Identification of C_4 genetic determinants by comparative transcriptomics and forward genetic approaches

Inaugural-Dissertation

To obtain the doctoral degree
in the Faculty of Mathematics and Natural Sciences
at the Heinrich Heine University, Duesseldorf

Presented by

Kumari Billakurthi

from Srikakulam, India

Duesseldorf, September 2018

From the institute of "Developmental and Molecular Biology of Plants"
at the Heinrich Heine University, Duesseldorf

Published with the permission of
Faculty of Mathematics and Natural Sciences at
Heinrich Heine University, Duesseldorf

Supervisor: **Prof. Dr. Peter Westhoff**
Co-supervisor: **Prof. Dr. Maria von Korff Schmising**

Date of the oral examination: 17.12.2019

Declaration of the Doctoral Thesis

Herewith I declare that this thesis was written independently by myself and without any unauthorized help. I have listed the contribution of all authors and have cited all references properly. I assure that this thesis has not been submitted elsewhere for examination.

Duesseldorf, 14.09.2018

Kumari Billakurthi

Table of contents

I. Introduction

1. Open-ended high photorespiration rates constrain C_3 photosynthesis efficiency

1.1 The C_3 cycle

Photosynthesis is one of the basic life processes endowed in plants to assimilate atmospheric inorganic carbon (CO_2) to organic carbon compounds (biomass) with the release of molecular oxygen (O_2) in presence of solar energy. Further, it is divided in two parts, light and dark reactions. Light reaction is characterized by the photosynthetic electron transport chain consisting of two photosystems, the cytochrome bf complex, and the ATP synthase whose coordinated interplay results in the generation of energy-rich molecules (ATP and NADPH) *via* photolysis of water and electron mobility between PS-I and PS-II. During the dark reaction which is also known as Calvin-Benson-Bassham or C_3 cycle (Calvin, 1948; Benson and Calvin, 1950; Edwards and Walker, 1983), the generated energy-rich molecules are utilized for carbon assimilation. Based on specialization in this carbon fixation pathway, plants are broadly categorized C_3, C_4, or CAM plants. However, the majority of the plants follow C_3 pathway signifying its dominance while C_4 and CAM photosynthesis is seen only in 3 and 6 percentage of land plants (Ehleringer *et al.*, 1991; Borland *et al.*, 2011; Sage and Stata, 2015), respectively.

Mesophyll cells are major sites of photosynthesis in C_3 plants, where C_3 cycle initiates with RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE (Rubisco) enzyme carboxylating 5-C Ribulose-1,5-bisphosphate (RubP). This generates the first stable product of the reaction i.e. two molecules of 3-C containing phosphoglycerate (2×3-PGA) which are further reduced to two molecules of glyceraldehyde-3-phosphate (2×G3P). One in six carbon atoms exeunt the cycle and utilized to synthesize sugars and the rest of the carbon molecules are used to regenerate RubP (Figure 1). Reduction and regeneration phases require energy - fixation of one CO_2 molecule consumes 3ATP and 2 NADPH molecules (Edwards and Walker, 1983). Intermediate compounds of this cycle directly enter other primary and secondary metabolic pathways such as starch, shikimate, nucleotide, and isoprenoid biosynthetic pathways (Figure 1; Raines, 2011). This explains the crucial role of C_3 cycle in carbon metabolism and plant growth.

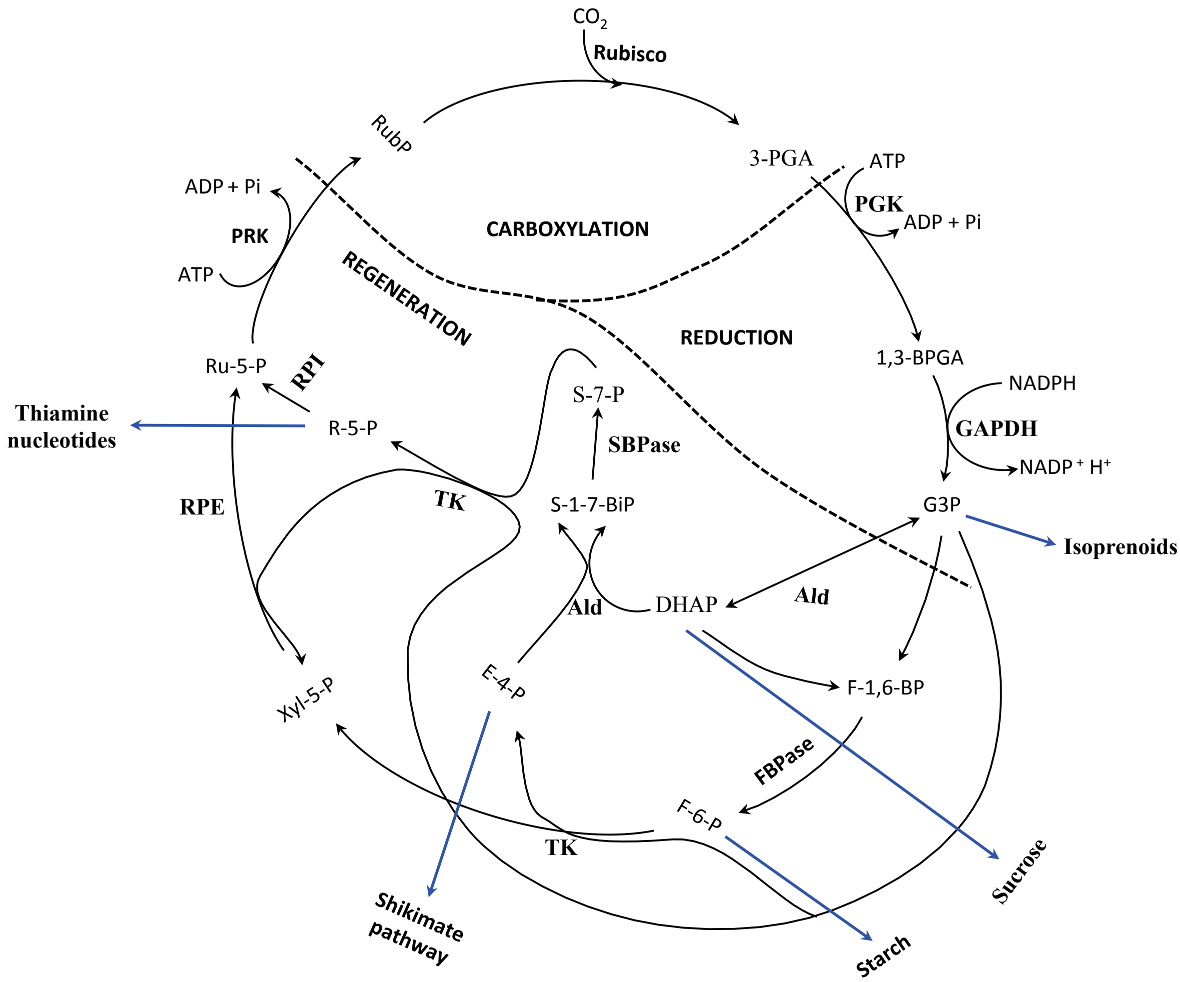


Figure 1. The Calvin-Benson-Bassham or C3 cycle.
The glyceraldehyde-3-phosphate (G3P) that is resulting from the reductive phase enters into regenerative phase. Subsequent reactions catalyzed by ALDOSE (Ald), FRUCTOSE-1,6-BISPHOSPHATASE (FBPase), TRANSKETOLASE (TK), SEDOHEPTULOSE-1,7-BISPHOSPHATASE (SBPase), RIBULOSE-5-P EPIMERASE (RPE), RIBOSE-5-P ISOMERASE (RPI) and PHOSPHORIBULOKINASE (PRK), regenerating RubP from G3P. Intermediate products of C_3 cycle act as precursor molecules in several metabolic pathways (blue arrows). 3-PGA – 3-phosphoglycerate; 1,3-BPGA – 1,3-bisphosphoglycerate; DHAP – dihydroxyacetone phosphate; F-1,6-BP – fructose-1,6-bisphosphate; F-6-P – fructose-6-phosphate; E-4-P – erythrose-4-P; S-1-7-P – sedoheptulose-1,7-bisphosphate; S-7-P – sedoheptulose 7-bisphosphate; Xyl-5-P – xylulose-5-P; R-5-P – ribose-5-P; Ru-5-P – ribulose-5-P; RubP - ribulose-1,5-bisphosphate; PGK – PHOSPHOGLYCERATE KINASE; GAPDH – G3P DEHYDROGENASE. Source: adapted from Raines, 2011.

Catalysis by Rubisco is the rate-limiting step in carbon assimilation. For a period of over 50 years, it was regarded as a sluggish enzyme in comparison to super catalysts like CARBONIC ANHYDRASE, TRIOSEPHOSPHATE ISOMERASE, and SUPEROXIDE DISMUTASE, but Rubisco kinetic properties are comparable to that of other enzymes operating in carbon metabolism (Bar-Even *et al.*, 2011; Bathellier *et al.*, 2018). On the other hand, Rubisco can fix both CO_2 and O_2 (Ogren and Bowes, 1971; Bowes *et al.*, 1971;

Andrews *et al.*, 1973) on account of their similar electrostatic properties (Whitney *et al.*, 2011). This substrate promiscuity limits the photosynthetic efficiency of C_3 plants. Therefore, the net CO_2 assimilation by Rubisco depends on the rate of carboxylation to oxygenation. Further, looking at the pedigree of this enzyme that evolved around 3.5 billion years ago, the earth's atmosphere was CO_2 saturated (>1000 ppm) with negligible O_2 levels (Sage, 1999; Whitney *et al.*, 2011). This corresponds to the unaccounted, Rubisco oxygenation reaction then (Sage, 1999). Consequently, with the evolution of oxygenic cyanobacterial and algal photosynthesis and also the emergence of land plants, atmospheric O_2 levels raised rapidly with the simultaneous decline in CO_2 levels (Whitney *et al.*, 2011). Today, with the higher percentage of O_2 in the environment (21 % O_2 and 0.04 % CO_2), Rubisco catalyzes two oxygenation reactions for every five carboxylation reactions (Walker *et al.*, 2016) in a C_3 leaf. This makes Rubisco an inefficient enzyme to effectively assimilate atmospheric CO_2 into the C_3 cycle.

1.2 The phenomenon of photorespiration

Fixation of O_2 to RubP instead of CO_2 by Rubisco results in one molecule each of 3-PGA and 2-phosphoglycolate (2-PG) (Ogren and Bowes, 1971; Bowes *et al.*, 1971; Andrews *et al.*, 1973). This newly generated 2-PG and its oxidized product glyoxylate, potentially inhibit TRIOSEPHOSPHATE ISOMERASE and Rubisco enzymes in the C_3 cycle and hence are phytotoxic at higher concentrations (Anderson, 1971; Campbell and Ogren, 1990). This sets the stage for the onset of the photorespiratory process that recycles the generated 2-PG molecule to 3-PGA back. However, photorespiration is a complex pathway, needing several enzymes, transporters, and participation of three different organelles; chloroplast, peroxisome and mitochondria (Bauwe *et al.*, 2010; Peterhansel *et al.*, 2010). Briefly, 2-PG is first dephosphorylated to glycolate in chloroplasts, which is later transported to peroxisomes and oxidized to glyoxylate with the help of GLYCOLATE OXIDASE (GOX) enzyme. Glycine produced here, as a result of transamination reaction, is then transported to mitochondria only to be decarboxylated and deaminated by GLYCINE DECARBOXYLASE COMPLEX (GDC). This reaction results in the release of CO_2, NH_4^+ along with methylene-tetrahydrofolate (CH_2-THF) and NADH as byproducts. CH_2-THF then combines with another molecule of glycine *via* SERINE HYDROXYMETHYLTRANSFERASE to generate one molecule of serine. Finally, this serine is transferred back to peroxisomes to produce glycerate as a precursor to be converted to 3-PGA in chloroplasts, eligible to enter into C_3 cycle. This process of recycling fatal 2-PG to one molecule of harmless 3-PGA requires two oxygenation reactions (Bauwe *et al.*, 2010; Peterhansel *et al.*, 2010). However, in this process, 3 ATP and 2 NAD(P)H molecules are utilized for the regeneration of RubP from 1.5 molecules of 3-PGA (Walker *et al.*, 2016) as a result of one oxygenation reaction. Along with this, reassimilation of released ammonia in the plastids via GS/GOGAT cycle also consumes additional energy (Peterhansel *et al.*, 2010). Further, this energy expensive process of photorespiration decreases the photosynthetic efficiency of C_3 plants as previously fixed carbon is lost (Figure 2; Sharkey, 1988; Walker *et al.*, 2016).

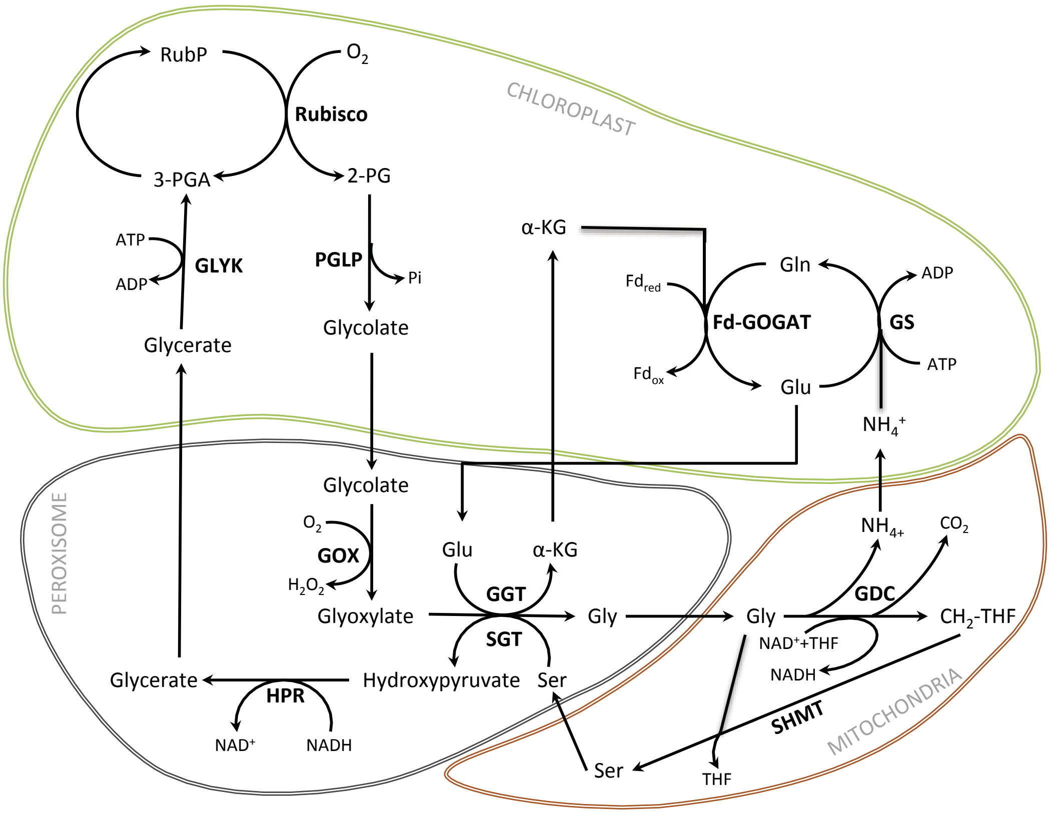


Figure 2. Photorespiratory cycle.

RubP – Ribulose-1,5-bisphosphate; 3-PGA – 3-phosphoglycerate; 2-PG – 2-phosphoglycolate; Gly – glycine; α-KG – α-ketoglutarate; CH_2-THF – methylene-tetrahydrofolate; Ser – serine; Glu – glutamate; Gln – glutamine; Rubisco – RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE; PGLP – PHOSPHOGLYCOLATE PHOSPHATASE; GOX – GLYCOLATE OXIDASE; GGT – GLUTAMATE:GLYOXYLATE AMINOTRANSFERASE; SGT – SERINE:GLYOXYLATE AMINOTRANSFERASE; SHMT – SERINE HYDROXYMETHYLTRANSFERASE; GDC – GLYCINE DECARBOXYLASE COMPLEX; HPR – HYDROXYPYRUVATE REDUCTASE; GLYK – GLYCERATE KINASE; GS – GLUTAMINE SYNTHETASE; Fd-GOGAT – FERREDOXIN-DEPENDANT, GLUTAMINE:OXOGLUTARATE AMINOTRANSFERASE. Source: adapted from Peterhansel *et al.*, 2010.

Carbon loss *via* photorespiration is about 26 % of net assimilation at a leaf temperature of 25°C, under current ambient CO_2 concentrations (350 ppm), which corresponds to 32 and 28 percentage of total ATP and NADH consumption (Walker *et al.*, 2016). This United States based study, estimated yield loss of 36 and 20 percentage in soybean and wheat crops, respectively, due to photorespiration. This means, if photorespiration is reduced even by 5 %, a benefit of no less than 540 million US$ in-terms of saved crop yield can be made (Walker *et al.*, 2016). The rise in temperature further accelerates yield loss through photorespiration (Laing *et al.,* 1974). With increasing temperature, CO_2/O_2 specificity of Rubisco and relative solubility of CO_2 to that of O_2 decreases (Brooks and Farquhar, 1985; Jordan and Ogren, 1984; Galmes *et al.*, 2005). This results in doubling the oxygenation rate of RubP from 15 to 35°C than carboxylase activity under ambient CO_2 levels (Sage *et al.*, 2018). Stomatal closure at high temperatures and in drought further increases internal O_2 concentrations (Parry *et al.*, 2007; Wingler *et al.*, 2000) and subsequent photorespiration.

2. Need and an approach for establishing enhanced photosynthetic efficiency in C_3 crops

Until the mid of the present century, the current world population is expected to increase by two billion (Food and Agriculture Organization (FAO), 2017). To meet the food demands of this rapidly growing population, photosynthetic efficiency needs to be enhanced by at least 1.7 % per year (Ray *et al.*, 2013; Foyer *et al.*, 2017). At the same time, fast-paced urbanization, increasing temperature, drought conditions and other human activities decreasing the per capita arable land (Food and Agriculture Organization (FAO), 2017). To add to these problems, photosynthetic efficiency of the majority of crop plants remained stagnant or decreased in the last few years (Ray *et al.*, 2012).

Net crop yield mainly depends on four factors (Long *et al.*, 2006; Zhu *et al.*, 2010; Jansson *et al.*, 2018); i) the amount of photosynthetically active radiation (PAR) that hits the leaf surface; ii) the efficiency of a plant in capturing PAR, called light interception capacity; iii) efficiency of plant in converting the absorbed solar energy (PAR) into biomass, termed photosynthetic light conversion efficiency and iv) allocation of this biomass into harvestable plant parts termed as harvest index or partitioning efficiency. While many crops already attained their theoretical maxima of light interception and partitioning efficiencies, light conversion efficiency is still far below to its threshold (Long *et al.*, 2006; Zhu *et al.*, 2010;

Jansson *et al.*, 2018). Therefore, this makes it the number one target for improving agricultural productivity.

Reducing RubP oxygenation reaction and subsequent photorespiration rates would substantially improve the photosynthetic performance of C_3 plants (Parry *et al.*, 2007). Oxygenase activity is primarily affected by internal CO_2 and O_2 concentration and its specificity towards CO_2 (Parry *et al.*, 2007). Increasing mesophyll conductance (g_m) i. e. CO_2 diffusion from intercellular airspace to the chloroplast stroma might be one good approach to magnify CO_2 levels at the site of carboxylation (Sage, 2013). However, achieving this by molecular approach is extremely challenging due to the incomplete knowledge of the underlying mechanism (Parry *et al.*, 2007). Interestingly, increasing CO_2 specificity of Rubisco reduces photorespiration to some extent albeit at a lower turnover rate (Sage, 2013; Galmés *et al.*, 2014). Previous attempts on engineering naturally existing highly specific Rubisco from marine red-algae into higher plants failed due to different gene regulatory mechanisms (Whitney *et al.*, 2001). While, photorespiratory knockout mutants of both C_3 (*Arabidopsis thaliana*) and C_4 (*Zea mays*) species had deleterious effects on plant growth leading to seedling lethality (Engel *et al.*, 2007; Schwarte and Bauwe, 2007; Zelitch *et al.*, 2009). This suggests the need for an intact photorespiratory cycle as it plays a crucial role in one-carbon metabolism and nitrogen assimilation (Bauwe *et al.*, 2010; Busch *et al.*, 2018).

Fortunately, there exist two possible ways of reducing photorespiratory yield loss in C_3 plants. One strategy could be targeting bacterial glycolate pathway into chloroplast of C_3 plants (Kebeish *et al.*, 2007; Leegood, 2008; Peterhansel and Maurino, 2011) to bypass the photorespiration cycle while the second promising method is to engineer naturally existing CO_2 concentration mechanism (CCM) from C_4 plants (Hatch, 1987). It is assumed that introducing C_4 CCM mechanism into C_3 crop plants such as rice, substantially improves their photosynthetic performance by about 50 % (Hibberd *et al.*, 2008; von Caemmerer et al., 2012).

3. C_4 photosynthesis – an evolutionary tool counteracting photorespiration rates

3.1 Spatial distribution of C_4 cycle enzymes facilitates CO_2 concentration mechanism

Following C_3 development, plants evolved with an advanced mechanism called C_4 photosynthesis pathway. In C_4 plants photosynthetic reactions distributed to two different cell compartments, the mesophyll and bundle sheath cells (Hatch, 1987) with an exception of single cell C_4 species (Edwards and Voznesenskaya, 2011). Unlike bundle sheath cells of C_3 plants that are small and barely occupied with chloroplasts (Kinsman and Pyke, 1998; Leegood, 2007), bundle sheath cells of C_4 species are distinct with densely packed chloroplasts and mitochondria (Sage *et al.*, 2012; Lundgren *et al.*, 2014; Sage *et al.*, 2014). Ideally, each bundle sheath is directly in contact with mesophyll cells and clustered as a ring around the vascular tissue (Figure 3). This type of anatomy was termed as Kranz anatomy by Haberlandt (1904) after its appearance like a Kranz, a German word for the wreath. Moreover, bundle sheath and mesophyll cells of C_4 species are highly connected with plasmodesmata facilitating metabolite exchange between them (Botha, 1992; Danila *et al.*, 2016; Danila *et al.*, 2018).

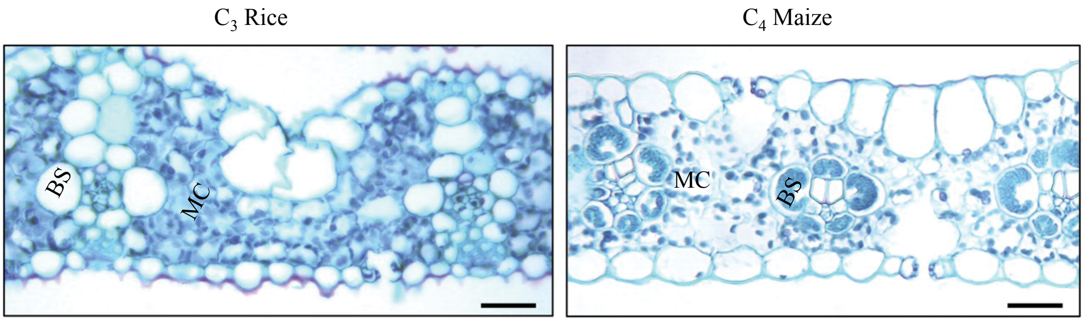


Figure 3: Cross sections of Rice and Maize leaves. Source: taken from Langdale, 2011. BS – bundle sheath cell; MC – mesophyll cell.

This modified anatomy is the basis of C_4 cycle. In all C_4 species, inorganic CO_2 is first converted into bicarbonate (HCO_3^-) by CARBONIC ANHYDRASE, located in the mesophyll cytosol. Then the oxygen-insensitive PHOSPHO*ENOL*PYRUVATE CARBOXYLASE uses HCO_3^- as a substrate and carboxylates 3-carbon compound phospho*enol*pyruvate (PEP) in the cytosol of mesophyll cells resulting in 4-carbon compound oxaloacetate (OAA). Since the 4-C compound is the first stable product, this pathway was named after it as a C_4 cycle (Hatch and Slack, 1966). OAA is converted either to malate or aspartate and the C_4 acids are then transported into bundle sheath cells in order to be decarboxylated further to pyruvate. The bundle sheath compartment refixes the released CO_2 with the help of Rubisco and thus helps to onset the Calvin cycle. This spatially separated carbon fixation and refixation enhance CO_2 pool in the vicinity of Rubisco, which is exclusively located in the bundle sheath cell compartment (Hatch, 1971; Hatch, 1987), and thereby drastically reduces photorespiration in C_4 species (Ku *et al.*, 1991).

C_4 species are classically divided into NADP-ME, NAD-ME and PEP-CK biochemical subtypes based on their dependency on the type of decarboxylation enzyme (Hatch, 1987; Furbank, 2011). In NADP-ME species, OAA is reduced to malate by NADPH-MALATE DEHYDROGENASE in the mesophyll chloroplast and subsequently transported into bundle sheath chloroplasts. The malate is then decarboxylated to pyruvate by NADP-MALIC ENZYME with pyruvate finally shuttling back to mesophyll cell to be converted to PEP by PYRUVATE, ORTHOPHOSPHATE DIKINASE (PPDK) (Figure 4). In NAD-ME and PEP-CK subtypes, aspartate mediates carbon flow from mesophyll to bundle sheath cells. In these species, OAA is converted to Asp by transamination catalyzed by ASPARTATE-AMINOTRANSFERASE in the mesophyll cytosol. However, in the case of NAD-ME species, Asp is reconverted to malate in the bundle sheath mitochondria by consecutive deamination and reduction reactions, which is then decarboxylated to pyruvate by NAD-MALIC ENZYME. Here, the released CO_2 diffuses into chloroplasts. Later, pyruvate is first converted to alanine by ALANINE-AMINOTRANSFERASE and transported to mesophyll cells where it is recycled back to pyruvate and subsequently to PEP. Whereas PEP-CK species, convert Asp back to OAA, which is then directly decarboxylated to PEP by PEP-CARBOXYKINASE in the bundle sheath cytosol with PEP diffusing back to mesophyll cells (Hatch, 1987; Furbank, 2011). In addition to this, several studies reported that some C_4 species can utilize a combination of different decarboxylases (Muhaidat *et al.*, 2007; Furbank, 2011; Wang *et al.*, 2014; Rao and Dixon, 2016). For example, a considerable amount of PEP-CK protein and its activity was reported in bundle sheath strands of C_4 maize (Wingler *et al.*, 1999), which actually is assigned to an NADP-ME subtype. However, C_4 species, which solely depends on PEP-carboxykinase, are yet to be confirmed. Many studies showed PEP-carboxykinase rather functions in addition to NADP-ME or NAD-ME (Furbank, 2011; Rao and Dixon, 2016).

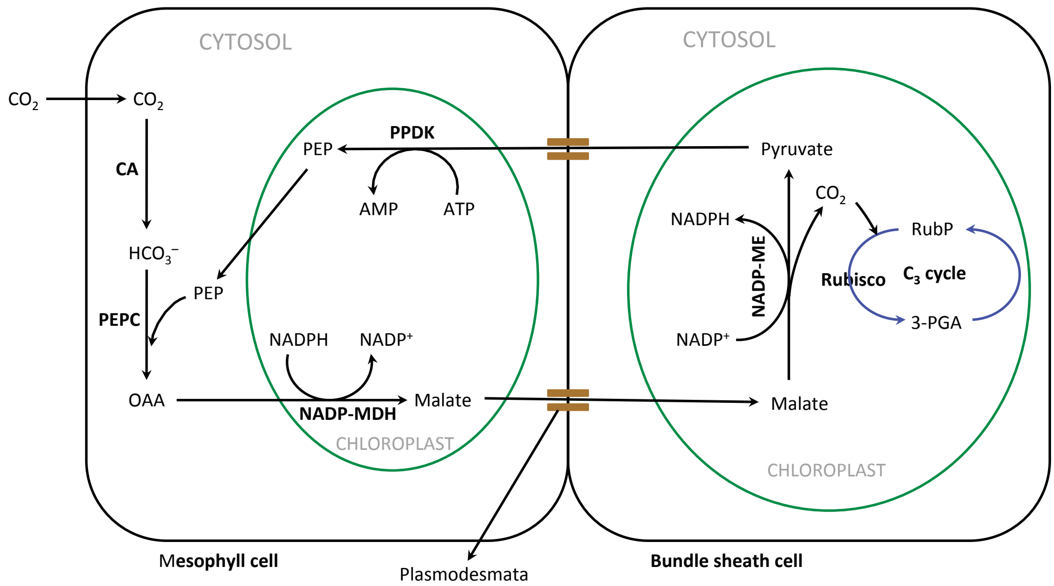


Figure 4. Reactions in NADP-ME subtype C_4 cycle.
CA – CARBONIC ANHYDRASE; HCO_3^- – bicarbonate; OAA – oxaloacetate; PEP – phospho*enol*pyruvate; NADP-MDH – NADP-MALATE DEHYDROGENASE; NADP-ME – NADP-MALIC ENZYME; PPDK – PYRUVATE, ORTHOPHOSPHATE DIKINASE.

Apart from reduced photorespiratory rates, two other benefits of C_4 CO_2 assimilation are enhanced nitrogen and water use efficiencies. As this mechanism increases the efficiency of Rubisco, C_4 plants show a reduction in total amount of Rubisco by 50-80 % in comparison to that required to assimilate equivalent amount of carbon in C_3 species (Sage and Zhu, 2011). In C_3 plants, 10-27 % total nitrogen is allocated for Rubisco against 5-9 % in case of C_4 (Sage and Pearcy, 1987). Further, C_4 species are advanced in keeping their stomata closed for longer duration in comparison to C_3 plants and therefore, produce more biomass for the same amount of water lost during transpiration (Ehleringer and Monson, 1993; Ghannoum and Evans, 2011). This indicates significant positive aspects of C_4 over C_3 plants in hot and dry climatic conditions.

3.2 Evolution of C_4 species

C_4 plant species evolved independently from existing C_3 ancestors, at least 61 times in angiosperm lineages (Sage *et al.*, 2011; Sage, 2016). Based on molecular phylogeny the estimated *de novo* origins in eudicots and monocots are 34 and 27, respectively (Sage, 2016). As of today, 8145 C_4 species are identified, which are distributed within 19 families. Further, the monocots contain 5044 species in grasses and 1322 species in sedges while in eudicots, 1777 C_4 species are reported (Sage, 2016).

The first evolved C_4 species appeared in the grass lineage Chloridoidae, approximately 30 million years (Mya) ago when earth's atmospheric CO_2 levels declined from ≈ 800 µmol mol^{-1} to nearly present levels (400 µmol mol^{-1}), in the Oligocene epoch (34-23 Mya). However, recently C_4 photosynthesis evolved in more than 20 lineages dating to the late-Miocene period (5-12 Mya), characterized by dry, arid and saline environmental conditions (Christin *et al.*, 2008; Besnard *et al.*, 2009; Sage *et al.*, 2011; Sage, 2016). This altogether confirms that a decrease in CO_2 levels in combination with other environmental constraints of the late-Miocene period led to the high photorespiratory rates in C_3 plants, ultimately leading to the evolution of C_4 photosynthesis in many angiosperm lineages (Ehleringer and Monson, 1993; Christin *et al.*, 2008; Sage, 2016).

3.3 Gradual evolution of C_4 photosynthesis

Based on proposed models, C_4 evolution from a C_3 ancestor proceed through two major intermediate stages, proto-Kranz and C_2-Kranz anatomies from preconditioning of a C_3 state to a completely established C_4 cycle (Figure 5; Sage, 2004; Gowik and Westhoff, 2011; Sage *et al.*, 2012; Sage *et al.*, 2014). In such models, preconditioning of C_3 ancestral state involves gene/genome duplications and increased vein density. Gene duplications and consequent small modifications reinforce the evolution of new gene functions while old copy still maintaining primitive function (Monson, 2003). The ever-increasing hot and arid climatic conditions not only increase the rates of photorespiration but also result in high water loss *via* transpiration and closure of stomata for preserving water is associated with the reduction in photosynthetic rates (Sage, 2001; Osborne and Sack, 2012), thus facilitating the evolution of C_4 species (Ehleringer *et al.*, 1997) as a mechanism of plant survival. To balance this trade-off between plant carbon and water relations, vein density was increased in closely related C_3 species (Sage, 2001). This is evidentially supported by detailed analysis of C_3 sister clades of many C_4 lineages in both eudicots (Muhaidat *et al.*, 2007) and grasses (Christin *et al.*, 2013; Griffiths *et al.*, 2013). For example, vein density of closely related C_3 *Cleome* members was comparable to the vein density of C_4 *Gynandropsis gynandra* species (Marshall *et al.*, 2007). However, this increased vein density is often associated with reduced photosynthetic efficiency per leaf area as a consequence of loss of mesophyll tissue (Gowik and Westhoff, 2011) in C_3 ancestor members. Triggering the need for compensation, it is thus assumed that in the initial stages, bundle sheath cells of C_2/C_4 closely related C_3 species show enlargement in size (Sage *et al.*, 2014). This phenomenon is frequently observed in eudicots *Flaveria* (Sage *et al.*, 2013), *Cleome* (Marshall *et al.*, 2007), *Heliotropium* (Muhaidat *et al.*, 2011) and *Euphorbia* (Sage *et al.*, 2011) genus and also in PACMAD clade of grasses (Christin *et al.*, 2013). Generally, bundle sheath cells of such C_3 species are comparatively more exposed to intercellular airspace with chloroplasts and mitochondria arranged along cell periphery similar to their arrangement in mesophyll cells. Such cellular organization enables bundle sheath cells to actively trap photorespiratory CO_2 escaping from mesophyll cells and re-assimilate the same into Calvin cycle (Sage *et al.*, 2014). Thus, the high evaporation and photorespiratory rates lead by hot and dry climate sets-up the stage for C_4 evolution in plants.

The next sequential step in C_4 evolution is the development of proto-Kranz anatomy, characterized by enlarged, activated bundle sheath cells with increased chloroplast and mitochondrial number and size. The most notable feature of this cellular state is the shift in localization of mitochondria from cell the periphery into the inner bundle sheath wall, against the vascular tissue (Muhaidat*et al.*, 2011; Sage *et al.*, 2013). Only a few chloroplasts shift to centripetal cell wall while most ordinate towards intercellular airspace. In this state, since GLYCINE DECARBOXYLASE is expressed in mitochondria, photorespiratory glycine produced in the outer chloroplasts should move to the inner mitochondrial space for further processing. However, diffusion barrier conferred by large vacuole restrains ease of CO_2 efflux. Therefore, the released CO_2 in the inner bundle sheath cells generates localized CO_2 pool enhancing Rubisco specificity in the adjacent chloroplast channels. Additionally, peripheral chloroplasts also trap photorespiratory CO_2 over-flown from mesophyll cells (Sage *et al.*, 2014). Hence, Proto-Kranz anatomy contributes to the net carbon assimilation in C_3 plants, albeit in lower amounts (Sage *et al.*, 2018), which further favor C_4 evolution. In conclusion, the increased vein density in association with activated bundle sheath cells supremely drives the complete establishment of the C_4 cycle (Christin *et al.*, 2013; Sage *et al.*, 2014; Bräutigam and Gowik, 2016).

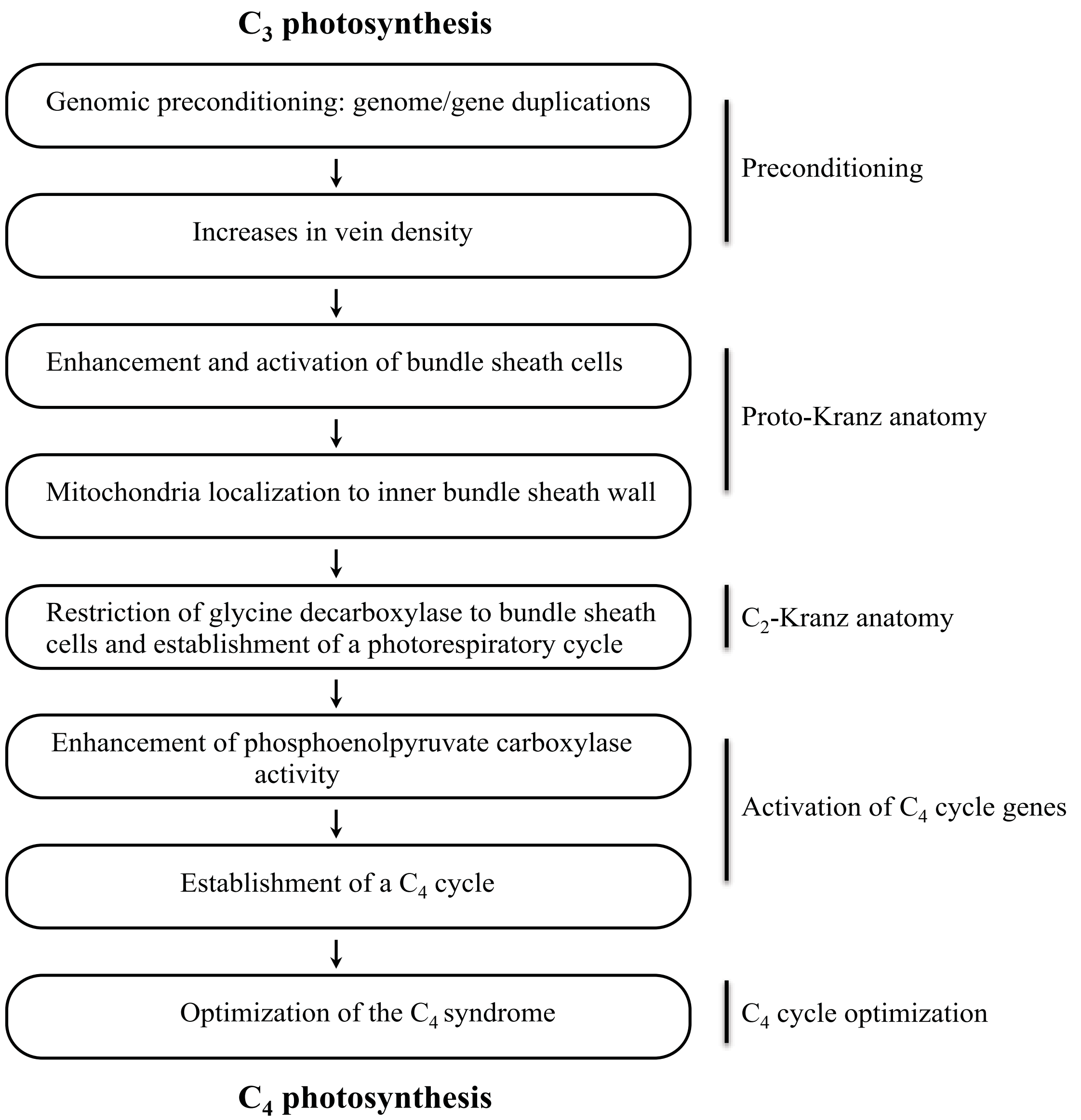


Figure 5. Step-wise evolution of C_4 photosynthesis from C_3 ancestor state.
Source: adapted from Gowik and Westhoff, 2011.

With previous steps setting the platform, the process of C_4 evolution head towards C_2-Kranz state, where GLYCINE DECARBOXYLASE (GDC) is seen restricted to the bundle sheath cells. This restriction of GDC to bundle sheath cells poses the need of transport of photorespiratory glycine from mesophyll cells in order to process it further. While in the completely established C_2-state, chloroplasts show layered organization right behind mitochondria in the majority of the species. This organization again facilitates rapid fixation of released CO_2 from glycine decarboxylation. Since glycine, which is a 2-carbon molecule shuttles between mesophyll and bundle sheath cells, this phase was termed as C_2-state. It is a common mechanistic feature in C_3-C_4 intermediate species and was first proposed by Monson *et al.*, (1984). Under high photorespiratory conditions, photorespiratory glycine shuttling between C_3-C_4 intermediates, raise bundle sheath CO_2 levels by 3-folds in comparison to mesophyll cell CO_2 levels (Keerberg *et al.*, 2014), thus enhancing photosynthetic efficiency by nearly 30 % (Sage *et al.*, 2018). These conditions altogether finally led to the successful establishment of a completely active C_4 cycle. Thus photorespiratory CO_2 concentration mechanism acts as an evolutionary bridge between C_3 and C_4 species (Mallmann *et al.*, 2014; Bräutigam and Gowik, 2016). Mallmann *et al.*, (2014) hypothesized that most of the C_4 cycle evolved as a side effect of photorespiratory ammonia recycling mechanism.

Proceeding further, in the final phase of C_4 development, mesophyll chloroplast number showed a decrease in C_4-like and C_4 species while no significant change was observed between C_3 and C_2 species. This implies that a reduction in mesophyll chloroplast number as a late evolutionary event (Stata *et al.*, 2014). While complete C_4 cycle can be established by compartmentalization and up-regulation of different C_4 enzymes (Sage *et al.*, 2012), C_4 cycle genes are regulated at different levels i. e. epigenetic, transcriptional, post-transcriptional, translational and post-translational levels (Reeves *et al.*, 2017), empowering their cell-specific expression. For example, mesophyll specificity of phospho*enol*pyruvate carboxylase A gene (*ppcA*) in C_4 *Flaveria trinervia* was achieved by minimum alterations in the 41 bp MEM1 element (mesophyll enhanced module 1) that is localized in the promoter region (Westhoff and Gowik, 2004), while bundle sheath cell specific expression of glycine decarboxylase P subunit gene (*GLDPA*) in C_4 *Flaveria* species was achieved by complex transcriptional and post-transcriptional regulatory mechanisms (Wiludda *et al.*, 2012). Finally changes in kinetic properties of C_4 enzymes for efficient operation in new mesophyll or bundle sheath cell environment led to optimized C_4 photosynthesis (Sage *et al.*, 2012).

4. Regulation of Kranz anatomy and its development

The above literature clearly evidences C_4 photosynthesis evolution as a highly coordinated, stepwise development of complex but truly efficient pathway equipped with modified leaf anatomy and advanced biochemistry. Following its discovery (Hatch and Slack, 1966), regulation of C_4 cycle genes is comparatively better understood (Reeves *et al.*, 2017). However, final establishment of the functional C_4 cycle can only be possible with C_4 specific Kranz anatomy. This emphasizes the great need of understanding the regulation of such an advanced leaf anatomy, prior to our ultimate aim of introducing C_4 cycle into C_3 crops.

Generally, leaf development initiates with the formation of leaf primordium from the periclinal division of L1, L2 (in monocots) or L1, L2 and L3 (in dicots) cell layers within shoot apical meristem. Briefly, the L1 layer forms leaf epidermis while periclinal division of L2 (in monocots) or L2 and L3 layers (in dicots) contribute to the formation of ground meristematic tissue. This tissue later differentiates into the vasculature, bundle sheath and mesophyll tissues (Langdale and Nelson, 1991). Vascular tissue differentiates from ground meristematic procambial cells while the initiation of procambium for all vein orders continues to follow the auxin efflux transporter PIN1 expression pattern (Scarpella *et al.*, 2006). Exclusive studies in eudicot C_3 and C_4 *Flaveria* species (McKown and Dengler, 2009), and on different C_3 and C_4 grasses (Ueno *et al.*, 2006) reported higher vein density in C_4 species resulting from increased density of higher vein orders. However, a recent investigation by Huang *et al.*, (2017), specifies the key role of elevated auxin biosynthesis and transport in the developing leaf correlating with increased vein density in C_4 *Gynandropsis gynandra*. The same study also showed up-regulation of auxin biosynthesis genes and higher auxin content in maize foliar primordia than in husk primordia. Interestingly, maize foliar leaves exhibit Kranz anatomy while husk leaves are characterized with C_3 leaf anatomy (Wang *et al.*, 2013). Two different forward genetic approaches in C_4 *Sorghum bicolor* (Rizal *et al.*, 2015) revealed the pivotal role of brassinosteroids in establishing C_4 vein pattern.

Both, in C_3 and C_4 species, procambial initiation precedes with the specification of bundle sheath and mesophyll cells (McKown and Dengler, 2009). However, a knowledge gap exists in understanding the bundle sheath ontogeny in dicots compared to grasses. In C_4 NADP-ME grasses (single-sheath), the vascular tissue is encircled by an immediate layer of bundle sheath cells whereas C_4 NAD-ME and C_3 grasses possess double-sheath structure along with a layer of mestome sheath separating bundle sheath from vascular tissue (Brown, 1975; Hattersley and Watson, 1976). Further, bundle sheath tissue in single-sheath grasses is derived from procambial initials whereas in double sheath grasses it differentiates from ground meristem (Dengler *et al.*, 1985; Bosabalidis *et al.*, 1994). The only available study in dicot C_3 and C_4 *Cleome* species reported only adaxial bundle sheath cells to be of procambial origin (Koteyeva *et al.*, 2014). However, Langdale and Nelson (1991) hypothesized, the position of bundle sheath and mesophyll cells playing a significant role in C_4 specific bundle sheath and mesophyll differentiation than cell lineage. However, they couldn't imagine the possible responsible signal. Based on the research in last few years, Fouracre *et al.* (2014) assumed it to be the movement of SHORT-ROOT (SHR) protein from the vein into bundle sheath cells as usually observed in roots. During root radial patterning, SHORT-ROOT protein glides from stele to adjacent cell layer, activating endodermal cell specification and SCARECROW (SCR) mediated cell divisions of cortex/endodermal initials (Nakajima *et al.*, 2001). Leaf bundle sheath is analogous to root endodermis and starch sheath of the hypocotyl. *SHR*/*SCR* regulatory mechanism is conserved between roots, shoot and leaves (Wysocka-Diller *et al.*, 2000; Lim *et al.*, 2005). In C_4 maize, loss of function *SCR* (*zmscr*) and *SHR* (*zmshr1*) genes is reported to reduce vein density, enhance unusual vein pattern and altered bundle sheath and mesophyll differentiation (Slewinski *et al.*, 2012; Slewinski *et al.*, 2014). This clearly suggests functional *SCR* and *SHR* genes as crucial requirements for functional development of Kranz anatomy.

In NADP-ME C_4 species, ultrastructure of the chloroplasts further differs between mesophyll and bundle sheath cells. Bundle sheath chloroplasts are agranal and exhibits reduced PSII activity (Woo *et al.*, 1970). In *Zea mays*, genes encoding GOLDEN2-LIKE transcription factors (*ZmGLK1* and *ZmGLK2* or *ZmG2*) are known to differentially regulate mesophyll and bundle sheath chloroplast development. Bundle sheath chloroplast development in *Zea mays* was aborted in *zmg2* mutant while no effect in mesophyll chloroplast development (Langdale and Kidner, 1994; Hall *et al.*, 1998). In C_3 plants, *GLK1* and *GLK2* genes are functionally redundant (Wang *et al.*, 2013). They activate the expression of the number of genes that are required for chlorophyll biosynthesis, light harvesting complex formation and electron transport chain (Waters *et al.*, 2009). Constitutive expression of *ZmGLK1* or *ZmGLK2* in rice, resulted in increased volume of both chloroplasts and mitochondria in bundle sheath cells and also in increased plasmodesmata connections between mesophyll and bundle sheath cells (Wang *et al.*, 2017).

High vein density and activated bundle sheath are key early evolutionary events that ultimately led to the development of C_4 Kranz anatomy (Sage *et al.*, 2012; Christin *et al.*, 2013). However, except above discussed few genes, knowledge on the molecular regulation of Kranz anatomy development is limited.

