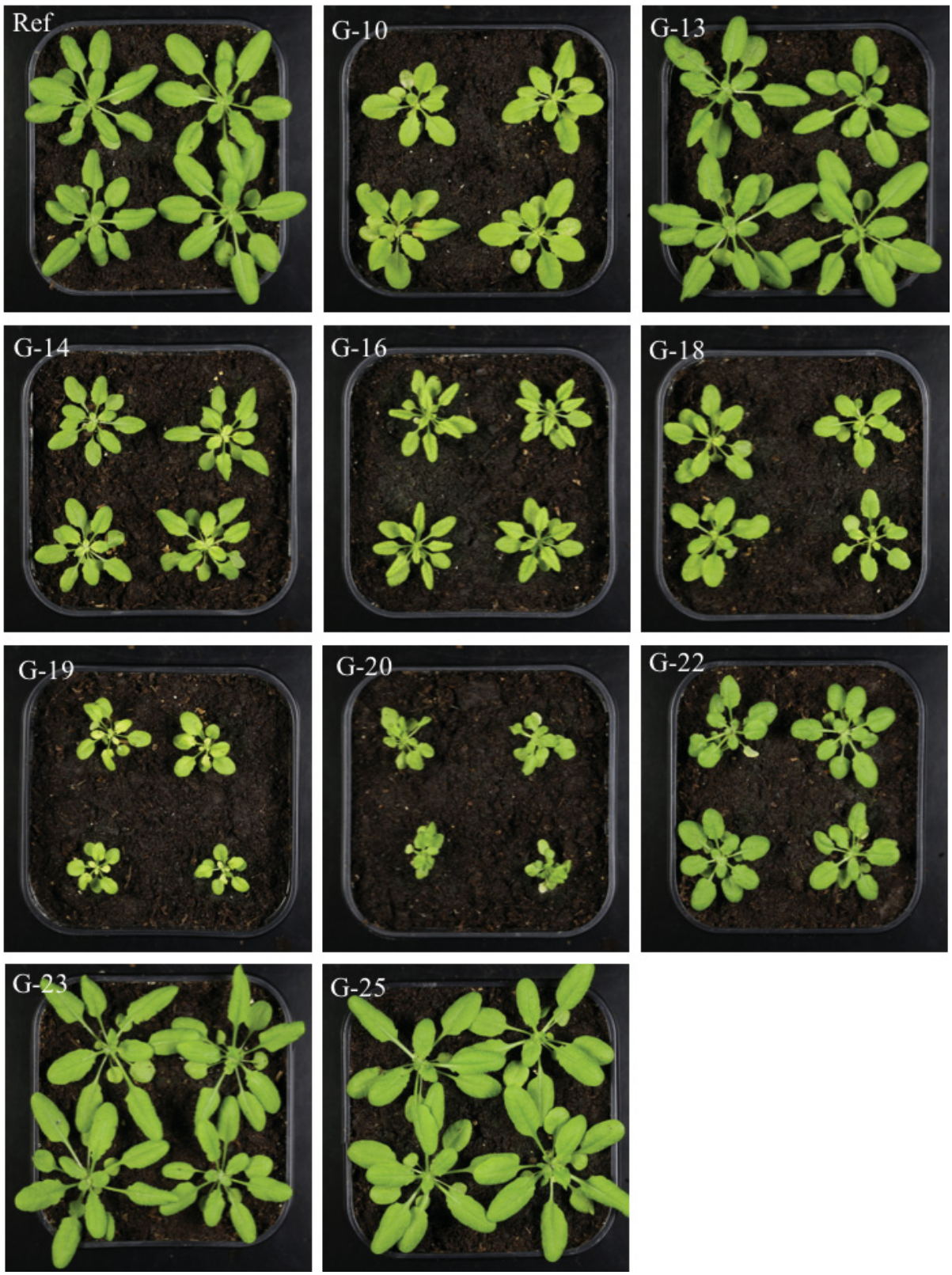
Overview of all the 25 mutant lines that were analyzed with high-resolution LM and compared to the non-mutagenized reference line. Each mutant line was checked for GFP signal in the leaves, the number of bundle sheath cells, and the overall size of the vascular tissue. +, more signal/number/size; ++, more signal and enlarged bundles; 0, no change in signal/number/size; -, less signal/number/size; ?, insufficient sample size to draw a conclusion; diffuse, reporter gene signal also detectable in the mesophyll; BSC, bundle sheath cell; V, vascular tissue.

Mutant Nr.	GFP signal	BSC Number	V Total size
1	+	?	?
2	+	0	0
3	+	0	0
4	+	0	0
5	+	0	0
6	+	?	?
7	+	0	0
8	+	?	?
9	+	0	0
10	+	?	?
11	+	0	0
12	+	?	?
13	+	?	?
14	++	+	+
15	++	+	0
16	++	?	?
17	++	+	+
18	++	+	+
19	++	+	0
20	++	+	+
21	-	0	0
22	-	0	0
23	-	0	0
24	diffuse	?	?
25	diffuse	0	0



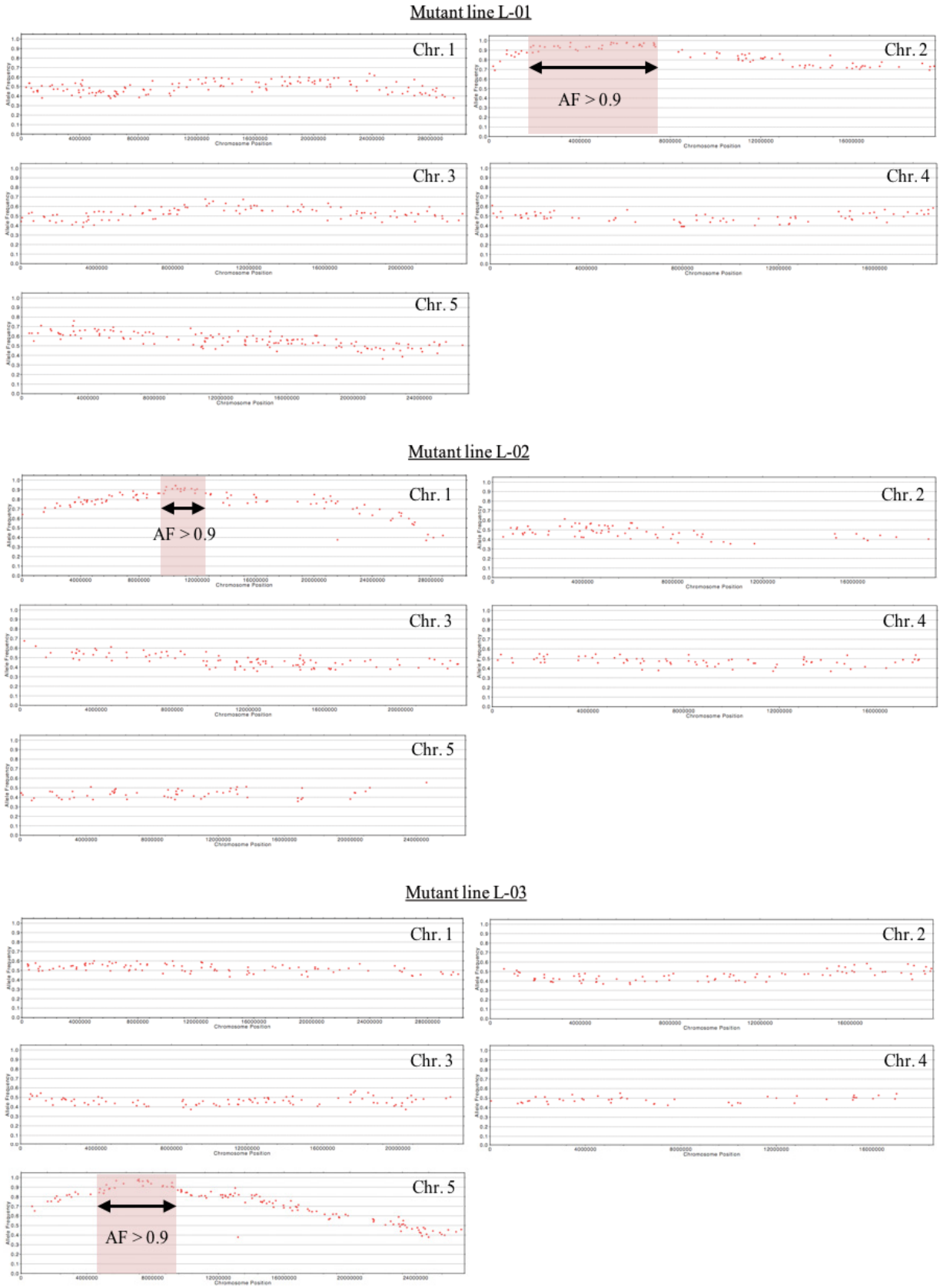
Supplementary Figure 1. Relative reporter gene signal intensity of all 56 EMS mutant lines.

The relative signal intensity of all 56 mutant lines (44 *GFP*, 12 *LUC*) was compared to the appropriate reference line. The reporter gene signal was measured in whole leaves of 14–17-day-old plants and normalized to the leaf area. At least 50 plants per mutant line were analyzed.



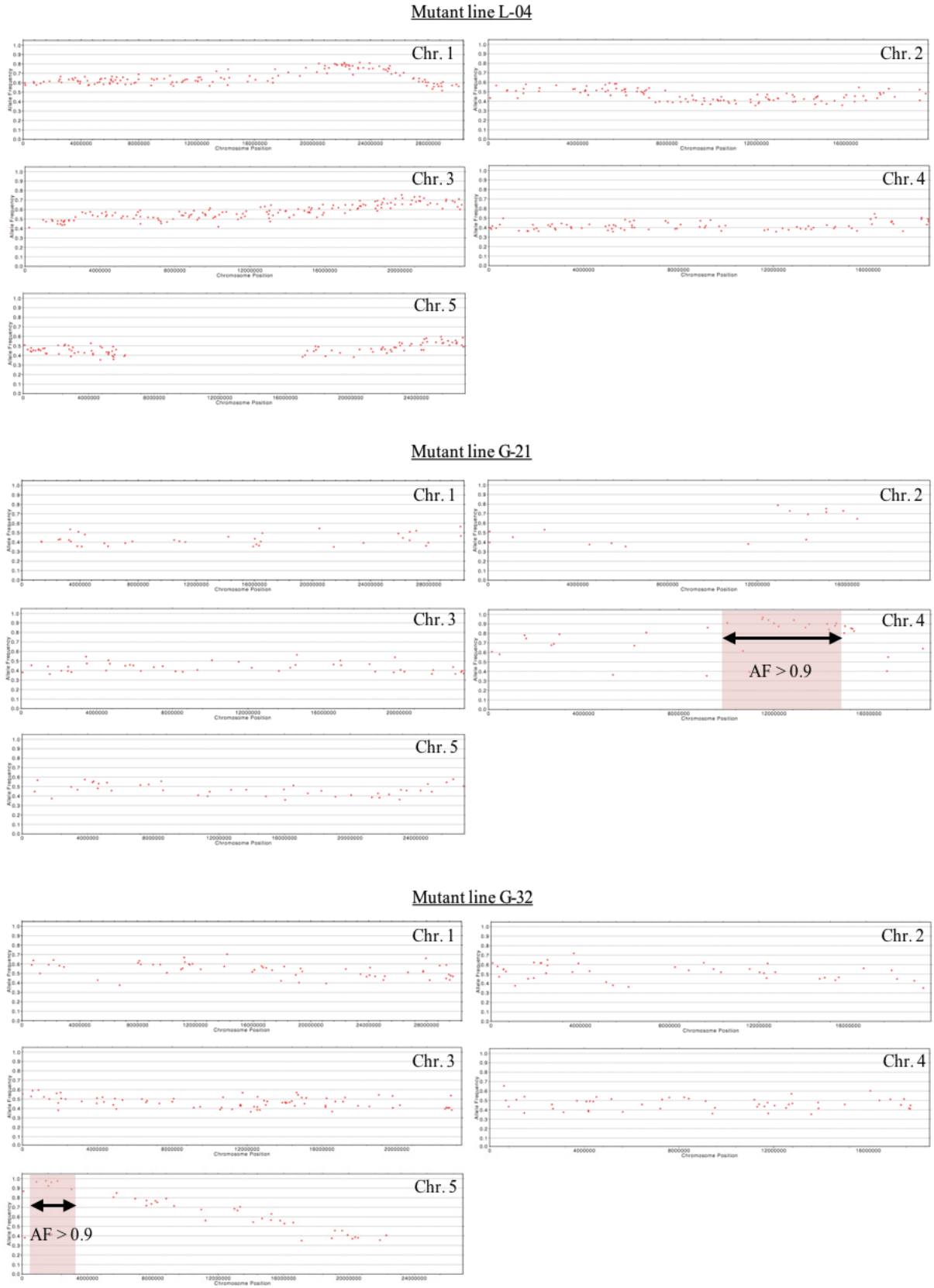
Supplementary Figure 2. Growth of the ten mutants that were analyzed by TEM.