5. Way forward with comparative transcriptomics and forward genetics to reveal novel players of C_4 leaf anatomy

5.1 Comparative transcriptomics of developmental leaf gradients from closely related Flaveria robusta (C_3) and Flaveria bidentis (C_4)

The advancement in Next Generation Sequencing (NGS) technologies in the last few years, led to several comparative transcriptome studies between C_3 and C_4 leaves, providing useful insights into C_4 photosynthesis (Bräutigam *et al.*, 2011; Gowik *et al.*, 2011; Wang *et al.*, 2013; Külahoglu *et al.*, 2014; Kümpers *et al.*, 2017). For instance, comparative leaf transcriptomics of C_3, C_4 *Cleome* and *Flaveria* species revealed novel plastid localized sodium-pyruvate symporter (*BASS2*) being upregulated in respective C_4 species (Bräutigam *et al.*, 2011; Furumoto *et al.*, 2011; Gowik *et al.*, 2011) while subsequent function of *BASS2* was also experimentally proven (Furumoto *et al.*, 2011). Another interesting example is comparative transcriptome and anatomical study of developmental leaf gradients of C_3 and C_4 *Cleome* species (Külahoglu *et al.*, 2014) concluding delayed mesophyll cell differentiation during C_4 leaf development than the development in C_3 leaf. Further, they hypothesized such a delay could be related to the initiation of more higher vein orders. This agrees to the fact that the comparative transcriptomics is a powerful approach in identifying novel C_4 genes in order to provide more detailed insight of Kranz anatomy development. Additionally, the genus *Flaveria* is now widely recognized as an excellent model system for studying C_4 evolution in eudicots. This acceptance is due to that the genus *Flaveria* not only contains C_3 and C_4 species but also contain the C_4-like and large number of C_3-C_4 intermediate species (McKown *et al.*, 2005; McKown and Dengler, 2007; Lyu *et al.*, 2015). Moreover, it is also assumed to be the youngest C_4 lineage that might have evolved just around 5 Mya (Sage *et al.*, 2012). Therefore, comparative transcriptomes from developmental leaf gradients of closely related C_3 *F. robusta* and *F. bidentis* definitively would improve the current knowledge of Kranz anatomy regulation**.**

5.2. Forward genetic approaches: EMS and activation tagging to identify bundle sheath mutants using Arabidopsis thaliana

Although comparative transcriptomics (reverse genetics) is beneficial in dissecting novel genes, it is often biased and fails to detect low expressing genes. In this context, forward genetics is an unbiased and powerful tool to excavate the functional annotation of novel genes. The benefit with this advanced technique is that it starts with the selection of desired phenotypes by random mutagenesis of whole plant genome and then enquires about genotype, i. e. the gene responsible for the obtained phenotype. The chemical ethyl methanesulfonate (EMS) and insertional mutagens (T-DNA or transposable element) have been widely used in *Arabidopsis* to generate the large collection of wide variety of mutant phenotypes (Alonso and Ecker, 2006).

5.2.1. A. thaliana as a model species for plant geneticists

Forward genetics relies on well-characterized and easily accessible plants. In this context, *A. thaliana*, a weed belonging to the mustard family (Brassicaceae), is present day's most popular model plant system (Koornneef and Meinke, 2010). This plant offers a wide range of advantages making it extremely suitable for plant genetic investigations. The primary advantages are the small size and the short generation time (around 6 weeks), making *Arabidopsis,* the best substitute for generating large-scale genetic screens within limited space and time. Secondly, its high fertility allows production of up to 10,000 seeds per plant by self-fertilization. Additionally, this also manifests effortless management of mutant phenotypes and their out-crossing with other ecotypes (Meinke *et al.*, 1998) when required. Most importantly *A. thaliana* has the smallest genome (125 Mb) and is completely sequenced, reporting less repetitive DNA than any other plant species and is openly accessible (The Arabidopsis Genome Initiative, 2000; TAIR10). Furthermore, *Arabidopsis* can easily be subjected to ethyl methanesulfonate (EMS) chemical mutagenesis and T-DNA insertional mutagenesis (Meinke *et al.*, 1998). For this, mature seeds (M0 generation) are incubated with EMS solution and homozygous recessive mutants can be obtained in the M2 generation by self-fertilizing M1 plants. Later, T-DNA transfer into *A. thaliana* is achieved simply through *Agrobacterium tumefaciens* mediated floral dip method. This avoids pitfalls of regenerating transgenics by tissue culture. To sum up, small size, short generation time, high fecundity, small and less repetitive genome and the ease to generate mutant phenotypes make *A. thaliana* amenable to high-throughput forward genetic screens.

5.2.2. Versatility of A. thaliana (C_3) in identifying bundle sheath mutants

Bundle sheath surrounds vascular tissue as a single cell layer. In C_4 species, it contains distinct cells with loaded chloroplasts and hence, photosynthetically highly active while its exact role in C_3 plants is poorly understood (Leegood, 2008). Bundle sheath cells of *A. thaliana* are similar to other C_3 plants, being smaller in size with fewer chloroplasts. On average about 22 chloroplasts per bundle sheath cell whereas, this number is 76 per mesophyll cell (Kinsman and Pyke, 1998), thus contributing less to the net photosynthetic activity. Nonetheless, this basic structure of *A. thaliana* bundle sheath indicates the presence of an essential blueprint for achieving Kranz anatomy. Therefore, mutagenesis in *A. thaliana* would be helpful for broadening our knowledge about bundle sheath development (Westhoff and Gowik, 2010).

In this connection, analyses of 5' flanking sequences and/or *cis*-regulatory motifs of few genes already suggest partial conservation in gene regulatory network of C_4 angiosperms and C_3 *Arabidopsis*. The bundle sheath and vascular tissue specificity of the glycine decarboxylase P subunit (*GLDPA*) promoter from *Flaveria trinervia,* an Asteracean C_4 plant, retained its specificity when expressed in the Brassicacean C_3 species *Arabidopsis* (Engelmann *et al.*, 2008; Wiludda *et al.*, 2012). Similarly, sulfate transporter *SULTR2;2* promoter from *Arabidopsis* (Takahashi *et al.*, 2000) confined its bundle sheath specificity when expressed in C_4 *Flaveria bidentis* (Kirschner *et al.*, 2018). However, in contrast to this, mesophyll specificity of phospho*enol*pyruvate carboxylase A gene promoter (*p-ppcA*) from a C_4 *Flaveria trinervia* (Stockhaus et al. 1997) was lost upon expression in *Arabidopsis* and it is active in the whole leaf tissue (Akyildiz *et al.*, 2007). Interestingly, an elaborate recent study (Reyna-Llorens *et al.*, 2018) found the combined action of two bundle sheath specific motifs (BSM1a and BSM1b), responsible for bundle sheath specific accumulation of NAD-dependent malic enzyme isoforms (NAD-ME1 and NAD-ME2) in *Gynandropsis gynandra* (C_4). BSM1a and BSM1b motifs located within the coding region of *NAD-ME1* and *NAD-ME2* genes. Additionally, the presence of such motifs in C_3 *Arabidopsis* and their specific role in C_4 *G. gynandra* was also confirmed. Altogether, these investigations highlight the importance of *cis*-regulatory modules, cell-specific *trans*-acting transcription factors, and conserved gene regulatory mechanisms within C_4 angiosperms and C_3 *Arabidopsis* species. Knowing all these advantages regarding *A. thaliana*, preferentially makes it a suitable system for isolating candidate genes regulating bundle sheath specific cell development in plants.

5.2.3. Ethyl methanesulfonate (EMS) and activation tagging mutagenesis in A. thaliana

In *Arabidopsis*, mature seeds (M0 generation) are treated with EMS and EMS targets the diploid cells of the fully developed embryo. The segregation ratio of the mutant phenotype in the M2 generation greatly depends on genetically effective cell number (GECN) in the embryo that contributes to the germline. In case of *Arabidopsis,* this number is two (Koornneef, 2002). Hence, EMS induced recessive mutations segregate in 1:7 ratio. EMS alkylates Guanine (G) nucleotide, forming O^6-ethlyguanine which pairs with Thymine (T) instead of Cytosine (C) and through subsequent DNA replication G/C base pair can be substituted with A/T base pair. Thus, 99 % of EMS generated single base pair changes are G/C to A/T transitions (Greene *et al.*, 2003). These EMS induced single nucleotide polymorphisms (SNPs) result either in i) altering a particular amino acid codon into a stop codon (nonsense mutations) or ii) generating a codon that codes for a different amino acid (missense mutations) or iii) a codon that still codes for the same amino acid (silent mutations) (McCallum *et al.*, 2000), where missense mutations can be both conservative or non-conservative in nature. In the case of conservative substitutions, one amino acid is replaced with another having similar biochemical properties while, in case of non-conservative substitutions, with an amino acid of completely different properties. Further, nonsense and non-conservative missense mutations have deleterious effects and in most of the cases result in loss of gene function (Koornneef, 2002). However, based on *Arabidopsis* codon usage, nonsense and missense mutations are estimated to be 5 % to 65 % respectively (McCallum *et al.*, 2000). Nevertheless, EMS mutagenesis is extremely efficient and generates SNPs averaging 700 per genome in *Arabidopsis*. Therefore, less than 50,000 M1 mutagenized lines are enough to find a mutation in any given G/C pair with the probability of 95 % (Jander *et al.*, 2003). However, despite its high efficiency, identification of causative SNPs is equally tedious as it depends on whole genome sequencing of backcross or outcross population (James *et al.*, 2013). EMS mutagenesis also limits the functional annotation of gene families as loss of gene function can be compensated by another functional copy of the gene.

On the other hand, T-DNA insertional mutagenesis occurs through the transfer and random integration of *Agrobacterium tumefaciens* T-DNA into the plant genome. This method was initially used for creating gene knockout mutants (Van Lijsebettens *et al.*, 1991; Krysan *et al.*, 1999) and later modified to randomly activate the genes by a method called activation tagging. In activation tagging, T-DNA region harboring strong enhancer or promoter elements transferred into plant genome for overexpression or activation of gene/genes in the proximity of its integration site (Kondou *et al.*, 2010). This sometimes can also generate knockout mutants, if T-DNA lands within the gene sequence. However, random distribution of T-DNA tags in *A. thaliana* genome estimated two to three fold higher chances of T-DNA integration in the 500 bp interval region of 5’ and 3’ regulatory elements than in similar intervals within the genes (Szabados *et al.*, 2002). Activation tagging by using the cauliflower mosaic virus 35S enhancer elements was first described by Hayashi *et al.*, (1992). The purpose of that study was to find auxin overexpression mutants of *Nicotiana tabacum*. Later, this method has been extensively used in *Arabidopsis* (Weigel *et al.*, 2000; Aukerman and Sakai, 2003; Nakazawa *et al.*, 2003; Palatnik *et al.*, 2003). Although the efficiency of this method is low when compared to EMS mutagenesis, as the former requires at least 1,60,000 T-DNA tags to saturate *Arabidopsis* genome with 95 % probability (Szabados *et al.*, 2002), this method is helpful for creating both, the overexpression and knockout mutants. Of which, overexpression phenotypes are highly useful for functional annotation of gene families. Additionally, activation tagging is convenient in finding a gene sequence responsible for the phenotype by an inverse PCR or TAIL-PCR (Earp *et al.*, 1990; Liu and Whittier, 1995) method. Finally, the combination of methods, EMS and activation tagging is robust in order to identify genes responsible for bundle sheath structure in *A. thaliana*.

II. Objectives

C_4 photosynthesis operating plants exhibit higher photosynthetic, nitrogen and water-use efficiencies in comparison to those, only with C_3 photosynthetic pathway. This forms the fundamental basis of their adaptability and survival under adverse environmental constraints characterized by high light intensity, increased temperature and drought. Further, this higher adaptability is a result of the extraordinary evolutionary trait in plants leading to the development of C_4 specific Kranz leaf anatomy. However, genes involved in regulation of this anatomy are poorly understood. In order to fill these bottleneck knowledge gaps, we need to provide molecular insights of C_4 leaf anatomy development. Here, we aim to identify novel regulators of Kranz leaf anatomy development with two different approaches.

Approach 1. In order to broaden our understanding of C_4 leaf development and of differentially expressed genes during this pivotal developmental process, transcriptome datasets were prepared from leaf primordia along with shoot apical meristem and from nine different developmental stages of closely related C_3 *Flaveria robusta* and C_4 *Flaveria bidentis* species. Leaf anatomy of these developmental series was assessed and transcriptome dataset was further deconvoluted by applying non-negative matrix factorization in collaboration with T. J. Wrobel from the group of Prof. Andreas P. M. Weber at Heinrich Heine University. The results of this work are presented in Manuscript I of this study.

Approach 2. Here, we took the advantage of forward genetics using *A. thaliana* to find bundle sheath mutants and further identify causative genes. In our study, to easily detect alterations in bundle sheath structure, either *A. thaliana* reference lines whose bundle sheath plus vascular tissue were labeled with green fluorescent protein (*GFP*) or Luciferase reporter gene (*LUC*) were used as a genetic background. EMS mutagenesis was performed both on *LUC* and *GFP* reference line background with only *GFP* reference line subjected to activation tagging screen. The outcomes of this experiment are documented in Manuscript II + III of our study.

III. Summary

C_4 photosynthetic pathway is a coordinated evolution between development of Kranz leaf anatomy and division-of-labor i.e. spatial distribution of C_4 cycle reactions within mesophyll and bundle sheath cells. However, for establishment of efficient C_4 cycle, C_4 Kranz anatomy is a prerequisite, which in-turn regulates the functional operation of the former. Therefore, in the presented study, two different approaches were chosen for identifying novel gene regulatory mechanisms underpinning C_4 specific leaf development.

Transcriptome dataset and leaf anatomy of developmental leaf gradients from closely related C_3 *Flaveria robusta* and C_4 *Flaveria bidentis* was analyzed to distinguish the developmental progression of C_3 and C_4 leaves and to identify differentially expressed genes during C_4 leaf development. Non-negative matrix factorization of the data discerned four different transcriptome patterns within the developing leaf of both the species. As expected, genes involved in Calvin-Benson-Bessham and photorespiration pathways were upregulated in C_3 *F. robusta* while genes in the C_4 cycle and cyclic electron transport complex were upregulated in C_4 *F. bidentis*. Importantly, in both species, the expression of genes in these pathways peaks at a same developmental stage. Along with this, further, in both the species mesophyll and bundle sheath differentiation occurred at the same developmental stage as assessed *via* leaf anatomy. Strikingly, transcripts of auxin synthesis and auxin homeostasis related genes showed higher abundance in early C_4 leaf development compared to *F. robusta* that might be further related to increased vein density in C_4 *F. bidentis* and maybe one future aspect of the current study.

To identify bundle sheath mutants, EMS mutagenesis and activation tagging genetic screens were performed on *A. thaliana LUC* and/or *GFP* reference line background. For this, primarily, several thousands of plants were screened based on reporter gene signal intensity with the further selection of mutant lines with deviated reporter signal intensity. In the following step, leaf anatomy of identified mutant lines was analyzed with light microscopy and transmission electron microscopy. With this approach, we could identify five EMS mutant lines with altered bundle sheath plus vascular tissue structure. Interestingly, the identified activation tagging *bsom4* mutant line contains numerous plasmodesmata connections between all leaf tissue cells of *A. thaliana*. The unknown *BSOM4* gene (bundle sheath ontogeny and morphology 4; AT1G29480) was proven to be responsible for the observed phenotype and this gene was further characterized in detail.

IV. Literature

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

Alonso, J.M. and Ecker, J.R. (2006) Moving forward in reverse: Genetic technologies to enable genome-wide phenomic screens in *Arabidopsis*. *Nat. Rev. Genet.*, **7**, 524–536.

Anderson, L.E. (1971) Chloroplast and cytoplasmic enzymes II. Pea leaf triose phosphate isomerases. *Biochim. Biophys. Acta - Enzymol.*, **235**, 237–244.

Andrews, T.J., Lorimer, G.H. and Tolbert, N.E. (1973) Ribulose diphosphate oxygenase. I. Synthesis of phosphoglycolate by Fraction-1 protein of leaves. *Biochemistry*, **12**, 11–18.

Aukerman, M.J. and Sakai, H. (2003) Regulation of flowering time and floral organ identity by a MicroRNA and its *APETALA2*-like target genes. *Plant Cell*, **15**, 2730–2741.

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D.S. and Milo, R. (2011) The Moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, **50**, 4402–4410.

Bathellier, C., Tcherkez, G., Lorimer, G.H. and Farquhar, G.D. (2018) Rubisco is not really so bad. *Plant. Cell Environ.*, **41**, 705–716.

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

Benson, A.A. and Calvin, M. (1950) Carbon dioxide fixation by green plants. *Annu Rev. Plant. physiol.*, **1**, 25–42.

Besnard, G., Muasya, A.M., Russier, F., Roalson, E.H., Salamin, N. and Christin, P.-A. (2009) Phylogenomics of C_4 photosynthesis in Sedges (Cyperaceae): Multiple appearances and genetic convergence. *Mol. Biol. Evol.*, **26**, 1909–1919.

Borland, A.M., Barrera Zambrano, V.A., Ceusters, J. and Shorrock, K. (2011) The photosynthetic plasticity of crassulacean acid metabolism: an evolutionary innovation for sustainable productivity in a changing world. *New Phytol.*, **191**, 619–633.

Bosabalidis, A.M., Evert, R.F. and Russin, W.A. (1994) Ontogeny of the vascular bundles and contiguous tissues in the maize leaf blade. *Am. J. Bot.*, **81**, 745–752.

Botha, C.E.J. (1992) Plasmodesmatal distribution, structure and frequency in relation to assimilation in C_3 and C_4 grasses in southern Africa. *Planta*, **187**, 348–358.

Bowes, G., Ogren, W.L. and Hageman, R.H. (1971) Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochem. Biophys. Res. Commun.*, **45**, 716–722.

Bräutigam, A. and Gowik, U. (2016) Photorespiration connects C_3 and C_4 photosynthesis. *J. Exp. Bot.*, **67**, 2953–2962.

Bräutigam, A., Kajala, K., Wullenweber, J., et al. (2011) An mRNA blueprint for C_4 Photosynthesis derived from comparative transcriptomics of closely related C_3 and C_4 species. *Plant Physiol.*, **155**, 142–156.

Brooks, A. and Farquhar, G.D. (1985) Effect of temperature on the CO_2/O_2 specificity of ribulose-1,5-bisphosphate carboxylase/oxygenase and the rate of respiration in the light. *Planta*, **165**, 397–406.

Brown, W. V. (1975) Variations in anatomy, associations, and origins of Kranz tissue. *Am. J. Bot.*, **62**, 395–402.

Busch, F.A., Sage, R.F. and Farquhar, G.D. (2018) Plants increase CO_2 uptake by assimilating nitrogen via the photorespiratory pathway. *Nat. Plants*, **4**, 46–54.

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

Calvin, M. (1948) The path of carbon in photosynthesis. *Science.*, **107**, 476–480.

Campbell, W.J. and Ogren, W.L. (1990) Glyoxylate inhibition of ribulosebisphosphate carboxylase/oxygenase activation in intact, lysed, and reconstituted chloroplasts. *Photosynth. Res.*, **23**, 257–268.

Christin, P.-A., Besnard, G., Samaritani, E., Duvall, M.R., Hodkinson, T.R., Savolainen, V. and Salamin, N. (2008) Oligocene CO_2 decline promoted C_4 Photosynthesis in grasses. *Curr. Biol.*, **18**, 37–43.

Christin, P.-A. and Osborne, C.P. (2014) The evolutionary ecology of C_4 plants. *New Phytol.*, **204**, 765–781.

Christin, P.-A., Osborne, C.P., Chatelet, D.S., et al. (2013) Anatomical enablers and the evolution of C_4 photosynthesis in grasses. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 1381 – 1386.

Danila, F.R., Quick, W.P., White, R.G., Furbank, R.T. and Caemmerer, S. von (2016) The metabolite pathway between bundle sheath and mesophyll: Quantification of plasmodesmata in leaves of C_3 and C_4 monocots. *Plant Cell*, **28**, 1461–1471.

Danila, F.R., Quick, W.P., White, R.G., Kelly, S., Caemmerer, S. von and Furbank, R.T. (2018) Multiple mechanisms for enhanced plasmodesmata density in disparate subtypes of C_4 grasses. *J. Exp. Bot.*, **69**, 1135–1145.

Dengler, N.G., Dengler, R.E. and Hattersley, P.W. (1985) Differing ontogenetic origins of PCR (" Kranz ") sheaths in leaf blades of C_4 grasses (POACEAE). *Am. J. Bot.*, **72**, 284–302.

Earp, D.J., Lowe, B. and Baker, B. (1990) Amplification of genomic sequences flanking transposable elements in host and heterologous plants: a tool for transposon tagging and genome characterization. *Nucleic Acids Res.*, **18**, 3271–3179.

Edwards, G.E. and Voznesenskaya, E. V. (2011) C_4 photosnythesis: Kranz forms and single-cell C_4 in terrestrial plants. In Raghavendra, A. and Sage, R.F (eds), *C_4 photosynthesis and related CO_2 concentrating mechanisms*. Dordrecht, Springer Netherlands: **32**, 29-61.

Edwards, G.E. and Walker, D.A (1983) *C_3, C_4: mechanisms, and cellular and environmental regulation, of photosynthesis.* Oxford, UK: Blackwell Sci.

Ehleringer, J.R., Cerling, T.E. and Helliker, B.R. (1997) C_4 photosynthesis, atmospheric CO_2, and climate. *Oecologia*, **112**, 285–299.

Ehleringer, J.R. and Monson, R.K. (1993) Evolutionary and ecological aspects of photosynthetic pathway variation. *Annu. Rev. Ecol. Syst.*, **24**, 411–39.

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

Engelmann, S., Wiludda, C., Burscheidt, J., et al. (2008) The gene for the P-subunit of glycine decarboxylase from the C_4 species *Flaveria trinervia*: analysis of transcriptional control in transgenic *Flaveria bidentis* (C_4) and *Arabidopsis* (C_3). *Plant Physiol.*, **146**, 1773–1785.

Food and Agriculture Organization (FAO) (2017) *The future of food and agriculture: Trends and challenges*. Rome.

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

Foyer, C.H., Ruban, A. V and Nixon, P.J. (2017) Photosynthesis solutions to enhance productivity. *Philos. Trans. R. Soc.*, **372**, 20160374.

Furbank, R.T. (2011) Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types? *J. Exp. Bot.*, **62**, 3103–3108.

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

Galmés, J., Andralojc, P.J., Kapralov, M. V., Flexas, J., Keys, A.J., Molins, A., Parry, M.A.J. and Conesa, M.À. (2014) Environmentally driven evolution of Rubisco and improved photosynthesis and growth within the C_3 genus *Limonium* (Plumbaginaceae). *New Phytol.*, **203**, 989–999.

Galmes, J., Flexas, J., Keys, A.J., Cifre, J., Mitchell, R.A.C., Madgwick, P.J., Haslam, R.P., Medrano, H. and Parry, M.A.J. (2005) Rubisco specificity factor tends to be larger in plant species from drier habitats and in species with persistent leaves. *Plant, Cell Environ.*, **28**, 571–579.

Ghannoum, O. and Evans, J.R. (2011) Nitrogen and water use efficiency of C_4 plants. In Raghavendra, A., and Sage, R.F. (eds), *C_4 photosynthesis and related CO_2 concentration mechanisms*. Dordrecht, Springer Netherlands: **32**, 129–146.

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

Gowik, U. and Westhoff, P. (2011) The path from C_3 to C_4 photosynthesis. *Plant Physiol.*, **155**, 56–63.

Greene, E.A., Codomo, C.A., Taylor, N.E., et al. (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics*, **164**, 731–740.

Griffiths, H., Weller, G., Toy, L.F.M. and Dennis, R.J. (2013) You're so vein: Bundle sheath physiology, phylogeny and evolution in C_3 and C_4 plants. *Plant, Cell Environ.*, **36**, 249–261.

Haberlandt, G. (1904) *Physiologische Pflanzenanatomie*, Leipzig, Germany: Verlag von Wilhelm Engelmann.

Hall, L.N., Rossini, L., Cribb, L. and Langdale, J.A. (1998) GOLDEN 2: a novel transcriptional regulator of cellular differentiation in the maize leaf. *Plant Cell*, **10**, 925–36.

Hall, N.P. and Keys, A.J. (1983) Temperature dependence of the enzymic carboxylation and oxygenation of ribulose 1,5-bisphosphate in relation to effects of temperature on photosynthesis. *Plant Physiol.*, **72**, 945–8.

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

Hatch, M.D. (1971) The C_4-Pathway of photosynthesis. Evidance for an intermediate pool of carbon dioxide and the identity of the donor C_4-dicarboxylic acid. *Biochem. J*, **125**, 425–432.

Hatch, M.D. and Slack, C.R. (1966) Photosynthesis by sugar-cane leaves. A new carboxylation reaction and the pathway of sugar formation. *Biochem. J.*, **101**, 103–111.

Hattersley, P.W. and Watson, L. (1976) C_4 grasses: An anatomical criterion for distinguishing between NADP-Malic enzyme species and PCK or NAD-malic enzyme species. *Aust. J. Bot.*, **24**, 297–308.

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

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

Huang, C.-F., Yu, C.-P., Wu, Y.-H., Lu, M.-Y.J., Tu, S.-L., Wu, S.-H., Shiu, S.-H., Ku, M.S.B. and Li, W.-H. (2017) Elevated auxin biosynthesis and transport underlie high vein density in C_4 leaves. *Proc. Natl. Acad. Sci.*, **114**, E6884-E6891.

James, G.V., Patel, V., Nordström, K.J.V., Klasen, J.R., Salomé, P.A., Weigel, D. and Schneeberger, K. (2013) User guide for mapping-by-sequencing in *Arabidopsis*. *Genome Biol.*, **14**, R61.

Jander, G., Baerson, S.R., Hudak, J.A., Gonzalez, K.A., Gruys, K.J. and Last, R.L. (2003) Ethylmethanesulfonate saturation mutagenesis in *Arabidopsis* to determine frequency of herbicide resistance. *Plant Physiol.*, **131**, 139–146.

Jansson, C., Vogel, J., Hazen, S., Brutnell, T. and Mockler, T. (2018) Climate-smart crops with enhanced photosynthesis. *J. Exp. Bot.,* **69,** 3801-3809**.**

Jordan, D.B. and Ogren, W.L. (1984) The CO_2/O_2 specificity of ribulose 1,5-bisphosphate carboxylase/oxygenase. *Planta*, **161**, 308–313.

Kebeish, R., Niessen, M., Thiruveedhi, K., et al. (2007) Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in *Arabidopsis thaliana*. *Nat. Biotechnol.*, **25**, 593–599.

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

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

Kirschner, S., Woodfield, H., Prusko, K., Koczor, M., Gowik, U., Hibberd, J.M. and

Westhoff, P. (2018) Expression of *SULTR2;2* in the *Arabidopsis* bundle sheath is mediated by a highly conserved positive regulator. *bioRxiv*. http://dx.doi.org/10.1101/250738.

Kondou, Y., Higuchi, M. and Matsui, M. (2010) High-throughput characterization of plant gene functions by using gain-of-function technology. *Annu. Rev. Plant Biol.*, **61**, 373–393.

Koornneef, M. (2002) *Classical mutagenesis in higher plants*. In: *Molecular Plant Biology*/Gilmartin, P.M. and Bowler, C., **1**, 1-11. Oxford, GB: Oxford University Press.

Koornneef, M. and Meinke, D. (2010) The development of *Arabidopsis* as a model plant. *Plant J.*, **61**, 909–921.

Koteyeva, N.K., Voznesenskaya, E. V, Cousins, A.B. and Edwards, G.E. (2014) Differentiation of C_4 photosynthesis along a leaf developmental gradient in two *Cleome* species having different forms of Kranz anatomy. *J. Exp. Bot.*, **65**, 3525–3541.

Krysan, P.J., Young, J.C. and Sussman, M.R. (1999) T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell*, **11**, 2283–90.

Ku, M.S.B., Wu, J., Dai, Z., Scott, R.A., Chu2, C. and Edwards, G.E. (1991) Photosynthetic and photorespiratory characteristics of *Flaveria* Species. *Plant Physiol*, **96**, 518–528.

Külahoglu, C., Denton, A.K., Sommer, M., et al. (2014) Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae C_3 and C_4 plant species. *Plant Cell*, **26**, 3243–60.

Kümpers, B.M.C., Burgess, S.J., Reyna-llorens, I., Smith-unna, R., Boursnell, C. and Hibberd, J.M. (2017) Shared characteristics underpinning C_4 leaf maturation derived from analysis of multiple C_3 and C_4 species of *Flaveria*. *J. Exp. Bot.,* **68**, 177–189.

Laing, W.A., Ogren, W.L. and Hageman, R.H. (1974) Regulation of soybean net photosynthetic CO_2 fixation by the interaction of CO_2, O_2 and Ribulose 1,5-Diphosphate Carboxylase. *Plant Physiol.*, **54**, 678–685.

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

Langdale, J.A. and Kidner, C.A. (1994) bundle sheath defective, a mutation that disrupts cellular differentiation in maize leaves. *Development*, **120**, 673-681.

Langdale, J.A. and Nelson, T. (1991) Spatial regulation of photosynthetic development in C_4 plants. *Trends Genet.*, **7**, 191–196.

Leegood, R.C. (2008) Roles of the bundle sheath cells in leaves of C_3 plants. *J. Exp. Bot.*, **59**, 1663–1673.

Lijsebettens, M. Van, Boer, B. den, Hernalsteens, J.-P. and Montagu, M. Van (1991) Insertional mutagenesis in *Arabidopsis thaliana*. *Plant Sci.*, **80**, 27–37.

Lim, J., Jung, J.W., Lim, C.E., Lee, M.-H., Kim, B.J., Kim, M., Bruce, W.B. and Benfey, P.N. (2005) Conservation and diversification of *SCARECROW* in maize. *Plant Mol. Biol.*, **59**, 619–630.

Liu, Y.-G. and Whittier, R.F. (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, S.P., Marshall-Colon, A. and Zhu, X.-G. (2015) Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell*, **161**, 56–66.

Long, S.P., Zhu, X.-G., Naidu, S.L. and Ort, D.R. (2006) Can improvement in photosynthesis increase crop yields? *Plant, Cell Environ.*, **29**, 315–330.

Lundgren, M.R., Osborne, C.P. and Christin, P.-A. (2014) Deconstructing Kranz anatomy to understand C_4 evolution. *J. Exp. Bot.*, **65**, 3357–3369.

Lyu, M.-J.A., Gowik, U., Kelly, S., et al. (2015) RNA-Seq based phylogeny recapitulates previous phylogeny of the genus *Flaveria* (Asteraceae) with some modifications. *BMC Evol. Biol.*, **15**, 116.

Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M.J., Weber, A.P.M., Westhoff, P. and Gowik, U. (2014) The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*. *Elife*, **2014**, 1–23.

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

McCallum, C.M., Comai, L., Greene, E.A. and Henikoff, S. (2000) Targeted screening for induced mutations. *Nat. Biotechnol.*, **18**, 455–457.

McKown, A.D. and Dengler, N.G. (2007) Key innovations in the evolution of Kranz anatomy and C_4 vein pattern in *Flaveria* (Asteraceae). *Am. J. Bot.*, **94**, 382–399.

McKown, A.D. and Dengler, N.G. (2009) Shifts in leaf vein density through accelerated vein formation in C_4 *Flaveria* (Asteraceae). *Ann. Bot.*, **104**, 1085–1098.

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

Meinke, D.W., Cherry, J.M., Dean, C., Rounsley, S.D. and Koornneef, M. (1998) *Arabidopsis thaliana*: a model plant for genome analysis. *Science*, **282**, 662–682.

Monson, R.K. (2003) Gene duplication, neofunctionalization, and the evolution of C_4 photosynthesis. *Int. J. Plant Sci.*, **164**, S43–S54.

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

Muhaidat, R., Sage, R.F. and Dengler, N.G. (2007) Diversity of Kranz anatomy and biochemistry in C_4 eudicots. *Am. J. Bot.*, **94**, 362–381.

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

Nakajima, K., Sena, G., Nawy, T. and Benfey, P.N. (2001) Intercellular movement of the putative transcription factor SHR in root patterning. *Nature*, **413**, 307–311.

Nakazawa, M., Ichikawa, T., Ishikawa, A., Kobayashi, H., Tsuhara, Y., Kawashima, M., Suzuki, K., Muto, S. and Matsui, M. (2003) Activation tagging, a novel tool to dissect the functions of a gene family. *Plant J.*, **34**, 741–750.

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

Osborne, C.P. and Sack, L. (2012) Evolution of C_4 plants: a new hypothesis for an interaction of CO_2 and water relations mediated by plant hydraulics. *Philos. Trans. R. Soc. B.*, **367**, 583–600.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C. and Weigel, D. (2003) Control of leaf morphogenesis by microRNAs. *Nature*, **425**, 257–263.

Parry, M.A.J., Madgwick, P.J., Carvalho, J.F.C. and Andralojc, P.J. (2007) Prospects for increasing photosynthesis by overcoming the limitations of Rubisco. *J. Agric. Sci.*, **145**, 31–43.

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

Peterhansel, C. and Maurino, V.G. (2011) Photorespiration redesigned. *Plant Physiol.*, **155**, 49–55.

Raines, C.A. (2011) Increasing photosynthetic carbon assimilation in C_3 plants to improve crop yield: current and future strategies. *Plant Physiol.*, **155**, 36–42.

Rao, X. and Dixon, R.A. (2016) The differences between NAD-ME and NADP-ME subtypes of C_4 Photosynthesis: more than decarboxylating enzymes. *Front. Plant Sci.*, **7**, 1525.

Ray, D.K., Mueller, N.D., West, P.C. and Foley, J.A. (2013) Yield trends are insufficient to double global crop production by 2050. *PLoS One*, **8**, e66428.

Ray, D.K., Ramankutty, N., Mueller, N.D., West, P.C. and Foley, J.A. (2012) Recent patterns of crop yield growth and stagnation. *Nat. Commun.*, **3**, 1293.

Reeves, G., Grangé-Guermente, M.J. and Hibberd, J.M. (2017) Regulatory gateways for cell-specific gene expression in C_4 leaves with Kranz anatomy. *J. Exp. Bot.*, **68**, 107–116.

Reyna-Llorens, I., Burgess, S.J., Reeves, G., Singh, P., Stevenson, S.R., Williams, B.P., Stanley, S. and Hibberd, J.M. (2018) Ancient duons may underpin spatial patterning of gene expression in C_4 leaves. *Proc. Natl. Acad. Sci.*,**115,** 1931-1936.

Rizal, G., Thakur, V., Dionora, J., et al. (2015) Two forward genetic screens for vein density mutants in sorghum converge on a cytochrome P450 gene in the brassinosteroid pathway. *Plant J.*, **84**, 257–266.

Sage, R.F. (2016) A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *J. Exp. Bot.*, **67**, 4039–4056.

Sage, R.F. (2001) Environmental and evolutionary preconditions for the origin and diversification of the C_4 photosynthetic syndrome. *Plant Biol.*, **3**, 202–213.

Sage, R.F. (2013) Photorespiratory compensation: a driver for biological diversity. *Plant Biol.*, **15**, 624–638.

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

Sage, R.F. (1999) Why C_4 Photosynthesis?. In C_4 *Plant Biol.*, Sage R.F and Monson R.K. (eds). San Diego:Academic press: 3–16.

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

Sage, R.F., Khoshravesh, R. and Sage, T.L. (2014) From proto-Kranz to C_4 Kranz: building the bridge to C_4 photosynthesis. *J. Exp. Bot.*, **65**, 3341–3356.

Sage, R.F., Monson, R.K., Ehleringer, J.R., Adachi, S. and Pearcy, R.W. (2018) Some like it hot: the physiological ecology of C_4 plant evolution. *Oecologia*, **187,** 941-966.

Sage, R.F. and Pearcy, R.W. (1987) The Nitrogen use efficiency of C_3 and C_4 Plants. *Plant Physiol.*, **85**, 355–359.

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

Sage, R.F. and Stata, M. (2015) Photosynthetic diversity meets biodiversity: The C_4 plant example. *J. Plant Physiol.*, **172**, 104–119.

Sage, R.F. and Zhu, X.-G. (2011) Exploiting the engine of C_4 photosynthesis. *J. Exp. Bot.*, **62**, 2989–3000.

Sage, T.L., Busch, F.A., Johnson, D.C., et al. (2013) Initial events during the evolution of C_4 photosynthesis in C_3 species of *Flaveria*. *Plant Physiol.*, **163**, 1266–76.

Sage, T.L., Sage, R.F., Vogan, P.J., Rahman, B., Johnson, D.C., Oakley, J.C. and Heckel, M.A. (2011) The occurrence of C_2 photosynthesis in *Euphorbia* subgenus *Chamaesyce* (Euphorbiaceae). *J. Exp. Bot.*, **62**, 3183–3195.

Scarpella, E., Marcos, D., Friml, J. and Berleth, T. (2006) Control of leaf vascular patterning by polar auxin transport. *Genes Dev.*, **20**, 1015–1027.

Schwarte, S. and Bauwe, H. (2007) Identification of the photorespiratory 2-phosphoglycolate phosphatase, PGLP1, in *Arabidopsis*. *Plant Physiol.*, **144**, 1580–6.

Sharkey, T.D. (1988) Estimating the rate of photoresperation in leaves. *Physiol.Plant.*, **73**, 147–152.

Slewinski, T.L., Anderson, A.A., Price, S., Withee, J.R., Gallagher, K. and Turgeon, R. (2014) Short-Root1 plays a role in the development of vascular tissue and kranz anatomy in maize leaves. *Mol. Plant*, **7**, 1388–1392.

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

Stata, M., Sage, T.L., Rennie, T.D., Khoshravesh, R., Sultmanis, S., Khaikin, Y., Ludwig, M. and Sage, R.F. (2014) Mesophyll cells of C_4 plants have fewer chloroplasts than those of closely related C_3 plants. *Plant. Cell Environ.*, **37**, 2587–2600.

Stockhaus, J. (1997) The promoter of the gene encoding the C_4 form of phospho*enol*pyruvate carboxylase directs mesophyll-specific expression in transgenic C_4 *Flaveria* spp. *Plant Cell*, **9**, 479–489.

Szabados, L., Kovács, I., Oberschall, A., et al. (2002) Distribution of 1000 sequenced T-DNA tags in the *Arabidopsis* genome. *Plant J.*, **32**, 233–242.

Takahashi, H., Watanabe-Takahashi, A., Smith, F.W., Blake-Kalff, M., Hawkesford, M.J. and Saito, K. (2000) The roles of three functional sulphate transporters involved in uptake and translocation of sulphate in *Arabidopsis thaliana*. *Plant J.*, **23**, 171–182.

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

Walker, B.J., Vanloocke, A., Bernacchi, C.J. and Ort, D.R. (2016) The costs of photorespiration to food production now and in the future. *Annu. Rev. Plant Biol*, **67**, 107–129.

Wang, P., Fouracre, J., Kelly, S., et al. (2013) Evolution of *GOLDEN2-LIKE* gene function in C_3 and C_4 plants. *Planta*, **237**, 481-495.

Wang, P., Kelly, S., Fouracre, J.P. and Langdale, J.A. (2013) Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C_4 Kranz anatomy. *Plant J.*, **75**, 656–670.

Wang, P., Khoshravesh, R., Karki, S., et al. (2017) Re-creation of a key Step in the evolutionary switch from C_3 to C_4 leaf anatomy. *Curr. Biol.*, **27**, 3278–3287.

Wang, Y., Bräutigam, A., Weber, A.P.M. and Zhu, X.-G. (2014) Three distinct biochemical subtypes of C_4 photosynthesis? A modelling analysis. *J. Exp. Bot.*, **65**, 3567–3578.

Waters, M.T., Wang, P., Korkaric, M., Capper, R.G., Saunders, N.J. and Langdale, J.A. (2009) GLK transcription factors coordinate expression of the photosynthetic apparatus in *Arabidopsis*. *Plant Cell*, **21**, 1109–1128.

Weigel, D., Ahn, J.H., Blázquez, M.A., et al. (2000) Activation tagging in *Arabidopsis*. *Plant Physiol.*, **122**, 1003–1013.

Westhoff, P. and Gowik, U. (2004) Evolution of C_4 phospho*enol*pyruvate carboxylase. Genes and proteins: A case study with the genus *Flaveria*. *Ann. Bot.*, **93**, 13–23.

Westhoff, P. and Gowik, U. (2010) Evolution of C_4 photosynthesis–looking for the master switch. *Plant Physiol.*, **154**, 598–601.

Whitney, S.M., Baldet, P., Hudson, G.S. and Andrews, T.J. (2001) Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *Plant J.*, **26**, 535–547.

Whitney, S.M., Houtz, R.L. and Alonso, H. (2011) Advancing our understanding and capacity to engineer nature's CO_2-sequestering enzyme, Rubisco. *Plant Physiol.*, **155**, 27–35.

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

Wingler, A., Lea, P.J., Quick, W.P. and Leegood, R.C. (2000) Photorespiration: metabolic pathways and their role in stress protection. *Philos. Trans. R. Soc. Lond. B. Biol.*, **355**, 1517–1529.

Wingler, A., Walker, R.P., Chen, Z.-H. and Leegood, R.C. (1999) Phospho*enol*pyruvate carboxykinase is involved in the decarboxylation of aspartate in the bundle sheath of maize. *Plant Physiol.*, **120**, 539–545.

Woo, K.C., Anderson, J.M., Boardman, N.K., Downton, W.J., Osmond, C.B. and Thorne, S.W. (1970) Deficient Photosystem II in agranal bundle sheath chloroplasts of C_4 Plants. *Proc. Natl. Acad. Sci. U. S. A.*, **67**, 18–25.

Wysocka-Diller, J.W., Helariutta, Y., Fukaki, H., Malamy, J.E. and Benfey, P.N. (2000) Molecular analysis of SCARECROW function reveals a radial patterning mechanism common to root and shoot. *Development*, **127**, 595–603.

Zelitch, I., Schultes, N.P., Peterson, R.B., Brown, P. and Brutnell, T.P. (2009) High glycolate oxidase activity is required for survival of maize in normal air. *Plant Physiol.*, **149**, 195–204.

Zhu, X.-G., Long, S.P. and Ort, D.R. (2010) Improving photosynthetic efficiency for greater yield. *Annu. Rev. Plant Biol.*, **61**, 235–261.

V. Manuscripts

1. Kumari Billakurthi, Thomas J. Wrobel, Udo Gowik, Andrea Bräutigam, Andreas P.M. Weber and Peter Westhoff (2018). **Comparative transcriptomics from developmental leaf gradients of closely related C_3 and C_4 *Flaveria* species.** In preparation for publication (The Plant Cell).

2. Florian Döring, Kumari Billakurthi, Udo Gowik, Stefanie Sultmanis, Roxana Khoshravesh, Shipan Das Gupta, Tammy L. Sage and Peter Westhoff (2018). **Reporter-based forward genetic screen to identify bundle sheath anatomy mutants in *A. thaliana*.** Submitted to "The Plant Journal" for publication.

3. Kumari Billakurthi, Tammy L. Sage and Peter Westhoff. **Activation tagging in *Arabidopsis thaliana* identifies novel *BSOM4* gene as a player in plasmodesmata development.** Unpublished work.

Manuscript I

Comparative transcriptomics from developmental leaf gradients of C_3 and C_4 *Flaveria* species

Comparative transcriptomics from developmental leaf gradients of C_3 and C_4 *Flaveria* species

Kumari Billakurthi[1,5±], Thomas J. Wrobel[2,5±], Udo Gowik[3], Andrea Bräutigam[4], Andreas P.M. Weber[2,5*] and Peter Westhoff[1,5*]

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

[2]Institute of Plant Biochemistry, Universitaetsstrasse 1, Heinrich-Heine-University, 40225 Duesseldorf, Germany

[3]Department of Biology and Environmental Sciences, Ammerlaender Heerstrasse 114, Carl Von Ossietzky University, 26129 Oldenburg, Germany.

[4]Faculty of Biology and Computational Sciences, Bielefeld University, 33501 Bielefeld, Germany.

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

*To whom correspondence should be addressed

± These authors contributed equally to this work

Kumari Billakurthi

e-mail: kumari.billakurthi@hhu.de

Thomas J. Wrobel

e-mail: Thomas.jan.wrobel@hhu.de

Udo Gowik

e-mail: udo.gowik@uni-oldenburg.de

Andrea Bräutigam

e-mail: andrea.braeutigam@uni-bielefeld.de

Andreas P.M. Weber

e-mail: andreas.weber@uni-duesseldorf.de

Peter Westhoff

e-mail: west@hhu.de

(I) Abstract

C_4 species have evolved more than 60 times independently from C_3 ancestors. This multiple and parallel evolution of the complex C_4 syndrome indicates common underlying evolutionary mechanisms that might be identified by studying closely related C_3 and C_4 species. To function efficiently C_4 plants exhibit a certain leaf anatomy that is characterized by enlarged, chloroplast rich bundle sheath cells and narrow vein spacing, termed Kranz anatomy. To elucidate molecular mechanisms generating Kranz anatomy, we analyzed developmental series of leaves from the C_4 plant *Flaveria bidentis* and the closely related C_3 species *Flaveria robusta* utilizing leaf clearing and whole transcriptome sequencing. Applying non-negative matrix factorization on the data allowed us to define four different zones with distinct transcriptome patterns in growing leaves of both species. Comparing these transcriptome patterns points towards an important role of auxin metabolism and especially auxin homeostasis for establishing the high vein density typical for C_4 leaves.

(II) Introduction

In C_3 plants, the carbon assimilation process starts with the carboxylation of 5-carbon compound Ribulose-1,5-bisphosphate (RubP) by the carboxylase activity of Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), resulting in two 3-carbon molecules of 3-phosphoglycerate (3-PGA) (Calvin, 1948; Benson and Calvin, 1950). Additionally, Rubisco can catalyze the oxygenation of RubP yielding one 3-PGA and one molecule of 2-phosphoglycolate (2-PG). 2-PG and its derivatives glycolate and glyoxylate are toxic to plants upon accumulation (Anderson, 1971; Campbell and Ogren, 1990). Therefore 2-PG is recycled to 3-PGA in a process called photorespiration that releases CO_2 as well as ammonia and consumes ATP and reducing equivalents (Bauwe *et al.*, 2010; Peterhansel *et al.*, 2010). Under current ambient CO_2 concentrations (350 ppm) at 25°C, photorespiration is estimated to decrease the yield of soybean or wheat in the US by 36 % and 20 %, respectively (Walker *et al.*, 2016). Environmental constraints such as high temperatures and drought further increases Rubisco oxygenase activity (Laing *et al.*, 1974; Jordan and Ogren, 1984; Brooks and Farquhar, 1985; Parry *et al.*, 2007). C_4 plants bypass this dilemma by increasing CO_2 concentrations up to 15 fold in the vicinity of Rubisco and thereby suppressing the oxygenation reaction of Rubisco and supercharging photosynthesis (Sage *et al.*, 2012). In most C_4 plants CO_2 fixation is compartmentalized between two cell types - the bundle sheath and the mesophyll cells. In the mesophyll phospho*enol*pyruvate (PEP) is carboxylated by phospho*enol*pyruvate carboxylase (PEPC) resulting in the 4-carbon compound oxaloacetate (OAA). OAA is then converted to either malate or aspartate, which is transferred to the bundle sheath. Here the 4-carbon compounds are decarboxylated and the released CO_2 is routed to Calvin-Benson-Bassham cycle (CBB). The resulting pyruvate is transferred back to the mesophyll where the primary CO_2 acceptor PEP is regenerated by pyruvate orthophosphate dikinase (PPDK) (Hatch, 1987).

One striking difference between C_3 and C_4 plants is the overall leaf anatomy. As bundle sheath and mesophyll cells operate as one photosynthetic unit, direct contact of both cell types is necessary to ensure efficient photosynthesis. The bundle sheath is composed of the cells directly adjacent to the vasculature. Therefore, leaves of most C_4 species exhibit high vein densities with a characteristic pattern in which two veins, each surrounded by bundle sheath cells, are separated by only two layers of mesophyll cells in a vein-bundle sheath-mesophyll-mesophyll-bundle sheath-vein layout. Bundle sheath cells of C_4 plants are often larger compared to C_3 species and contain more chloroplasts. This characteristic tissue pattern is also called Kranz anatomy (Haberlandt, 1904).

C_4 plants evolved more than 60 times independently from C_3 progenitors (Sage *et al.*, 2011; Sage, 2016). It is widely accepted that the changes in leaf anatomy, leading to higher vein density and enlarged, chloroplast rich bundle sheath cells, belong to the earliest steps of C_4 evolution (Sage *et al.*, 2012; Christin *et al.*, 2013; Lundgren *et al.*, 2014; Christin and Osborne, 2014; Bräutigam and Gowik, 2016). In the last years, several studies set out to determine the factors governing the different developmental programs in C_3 and C_4 leaf development using next-generation sequencing of transcriptomes. They either analyzed whole leaves of sequential stages covering the growth process (Külahoglu *et al.*, 2014; Wang, Kelly, *et al.*, 2013) or slices from the leaves which were assumed to contain all stages of development (Aubry *et al.*, 2014; Kümpers *et al.*, 2017). Combined with the in-depth analysis of several candidate genes, these approaches provided some insights according to the changes of leaf development during C_4 evolution.

It became obvious that the leaf bundle sheath is equivalent to the root endodermis and the starch sheath of hypocotyls (Wysocka-Diller *et al.*, 2000; Lim *et al.*, 2005; Slewinski, 2013). The mechanisms controlling the development and differentiation of root endodermis, namely the *SHORT-ROOT*/*SCARECROW* regulatory system, are at least partially conserved in bundle sheath development and differentiation (Fouracre *et al.*, 2014). It could be shown that *scarecrow* and *short-root* mutants of maize, as well as *Arabidopsis thaliana* exhibit a distorted bundle sheath development (Slewinski *et al.*, 2012; Slewinski *et al.*, 2014; Cui *et al.*, 2014). The overexpression of the maize *SCARECROW* gene (*ZmSCR1*) in Kitaake rice, on the other hand, did not lead to remarkable changes in leaf anatomy (Wang, *et al.*, 2017a).

Several transcription factors related to chloroplast development were identified e.g. the GOLDEN2-LIKE (GLK) proteins (Hall *et al.*, 1998; Fitter *et al.*, 2002; Waters *et al.*, 2009; Wang *et al.*, 2013) or B-class GATA transcription factors (Chiang *et al.*, 2012; Hudson *et al.*, 2013) mainly by their mutant phenotypes although they also showed up with several transcriptome centric approaches. Constitutive expression of one of the maize *GOLDEN2-LIKE* genes in rice induced chloroplast and mitochondrial development in bundle sheath cells. The organelles increased in size and photosynthetic enzymes were induced mimicking the situation in C_4 leaves (Wang *et al.*, 2017b). So far, these *GLK* genes are the only known genes that can shift one of the signature anatomical traits towards the C_4 state when solely overexpressed in a C_3 leaf.

In leaves, polar auxin transport is central to the development of a functional vascular network and controls its structure and density. According to the widely accepted auxin canalization model (Sachs, 1969; Mitchison, 1981; Rolland-Lagan and Prusinkiewicz, 2005), vein differentiation is initiated by elevated auxin concentration due to polar auxin transport along strands of undifferentiated ground meristematic cells and is followed by the differentiation of photosynthetically active cells (Sud and Dengler, 2000; Scarpella *et al.*, 2004; McKown and Dengler, 2010). Külahoglu *et al.*, (2014) observed that the differentiation of mesophyll cells is delayed in the leaves of the C_4 species *Gynandropsis gynandra* compared to that of the closely related C_3 species *Tarenaya hassleriana* prompting the hypothesis that more veins can be initiated in the C_4 leaf due to this prolonged time slot.

In the present study we analyzed the transcriptomes of series of developing leaves from the closely related C_4 and C_3 species *Flaveria bidentis* and *Flaveria robusta*. McKown and Dengler (2009) analyzed leaf development in these two species in great detail and could show that increased vein density in the C_4 species is caused by the formation of an additional minor vein order. In contrast to Kümpers *et al.*, (2017) who followed leaf development in C_3 and C_4 *Flaveria* species by dissecting leaves of one stage containing different developmental areas, we used a series of complete leaves of different ages. Transformation of the resulting expression data from nine different leaf stages by applying nonnegative matrix factorization (NMF), revealed up to four highly diverse transcriptional patterns within the leaves of both species. Clearing of the leaves and determination of vein density allowed us to precisely correlate the different transcriptional programs to different developmental zones in the leaves.

(III) Materials and Methods

Leaf material and RNA isolation

Flaveria bidentis and *Flaveria robusta* plants were grown from mid-October to mid-December in rooftop greenhouses with at 16 h additional light per day, photon flux density (PFD) of ~300 µmol m^{-2} s^{-1} and at 24 °C (during the day) and 21 °C (night). Leaf primordia along with shoot apical meristem (stage 0) and 9 leaves starting from first visible leaf pair to fully developed leaf pairs were harvested during noon and immediately frozen in liquid nitrogen. For each replicate leaf material was pooled from five plants and primordia (or shoot apices) were pooled from 25 plants. Approximate length of the different leaf stages was 0.3, 0.5, 1.5, 2, 3, 4, 6, 8 and 10 cm. Total RNA was isolated using the RNeasy Plant Mini Kit (QIAGEN) and DNase digestion was performed with RNase-Free DNase Set (QIAGEN). RNA Integrity Number (RIN), and quantity were determined with the 2100 Bioanalyzer (Agilent Technologies).

Library preparation, sequencing, and mapping

Libraries were prepared with the TruSeq RNA Library Prep Kit v2 (Illumina) using 0.5 µg of total RNA as a starting material. Quality and quantity of the libraries were determined with the 2100 Bioanalyzer and the Qubit (Thermo Fisher Scientific), respectively. Single end reads of 100 bp length were sequenced with the HiSeq2000 Illumina platform and samples were multiplexed with 6 libraries per lane. On an average ~ 33 million reads were obtained per sample. Reads were mapped to a minimal set of coding sequences of the *Arabidopsis thaliana* (http://www.Arabidopsis.org/), as described by Bräutigam *et al.*, (2011). Read count was normalized to reads per kilobase of a transcript, per million mapped reads (RPKM).

RNAseq analysis and factorization

Unless stated otherwise all data analysis was performed with the R statistical package (version 3.3.3 "Another Canoe" www.r-project.org). The transcriptome was filtered by expression and all genes with an expression above ten RPKM in at least one sample across both species were considered for further analysis. The transcriptional investment was calculated based on reduced Mapman categories (Külahoglu *et al.*, 2014; Brilhaus *et al.*, 2016). PCA analysis was performed on Z-scored expression values. We factorized our data using the Lee and Suad, the Brunet, the Kullback-Leibler and the non-smooth NMF algorithms provided within the NMF package for R (Gaujoux and Seoighe, 2010). We analyzed the algorithm performance and an optimal number of areas by repeating factorization 50 times for two to ten areas (Supplementary Figure 2). To improve calculation speed factorization was performed on the transpose of the Gene-by-Sample matrix. The coefficient matrix was restricted to represent tissue structure by dividing each value in a column with its sum during every iteration. The final factorizations were performed 500 times using the Kullback-Leibler algorithm for five areas, when the primordia were included and for four areas for the dataset without primordia. We clustered the factorized data using the K-means algorithm described by Hartigan and Wong (1979) after Z-scoring by genes across both the species. A suitable number of clusters was determined using the elbow method and clustering was performed 2000 times. GO enrichment was determined using the TopGo package in R (Alexa and Rahnenführer, 2018). GO annotations were taken from TAIR10 and enrichment was detected via the classic Fisher algorithm. All generated p-values were corrected for multiple testing using Benjamini-Hochberg correction (Benjamini and Yosef, 1995).

The comparative PCA (Figure 3B) was performed on the data provided in the supplementary dataset of Kümpers *et al.,* (2017). Genes were filtered for AGIs present in both studies and by gene normalization was performed separately for the Kümpers dataset, the initial dataset, and the factorized dataset.

Araldite embedding

For Araldite embedding we cut 1 x 2 mm sized segments from the second quarter of the leaves. The samples were transferred into 4 % paraformaldehyde solution until the majority of the segments dropped to the bottom of the tube or for at least 24 hours. The material was transferred into 0.1 % v/v glutaraldehyde in 1x PBS, incubated for 20 min at room temperature and vacuum infiltrated three times for five minutes each. Subsequently, an ethanol dilution series from 30 % to 90 % ethanol, with an increment of 10 % for each step was applied. Each step was carried out for 20 minutes and the final dehydration steps were carried out in 96 % ethanol (two times), 100 % ethanol (two times) and 100 % acetone (two times) for one hour. To remove residual water the 100 % ethanol and acetone solutions were dried using zeolite capsules. Finally, the samples were then transferred into open vessels and overlayered with araldite. To ensure the acetone evaporation the samples were incubated for four hours at room temperature under a fume hood. The samples were transferred into fresh araldite in a silicon mold and incubated overnight. The final polymerization was performed for 24-48 hours at 65 °C. Samples were cut with a microtome stained in bromophenol blue and analyzed at 20x magnification.

Leaf clearing and determination of the vascular structure

Leaf clearing was performed as described by Hasegawa *et al.* (2016). Leaves of a suitable size were cleared in 3:1 ethanol-acetic acid (v/v) until no green color was visible and transferred to 97 % 2,2'-thiodiethanol solution containing 0.0025 % (w/v) propyl gallate in PBS for 20 minutes. All pictures were taken at four times magnification with an estimated overlap of 30 to 50 % and stitched in Adobe Photoshop (Version 2014.0.0) using the build in merge function. The vascular density estimation was performed in 200 µm steps for leaves up to 4 cm in length and in 500 µm steps for larger leaves.

Auxin treatment

The effect of auxin on leaf structure and vascular density of *F. robusta* and *F. bidentis* was tested by spraying with NAA and 2,4-D. *Flaveria* seeds were surface sterilized, germinated for two weeks on half strength MS medium with 0.8 % (w/v) agar in a 16 h light/ 8 h dark cycle (22/18 °C) and transferred to soil. One week after the transfer, plants were sprayed for three weeks on a daily basis with either 1-Naphthaleneacetic acid (NAA, Sigma Aldrich) or 2,4-Dichlorophenoxyacetic acid (2,4-D, PESTANAL Sigma Aldrich). Both phytohormones were dissolved in DMSO and 0.1 % v/v DMSO was used as negative control.

(IV) Results

Transcriptomes of developmental leaf gradients from C_3 and C_4 *Flaveria* species

We performed RNA-seq on leaf developmental gradients from the C_4 species *F. bidentis* and the C_3 species *F. robusta*. The leaves analyzed span the whole range of development from meristematic/primordial tissues (stage 0) to fully expanded leaves of about 10 cm in length (stage 9) (Figure 1A). RNA-seq analysis was performed in triplicates and generated on average 33 million reads per sample. Reads were mapped to a minimal set of coding sequences of the *Arabidopsis thaliana* transcriptome (http://www. Arabidopsis.org/), as described by Bräutigam *et al.*, (2011) with an average mapping efficiency of 53 %

(Supplemental Table 1). We were able to detect transcripts of 17504 genes out of which 10864 were expressed with at least ten RPKM (reads per kilobase of a transcript, per million reads) in at least one sample and used for further analysis.

Principal component analysis (PCA) of the dataset demonstrates high consistency between replicates (Figure 1B). The different leaf samples mainly vary by developmental stage of the leaves shown by a separation in principal component 1 explaining 45 % of the total variation (Figure 1B). Species-specific variations including those related to the different types of photosynthesis are mainly represented by principal component 2 that explains 17 % of the total variation (Figure 1B).

Transcriptional investment, according to functional categories inferred from MapMan bins (Usadel *et al.*, 2009) is largely in accordance with the PCA. Leaves from a comparable developmental stage showed higher similarity between the different species than young and mature leaves of the same species (Figure 1C). The progression through leaf development is quite comparable for both species. While transcripts related to protein and DNA synthesis occupy a predominant role in young leaves, photosynthesis becomes more and more important with progressing leaf age (Figure 1C). *Flaveria bidentis* exhibits a higher transcript investment into DNA synthesis in young tissues and into carbon metabolism in older leaves, while *F. robusta* has a slightly higher investment into photosynthesis.

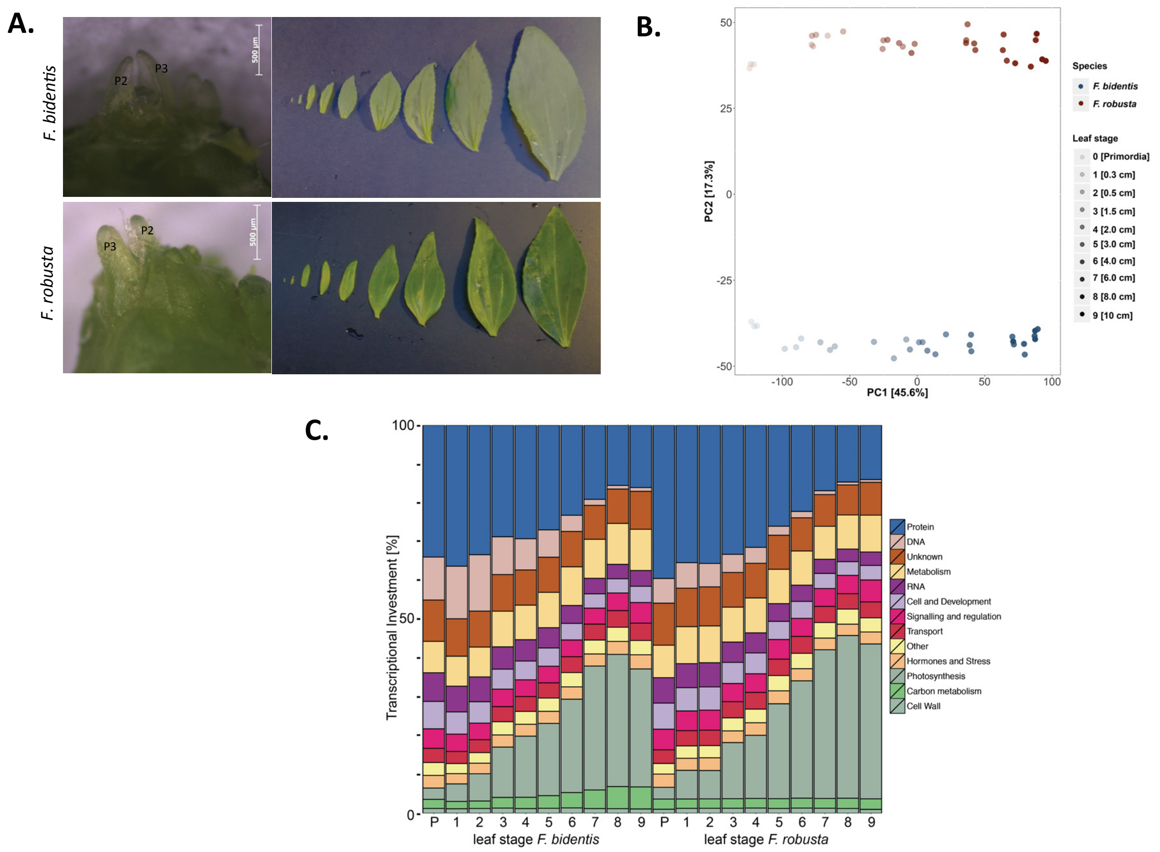


Figure 1. Leaf developmental transcriptomics of both C_3 and C_4 *Flaveria* species.
A. Sampled leaf developmental series from *F. bidentis* and *F. robusta.* P2, P2 - leaf primordia. Scale bar – 500µm. **B.** PCA of the developmental gradient for all leaf stages sequenced in this study. **C.** Transcriptional investment plot of the developmental gradient. The relative investment into different gene categories was calculated based on RPKM values.

Changes of leaf anatomy and venation patterns in the two *Flaveria* species during development

To follow the actual developmental progression of leaves from both species we performed cross sections within the upper quarter of the respective leaf for stages two to nine (Supplemental Figure 1). Both species differentiate at a similar pace. While in stage two, leaves of both species mainly consist out of undifferentiated ground tissue, in stages three and four the first differentiating veins can be observed. A completely developed bundle sheath and palisade parenchyma tissues are present at leaf stage five (3 cm, Supplemental Figure 1) in both species. Leaves of the C_4 species *F. bidentis* are thinner, with fewer cell layers in all stages and stop further vertical expansion earlier than their C_3 counterpart, between stages six (4 cm) and seven (6 cm).

In order to gain the better understanding of the developing vein system, all leaf stages were cleared using TOMEI-I (Hasegawa *et al.*, 2016) and photographed with four times magnification using a light microscope. The microscope pictures were stitched together and present an overview of the process of vascularization in Figure 2A. We quantified the vascular density along the leaf lamina from the base to the tip in a stepwise manner for leaf stages one to nine and calculated the vein density as a function of the relative leaf area. We could observe three clearly distinct areas that were not necessarily present in all different stages. Figure 2B shows an example of stage six leaves (4 cm in size) from both species that contain all different areas. The proximal leaf areas that are characterized mostly by meristematic cells exhibit a very low vein density in both species (Figure 2B) and represent the cell division zone of the leaves. This zone with dividing cells is followed by an area with a marked increase in vascular density indicating that the majority of veins differentiate in this area. The vein density in this area is higher in the C_4 species (19.8 mm/cm^2) compared to the C_3 species (8.5 mm/cm^2) (Figure 2B). Towards the tips of the older leaves from both species vein densities decreases to 11.2 mm/cm^2 in *F. bidentis* and 5.7 mm/cm^2 in *F. robusta* due to cell expansion (Figure 2B).

These characteristic vein densities were used to estimate the relative areas associated with cell division, cell/vein differentiation and mature, photosynthetic tissues. To this end differentiating tissues were considered to have vascular densities above 15 mm/cm^2 in *F. bidentis* and 7 mm/cm*2* in *F. robusta*. The area at the base of the leaves with lower vein density was categorized as consisting mainly out of dividing cells, whereas the distal area at the tips of the leaves was considered to consist of mature and photosynthetic tissues. Based on this area estimates, cell division rapidly decreases in the first leaf stages reaching 0 % of the leaf area in stage seven (6 cm). The majority of differentiation takes place between stage three (1.5 cm) and seven (6 cm), peaking at stage five (3 cm) in both species (Figure 2C).

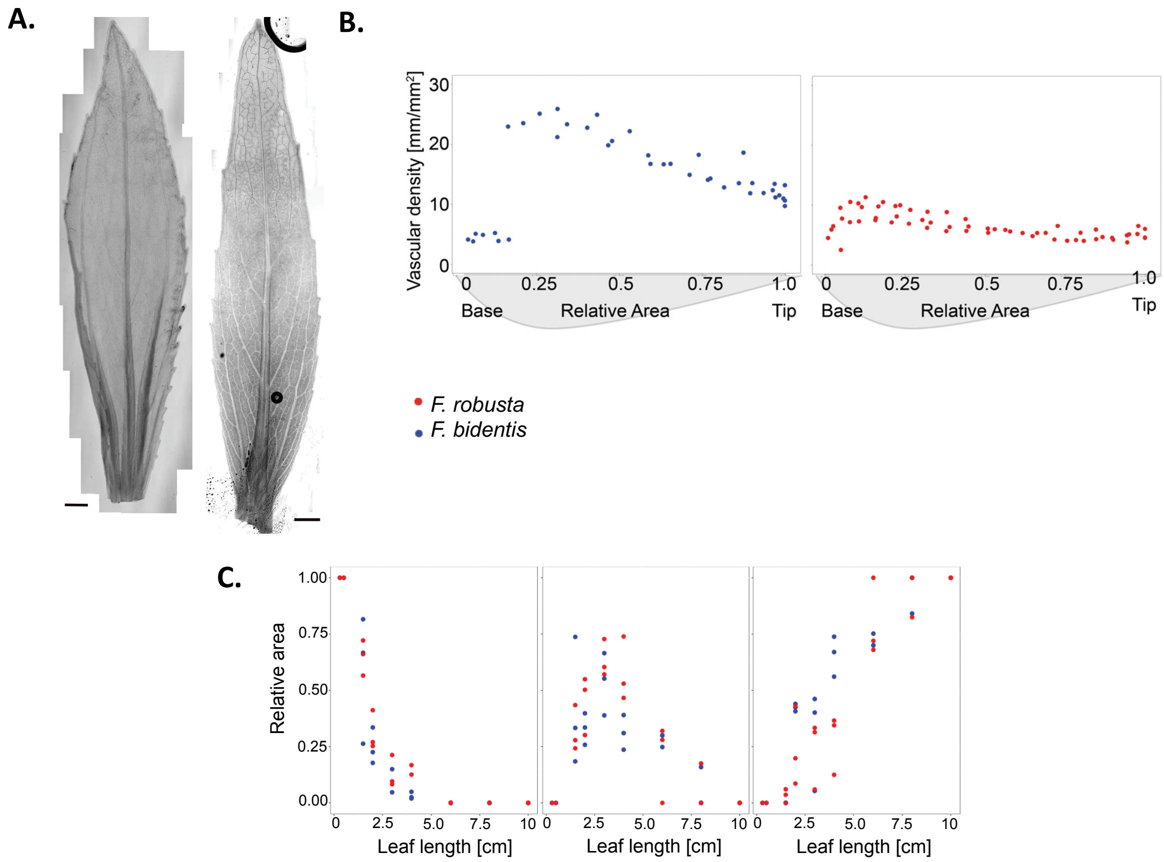


Figure 2. Anatomical changes during *Flaveria* leaf development
A. Stitched pictures of 2 cm leaf of *F. bidentis* (left) and *F. robusta* (right) that were used to determine the progression of vascular development. Scale bar – 1mm. **B.** Vascular density of 4 cm sized leaves presented as a function of the relative area. **C.** Relative proportions of Dividing (left), Differentiating (middle) and Photosynthetic (right) areas inferred from vascular density. *F. bidentis* in blue and *F. robusta* in red.

Estimation of gene expression patterns in the different developmental areas

In order to deduce the gene expression patterns in the different leaf areas identified by differences in vein patterning, area specific expression has to be calculated from whole leaf expression data. In this light, every leaf is a composite of tissues with different developmental states, represented by the areas with different vein patterning. The proportions of the areas change depending on the developmental state of the leaf. When we assume that the expression of individual genes stays constant in the respective area of all leaf stages, the expression (E) of each gene (G) within a leaf sample (S) can be written as the sum of products of its area specific expression (e(A, G)) multiplied by the relative abundance of the area in a sample (A(S, A)) (Form.1A). In a sample, the relative abundances of the areas are identical for all genes, while the individual expression of all genes are area specific. This allows the formalization as matrix multiplication (Form.1B). As we consider relative tissue abundances, their sum has to add up to one. Using these constraints we applied Nonnegative matrix factorization (NMF) (Lee and Seung, 1999; Brunet *et al.*, 2004) to deconvolute the gene expression data from primordia and all leaf developmental stages (one to nine) from both species for all genes.

Formula 1A:

$$E_{(G,S)}=\sum_{i=1}^{n} A_{(S,A)} * E_{(A,G)}$$

Formula 1B:

$$\begin{bmatrix} E_{1,1}, E_{1,2}, E_{1,3} \ldots E_{1,S} \\ E_{2,1}, E_{2,2}, E_{2,3} \ldots E_{2,S} \\ \ldots \\ E_{G,1}, E_{1,2}, E_{1,3} \ldots E_{G,S} \end{bmatrix} = \begin{bmatrix} e_{1,1}, e_{1,2}, e_{1,3} \ldots e_{1,a} \\ e_{2,1}, e_{2,2}, e_{2,3} \ldots e_{2,a} \\ \ldots \\ e_{G,1}, e_{G,2}, e_{G,3} \ldots e_{G,a} \end{bmatrix} X \begin{bmatrix} A_{1,1}, A_{1,2}, A_{1,3} \ldots A_{1,S} \\ A_{2,1}, A_{2,2}, A_{2,3} \ldots A_{2,S} \\ \ldots \\ A_{a,1}, A_{a,2}, A_{a,3} \ldots A_{a,S} \end{bmatrix}$$

We determined the optimal number of areas to assume for NMF and the optimal algorithm to use, by calculating the cophenetic correlation coefficients as described by Brunet *et al.,* (2004). To this end, NMF was performed 50 times using different NMF algorithms for two to ten areas. Ideal predictions were determined for five different areas when either the Kullbeack-Leibler, the Lee and Suad or the Brunet algorithm was used (Supplemental Figure 2). Non-Smooth NMF increases the sparseness and therefore the number of zeros in both matrices using a smoothness matrix (Pascual-montano *et al.*, 2006), which can provide increased biological interpretability. Correspondingly the results varied depending on the sparsity forced upon the predicted basis and coefficient matrices, with a medium sparsity creating the best results (Supplemental Figure 2). The algorithm using Kullbeack-Leibler distance (KL) exhibited a low sum of squares for five leaf areas and was chosen for further analysis.

NMF was performed for the whole dataset 500 times. In both species, the primordial samples occupy a special position within the dataset (Supplemental Figure 3). Their presence creates an area that contributes to the majority of the transcription in the primordial sample and does not reoccur within the remaining leaf stages. Analysis of k-means clustering coupled to GO enrichment shows the genes peaking in this area to be enriched in hormone response, histone organization, shoot system development, meristem development, and meiotic cell cycle. The primordial samples represent a separate functional state and are only marginally present in the remaining leaves of the developmental gradient. Therefore we omitted them for the follow-up analysis and performed the final NMF with four areas.

We calculated the quality of the fit provided from the NMF, area abundances and the area specific expressions for each gene via the regression coefficient (R^2). Overall an average R^2 of 0.81 was achieved for the *F. bidentis* data and an average R^2 of 0.75 for the *F. robusta* data. This corresponds to 8690 of 10864 (80 %) genes with a $R^2 > 0.7$ in *F. bidentis* and 7558 of 10864 (70 %) genes with a R2 > 0.7 in *F. robusta*.

While analysis of the anatomy points towards three different developmental areas (Figure 2C), prediction of different areas based on transcriptome data results in four different areas with a quite different gene expression profiles (Figure 3A). Comparison of the predicted to the measured relative areas (Figure 3A and 2C, respectively) strongly suggests that the area with dividing cells is covered by the predicted area A1 according to high Pearson correlations of 0.98 for *F. bidentis* and 0.97 for *F. robusta*. Similar results were obtained for the differentiating area and the predicted area A2 with correlation coefficients of 0.95 and 0.92 for *F. bidentis* and *F. robusta*, respectively. The third measured area exhibits a weak correlation with either predicted area A3 or A4. However, it highly correlates with the added value of the two predicted areas (0.98 *F. bidentis* and 0.94 *F. robusta*) suggesting that this anatomically uniform area comprise two different zones that can be distinguished by different gene expression profiles in both species.

We further confirmed the validity of our modelling approach by comparing the modelled expression data to the dataset published by Kümpers *et al.*, (2017), who sequenced the transcriptomes of 2 cm leaves from *F. bidentis* and *F. robusta* that were divided into six slices along the leaf from base to tip. The PCA of both datasets illustrates that they span a similar developmental space, with our modelled tissues representing central states within this progression (Figure 3B).

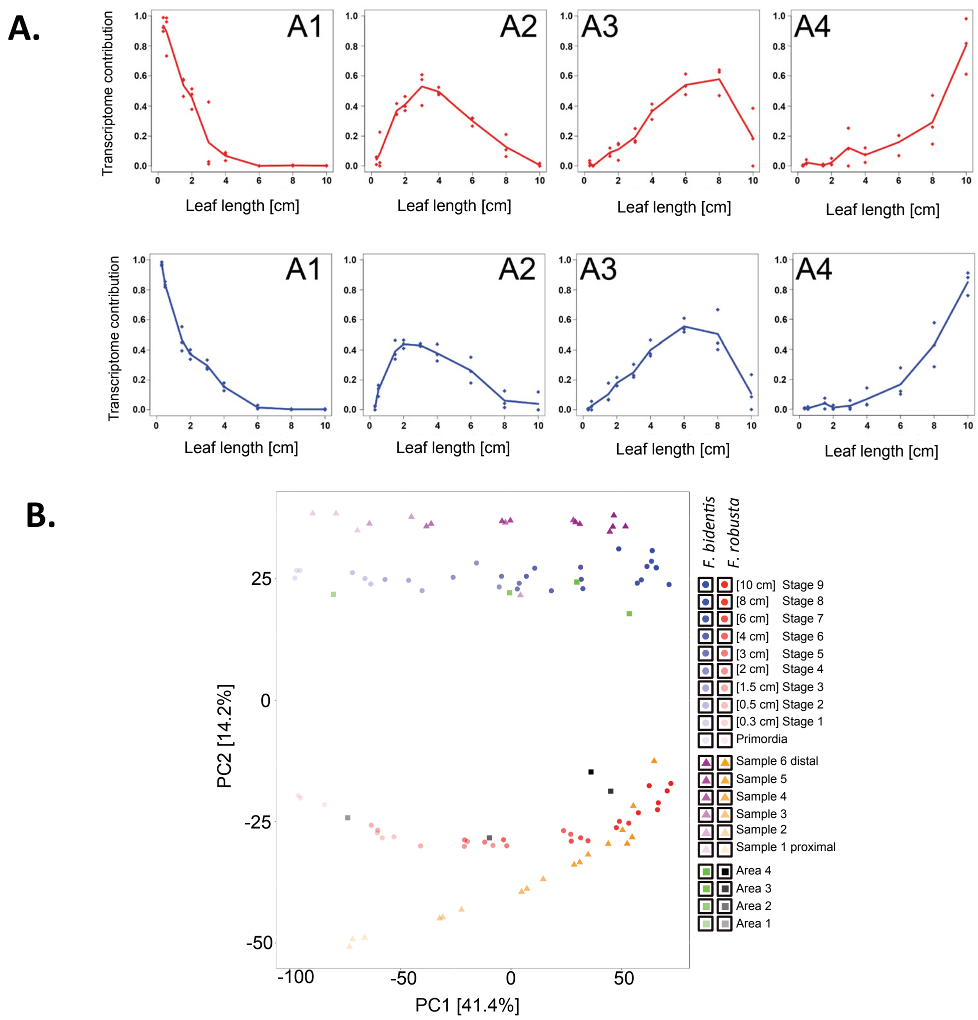


Figure 3. Unsupervised deconvolution
A. The relative contributions of each modelled area to the leaf transcriptomes in the developmental gradient. *F. bidentis* in blue and *F. robusta* in red. **B.** PCA of the developmental gradient for all leaf stages sequenced in this study (Duesseldorf; round dots) as well as the *Flaveria* slice transcriptomics from Kümpers *et al.* 2017 (triangles). The modelled areas are depicted in green (*F. bidentis*) and black (*F. robusta*).

Functional relevance of modelled leaf areas

To evaluate the biological significance of the predicted areas and modelled expression data, we applied k-means clustering to the deconvoluted dataset for both species separately. We obtained seven clusters for each species (Supplemental Figure 4) and GO enrichment analysis was performed using genes with an R^2 score higher than 0.7 when modelled data were compared to real expression profiles in the analyzed species. Five clusters showed a remarkable expression pattern with an expression peak at the very early, early, middle, late and very late stage of the leaf development, respectively (Figure 4). Accordingly, these clusters were termed E1, E2, M, L1, and L2. Area A1 is modelled to compose nearly 100 % in young leaves and declines to 0 % by leaf stage seven (6 cm) in both species (Figure 3A). Cluster E1 (Figure 4) contains genes specifically up regulated within in the first area (A1). A variety of GO categories are overrepresented in this cluster in both species. Most of them are characteristic for processes during early leaf development covering cell division, DNA replication, meristem initiation or the determination of bilateral symmetry (Figure 4). For example, genes encoding CUP SHAPED COTYLEDON3 (CUC3), REVOLUTA (REV), YABBY3 (YAB3), components of the COP9 SIGNALOSOME COMPLEX, HISTONES, CYCLINS (CYCs), CYCLIN DEPENDENT KINASES (CDKs) or the SISTER CHROMATID COHESION PROTEIN 3 (SCC3) are in cluster E1 in both species. Cluster E2 exhibits high expression in Areas A1 and A2 and is mainly characterized by ribosome biogenesis, peptide biosynthesis, and protein biosynthesis.

Cluster M contains genes specific for area 2 (A2), contributing to 44 % +/- 3 % and 52.9 % +/- 11 % of the total transcriptome in *F. bidentis* and *F. robusta*, respectively. The area A2 peaks in leaf stage four (2cm) of *F. bidentis* and in leaf stage five (3 cm) of *F. robusta* (Figure 3A). Enriched GO terms in this cluster for both species are chloroplast organization, plastid transcription and translation as well as chlorophyll biosynthesis (Figure 4).

None of the clusters was specific for area three (A3) although cluster L1 peaked in this area but also exhibit increased expression in area A4 in both species. GO enrichment in this cluster comprises photosynthetic light reactions and carbon fixation (Figure 4). Cluster L2 peaks in area A4, which appears first at leaf stage five (3 cm) in both species and increases rapidly to contribute the majority of the leaf transcriptome in both species in the final stages (Figure 3A). GO terms enriched for this cluster are cell death, aging, several transport processes and cellular response to starvation including the mobilization of nutrients (Figure 4), which are all known to be related to senescence.

Based on these results we conclude that our deconvolution approach was quite reasonable as the modelled areas A1 to A4, according to their gene expression patterns, represent dividing tissues (A1), differentiating tissues with the onset of large-scale plastid development (A2), mature photosynthetic tissue (A3 and A4) and the onset of senescence (A4).

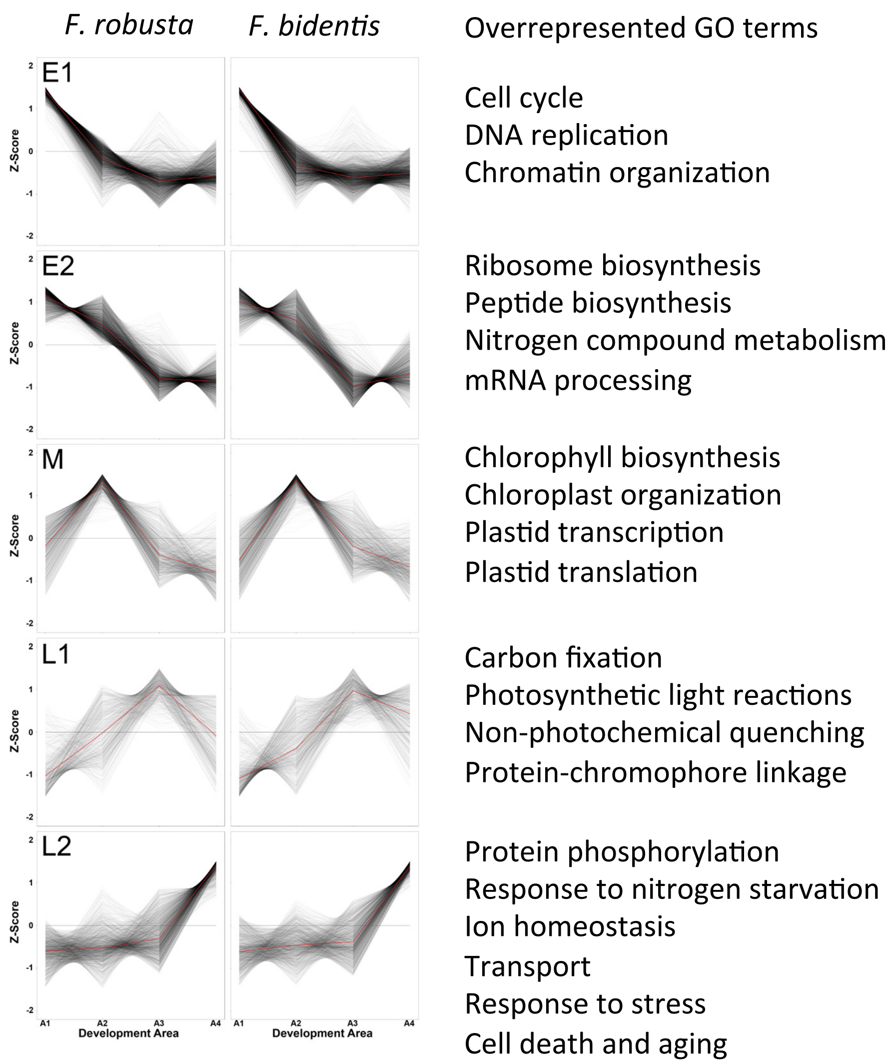


Figure 4. Functional relevance of modelled leaf areas
Analysis of k-means clustering based on the modelled gene coefficients. An excerpt of the overrepresented GO terms that are identical in both species is presented. A1-4 corresponds to the Areas presented in Figure 3A.

Photosynthesis and photorespiration

Genes related to photosynthesis and photorespiration are mainly found in clusters L1 and L2 (Figure 4 A; Supplementary Figures 5, 6 and 7) and are highly expressed in areas A3 and A4. *F. bidentis* is known to be an NADP-ME C_4 species and all landmark genes known to be related to the C_4 pathway in *Flaveria* (Gowik *et al.*, 2011) like *CARBONIC ANHYDRASE* (*CA*), *PHOSPHOENOLPYRUVATE CARBOXYLASE* (*PEPC*), *NADP-MALATE DEHYDROGENASE* (*NADP-MDH*), *ASPARTATE AMINOTRANSFERASE* (*AspAT*), *NADP-MALIC ENZYME* (*NADP-ME*), *ALANINE AMINOTRANSFERASE* (*AlaAT*), *PYRUVATE, ORTHOPHOSPHATE DIKINASE* (*PPDK*), *PPDK-REGULATORY PROTEIN* (*PPDK-RP*) are highly expressed in areas A3 and A4 of leaves of the C_4 species *F. bidentis* but not in leaves of the C_3 species *F. robusta* (Supplementary Figure 5).

Consistent with the results of (Nakamura *et al.*, 2013) transcripts related to both complexes involved in the cyclic electron transport, the NADH dehydrogenase-like (NDH) complex and *PGR5*/*PGRL1* are more abundant in the C_4 species while the expression of most transcripts peak in area A3 (Supplementary Figure 6). Genes encoding the proteins of the Calvin-Benson-Bassham (CBB) cycle are most highly expressed in areas A3 and A4 of both species. Most of the CBB genes and especially the Rubisco small subunit genes are more highly expressed in the C_3 species (Figure 5A). As to be expected, transcripts related to photorespiration are more abundant in the C_3 species *F. robusta*. The expression of photorespiratory genes is highest in leaf areas A3 and A4 (Supplementary Figure 7). The expression of genes related to photosynthesis and photorespiration is in accordance with our modelling result, that areas A3 and A4 represent mature photosynthetic active tissues.

Chloroplast development and division

Chloroplast differentiation is a key regulatory step in cell development to exit from the proliferation stage (Andriankaja *et al.*, 2012). We have identified a couple of transcription factors that positively regulate chloroplast development and/or biogenesis being upregulated during C_4 leaf differentiation compared to C_3 leaves (Figure 5B). In the light, photomorphogenesis is promoted by transcription factors like the ELONGATED HYPOCOTYL 5 (HY5) (Waters and Langdale, 2009). It is known that B-class GATA transcription factors (Chiang *et al.*, 2012; Hudson *et al.*, 2013) and GLODEN2-LIKE transcription factors (GLK1 and GLK2) (Hall *et al.*, 1998; Fitter *et al.*, 2002; Waters *et al.*, 2009; Wang *et al.*, 2013) positively regulate the chloroplast development. *CGA1* (ATGATA22; *CYTOKININ-RESPONSIVE GATA FACTOR 1*) is upregulated in the C_4 species *F. bidentis*. In both species, *CGA1* attained maximum expression in leaf area A2 whereas transcript abundances are two-fold higher in area A2 of the C_4 leaves (Figure 5B). CGA1 transcription factor positively regulates chloroplast development and division (Chiang *et al.*, 2012). Transcripts of *GLK-2* did not differ much between C_3 and C_4 *Flaveria* species while *GLK-1* levels are slightly up in *F. robusta* (Figure 5B). We have further identified genes encoding other transcription factors like STH2 or BBX21 (SALT TOLERANCE HOMOLOG 2 or B-BOX CONTAINING PROTEIN 21) and HEC2 (HECTATE2) being upregulated in the C_4 leaves. STH2/BBX21 and HEC2 indirectly promote photomorphogenesis by positively regulating the transcription of *HY5* and by inhibiting the expression of *PHYTOCHROME INTERACTING FACTOR1* (*PIF1*) respectively (Xu *et al.*, 2016; Zhu *et al.*, 2016) (Figure 5B). We can only speculate that up-regulation of these factors in the C_4 species is related to chloroplast development in the bundle sheath, since overall chloroplast development seems to be quite similar in both species as indicated by genes related to plastid division and chloroplast gene expression (Figure 5B).

Plastid division mainly occurs in the area A2 in both species, as is reflected in the transcription patterns of *FtsZ1* and *FtsZ2* (Figure 5B). Chloroplast transcription peaks in a similar manner in area A2, with the highest predicted expression of the plastidic RNA polymerase (*RpoTp*) as well as *SIGMA TRANSCRIPTION FACTORS*, two (*SIG2*) and six (*SIG6*) (Figure 5B). The mitochondrial RNA polymerase gene (*RpoTm*) on the other hand shows the highest expression in area A1, while *RopTmp* believed to interact with both organelles (Kühn *et al.*, 2007) exhibits similar levels in both areas (Figure 5B).

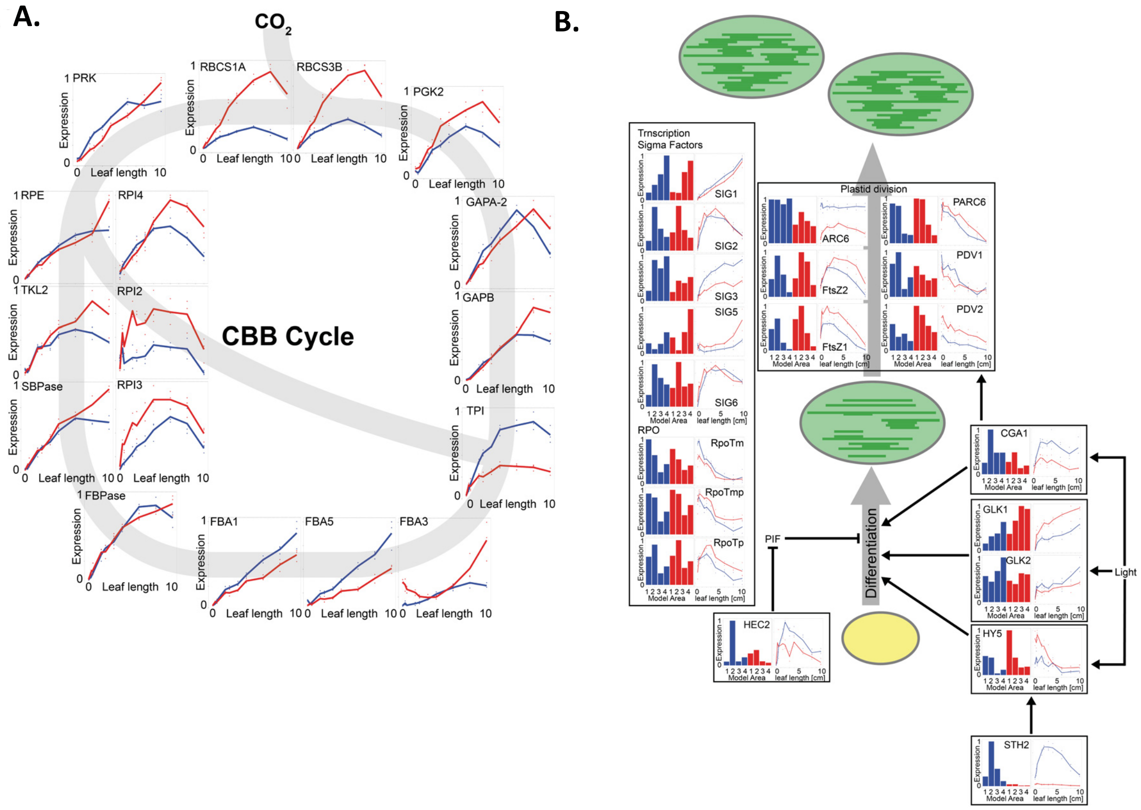


Figure 5. Expression patterns of the Calvin-Benson-Bassham and chloroplast differentiation/division genes.

A. Expression of CBB cycle genes throughout the leaf developmental series, normalized to maximum expression per species. *RBCS1A* and *RBCS3B – RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN 1A* and *3B*, respectively; *PGK – PHOSPHOGLYCERATE KINASE*; *GAPA2 – GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE A SUBUNIT 2*; *GAPB – GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE B SUBUNIT*; *TPI – TRIOSEPHOSPHATE ISOMERASE*; *FBA – FRUCTOSE-BISPHOSPHATE ALDOLASE*; *FBPase – FRUCTOSE-1,6-BISPHOSPHATASE*; *SBPase – SEDOHEPTULOSE-1,7-BISPHOSPHATASE*; *TKL – TRANSKETOLASE*; *RPI – RIBOSE-5-P ISOMERASE*; *RPE – RIBULOSE-5-P EPIMERASE*; *PRK – PHOSPHORIBULOKINASE* **B.** Expression of genes connected to plastid differentiation and division. Modelled and original data is normalized to its maximum within each species. *ARC6 – ACCUMULATION AND REPLICATION OF CHLOROPLASTS 6*; *FtsZ* – homolog of bacterial Filamenting temperature-sensitive Z gene; *PARC6* – paralog of ARC6; *PDV1* and *PDV2 – PLASTID DIVISION 1* and *2*, respectively; *CGA1 – CYTOKININ-RESPONSIVE GATA FACTOR 1; GOLDEN-LIKE* genes (*GLK1* and *GLK2*); *HY5 – ELONGATED HYPOCOTYL 5; STH2 – SALT TOLERANCE HOMOLOG 2; HEC2 – HECTATE2;* plastidic RNA polymerase gene (*RpoTp*); mitochondrial RNA polymerase gene (*RpoTm*); *RpoTmp* – plastid RNA polymerase, localized both in mitochondria and chloroplast. *F. bidentis* in blue and *F. robusta* in red.

The timing of mesophyll and bundle sheath differentiation is comparable in both species

Analysis of leaf development in C_3 and C_4 Cleomaceae species revealed that increased vein formation in the C_4 species *Gynandropsis gynandra* is related to a delay in mesophyll cell differentiation during leaf development (Külahoglu *et al.*, 2014). We could not observe something comparable happening during leaf development in C_3 and C_4 *Flaveria*. Leaf development proceeds quite similar in both species regarding to changes either in anatomy or in gene expression. Completely developed bundle sheath and palisade parenchyma appear in the same leaf stage in both species (stage five; Supplementary Figure 1). Expression of photosynthetic genes and genes related to chloroplast development peak in the same leaf stages (Figure 4 and Figure 5). Overall transcriptional investment is quite similar in both species for all analyzed leaf stages (Figure 1A). When analyzing our modelled data, the predicted leaf areas are again very similar for both species (Figure 3A). Leaf area A1 is modelled to compose nearly 100 % in young leaves and declines to 0 % by leaf stage seven (6 cm) in both species. Area A2 peaks in leaf stage four (2cm) of *F. bidentis* and in leaf stage five (3cm) of *F. robusta* (Figure 3A) whereas area A3 peaks in leaf stages seven and eight of both species respectively. The leaf area A4 first appears in leaf stage five and increases in both species with leaf size (Figure 3A). These findings are in line with the results of Kümpers *et al.*, (2017) who also not reported any delays in leaf development or cell differentiation in C_4 compared to C_3 *Flaveria*.

Auxin metabolism is upregulated in developing and differentiating tissues of the C_4 species compared to the C_3 species

In leaves, polar auxin transport is central to the development of a functional vascular network and controls its structure and density. The local auxin maxima are created by auxin synthesis in dividing tissues and through transport facilitated by the concerted expression of different PIN proteins (Verna *et al.*, 2015). Accordingly, genes related to the main Indole-3-acetic acid (IAA) synthesis pathway, the Indole-3-pyruvic acid (IPA) pathway (Mashiguchi *et al.*, 2011), e.g. *TAR2* (*TAA RELATED PROTEIN 2*) and *YUCCA1* (*YUC1*) are mainly expressed in leaf area A1 and correspondingly leaf stages one and two of both species (Figure 6). *YUC1*, encoding the enzyme catalyzing the rate-limiting step of the IPA pathway (Mashiguchi *et al.*, 2011) is higher expressed in the C_4 species compared to the C_3 species indicating that IAA synthesis might be enhanced. Additionally, AMIDASE 1(AMI1) catalyzing the synthesis of IAA from Indole-3-acetamide is highly expressed in the differentiating area of both the species and might contribute to auxin homeostasis.

IAA can be inactivated by sequestration into amino acid conjugates (Nakazawa *et al.*, 2001; Takase *et al.*, 2004; Staswick *et al.*, 2005). This reaction is catalyzed by GRETCHEN HAGEN 3 (GH3) enzymes (Staswick *et al.*, 2002; Staswick *et al.*, 2005). Genes encoding these enzymes are known to be induced by auxin (Hagen and Guilfoyle, 1985). Accordingly, we observe expression of *GH3* genes mainly in leaf area A1 in both the species where auxin is synthesized. The majority of the *GH3* genes covered by our dataset exhibit a marked increase in expression in the C_4 species (Figure 6). IAA can be set free from amino acid conjugates and thus be reactivated by hydrolyzing enzymes (Bartel and Fink, 1995; Davies *et al.*, 1999; Rampey *et al.*, 2004). Genes encoding such enzymes like *ILL* (*IAA-AMINO ACID CONJUGATE HYDROLASE*) and *ILR* (*IAA-LEUCINE RESISTANT*) genes are more highly expressed in leaf area A1 and especially leaf area A2 of the C_4 species compared to the C_3 species (Figure 6).