Overview of the ten mutant lines that were analyzed by TEM plus the reference line (Ref). All the plants were 28 days old.



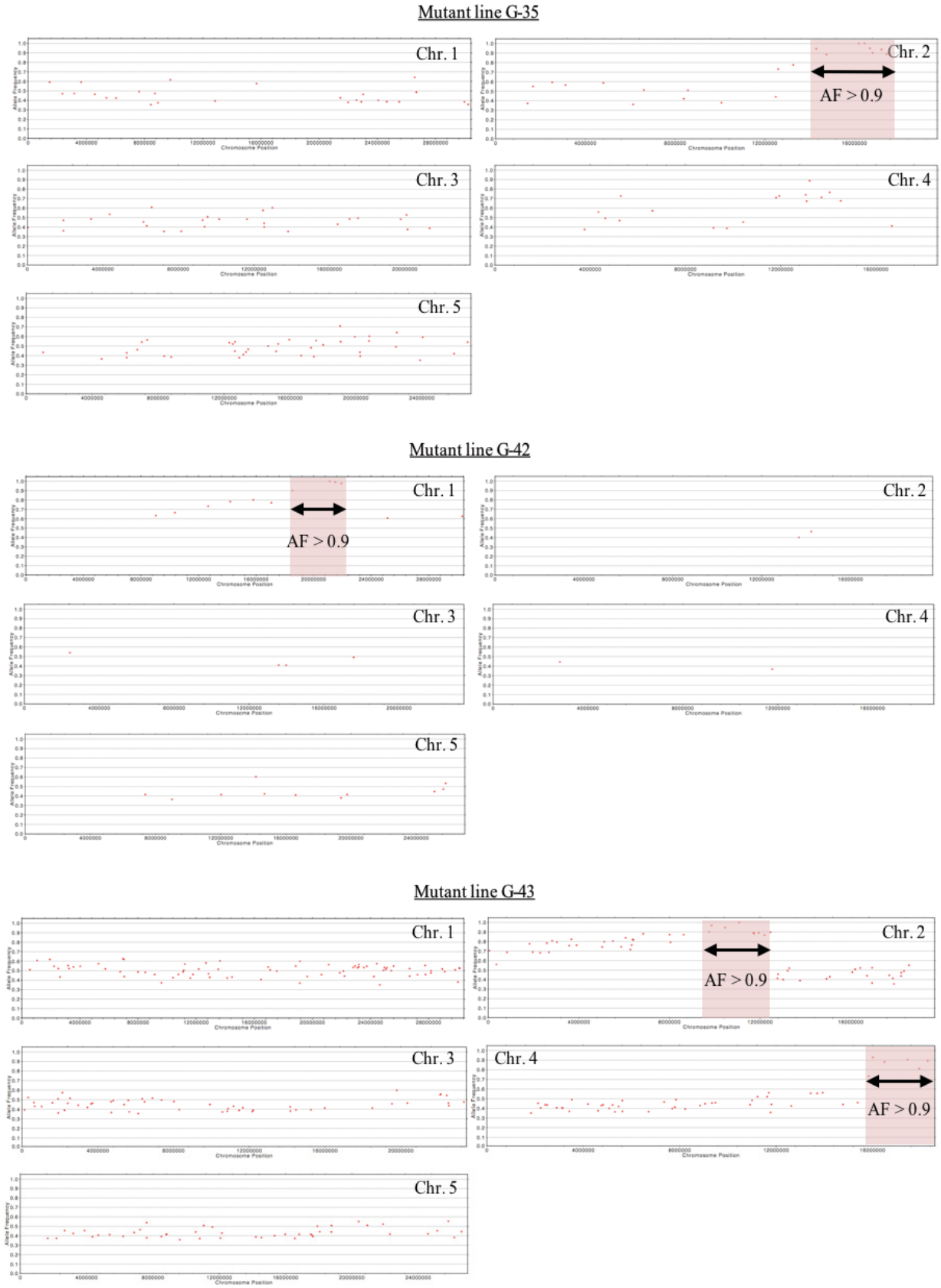
Supplementary Figure 3A. Allelic frequencies for mutant lines L-01, L-02, and L-03.

Allelic frequencies (AF) for all SNPs resolved using the *LUC* or *GFP* reporter gene line parent and BCF2 mutant whole genome sequence data. Genes containing a non-synonymous SNP with AF>0.9 were considered as candidate genes, and genomic regions with AF>0.9 are highlighted in the diagrams.



Supplementary Figure 3B. Allelic frequencies for mutant lines L-04, G-21, and G-32.

Allelic frequencies (AF) for all SNPs resolved using the *LUC* or *GFP* reporter gene line parent and BCF2 mutant whole genome sequence data. Genes containing a non-synonymous SNP with AF>0.9 were considered as candidate genes, and genomic regions with AF>0.9 are highlighted in the diagrams.



Supplementary Figure 3C. Allelic frequencies for mutant lines G-35, G-42, and G-43.

Allelic frequencies (AF) for all SNPs resolved using the LUC or *GFP* reporter gene line parent and BCF2 mutant whole genome sequence data. Genes containing a non-synonymous SNP with AF>0.9 were considered as candidate genes, and genomic regions with AF>0.9 are highlighted in the diagrams.

The authors' contributions

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

KB provided 20 EMS mutant lines and did all the light microscopic and transmission electron microscopic analysis of the selected mutant lines.

KB, SG and **SS** helped with the fixation and embedding of leaf samples for microscopy.

UG analyzed the sequencing data.

SG provided nine EMS mutant lines.

SS, RK and **TS** helped with the operation of the microscopes and the analysis of the sections.

PW participates in drafting of the manuscript.

Manuscript 3

Altered expression of *BSOM2*, a gene coding for an ABA efflux transporter, results in increased bundle sheath and vascular tissue in the leaf of *Arabidopsis thaliana*

Introduction

Kranz anatomy is a key feature of most C_4 plants. It describes the concentric formation of two different tissues: the bundle sheath and the mesophyll. One layer of compactly arranged bundle sheath cells surrounds the vasculature in a wreath-like mode and is itself enclosed by mesophyll cells that are located toward the outer face of the leaf. This special arrangement of both cell types is the basis for a functional CO_2-concentrating mechanism (CCM) in C_4 plants. CCMs lead to increased CO_2 concentrations around the RubisCO enzyme, which is exclusively located in the bundle sheath of plants operating a dual-celled C_4 photosynthesis. Since the primary fixation of atmospheric carbon is spatially separated from the secondary fixation of CO_2 by RubisCO, photorespiration is efficiently repressed in C_4 plants (Von Caemmerer and Furbank, 2003; Furbank, 2011).

Most reactions of C_4 photosynthesis take place in the bundle sheath, while the mesophyll mostly functions in primary fixation of atmospheric CO_2 by the action of phospho*enol*pyruvate carboxylase (PEPC). Nevertheless, the cumulative volume of the bundle sheath relative to the mesophyll always seems to be within a certain range in C_4 plants (Dengler et al., 1994; Muhaidat et al., 2007). In an optimal C_4 plant, almost all the bundle sheath cells are in contact with a mesophyll cell and they are intimately connected to each other due to multiple plasmodesmata (Dengler and Nelson, 1999). This illustrates how closely these two cell types are interlinked and mutually dependent in C_4 plants.

In C_3 plants, both cell types are photosynthetically active; however, bundle sheath cells are small and contain only a few chloroplasts. Since CO_2 refixation rates directly depend on the number of chloroplasts and hence the amount of RubisCO within the cell type designated for this function, bundle sheath cells tend to be large in C_4, which allows for more accommodation of chloroplasts (Black and Mollenhauer, 1971; Muhaidat et al., 2011; Lundgren et al., 2014). Additionally, a larger bundle sheath might also provide better protection against cavitation in hot and arid environments (Sage, 2001b; Griffiths et al., 2013; Lundgren et al., 2014). It has been shown that the number of chloroplasts in the mesophyll is reduced during C_4 evolution resulting in lower chloroplast coverage in mesophyll cells (Stata et al., 2014). With respect to C_4 engineering, it is therefore of great importance to have control over mesophyll and bundle sheath chloroplast characteristics in a cell-specific manner. However, the genetic situation has remained mainly unclear so far. Studies in maize showed that the GOLDEN2-LIKE (GLK) transcription factors GLK1 and GLK2 play an important role in bundle sheath chloroplast development. While in *Arabidopsis thaliana*, the GLK proteins mostly act redundantly and show no cell-specificity, it became apparent that in maize, GOLDEN2 specifically affects chloroplast development in the

bundle sheath. Accordingly, GOLDEN2 is only expressed in bundle sheath cells in maize, while its homologue ZmGLK1 is more abundant in mesophyll cells (Rossini et al., 2001; Waters et al., 2008; Waters and Langdale, 2009). Therefore, one can conclude that the GLK proteins seem to have essential functions in the gene regulation of bundle sheath/mesophyll differentiation.

Nevertheless, more information is needed about genes that are associated with the development of bundle sheath cells and bundle sheath chloroplasts, genes that could be manipulated in order to introduce C_4 anatomy into C_3 plants. Many properties that define C_4 anatomy are already present in at least some C_3 plants, but all the characteristics need to be assembled at one place to provide a framework for the introduction of the C_4 cycle. In addition, no C_3 plants have been known so far to supply a high concentration of chloroplasts in bundle sheath cells, something that seems to be specific to C_4 plants and partly also to C_2 plants with C_3-C_4 intermediate photosynthesis (Lundgren et al., 2014).

In this study, we used a forward genetic activation tagging approach to identify novel genes for bundle sheath ontogeny and maintenance (*BSOM* genes) in *A. thaliana*. Studies with the p*GLDPA*$_{Ft}$ promoter (glycine decarboxylase P gene of *Flaveria trinervia*), a promoter that is highly active in the bundle sheath and vasculature in *F. trinervia*, revealed that the expression specificity was maintained in the Brassicacean C_3 species *A. thaliana* (Engelmann et al., 2008). Assuming that gene-regulatory systems of the bundle sheath have been at least partly conserved during evolution, it indicates that some kind of cryptic Kranz anatomy is already present in C_3 species, and that the bundle sheath is an evolutionary ancient invention in angiosperms (Westhoff and Gowik, 2010). For this reason, we think that *BSOM* genes discovered in Arabidopsis by forward genetics might be useful to activate the bundle sheath in other C_3 plants. Since bundle sheath cells are only detectable under microscopic examination, we used an Arabidopsis reporter line with labeled bundle sheath cells that was previously described by Döring et al. (2017, unpublished; Manuscript 2).

Material and Methods

Plant material

The activation tagging screen was performed with *Arabidopsis thaliana* (Ecotype Columbia). Since bundle sheath cells are small in Arabidopsis, and it is a tedious task to identify them microscopically in cross sections, we used a reporter gene line that expresses a chloroplast targeted variant of the green fluorescent protein (GFP) in the bundle sheath and vasculature of *A. thaliana* leaves. This reporter gene line—here referred to as reference line—has a stable T-DNA insertion of p*GLDPA*$_{Ft}$::RbcS.TP-sGFP. It was previously described by Döring et al. (2017, unpublished; Manuscript 2) and is used here as a genetic background for the activation tagging screen. The seeds were surface-sterilized with a chloric solution (20% Dan Klorix [Colgate-Palmolive, Hamburg, Germany], 0.02% Triton X-100) and incubated at 4 °C for at least 48 h. Subsequently, they were sown on soil in large flats (Floraton 1, Floragard, Oldenburg, Germany) or in petri dishes with ½ Murashige and Skoog (MS)-medium containing 0.6% agar and 1% sucrose. All plants were grown either in our greenhouse with supplementary light for 14 h per day at a photon flux density (PFD) of ~300 µmol m^{-2} s^{-1} or in climate chambers with 16 h light/8 h of darkness periods (~60 µmol m^{-2} s^{-1}) and a constant temperature of 21–22 °C. Mutant seeds of SALK line SALK_144096 were obtained from the Arabidopsis Biological Resource Center (ABRC).

Construction of the activation tagging construct

The 3.2 kb long region upstream of the *GLDT* gene of *Flaveria trinervia* was amplified by PCR with primers that add *Sac*I and *Pme*I restriction sites to the distal and the proximal parts of the promoter, respectively. This fragment was cut with both enzymes and cloned into the Gateway vector pMDC123 (Curtis and Grossniklaus, 2003), previously digested with *Sac*I and *Pme*I. Consequently, the complete Gateway cassette of pMDC123 was removed and replaced by the promoter region with the proximal part of the promoter adjacent to the LB of the T-DNA. The resulting vector pMDC123-p*GLDT*$_{Ft}$(Ac.T.) was transformed into *Agrobacterium tumefaciens* strain AGL1 by electroporation (Lazo et al., 1991). Transgenic Arabidopsis plants were generated according to Zhang et al. (2006) by the floral dip method. M1 plants were directly sown on soil in large flats in the greenhouse.

Activation tagging screen and quantification of GFP signal

Approximately 10,000–20,000 seeds of the M1 generation were sown on each flat (45 cm x 25 cm). At three different time points (1, 5, and 10 days after sowing), the flats were watered with a PPT solution containing 80 mg/l glufosinate-ammonium (Bayer Agrar, Germany) and 0.1% Tween 20. Only transformants with an intact *BAR* resistance gene that provides resistance to the herbicide PPT survived. Each transformant (M1 generation) was individually screened for increased or decreased GFP signal in the leaves. Therefore, the first leaf pair of 14–17-day-old plants was analyzed with a fluorescence microscope (Axio Imager M2m, Zeiss, Oberkochen, Germany). The total GFP signal per leaf was quantified and normalized to leaf area with ImageJ (Collins, 2007) and only mutant lines with at least 30% stronger or weaker signals compared to the non-mutagenized reference line were selected and further analyzed in the M2 generation. Here, we again checked for the aberrant phenotype and its segregation. Stable and non-segregating mutant lines were selected in M3 and M4.

Localization of the T-DNA insertion

The flanking regions of the T-DNA containing the activation tag in our *BSOM2* mutant line were cloned with Thermal Asymmetric Interlaced PCR (TAIL-PCR). For this purpose, genomic DNA was extracted according to Edwards et al. (1991) and used as a template in the PCR reactions. The PCR reactions were performed as described by Liu and Whittier (1995). We used three different nested primers for both LB and RB and a set of nine independent arbitrary degenerated primers (AD1–9). The PCR products were sequenced and the results were blasted against the *A. thaliana* genome to identify the T-DNA insertion. The results were confirmed with appropriate primers matching to the T-DNA LB or RB and the adjacent genomic sequences.

Reconstruction of the *BSOM2* mutant phenotype by overexpression of *BSOM2* with p*GLDT*$_{Ft}$ and p(2x*35S*)

The pAUL1 Gateway vector (Lyska et al., 2013) was used to create the p*GLDT*$_{Ft}$ overexpression construct. The p*GLDT*$_{Ft}$ sequence was amplified from genomic DNA of *F. trinervia* by PCR, and *Pvu*I and *Asc*I restriction sites were added to the distal and proximal parts of the promoter, respectively. The PCR product as well as pAUL1 were digested with *Pvu*I + *Asc*I, and subsequently ligated to result in pAUL1-p*GLDT*$_{Ft}$. The coding region of *BSOM2* (AT5G52050) was amplified by PCR with specific primers that added Gateway compatible *att*B sites to the product. The *att*B-flanked PCR products were recombined into pDONR^{TM}221 (Invitrogen) and

subsequently introduced into pAUL1-p*GLDT*$_{Ft}$, which resulted in pAUL1-p*GLDT*$_{Ft}$::*BSOM2*. As a feature of pAUL1, the coding region of *BSOM2* was fused with its N-terminal side to a 3xHA tag. For the assembly of the p(2x*35S*)::*BSOM2* construct, the *att*B-flanked PCR product of *BSOM2* was used again, and recombined with the Gateway vector pMDC32 (Curtis and Grossniklaus, 2003), resulting in pMDC32-p(2x*35S*)::*BSOM2*. All the constructs were verified by DNA sequencing.

Expression analysis of *BSOM2* with quantitative real-time PCR (qPCR)

The total RNA was isolated with the RNeasy Mini Kit (Qiagen, Hilden, Germany) from leaves of 21-day-old plants of the reference line and homozygote T3 lines containing either p*GLDT*$_{Ft}$::*BSOM2* or p(2x*35S*)::*BSOM2*. The RNA was transcribed into cDNA with the QuantiTect Reverse Transcription Kit (Qiagen, Hilden, Germany). The purity and integrity of the total RNA and the cDNA were verified spectroscopically with a NanoDrop ND-1000 and by agarose gel electrophoresis. The quantitative real-time PCR (qRT-PCR) followed standard procedures and was performed with an ABI7500 fast Real Time PCR system. The primers were designed to target the CDS of the genes, and to create amplicons of 110–120 bp length. Agarose gel electrophoresis and melting curve analysis served to verify the specificity of the PCRs.

Microscopic imaging

All the steps for sampling and preparation for light microscopy (LM) were performed as described by Akhani and Khoshravesh (2013). We sampled the fully expanded 2nd leaf pair of four-week-old *A. thaliana* plants. For this purpose, we cut off the edges of the leaf and fragmented it into 1–2 mm^2 small sections, which were immediately fixed with a fixative (1% glutaraldehyde, 1% paraformaldehyde, 0.1 M sodium cacodylate). All the samples were post-fixed with OsO_4 followed by a series of ethanol dehydrations and Araldite resin infiltration and embedding. The embedded tissue blocks were trimmed with a trimming microtome, and sections with a thickness of 1.5–2 micron were obtained by the use of a microtome (Microm HM 330 Microtome). All the sections were subsequently stained with toluidine blue for light microscopy.

Results

The activation tagging screen resulted in eight stable mutant lines with altered GFP signal in the bundles

The reference line, which was previously described by Döring et al. (2017, unpublished; Manuscript 2), was used as a genetic background in the activation tagging screen and it allowed for an efficient identification of the bundle sheath in Arabidopsis leaves. A chloroplast transit peptide attached to the *sGFP* gene made it possible to detect chloroplasts within the cells of the bundle sheath and vasculature (Figure 1).

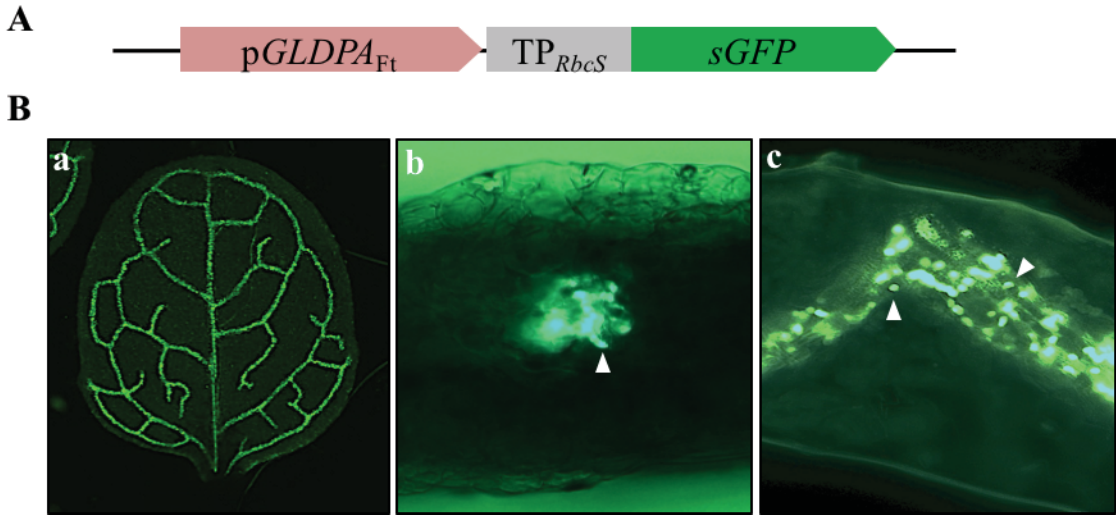


Figure 1. The Arabidopsis reporter gene line with GFP-labeled bundle sheath cells.

(A) The reporter gene construct in the reference line. A chloroplast-targeted sGFP is under the control of the p*GLDPA*$_{Ft}$ promoter region of the C_4 dicot *F. trinervia*. (B) Fluorescence imaging in leaves of the reference line, showing the whole leaf (a) and cross sections (b and c). Arrowheads indicate single chloroplasts. Source: Adapted from Döring et al. (2017, unpublished; Manuscript 2).