Upregulation of auxin conjugating and reactivating genes in the division and differentiation areas of C_4 leaves implies a higher capacity of auxin homeostasis in the C_4 species *F. bidentis*. To test this assumption, we treated three week old C_3 and C_4 *Flaveria* plants for three weeks with the synthetic auxin NAA (1-Naphthaleneacetic acid) and 2,4-D (2,4-Dichlorophenoxyacetic acid) via leaf spraying. In both species, 2,4-D application was lethal in the concentration above 10 µM. At 10 µM it caused a severe phenotype with distorted leaf shape and significantly increased vascular density in both species (Figure 7). With 1µM 2,4-D leaf shape was still distorted in both species. Strikingly the vascular structure of the C_4 plant *F. bidentis* was highly perturbed with fused veins forming plate-like structures, while the C_3 species showed an increase in vascular density (Figure 7B). NAA changed the overall leaf shape and significantly increased vascular density in *F. robusta* (C_3) at its highest concentration of 790 µM but not in *F. bidentis* (C_4) (Figure 7C). Similar effects, specifically an increased vein density, are reported when *A. thaliana* leaves accumulating auxin (Mattsson *et al.*, 1999; Sieburth, 1999).

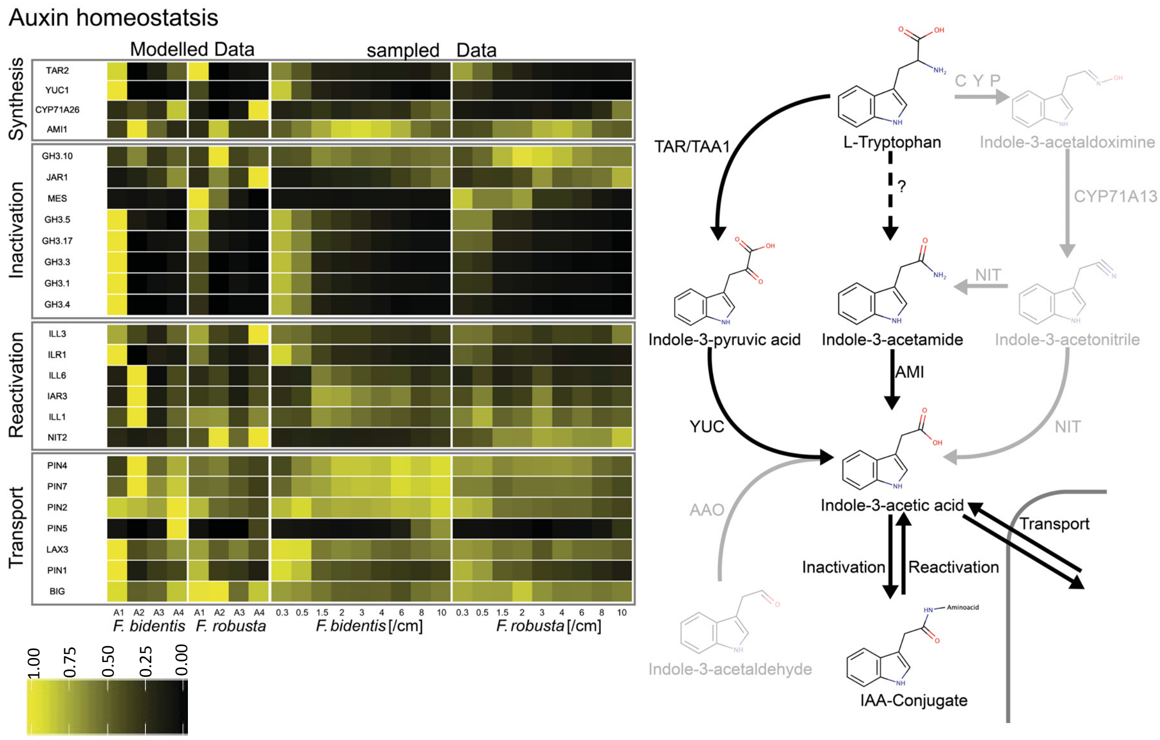


Figure 6. Heatmap of genes related to Auxin homeostasis
Gene expression is presented separately for the modelled and the sampled data, normalized to the respective maximum in both species. Representation of auxin biosynthesis Indole-3-acetic acid pathway (right panel).
TAR2 – TRYPTOPHAN AMINOTRANSFERASE RELATED 2; *YUC1 – YUCCA1*; *CYP71A26* – Putative cytochrome P450; *AMI1 – AMIDASE 1*; *GH3 – GRETCHEN HAGEN 3*; *JAR1 – JASMONATE RESISTANT 1*; *MES – METHYL ESTERASE*; *ILL – IAA-AMINO ACID CONJUGATE HYDROLASE*; *ILR – IAA-LEUCINE RESISTANT*; *IAR – IAA-ALININE RESISTANT*; *NIT2 – NITRILASE 2*; *PIN – PIN-FORMED*; *LAX3 – LIKE AUX1-3*; *BIG* – Calossin-like protein.

The outcome of the spraying experiments confirms our idea according to altered auxin homeostasis in *F. bidentis* compared to *F. robusta*. Obviously, the C_4 plant can tolerate higher auxin levels than the C_3 plant without clear alterations of leaf anatomy. 2,4-D generates effects at lower concentrations compared to NAA since it cannot be sequestered effectively into amino acid conjugates by GH3 proteins while NAA is a GH3 substrate (Staswick *et al.*, 2005).

Genes related to auxin transport are also differentially expressed in both species. While PIN1 expression is highest in dividing tissues (area A1) and decreases rapidly during differentiation (area A2) in both species its expression is higher in the *F. bidentis* (C_4). PIN7 and PIN4 expression peak during differentiation (area A2) and PIN2 and PIN5 expression peak in area A4 of the C_4 leaves while these genes exhibit much lower transcript levels in the C_3 leaves (Figure 6).

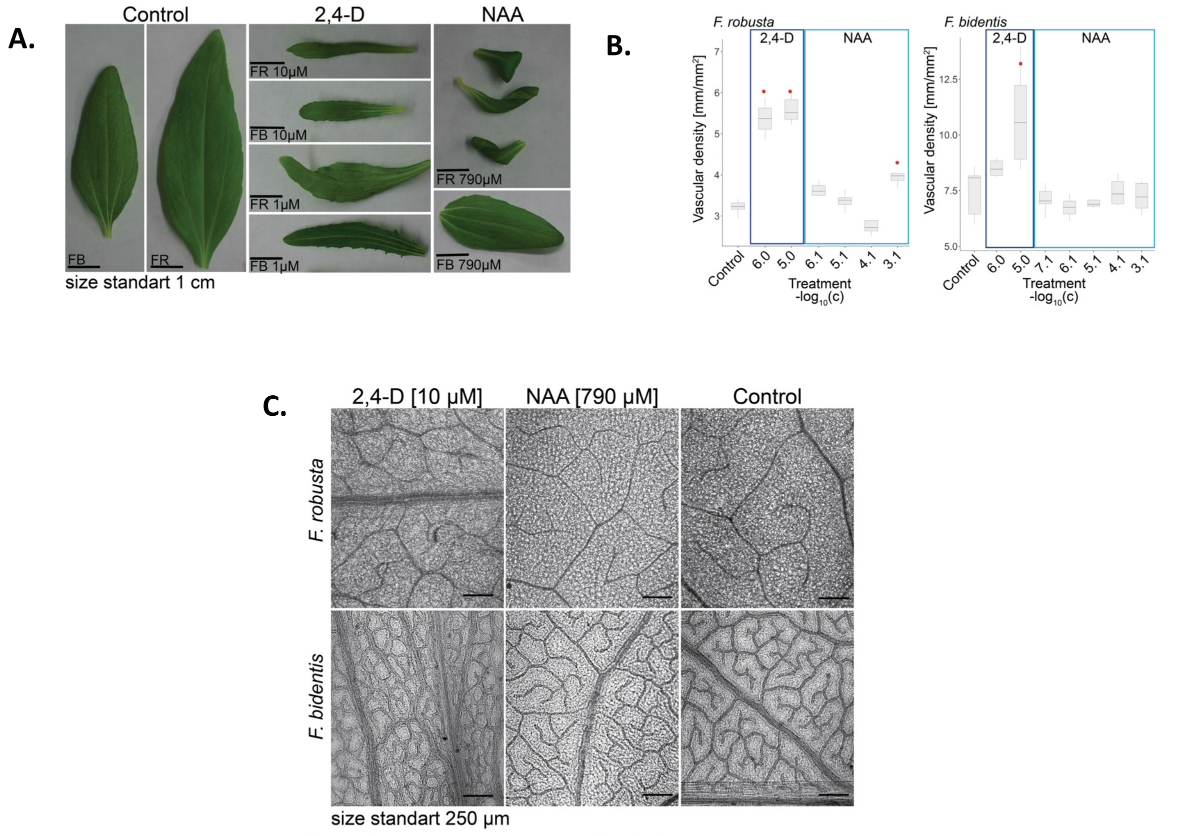


Figure 7. Anatomical changes of *F. bidentis* and *F. robusta* leaves sprayed with synthetic auxins
A. Leaf shapes of plants sprayed with 2,4-D, NAA or the 0.1 % DMSO control. **B.** Vascular structure of plants sprayed with 10 µM 2,4-D, 700 µM NAA or the 0.1 % DMSO control.

Genes related to auxin signaling and auxin induced pathways are differentially expressed in C_3 and C_4 leaves

Genes encoding the auxin receptors TIR1 (Transport inhibitor response 1), AFB1 and AFB3 (Auxin signaling F-Box 1 and 3) are specifically upregulated in leaf area A2 of the C_4 species *F. bidentis* (Supplemental Figure 8) whereas their expression is more uniform but lower in areas A1 and A2 of the *F. robusta* (C_3) leaf (Supplemental Figure 8). Expression of other, more downstream, components of auxin signaling e.g. *IAA* and *ARF* genes appear to be more similar in both species (Supplemental Figure 8). Although we found that expression patterns of *IAA5* and *IAA6* as well as of *ARF3*, *ARF19* and *ARF16* show peaks in leaf area A2 in the C_4 species that are missing in the C_3 species (Supplemental Figure 8).

In accordance with the expression patterns of genes related to auxin synthesis, homeostasis and transport we observed a tissue specific induction of genes known to be regulated by auxin, like the *SAUR* gene family (small auxin up regulated RNAs) (Supplemental Figure 9). The majority of these genes were most highly expressed in leaf area A1 of both species. Although we identified several genes like *SAUR10*, *SAUR*12, *SAUR*16, *SAUR*53 or the auxin-induced dormancy related gene AT1G54070 that were exclusively or most highly expressed in leaf area A2 of the C_4 species *F. bidentis* (Supplemental Figure 9). This finding further confirms the assumption that increased expression of genes related to auxin synthesis and homeostasis in leaf areas A1 and A2 of *F. bidentis* translate to higher auxin levels especially in leaf area A2 of the C_4 species compared to the C_3 species.

Genes related to bundle sheath differentiation are upregulated in the C_4 species

Bundle sheath tissue in leaves is equivalent to the root endodermis and the starch sheath of hypocotyls and the development of bundle sheath cells is at least partially conserved to root endodermis development and differentiation (Wysocka-Diller *et al.*, 2000; Lim *et al.*, 2005; Slewinski, 2013; Cui *et al.*, 2014; Kyung Yoon *et al.*, 2016). It is known that the interplay of the GRAS-type transcription factors SCARECROW (SCR) and SHORT-ROOT (SHR) is essential to specify the root endodermis identity as well as for the correct bundle sheath development in the C_3 species *Arabidopsis* and the C_4 species maize (Slewinski *et al.*, 2012; Cui *et al.*, 2014; Slewinski *et al.*, 2014). In both *Flaveria* species, transcript abundances of genes encoding SCR and SHR attain a maximum in dividing tissue (area A1) and decline during further leaf development. Transcripts of both, *SCR* and *SHR* are upregulated in *F. bidentis* (C_4) leaf areas A1 and A2 compared to the C_3 species (Supplementary Figure 10). The activity of *SCR* and *SHR* are regulated by interactions with intermediate domain (IDD) transcription factors that also fulfill roles in several other developmental processes (Welch *et al.*, 2007; Ogasawara *et al.*, 2011; Coelho *et al.*, 2018). Recently, *IDD*s were proposed to be also involved in bundle sheath development recently (Coelho *et al.*, 2018). Transcripts of *IDD2*, *IDD5*, *IDD10* (*JKD*) and *IDD12* genes were specifically more abundant in differentiating tissue (area A2) of the C_4 species, while *IDD15*, *IDD7*, *IDD16,* and *MGP* abundance peaked in area A1 of both species but were again higher in the C_4 species, though MGP transcript levels are similar in both species (Supplementary Figure 10). Expression patterns of *SCARECROW-LIKE* gene family members (*SCL*) are more similar in both species. Only *SCL3* is upregulated in area 2 in C_4 species *F. bidentis* (Supplementary Figure 10). *SCL3* gene promotes root endodermal cell elongation (Heo *et al.*, 2011). Genes involved in bundle sheath development and differentiation can be expected to be upregulated in developing leaves of C_4 species compared to C_3 species since more bundle sheath cells has to develop due to the higher vein density in C_4 leaves.

(V) Discussion

C_4 plants evolved more than 60 times independently from C_3 ancestors. Since C_4 photosynthesis demands a special leaf anatomy for functioning efficiently, termed Kranz anatomy, leaf anatomy had to change during most of these independent evolutionary events. Kranz anatomy is characterized by high vein density and enlarged bundle sheath cells with elevated numbers of organelles especially chloroplasts. To get an insight on how leaf development changes at the molecular level to achieve these alterations of leaf anatomy, we analyzed the anatomy and transcriptomes of developing leaves from the closely related C_3 and C_4 species *F. robusta* and *F. bidentis*. Analyses of vein densities in a developmental series of leaves from both species point towards the existence of three clearly distinguishable areas in leaves of both the species - a cell division area, a cell differentiation area and an area with mature, photosynthetic cells. We used nonnegative matrix factorization to deconvolute transcriptome data from whole leaves of sequential stages covering the growth process in both the species. This revealed the existence of four highly distinct transcriptional patterns within the leaves of both the species.

NMF as a tool for the deconstruction of developmental processes

Nonnegative matrix factorization was introduced as a tool to recognize patterns for the decomposition of images into meaningful features (Lee and Seung, 1999), spectral data analysis or denoising of audio signals. In biology it has been applied to a variety of datasets ranging from EEG based imaging (Lu and Yin, 2015; Delis *et al.*, 2016) and muscle electrographs to transcriptomic datasets from microarrays or RNA-Seq experiments (Kong *et al.*, 2011; Shao and Höfer, 2017; Duren *et al.*, 2018; Ebied *et al.*, 2018). Regarding transcriptome data, the initial idea was published by Brunet *et al.,* (2004). They applied NMF on a gene by sample matrix to identify metagenes from microarray data of different tumor cells. An alternative option is the use of NMF to factorize sample by gene matrices to create so-called metasamples. Recently, Shao and Höfer (2017) analyzed metasamples in single cell gene expression datasets and were able to demonstrate that these can be used to correctly classify cell types and detect their signature genes.

In our work, we used a restricted metagene approach. The expression data of metagenes were normalized to a sum of 1 in every leaf sample reflecting the attempt to detect the relative composition of tissues within a leaf. The assumption, that the expression of individual genes stays constant in a respective developmental area throughout all the different leaves, is of course, an oversimplification of the real situation and might be inaccurate for certain genes. Therefore, we critically assessed the results of the NMF. Using this method, we were able to obtain a realistic representation of the tissue composition of the different leaves solely from the transcriptome data in both species. The cell division area as well as the cell differentiation area, as defined by their anatomical features, could be clearly identified as indicated by the high correlations of modelled and measured area sizes (Figures 2 and 3). Additionally, the factorized data indicate that the area of mature, photosynthetic cells can be subdivided into two areas with clearly different transcript patterns. Clustering of gene expression data and analysis of enriched GO terms indicate that this is due to the onset of senescence in this area of older leaves. The modelled expression data correctly capture the signature expression differences between both species as indicated by genes and metabolic pathways we can have clear expectations for, like C_4 photosynthesis, photorespiration or the CBB cycle (Figure 5A, Supplementary Figures 5, 6 and 7). Finally, we compared the modelled expression data with the results of Kümpers *et al.*, (2017) who analyzed base to tip maturation gradients of C_3 and C_4 *Flaveria* leaves. We found both datasets to be highly similar when compared via PCA (Figure 3B). Overall our approach to deconvolute transcriptome data from a developmental series of whole leaves to identify distinct transcriptome patterns within the leaves appears to be justified and valid.

Leaf development proceeds very similar in both *Flaveria* species

Both, the analyses of the leaf anatomy as well as the analysis of the corresponding leaf transcriptomes indicate that leaf development in the C_4 species *F. bidentis* is quite comparable to the leaf development in the C_3 species *F. robusta*. Leaves of the same age have a comparable size and the distribution of anatomically distinguishable developmental areas is very similar in leaves of the same size from both species (Figure 2). The main difference we could identify is the overall higher vein density in leaves of the C_4 species.

Accordingly, the transcriptomes of leaves of the same stage from both species are quite similar. Transcriptional investment is very similar when leaves from the same age are compared and the PCA of all our RNA-Seq data indicates that the variation between the samples from different developmental leaf stages is up to three times higher than the species-specific variation (Figure 1B, 1C). Key developmental events like chloroplast division and differentiation occur simultaneously in both species. These findings are in line with earlier studies of leaf development in C_3 and C_4 *Flaveria* species (McKown and Dengler, 2009; Kümpers *et al.*, 2017).

Changes in auxin homeostasis could be related to higher vein density in the C_4 *Flaveria*

The transcriptome analysis of developing leaves from a C_3 and a C_4 *Flaveria* species indicate that the young leaves of the C_4 plants exhibit a higher capacity for auxin synthesis and auxin homeostasis. Genes related to the main auxin synthesis pathway in dicot leaves (Mashiguchi *et al.*, 2011) as well as genes related to the sequestration of auxin into amino acid conjugates, e.g. several *GH3* genes (Staswick *et al.*, 2002; Staswick *et al.*, 2005), are more highly expressed in the cell division area of the C_4 leaves than in the C_3 leaves (Figure 6). On the other hand we found genes related to re-activating the auxins due to hydrolase activity, e.g. *ILL*, *IAR* and *ILR* genes (Bartel and Fink, 1995; Davies *et al.*, 1999; Rampey *et al.*, 2004), up-regulated in the cell differentiation area of the C_4 compared to the C_3 leaves. This implies higher auxin availability in the cell division and differentiation areas of the C_4 compared to C_3 leaves, where vein formation takes place.

To test if developing *F. bidentis* leaves truly exhibit higher capacity of auxin homeostasis than the leaves of *F. robusta* we examined the resistance of both plant species towards externally applied synthetic auxins. Spraying the plants with NAA changed the overall leaf shape and significantly increased vascular density in *F. robusta* but did not alter leaves of *F. bidentis.* Application of 2,4-D, on the other hand, led to distorted leaf shape and significantly increased vascular density in both species in much lower concentrations (Figure 7). The higher sensitivity of both species toward 2,4-D compared to NAA is most likely due to the fact that NAA can be inactivated by GH3 proteins while 2,4-D, like other halogenated auxins, is no substrate of these proteins (Staswick *et al.*, 2005). This indicates that capacity for auxin homeostasis is indeed higher in the C_4 species, while its obvious sensitivity towards 2,4-D confirms an involvement of GH3 proteins.

Veins form from procambial strands. These are induced by elevated auxin concentration due to polar auxin transport along strands of undifferentiated ground meristematic cells (Scarpella *et al.*, 2004). According to the auxin canalization model (Sachs, 1969; Mitchison, 1981; Rolland-Lagan and Prusinkiewicz, 2005), such strands can form by local auxin maxima and induction of polar auxin transport due to PIN proteins. When more auxin is available in the cell division and differentiation areas of *F. bidentis* more local auxin maxima could form and lead to the formation of procambial strands and, in the following, to the formation of more veins than in the C_3 species *F. robusta*. This in line with the observations of McKown and Dengler (2009). They described that the existence of an additional minor vein order (7 vein orders in *F. bidentis* compared to 6 vein orders in *F. robusta*) is the main reason for higher vein density in the C_4 species. They observed an accelerated vein formation during C_4 leaf development and concluded that this could be explained either by increased auxin production, modified leaf ground meristem cell competency to becoming procambium, or a combination of these developmental parameters (McKown and Dengler, 2009). According to our data, the C_4 *Flaveria* can produce more auxin. Due to its enhanced auxin homeostasis capacity, it also can maintain higher auxin concentrations in the early cell differentiation area, where the higher order minor veins are induced (Figure 8).

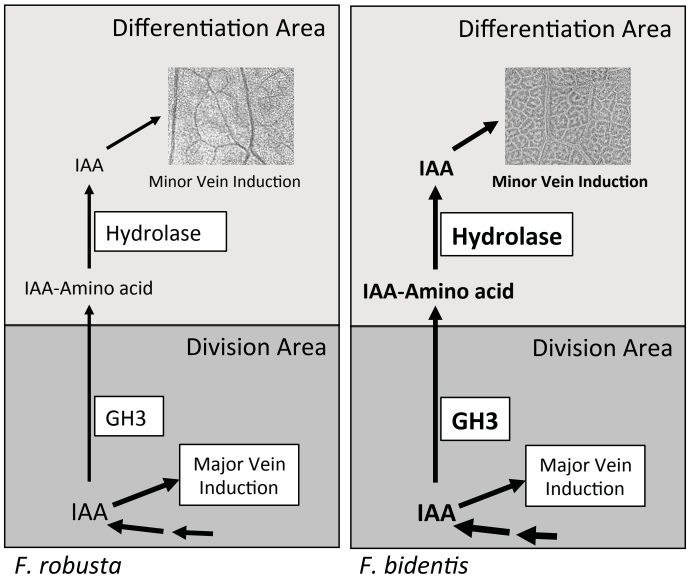


Figure 8. Possible Auxin homeostasis mechanism for higher vein density in C_4 *F. bidentis*
Auxin conjugating GH3 enzymes and Auxin-amino acid conjugate hydrolases are upregulated in division and differentiation area of C_4 *F. bidentis*, respectively.

Polyphyletic evolution of C_4 - comparison of *Flaveria* to other C_4 origins

Given that C_4 photosynthesis is a prime example of convergent evolution one can ask if the differences in leaf development we observed for C_3 and C_4 *Flaveria* species are typical for *Flaveria* only or can be found also in other C_4 species.

It appears quite clear that the interplay of the transcription factors SCARECROW (SCR) and SHORT ROOT (SHR) is important for bundle sheath differentiation and to specify its identity in C_3 and C_4 plants (Slewinski *et al.*, 2012; Cui *et al.*, 2014; Slewinski *et al.*, 2014). Together with several IDD transcription factors, these genes were found to be upregulated in dividing and differentiating tissues of the C_4 leaves compared to the C_3 leaves of the two *Flaveria* species. Since these factors were identified as candidate for bundle sheath differentiation in maize (Fouracre *et al.*, 2014), one can assume that they represent an important common factor for the establishment of Kranz anatomy. This might be due to the fact that they are also important for bundle sheath development in C_3 species. One could assume that they were recruited to fulfill the same role in C_4 leaf development and are found to be upregulated in developing C_4 leaves because C_4 leaves contain more bundle sheath cells compared to C_3 leaves.

We identified several factors like *CGA1, STH2 or BBX21 and HEC2* genes with different expression patterns in the developing leaves of the two *Flaveria* species. CGA1, a GATA transcription factor promotes chloroplast differentiation and division and BBX21, HEC2 positively regulate photomorphogenesis (Chiang *et al.*, 2012; Xu *et al.*, 2016; Zhu *et al.*, 2016). The differences in gene expression are most likely related to increased chloroplast development in the bundle sheath cells. Although, the expression pattern of the known transcription factor ZmGLK2, which is essential for maize bundle sheath chloroplast development (Hall *et al.*, 1998; Rossini *et al.*, 2001) didn't differ between C_3 and C_4 *Flaveria* (Figure 5B).

Regarding vein density, on the other hand, obviously different alterations of the C_3 developmental program can lead to the same effect, namely the high vein density usually found in C_4 leaves. In *Flaveria* a combination of enhanced auxin synthesis and high auxin homeostasis capacity seem to accelerate vein formation by enhancing the induction of procambial strains in competent leaf areas. For C_3 and C_4 species of Cleomaceae it could have been shown that high vein density is related to a delay in mesophyll cell differentiation during leaf development in the C_4 species *Gynandropsis gynandra* (Külahoglu *et al.*, 2014), probably in combination with elevated auxin synthesis (Huang *et al.*, 2017). This leads to induction of more procambial strands and following more veins compared to the C_3 species *Tarenaya hassleriana,* exhibiting a faster mesophyll differentiation (Külahoglu *et al.*, 2014). Overall it appears that vascular density can be represented as a function of auxin concentration, auxin sensitivity and the duration of an inducible/competent state. In this framework, *Cleome* increases vascular density by prolonging the inducible state, while *Flaveria* keeps the length of the inducible state identical but increases auxin availability in tissues with the competence to differentiate into veins.

Lessons for converting a C_3 to a C_4 species

There has been much interest in engineering C_3 plants like rice to express C_4 traits increasing their photosynthetic efficiency and productivity (Hibberd *et al.*, 2008; Schuler *et al.*, 2016). A critical step in engineering a C_4 plant would be to increase leaf vein density. Obviously, auxin metabolism and/or auxin signaling pathways in the developing leaves have to be altered to achieve this. Our analysis, as well as earlier work (Huang *et al.*, 2017; Külahoglu *et al.*, 2014) demonstrates that different possibilities to alter auxin effects exist, it could be either a combination of enhanced auxin synthesis and prolongation of the time auxin is effective in vein initiation, as found in *Cleome,* or a combination of enhanced auxin synthesis and alterations in auxin homeostasis as we found for *Flaveria*. If there are even more different ways towards high leaf vein density and which way is the most suitable for engineering approaches has to be elucidated in future research.

(VI) Acknowledgments

We thank the 'Cluster of Excellence on Plant Sciences' (CEPLAS) for supporting this work. 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.

(VII) References

Alexa, A. and Rahnenführer, J. (2018) Gene set enrichment analysis with topGO. *http://www.mpi-sb.mpg.de/~alexa*.

Anderson, L.E. (1971) Chloroplast and cytoplasmic enzymes II. Pea leaf triose phosphate isomerases. *Biochim. Biophys. Acta*, **235**, 237–244.

Andriankaja, M., Dhondt, S., DeBodt, S., et al. (2012) Exit from proliferation during leaf development in *Arabidopsis thaliana*: A Not-So-Gradual Process. *Dev. Cell*, **22**, 64–78.

Aubry, S., Kelly, S., Kümpers, B.M.C., Smith-Unna, R.D. and Hibberd, J.M. (2014) Deep evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in two independent origins of C_4 Photosynthesis. *PLoS Genet.*, **10**, e1004365.

Bartel, B. and Fink, G.R. (1995) ILR1, an amidohydrolase that releases active indole-3-acetic acid from conjugates. *Science*, **268**, 1745–1748.

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

Benjamini, Y. and Yosef, H. (1995) Controlling The False Discovery Rate - A practical And powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B*, **57**, 289–300.

Benson, A.. and Calvin, M. (1950) Carbon dioxide fixation by green plants. *Annu. Rev. Plant physiol*, **1**, 25–42.

Bräutigam, A. and Gowik, U. (2016) Photorespiration connects C_3 and C_4 photosynthesis. *J. Exp. Bot.*, **67**, 2953–2962.

Bräutigam, A., Kajala, K., Wullenweber, J., et al. (2011) An mRNA blueprint for C_4 photosynthesis derived from comparative transcriptomics of closely related C_3 and C_4 species. *Plant Physiol.*, **155**, 142–156.

Brilhaus, D., Bräutigam, A., Mettler-Altmann, T., Winter, K. and Weber, A.P.M. (2016) Reversible burst of transcriptional changes during induction of Crassulacean Acid Metabolism in *Talinum triangulare*. *Plant Physiol.*, **170**, 102–122.

Brooks, A. and Farquhar, G.D. (1985) Effect of temperature on the CO_2/O_2 specificity of ribulose-1,5-bisphosphate carboxylase/oxygenase and the rate of respiration in the light. *Planta*, **165**, 397–406.

Brunet, J.-P., Tamayo, P., Golub, T.R. and Mesirov, J.P. (2004) Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 4164–4169.

Calvin, M. (1948) The path of carbon in photosynthesis. *Science*, **107**, 476–480.

Campbell, W.J. and Ogren, W.L. (1990) Glyoxylate inhibition of ribulosebisphosphate carboxylase/oxygenase activation in intact, lysed, and reconstituted chloroplasts. *Photosynth. Res.*, **23**, 257–268.

Chiang, Y.-H., Zubo, Y.O., Tapken, W., Kim, H.J., Lavanway, A.M., Howard, L., Pilon, M., Kieber, J.J. and Schaller, G.E. (2012) Functional characterization of the GATA transcription factors GNC and CGA1 reveals their key role in chloroplast development, growth, and division in Arabidopsis. *Plant Physiol.*, **160**, 332–348.

Christin, P.-A. and Osborne, C.P. (2014) The evolutionary ecology of C_4 plants. *New Phytol.*, **204**, 765–781.

Christin, P.-A., Osborne, C.P., Chatelet, D.S., et al. (2013) Anatomical enablers and the evolution of C_4 photosynthesis in grasses. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 1381 – 1386.

Coelho, C.P., Huang, P., Lee, D.-Y. and Brutnell, T.P. (2018) Making roots, shoots, and seeds: IDD gene family diversification in plants. *Trends Plant Sci.*, **23**, 66–78.

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

Davies, R.T., Goetz, D.H., Lasswell, J., Anderson, M.N. and Bartel, B. (1999) *IAR3* encodes an auxin conjugate hydrolase from *Arabidopsis*. *Plant Cell*, **11**, 365–376.

Delis, I., Onken, A., Schyns, P.G., Panzeri, S. and Philiastides, M.G. (2016) Space-by-time decomposition for single-trial decoding of M/EEG activity. *Neuroimage*, **133**, 504–515.

Duren, Z., Chen, X., Zamanighomi, M., Zeng, W., Satpathy, A.T., Chang, H.Y., Wang, Y. and Wong, W.H. (2018) Integrative analysis of single-cell genomics data by coupled nonnegative matrix factorizations. *Proc. Natl. Acad. Sci. U. S. A.*, **115**, 7723–7728.

Ebied, A., Kinney-Lang, E., Spyrou, L. and Escudero, J. (2018) Evaluation of matrix factorisation approaches for muscle synergy extraction. *Med. Eng. Phys.*, **57**, 51–60.

Ifuku, K., Endo, T., Shikanai, T. and Aro., E-M. (2011) Structure of the chloroplast NADH dehydrogenase-like complex: Nomenclature and nuclear-encoded subunits. *Plant Cell Physiol.,***52,** 1560-1568.

Fitter, D.W., Martin, D.J., Copley, M.J., Scotland, R.W. and Langdale, J.A. (2002) *GLK* gene pairs regulate chloroplast development in diverse plant species. *Plant J.*, **31**, 713–727.

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

Gaujoux, R. and Seoighe, C. (2010) A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics*, **11**, 367.

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

Haberlandt, G. (1904) *Physiologische Pflanzenanatomie*, Leipzig, Germany: Verlag von Wilhelm Engelmann.

Hagen, G. and Guilfoyle, T.J. (1985) Rapid induction of selective transcription by auxins. *Mol. Cell. Biol.*, **5**, 1197–1203.

Hall, L.N., Rossini, L., Cribb, L. and Langdale, J.A. (1998) GOLDEN 2: a novel transcriptional regulator of cellular differentiation in the maize leaf. *Plant Cell*, **10**, 925–936.

Hall, N.P. and Keys, A.J. (1983) Temperature dependence of the enzymic carboxylation and oxygenation of ribulose 1,5-bisphosphate in relation to effects of temperature on photosynthesis. *Plant Physiol.*, **72**, 945–948.

Hartigan, J.A. and Wong, M.A. (1979) Algorithm AS 136: A K-Means Clustering

Algorithm. *J. R. Stat. Soc. Ser. C.*, **28**, 100–108.

Hasegawa, J., Sakamoto, Y., Nakagami, S., Aida, M., Sawa, S. and Matsunaga, S. (2016) Three-Dimensional imaging of plant organs using a simple and rapid transparency technique. *Plant Cell Physiol.*, **57**, 462–472.

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

Heo, J.-O., Chang, K.S., Kim, I.A., Lee, M.-H., Lee, S.A., Song, S.-K., Lee, M.M. and Lim, J. (2011) Funneling of gibberellin signaling by the GRAS transcription regulator SCARECROW-LIKE 3 in the *Arabidopsis* root. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 2166–2171.

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

Huang, C.-F., Yu, C.-P., Wu, Y.-H., Lu, M.-Y.J., Tu, S.-L., Wu, S.-H., Shiu, S.-H., Ku, M.S.B. and Li, W.-H. (2017) Elevated auxin biosynthesis and transport underlie high vein density in C_4 leaves. *Proc. Natl. Acad. Sci.*, **114**, E6884–E6891.

Hudson, D., Guevara, D.R., Hand, A.J., Xu, Z., Hao, L., Chen, X., Zhu, T., Bi, Y.-M. and Rothstein, S.J. (2013) Rice cytokinin GATA transcription Factor1 regulates chloroplast development and plant architecture. *Plant Physiol.*, **162**, 132–44.

Jordan, D.B. and Ogren, W.L. (1984) The CO_2/O_2 specificity of ribulose 1,5-bisphosphate carboxylase/oxygenase. *Planta*, **161**, 308–313.

Kong, W., Mou, X. and Hu, X. (2011) Exploring matrix factorization techniques for significant genes identification of Alzheimer's disease microarray gene expression data. *BMC Bioinformatics*, **12**, S7.

Kühn, K., Bohne, A.-V., Liere, K., Weihe, A. and Börner, T. (2007) *Arabidopsis* phage-type RNA polymerases: accurate in vitro transcription of organellar genes. *Plant Cell*, **19**, 959–971.

Külahoglu, C., Denton, A.K., Sommer, M., et al. (2014) Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae C_3 and C_4 plant species. *Plant Cell*, **26**, 3243–60.

Kümpers, B.M.C., Burgess, S.J., Reyna-llorens, I., Smith-unna, R., Boursnell, C. and Hibberd, J.M. (2017) Shared characteristics underpinning C_4 leaf maturation derived from analysis of multiple C_3 and C_4 species of. *J. Exp. Bot.*, **68**, 177–189.

Kyung Yoon, E., Dhar, S., Lee, M.-H., et al. (2016) Conservation and diversification of the *SHR-SCR-SCL23* regulatory network in the development of the functional endodermis in *Arabidopsis* shoots. *Mol. Plant*, **9**, 1197–1209.

Laing, W.A., Ogren, W.L. and Hageman, R.H. (1974) Regulation of soybean net photosynthetic CO_2 fixation by the interaction of CO_2, O_2 and Ribulose 1,5-Diphosphate Carboxylase. *Plant Physiol.*, **54**, 678–685.

Lee, D.D. and Seung, H.S. (1999) Learning the parts of objects by non-negative matrix factorization. *Nature*, **401**, 788–791.

Lim, J., Jung, J.W., Lim, C.E., Lee, M.-H., Kim, B.J., Kim, M., Bruce, W.B. and Benfey, P.N. (2005) Conservation and diversification of *SCARECROW* in maize. *Plant Mol. Biol.*, **59**, 619–630.

Lu, N. and Yin, T. (2015) Motor imagery classification via combinatory decomposition of ERP and ERSP using sparse nonnegative matrix factorization. *J. Neurosci. Methods*, **249**, 41–49.

Lundgren, M.R., Osborne, C.P. and Christin, P.-A. (2014) Deconstructing Kranz anatomy to understand C_4 evolution. *J. Exp. Bot.*, **65**, 3357–3369.

Mashiguchi, K., Tanaka, K., Sakai, T., et al. (2011) The main auxin biosynthesis pathway in *Arabidopsis*. *Proc. Nat. Acad. Sci.*, **108**, 18512–18517.

Mattsson, J., Sung, Z.R. and Berleth, T. (1999) Responses of plant vascular systems to auxin transport inhibition. *Development*, **126**, 2979–2991.

McKown, A.D. and Dengler, N.G. (2009) Shifts in leaf vein density through accelerated vein formation in C_4 *Flaveria* (Asteraceae). *Ann. Bot.*, **104**, 1085–1098.

McKown, A.D. and Dengler, N.G. (2010) Vein patterning and evolution in C_4 plants. *Botany*, **88**, 775–786.

Mitchison, G.J. (1981) The polar transport of auxin and vein patterns in plants. *Phil. Trans. R. Soc. Lond. B*, **295**, 461–471.

Nakamura, N., Iwano, M., Havaux, M., Yokota, A. and Munekage, Y.N. (2013) Promotion of cyclic electron transport around photosystem I during the evolution of NADP-malic enzyme-type C_4 photosynthesis in the genus *Flaveria*. *New Phytol.*, **199**, 832–842.

Nakazawa, M., Yabe, N., Ichikawa, T., Yamamoto, Y.Y., Yoshizumi, T., Hasunuma, K. and Matsui, M. (2001) *DFL1*, an auxin-responsive *GH3* gene homologue, negatively regulates shoot cell elongation and lateral root formation, and positively regulates the light response of hypocotyl length. *Plant J.*, **25**, 213–221.

Ogasawara, H., Kaimi, R., Colasanti, J. and Kozaki, A. (2011) Activity of transcription factor JACKDAW is essential for SHR/SCR-dependent activation of SCARECROW and MAGPIE and is modulated by reciprocal interactions with MAGPIE, SCARECROW and SHORT ROOT. *Plant Mol. Biol.*, **77**, 489–499.

Parry, M.A.J., Madgwick, P.J., Carvalho, J.F.C. and Andralojc, P.J. (2007) Prospects for increasing photosynthesis by overcoming the limitations of Rubisco. *J. Agric. Sci.*, **145**, 31–43.

Pascual-montano, A., Carazo, J.M., Kochi, K., Lehmann, D. and Pascual-marqui, R.D. (2006) Nonsmooth Nonnegative Matrix Factorization (nsNMF). *IEEE Trans. Pattern Anal. Mach. Intell.*, **28**, 403–415.

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

Rampey, R.A., LeClere, S., Kowalczyk, M., Ljung, K., Sandberg, G. and Bartel, B. (2004) A family of auxin-conjugate hydrolases that contributes to free indole-3-acetic acid levels during *Arabidopsis* germination. *Plant Physiol.*, **135**, 978–988.

Rolland-Lagan, A.-G. and Prusinkiewicz, P. (2005) Reviewing models of auxin canalization in the context of leaf vein pattern formation in *Arabidopsis*. *Plant J.*, **44**, 854–865.

Rossini, L., Cribb, L., Martin, D.J. and Langdale, J.A. (2001) The maize *Golden2* gene defines a novel class of transcriptional regulators in plants. *Plant Cell*, **13**, 1231–1244.

Sachs, T. (1969) Polarity and the induction of organized vascular tissues. *Ann. Bot.*, **33**, 263–275.

Sage, R.F. (2016) A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *J. Exp. Bot.*, **67**, 4039–4056.

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

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

Scarpella, E., Francis, P. and Berleth, T. (2004) Stage-specific markers define early steps of procambium development in *Arabidopsis* leaves and correlate termination of vein formation with mesophyll differentiation. *Development*, **131**, 3445–3455.

Schuler, M.L., Mantegazza, O. and Weber, A.P.M. (2016) Engineering C_4 photosynthesis into C_3 chassis in the synthetic biology age. *Plant J.*, **87**, 51–65.

Shao, C. and Höfer, T. (2017) Robust classification of single-cell transcriptome data by nonnegative matrix factorization. *Bioinformatics*, **33**, 235–242.

Sieburth, L.E. (1999) Auxin is required for leaf vein pattern in *Arabidopsis*. *Plant Physiol.*, **121**, 1179–90.

Slewinski, T.L. (2013) Using evolution as a guide to engineer kranz-type C_4 photosynthesis. *Front. Plant Sci.*, **4**, 212.

Slewinski, T.L., Anderson, A.A., Price, S., Withee, J.R., Gallagher, K. and Turgeon, R. (2014) Short-Root1 plays a role in the development of vascular tissue and kranz anatomy in maize leaves. *Mol. Plant*, **7**, 1388–1392.

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

Staswick, P.E., Serban, B., Rowe, M., Tiryaki, I., Maldonado, M.T., Maldonado, M.C. and Suza, W. (2005) Characterization of an *Arabidopsis* enzyme family that conjugates amino acids to indole-3-acetic acid. *Plant Cell*, **17**, 616–627.

Staswick, P.E., Tiryaki, I. and Rowe, M.L. (2002) Jasmonate response locus *JAR1* and several related *Arabidopsis* genes encode enzymes of the firefly luciferase superfamily that show activity on jasmonic, salicylic, and indole-3-acetic acids in an assay for adenylation. *Plant Cell*, **14**, 1405–1415.

Sud, R.M. and Dengler, N.G. (2000) Cell lineage of vein formation in variegated leaves of the C_4 grass *Stenotaphrum secundatum*. *Ann. Bot.*, **86**, 99–112.

Takase, T., Nakazawa, M., Ishikawa, A., Kawashima, M., Ichikawa, T., Takahashi, N., Shimada, H., Manabe, K. and Matsui, M. (2004) *ydk1-D*, an auxin-responsive *GH3* mutant that is involved in hypocotyl and root elongation. *Plant J.*, **37**, 471–483.

Usadel, Bj., Poree, F., Nagel, A., Lohse, M., Czedik-Eysenberg, A. and Stitt, M. (2009) A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize. *Plant. Cell Environ.*, **32**, 1211–1229.

Verna, C., Sawchuk, M.G., Linh, N.M. and Scarpella, E. (2015) Control of vein network topology by auxin transport. *BMC Biol.*, **13**, 94.

Walker, B.J., Vanloocke, A., Bernacchi, C.J. and Ort, D.R. (2016) The costs of photorespiration to food production now and in the future. *Annu. Rev. Plant Biol*, **67**, 107–29.

Wang, P., Fouracre, J., Kelly, S., et al. (2013) Evolution of *GOLDEN2-LIKE* gene function in C_3 and C_4 plants. *Planta*, **237**, 481–495.

Wang, P., Karki, S., Biswal, A.K., et al. (2017a) Candidate regulators of early leaf development in maize perturb hormone signalling and secondary cell wall formation when constitutively expressed in rice. *Sci. Rep.*, **7**, 4535.

Wang, P., Kelly, S., Fouracre, J.P. and Langdale, J.A. (2013) Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C_4 Kranz anatomy. *Plant J.*, **75**, 656–670.

Wang, P., Khoshravesh, R., Karki, S., et al. (2017b) Re-creation of a key step in the evolutionary switch from C_3 to C_4 leaf Anatomy. *Curr. Biol.*, **27**, 3278–3287.

Waters, M.T. and Langdale, J.A. (2009) The making of a chloroplast. *EMBO J.*, **28**, 2861–2873.

Waters, M.T., Wang, P., Korkaric, M., Capper, R.G., Saunders, N.J. and Langdale, J.A. (2009) GLK transcription factors coordinate expression of the photosynthetic apparatus in *Arabidopsis*. *Plant Cell*, **21**, 1109–1128.

Welch, D., Hassan, H., Blilou, I., Immink, R., Heidstra, R. and Scheres, B. (2007) *Arabidopsis* JACKDAW and MAGPIE zinc finger proteins delimit asymmetric cell division and stabilize tissue boundaries by restricting SHORT-ROOT action. *Genes Dev.*, **21**, 2196–2204.

Wysocka-Diller, J.W., Helariutta, Y., Fukaki, H., Malamy, J.E. and Benfey, P.N. (2000) Molecular analysis of *SCARECROW* function reveals a radial patterning mechanism common to root and shoot. *Development*, **127**, 595–603.

Xu, D., Jiang, Y., Li, J., Lin, F., Holm, M. and Deng, X.W. (2016) BBX21, an *Arabidopsis* B-box protein, directly activates *HY5* and is targeted by COP1 for 26S proteasome-mediated degradation. *Proc. Natl. Acad. Sci. U. S. A.*, **113**, 7655–60.

Zhu, L., Xin, R., Bu, Q., Shen, H., Dang, J. and Huq, E. (2016) A Negative feedback loop between PHYTOCHROME INTERACTING FACTORs and HECATE proteins fine-tunes photomorphogenesis in *Arabidopsis*. *Plant Cell*, **28**, 855-874.

(VIII) Supplementary Information

Supplementary Figure 1 – Cross sections of the developmental leaf gradients from *F. bidentis* and *F. robusta*.

Supplementary Figure 2 – Determination of the area number for nonnegative matrix factorization.

Supplementary Figure 3 – Factorization of the dataset with primordia into five areas.

Supplementary Figure 4 – k-means clusters on factorized gene expression for the dataset without leaf primordia.

Supplementary Figure 5 – Heatmap of the genes involved in the C_4 cycle.

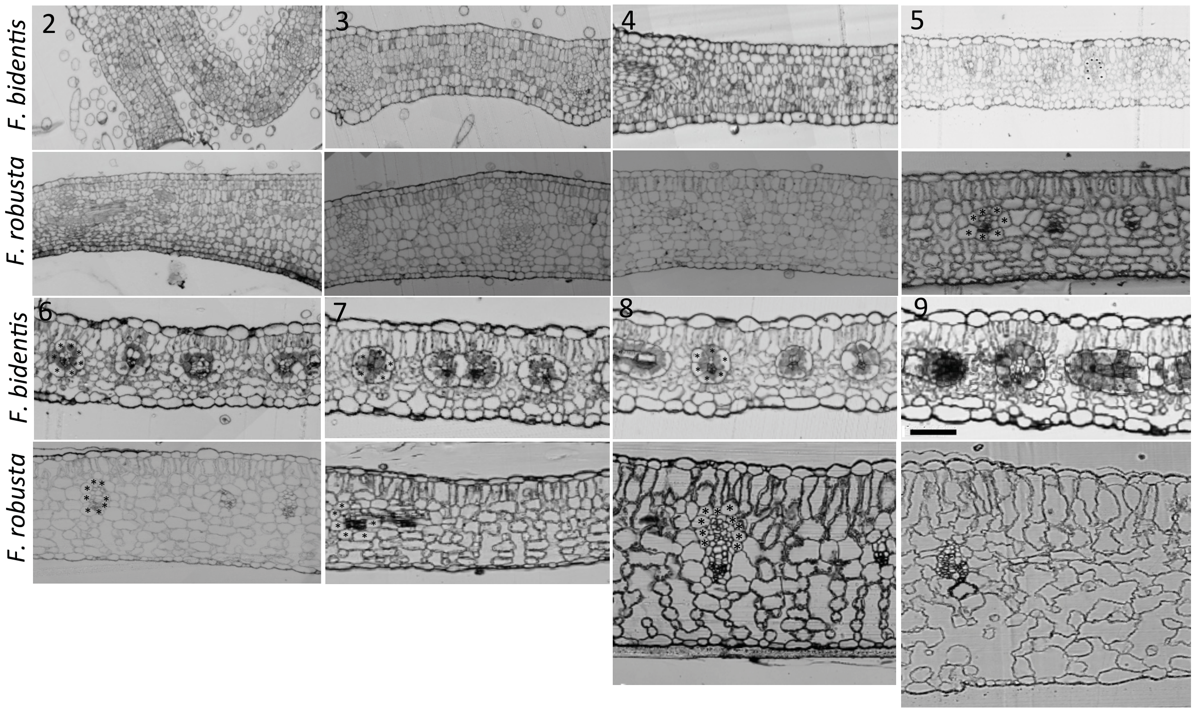
Supplementary Figure 6 – Heatmap of the genes related to cyclic electron transport.

Supplementary Figure 7 – Heatmap of the genes involved in photorespiration.

Supplementary Figure 8 – Heatmap of the genes related to Auxin signaling pathway.

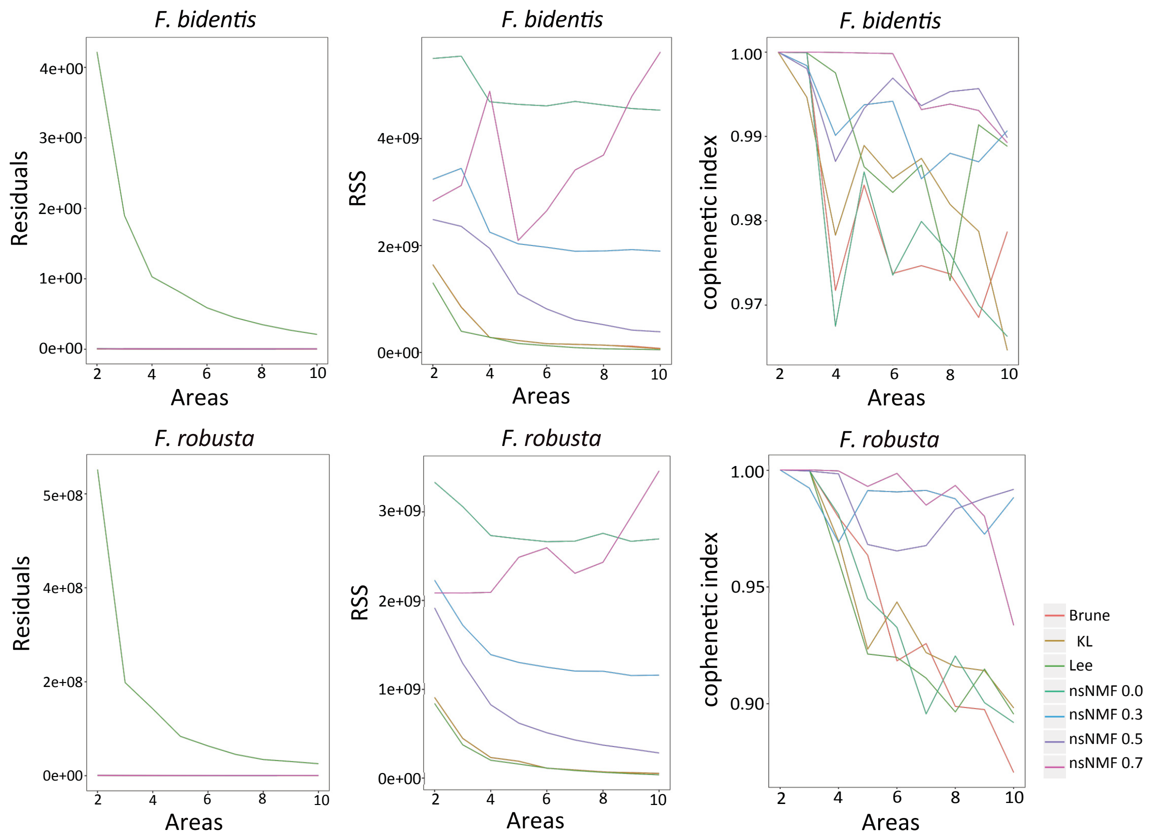
Supplementary Figure 9 – Heatmap of Auxin induced genes.

Supplementary Figure 10 – Heatmap of the genes related to bundle sheath development

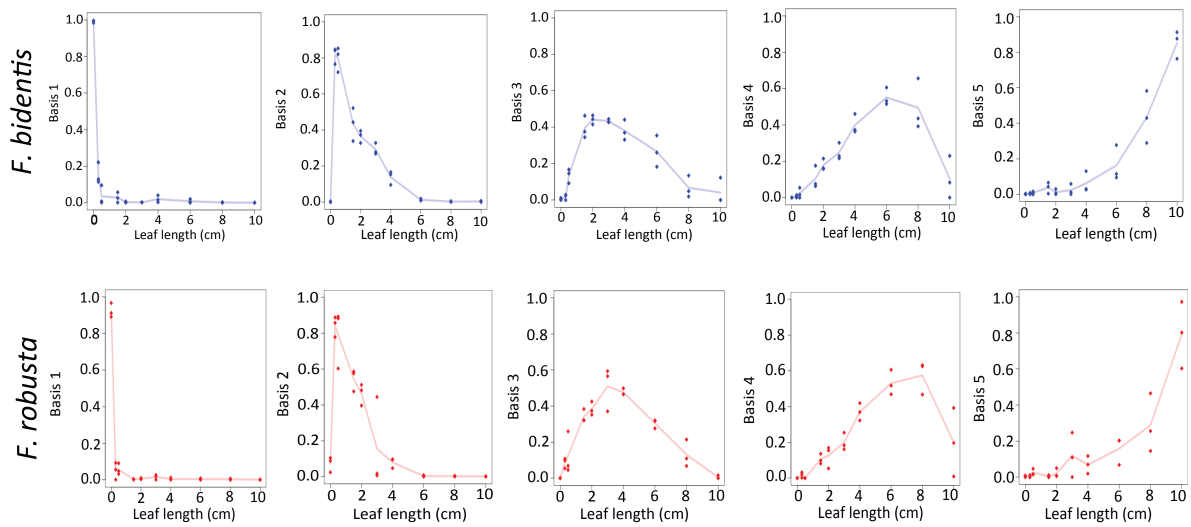


Supplementary Figure 1. Cross sections of the developmental leaf gradients from *Flaveria bidentis* and *Flaveria robusta*.

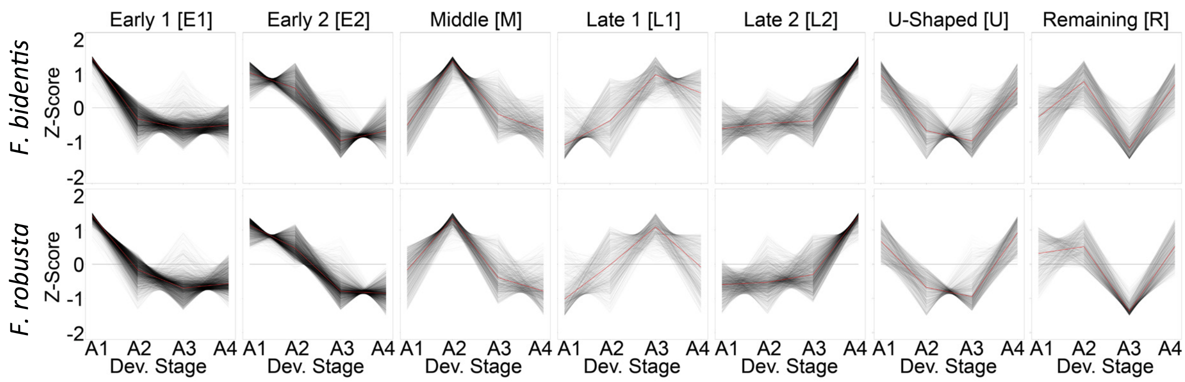
Cross sections were taken in the third quarter of the leaf for every stage in the developmental gradient. The numbers represent the developmental stage with the corresponding leaf lengths of 0.5 cm, 1.5 cm, 2.0 cm, 3.0 cm, 4.0 cm, 6.0 cm, 8.0 cm, and 10.0 cm. The *F. bidentis* section is depicted at the top and *F. robusta* at the bottom of each pair. Scale bar – 10µm.



Supplementary Figure 2. Determination of the area number for nonnegative matrix factorization. The quality of NMF factorizations for two to ten factors was determined using the sum of residuals (**A**), the sum of the square of residuals (**B**) and the cophenetic index (**C**).

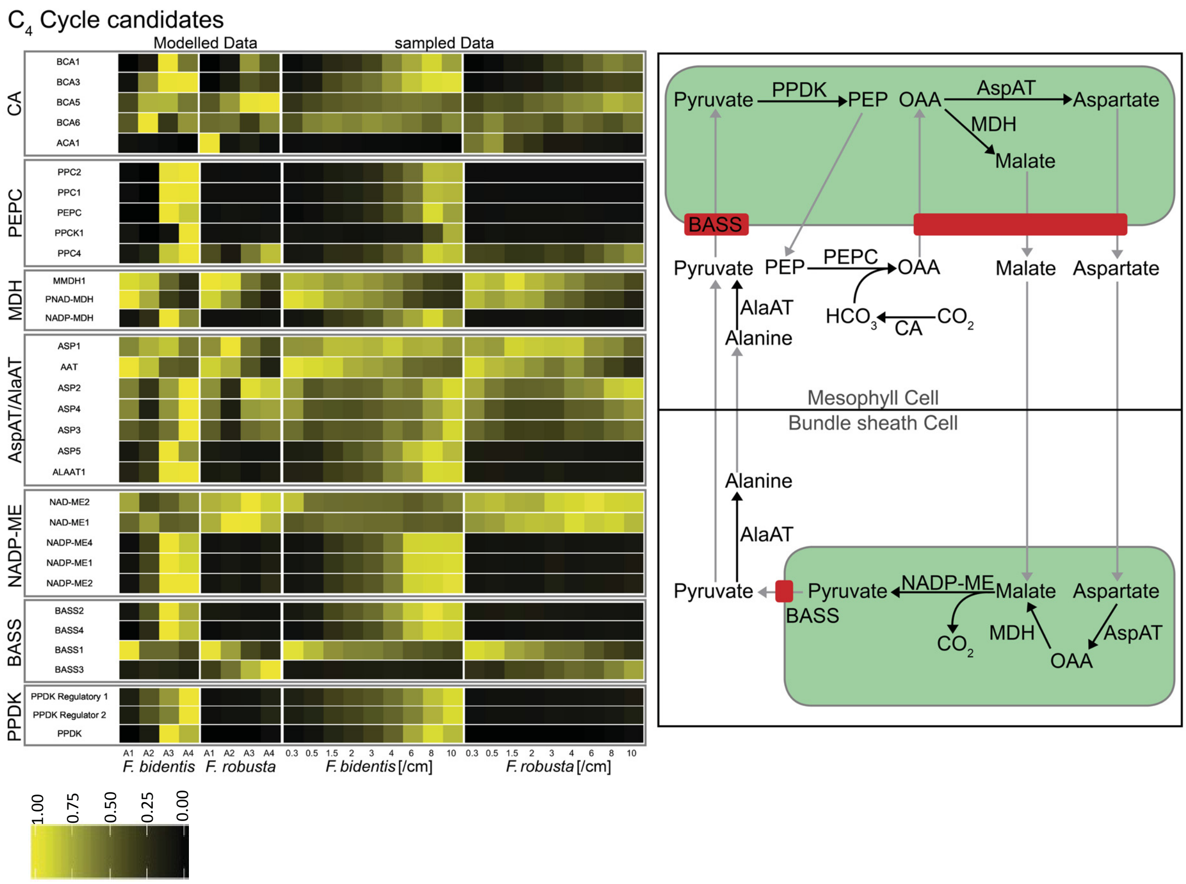


Supplementary Figure 3. Factorization of the dataset with primordia into 5 areas.
The dataset containing the primordia samples was factorized into five areas. The relative proportions of an area contributing to the sample set for both species is presented. The samples are sorted according to their corresponding leaf length with an assigned length of 0 cm for primordia.

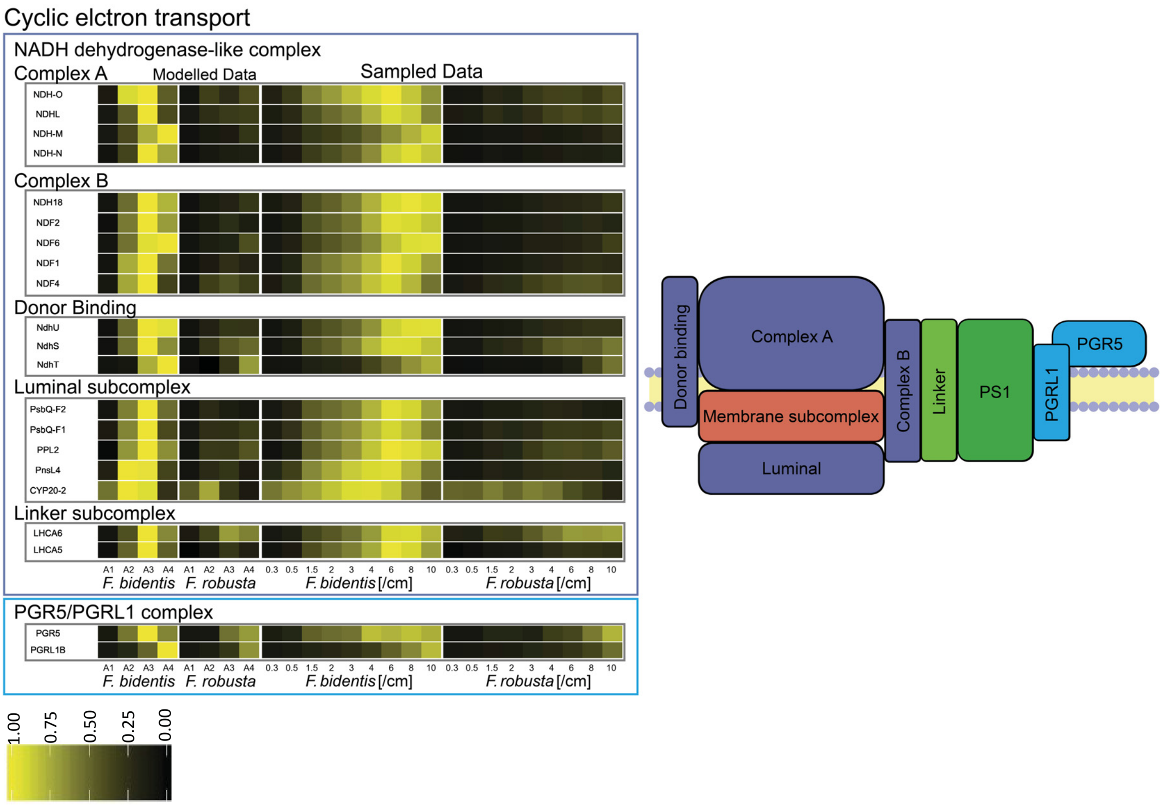


Supplementary Figure 4. k-means clusters based on factorized gene expression for the dataset without primordia.

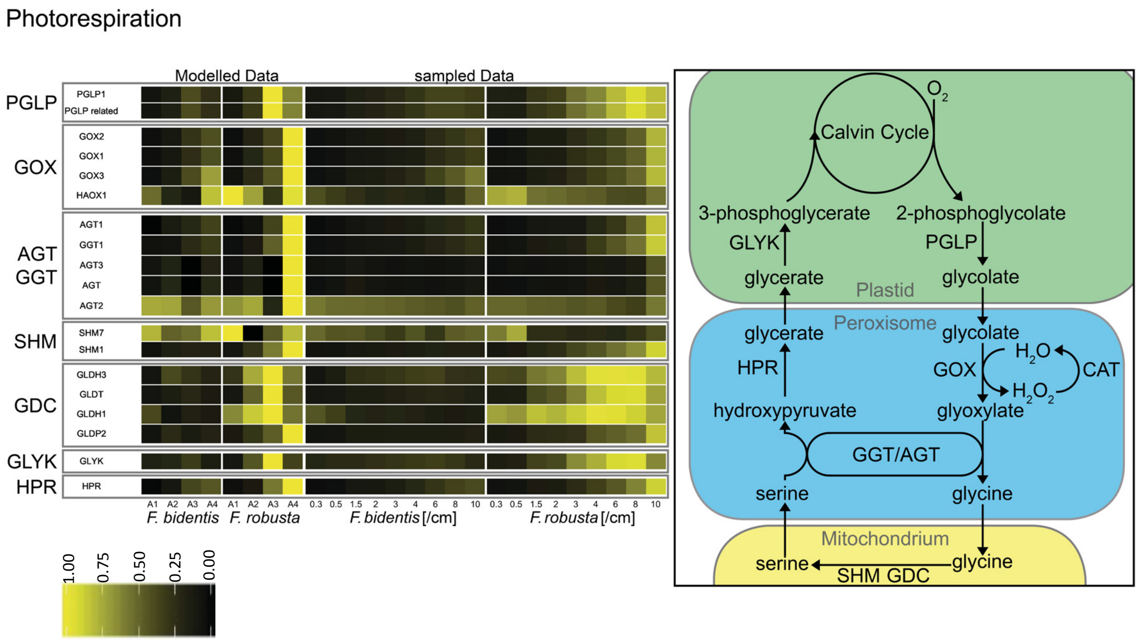
The dataset without the primordia samples was factorized into four areas and k-means clustered based on the intensity of the area specific expressions into seven clusters.



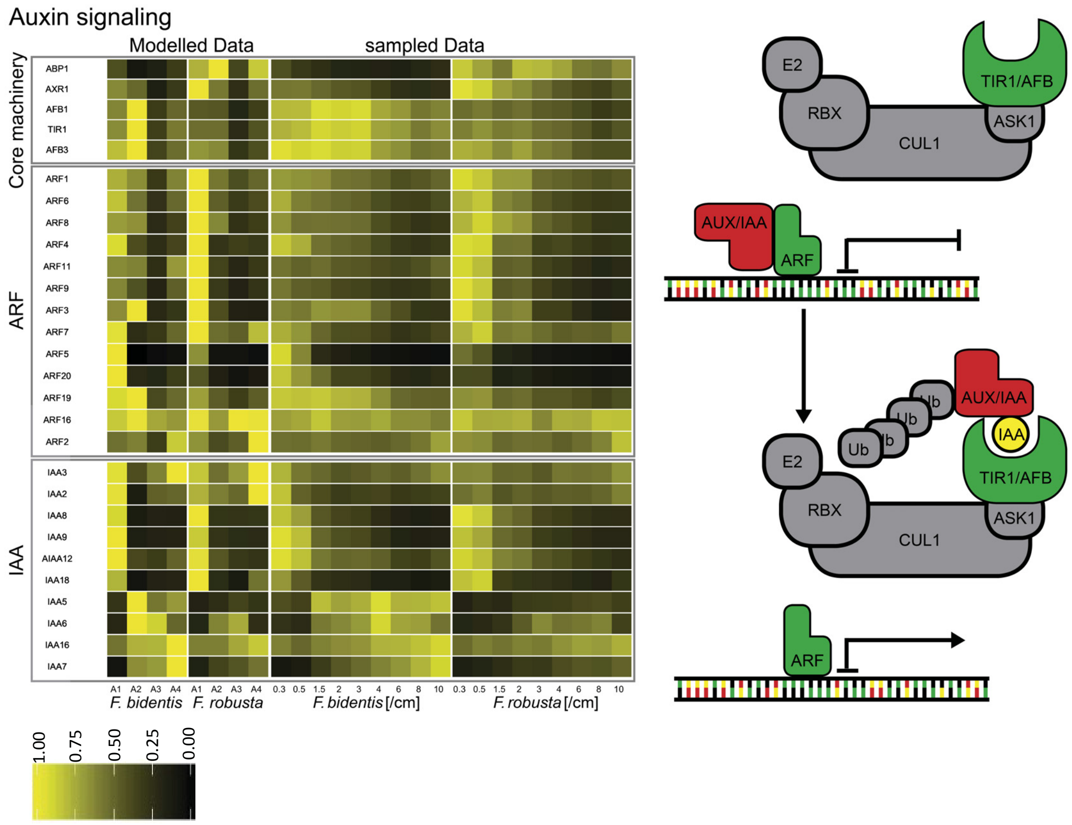
Supplementary Figure 5. Heatmap of the genes involved in the C_4 cycle.
CA – CARBONIC ANHYDRASE; *BCA – BETA CARBONIC ANHYDRASE*; *ACA – ALPHA CARBONIC ANHYDRASE*; *PEPC/PPC – PHOSPHOENOLPYRUVATE CARBOXYLASE*; *PPCK1 – PHOSPHOENOLPYRUVATE CARBOXYKINASE*; *MDH – MALATE DEHYDROGENASE*; *MMDH – MITOCHONDRIAL MALATE DEHYDROGENASE* (NAD-dependent); *PNAD-MDH – PEROXISOMAL NAD-MALATE DEHYDROGENASE*; *AspAT – ASPARTATE AMINOTRANSFERASE*; *AlaAT – ALANINE AMINOTRANSFERASE*; *NADP-ME – NADP-MALIC ENZYME*; *BASS – BILE ACID:SODIUM SYMPORTER* and *PPDK – PYRUVATE, ORTHOPHOSPHATE DIKINASE*. The diagrammatic representation of the NADP-ME subtype C_4 cycle (right panel).



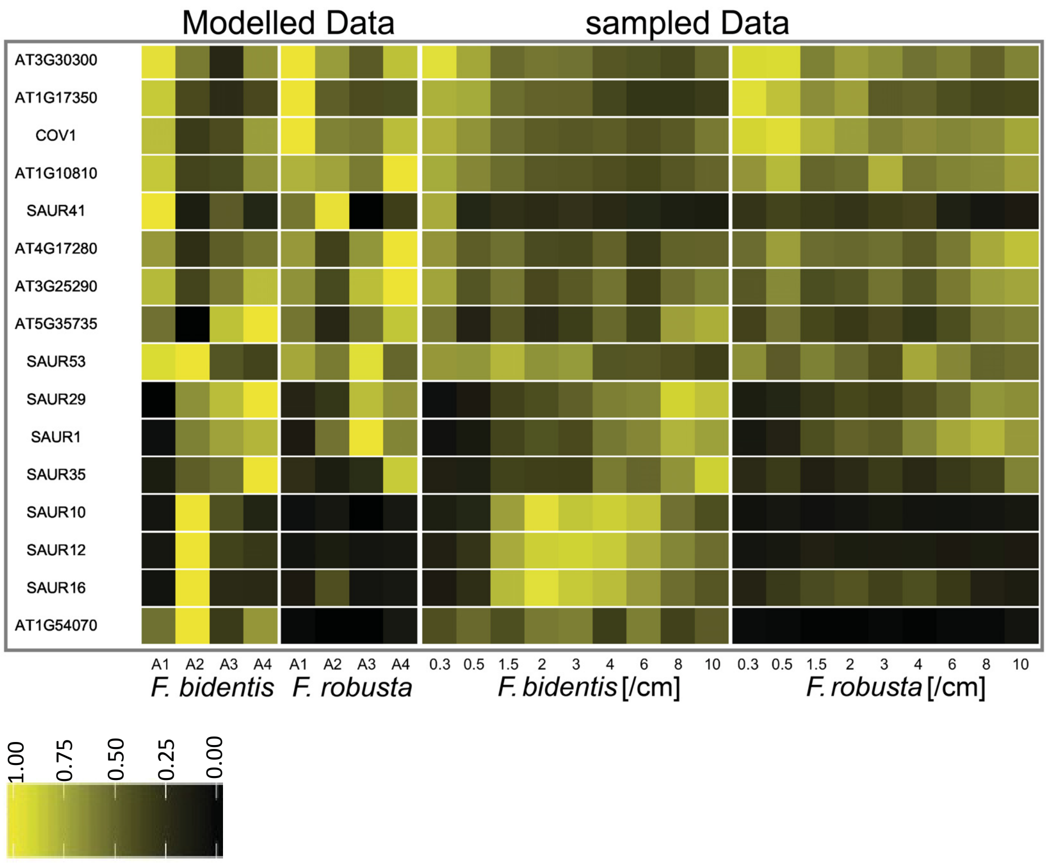
Supplementary Figure 6. Heatmap of the genes related to cyclic electron transport.
NDH – NAD(P)H DEHYDROGENASE; *NDF – NDH-DEPENDENT CYCLIC ELECTRON FLOW 1*; *PPL2 – PSB-LIKE PROTEIN 2*; *LHCA5* and *LHCA6* – components of the light harvesting complex of PSI; PSI – photosystem I; *PGR5 – PROTON GRADIENT REGULATION 5*; *PGRL1 – PROTON GRADIENT REGULATION 5-LIKE 1*. The structure of the NDH complex is depicted (right panel; Source: adapted from Ifuku *et al.,* 2011) and *PGR5*, *PGRL1* are involved in ferredoxin:plastoquinone oxidoreductase (FQR) dependent cyclic electron transport.



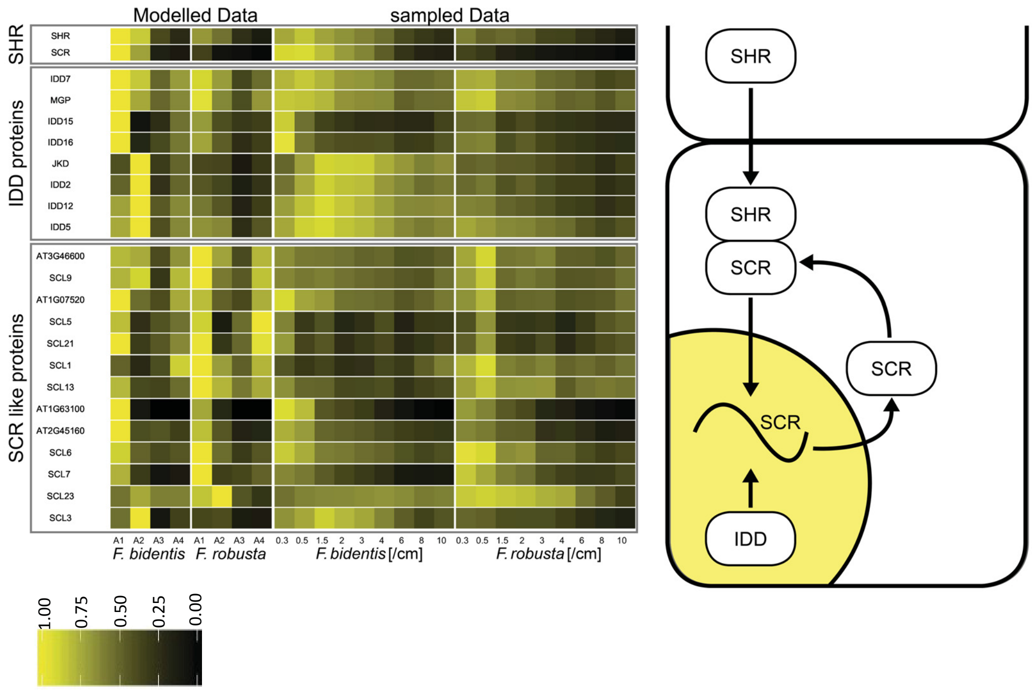
Supplementary Figure 7. Heatmap of the genes involved in photorespiration.
PGLP – PHOSPHOGLYCOLATE PHOSPHATASE; GOX – GLYCOLATE OXIDASE; HAOX1 – 2-HYDROXY-ACID OXIDASE; CAT – CATALASE; AGT – ALANINE:GLYOXYLATE AMINOTRANSFERASE; GGT – GLUTAMATE:GLYOXYLATE AMINOTRANSFERASE; SHM – SERINE HYDROXYMETHYLTRNASFERASE; GDC – GLYCINE DECARBOXYLASE COMPLEX; GLYK – GLYCERATE KINASE; HPR – HYDROXYPYRUVATE REDUCTASE. The photorespiration pathway is depicted (right panel).



Supplementary Figure 8. Heatmap of the genes related to Auxin signaling pathway.
ABP1 – ENDOPLASMIC RETICULUM AUXIN BINDING PROTEIN 1; *AXR1 – AUXIN RESISTANT 1*; *TIR1 – TRANSPORT INHIBITOR RESPONSE 1*; *AFB3 – AUXIN SIGNALING F-BOX 3*; *ARF – AUXIN RESPONSIVE FACTOR*; *IAA – INDOLE-3-ACETIC ACID* (Auxin); Diagrammatic representation of the Auxin signaling pathway (right panel) CUL1 – CULLIN 1; E2 – UBIQUITIN-CONJUGATING ENZYME; RBX – RING-BOX PROTEIN; *ASK1* – SKP1 homolog in *Arabidopsis*; Ub – UBIQUITIN. CUL1, E2, RBX, ASK1 – components of the auxin induced E3 ubiquitin ligase complex.



Supplementary Figure 9. Heatmap of the genes involved in Auxin induced pathway.
COV1 – CONTINUOUS VASCULAR RING; *SAURs – SMALL AUXIN UP RNAs*.



Supplementary Figure 10. Heatmap of the genes related to bundle sheath development. *SHR – SHORT-ROOT*; *SCR – SCARECROW*; *IDD – INTERMEDIATE DOMAIN* family genes; *JKD – JACKDAW*; *MGP – MAGPIE; SCL – SCARECROW-LIKE* genes. AT3G46600, AT1G07520, AT1G63100, and AT2G45260 - *SCARECROW* transcription factor family proteins. The *SCARECROW*/*SHORT-ROOT*/*IDD* regulation mechanism was depicted in the right panel.

Supplementary Table 1

The number of clean reads, average read count per sample and average mapping efficiency to *A. thaliana*. Fb – *F. bidentis*; Fr – *F. robusta.* P – primordia; 1 to 9 – Leaf developmental series.

Sample	Average number of celan reads/sample	Number of reads mapped to the *A. thaliana* reference	Mapping efficiency
Fb_P	33965975	18096619	53.3
Fb_1	31941648	14676635	45.9
Fb_2	33154082	17333775	52.3
Fb_3	33965975	18096619	53.3
Fb_4	32749149	17903680	54.7
Fb_5	33395773	18348598	54.9
Fb_6	33350589	18480253	55.4
Fb_7	32888130	18674681	56.8
Fb_8	33224947	18685086	56.2
Fb_9	29627029	15911848	53.7
Fr_P	30636725	14801778	48.8
Fr_1	33732554	17818741	52.8
Fr_2	33291967	17193595	51.6
Fr_3	32498576	18242083	56.1
Fr_4	34885421	19338385	55.4
Fr_5	34418159	16117395	46.8
Fr_6	35002125	19879212	56.8
Fr_7	35296555	20094107	56.9
Fr_8	35339924	20392894	57.7
Fr_9	28155274	15323716	54.4

(IX) Author contributions

KB generated the transcriptome dataset.

TJW assessed the anatomy of leaf developmental gradients and performed the deconvolution of the transcriptome dataset.

TJW performed auxin spray experiments.

KB and TJW analyzed the data.

UG and **AB** helped with data analysis.

KB, TJW and **UG** wrote the manuscript.

AB, **A.P.M.W** and **PW** will participate in drafting of the manuscript.

Manuscript II

Reporter-based forward genetic screen to identify bundle sheath anatomy mutants in *A. thaliana*

Reporter-based forward genetic screen to identify bundle sheath anatomy mutants in *A. thaliana*

Florian Döring[1,a]*, Kumari Billakurthi[1,2,a], Udo Gowik[1,3], Stefanie Sultmanis[4], Roxana Khoshravesh[4], Shipan Das Gupta[1], Tammy L. Sage[4] and Peter Westhoff[1,2]

[1]Institute of Plant Molecular and Developmental Biology, Universitaetsstrasse 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 Biology and Environmental Sciences, Ammerlaender Heerstrasse 114, Carl Von Ossietzky University, 26129 Oldenburg, Germany.
[4]Department of Ecology and Evolutionary Biology, The University of Toronto, Toronto, ON, Canada M5S 3B2
a – These authors contributed equally to this work.
*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: udo.gowik@uni-oldenburg.de
Stefanie Sultmanis
e-mail: stefanie.sultmanis@mail.utoronto.ca
Roxana Khoshravesh
e-mail: r.khoshravesh@utoronto.ca
Shipan Das Gupta
e-mail: shipan.gupta@hhu.de
Tammy Sage
e-mail: tammy.sage@utoronto.ca
Peter Westhoff
e-mail: west@hhu.de

(I) Summary

The evolution of C_4 photosynthesis proceeded stepwise with each small step increasing the fitness of the plant. An important precondition for the introduction of a functional C_4 cycle is the photosynthetic activation of the C_3 bundle sheath by increasing its volume and organelle number. To engineer C_4 photosynthesis into existing C_3 crops, information about genes controlling bundle sheath cell size and organelle content is therefore needed. However, very little is known about the genes that could be manipulated in order to create a more C_4–like bundle sheath. To this end, we established an EMS-based forward genetic screen in the Brassicacean C_3 species *Arabidopsis thaliana*. To ensure a high–throughput primary screen, the bundle sheath cells of *A. thaliana* were labeled by a luciferase (LUC68) or by a chloroplast-targeted green fluorescent reporter protein (sGFP) using a bundle sheath specific promoter. The signal strength of the reporter genes was used as a proxy to search for mutants with altered bundle sheath anatomy. Here we show that our genetic screen identified predominantly mutants that were primarily affected in the architecture of the vascular bundle, which secondarily lead to an increase in bundle sheath volume. By using the mapping-by-sequencing approach we could identify the genomic segments containing mutated candidate genes.

Keywords: C_4 photosynthesis, EMS mutagenesis, bundle sheath cell, *Arabidopsis thaliana*, *GFP*, *LUC*

(II) Significance statement

We present a robust method using *Arabidopsis thaliana* as a model system to efficiently induce mutants affected in bundle sheath anatomy by ethyl methane sulfonate and to map the responsible genomic region by genotyping by sequencing. The method's key feature is the use of easily detectable reporter genes for the primary screening of mutant candidates. We believe that this method is generally applicable in searching for mutants that are affected in inner leaf anatomy.

(III) Introduction

C_4 plants surpass C_3 species in their photosynthetic performance under conditions of high light, hot temperatures and drought (Ehleringer *et al.*, 1991). This is due to their unique mode of photosynthesis that is characterized by a division of labor between two different cell leaf types, mesophyll and, most commonly, bundle sheath cells (Edwards and Voznesenskaya, 2011) The two cell types are arranged in a wreath-like structure around the vasculature termed Kranz anatomy (Haberlandt, 1904) and build a single integrated metabolic system (Hatch, 1987). Atmospheric CO_2 is first fixed in the mesophyll cells by phospho*enol*pyruvate carboxylase, an oxygen-insensitive carboxylase. The resulting C_4 acid is then transported into the bundle sheath cells where it is decarboxylated by one or a combination of NADP-, NAD-dependent malic enzyme and phospho*enol*pyruvate carboxykinase (Furbank, 2011; Wang *et al.*, 2014). The released CO_2 is thereby concentrated at the site of Ribulose-1, 5-bisphosphate carboxylase/oxygenase and finally channeled into the Calvin-Benson cycle. Due to this CO_2 pumping mechanism photorespiration is largely abolished resulting in the superior photosynthetic efficiency of C_4 plants (Zhu *et al.*, 2010).

C_4 photosynthesis occurs only in the angiosperms (Ehleringer *et al.*, 1997) and has evolved more than sixty times independently (Sage *et al.*, 2011; Sage, 2016). The polyphyletic origin of the C_4 photosynthetic pathway indicates that, from the genetic point of view, it must have been relatively easy to evolve a C_4 from a C_3 species. Indeed quantitative modeling showed that C_4 evolution proceeded step by step and that each of the small evolutionary changes contributed to increasing the general fitness of the plant (Heckmann *et al.*, 2013).

Because of the high photosynthetic performance and the possible application of this knowledge in plant breeding the molecular genetics and evolutionary basis of C_4 photosynthesis has been studied intensively in the last decade (von Caemmerer *et al.*, 2012). If existing C_3 crops, such as rice or wheat, could be converted, by genetic engineering, to operate a C_4 photosynthetic pathway the predicted enhancement in photosynthetic efficiency could possibly be used to boost crop yields (Sheehy *et al.*, 2007; Zhu *et al.*, 2010).

The lack of an appropriate C_4-like bundle sheath in C_3 plants is one major obstacle that has to be overcome in this endeavor. While the bundle sheath of C_4 species is enlarged and rich in chloroplasts and mitochondria, the corresponding tissue of C_3 species is usually not very prominent and relatively poor in organelles (Sage *et al.*, 2014). This indicates 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 well understood. It is assumed they function in phloem loading and unloading and contribute to the mechanical support of the leaf (Van Bel 1993; Kinsman & Pyke 1998; Griffiths *et al*., 2013). Transcript profiling of *Arabidopsis thaliana* bundle sheath cells indicated that the bundle sheath tissue, at least in Brassicacean species, is highly active in sulfur and glucosinolate metabolism (Aubry *et al*., 2014). The cross-species expression specificity of the bundle sheath specific promoter of the gene encoding the P-subunit of glycine decarboxylase (*GLDPA*) from the Asteracean C_4 species *Flaveria trinervia* showed that bundle sheath specific expression was maintained in the C_3 species *Arabidopsis thaliana* (Engelmann *et al*., 2008; Wiludda *et al*., 2012). Conversely, the bundle sheath specific expression of the promoter of the sulfate transporter gene *SULTR2;2* from *A. thaliana* (Takahashi *et al.*, 2000) maintained its bundle sheath specificity in the Asteracean C_4 species *Flaveria bidentis* (Kirschner *et al.*, 2018) These findings indicated that the transcription-regulatory system of bundle sheath cells, i.e. the interplay of *cis*-regulatory elements with their cognate transcription factors, is at least partially conserved in dicotyledonous angiosperms and that a cryptic Kranz anatomy is already present in C_3 species (Westhoff and Gowik, 2010).

Mutant analysis (Slewinski *et al*., 2012; Slewinski *et al*., 2014) and transcript profiling experiments (Wang *et al*., 2013) with maize indicated that the *SHORTROOT-SCARECROW* transcriptional regulatory module (Sparks *et al*., 2016) is not only a key component in root radial patterning (Petricka *et al*., 2012), but also regulating the establishment of Kranz anatomy (Slewinski, 2013; Fouracre *et al*., 2014). In addition, the GOLDEN2-like transcriptional regulator proteins play a role and can induce a C_3 to C_4 switch in bundle sheath characteristics (Wang *et al*., 2017).

Forward genetic screens proved to be powerful and unbiased tools to dissect biological processes and identify their underlying genes and regulatory networks. Here, we present the design of a simple screening method on the tractable genetic model plant *A. thaliana* with the aim to identify bundle sheath developmental genes which are involved in the ontogeny and the functional maintenance of the bundle sheath and might be possible engineering targets to the increase size and organelle number of this cell type. A successful forward genetic screen is defined by a high throughput and a reliable, robust primary screen for mutants in which thousands of plants have to be analyzed (Page and Grossniklaus, 2002). Since the bundle sheath of *A. thaliana* is not very prominent and its cells contain only a few chloroplasts, we used reporter genes to label the bundle sheath or its chloroplasts. *Arabidopsis* lines expressing these reporter genes should allow an easy primary screen for mutants that were potentially affected in bundle sheath size or chloroplast numbers. To this end, we used the *GLDPA* promoter of *F. trinervia* which is active in the bundle sheath but not in the mesophyll of *Arabidopsis thaliana* (Engelmann *et al.*, 2008) to drive the expression of reporter genes that encode either the firefly luciferase 68 (LUC68) or a chloroplast-targeted green fluorescent protein (sGFP). *Arabidopsis* lines homozygous for the p*GLDPA*$_{\mathrm{Ft}}$::*LUC68* or the p*GLDPA*$_{\mathrm{Ft}}$::*sGFP* reporter genes were generated and mutagenized with the chemical mutagen ethyl methanesulfonate (EMS). The level of reporter gene expression served as a proxy to collect mutants with altered bundle sheath anatomy. Mutant lines that contain intact reporter gene and whose reporter expression deviated strongly from the non-mutagenized reference lines were then further analyzed by light and electron microscopy for alterations in bundle sheath anatomy.

(IV) Results

Design of the mutant screen

All mutants had to be viable and sustain autotrophic growth in soil because we did not intend to keep the seeds of the M1 plants (= M2 seeds) separately for maintaining the mutant alleles. Moreover, we aimed at designing a non-destructive screen with a robust and quick and quantifiable detection method. In a second step, identified primary mutants were analyzed by light microscopy to reveal alterations in bundle sheath anatomy. To facilitate this step, the bundle sheath cells of *A. thaliana* were labeled by a reporter gene that allowed a non-destructive and large-scale phenotyping of segregating M2 populations. We generated a *LUC* and chloroplast-targeted *GFP* reporter lines (Figure 1) by expressing the respective genes under the control of the 1571-bp 5'-flanking region of the glycine decarboxylase P protein gene (*GLDPA*) of the C_4 Asteracean species *F. trinervia* (Figure 1a). This promoter region is highly active in bundle sheath cells and vascular tissue of *A. thaliana* (Wiludda *et al.*, 2012; Engelmann *et al.*, 2008). Both reporter gene constructs were transformed into *A. thaliana* (Ecotype Columbia-0). Transgenic lines with high reporter gene activity relative to the control/reference line were identified, and homozygous lines were obtained by selfing. Figure 1b and c shows the expression pattern of both reporter genes in leaves of *Arabidopsis* plants. The *GFP* reporter gene was targeted to the chloroplasts of the bundle sheath and vascular tissue cells by adding the transit peptide of rubisco small subunit (Kim *et al.*, 2010) to the gene. This allowed to differentiate between single chloroplasts within the cells and is shown in leaf transverse sections (Figure 1d).

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

Approximately 160,000 seeds were mutagenized with EMS (40,000 *LUC* reporter gene line seeds; 120,000 *GFP* reporter gene line seeds) and sown in the soil in large trays under greenhouse conditions. A survival rate of about 50 % was observed in the M1 generation and seeds were harvested in pools of 30–50 plants from the remaining 80,000 M1 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 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).

The general workflow of the EMS-based genetic screen, which aims at identifying mutants altered in bundle sheath anatomy and function, is depicted in Figure 2. Seeds from each M2 pool were sown individually on large trays 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). In total, 755 primary mutants were identified; 258 mutants with the *LUC* background and 497 mutants with the *GFP* background. The phenotype of each mutant line was assessed in the following M3 generation for its stability, whereby only mutant lines with a strong deviation in reporter gene signal intensity compared to the reference line were selected. Thereafter, 85 mutant lines with a *LUC* background and 145 mutants with a *GFP* background remained, reporter gene sequences were amplified via PCR and sequenced in the background of every identified mutant in order to exclude aberrant phenotypes that are based on reporter gene mutations. Almost 75 % of the mutant lines had to be discarded due to reporter gene mutations. Nevertheless, twelve mutant lines with the *LUC* reporter gene and 45 mutant lines with the *GFP* reporter gene were both stable and contained intact reporter genes.

The EMS-based *LUC* reporter screen resulted in six mutant lines with increased and six mutant lines with decreased reporter gene activity (Figure 3b and c) while the GFP reporter screen resulted in 22 and 19 lines with increased and decreased reporter gene activity relative to control/reference lines, respectively (Figure 3e and f). Moreover, four mutant lines possessed a diffused GFP signal in which the reporter gene signal was clearly detectable in the mesophyll tissue (Figure 3g). Figure S1 shows the relative LUC and GFP signal intensity of all mutant lines. Some of the mutants with a diffused GFP signal were kept for further analyses 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 (Engelmann *et al.*, 2008; Wiludda *et al.*, 2012). Intriguingly, seven mutants with increased GFP signal intensity also contained bundle sheath strands (vascular tissue plus bundle sheath) with an increase in diameter in comparison to the reference line (Figure 3h and i), which might be caused by an increase in either vascular tissue or bundle sheath tissue, or a combination of both.

Microscopic analysis of EMS-generated mutant lines

Our primary screening criterion was based on the reporter gene activity and therefore, we could not clearly assign changes in reporter gene expression to anatomical alterations of bundle sheath cells in identified mutants. To address this question, we selected 27 mutant lines from the primary screen with the strongest phenotypes in terms of signal intensity and width of the bundle sheath strands (G01–G25 and L01, L02) (Table S1) to image 1.5 µm thin sections of resin-embedded leaf tissue with the light microscope. Among this subset of mutant lines, twenty mutant lines possessed an increased GFP signal, three mutant lines showed less GFP signal, and two mutant lines exhibited a diffused GFP reporter signal. In addition, two mutant lines of the *LUC* reporter screen with an increased reporter signal were included in the survey.

Transverse sections of each replicate were compared to those of the reference line, and the 3° higher-order veins were analyzed with respect to the anatomy of the bundle sheath and vascular tissue. Five mutant lines (G14, G15, G17, G19, and G20) were identified, whose bundle sheath tissue contained more cells in the radial direction as compared to the reference line (Figure 4a versus b-f), although no differences could be found in paradermal view (Figure S2). Mutant lines G14, G15, G17, and G19 showed an increased number of chloroplast containing cells within the phloem tissue (companion cell and vascular parenchyma; Maeda *et al.*, 2008) as well as an increase in sieve elements relative to the reference line (Figure 4a versus b-e). Mutant line G20 showed an amplification of the tracheary elements (Figure 4f).

Due to expected chloroplast targeting of the *GFP* reporter, we hypothesized that changes in GFP signal intensity in identified mutant lines might derive from increased chloroplast number, size, or structure. However, further analyses of selected mutant lines revealed that none indicated a difference in chloroplast number or sizes in bundle sheath cells.

Analyses of all chloroplast-containing cells of nine mutant lines (G10, G13, G14, G18, G19, G20, G22, G23, and G25) and the reference line with transmission electron microscopy (TEM) indicated no obvious changes in chloroplast ultrastructure with the exception of line G19. In this line, the majority of mesophyll and bundle sheath cell chloroplasts contained prominent nucleoids and two to four grana with extensive stacking of long thylakoids (Figure S3 e, g, h, i, and j). This phenotype was more prominent in mesophyll cells. The lumens of the grana thylakoids in line G19 were narrow and disorganized relative to the reference line (Figure S3 j versus i). Some chloroplasts also contained numerous vesicles (Figure S3h). Numerous prominent nucleoids that contain chloroplast DNA and plastid nucleoid associated proteins (Powikrowska *et al.*, 2014) were associated with grana and vesicles (Figure S3, panels e, g, h, j).

Leaf morphology and growth characteristics of mutant lines G14, G15, G17, G19, and G20

Mutant lines G14, G19, and G20 showed impaired leaf morphology and growth characteristics relative to the reference line. The growth of the mutant lines was strongly reduced and their leaves were smaller in size compared to those of the reference line. Furthermore, in mutant line G20 the first leaf pair, but not the cotyledons, displayed a partial reticulate leaf pattern, i.e. there were prominent, green bundles on a pale lamina. These green bundles on a pale lamina were specific to the tip of the leaf, whereas the lamina of the leaf base remained mostly green. All other leaves did not develop this reticulated leaf phenotype, however, we observed slightly pale leaves in general, especially in the emerging leaves (Figure 5).

Mapping of the EMS induced point mutations within the genome

To identify affected genes causing the mutant phenotypes a mapping-by-sequencing approach (Schneeberger *et al.*, 2009) was conducted. However, we could not follow the standard procedures for gene mapping/identification using outcross populations because we depended on the reporter gene expression in the bundle sheath to identify the individual mutant phenotype in the segregating mapping population. It has been shown before, that the use of backcross populations results in sufficient genetic diversity to identify the causative point mutation (Abe *et al.*, 2012; James *et al.*, 2013). Therefore, we backcrossed our mutant lines with the non-mutagenized reference line. F1 plants were propagated, resulting in the F2 backcross population, which showed a 3:1 segregation of the recessive mutant phenotype according to the Mendelian law.

To obtain a proof of concept of our mapping strategy, the first five homozygous EMS-generated mutant lines G21, G32, G35, L02 and L03 were selected for bulked segregate analyses and studied in parallel to the light and electron microscopy described above. Neighboring EMS-induced point mutations close to the causable SNP /the causative mutation were expected to be in linkage disequilibrium, and therefore, should not recombine. We analyzed the sequencing data of the bulked homozygous mutant plants by using a SHOREmap backcross scheme (Sun and Schneeberger, 2015) by which the peak of the confidence interval was mapped by analyzing the allele frequencies (AFs) of the EMS-induced SNPs. By pursuing this approach, clear candidate regions (AF > 0.9) could be identified in all five mutants (Figure 6, Figure S4, a, b, and c). Between two and ten mutations altering the coding region (exons) or the splicing sites of protein encoding genes could be identified for each of these mutant lines. The candidate genes will be analyzed by targeted gene knock-out experiments at a later date.

(V) Discussion

The photosynthetic activation of the bundle sheath, which is characterized by an increase in cell size and chloroplasts volume, is considered to be a key step in the evolution towards C_4 photosynthesis (Sage *et al*., 2014; Westhoff and Gowik, 2010). With the exception of the SCARECROW/SHORT-ROOT (SCR/SHR) and Golden*2*-like (GLK1/GLK2) transcription factors our knowledge on the gene regulatory networks that are additionally involved in the activation of the bundle sheath is rather poor (Slewinski *et al*., 2012; Cui *et al*., 2014; Slewinski *et al*., 2014; Rossini *et al*., 2001; Wang *et al*., 2017). To this end, we developed a forward genetic mutant screen using *A. thaliana* aiming to identify mutants with an increased bundle sheath volume and/or increased numbers of chloroplasts within the bundle sheath cells.

Both reporter gene lines, *LUC* and *GFP* were subjected to EMS mutagenesis and the primary screen of M2 plants was performed in parallel. For the primary screen, we used the reporter signal as a proxy, whereby M2 mutant plants with deviating reporter gene activity were selected (Figure 3). The *GFP* reporter gene turned out to be better suited for a high throughput screen than the *LUC* reporter gene. *LUC* reporter plants required an extra incubation step of leaves with the substrate D-Luciferin to generate the luminescence signal.

In total, we could screen twice as many plants in a given time with the *GFP* reporter compared to the *LUC* reporter. Furthermore, the spatial resolution of the reporter gene signal in the primary screen was higher in *GFP* plants in comparison to *LUC* reporter plants (Figure 1 and 3). Therefore, already in early screening stages we exclusively continued the screening process with the *GFP* reporter line.

Out of 755 identified primary mutants, more than 93 % had to be eliminated due to the instability of the phenotype or point mutations within the reporter gene construct (Figure 2). After sorting out these lines, we were still left with a reasonable number of 57 mutant lines (12 *LUC* and 45 *GFP*) (Figure 2). Furthermore, byproducts of this genetic screen such as mutant lines with mutations in the 1571-bp 5'-flanking region of the *GLDPA* gene could be helpful in elucidating the regulation of this complex 5' flanking segments with respect to the balance of transcriptional versus post-transcriptional gene control (Engelmann *et al.*, 2008; Wiludda *et al.*, 2012).