About 2500–3000 plants of the reference line were transformed with the activation tag. It contained the p*GLDT*$_{Ft}$ promoter of the Asteraceaen C_4 species *F. trinervia*, which, like the p*GLDP*A$_{Ft}$ promoter, only shows expression in the bundle sheath and vasculature of the C_3 Brassicacean species *A. thaliana*. To easily screen for transformants, the T-DNA also contained a *BAR* gene that provides resistance to the herbicide Phosphinothricin (PPT) (Figure 2A). The complete activation tag was introduced into the plant genome of the reference line by *Agrobacterium*-mediated transformation. The M1 seeds of these plants were grown on soil in large flats, which were inoculated with PPT solution. Each transformant was screened individually for differences in reporter gene activity in the first leaf pair. The intensity of the GFP fluorescence signal was quantified with ImageJ and normalized to the total leaf area for the

mutants with an obvious disparity in reporter gene signal. The mutants with less than 30% difference in reporter gene activity relative to the reference line were discarded. This resulted in 175 primary mutants that were analyzed once again in the following M2 and M3 generations to check the stability and segregation of the aberrant phenotypes. It turned out that the majority of the collected M1 mutant lines (167/173) did not show the specific change in reporter gene activity in the following generations. Nevertheless, the activation tagging screen resulted in eight mutant lines, in which the aberrant phenotypes could be confirmed in the M2 and M3 generations, and in which there was at least 30% difference in the reporter gene signal (Figure 2B + 3).

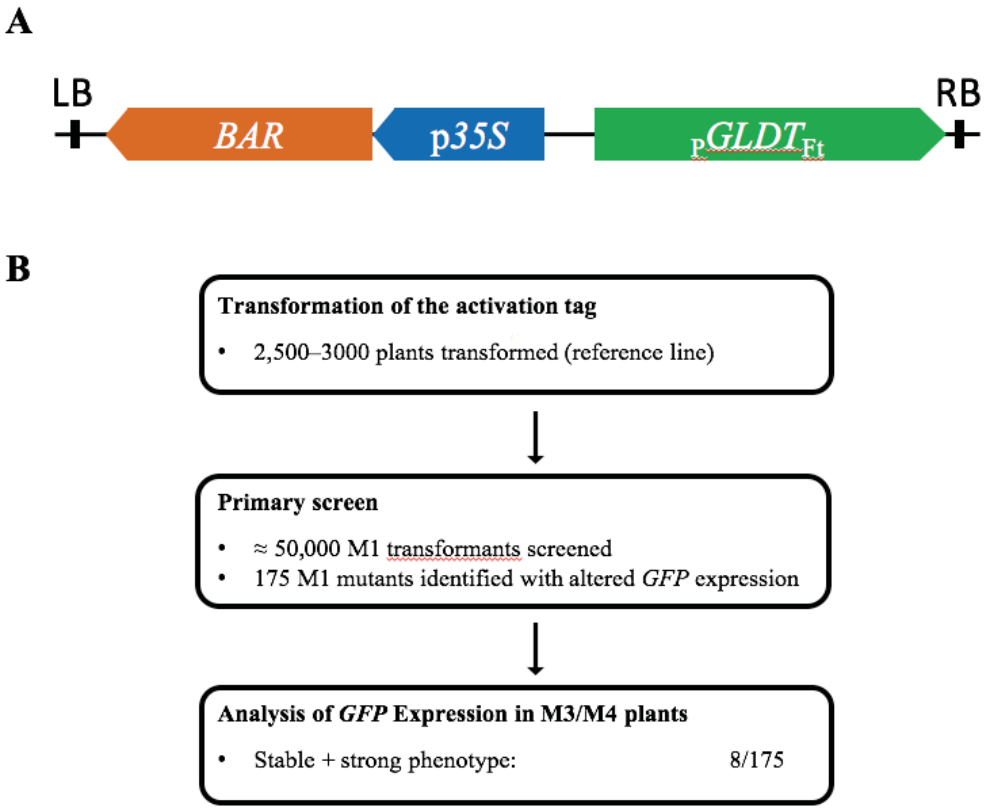


Figure 2. Workflow of the activation tagging screen.

(A) The activation tag containing the p*GLDT*$_{Ft}$ promoter and a *BAR* resistance gene to the herbicide PPT. (B) Workflow of the activation tagging screen.

We checked the segregation of the T-DNA in all the mutant lines. It is known that *Agrobacterium*-mediated transformation often leads to multiple insertions of the T-DNA in the genome, which makes the mapping of the T-DNA with PCR-based methods more difficult (Castle et al., 1993). Ideally, if there is only one T-DNA insertion per M1 line that causes the phenotype, we should observe a Mendelian segregation ratio of 3:1 in the M2 generation resulting in ¼ plants without PPT resistance. We observed this in five mutant lines (AT-1, AT-3, AT-4, AT-7, and AT-8), whereas two M1 mutant lines (AT-5 and AT-6) segregated at least 9:1 in M2, thus possibly having multiple insertions of the T-DNA in the genome. Mutant line AT-2 did not

produce enough seeds to calculate the segregation of the T-DNA insertion. Seven mutant lines possessed a stronger GFP signal in the bundles (varying from +32% to +100%), while only one had less signal (-33%) in this compartment (Figure 3). Additionally, we also measured the width of whole bundle sheath strands. For this purpose, we used the first leaf pair of 17-day-old plants and analyzed the 3° veins at comparable positions. The distance from one end of the bundle sheath strand to the other end was measured with the software Zen 2012–blue edition (Zeiss, Oberkochen, Germany) in the leaves of 50 plants of each mutant line.

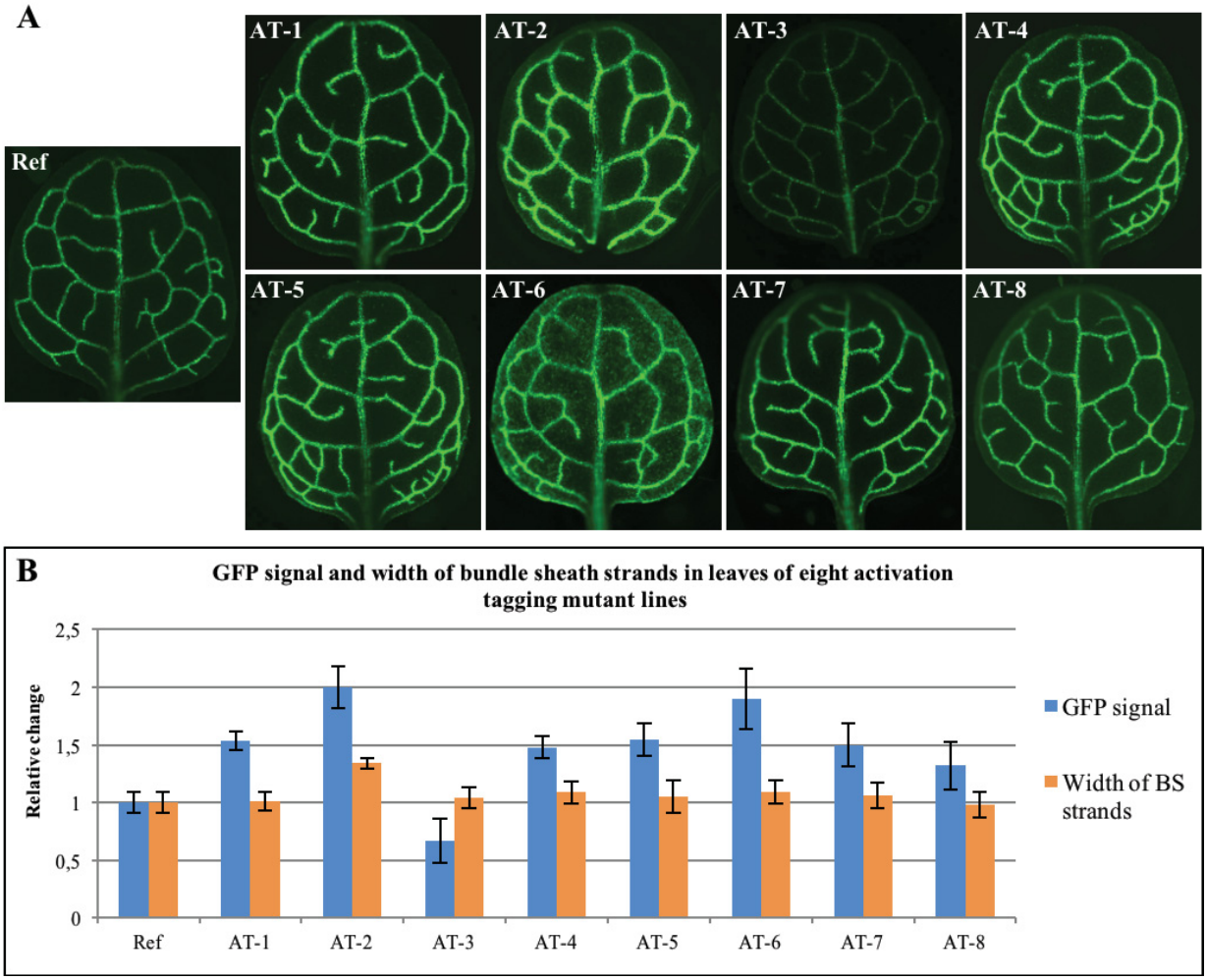


Figure 3. Overview of the eight stable activation tagging mutant lines.

(A) Fluorescence images of whole leaves of all activation tagging mutant lines (AT-1–8) and the reference line (Ref). (B) Relative changed of the GFP signal and width of bundle sheath strands in whole leaves of the activation tagging mutant lines compared to the reference line. BS, bundle sheath.

The only significant difference ($p < 0.05$) between mutant line and reference line was detected for the mutant line AT-2 in which the width of bundle sheath strands was enlarged by 34%. Interestingly, AT-2 also had the highest reporter gene expression among all the activation tagging mutants. Our forward genetic screen aimed to find the genes that activate the bundle sheath, and

hence, we chose to focus on the AT-2 mutant, which, in this regard, indicated the highest potential for an interesting candidate gene.

Mutant line AT-2 (*BSOM2*) possessed more bundle sheath cells and enlarged vascular tissue

The difference between the mutant line AT-2 and the reference line was not only restricted to the reporter gene signal and the width of bundle sheath strands, but also included the general growth of the plant. Firstly, the rosette leaves of the mutant were considerably smaller at every stage of plant growth (Figure 4A). Secondly, the mutant started flowering approximately eight days before the reference line under the given growth conditions, and thirdly, the siliques remained almost exclusively empty. Nevertheless, we were able to identify the T-DNA insertion containing the activation tag by Thermal Asymmetric Interlaced PCR (TAIL-PCR). TAIL-PCR helps to analyze unknown DNA fragments that are adjacent to known sequences. We used a mix of specific primers to the right border of the T-DNA and arbitrary degenerated primers that bind to the genomic DNA at many positions. After two following nested PCR reactions, we could amplify specific PCR products that were sequenced and mapped to the Arabidopsis genome. We found the T-DNA insertion on the long arm of chromosome five, 800 bp in front of the gene AT5G52050 (Figure 4D; results for AT-1 and AT-3 are shown in Supplementary Figure 1). The p*GLDT*$_{Ft}$ promoter, as part of the T-DNA sequence, was inserted in 5'→3'-orientation according to the gene AT5G52050. There were no other genes within 5.4 kb upstream of the T-DNA insertion, and the first annotated sequence downstream was the retrotransposon AT5G52055 (distance of 7.3 kb). The insertion site was validated on both sides of the T-DNA by PCR with primer pairs binding in the flanking genomic region and in the distal or proximal ends of the T-DNA. The candidate gene AT5G52050 is termed *BSOM2* as it emanates from the activation tagging mutant line AT-2 and follows the designated nomenclature of *BSOM* genes (bundle sheath ontogeny and maintenance) genes. BSOM2 has already been described by a Chinese group (Zhang et al., 2014). They found out that it is a member of the DTX/Multidrug and Toxic Compound Extrusion (MATE) family and functions as an abscisic acid (ABA) efflux transporter. Promoter-GUS (β-glucuronidase) fusions indicated expression of the gene mainly in the vascular tissue. They could also show that the complete knockout of *BSOM2* by T-DNA resulted in growth retardation and ABA hypersensitivity (Zhang et al., 2014). Surprisingly, the growth of the *bsom2* mutant strongly resembled the growth of our original *BSOM2* mutant line (AT-2), in which we assumed an overexpression of *BSOM2* in the bundle sheath and vasculature. To verify this, we overexpressed *BSOM2* with the p*GLDT*$_{Ft}$ promoter in the reference line and compared it to the original mutant.

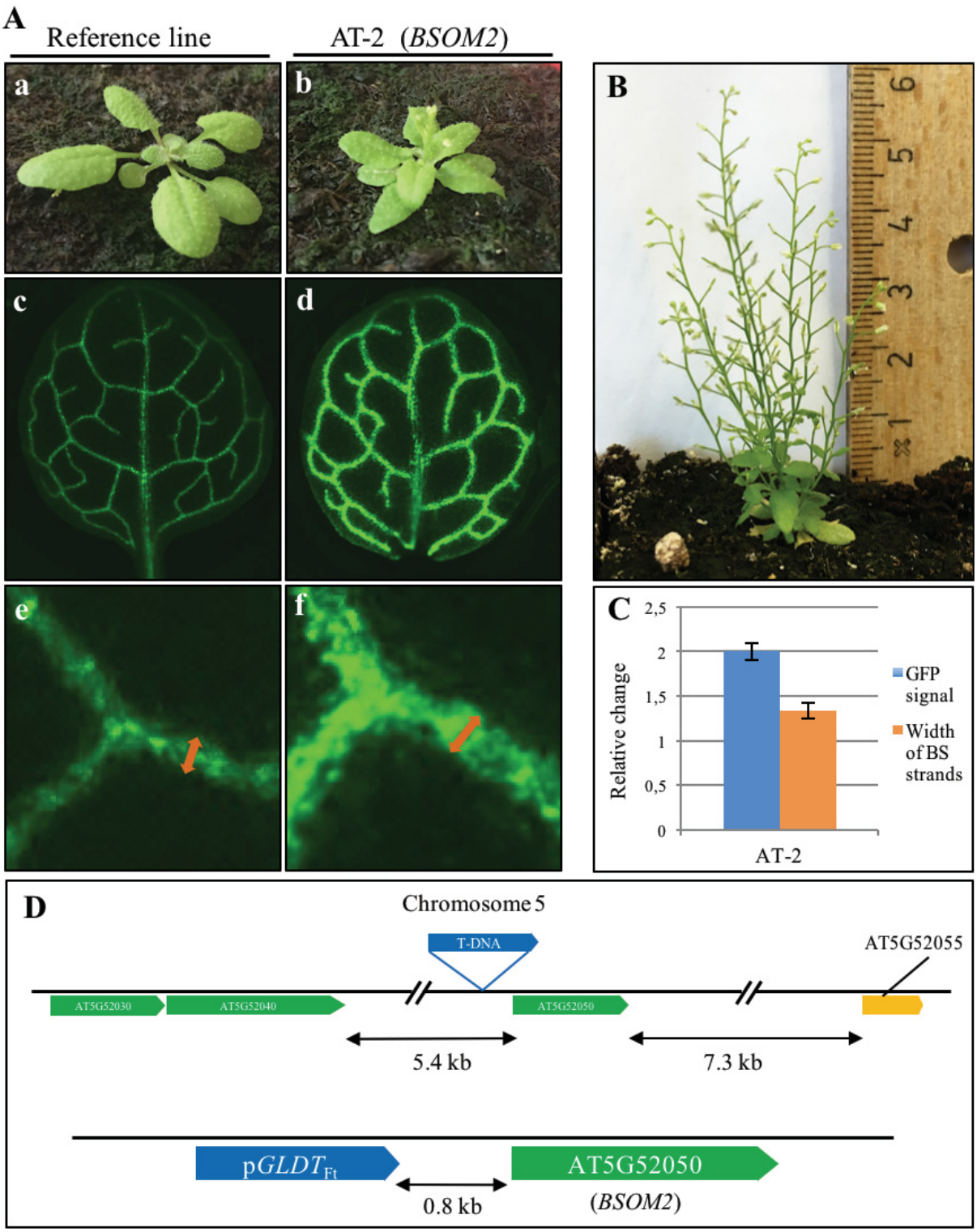


Figure 4. Characteristics of the *BSOM2* mutant line (AT-2).

(A) Plant growth and GFP signal in the *BSOM2* mutant line compared to the reference line. The mutant is smaller and flowers earlier than the reference line (a and b). Fluorescence images of whole leaves show a strongly increased GFP signal in the mutant leaf (c and d). Additionally, the width of bundle sheath strands is enlarged in the mutant line (e and f). (B) The full-grown *BSOM2* mutant line. (C) Relative signal intensity and width of bundle sheath strands in single leaves of the mutant line compared to the reference line. (D) Localization of the T-DNA insertion in the *BSOM2* mutant line. The activation tag is inserted on the long arm of chromosome five in front of the gene AT5G52050. AT5G52055 represents a retrotransposon. A close-up view of the T-DNA insertion shows that p*GLDT*$_{Ft}$ is inserted in the same orientation as the following gene AT5G52050. The distance between the proximal part of the promoter and the ATG of the gene is 800 bp.

The *BSOM2* mutant phenotype can be reconstructed

To confirm that the insertion of the p*GLDT*$_{Ft}$ promoter in front of AT5G52050 (*BSOM2*) was the causative event for the aberrant phenotype in the AT-2 mutant line, we tried to reconstruct the event by fusing the p*GLDT*$_{Ft}$ promoter to the coding region of the *BSOM2* gene. Additionally, we made one construct using the p(2x*35S*) promoter with no tissue-specificity. Both the constructs were stably transformed into the reference line and the transformants were checked via PCR for T-DNA insertion. Positive lines were screened for the *BSOM2* phenotype, in particular for more GFP signal and enlarged bundle sheath strands. More than 75% of the analyzed transformants possessed the characteristics of the *BSOM2* mutant (Figure 5). Homozygous lines with only one T-DNA copy were isolated in the T2 generation. The qPCR analysis of these lines showed that *BSOM2* mRNA levels indeed increased in the leaves of both the p(2x*35S*) and the p*GLDT*$_{Ft}$ overexpression lines by 80% and 240%, respectively (Figure 5C).

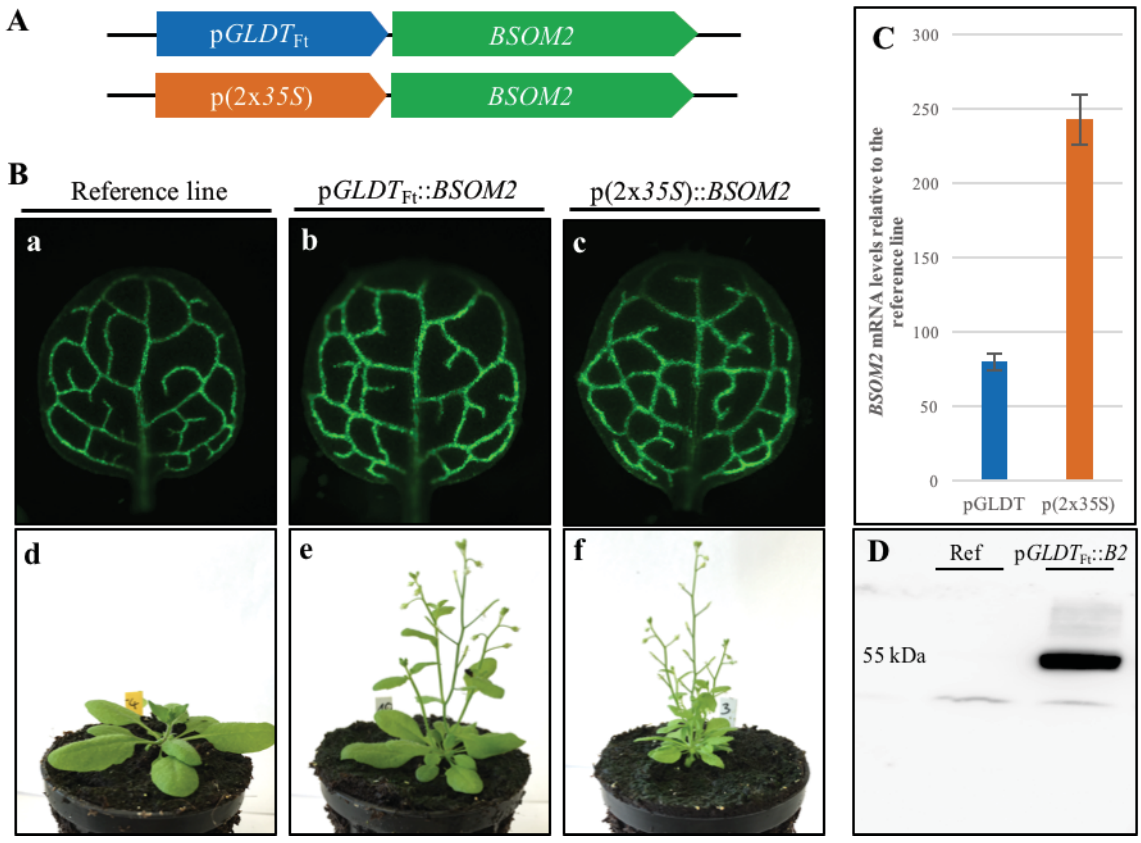


Figure 5. Reconstruction of the *BSOM2* mutant phenotype.

(A) The design of the constructs to overexpress the *BSOM2* gene with p*GLDT*$_{Ft}$ or p(2x*35S*) (B) Images of whole leaves and 28-day-old plants of the reference line (a and d), the p*GLDT*$_{Ft}$::*BSOM2* overexpression line (b and e), and the p(2x*35S*)::*BSOM2* overexpression line (c and f). The overexpression of the *BSOM2* gene with both p*GLDT*$_{Ft}$ and p(2x*35S*) resulted in retarded growth of the plant (e and f) as well as steeply increased GFP signal in the bundle sheath strands of the leaves (b and c). (C) The relative mRNA expression levels of the *BSOM2* gene in the p*GLDT*$_{Ft}$ and the p(2x*35S*) overexpression line. (D) Western blot analysis of the BSOM2 protein in membrane fractions isolated of leaves from the reference line and the p*GLDT*$_{Ft}$::*BSOM2* line.

Western blot analysis was carried out using antibodies against the HA-epitope with protein extracts of the p*GLDT*$_{Ft}$::*BSOM2* overexpression line and the reference line. While there was no recombinant protein detectable in the reference line, we observed accumulation of the protein in the overexpression line underlying the results of the qPCR at the protein level. From these experiments, we can conclude that it is possible to reconstruct the aberrant phenotype of activation tagging line AT-2 by overexpressing *BSOM2* either in the bundle sheath and vasculature or in all cell types. Hence, we could confirm that we had identified the right candidate gene.