To investigate the correlation of altered reporter gene expression with an altered bundle sheath anatomy, 27 mutant lines (25 from the *GFP*-based screen and two from the *LUC* experiments) with strong deviations in reporter gene activity were used for light microscopic analysis (Table S1). Out of these 27 mutant lines, five mutant lines (G14, G15, G17, G19, and G20) showed altered bundle sheath structure. All of the five mutants possessed more bundle sheath cells, which was accompanied by an apparent increase in vascular tissue per vein (Figure 4).

We did not find any mutant with either increased chloroplast number per bundle sheath cell or altered chloroplast size, although the five mutant lines mentioned above exhibited a strong increase in GFP fluorescence which, due to the RbcS transit peptide, was localized in the chloroplasts (Figure S1). This enhanced GFP signal could derive either by an increase in xylem and/or phloem parenchyma and companion cells, all of which contain chloroplasts and/or could be explained by an increased number of bundle sheath cells per se in the mutant lines (Figure 4; see below).

Mutant line G19 showed deviations in the stacking of thylakoids and the accumulation of nucleoids (Figure S3). Regulation of thylakoid organization/stacking is controlled by numerous factors (Gao *et al.*, 2006; Armbruster *et al.*, 2013; Pribil *et al.*, 2014). Among those, CURT1 proteins are involved in bringing about membrane curvature at the grana margins, and increased amounts of CURT1 proteins give rise to grana with a large number of thylakoids (Armbruster *et al.*, 2013). Thylakoid formation is also linked with nucleoid distribution (Kobayashi *et al.*, 2013) and the spatial relationship is important for the assembly of the photosynthetic apparatus (Powikrowska *et al.*, 2014). Isolation of the gene responsible for the phenotype of mutant G19 may identify an additional genetic factor regulating thylakoid biogenesis.

All the five mutants showed to harbor an enlarged bundle sheath compartment. This enlargement was not caused by an increase in the bundle sheath cell sizes but rather originated from increase of the bundle sheath cell numbers. In summary, almost 20 % of the mutant lines (5/27) of which an altered reporter gene signal was detected in the primary screen could be clearly linked to anatomical changes within the bundle sheath and/or the vascular tissue. In the remaining mutant lines with no obvious aberration in bundle sheath/vascular tissue anatomy the increase/decrease of reporter gene signal might be caused by a transcriptional or translational perturbation of reporter gene expression. We conclude therefore that our screening strategy, i.e. using the activity of a reporter gene driven by a tissue-specific promoter as a rapid proxy in the primary screen was successful to identifying mutants affected in the anatomy of the bundle and its sheath.

As stated above the EMS mutant screen did not result in any mutant lines that were exclusively affected in bundle sheath anatomy. The increase in bundle sheath cell number was always associated with an expansion of the vascular tissue, probably due to enhanced cell division within the vascular tissue. The ontogenetic relation of the vascular bundle and the surrounding bundle sheath layer is already well described in grasses (Dengler *et al.*, 1985; Bosabalidis *et al.*, 1994). All C_3 grasses, as well as many C_4 grasses develop a double-sheath, i. e. the vascular tissue is encircled by a mestome sheath which itself is enclosed by a layer of parenchymatous sheath cells. In contrast, C_4 grasses of the NADP-malic enzyme subtype are single-sheath and do not have a mestome-sheath (Brown 1975; Hattersley & Watson 1976; Rao & Dixon 2016). Dengler et al. (1985) provided a detailed study on the origin of the bundle sheath in single-sheath C_4 and double-sheath C_4 and C_3 grasses. They reported that that the vasculature and its adjacent cell layer were clonally related and derive from procambial initials in both double-sheath (*Panicum effusum*, *Eleusine coracana* and *Sporobolus elongatus*) and single-sheath C_4 species (*Panicum bulbosum*, *Digitaria brownii* and *Cymbopogon procerus*). Nevertheless, it is not completely understood whether this situation is also true for minor veins. However, further studies in maize reported that both major and minor veins and associated bundle sheath cells are derived from a single cell lineage in the median layer of the leaf primordium (Bosabalidis *et al.*, 1994).

In contrast to grasses, our knowledge on the ontogeny of the bundle sheath in dicots is limited. It has been reported that the bundle sheath of C_3 and C_4 *Cleome* species, originate from more than one layers of ground meristem cells and only adaxial bundle sheath cells are of procambial origin (Koteyeva *et al.*, 2014). Since bundle sheath and vascular tissue either completely or partially arise from the same cell lineage, changes in vascular tissue might subsequently result in changes in bundle sheath anatomy as well. Moreover, the *GLDPA* promoter of the C_4 species *Flaveria trinervia,* which was used to drive the reporter gene expression in this study, contributes to the mutant characteristics and phenotypic spectrum obtained. The $GLDPA_{\mathrm{Ft}}$ promoter is highly active in both, the bundle sheath and the vascular tissue of *Arabidopsis* (Engelmann e*t al.*, 2008, Wiludda *et al.*, 2012). Hence, the use of this promoter inevitably produced mutants primarily affected in vascular tissue. Therefore, the use of an alternative bundle sheath specific promoter for a mutagenesis screen might result in mutants affected only in bundle sheath anatomy. A promoter, which drives expression exclusively in the bundle sheath of *Arabidopsis* has yet to be identified since all bundle sheath promoters known for dicots are, in varying degree, also active in the vasculature (Engelmann e*t al.*, 2008; Kirschner *et al.*, 2018). Alternatively, the specificity problem could be overcome by additional labeling of the vascular tissue with a second reporter gene. This two-reporter gene system would help to separate mutants only affected in the bundle sheath from mutants affected both in the bundle sheath and the vasculature. The promoters of the *SHORT-ROOT* (*SHR*), *Sultr2;1* and *SWEET1* genes of *Arabidopsis* which encode a GRAS family transcription factor, a sulfate transporter and a sucrose efflux transporter, respectively, are specifically active in the vasculature of developed *Arabidopsis* leaves and might serve as suitable candidates to additionally label the vasculature (Cui *et al*., 2014; Takahashi *et al*., 2000; Chen *et al*., 2012).

Although our screen did not directly deliver the type of mutants we were aiming for, the mutants obtained should nevertheless be helpful in understanding the ontogeny of the bundle sheath in the context of vascular tissue. In *Arabidopsis*, vascular cell proliferation and balance of xylem and phloem tissue production within a vascular strand is controlled by numerous factors (Schuetz *et al.*, 2013; Furuta *et al.*, 2014). Our screening strategy might, therefore, be a straightforward approach in identifying genes that are primarily involved in the differentiation of the vasculature and its ontogeny.

A successful forward genetic approach requires that the mutant genes identified can be molecularly identified, i.e. the causative genes have to be mapped precisely in order to facilitate their identification and verification by state of the art tools such as gene knock-out or replacement by the CRISPR/Cas9 technology (Hahn *et al.*, 2017). By using a back-cross procedure combined with bulk whole genome sequencing of F2 mutant plants we were able of locating the causative mutations in an interval of 0.2 – 3.4 Mbp containing 2 to 10 mutated candidate genes. This is suitable for inducing CRISPR/Cas9-mediated knock-outs of the candidate genes. Moreover, the mapping resolution may be improved by enlarging the mapping population and thus increasing the numbers of pooled F2 mutant plants for bulk sequencing (James *et al.*, 2013). Our forward genetic approach relied on the use of reporter genes for the rapid and easy identification of mutant candidates in a primary proxy screen. It was coupled with a powerful mapping by sequencing strategy. We believe that this combination is very useful if high-throughput phenotyping of structural deviations at the cellular or tissue level is not possible.

(VI) Experimental procedures

Plant material

Arabidopsis thaliana (Ecotype Columbia-0) was used as a genetic background for both reporter gene lines. 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 operated 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 bleach 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 stratified at 4 °C in the dark for at least 48 h before sowing in either soil (Floraton 1, Floragard, Oldenburg, Germany) or petri dishes with ½ Murashige and Skoog (MS)-medium containing 0.6 % agar and 1 % sucrose.

Generation of reporter gene lines

The pGreen Gateway vector containing the firefly luciferase 68 gene (pGreen-*LUC68*) served as a backbone for the Luciferase (*LUC*) reporter construct and was kindly provided by Franziska Turck (Adrian *et al.*, 2010). The 1571-bp 5'-flanking region of the glycine decarboxylase P protein gene (*GLDPA*) of the C_4 Asteracean species *Flaveria trinervia* was amplified by PCR from a *GLDPA*$_{Ft}$-GUS template (Engelmann *et al.*, 2008) with specific oligonucleotides listed in Table S2 that added *att*B1 and *att*B2 sites to the PCR product. To introduce the *GLDPA*$_{Ft}$ promoter sequence (p*GLDPA*$_{Ft}$) into the Gateway entry vector pDONR221 the BP Clonase reaction (Gateway® BP Clonase® enzyme mix, ThermoFisher Scientific) was carried out as described by the manufacturer. The resulting pENTRY221-

p*GLDPA*$_{Ft}$ was subsequently used for the LR Clonase reaction (Gateway® LR Clonase® enzyme mix, ThermoFisher Scientific) to transfer p*GLDPA*$_{Ft}$ into pGreen-*LUC68* (pGreen–p*GLDPA*$_{Ft}$::*LUC68*).

The binary plant transformation vector pBI121 (Clonetech laboratories; Jefferson et al., 1987) was used to assemble the *GFP* reporter gene construct that included the p*GLDPA*$_{Ft}$ region (Engelmann *et al.*, 2008), the transit peptide sequence of the gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase of *A. thaliana* (AT1G67090; TP$_{RbcS}$; Kim *et al.*, 2010) and the s*GFP* gene sequence fused in frame to the TP$_{RbcS}$ segment by using standard cloning procedures. The resulting final reporter gene construct was named p*GLDPA*$_{Ft}$::*TP*$_{RbcS}$ -s*GFP*.

Transformation of *A. thaliana*

Both reporter gene constructs were transferred into the *Agrobacterium tumefaciens* strain AGL1 (Lazo *et al.*, 1991) by electroporation, and subsequently transformed into *A. thaliana* (Ecotype Columbia-0) by using the floral dip method (Logemann *et al.*, 2006). T1 plants containing an intact reporter gene were first selected by Kanamycin resistance followed by PCR amplification and sequencing of the entire reporter gene construct. Positive lines were propagated into the T3 generation, and homozygous plants were selected for the mutant screens.

Ethyl methanesulfonate (EMS) mutagenesis

Approximately 40,000 seeds (~1.6 g) of the p*GLDPA*$_{Ft}$::*LUC* reporter gene line and 120,000 (~4.8 g) seeds of the p*GLDPA*$_{Ft}$::*GFP* reporter gene line were used for EMS mutagenesis. The seeds were initially washed with 0.1 % (v/v) TWEEN® 20 for 15 min, after which EMS (Sigma-Aldrich) was added to a final concentration of 0.25 % (v/v). The mixture was incubated for 16 h on a rotating platform at room temperature in the dark. 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 two days at 4 °C, M1 seeds were sown evenly in soil. M2 seeds were harvested from a pool of 30–50 M$_1$ plants. M2 plants were grown for about 14–17-days and used for the mutagenesis screen described below.

Mutant screen

The first leaf pair was analyzed for both, *LUC* and *GFP* aberrant reporter gene expression. In general, 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 NC100 U (Berthold Technologies, Bad Wildbad, Germany) using the *in vivo* imaging software indiGO (Berthold Technologies, Bad Wildbad, Germany). Before screening, leaves were incubated in a 1 mM luciferin solution for 5 min after which LUC activity was detected (exposure time: 120 s). The resultant signal in the bundle sheath/vasculature of the EMS-mutagenized M2 populations was compared to the non-mutagenized reporter line. In terms of the mutant screen with the *GFP* reporter gene line, plants of the M2 generation were screened for aberrant GFP expression under a fluorescence binocular microscope (Nikon SMZ25, Duesseldorf, Germany). All primary mutants selected at the M2 stage were analyzed again at the M3 stage, to confirm the aberrant mutant phenotype. Additionally, signal intensity was measured for whole leaves and normalized to the leaf area using 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.

DNA was isolated from the mutant lines to check for point mutations in the reporter gene construct. The complete region (p$GLDPA_{\mathrm{Ft}}$::*RbcS.TP* or p$GLDPA_{\mathrm{Ft}}$::*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 exhibiting point mutations within the reporter gene constructs were discarded.

Microscopic analysis

Internal leaf anatomy was assessed on sections sampled from the middle of the second leaf pair (one leaf per plant: three plants per line). Plants were sampled between 09:00 to 11:00 a.m. and prepared for light and transmission (TEM) microscopy as described by (Khoshravesh *et al.*, 2017). Images for light microscopy were captured on a Zeiss Axiophot microscope equipped with a DP71 Olympus camera and image analysis software (Olympus cellSens, 2009). Images for TEM were captured on the Phillips 201 TEM equipped with an Advantage HR camera system (Advanced Microscopy Techniques).

Mapping by sequencing

Stable M4 mutant lines with intact reporter gene sequences were backcrossed with the corresponding non-mutagenized reporter gene line. The resulting BC1 plants were selfed and the BC1-F2 plants were examined for the individual aberrant phenotype. Genomic DNA was isolated from pooled leaf samples of 50–60 BC1-F2 mutant plants using the DNeasy Plant Maxi Kit (Qiagen, Hilden, Germany). DNA was eluted in 750 µl sterile water in two steps and concentrated to at least 50 ng/µl by vacuum infiltration.

Sequencing libraries of the pooled mutant DNA as well as of the two 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 350bp. The DNA library was prepared with the TruSeq DNA PCR-Free LT Library Preparation Kit (Illumina) according to the manufacturer's manual. The DNA concentrations of the libraries were determined with the KAPA Library Quantification Kit Illumina® platforms (Kapabiosystems).

Paired-end sequencing (2 x 150 bp) was performed using an Illumina HiSeq3000 system and was carried out by the "Genomics and Transcriptomics laboratory" of the Biologisch-Medizinisches Forschungszentrum (BMFZ) of the Heinrich Heine University of Duesseldorf with 80 to 500 fold coverage. EMS induced mutations potentially responsible for the mutant phenotypes were identified by using SHOREmap v3.0 following the backcrossing procedure as described (http://bioinfo.mpipz.mpg.de/shoremap/guide.html; Sun and Schneeberger, 2015). Read mapping and SNP calling were performed by using SHORE v0.9.3 and Genomemapper v0.4.4.

(VII) Accession numbers

(VIII) Acknowledgments

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.

(IX) Supporting Information

Supporting figures: Figure S1, Figure S2, Figure S3 and Figure S4

Supporting tables: Table S1 and Table S2

(X) References

Abe, A., Kosugi, S., Yoshida, K., et al. (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.*, **30**, 174–178.

Adrian, J., Farrona, S., Reimer, J.J., Albani, M.C., Coupland, G. and 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.

Armbruster, U., Labs, M., Pribil, M., et al. (2013) *Arabidopsis* CURVATURE THYLAKOID1 proteins modify thylakoid architecture by inducing membrane curvature. *Plant Cell*, **25**, 2661–78.

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

Bel, A.J.E. Van (1993) Strategies of phloem loading. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **44**, 253–281.

Bosabalidis, A.M., Evert, R.F. and Russin, W.A. (1994) Ontogeny of the vascular bundles and contiguous tissues in the maize leaf blade. *Am. J. Bot.*, **81**, 745–752.

Brown, W. V. (1975) Variations in anatomy, associations, and origins of Kranz tissue. *Am. J. Bot.*, **62**, 395–402.

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

Chen, L.-Q., Qu, X.-Q., Hou, B.-H., Sosso, D., Osorio, S., Fernie, A.R. and Frommer, W.B. (2012) Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science*, **335**, 207–211.

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

Dengler, N.G., Dengler, R.E. and Hattersley, P.W. (1985) Differing ontogenetic origins of PCR (" Kranz ") sheaths in leaf blades of C_4 grasses (POACEAE). *Am. J. Bot.*, **72**, 284–302.

Edwards, G.E. and Voznesenskaya, E. V. (2011) C_4 photosnythesis: Kranz forms and single-cell C_4 in terrestrial plants. In: Raghavendra, A.S. and Sage, R.F. (eds.), *C_4 photosynthesis and related CO_2 concentrating mechanisms*. Dordrecht, Springer Netherlands **32,** 29-61.

Ehleringer, J.R., Cerling, T.E. and Helliker, B.R. (1997) C_4 photosynthesis, atmospheric CO_2, and climate. *Oecologia*, **112**, 285–299.

Ehleringer, J.R., Sage, R.F., Flanagan, L.B. and Pearcy, R.W. (1991) Climate change and the evolution of C_4 photosynthesis. *Trends Ecol. Evol.*, **6**, 95–99.

Engelmann, S., Wiludda, C., Burscheidt, J., et al. The gene for the P-subunit of glycine decarboxylase from the C_4 Species *Flaveria trinervia*: analysis of transcriptional control in transgenic *Flaveria bidentis* (C_4) and *Arabidopsis* (C_3). *Plant Physiol.*, **146**, 1773–1785.

Furbank, R.T. (2011) Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types? *J. Exp. Bot.*, **62**, 3103–3108.

Furuta, K.M., Hellmann, E. and Helariutta, Y. (2014) Molecular control of cell specification and cell differentiation during procambial development. *Annu. Rev. Plant Biol.*, **65**, 607–638.

Gao, H., Sage, T.L. and Osteryoung, K.W. (2006) FZL, an FZO-like protein in plants, is a determinant of thylakoid and chloroplast morphology. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 6759–6764.

Gowik, U. and Westhoff, P. (2011) The path from C_3 to C_4 photosynthesis. *Plant Physiol.*, **155**, 56–63.

Griffiths, H., Weller, G., Toy, L.F.M. and Dennis, R.J. (2013) You're so vein: Bundle sheath physiology, phylogeny and evolution in C_3 and C_4 plants. *Plant, Cell Environ.*, **36**, 249–261.

Haberlandt, G. (1904) *Physiologische Pflanzenanatomie*, Leipzig, Germany: Verlag von Wilhelm Engelmann.

Hahn, F., Mantegazza, O., Greiner, A., Hegemann, P., Eisenhut, M. and Weber, A.P.M. (2017) An Efficient visual screen for CRISPR/Cas9 activity in *Arabidopsis thaliana*. *Front. Plant Sci.*, **08**, 39.

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

Hattersley, P. and Watson, L. (1976) C_4 Grasses: an Anatomical criterion for distinguishing between NADP-malic enzyme species and PCK or NAD-malic enzyme species. *Aust. J. Bot.*, **24**, 297-308.

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

James, G.V., Patel, V., Nordström, K.J.V., Klasen, J.R., Salomé, P.A., Weigel, D. and Schneeberger, K. (2013) User guide for mapping-by-sequencing in *Arabidopsis*. *Genome Biol.*, **14**, R61.

Jander, G., Baerson, S.R., Hudak, J.A., Gonzalez, K.A., Gruys, K.J. and Last, R.L. (2003) Ethylmethanesulfonate saturation mutagenesis in *Arabidopsis* to determine frequency of herbicide resistance. *Plant Physiol.*, **131**, 139–146.

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

Khoshravesh, R., Lundsgaard-Nielsen, V., Sultmanis, S. and Sage, T.L. (2017) Light microscopy, transmission electron microscopy, and immunohistochemistry protocols for studying photorespiration. In: Fernie A., Bauwe H., Weber A. (eds) Photorespiration. *Methods Mol. Biol.*, **1653,** 243–270. Humana Press, New York, NY

Kim, S., Lee, D.-S., Choi, I.S., Ahn, S.-J., Kim, Y.-H. and 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**, 489–497.

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

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

Kirschner, S., Woodfield, H., Prusko, K., Koczor, M., Gowik, U., Hibberd, J.M. and Westhoff, P. (2018) Expression of *SULTR2;2* in the *Arabidopsis* bundle sheath is mediated by a highly conserved positive regulator. *bioRxiv*. http://dx.doi.org/10.1101/250738

Kobayashi, K., Narise, T., Sonoike, K., et al. (2013) Role of galactolipid biosynthesis in coordinated development of photosynthetic complexes and thylakoid membranes during chloroplast biogenesis in *Arabidopsis*. *Plant J.*, **73**, 250–261.

Koteyeva, N.K., Voznesenskaya, E. V, Cousins, A.B. and Edwards, G.E. (2014) Differentiation of C_4 photosynthesis along a leaf developmental gradient in two *Cleome* species having different forms of Kranz anatomy. *J. Exp. Bot.*, **65**, 3525–3541.

Lazo, G.R., Stein, P.A. and Ludwig, R.A. (1991) A DNA transformation-competent *Arabidopsis* genomic library in Agrobacterium. *Biotechnol. (Nature Publ. Company)*, **9**, 963–967.

Leegood, R.C. (2008) Roles of the bundle sheath cells in leaves of C_3 plants. In *Journal of Experimental Botany*. **59,** 1663–1673.

Logemann, E., Birkenbihl, R.P., Ülker, B. and Somssich, I.E. (2006) An improved method for preparing Agrobacterium cells that simplifies the *Arabidopsis* transformation protocol. *Plant Methods*, **2**, 1–5.

Maeda, H., Sage, T.L., Isaac, G., Welti, R. and Dellapenna, D. (2008) Tocopherols modulate extraplastidic polyunsaturated fatty acid metabolism in *Arabidopsis* at low temperature. *Plant Cell*, **20**, 452–470.

Page, D.R. and Grossniklaus, U. (2002) The art and design of genetic screens: *Arabidopsis thaliana*. *Nat. Rev. Genet.*, **3**, 124–136.

Petricka, J.J., Winter, C.M. and Benfey, P.N. (2012) Control of *Arabidopsis* Root Development. *Annu. Rev. Plant Biol.*, **63**, 563–590.

Pollock, D.D. and Larkin, J.C. (2004) Estimating the degree of saturation in mutant screens. *Genetics*, **168**, 489–502.

Powikrowska, M., Oetke, S., Jensen, P.E. and Krupinska, K. (2014) Dynamic composition, shaping and organization of plastid nucleoids. *Front. Plant Sci.*, **5**, 424.

Pribil, M., Labs, M. and Leister, D. (2014) Structure and dynamics of thylakoids in land plants. *J. Exp. Bot.*, **65**, 1955–1972.

Rao, X. and Dixon, R.A. (2016) The differences between NAD-ME and NADP-ME subtypes of C_4 Photosynthesis: More than Decarboxylating Enzymes. *Front. Plant Sci.*, **7**, 1525.

Rossini, L., Cribb, L., Martin, D.J. and Langdale, J.A. (2001) The maize *Golden2* gene defines a novel class of transcriptional regulators in plants. *Plant Cell*, **13**, 1231–1244.

Sage, R.F. (2016) A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *J. Exp. Bot.*, **67**, 4039–4056.

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

Sage, R.F., Khoshravesh, R. and Sage, T.L. (2014) From proto-Kranz to C_4 Kranz: building the bridge to C_4 photosynthesis. *J. Exp. Bot.*, **65**, 3341–3356.

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

Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*, **9**, 671–675.

Schuetz, M., Smith, R. and Ellis, B. (2013) Xylem tissue specification, patterning, and differentiation mechanisms. *J. Exp. Bot.*, **64**, 11–31.

Sheehy, J.E., Mitchell, P.L. (Peter L.. and Hardy, B. (2007) Charting new pathways to C_4 rice, International Rice Research Institute.

Slewinski, T.L. (2013) Using evolution as a guide to engineer kranz-type C_4 photosynthesis. *Front. Plant Sci.*, **4**, 212.

Slewinski, T.L., Anderson, A.A., Price, S., Withee, J.R., Gallagher, K. and Turgeon, R. (2014) *SHOR-ROOT1* plays a role in the development of vascular tissue and kranz anatomy in maize leaves. *Mol. Plant*, **7**, 1388–1392.

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

Sparks, E.E., Drapek, C., Gaudinier, A., et al. (2016) Establishment of expression in the *SHORTROOT-SCARECROW* transcriptional cascade through opposing activities of both activators and repressors. *Dev. Cell*, **39**, 585–596.

Sun, H. and Schneeberger, K. (2015) SHOREmap v3.0: Fast and accurate identification of causal mutations from forward genetic screens. In J. M. Alonso and A. N. Stepanova. (eds) *Plant Functional Genomics: Methods Mol. Biol.,* ***1284,*** 381–395. Human Press, New York, NY.

Takahashi, H., Watanabe-Takahashi, A., Smith, F.W., Blake-Kalff, M., Hawkesford, M.J. and Saito, K. (2000) The roles of three functional sulphate transporters involved in uptake and translocation of sulphate in *Arabidopsis thaliana*. *Plant J.*, **23**, 171–182.

Wang, P., Kelly, S., Fouracre, J.P. and Langdale, J.A. (2013) Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C_4 Kranz anatomy. *Plant J.*, **75**, 656–670.

Wang, P., Khoshravesh, R., Karki, S., Furbank, R., Sage, T.L. and Langdale, J.A. (2017) Re-creation of a key step in the evolutionary switch from C_3 to C_4 leaf anatomy. *Curr. Biol.*, **27**, 3278–3287.

Wang, Y., Bräutigam, A., Weber, A.P.M. and Zhu, X.-G. (2014) Three distinct biochemical subtypes of C_4 photosynthesis? A modelling analysis. *J. Exp. Bot.*, **65**, 3567–3578.

Westhoff, P. and Gowik, U. (2010) Evolution of C_4 photosynthesis–looking for the master switch. *Plant Physiol.*, **154**, 598–601.

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

Zhu, X.-G., Long, S.P. and Ort, D.R. (2010) Improving photosynthetic efficiency for greater yield. *Annu. Rev. Plant Biol.*, **61**, 235–261.

(XI) Tables

Supporting tables

Table S1. Reporter gene expression of 27 EMS mutant lines, which were used for light microscopic analysis. +, more signal; -, less signal.

Table S2. Oligonucleotides used in this study

(XII) Figure legends

Figure 1. Labeling the bundle sheath of leaves of *Arabidopsis thaliana* by using Luciferase (*LUC*) and green fluorescent Protein (*GFP*) reporter genes, respectively, which are driven by the promoter of the gene encoding the P subunit of glycine decarboxylase of the C_4 plant *Flaveria trinervi*a (p*GLDPA*$_{\mathrm{Ft}}$).
(a) The constructs used to generate the p*GLDPA*$_{\mathrm{Ft}}$-*LUC* and p*GLDPA*$_{\mathrm{Ft}}$-*GFP* reference lines. (b) Luminescence of a leaf of the *LUC* reference line. (c) GFP fluorescence of a leaf of the *GFP* reference line. (d) Longitudinal section of veins of the reference line showing GFP fluorescence localized in the chloroplasts (arrowheads) of the bundle sheath and vasculature.

Figure 2. Workflow of the EMS screen with both reporter gene lines (*LUC* and *GFP*). The number of plants/mutants of each step of the mutant screen using the *LUC* and *GFP* reporter gene lines are shown on the left side and right side, respectively.

Figure 3. Results of the EMS primary screen.
(a) Luminescence signal emitted by the reference line compared to (b) a mutant line with increased and (c) decreased reporter gene signal relative to control/reference line. (d) GFP fluorescence of the reference line compared to (e–g) three different classes of primary mutants with either (e) more, (f) less, or (g) a diffused reporter gene activity in the leaves. (h and i) A close-up view of 3° veins of (h) the reference line and (i) mutant line G-19. The width of the veins is emphasized by white arrows.

Figure 4. Light micrographs illustrating cross-sections of a third-order vein. (a) Reference line. (b) Line 14. (c) Line 15. (d) Line 17. (e) Line 19. (f) Line 20. Bar – 10 µm. BS, bundle sheath; X, vessel element; * marks companion cell; white arrow marks sieve tube element.

Figure 5. Overview of mutant line G14, G15, G17, G19, G20, and the reference line. All plants were 28 days old.

Figure 6. Allelic frequencies for mutant lines G21, G32, G35, L02, and L03. Allelic frequencies (AF) for all SNPs resolved using the 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. Only the chromosome with allelic frequencies >0.9 is shown for each mutant line.

Legends of supporting figures

Figure S1. Relative reporter gene signal intensity of all (a) 45 *GFP* and (b) 12 *LUC* reporter mutant lines compared to the appropriate reference line. Reporter gene signal was measured in whole leaves of 14–17 days old plants and normalized to leaf area. At least 50 plants per mutant line were analyzed.

Figure S2. Paradermal sections of the reference line (a) and mutant lines 14 (b), 15 (c), 17 (d), 19 (e), and 20 (f). Bar, 10µm. BS, bundle sheath; M, mesophyll.

Figure S3. Structural features of reference and mutant line G-19. (a) GFP fluorescence of the reference line. (b) GFP fluorescence of mutant line. (c) Light micrograph illustrating vascular tissue size of G-19 relative to reference line as viewed in Figure 3. (d, i) Chloroplasts from mesophyll of the reference line. (e, h, j) Chloroplasts from mesophyll of mutant line. (f) Chloroplast from bundle sheath of reference line. (g) Chloroplast from bundle sheath of mutant line. Chloroplasts are labeled by arrowheads, chloroplast vesicles by arrows. Double black asterisks mark grana of the reference line, white asterisks those of mutant line. BS, bundle sheath; N, nucleoid. Bar, c = 10 µm; d, e, f, h = 250 nm; g, i, j = 100 nm.

Figure S4. Allelic frequencies for mutant line L02 and L03 (Panel a), G21 and G32 (Panel b), and G25 (Panel c).

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.

(XIII) Figures

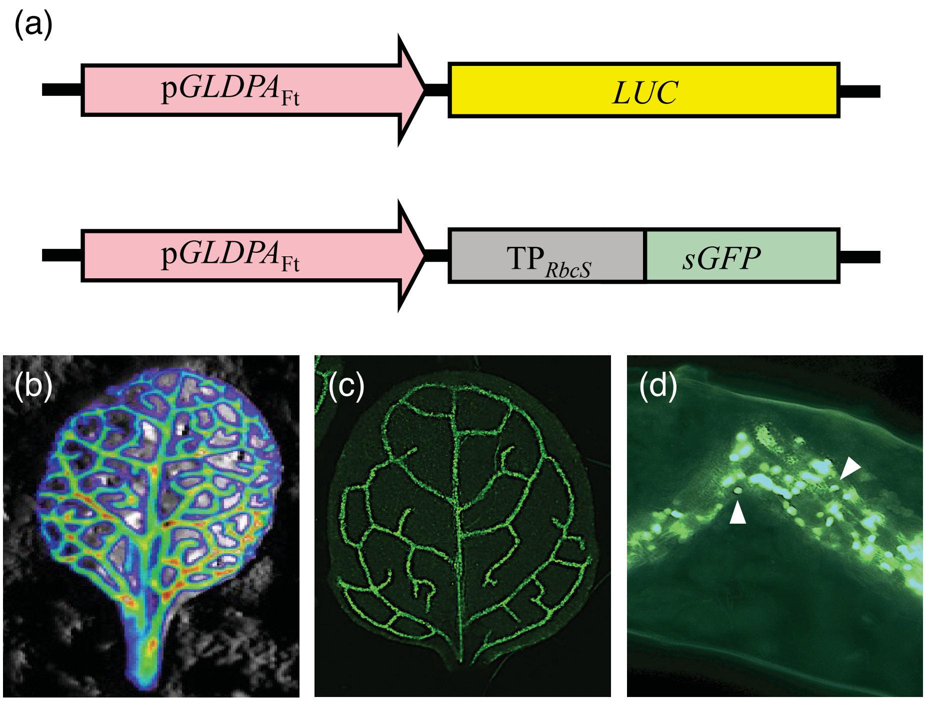


Figure 1. Labeling the bundle sheath of leaves of *Arabidopsis thaliana* by using Luciferase (*LUC*) and green fluorescent Protein (*GFP*) reporter genes, respectively, which are driven by the promoter of the gene encoding the P subunit of glycine decarboxylase of the C_4 plant *Flaveria trinervi*a (p*GLDPA*$_{Ft}$). (**a**) The constructs used to generate the p*GLDPA*$_{Ft}$-*LUC* and p*GLDPA*$_{Ft}$-*GFP* reference lines. (**b**) Luminescence of a leaf of the *LUC* reference line. (**c**) GFP fluorescence of a leaf of the *GFP* reference line. (**d**) Longitudinal section of veins of the reference line showing GFP fluorescence localized in the chloroplasts (arrowheads) of the bundle sheath and vasculature.

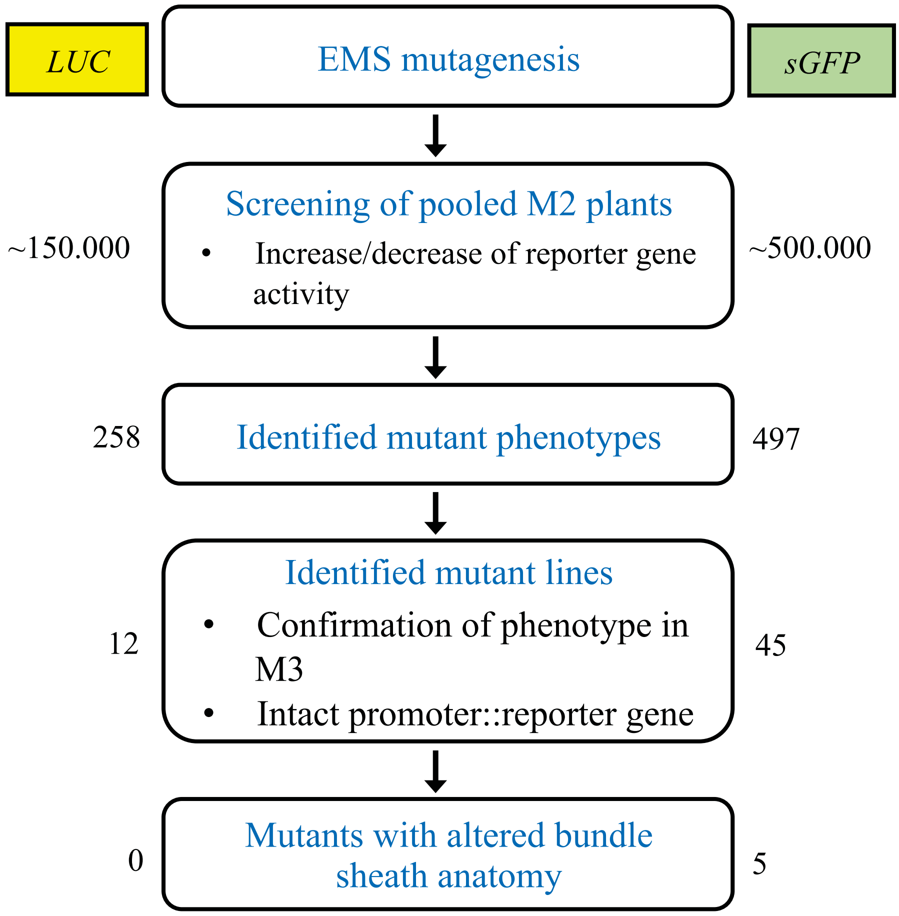


Figure 2. Workflow of the EMS screen with both reporter gene lines (*LUC* and *GFP*).
The number of plants/mutants of each step of the mutant screen using the *LUC* and *GFP* reporter gene lines are shown on the left side and right side, respectively.

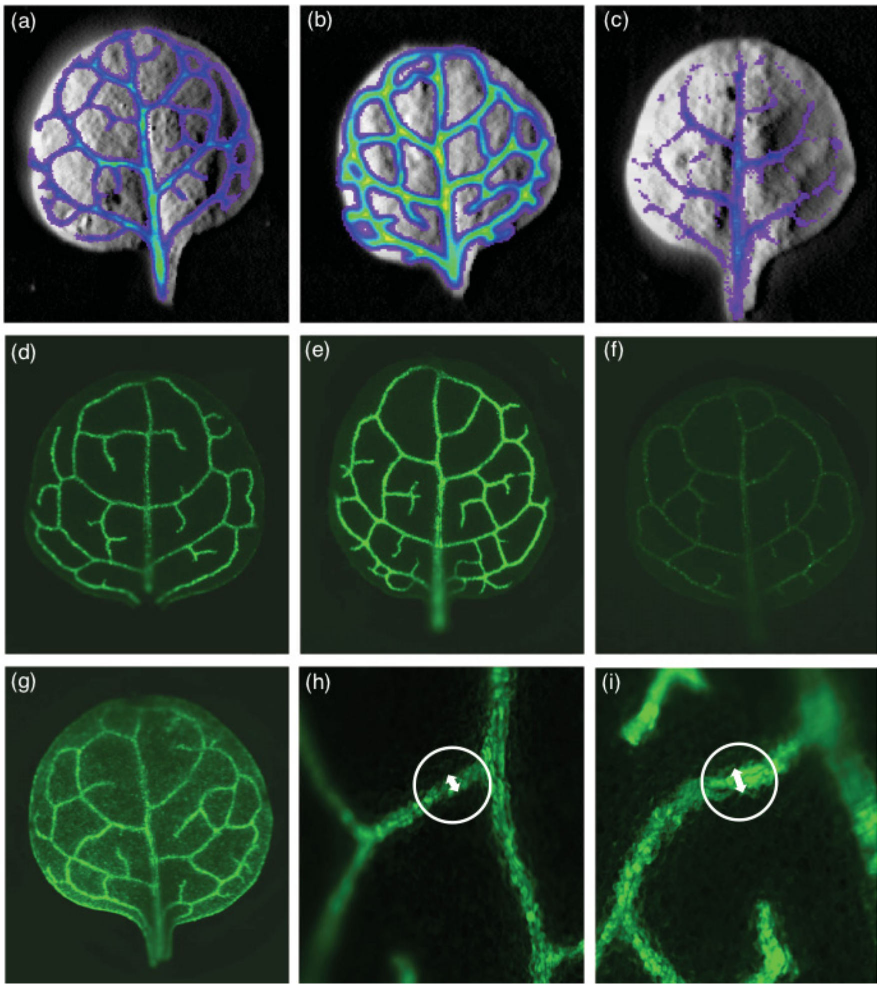


Figure 3. Results of the EMS primary screen.
(**a**) Luminescence signal emitted by the reference line compared to (**b**) a mutant line with increased and (**c**) decreased reporter gene signal relative to control/reference line. (**d**) GFP fluorescence of the reference line compared to (**e–g**) three different classes of primary mutants with either (**e**) more, (**f**) less, or (**g**) a diffused reporter gene activity in the leaves. (**h** and **i**) A close-up view of 3° veins of (**h**) the reference line and (**i**) mutant line G-19. The width of the veins is emphasized by white arrows.

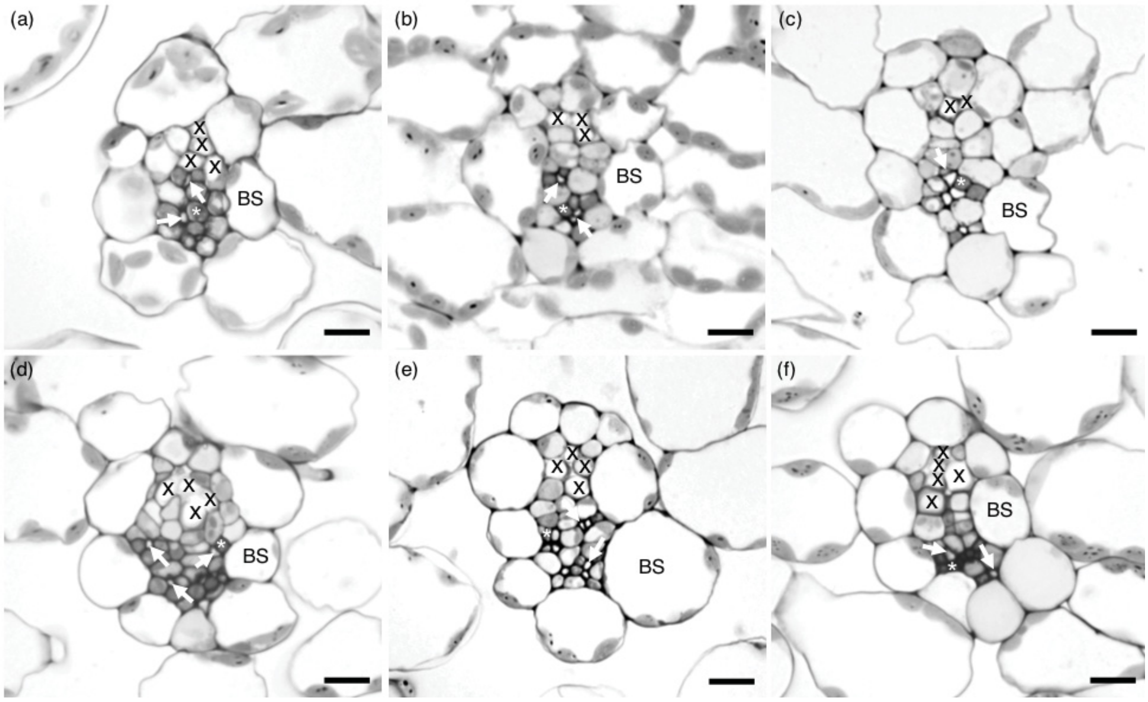


Figure 4. Light micrographs illustrating cross-sections of a third-order vein.
(**a**) Reference line (**b**) Line 14 (**c**) Line 15 (**d**) Line 17 (**e**) Line 19 (**f**) Line 20. Bar – 10µm. BS, bundle sheath; X, vessel element; * marks companion cell; white arrow marks sieve tube element.

Figure 5. Overview of mutant line G14, G15, G17, G19, G20, and the reference line. All plants were 28 days old.

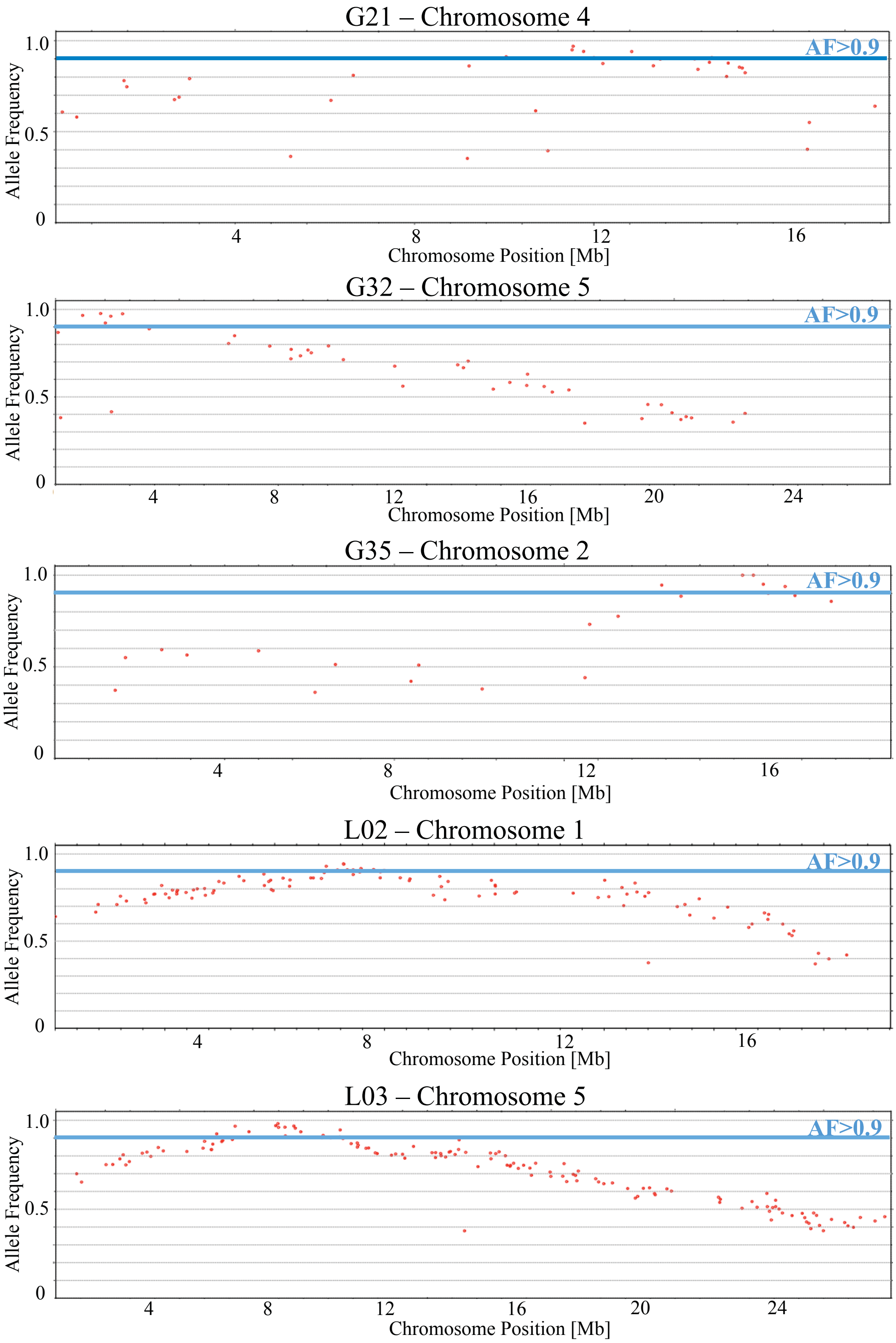


Figure 6. Allelic frequencies for mutant lines G21, G32, G35, L02 and L03.
Allelic frequencies (AF) for all SNPs resolved using the reporter gene line parent and BCF2 mutant whole genome sequence data. Genes containing a non-synonymous SNP with AF>0.9 were considered as candidate genes. Only the chromosome with allelic frequencies >0.9 is shown for each mutant line.

Supporting figures

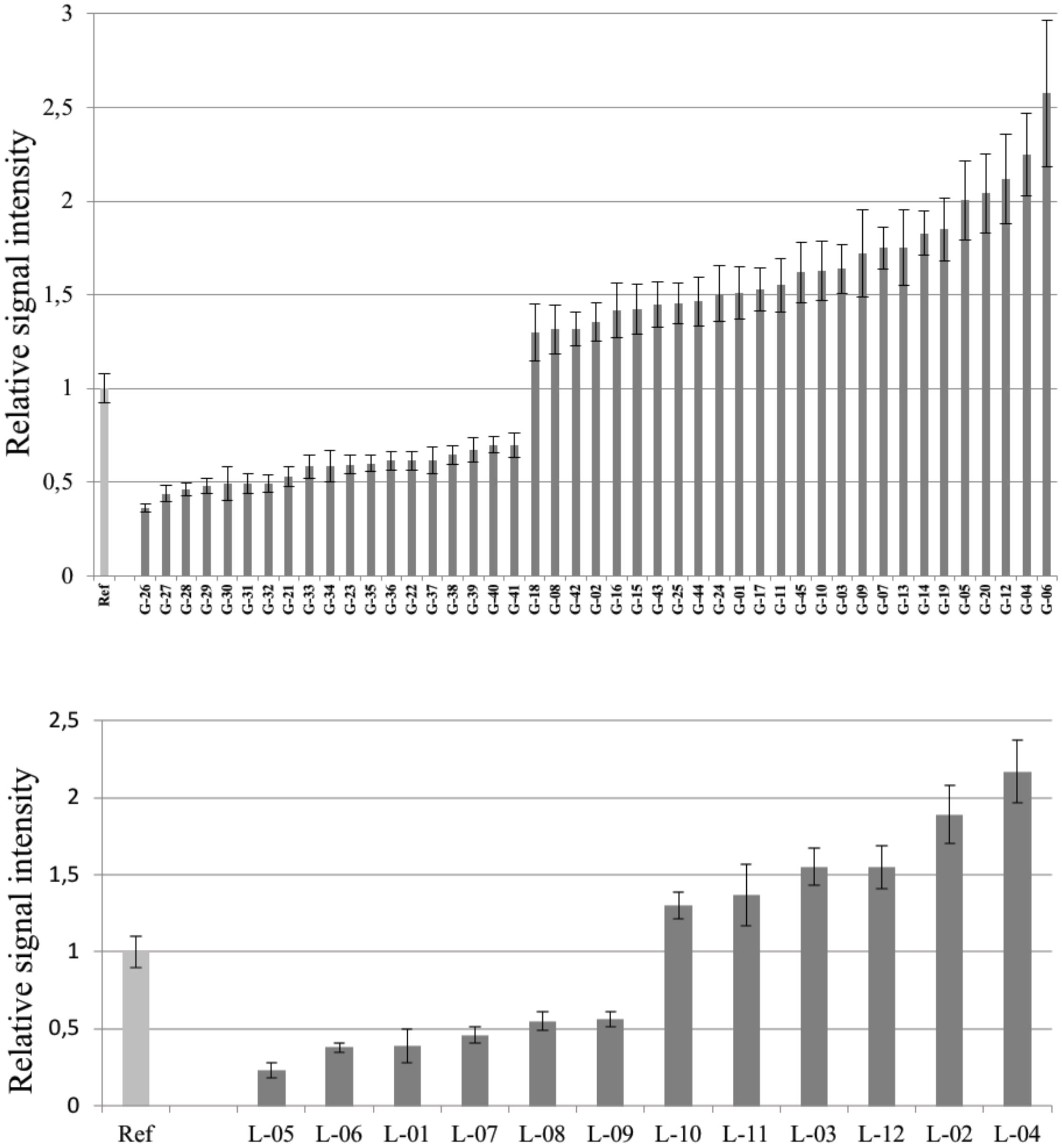


Figure S1. Relative reporter gene signal intensity of all (**a**) 45 *GFP* and (**b**) 12 *LUC* reporter mutant lines were compared to the appropriate reference line. Reporter gene signal was measured in whole leaves of 14–17 days old plants and normalized to leaf area. At least 50 plants per mutant line were analyzed.