Microscopic characterization of *BSOM2* overexpression and knockout lines revealed alterations in bundle sheath and vascular tissue anatomy

We could reconstruct all properties of the original *BSOM2* mutant by overexpressing the *BSOM2* gene with the p*GLDT*$_{Ft}$ promoter that we had used for the activation tagging screen. These characteristics included more GFP signal in the bundle sheath strands, larger bundles, impaired plant growth, early flowering, and steeply reduced seed production. However, the specific effect on bundle sheath cell anatomy has remained unclear so far. To address this question, we prepared sections for high-resolution light microscopy (LM) of the reference line, the homozygous T3 plants of the p*GLDT*$_{Ft}$::*BSOM2* overexpression line, and the available homozygous T-DNA knockout line (SALK_144096), hereinafter referred to as *bsom2*. The T-DNA insertion in the only exon of *BSOM2* in *bsom2* was confirmed beforehand by PCR, and a homozygous line was used to prepare samples for LM of the knockout mutant *bsom2*. While sampling three biological replicates of each line, it turned out that the replicates hardly contained any comparable veins of the same vein order. Nonetheless, we found well developed 3° veins in at least one replicate of each sample (reference line, p*GLDT*$_{Ft}$::*BSOM2*, and *bsom2*) that we could compare against one another (Figure 6). It could be shown that the anatomy of the bundle sheath strands was tremendously affected in both the *BSOM2* overexpression line and the knockout line *bsom2*. The overexpression of the *BSOM2* gene led to a massive increase in vascular tissue, whereby the cells of both phloem and xylem increased in number as compared to the reference line (Figure 6). Additionally, we observed more bundle sheath cells surrounding the vasculature, which might be a logical consequence, since there were no gaps detectable in the rows of bundle sheath cells around the vasculature. Furthermore, the bundle sheath cells in the overexpression line appeared to be slightly smaller compared to the reference line. Although the sample size of the independent bundles of the same vein order was too low to perform statistically relevant measurements, it seemed obvious that the overall area of bundle sheath tissue largely increased in the *BSOM2*

overexpression line. Surprisingly, a very similar phenotype could be observed in the knockout line *bsom2* in which both the size of the vasculature and the total bundle sheath tissue increased as well; however, not to the same extent as in the overexpression line (Figure 6). While it seemed that the bundle sheath cell number was comparable between p*GLDT*$_{Ft}$::*BSOM2* and *bsom2*, the bundle sheath cell size tended to be smaller in the knockout line. The cross sections of *bsom2* indicated large gaps in the mesophyll, both in palisade and spongy parenchyma. However, this phenomenon is most likely due to issues during the embedding process of the leaf samples as it could be also observed in some samples of the reference line. To conclude, it seems that both the overexpression and the knockout of the *BSOM2* gene result in a similar phenotype in which the bundle sheath cell number is increased and the vasculature is enlarged.

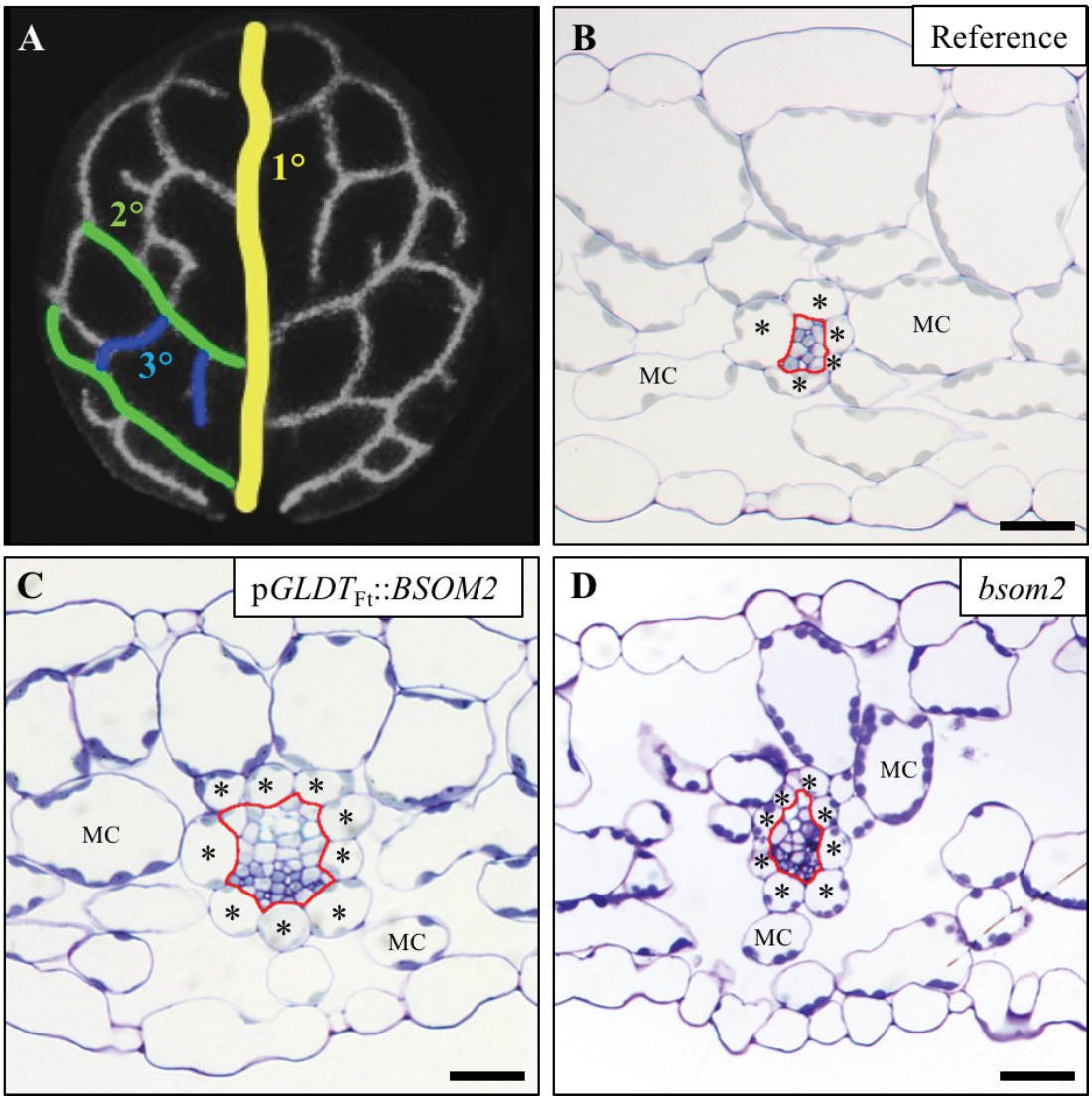


Figure 6. Analysis of bundle sheath strands in *BSOM2* overexpression/knockout lines.

(A) The vein order of a young Arabidopsis leaf showing 1°, 2°, and 3° veins. Leaf cross sections of (B) the reference line, (C) the *BSOM2* overexpression line, and (D) the T-DNA knockout line *bsom2*. Each sample shows a representative 3° vein. Bundle sheath cells are indicated with an asterisk. The complete vascular tissue is framed by a red line. MC, mesophyll cell. Scale bars = 25 µm.

Discussion

We wanted to find genes involved in the activation of the bundle sheath by using an activation tagging approach. Döring et al. (2017, unpublished; Manuscript 2) described a similar approach in which they used the chemical mutagen ethyl methanesulfonate (EMS) to introduce single-nucleotide polymorphisms (SNPs) randomly in the genome of *A. thaliana*. EMS-based genetic screens benefit from high mutation frequencies, often resulting in large number of mutant lines, but they might not cover all genes in the process of interest. Redundant genes, for example, will probably not show any phenotype when knocked out or knocked down. Additionally, it is still tedious to reliably link the phenotype to a genotype in EMS mutant lines. In this study, we described a forward genetic screen based on the random insertion of the p*GLDT*$_{Ft}$ promoter into the genome of an Arabidopsis reporter line with labeled bundle sheath cells that was previously described by Döring et al. (2017, unpublished; Manuscript 2).

Activation tagging with the p*GLDT*$_{Ft}$ promoter discovered a mutant with more bundle sheath cells

The random insertion of enhancers or promoters via T-DNA was first described by Hayashi et al. (1992), and became known as activation tagging. Activation tagging screens are a powerful tool in forward genetics since they also reveal genes that cannot be found by the classical knockout and knockdown approaches. Depending on the insertion of the T-DNA in the genome, the result can be manifold. Beside up-regulation or ectopic activation of neighbored genes, T-DNA insertion might also lead to classical gene knockouts. One of the most used sequences for the activation tag is the enhancer element from the constitutively active promoter of the cauliflower mosaic virus (CaMV) *35S* gene that has been used in numerous studies in *A. thaliana* (Hayashi et al., 1992; Weigel et al., 2000; Tani et al., 2004; Pogorelko et al., 2008). Enhancers can facilitate the transcriptional activation of nearby genes. However, it has also been shown that insulator sequences in the genome can negatively affect adjacent enhancers or silencers (Chung et al., 1993; Weigel et al., 2000). Additionally, promoter selectivity of endogenous promoters can hinder CaMV *35S* enhancers to activate genes (Ohtsuki et al., 1998).

In the context of a genetic screen that aims to find mutants with alterations in bundle sheath anatomy, we decided to use a promoter that is highly active in the bundle sheath cells of *A. thaliana*. The genomic region of 3.2 kb upstream of the *GLDT* gene in *F. trinervia* was shown to be sufficient for a strong gene expression in the bundle sheath and vasculature in both *F. trinervia* and *A. thaliana* (Emmerling, 2017; unpublished). We transformed the p*GLDT*$_{Ft}$

promoter via T-DNA into the reference line and screened M1 plants for aberrant reporter gene signal. 175 plants with at least 30% increased or decreased GFP signals in the leaves were isolated and tested in following generations for stability of the phenotype, which eventually resulted in eight homozygous mutant lines with stable aberrant reporter gene activity in the leaves. We localized the T-DNA insertion in three mutants (Figure 4; Supplementary Figure 1). By analyzing one mutant line (AT-2) in more detail with high-resolution light microscopy, we could show that it is possible to obtain mutants from this screen with altered bundle sheath anatomy. Thus, the activation tagging approach of using the full-length $pGLDT_{Ft}$ promoter can be considered as a success.

BSOM2—an ABA efflux transporter

Our *BSOM2* mutant line was characterized by increased reporter gene activity and larger bundle sheath strands, which were the two main criteria we used as a proxy in the primary screen. By the selective overexpression of the *BSOM2* gene in the bundle sheath and vasculature of the reference line, we were able to reconstruct the aberrant phenotype. This made us to conclude that it was solely the overexpression of the *BSOM2* gene that caused larger bundle sheath strands and stronger GFP signal in our mutant (Figure 5). The BSOM2 protein has been already described as an ABA efflux transporter in *A. thaliana* that is localized to the plasma membrane and functions in the export of ABA from the cytosol to the outside of a cell (Zhang et al., 2014). Zhang et al. (2014) could show that transcription of the *BSOM2* gene took place mainly in the vasculature and to some extent in the guard cells, which was in accordance with the fact that most ABA biosynthesis enzymes are localized in vascular parenchyma cells (Koiwai et al., 2004; Endo et al., 2008). In this context, these studies might explain the obvious impact on the vasculature in both the *BSOM2* overexpression line and the *bsom2* knockout mutant (Figure 6). We observed a massive increase in leaf vascular tissue in the p$GLDT_{Ft}$::*BSOM2* overexpression line. This, as can be seen from Figure 6, is mostly due to more cells in both the phloem and the xylem. An enlarged vasculature might be the primary effect of *BSOM2* overexpression or *BSOM2* knockout; however, there is also an obvious impact on the bundle sheath cells. In both cases, bundle sheath cell number almost doubled as compared to the reference line, which is very exciting in terms of our initial aim to find mutants with, inter alia, more bundle sheath tissue (Figure 6). However, it remains highly questionable whether BSOM2 in particular can be helpful to increase bundle sheath tissue size in C_3 crops due the severe growth reduction caused by the overexpression of this gene. Furthermore, plant fecundity was dramatically lowered in these plants. In this regard, a precise and careful regulation of the *BSOM2* gene might be more successful rather than a plain

overexpression or knockout. However, further experiments have been initiated to reveal how overexpression as well as knockdown of *BSOM2* orthologous genes in rice affects the bundle sheath vasculature in the C_3 monocot.

The vein density in our samples for high-resolution microscopy was relatively low. It will be mandatory to again analyze more samples in order to validate our results and to allow statistically evaluable data analysis such as the measurement of vascular area, bundle sheath total cell area, and number of bundle sheath cells. The samples obtained in this study do not allow quantifying these parameters; nevertheless, they give a qualitative impression of the changes in bundle sheath cells and vascular tissue caused by the altered expression of *BSOM2*.

The plant hormone ABA and its relationship to the vascular tissue

ABA is a phytohormone that plays a key role in regulating plant development and adaptation to biotic and abiotic stresses (Zeevaart and Creelman, 1988; Cutler et al., 2010). Although ABA is produced in the vasculature, a correlation between the plant hormone and vascular differentiation in general is not known. It is most famous for its function in drought adaptive responses such as stomata closure (Finkelstein et al., 2002). ABA is synthesized from carotenoids and needs to be transported from the site of synthesis to the site where it is needed, for instance in the guard cells. Notwithstanding the fact that ABA can pass biological membranes by diffusion, there is evidence that it mostly moves in and out of cells by transport proteins (Wilkinson and Davies, 2010; Ng et al., 2014). In Arabidopsis, members of an ATP-binding cassette (ABC)-containing transporter protein family were described to actively shuttle ABA. In addition to the ABC family, Zhang et al. (2014) identified that BSOM2 as part of the DTX/Multidrug and Toxic Compound Extrusion (MATE) family was an essential player in ABA transport out of a cell. A complete knockout of the gene resulted in accumulation of ABA in the leaf, thereby increasing ABA responses such as more severe growth inhibition by ABA, which explained the retarded growth of the mutant plant (Zhang et al., 2014). Assuming that the veins are a major place for ABA synthesis, a knockout of an important ABA efflux transporter could lead to accumulation of ABA in the vasculature. In general, plant responses to ABA are triggered by binding to soluble Pyrabactin resistance/Pyrabactin-like/regulatory components of the ABA receptor (PYR/PYL/RCAR) receptors (Fujii et al., 2009; Ma et al., 2009; Park et al., 2009). However, a connection between high ABA levels in the vasculature and ABA-induced alterations in the size of this compartment remains highly speculative. In addition, even if it might explain the phenotype of the knockout mutant *bsom2*, the very similar effect on the vascular tissue observed in the *BSOM2* overexpression line would still remain unexplained.

Overexpression and knockout of the *BSOM2* gene have similar effects on the anatomy of bundle sheath cells and vascular tissue

Regardless of direct ABA involvement or not, it is quite puzzling that the overexpression of a gene results in more or less the same phenotype as the knockout mutant. The evidence available indicates that this is truly the case for the *BSOM2* overexpression and knockout line. It can be excluded that the T-DNA insertion in *bsom2* does not lead to a complete knockout of the gene since it was validated by RT-PCR, and, additionally, it was shown that the knockout mutant could be complemented (Zhang et al., 2014). Sometimes the introduction and overexpression of transgenes result in gene silencing, which could explain the similar phenotype of *bsom2* and p$GLDT_{Ft}$::*BSOM2* (Stam et al., 1997). However, we could show that the recombinant BSOM2 protein was expressed in the *BSOM2* overexpression line, which was validated at the mRNA and protein levels (Figure 5). In summary, the previous results of Zhang et al. (2014) together with the experiments described in this study, strongly suggest that both the overexpression and the knockout of *BSOM2* result in an enlargement of vascular tissue in the leaves of *A. thaliana*.

Notwithstanding the results of this work, there is still no satisfactory answer to the question of whether or not increased ABA levels in the vasculature have a direct or indirect effect on the enlargement of vascular tissue observed in the *BSOM2* overexpression line as well as in the *bsom2* knockout mutant. Moreover, the similar phenotypes in both lines with respect to the cells of the bundle sheath and vasculature and also to the general plant growth remain unexplained. In order to address these questions, further experiments need to be planned. Firstly, the expression pattern of *BSOM2* should be re-examined. While the results of Zhang et al. (2014) indicated promoter activity predominantly in vascular tissue and in guard cells, they still saw reporter gene activity in the bundle sheath and also slightly in mesophyll cells. The promoter activity should be analyzed once again by using a different reporter gene such as H2B-YFP since GUS-staining often results in the diffusion of the reporter signal into the surrounding cells and strongly depends on the incubation time of the samples. Nevertheless, RNA *in situ* hybridization might be a better solution to reliably pinpoint the tissue in which the *BSOM2* promoter is active. Immunolocalization of the BSOM2 protein in high-quality cross sections is needed to see where the protein is distributed within the leaf. This information would be crucial to understand the role of BSOM2 with respect to our bundle sheath and vascular tissue mutant.

ABA levels are significantly increased in the leaves of *bsom2* and, in addition to this, several ABA marker genes are more active in the knockout mutant (Zhang et al., 2014). This needs to be analyzed in the *BSOM2* overexpression line as well. If we could also measure more ABA in this line, it would support the theory that increased ABA levels are a key factor for an enlarged

vascular tissue and, in addition, more surrounding bundle sheath cells. Furthermore, ABA levels increase when plants are subjected to biotic stresses such as drought (Zeevaart and Creelman, 1988; Zhu, 2002). All the plants that were used in the experiments described in this study were well watered and grew up in the absence of any stress. It would be interesting to compare this data to plants that were confronted with increased ABA levels during plant growth to analyze whether or not there is an effect on the differentiation of the bundle sheath and vasculature.

Acknowledgements

This work was supported by the European Union through the 3to4 project—Converting C_3 to C_4 Photosynthesis for Sustainable Agriculture. We thank Jan Emmerling, who provided the p*GLDT*$_{Ft}$ promoter that was used in the activation tagging screen, and we thank Kumari Billakurthi who performed the high-resolution microscopic analysis.

References

Akhani H, Khoshravesh R (2013) The relationship and different C4 Kranz anatomy of Bassia eriantha and Bassia eriophora, two often confused Irano-Turanian and Saharo-Sindian species. Phytotaxa **93**: 1–24

Black C, Mollenhauer H (1971) Structure and distribution of chloroplasts and other organelles in leaves with various rates of photosynthesis. Plant Physiol **47**: 15–23

Von Caemmerer S, Furbank RT (2003) The C4 pathway: An efficient CO2 pump. Photosynth Res **77**: 191–207

Castle LA, Errampalli D, Atherton TL, Franzmann LH, Yoon ES, Meinke DW (1993) Genetic and molecular characterization of embryonic mutants identified following seed transformation in Arabidopsis. MGG Mol Gen Genet **241**: 504–514

Chung JH, Whiteley M, Felsenfeld G (1993) A 5' element of the chicken ß-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. Cell **74**: 505–514

Collins TJ (2007) ImageJ for microscopy. Biotechniques **43**: 25–30

Curtis MD, Grossniklaus U (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol **133**: 462–9

Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR (2010) Abscisic acid: emergence of a core signaling network. Annu Rev Plant Biol **61**: 651–679

Dengler N, Dengler RE, Donnelly PM, Hattersley PW (1994) Quantitative leaf anatomy of C3 and C4 grasses (Poaceae): undle sheath and mesophyll Surface area relationships. Ann Bot **73**: 241–255

Dengler NG, Nelson T (1999) Leaf structure and development in C4 plants. C4 Plant Biol. pp 133–172

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

Endo A, Sawada Y, Takahashi H, Okamoto M, Ikegami K, Koiwai H, Seo M, Toyomasu T, Mitsuhashi W, Shinozaki K, et al (2008) Drought induction of Arabidopsis 9-cis-epoxycarotenoid dioxygenase occurs in vascular parenchyma cells. Plant Physiol **147**: 1984–1993

Engelmann S, Wiludda C, Burscheidt J, Gowik U, Schlue U, Koczor M, Streubel M, Cossu R, Bauwe H, Westhoff P (2008) The gene for the P-subunit of glycine decarboxylase from the C4 species Flaveria trinervia: analysis of transcriptional control in transgenic Flaveria bidentis (C4) and Arabidopsis (C3). Plant Physiol **146**: 1773–1785

Finkelstein RR, Gampala SSL, Rock CD (2002) Abscisic acid signaling in seeds and seedlings. Plant Cell Online **14**: S15--S45

Fujii H, Chinnusamy V, Rodrigues A, Rubio S, Antoni R, Park S-Y, Cutler SR, Sheen J, Rodriguez PL, Zhu J-K (2009) In vitro reconstitution of an abscisic acid signalling pathway. Nature **462**: 660–4

Furbank RT (2011) Evolution of the C4 photosynthetic mechanism: Are there really three C4 acid decarboxylation types? J Exp Bot **62**: 3103–3108

Griffiths H, Weller G, Toy LFM, DENNIS RJ (2013) You're so vein: Bundle sheath physiology, phylogeny and evolution in C3 and C4 plants. Plant, Cell Environ **36**: 249–261

Hayashi H, Czaja I, Lubenow H, Schell J, Walden R (1992) Activation of a plant gene by T-DNA tagging: auxin-independent growth in vitro. Sci (New York, NY) **258**: 1350–1353

Koiwai H, Nakaminami K, Seo M, Mitsuhashi W, Toyomasu T, Koshiba T (2004) Tissue-specific localization of an abscisic acid biosynthetic enzyme, AAO3, in Arabidopsis. Plant Physiol **134**: 1697–707

Lazo GR, Stein PA, Ludwig RA (1991) A DNA transformation-competent Arabidopsis genomic library in Agrobacterium. Biotechnol (Nature Publ Company) **9**: 963–7

Liu Y-GG, Whittier RF (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics **25**: 674–681

Lundgren MR, Osborne CP, Christin P-A (2014) Deconstructing Kranz anatomy to understand C4 evolution. J Exp Bot **65**: 3357–69

Lyska D, Engelmann K, Meierhoff K, Westhoff P (2013) pAUL: a Gateway-based vector system for adaptive expression and flexible tagging of proteins in Arabidopsis. PLoS One **8**: 1–12

Ma Y, Szostkiewicz I, Korte A, Moes D, Yang Y, Christmann A, Grill E (2009) Regulators of PP2C phosphatase activity function as abscisic acid sensors. Science **324**: 1064–1068

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

Muhaidat R, Sage TL, Frohlich MW, Dengler NG, Sage RF (2011) Characterization of C3-C4 intermediate species in the genus Heliotropium L. (Boraginaceae): Anatomy, ultrastructure and enzyme activity. Plant, Cell Environ **34**: 1723–1736

Ng LM, Melcher K, Teh BT, Xu HE (2014) Abscisic acid perception and signaling: structural mechanisms and applications. Acta Pharmacol Sin **35**: 567–584

Ohtsuki S, Levine M, Cai HN (1998) Different core promoters possess distinct regulatory activities in the Drosophila embryo. Genes Dev **12**: 547–556