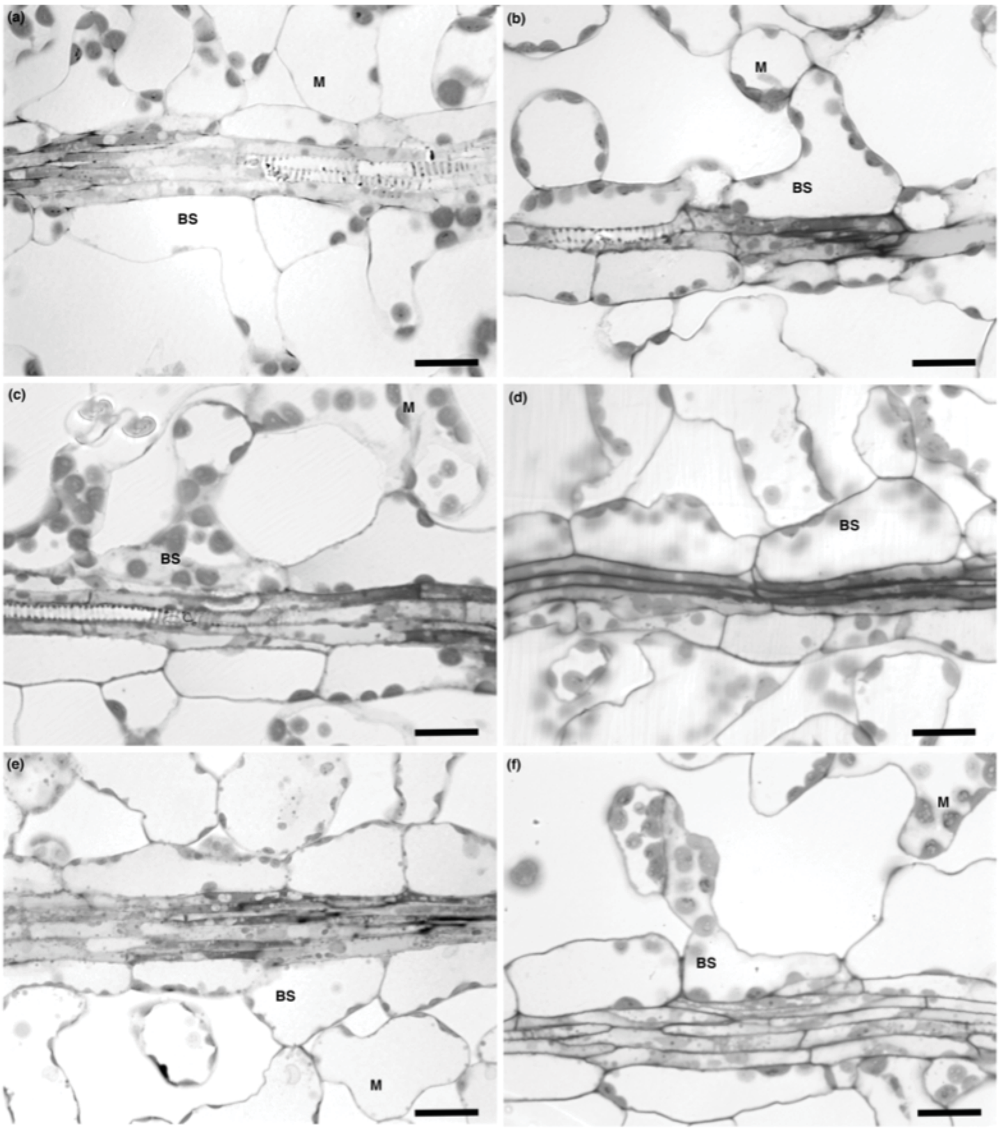


Figure S2. Paradermal sections of the reference line (**a**) and mutant lines 14 (**b**), 15 (**c**), 17 (**d**), 19 (**e**), and 20 (**f**). Bar – 10µm. BS, bundle sheath; M, mesophyll.

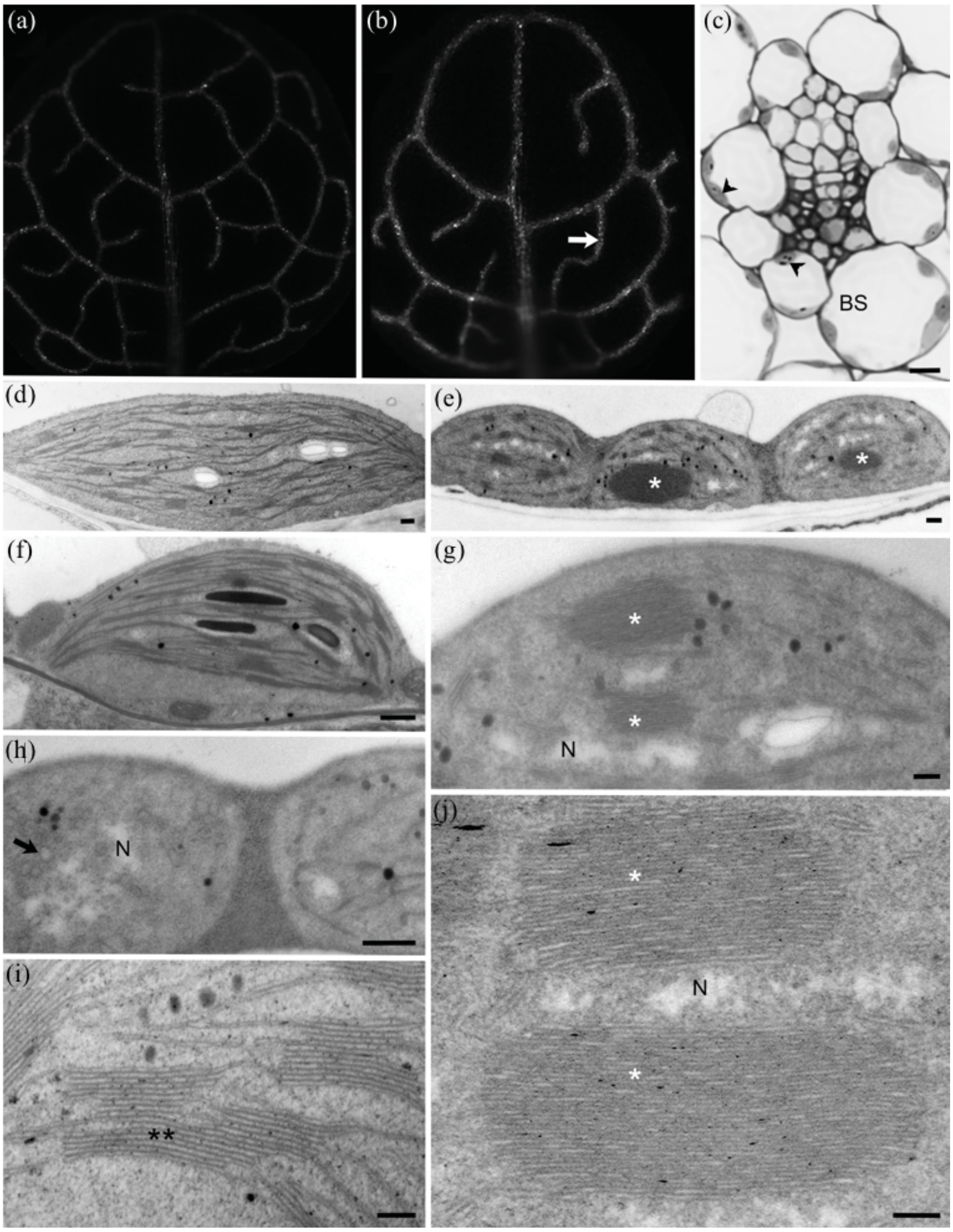
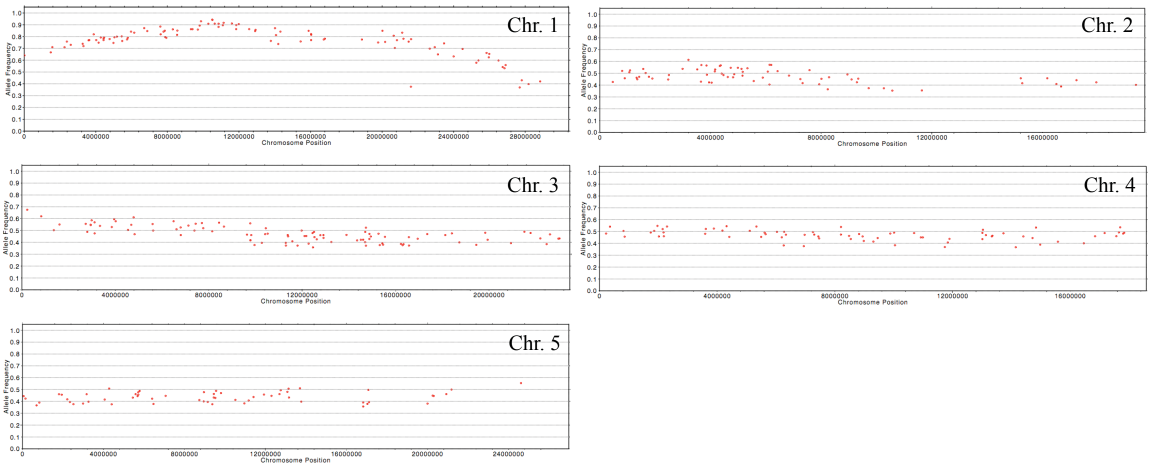
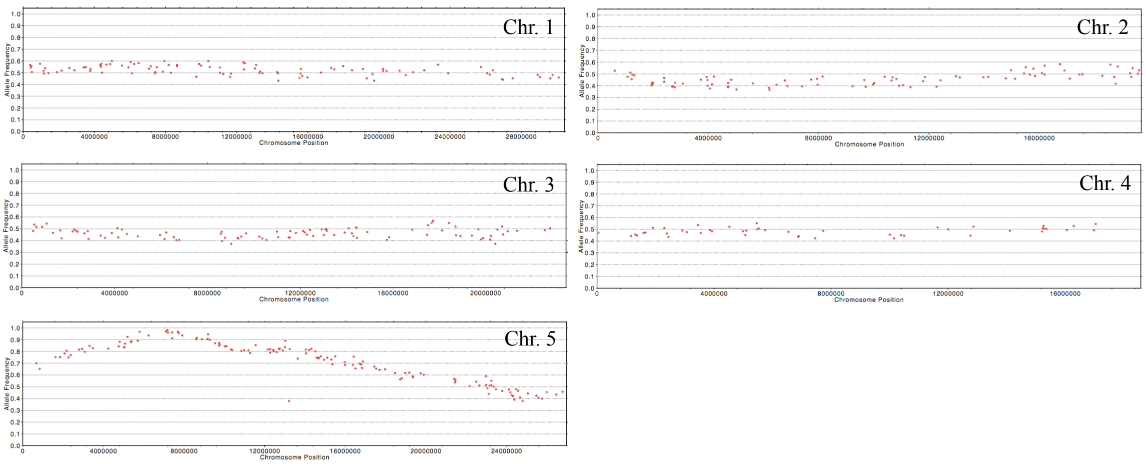


Figure S3. Structural features of reference and mutant line G-19.
(**a**) GFP fluorescence of the reference line. (**b**) GFP fluorescence of mutant line. (**c**) Light micrograph illustrating vascular tissue size of G-19 relative to reference line as viewed in Figure 3. (**d, i**) Chloroplasts from mesophyll of the reference line. (**e, h, j**) Chloroplasts from mesophyll of mutant line. (**f**) Chloroplast from bundle sheath of the reference line. (**g**) Chloroplast from bundle sheath of mutant line. Chloroplasts are labeled by arrowheads, chloroplast vesicles by arrows. Double black asterisks mark grana of reference line, white asterisks those of mutant line. BS, bundle sheath; N, nucleoid. Bar, c = 10 µm; d, e, f, h = 250 nm; g, i, j = 100 nm.

Mutant line L02



Mutant line L03



Mutant line G21

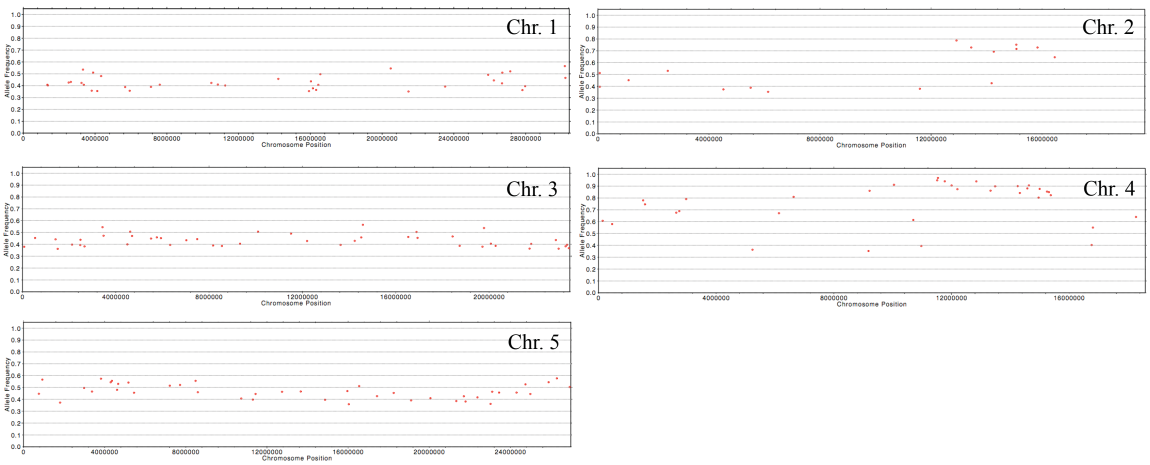
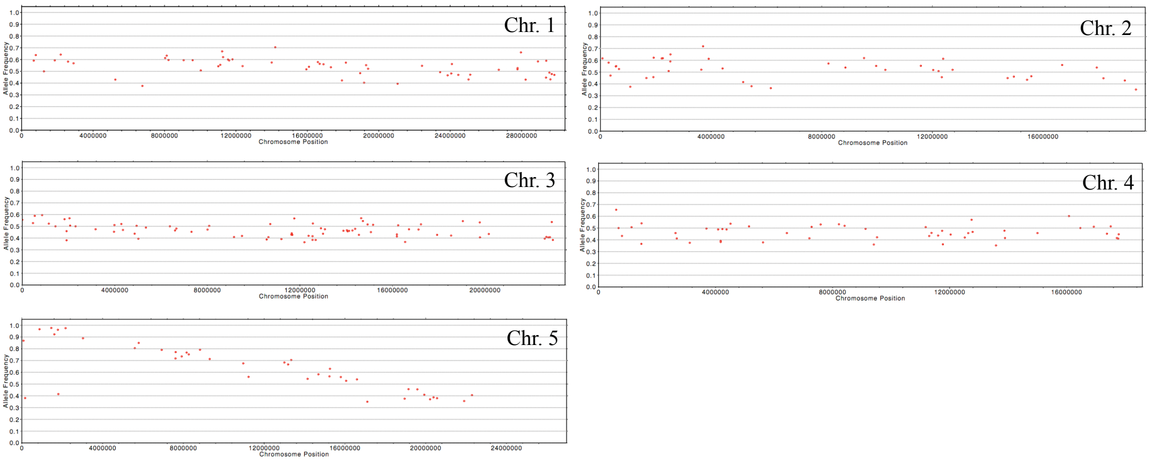


Figure S4. Allelic frequencies for mutant line L02, L03, G21, G32 and G25.
Allelic frequencies (AF) for all SNPs resolved using the *LUC* or *GFP* reporter gene line parent and BCF2 mutant whole genome sequence data. Genes containing a non-synonymous SNP with AF>0.9 were considered as candidate genes, and genomic regions with AF>0.9 are highlighted in the diagrams.

Mutant line G32



Mutant line G35

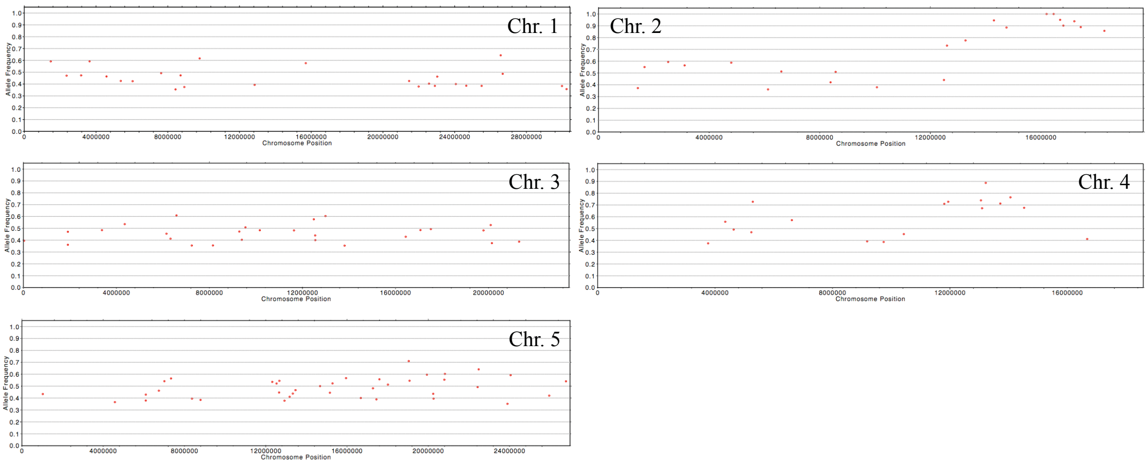


Figure S4. Allelic frequencies for mutant line L02, L03, G21, G32 and G25.
Allelic frequencies (AF) for all SNPs resolved using the *LUC* or *GFP* reporter gene line parent and BCF2 mutant whole genome sequence data. Genes containing a non-synonymous SNP with AF>0.9 were considered as candidate genes, and genomic regions with AF>0.9 are highlighted in the diagrams.

Supporting tables

Table S1. Reporter gene expression of 27 EMS mutant lines, which were used for light microscopic analysis. +, more signal; -, less signal.

Mutant Nr.	LUC/GFP signal
L02	+
L03	+
G01	+
G02	+
G03	+
G04	+
G05	+
G06	+
G07	+
G08	+
G09	+
G10	+
G11	+
G12	+
G13	+
G14	+; wider bundle
G15	+; wider bundle
G16	+; wider bundle
G17	+; wider bundle
G18	+; wider bundle
G19	+; wider bundle
G20	+; wider bundle
G21	-
G22	-
G23	-
G24	diffused
G25	diffused

Table S2. Oligonucleotides used in this study

Bold letters – gateway *attB* sites

Primer	Sequence (5'→ 3')	Orientation
p*GLDPA*-Ft-F	TACTCCTCTCAACTTTCAA	F
p*GLDPA*-Ft-R	AGTGTAAGATGGGGTCTAA	R
p*GLDPA*-Ft+*att*B1	**GGGGACAAGTTTGTACAAAAAAGCAGGCTTA**CTCCTCTCAACTTTCAA	F
p*GLDPA*-Ft+*att*B2	**GGGGACCACTTTGTACAAGAAAGCTGGGTA**GTGTAAGATGGGGTCTAA	R
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

Author contributions

KB and **FD** wrote the manuscript.

KB, FD and **SDG** performed the *GFP* reporter screen –

KB screened 300,000 M2 plants, analyzed 300 primary mutants and contributed 20 stable lines.

FD screened 140,000 M2 plants, analyzed 155 primary mutants and contributed 17 stable lines.

SDG screened 90,000 M2 plants, analyzed 92 primary mutants and contributed 8 stable lines.

KB performed the microscopic work of the selected 25 mutant lines –

1, 2, 21, 22, 23 and 24 mutant lines are from **FD** and 3, 4, 5, 6 and 25 mutant lines are from **SDG** and 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20 mutant lines are from **KB**.

FD performed *LUC* reporter screen.

UG analyzed the sequencing data.

SS, FD and SDG helped with the leaf tissue fixation and embedding.

SS and **RK** helped with the operation of microscopes.

TLS and **RK** helped in interpreting the results.

PW participated in the drafting of the manuscript.

Manuscript III

Activation tagging in *Arabidopsis thaliana* identifies novel *BSOM4* gene as a player in plasmodesmata development.

(I) Introduction

Photorespiration leads to nearly 35 % reduction in photosynthetic efficiency of C_3 plants (Cegelski and Schaefer, 2006; Raines, 2011; Walker *et al.*, 2016), which exceeds even further at higher temperatures (Brooks and Farquhar, 1985). This abated ecophysiological change, led to the evolution of C_4 photosynthesis in order to reduce photorespiration while simultaneously increasing water and nitrogen use efficiency (Ehleringer and Monson, 1993). In C_4 plants, the photosynthetic reaction occurs in two cell compartments; mesophyll and bundle sheath cells, where atmospheric CO_2 is primarily fixed in the mesophyll cells with the help of the phospho*enol*pyruvate carboxylase (PEPC) enzyme. Later, the resultant four-carbon compound (malate or aspartate) is transported to bundle sheath cells and decarboxylated, releasing a molecule of CO_2. This mechanism specifically accumulates CO_2 in the bundle sheath tissue where Rubisco and Calvin cycle are localized (Hatch 1987; Furbank 2011; Sage *et al.,* 2012). Consequently, C_4 plants efficiently over-rule the oxygenase activity of Rubisco, thus reducing the photorespiratory energy and carbon loss. Additionally, this superior photosynthetic machinery is a product of specialized Kranz leaf anatomy, where directly connected bundle sheath and mesophyll cells (Haberlandt, 1904) are organized as concentric rings around each vascular bundle in C_4 leaves. The rapid metabolite flux between these two cell types in-turn is facilitated via well-developed numerous plasmodesmata connections (Botha, 1992; Danila *et al.*, 2016; Danila *et al.*, 2018).

The C_3 pathway of carbon assimilation is evident in the majority of food crops such as rice, wheat, and soybean, just to name a few. Engineering C_4 carbon assimilatory cycle into C_3 crops was proposed to further enhance the photosynthetic efficiency of the latter by at least 50 % (Hibberd *et al.,* 2008) in best case scenario. However, in this complex process of C_4 establishment, development of C_4-Kranz anatomy in C_3 leaves is a prerequisite. Activation of bundle sheath that is characterized by enlarged cells with more organelle content is an earliest evolutionary event and driving force for the gradual development of Kranz anatomy (Gowik and Westhoff, 2011; Sage *et al.*, 2012; Sage *et al.*, 2014; Christin *et al.*, 2013).

In contrast to C_3 species, bundle sheath cells of C_4 plants, bear more and larger chloroplasts, while mesophyll cells carry less than in C_3 plants (Sage *et al.*, 2014; Stata *et al.,* 2014; Stata *et al.,* 2016). This implies the major role of cell-specific regulation in C_4 evolution whose knowledge on the genetic level is still limited. Until now only a few genes are known to differentially regulate mesophyll and bundle sheath cell development. Constitutive expression of maize GOLDEN2-like transcription factors (GLK1/GLK2) led to increased chloroplast and mitochondrial volume in bundle sheath cells of rice (Wang *et al.,* 2017). In reticulate mutants of *Arabidopsis* mesophyll plastid development was shown to be impaired (Kinsman & Pyke 1998; Lundquist *et al.,* 2014)*,* while bundle sheath tissue lost its identity in C_3 *Arabidopsis* and C_4 maize when GRAS family transcription factors SCARECROW and SHORT-ROOT were mutated (Slewinski *et al.*, 2012; Slewinski *et al.*, 2014; Cui *et al.*, 2014).

Although C_4 is a complex trait, (Gowik et al. 2011; Bräutigam *et al.,* 2011), existence of C_4-like anatomical features in closely related C_3 ancestors (Marshall *et al.,* 2007; Muhaidat *et al.,* 2011; Sage *et al.,* 2013; Christin *et al.,* 2013) and its repeated evolution in many angiosperm lineages (Sage, 2016), indicate the possibility of mimicking C_4-Kranz state with minimum changes to existing gene regulatory systems in C_3 plants.

Therefore, in this study we aimed to excavate novel genes involved in regulating bundle sheath activation, using a high-throughput activation-tagging genetic screen in C_3 *Arabidopsis thaliana*. We hypothesize that *Arabidopsis* serves as the best suitable model for isolation of candidate genes regulating bundle sheath specific cell development. In our genetic screen, promoter of the phospho*enol*pyruvate carboxylase A gene (p-$ppcA_{\mathrm{Ft}}$) from a C_4 dicot *Flaveria trinervia* (Stockhaus *et al.,* 1997), specific to C_3 *Arabidopsis* leaf tissue (Akyildiz *et al.*, 2007) served as activation-tag. For ease of detection of mutant lines, *A. thaliana GFP* reporter line whose bundle sheath cells were labeled with chloroplast targeted green fluorescent protein (GFP) (Döring *et al.,* 2018; Manuscript II) served as the genetic background of the activation tagging.

The GFP reporter signal intensity was used as a proxy to primarily select mutant lines with deviated reporter gene signal intensity and stable mutants were analyzed for changes in leaf anatomy. This screen identified one mutant line with two-fold increased GFP signal intensity and AT1G29480 (*BSOM4*) was proven to be responsible for the observed phenotype. Strikingly, microscopic analysis of leaf tissue cells in *bsom4* mutant line revealed abundant plasmodesmata in comparison to the reference line. However, further characterization of *BSOM4* gene suggesting that it might be functioning as a non-coding RNA gene.

(II) Results

Stable AT47 mutant line generated from activation tagging screen

Bundle sheath cells in the reference line (GFP reporter line) were labeled with a chloroplast targeted green fluorescent protein by using the bundle sheath specific $GLDPA_{\mathrm{Ft}}$ promoter and the transit peptide of rubisco small subunit (TP_{rbcS}). (Döring *et al.,* 2018; unpublished, Figure 1A). Since GFP in these lines is transported to individual chloroplasts of bundle sheath cells, we hypothesized that GFP signal intensity might correlate to chloroplast number and/or volume and by then to the cell size. Therefore, *A. thaliana* reference plants were chosen as a genetic background to transform with the activation tagging construct to randomly activate genes in the entire leaf tissue and select candidate mutant lines with altered bundle sheath anatomy. The activation tagging construct contained a $ppcA_{Ft}$ promoter and a *BAR* gene that is driven by a Cauliflower Mosaic Virus 35S promoter ($p35S_{CaMV}$) (Figure 1B). The p-$ppcA_{Ft,}$ which is specifically active in all chlorenchyma tissues (mesophyll and bundle sheath) of *A. thaliana* (Akyildiz *et al.*, 2007) was used to drive the activation of random genes, near to its integration site and a BAR gene that encodes a phosphinothricin acetyltransferase that eases screening transgenic plants against phosphinothricin (BASTA) selection. In this method, the T1 seeds were sown in the soil taken in big trays and watered directly with BASTA solution. Single first leaf (two-three weeks old plants) from each transgenic plant is dissected and observed under the microscope through GFP filter. At this step, transgenic lines showing either a strong increase/decrease in reporter signal intensity relative to the reference line were selected. The selected transgenic lines were further maintained in order to harvest seeds. The mutant phenotypes were analyzed in the following T2 generation for stability and segregation. Only mutant lines that were stable and exhibited a Mendelian single gene inheritance segregation ratio of 3:1 were maintained for detailed analysis. Mutant lines that were neither stable or showed different segregation ratios were discarded.

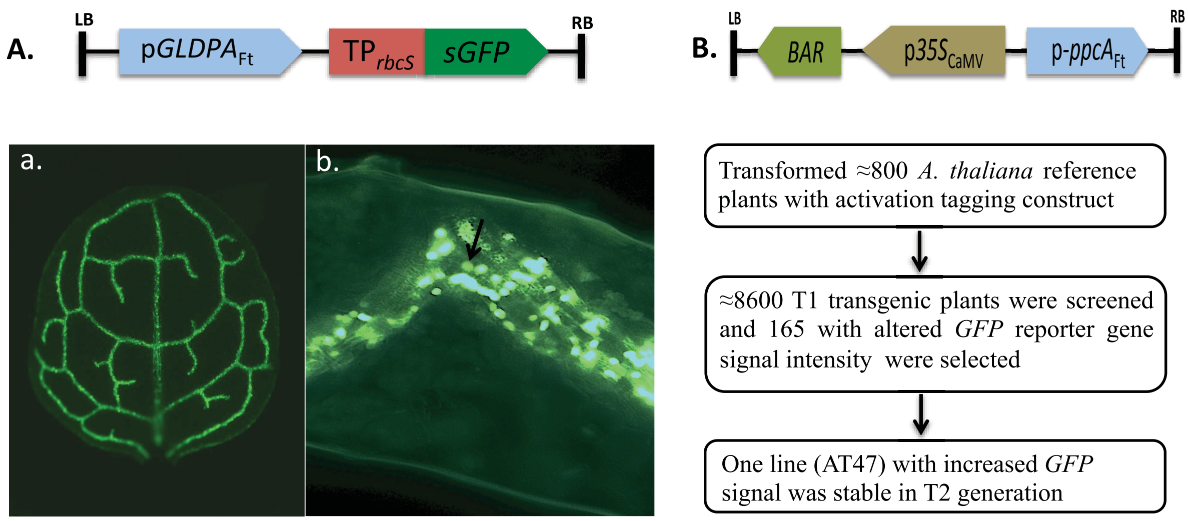


Figure 1: Activation tagging – forward genetic screen.

A. Fluorescence leaf image of *A. thaliana* reference line (**a**) and single labeled chloroplasts (arrow) in leaf longitudinal section (**b**). (Source: F. Döring et al., 2018; unpublished). p*GLDPA*$_{Ft}$ - Promoter of the glycine decarboxylase P subunit gene (*GLDPA*) from a C_4 *Flaveria trinervia;* TP$_{rbcS}$ - transit peptide from Rubisco small subunit; sGFP – green fluorescent protein. **B**. Work-flow of the screen. p-*ppcA*$_{Ft}$ - promoter of the phospho*enol*pyruvate carboxylase A gene (*ppcA*) from a C_4 *Flaveria trinervia*; BAR - BASTA resistance gene; p*35S*$_{CaMV}$ - *35S* promoter from Cauliflower Mosaic Virus.

Overall, approximately 800 *A. thaliana* reference plants were transformed with activation tagging construct. In the T1 generation, about 8600 transgenic plants were screened for altered GFP signal intensity, out of which 165 primary mutant lines with either enhanced or reduced signal intensity were selected. In the T2 generation, out of all 165 lines obtained in the T1 generation, only one mutant line (AT47 - activation tagging line 47) had a stable phenotype with 3:1 segregation ratio (Figure 1B).

T-DNA integration site in the AT47 (*bsom4*) mutant line

AT47 mutant line showed an increased GFP reporter signal intensity in the bundles of *A. thaliana* leaves (Figure 2A). Fluorescence signal intensity was quantified from first leaves of 20 homozygous mutant and reference plants and was approximately two-fold enhanced in the mutant line (Figure 2B) as analyzed using ImageJ software (Schneider *et al.*, 2012). No other abnormalities in externally visible features were observed in the mutant line.

In order to locate the T-DNA integration site of the AT47 line, inverse PCR was carried out. Upon mapping the sequenced PCR product to *Arabidopsi*s whole genome database (TAIR10), the T-DNA region is found to be located inside the coding region of an unassigned gene AT1G29480. It is inserted 15bp downstream to the predicted translational start site ATG (ATG$_{+1}$) and further p-*ppcA*$_{Ft}$ that is part of T-DNA is oriented in 5' to 3' direction as the gene is located on chromosome 1 (Figure 2C).

AT1G29480 is not yet characterized and predicted for encoding any hypothetical protein. According to the predicted gene model (TAIR10), this gene consists of two exons of 174bp and 504bp in length respectively and one 96bp intronic region (Figure 2C). The AT1G29480 gene is now termed as *BSOM4* (bundle sheath ontogeny and morphology gene 4) following the aim of our screen, to find novel genes that are responsible for bundle sheath activation and this was the fourth gene isolated from GFP reporter gene based activation tagging screen in our group. The T-DNA integration site in the mutant line (at +16) might lead to either knockdown or knockout of *BSOM4* gene function if gene prediction is considered to be true. With an aim to verify this, first at the transcriptional level, Semi-Quantitative PCR was conducted by using total RNA from leaves of both reference and mutant lines. While we detected a bright PCR product from the mutant *bsom4*, hardly any product was observed in case of the reference line (Figure S1). Thus insertion of p-$ppcA_{Ft}$ in the *BSOM4* gene resulted in its over-expression.

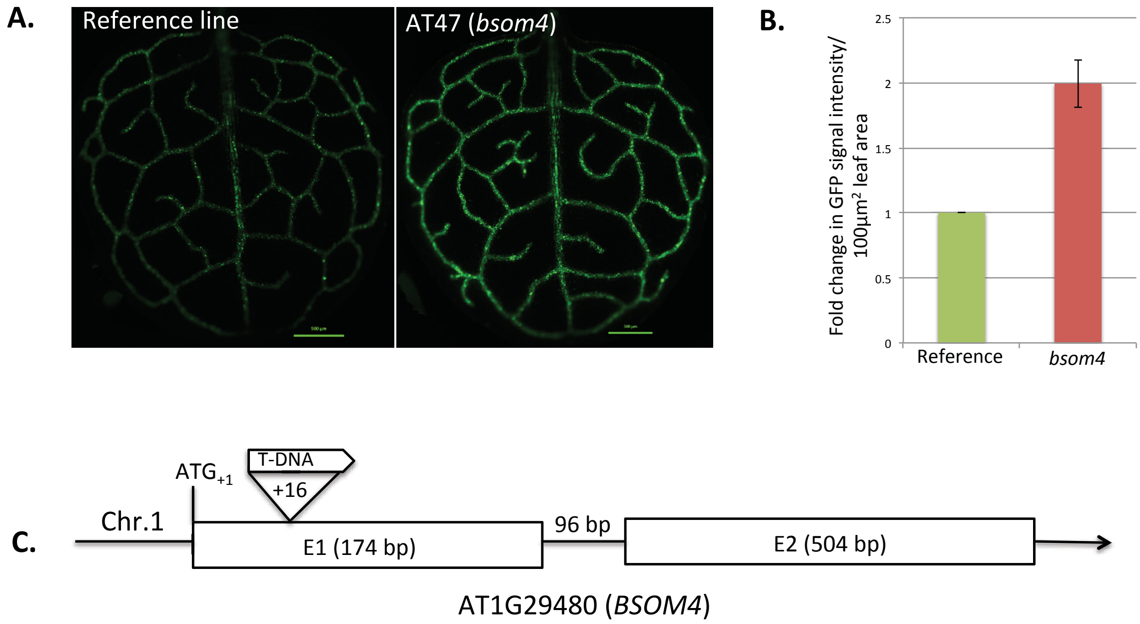


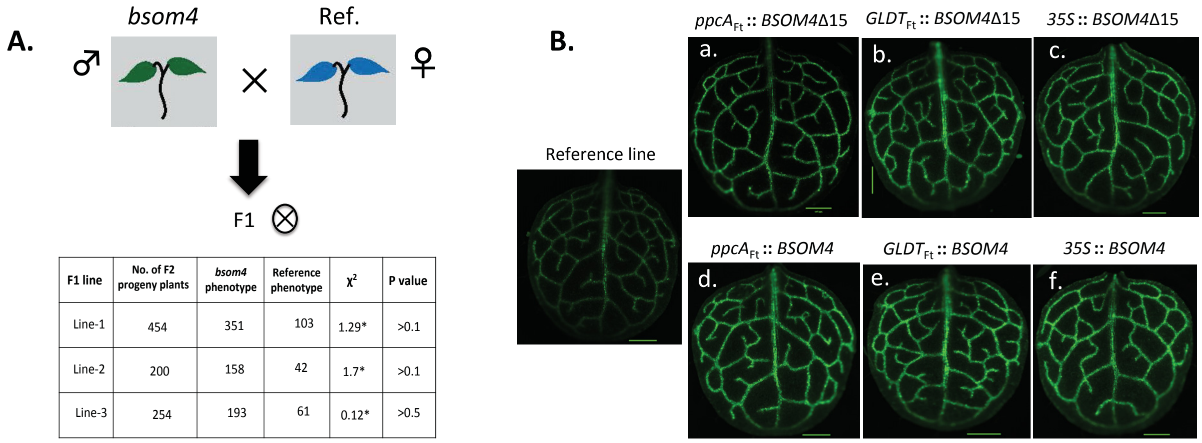
Figure 2: The mutant line *bsom4*
A. Fluorescence leaf images of the reference and *bsom4* mutant line and fold change in GFP signal intensity (**B**). Scale bar – 500 µm. **C.** Representation of the T-DNA insertional event in the *bsom4* mutant line, on chromosome-1.

Co-segregation of *bsom4* mutant phenotype and phenotype recapitulation

To testify if the identified gene (T-DNA integration) is the causative event for *bsom4* phenotype, we performed backcross and overexpression analyses. The homozygous mutant line was backcrossed to reference line and resulting F1 plants were allowed to self-fertilize to generate F2 backcross population. By considering GFP signal intensity as a proxy, progeny of three independent F1 plants were analyzed. Chi-square statistical test was performed and our results ($p > 0.05$) indicated that *bsom4* phenotype is co-segregating with T-DNA in a Mendelian fashion (3:1) (Figure 3A). T-DNA integration in these plants was further confirmed with BASTA resistance marker.

To carry out overexpression analysis, we took the predicted coding sequence avoiding the first 15 bp (*BSOM4*Δ15) as the T-DNA in the mutant line is between +15 and +16 bp and expressed under the control of p-$ppcA_{Ft.}$ Additionally, we expressed *BSOM4*Δ15 in bundle sheath cells using p$GLDT_{Ft}$ and ubiquitously with p*35S*. The promoter p$GLDT_{Ft}$ is a bundle-sheath preferential promoter both in the C_4 species *Flaveria bidentis* and in the C_3 plant *Arabidopsis*, but is also to varying degrees active in the vascular tissue (J. Emmerling, unpublished). The resulted p-*ppcA*$_{\mathrm{Ft}}$::*BSOM4*Δ15, p*GLDT*$_{\mathrm{Ft}}$::*BSOM4*Δ15 and p*35S*::*BSOM4*Δ15 constructs were transformed into the reference line. From each construct at least 30 T1 transgenics were analyzed and in all cases, we were able to reconstruct the mutant phenotype with respect to GFP signal intensity (Figure 3B: a, b and c).

We were also interested in understanding the inclusion of first 15 bps of *BSOM4* gene. Hence, the complete predicted coding sequence was also expressed under p-*ppcA*$_{\mathrm{Ft,}}$ *GLDT*$_{\mathrm{Ft}}$ and p*35S*, and transformed into reference lines. All the three p-*ppcA*$_{\mathrm{Ft}}$::*BSOM4*, p*GLDT*$_{\mathrm{Ft}}$::*BSOM4,* and p*35S*::*BSOM4* overexpression lines resulted in reproducing the mutant GFP reporter signal intensity (Figure 3B: d, e and f). We conclude from these results that the *bsom4* phenotype can be reproduced by either overexpression of the complete predicted coding sequence (*BSOM4*) or by a truncated version (*BSOM4*Δ15). Moreover, expression of *BSOM4* or *BSOM4*Δ15 in the bundle sheath and vascular cells of *A. thaliana* is also sufficient to reconstitute the mutant phenotype.



F1 line	No. of F2 progeny plants	*bsom4* phenotype	Reference phenotype	χ^2	P value
Line-1	454	351	103	1.29*	>0.1
Line-2	200	158	42	1.7*	>0.1
Line-3	254	193	61	0.12*	>0.5

Figure 3: Co-segregation and overexpression analysis
A. Analysis of F2 segregating backcross population. **B.** GFP reporter signal intensity in the reference line and in overexpression lines with truncated (**a, b** and **c**) or complete *BSOM4* sequence (**d, e** and **f**). Scale bar – 500 µm.

Light and Transmission electron microscopic analyses

Until this point, the mutant phenotype is described only concerning GFP signal intensity while occurrences of any alterations in bundle sheath anatomy led by phenotype is still to be answered. To seek this answer, leaf cross and paradermal sections of the mutant and the reference line were prepared for light microscopy and transmission electron microscopic (TEM) observations. Qualitative analyses with light microscopy could not detect any obvious changes in bundle sheath cell anatomy of *bsom4* line in comparison with the reference line. The bundle sheath cell size and/or number seemed to be normal. No differences to the reference line in either chloroplast number and/or volume within the bundle sheath cells were observed, when single cell isolates were analyzed (data not shown).

Nonetheless, we have further performed ultrastructural analysis of mutant and reference line using TEM. No structural changes were detected either in chloroplast or mitochondria. However, the number of plasmodesmata connections found to be increased in all cell types of the *bsom4* mutant (Figure 4A: b, c, and d, respectively). Due to less sample size, we couldn't get good comparable images of mesophyll-bundle sheath and bundle sheath-bundle sheath cell junction of the reference line. Hence, we were only able to provide mesophyll-mesophyll connections from reference line for qualitative comparisons (Figure 4A: a).

For further confirmation of *bsom4* phenotype single cell isolates that were stained with aniline blue fluorochrome were analyzed. This fluorochrome reacts with callose (β-1, 3-glucan), which is a structural component of plasmodesmata and emits a brilliant yellow fluorescence at around 455 nm (Stone *et al.*, 1984; Zavaliev and Epel, 2014). Aniline blue has been widely used to study plasmodesmata connections and callose deposition in different tissues (Radford *et al.*, 1998; Bougourd *et al.*, 2008). As a result of aniline blue staining, we observed more fluorescence signal as dots in the mesophyll and bundle sheath cells of the *bsom4* line when compared to respective cell types of the reference line (Figure 4B: g, h and e, f, respectively). These fluorescence dots represent the callose deposition at plasmodesmata connections and thereby, provide indirect estimation for the number of plasmodesmata connections. To conclude, leaf tissue cells in *bsom4* are connected with numerous plasmodesmata.

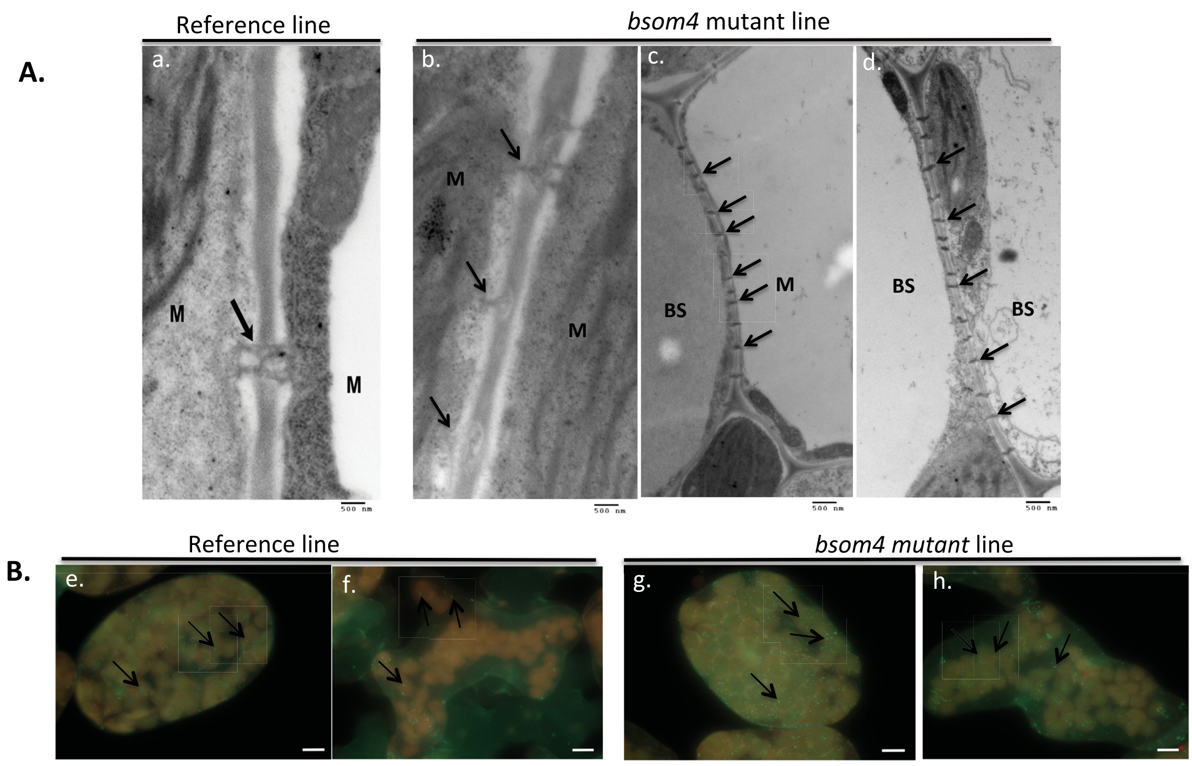


Figure 4: Microscopic analysis of analysis of *bsom4* and reference line
A. Transmission electron microscopic images showing plasmodesmata (PD) connections (arrows) at M-M junction of the reference line (a) and mutant line (b). c, d – PD at BS-M and BS-BS cell interface of *bsom4*, respectively. M – mesophyll cell; BS – bundle sheath cell. Scale bar – 500 nm. **B.** Aniline blue fluorescence at PD connections (arrows). e, g – mesophyll cells and f, h – bundle sheath cells.

Nucleotide sequence conservation of *BSOM4* gene

Based on BLASTN and BLASTP searches (NCBI and Phytozome 12) sequence similarity to *BSOM4* gene could only be detected in members of the Brassicaceae and nowhere else. All these *BSOM4* homologs are uncharacterized, their predicted gene models are shown in Figure 5A**.** Additionally, in *A. thaliana,* part of the intergenic sequence (nucleotide numbers 7822800:7821800; as per TAIR10) residing between AT4G13455 and AT4G13460 genes were found to be conserved to the exon-2 region of *BSOM4*. AT4G13455: a copia-like retrotransposon and AT4G13460: encodes a SU(VAR) 3-9 homolog, a SET domain protein. However, nucleotide sequence similarity of identified genes or genomic contigs to *BSOM4* depicted in Figure 5A. The sequence similarity between *BSOM4* and other Brassicaceae species is consistent with their evolutionary origin (Hohmann *et al.*, 2015).

In order to make better visualization of sequence conservation, predicted coding sequences (excluded introns) were extracted and global pair-wise alignment was performed using AVID alignment program (Bray *et al.*, 2003). The graphical representation shows that exon-2 (175 bp – 678 bp) of *BSOM4* is highly conserved (Figure 5B). As it is visible in Figure 5B, second predicted ORF (from ATG_{+418}) and nucleotide sequence upstream of it, is highly conserved. Here, we observed a gap between two conserved regions of intergenic sequence from chromosome 4 (nucleotide numbers 7822800:7821800), as one can clearly see it in Figure 5B. Our further analysis revealed this gap is of 113 bp in length and doesn't share any sequence similarity to the *BSOM4* gene. However, this 113 bp nucleotide sequence is conserved to the intronic sequence of the other similarity genes from closely related Brassicaceae species (Figure 5C). This intron in these genes is flanked by two conserved exons (red and blue; Figure 5C). Further, the next successive exon in these species starts with ATG that is conserved to ATG_{+418} of second ORF in *BSOM4* gene (Figure 5C).

A.