Park S-Y, Fung P, Nishimura N, Jensen DR, Fujii H, Zhao Y, Lumba S, Santiago J, Rodrigues A, Chow T-FF, et al (2009) Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. Science **324**: 1068–1071

Pogorelko G V., Fursova O V., Ogarkova OA, Tarasov VA (2008) A new technique for activation tagging in Arabidopsis. Gene **414**: 67–75

Rossini L, Cribb L, Martin DJ, Langdale JA (2001) The maize golden2 gene defines a novel class of transcriptional regulators in plants. Plant Cell **13**: 1231–44

Sage RF (2001) Environmental and evolutionary preconditionsfor the origin and diversification of the C4 photosynthetic syndrome. Plant Biol **3**: 202–213

Stam M, Mol JNM, Kooter JM (1997) The silence of genes in transgenic plants. Ann Bot **79**: 3–12

Stata M, Sage TL, Rennie TD, Khoshravesh R, Sultmanis S, Khaikin Y, Ludwig M, Sage RF (2014) Mesophyll cells of C4 plants have fewer chloroplasts than those of closely related C3 plants. Plant Cell Environ **37**: 2587–2600

Tani H, Chen X, Nurmberg P, Grant JJ, Santa Maria M, Chini A, Gilroy E, Birch PR, Loake GL (2004) Activation tagging in plants: A tool for gene discovery. Funct Integr Genomics **4**: 258–266

Waters MT, Langdale JA (2009) The making of a chloroplast. EMBO J **28**: 2861–2873

Waters MT, Moylan EC, Langdale JA (2008) GLK transcription factors regulate chloroplast development in a cell-autonomous manner. Plant J **56**: 432–444

Weigel D, Ahn JH, Blazquez MA, Borevitz JO, Christensen SK, Fankhauser C, Ferrandiz C, Kardailsky I, Malancharuvil EJ, Neff MM, et al (2000) Activation tagging in Arabidopsis. Plant Physiol **122**: 1003–1013

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

Wilkinson S, Davies WJ (2010) Drought, ozone, ABA and ethylene: new insights from cell to plant to community. Plant, Cell Environ **33**: 510–525

Zeevaart JAD, Creelman RA (1988) Metabolism and physiology of abscisic acid. Annu Rev Plant Physiol Plant Mol Biol **39**: 439–473

Zhang H, Zhu H, Pan Y, Yu Y, Luan S, Li L (2014) A DTX/MATE-type transporter facilitates abscisic acid efflux and modulates ABA sensitivity and drought tolerance in Arabidopsis. Mol Plant **7**: 1522–32

Zhang X, Henriques R, Lin S-S, Niu Q-W, Chua N-H (2006) Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. Nat Protoc **1**: 641–646

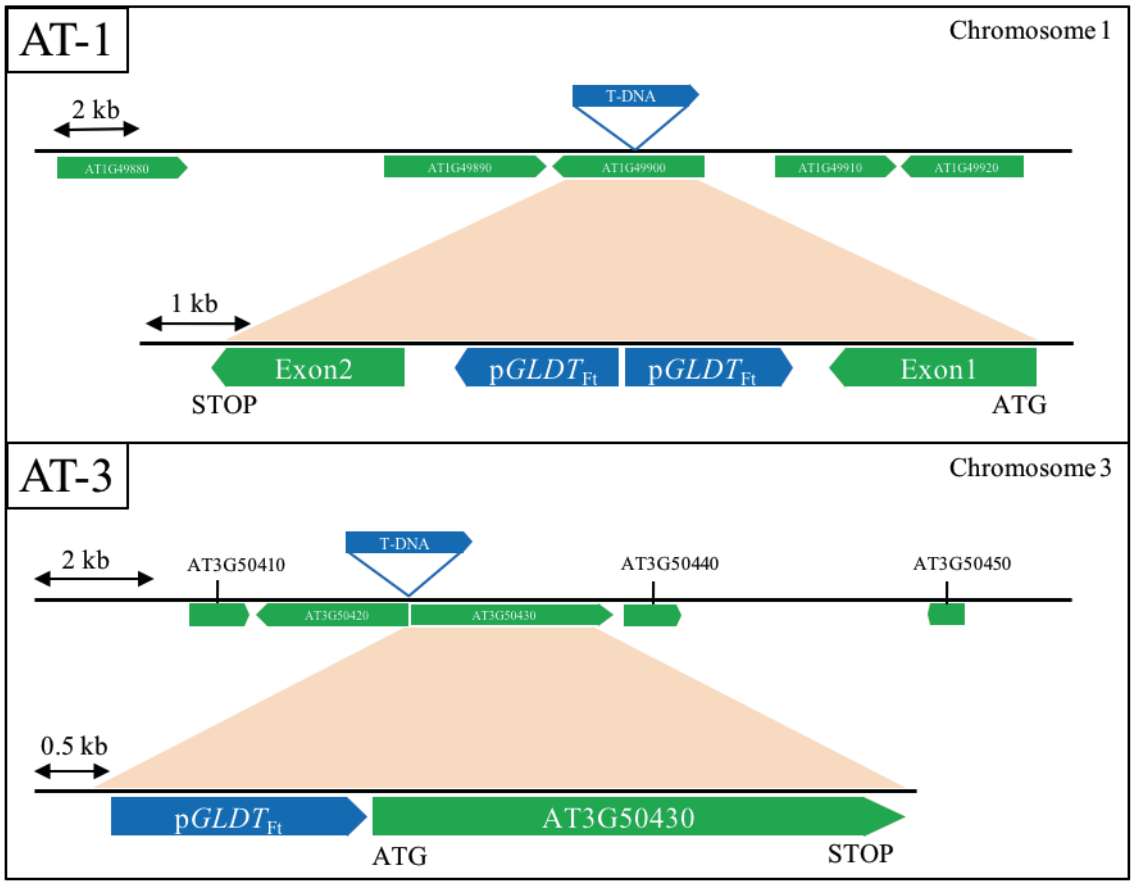
Zhu J-K (2002) Salt and drought stress signal transduction in plants. Annu Rev Plant Biol **53**: 247–273

Supplementary Data

Supplementary Table 1. Oligonucleotides used in this study.

F, forward primer; R, reverse primer.

Primer	Sequence (5'→ 3')	Orientation
p*GLDT*-Ft+*Sac*I	GAGCTCTCGACCCGTAAATAGGTCAA	F
p*GLDT*-Ft+*Pme*I	GTTTAAACGTGTGCTTTATTCTTTAGAAAC	R
p*GLD*T-Ft+*Pvu*I	GAGCTCTCGACCCGTAAATAGGTCAA	F
p*GLDT*-Ft+*Asc*I	GTTTAAACGTGTGCTTTATTCTTTAGAAAC	R
TAIL-LB-nested-1	CTGCTGAAGTCCCTGGAGG	R
TAIL-LB-nested-2	GGCATGACGTGGGTTTCTG	R
TAIL-LB-nested-3	AGGGTTTCGCTCATGTGTTG	R
TAIL-RB-nested-1	TGGGTTGATCATGTCGTAAAG	F
TAIL-RB-nested-2	CTTGAAGTTGGTGCGTTTGA	F
TAIL-RB-nested-3	TCCAAATTTGCTTCTTTGTGATT	F
TAIL-AD-1	NGTCGASWGANAWGAA	
TAIL-AD-2	TGWGNAGSANCASAGA	
TAIL-AD-3	AGWGNAGWANCAWAGG	
TAIL-AD-4	STTGNTASTNCTNTGC	
TAIL-AD-5	NTCGASTWTSGWGTT	
TAIL-AD-6	WGTGNAGWANCANAGA	
TAIL-AD-7	NGTCGASWGANAWGAA	
TAIL-AD-8	GTNCGASWCANAWGTT	
TAIL-AD-9	WGTGNAGWANCANAGA	
BSOM2-check-R	AATCCCCAAACCGAGGTAAC	
BSOM2-CDS+*att*B1	GGGGACAAGTTTGTACAAAAAAGCAGGCTATATGAGTCAATCAAATCGTGTCAC	F
BSOM2-CDS+*att*B2	GGGGACCACTTTGTACAAGAAAGCTGGGTTCTTATCAACCATCCCAGCCTC	R
qPCR-BSOM2-F	GTTGGAGCAGTTTTGGCGTT	F
qPCR-BSOM2-R	CCCAATCGGTCCTACACGTC	R
qPCR-Actin-F	CTCTCCTTGTACGCCAGTGG	F
qPCR-Actin-R	CAAGACGGAGGATGGCATGA	R



Supplementary Figure 1. T-DNA insertion in the activation tagging mutant lines AT-1 and AT-3.

In the mutant line AT-1, the T-DNA inserted within the gene AT1G49900 (*BSOM1*), which is composed of two exons and one intron. Two copies of the T-DNA inserted as an inverted repeat within the intron and caused the promoters to point to both exon 1 and exon 2. BSOM1 is a C2H2-type zinc finger transcription factor. In the mutant line AT-3, p*GLDT*$_{Ft}$ is inserted 30 bp in front of gene AT3G50430 (*BSOM3*). BSOM3 is an unknown protein.

The authors' contributions

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

KB prepared the embedded samples for light microscopy and performed the high-resolution light microscopic analysis.

TS helped with the operation of the microscopes and the analysis of the sections.

VI. Acknowledgements / Danksagungen

Ich danke…

...Prof. Dr. Peter Westhoff für die Möglichkeit, meine Doktorarbeit an seinem Institut durchführen zu können und für die hilfreiche Betreuung während dieser Zeit.

...Prof. Dr. Maria von Korff Schmising für die Übernahme des Koreferates.

...**Udo** für viele hilfreiche Ratschläge und die große Unterstützung bei meiner ersten Publikation.

...**Susanne** für die Beantwortung von 3923749872332 Fragen und die Lösung ebenso vieler Probleme.

...**Kumari,** కోసం సహకారం, కెనడా ప్రయాణం: సహాయం ధన్యవాదాలు + మంచి సంభాషణలు ధన్యవాదాలు

...**Sandra** für viele tolle Spieleabende, all die Postkarten und neuen Teesorten.

...allen ehemaligen und aktuellen Mitgliedern des „**KITA-Labors**“ für eine tolle Zeit und die Akzeptanz meines unberechenbaren Musikgeschmacks.

...dem restlichen **BotanikIV** Institut für viele nette Gespräche, Ratschläge, viel Spaß und nicht zuletzt für viele leckere Kuchen (eine extra Erwähnung verdient der Stachelbeer-Baiser Kuchen von Maria).

...den **Gärtnern** für die Pflege meiner Pflanzen und die Aufnahme meiner Fische, die sich augenscheinlich sehr wohl fühlen (Rückgabe ausgeschlossen).

...dem **3to4 Projekt** für die tollen Meetings an wunderschönen Orten weltweit.

…**meiner Familie** dafür, dass sie mich stets bei allem unterstützt hat und immer an mich geglaubt hat (Danke Mama auch für die große Hilfe bei meinen Steuererklärungen).

…**Loni** für ihre beruhigende und liebevolle Art, mit der sie mich in jeder Phase stets aufzumuntern wusste.

…**Nina** zunächst einmal dafür, dass sie bei uns ihre Bachelorarbeit angefangen hat und wir uns dadurch kennen lernen konnten; außerdem für die Kreuzworträtselpausen, die Besuche im botanischen Garten, endlose Gespräche und viele Gemeinsamkeiten. Ich hab dich lieb!