Species/gene ID	Predicted gene model	Similarity
Arabidopsis thaliana AT1G29480 (*BSOM4*)	174 bp 96 bp 504 bp	–
1. *A. thaliana* (Intergenic*)* Chr. 4: 7822800:7821800	1 kb	49%
2. *Arabidopsis halleri* Scaffold: 40600	294 bp 182 bp 321 bp 102 bp 324 bp	76%
3. *Arabidopsis halleri* Scaffold: 11325	80 bp 97 bp 124 bp 183 bp 315 bp 97 bp 318 bp	71%
4. *Arabidopsis lyrata* LOC9329631	291 bp 183 bp 321 bp 87 bp 324 bp	75%
5. *Arabidopsis lyrata* LOC9329632	291 bp 181 bp 318 bp 102 bp 324 bp	75%
6. *Arabidopsis lyrata* LOC9330342	291 bp 185 bp 318 bp 108 bp 324 bp	73%
7. *Camelina sativa* LOC104743603	273 bp 189 bp 312 bp 98 bp 156 bp	57%
8. *Camelina sativa* LOC104743605	122 bp 111 bp 49 bp 209 bp 32 bp 120 bp 181 bp 96 bp 174 bp	52%
9. *Capsella rubella* LOC17900878	279 bp 163 bp 285 bp 109 bp 219 bp 85 bp 36 bp	51%
10. *Brassica oleracea* LOC106344700	468 bp	46%
11. *Brassica oleracea* LOC106302713	534 bp	44%
12. *Brassica rapa* LOC103840538	594 bp	40%
13. *Brassica rapa* LOC103828966	529 bp 115 bp 26 bp	36%
14. *Arabis alpina* Chr.1 contig	859 bp	34%
15. *Arabis alpina* Chr.7 contig	839 bp	32%

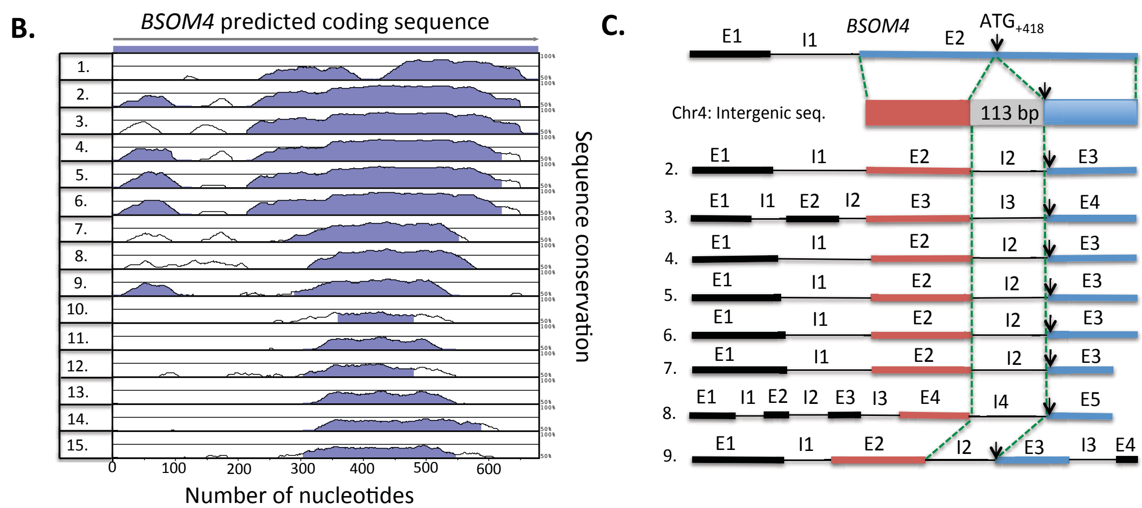


Figure 5: Gene sequences/genomic contigs that shared similarity to *BSOM4* gene
A. Predicted gene models and nucleotide sequence similarity with *BSOM4* gene. The similarity was based on pair-wise nucleotide sequence alignment using DiAlign (http://www.genomatix.de/cgi-bin/dialign/dialign.pl). **B.** mVISTA plot, depicting nucleotide sequence conservation of the predicted *BSOM4* coding sequence, based on AVID alignment program. **C.** Diagrammatic representation of the sequence conservation of intergenic sequence from chromosome-4 with *BSOM4* and its similar sequences. E - exon; I - intron; color codes (red and blue) indicate conservation between respective sequences and arrow marks representing the conserved second ATG. Green dotted lines indicating conservation of intergenic sequence to exon-2 of *BSOM4* and to the intronic sequence of the genes from other Brassicaceae species. In **B and C** – numbers are in correspondence to **A.**

The ambiguity in the predicted gene model of *BSOM4*

The T-DNA integration (at +16) in the *bsom4* mutant line could destroy translation of *BSOM4* from the predicted first ATG (ATG_{+1}) but we were able to mimic the mutant phenotype with respect to GFP signal intensity either by overexpression of a complete (*BSOM4*) or truncated (*BSOM4*Δ15) predicted coding sequence (Figure 3B). In addition to this, sequence alignment revealed that exon-2 is rather conserved (Figure 5B). This drove us to question, whether the predicted *BSOM4* gene model is true. Hence, we performed 5' RACE (Rapid Amplification of cDNA Ends) experiments in order to verify if intron is being spliced and the predicted two exons are fused together.

5' RACE was carried out as outlined in Materials and Methods using total RNA from leaves of both reference and *bsom4* lines. Primers binding sites are located in the exon-2 region (Figure 6A). No specific product could be obtained from the reference line reinforcing our previous observation that no *BSOM4* transcripts could be detected by semi-quantitative PCR (Figure S1). The 5' RACE experiments with RNA from *bsom4* yielded one major and three minor products. These products were cloned and transformed into *E. coli*. Ten random clones were sequenced from each product. The sequencing of the major product revealed that it is resulting from a splicing event, in which the intron from *BSOM4* gene and the intron located in the 5' leader of the *ppcA* gene that is contained in the p-*ppcA*$_{\mathrm{Ft}}$ (Stockhaus *et al.*, 1997; Ernst and Westhoff, 1997) has been spliced out (Figure 6B- a). In this major product transcription start site of the p-*ppcA*$_{\mathrm{Ft}}$ (Figure 6B-a) is consistent with what has been reported earlier (Ernst & Westhoff 1997). Whereas minor products are resulting from the unspliced intron of either the *BSOM4* gene or the p-*ppcA*$_{\mathrm{Ft}}$ or from the initiation of transcripts further upstream of the originally reported transcription start site of the p-*ppcA*$_{\mathrm{Ft}}$ (Figure 6B-b, c, and d, respectively). In general, we found no situation where the intron in both cases is not spliced out. This might suggest that the unspliced amplicons are not resulting from genomic DNA contamination but might be resulting from the on-going splicing event. To conclude, *BSOM4* contains a functional intron and the predicted gene model of it might be partially true.

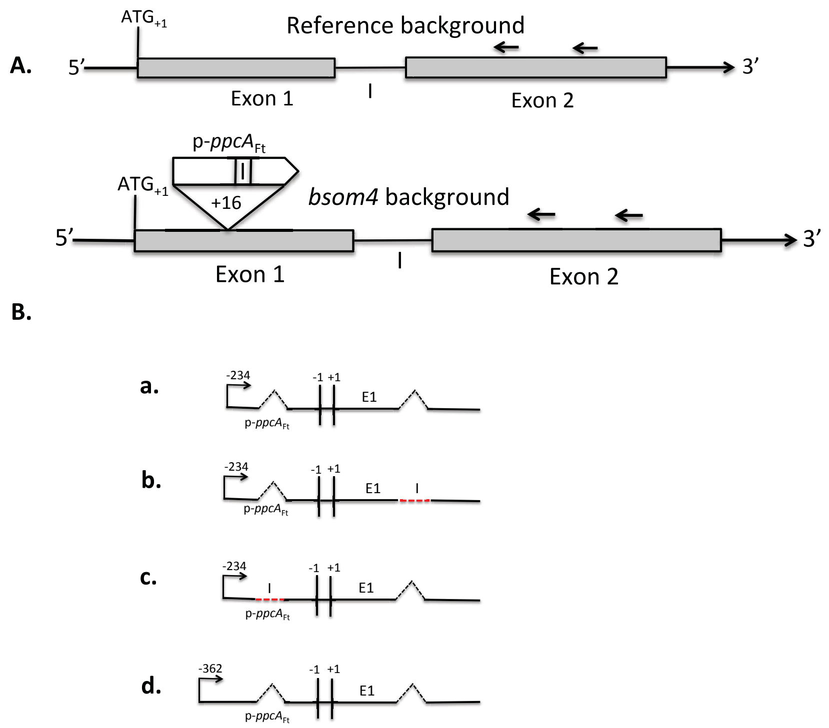


Figure 6: 5' Rapid amplification of cDNA ends (5' RACE)
A. 5' RACE was done using leaf RNA from both reference line and mutant line *bsom4*. Arrows represent primer-binding sites. **B.** 5' RACE sequencing result from major PCR product (**a**) and minor products (**b**, **c**, and **d**) from the *bsom4* mutant line. '+1' is in correspondence to ATG$_{+1}$ of *BSOM4* and '–1' with respect to phospho*enol*pyruvate carboxylase A gene promoter sequence (p-*ppcA*$_{\mathrm{Ft}}$). The unspliced intron is represented with the dotted red line. Forward arrows represent the position of transcription start site. E1 – exon 1.

No evidence that *BSOM4* encoding a protein

The above experiments revealed that the predicted ATG$_{+1}$ is not necessary for *BSOM4* function and that the two predicted exons are combined by a splicing event. So the question arose whether the second ATG (ATG_{+418}) that is located within exon-2 of *BSOM4* could act as a translational start site. Hence, second ORF (418 bp to 678 bp) (Figure 7B) was expressed under the control of p-*ppcA*$_{Ft}$, p*GLDT*$_{Ft}$, and p*35S* by transforming into the reference line. From each construct, at least 30 T1 transgenic lines were analyzed with respect to GFP signal intensity. Transgenics lines from all the three overexpression lines (p-*ppcA*$_{Ft}$::*BSOM4*Δ417, p*GLDT*$_{Ft}$::*BSOM4*Δ417 and p*35S*::*BSOM4*Δ417) showed similar GFP signal intensity to that of the reference line (Figure 7C and 7D). The mutant phenotype was thus not recapitulated by overexpression of the second ORF alone leading us to hypothesize that *BSOM4* gene might not code for a functional protein.

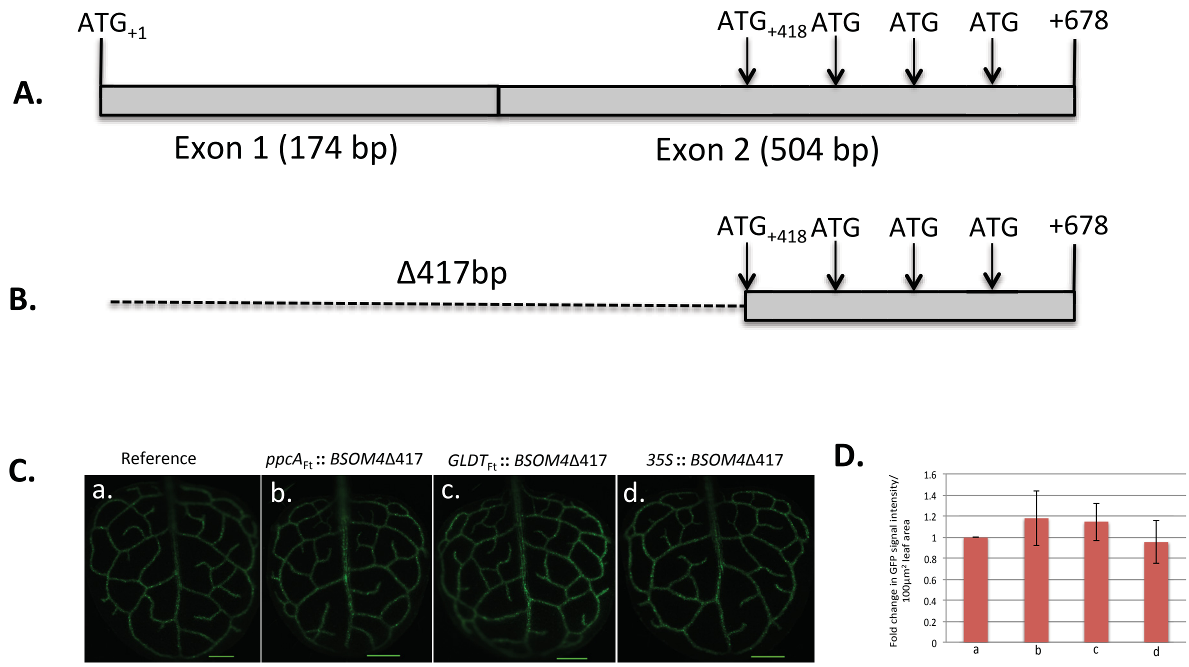


Figure 7: Overexpression of second predicted open reading frame (ORF)
A. Diagrammatic representation of all the predicted ORFs within *BSOM4* coding sequence that are in-frame to ATG_{+1}. Predicted second ORF was overexpressed (*BSOM4*Δ417) (**B**) under the control of *ppcA*, *GLDT,* and *35S* promoters. **C.** GFP fluorescence in whole leaves of reference (**a**) and *BSOM4*Δ417 deletion lines (**b, c, and d**) of *A. thaliana* and measured signal intensity from first leaves of 20 T1 plants (**D**).

To strengthen the hypothesis further protein detection experiments were designed. The complete *BSOM4* and the truncated *BSOM4*Δ15 reading frames were fused in frame at the carboxy-terminus to the 3XHA-StrepIII-2XpA tag as present in the pAUL3 vector (Lyska *et al.*, 2013), and expressed under the control of the *35S* promoter. The chimeric genes were stably transformed into wild-type *A. thaliana* and western blot analysis was carried out with total proteins from leaves to detect protein-A (PA) epitope by using an anti-mouse IgG that was conjugated with peroxidase. No specific signals could be detected at the expected positions in the western blot (data not shown).

In addition, the complete *BSOM4* and the truncated *BSOM4*Δ15 reading frames were fused in frame at the carboxy-terminus to *YFP* and the fusion constructs were transiently expressed in *Nicotiana benthamiana* and also stably transformed into *A. thaliana*. Also, in this case, no specific fluorescence signals indicating the production BSOM4 fusion protein could be detected. These results reinforce the notion that the *BSOM4* gene is not expressed into protein suggesting that the gene functions at the RNA level only.

To further confirm that *BSOM4* is not encoding any protein, the reference line was transformed with p*GLDT*$_{Ft}$::*BSOM4** construct, in which all in-frame ATG's of *BSOM4* were mutated by replacing Guanine nucleotide with Adenine and expressed under the control of the p*GLDT*$_{Ft}$. Thus, ORF structure of the *BSOM4* gene destroyed by replacing all possible translational start codons (ATG - codes for Methionine) with ATA (codes for Isoleucine) (Figure 8B). In the T1 generation, minimum 30 lines were analyzed, also referring to phenotype only with respect to GFP signal intensity. There is no difference in the reporter signal intensity of the p*GLDT*$_{Ft}$::*BSOM4** line and the *bsom4* mutant line (Figure 8C and 8D). To conclude, mutant phenotype was recapitulated by *BSOM4** sequence, which contains no functional ATG and therefore, strongly suggesting that *BSOM4* functions as a non-coding RNA.

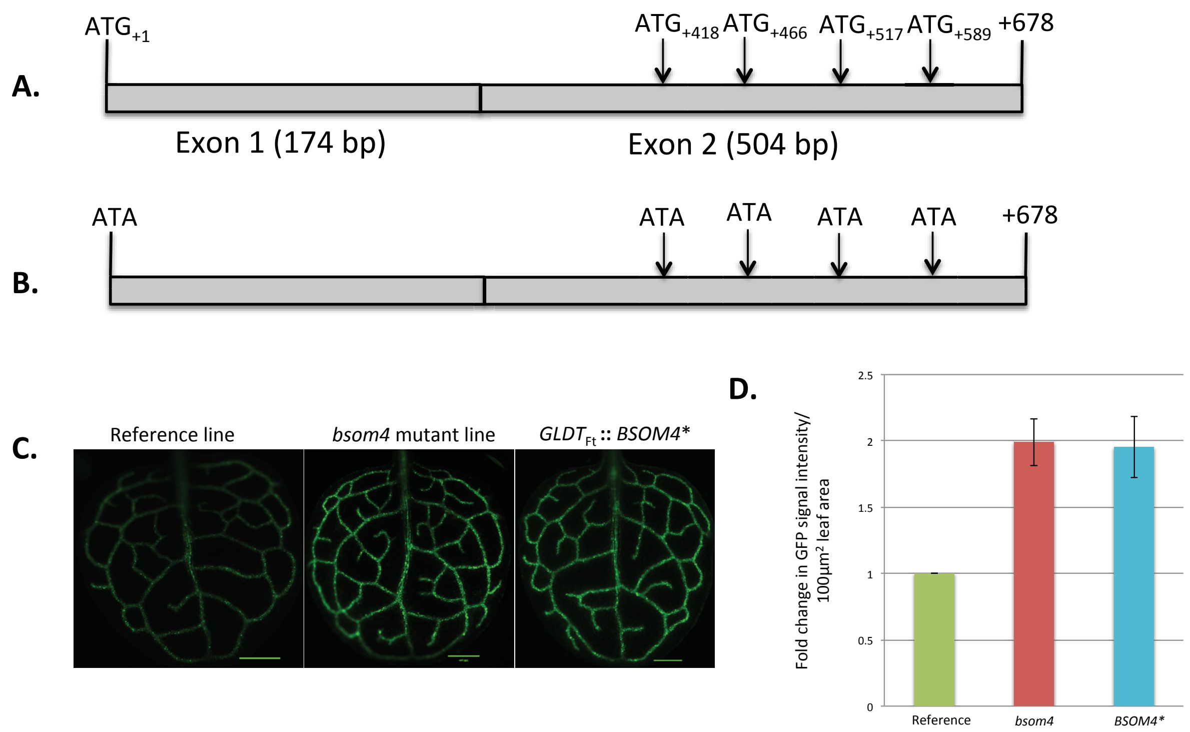


Figure 8: Overexpression of *BSOM4 sequence**
A. Representation of *BSOM4* sequence with all in-frame ATGs and replacement of all ATGs with ATA (**B**). **C.** Fluorescence leaf images of the reference, mutant and the p*GLDT*$_{Ft}$::*BSOM4** overexpression lines and GFP signal intensity was measured from first leaves of 20 T1 transgenics plants and relative signal intensity was depicted (**D**).

Analysis of *BSOM4* deletion constructs

It was further interesting to know which region of the *BSOM4* nucleotide sequence is important for its function. Therefore, deletion constructs were prepared with the deletion of 90 bp, 144 bp, 174 bp (E1) and 270 bp of *BSOM4* and by expressing rest of the region under the control of p*GLDT*$_{Ft}$ (Figure 9B). In addition, one more construct was generated in which the only region between T-DNA insertion and predicted second ORF was expressed (*BSOM4*:16-417; Figure 9B). The resulted constructs were transformed into the reference line and at least 30 T1 lines from each construct were analyzed with respect to GFP signal intensity. The transgenics from p*GLDT*$_{Ft}$::*BSOM4*Δ90, p*GLDT*$_{Ft}$::*BSOM4*Δ144 and p*GLDT*$_{Ft}$::*BSOM4*ΔE1 overexpression lines showed no difference from the phenotype of *bsom4* line. Whereas further deletion of exon-2 sequence in p*GLDT*$_{Ft}$::*BSOM4*Δ270 overexpression line, resulted in an approximately 50 % reduction in GFP signal intensity (Figure 9C, D). Furthermore, the mutant phenotype was not generated by the overexpression of *BSOM4*:16-417 nucleotide region alone (Figure 9C, D). These results lead us to conclude that exon-1 sequence is not necessary for *BSOM4* function and the complete exon-2 region is crucial for *BSOM4* function.

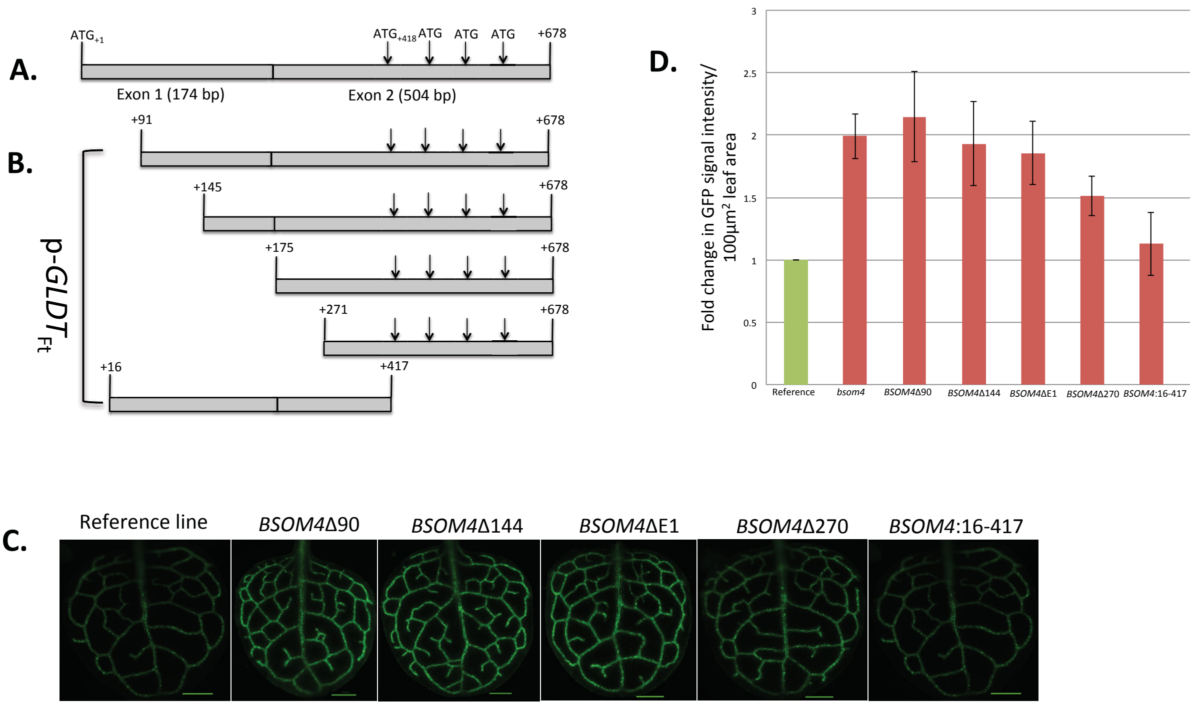


Figure 9: Overexpression of deletion constructs
A. Representation of all predicted ORFs from ATG **B.** Depicts *BSOM4* deletion constructs that were made in this study. **C.** Fluorescence leaf images of the reference and *BSOM4* deletion lines. GFP signal intensity was measured from first leaves of 20 T1 transgenics plants and fold change was depicted (**D**).

Overexpression of genes from other Brassicaceae species

Further to know the conservation of *BSOM4* gene function, we chose to overexpress *BSOM4* orthologous genes from closely related Brassicacean species, i.e. *Arabidopsis lyrata* (LOC9329632) and *Camelina sativa* (LOC104743603) and from the more distantly related *Brassica rapa* (LOC103828966) (Hohmann *et al.*, 2015). These genes share sequence conservation only part of exon-2 region of *BSOM4* (Figure 5B). Predicted coding sequences of these genes were expressed under the $GLDT_{\mathrm{Ft}}$ promoter and transformed into reference line. p$GLDT_{\mathrm{Ft}}$::AL-LOC9329632 and p$GLDT_{\mathrm{Ft}}$::CS-LOC104743603 overexpression lines could mimic the *bsom4* phenotype while p$GLDT_{\mathrm{Ft}}$::BR-LOC103828966 overexpression line showed only a partial recapitulation of the mutant event (Figure 10). This could be explained by their sequence conservation, the gene sequence from *B. rapa* being less conserved (Figure 5B). These experiments indicate that sequence conservation to exon-2 of *BSOM4* is sufficient and crucial to generate the *bsom4* phenotype.

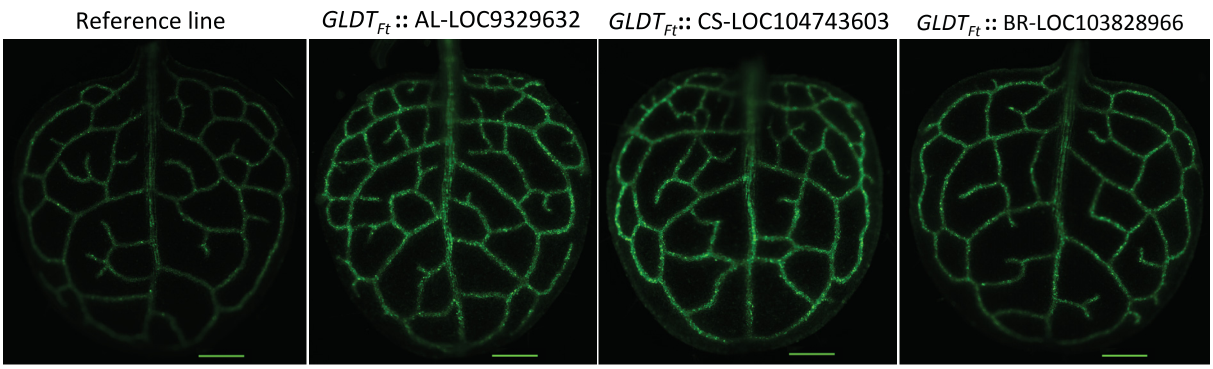


Figure 10: Overexpression of *BSOM4* similar sequences
The GFP reporter signal intensity of overexpression lines (**B, C, and D**) compared to reference (**A**) line. AL-LOC9329632 – gene from *Arabidopsis lyrata*; CS-LOC104743603 – gene from *Camelina sativa;* BR-LOC103828966 – gene from *Brassica rapa*.

(III) Discussion

Activation tagging approach is a gain-of-function mutagenesis screen and is advantageous to dissect the functional annotation of gene families. In this method, randomly integrated activation-tag i.e. a promoter or enhancer element activates nearby genes to its integration site. However, this method might also generate knock out mutants depending on T-DNA integration site within the genome. Hayashi *et al*. in (1992) developed activation tagging in *Nicotiana tabacum,* using enhancer elements from cauliflower mosaic virus *35S* gene and later it has been widely used across several plant species (Weigel *et al.,* 2000; Nakazawa *et al.,* 2003; Tani *et al.,* 2004; Jeong *et al.,* 2006). Further, the role of micro RNAs (miRNA) in plant development was first evidenced by using activation tagging (Palatnik *et al.,* 2003; Aukerman & Sakai 2003). However, the constitutive nature of *35S* enhancer or promoter generates ectopic expression of the endogenous gene in irrelevant tissues. This might sometimes lead to lethality or misinterpretation of a gene function (Kondou *et al.*, 2010). This can be avoided by using tissue-specific enhancers or promoters. Hence, we chose promoter from phospho*enol*pyruvate carboxylase A gene of the C_4 plant *Flaveria trinervia* (p-$ppcA_{\mathrm{Ft}}$) as an activation-tag, because this promoter is specific to the leaf tissue cells of *A. thaliana* (Akyildiz *et al.*, 2007).

The present study, using the $ppcA_{\mathrm{Ft}}$ promoter (Stockhaus *et al.*, 1997; Akyildiz *et al.*, 2007) for random activation of genes in the whole leaf tissue cells of *Arabidopsis thaliana* reference line background (GFP reporter line), resulted so far in one stable mutant line - *bsom4*. Our primary criterion to select for bundle sheath anatomy mutants was based on GFP signal strength. The *bsom4* mutant line exhibited 100 % enhanced GFP reporter signal intensity in comparison to the reference line (Figure 2B). The activation-tag, p-$ppcA_{\mathrm{Ft}}$ was localized 15 bp downstream to the predicted ATG_{+1} of *BSOM4* gene (Figure 3C) gene in this mutant line. This gene contains two predicted exons and one intron. The T-DNA insertional event led to the activation of this gene and as a result of this more transcripts were generated from *BSOM4*, while transcripts in leaves of the reference line were undetectable (Figure S1). Several published RNA-Seq datasets (NCBI - short read archive) from leaves of *A. thaliana* (Columbia-0) were used to check the expression pattern and no reads were found that mapped to *BSOM4*. Also, no transcripts of *BSOM4* were detected in leaves of different *A. thaliana* accessions (Kawakatsu *et al.*, 2016). Further, raw reads from published transcriptome dataset of different developmental stages of *A. thaliana* (Klepikova *et al.*, 2016) were mapped to *BSOM4* gene sequence and its transcripts were detected only in anthers. Additionally, *BSOM4* similar sequences from *A. lyrata* were also found to be expressed only in inflorescence material (Rawat *et al.*, 2015) while no transcripts were detected in leaves or shoot apical meristem.

The analysis of F2 segregating backcross population (Figure 3A) and recapitulation of the mutant event with overexpression of either complete or truncated predicted coding sequence of *BSOM4* ((*BSOM4*/*BSOM4*Δ15) under the control of p-$ppcA_{\mathrm{Ft}}$ (Figure 3B: d and a), allowed us to conclude that *BSOM4* gene was responsible for the *bsom4* phenotype. In addition, overexpression of *BSOM4* or *BSOM4*Δ15 with *GLDT*$_{\mathrm{Ft}}$ or *35S* promoter recapitulates the *bsom4* phenotype (Figure 3B: e, b and f, c). The *GLDT*$_{\mathrm{Ft}}$ promoter is exclusively active only in the bundle sheath cells and to various degrees in the vascular tissue cells of *A. thaliana* (J. Emmerling, unpublished). Hence, analysis of p*GLDT*$_{\mathrm{Ft}}$**::***BSOM4*/*BSOM4*Δ15 overexpression lines suggesting that the activity of *BSOM4* within these cell types is sufficient for reproducing the same mutant phenotype, i. e also in the leaf's mesophyll cells, as with the *ppcA*$_{\mathrm{Ft}}$ and 35 S promoters.

To investigate whether the increased GFP signal intensity is associated with any anatomical changes of bundle sheath cells, internal leaf anatomy of reference and *bsom4* lines was assessed with light and transmission electron microscopy. This provided striking insights on the numbers of plasmodesmata in *bsom4* mutant line. It was interesting to observe an increased number of plasmodesmata connections in *bsom4* mutant line between all chlorenchyma cell types (mesophyll-mesophyll, bundle sheath-mesophyll, and bundle sheath-bundle sheath) when compared to the reference line (Figure 4). Further quantitative estimation is strongly suggested. The possible reason for more plasmodesmata in the entire leaf tissue as a consequence of activated *bsom4* might be the activity of the *ppcA*$_{\mathrm{Ft}}$ promoter in all chlorenchyma cells of *Arabidopsis* (Akyildiz *et al.*, 2007). Although the exact reason for increased GFP signal intensity in *bsom4* is not clear yet. Nonetheless, high density of plasmodesmata between mesophyll to bundle sheath cells is one of the key characteristic features of C_4 leaf anatomy (Botha, 1992; Danila *et al.*, 2016; Danila *et al.*, 2018).

Phenotype recapitulation by overexpression of *BSOM4* or *BSOM4Δ15* suggests that ATG$_{+1}$ is not important for its function and sequence alignment of *BSOM4* with its similar sequences showed that exon2 nucleotide sequence is highly conserved (Figure 5B). This rendered *BSOM4* gene prediction for encoding a hypothetical protein (TAIR10) questionable, as there is no second ATG in exon-1 to function as a translational start codon. The second in-frame ATG is located in exon-2 region, at ATG$_{+418}$ (with respect to ATG$_{+1}$). However, a 5' RACE experiment revealed that intron of *BSOM4* is functional (Figure 6B) and therefore, the assumed gene prediction might be partially true. The absence of second in-frame ATG between ATG$_{+1}$ and ATG$_{+417}$ and high sequence conservation of a second predicted ORF (from ATG$_{+418}$) allowed to hypothesize that the second ORF alone could be sufficient to mimic the *bsom4* phenotype. However, differing to this hypothesis, overexpression of second ORF failed to generate the mutant GFP signal intensity (Figure 7). Additionally, testing directly for the occurrence of a BSOM4 protein by western blotting and translational *YFP* reporter gene fusions did not yield any evidence for the expression of *BSOM4* into protein. The above results suggested that *BSOM4* does not encode any protein but rather functions as non-coding RNA. This was further supported by the recapitulation of the mutant phenotype with the overexpression of *BSOM4** sequence, in which ORF structure was destroyed by replacing all in-frame ATGs with ATA codon (Figure 8). In plants, CTG may also act as a translational start codon in few cases (Simpson *et al.*, 2010). In case of *BSOM4*, one predicted ORF was detected in exon-1 region (from CTG_{+145}). However, the exon 1 sequence was shown to be not necessary for generating the *bsom4* mutant phenotype, the overexpression of exon 2 alone was sufficient (Figure 9). However, further deletion of 96 nucleotides from exon-2 (*BSOM4*Δ270; respect to ATG_{+1}) drastically reduced the GFP signal intensity by about 50 % in comparison to signal strength of the *bsom4* mutant line. Moreover, deletion of 3' nucleotide sequence of the exon-2 region (418 – 678 bp) failed to mimic the *bsom4* phenotype (Figure 9). This implies the importance of complete exon2 nucleotide sequence for *BSOM4* function and that exon1 is devoid of any such pivotal functional attributes. Beyond the over-accumulation GFP phenotype, *bsom4* line also possesses plentiful plasmodesmata connections between all leaf tissue cells. Due to time constraints and the "non-availability of an expert in electron microscopy at the University of Duesseldorf" we have not tested whether the GFP over-accumulation phenotype is always paralleled by the occurrence of enhanced numbers plasmodesmata. In C_4 species, the division of labor beholds the responsibility for enhanced plasmodesmata density at mesophyll and bundle sheath cell interface (Danila *et al.*, 2016; Danila *et al.*, 2018). The genetic information on the regulation of plasmodesmata number is still inadequate. Therefore, it would be advantageous to further analyze p*GLDT*$_{\mathrm{Ft}}$**::***BSOM4*/*BSOM4*Δ15 overexpression lines verifying whether the specificity of p*GLDT*$_{\mathrm{Ft}}$ restricts *bsom4* phenotype to mesophyll-bundle sheath cell interface.

Our results strongly suggesting that the *BSOM4* gene functions as a non-coding RNA. In general non-coding RNAs longer than 200 nucleotides are termed long non-coding RNAs, to distinguish them from the familiar short non-coding RNAs, i. e. tRNAs, microRNAs (miRNAs) and small nucleolar RNAs (snoRNAs). Most of the long non-coding RNAs (lncRNAs) are transcribed by RNA polymerase II so they are 5'-capped, 3'-polyadenylated and may contain functional introns (Erdmann *et al.*, 2000; Wen *et al.*, 2007; Quinn and Chang, 2016). Thus, their primary structure is indistinguishable from coding mRNAs. Their function is highly conserved at the level of secondary structure (Li *et al.*, 2016; Wang and Chekanova, 2017). This feature is commonly used to find orthologous lncRNAs (Quinn and Chang, 2016). They function in several aspects of gene regulation across animal and plant species. They may act as scaffolds in recruiting chromatin-modifying protein complexes, function as a decoy to mimic miRNA target sites, serve as precursor molecules for miRNAs (Quinn and Chang, 2016; Wang and Chekanova, 2017) or may function as enhancer molecules (eRNAs) (Li *et al.*, 2016). However, *BSOM4* gene sequence could not be related to any known family of non-coding RNAs, based on searches with the non-coding RNA web-tools PLncDB, GREENC, NONCODE, CANTATAdB, PNRD, PlantNATsDB, and Rfam. Further, no known miRNA motifs were detected (miRBASE) within this gene. Finally, comparison of *BSOM4* RNA structure with its similar sequences from other Brassicaceae species did not inform about any noticeable RNA structural features (Web-tools: mfold and RNAalifold).

In the present study, most of our analyses were done by using *bsom4* GFP signal intensity as mutant phenotype and not by the use of the plasmodesmata phenotype. These analyses are urgently needed to shed light on the function of the *BSOM4* gene, considering the fact that this is only found in the Brassicacean family.

(IV) Materials and methods

Plant transformations and growth conditions

The generated constructs in our study, containing the gene of interest were verified using restriction digestion and Sanger sequencing (LGC Genomics). Later, these constructs were mobilized into *Agrobacterium tumefaciens* GV3101 strain by electroporation (Mersereau *et al.*, 1990), and transformed into *Arabidopsis thaliana* (Ecotype Columbia-0) following the floral dip method (Logemann *et al.*, 2006). Seeds were harvested from the plants grown either in greenhouse conditions of 14h light/day at a photon flux density (PFD) of ~300 µmol m^{-2} s^{-1} and at 21-22 °C or from growth chambers operating at 16h light/day (PFD - ~70-100 µmol m^{-2} s^{-1}) and at a constant temperature of 21-22 °C.

Activation tagging construct and selection of transgenic plants

To generate the activation tagging construct - pMDC123-p-*ppcA*$_{\mathrm{Ft}}$, first, the *ppcA*$_{\mathrm{Ft}}$ promoter region (Stockhaus, 1997; Akyildiz *et al.*, 2007), a 2.181 kb 5' flanking region of phospohoenolpyruvate carboxylase A gene from *Flaveria trinervia* (p-*ppcA*$_{\mathrm{Ft}}$) was synthesized and cloned into a pUC57 vector (Biomatik). To the respective 5' and 3' ends of the cloned p-*ppcA*$_{\mathrm{Ft}}$ sequence; *SacI* and *PmeI* restriction sites were added. The resulting in pUC57-p-*ppcA*$_{\mathrm{Ft}}$ plasmid and pMDC123 Gateway vector (Curtis and Grossniklaus, 2003) were subsequently digested with *SacI* and *PmeI* to clone the released p-*ppcA*$_{\mathrm{Ft}}$ fragment into a pMDC123 vector and thereby gateway cassette was replaced with p*pcA*$_{\mathrm{Ft}}$ promoter region. Further, *Arabidopsis thaliana GFP* reporter gene line (Döring et al., 2018, unpublished) served as a genetic background for transformation with the activation-tagging construct - pMDC123-p-*ppcA*$_{\mathrm{Ft}}$. In the next step, T1 generation seeds were sown on soil and directly watered at two-three different times with BASTA solution (Bayer Agrar, Germany), containing 80-100 mg/lit BASTA and 0.1 % (v/v) Tween 20. Finally, the selected transgenic plants were screened for deviated GFP signal intensity using a GFP filter (Axio Imager M2m, Zeiss, Oberkochen, Germany).

Isolation of T-DNA flanking sequence

T-DNA flanking sequence of the *bsom4* line was isolated by inverse PCR (iPCR) method. For this purpose, genomic DNA (gDNA) was extracted from the T2 generation plants as described by Edwards *et al.,* (1991). A modified method from Earp *et al.,* (1990) was used to prepare the template for iPCR. In this method, 2.5 µg of gDNA was used as a starting material to be digested with *HphI* enzyme in a final reaction volume of 30 µl. Next, without any further cleanups, these digested DNA fragments were allowed to self-ligate using T4 DNA ligase enzyme. Approximately 10 U of T4 ligase was added to the 30 µl-digested product with final reaction volume adjusted up to 250 µl using T4 ligase buffer and deionized water. Finally, the ligated DNA products were precipitated by incubating with three volumes of 100 % cold ethanol and 0.1 volume of 3 M sodium acetate (pH-5.6) at -80 °C for 30 min. The incubated product was centrifuged and cleaned up with chilled 70 % ethanol and allowed to dry. Later, the dried pellet was dissolved in 20 µl water and used as a template for iPCR. Nested PCR was performed using P1, P2 and P3, P4 primer pairs, located at 3' end of the *ppcA*$_{\mathrm{Ft}}$ (Supplementary Table 1). Then the resulting PCR product was cloned into pJET1.2 a vector (Thermo Scientific) and sequenced (LGC Genomics). Finally, the T-DNA integration site was identified following a BLAST search against the *A. thaliana* genome (TAIR10) sequence.

Overexpression constructs

The complete predicted coding sequence of *BSOM4* (AT1G29480) or truncated version (*BSOM4*Δ15) was expressed under the control of *p-ppcA*$_{\mathrm{Ft}}$, *GLDT*$_{\mathrm{Ft}}$, and *35S*$_{\mathrm{CaMV}}$ promoters. The p-*ppcA*$_{\mathrm{Ft}}$ region was amplified from the pUC57-p-*ppcA*$_{\mathrm{Ft}}$ plasmid using specific primers (P5, P6; Supplementary Table 1) to add *HindIII* and *AscI* restriction sites to 5' and 3' ends of the promoter sequence, respectively. The p-*ppcA*$_{Ft}$ and pAUL1 gateway vectors (Lyska *et al.*, 2013) were digested with the same enzymes preceding to the cloning of p-*ppcA*$_{Ft}$ into pAUL1 by a subsequent ligation reaction, resulting into a pAUL1-p-*ppcA*$_{Ft}$ plasmid. The pAUL1-p*GLDT*$_{Ft}$ plasmid was kindly provided by F. Döring, in which 3.2 kb upstream sequence of a gene encoding glycine decarboxylase T subunit (*GLDT*) from *F. trinervia* was cloned into a pAUL1 vector.

Since *BSOM4* transcript levels are nearly undetectable in the rosette leaves of the reference line (*GFP* reporter line), the predicted complete coding sequence (CDS) of *BSOM4* gene (ATIG29480; TAIR10) was amplified from the mutant *bsom4* background, with specific primers (P7/P8), harboring 5'attB1 and 3'attB2 gateway cloning sites. The initial 15 bp (respect to ATG$_{+1}$) before the T-DNA integration site were added to forward primer. Firstly, total RNA was extracted from leaves (RNeasy Plant Mini Kit; QIAGEN) and cDNA was synthesized using by following the manufacturer's instructions (QuantiTect®Reverse Transcription Handbook, QIAGEN) and the amplified *BSOM4* CDS was cloned into a pJET1.2 vector (Thermo Scientific). In next step, *BSOM4* CDS was cloned into the destination vector according to the standard gateway cloning protocol (Thermo Scientific). A BP recombination event between pJET1.2-*BSOM4* and pDONOR^{TM}207 created an entry clone and a consecutive LR reaction between this entry clone and pAUL1-p-*ppcA*$_{Ft}$, pAUL1-p*GLDT*$_{Ft}$ and pMDC123 plasmids respectively, resulted in pAUL1-p-*ppcA*$_{Ft}$::*BSOM4*, pAUL1-p*GLDT*$_{Ft}$::*BSOM4*and pMDC123-*35S*::*BSOM4* overexpression constructs. In the same manner, the truncated version of *BSOM4* (*BSOM4*Δ15) was amplified (P9/P8) and cloned into relevant pAUL1 destination vectors. This resulted in generating truncated, pAUL1-p-*ppcA*$_{Ft}$::*BSOM4*Δ15, pAUL1-p*GLDT*$_{Ft}$::*BSOM4*Δ15 and pAUL1-*35S*::*BSOM4*Δ15 overexpression constructs, respectively. The nature of pAUL1 allowed in fusing *BSOM4* or *BSOM4*Δ15 to its C-terminal 3xHA tag. In this experiment, pMDC123-35S::*BSOM4* was served as a control, to be sure that C-terminal fusion, present in the pAUL1 vector, is not affecting the phenotype.

Deletion constructs

*BSOM4*Δ90, *BSOM4*Δ144, *BSOM4*Δ270, *BSOM4*ΔE1, *BSOM4*Δ417 and *BSOM4*:16-417 sequences were PCR amplified using P10/P11, P12/P11, P13/P11, P14/P11, P15/P11 and P9/P12 primer pairs, respectively. Using standard gateway cloning described above, these sequences were cloned individually into pAUL1-p*GLDT*$_{Ft}$ vector. Following were the resulted constructs: pAUL1-p*GLDT*$_{Ft}$::*BSOM4*Δ90, pAUL1-p*GLDT*$_{Ft}$::*BSOM4*Δ144, pAUL1-p*GLDT*$_{Ft}$::*BSOM4*Δ270, pAUL1-p*GLDT*$_{Ft}$::*BSOM4*ΔE1, pAUL1-p*GLDT*$_{Ft}$::*BSOM4*Δ417 and pAUL1-p*GLDT*$_{Ft}$::*BSOM4*:16-417.

Generation of pAUL1-p*GLDT*$_{Ft}$::*BSOM4 construct**

All in-frame ATGs (+1, +418, +466, +517 and +589) of *BSOM4* predicted coding sequence, were mutated by replacing Guanine residue (G) with Adenine (A) nucleotide. The modified sequence (*BSOM4**) was synthesized with attB1 and attB2 sites and cloned into a pUC57 vector (Biomatik), resulting in pUC57-*BSOM4** and by subsequent gateway reactions *BSOM4** was cloned into the corresponding pAUL1-p*GLDT*$_{Ft}$ vector to create pAUL1-p*GLDT*$_{Ft}$::*BSOM4**.

pAUL1-*GLDT*$_{Ft}$::AL-LOC9329632, pAUL1-*GLDT*$_{Ft}$::CS-LOC104743603 and pAUL1-*GLDT*$_{Ft}$::*BR*-LOC103828966) constructs

LOC9329632, LOC104743603, and LOC103828966 refer to gene IDs from *Arabidopsis lyrata*, *Camelina sativa* (cultivar DH55) and *Brassica rapa* (cultivar Chiifu-401-42), respectively. Predicted coding sequences from these genes were obtained from the National Center for Biotechnology Information (NCBI) and synthesized with attB sites and cloned into a pUC57 vector (Biomatik). Later, these sequences were individually introduced into pAUL1- p*GLDT*$_{Ft}$.

N-terminal and C-terminal translational fusion constructs

The pAUL3 gateway vector (Lyska *et al.*, 2013) was designed for C-terminal fusions with triple tag (3xHA-StrepIII-2xPA). *BSOM4* and *BSOM4*Δ15 sequences were individually fused to this tag by following standard gateway protocol described earlier in the text, resulting into: pAUL3-*35S*::*BSOM4* and pAUL3-*35S*::*BSOM4*Δ15 constructs, respectively. These constructs were transformed into the wild type Columbia-0. Total leaf proteins from T1 transgenics and from wild type plants were isolated using sodium dodecyl sulfate (SDS) buffer. This buffer contains – 10 mM Tris/HCl pH 7.8, 4 M urea, 5 % SDS (w/v), 15 % Glycerin, 10 mM β-mercaptoethanol, and protease inhibitor. The SDS-polyacrylamide gel electrophoresis and western blot was carried out as described by Schägger H and von Jagow G (1987) and Harlow E and Lane D (1999), respectively. The anti-mouse IgG fraction that was coupled to peroxidase (Sigma) was used against protein-A (PA) epitope to detect chimeric BSOM4 protein.

To create N-terminal and C-terminal YFP (yellow fluorescent protein) reporter fusion constructs, *BSOM4* and *BSOM4*Δ15 sequences were cloned separately into pUBN-Dest and pUBC-Dest gateway vectors (Grefen *et al.*, 2010). This generated p*UB10*-YFP::BSOM4, p*UB10*-BSOM4::YFP and p*UB10*-YFP::BSOM4Δ15, p*UB10*-BSOM4Δ15::YFP constructs, respectively.

Semi-Quantitative PCR (SQ-PCR)

Total RNA was extracted (RNeasy Plant Mini Kit; QIAGEN) from rosette leaves of the reference line and homozygous AT47 mutant line (four weeks old plants) that were grown in our growth chamber conditions. On-column DNase digestion was performed using the RNase-Free DNase Set (QIAGEN) and RNA quality was verified before proceeding to cDNA synthesis. cDNA was synthesized using 1 µg total RNA (QuantiTectReverse Transcription Kit; QIAGEN). The SQ-PCR was performed using gene-specific primers (P16/P17) and *ACTIN* primers (*ACTIN* 7 - F+R) that served as endogenous control.

5' RACE (rapid amplification of complementary DNA ends)

Above isolated 1µg total RNA was used as a starting material to synthesize RACE-ready cDNA, according to manufacturer's protocol provided with SMARTer™RACE cDNA Amplification Kit (Clontech). Nested PCR was carried out using gene-specific reverse primers (P18 and P17), Universal Primer A Mix (provided with Kit) and Phire Plant Direct Master Mix (Thermo Scientific). The obtained products were gel extracted and cloned into a pJET1.2 cloning vector (Thermo Scientific). Several clones were verified by performing a colony PCR of pJET1.2-F and pJET1.2-R and sequenced using the same primer pair (LGC Genomics).

Light microscopy and transmission electron microscopy

The mutant line *bsom4* and reference plants (GFP reporter line) were grown in growth chambers. To assess the internal leaf anatomy, second leaf pair from four weeks old grown plants was chosen. For this, the middle part of the leaf was cut into 1-2 mm^2 pieces and prepared for microscopic observation as described by Khoshravesh *et al*. (2017). Zeiss Axiophot light microscope equipped with Olympus CellSens imaging software and Phillips 201 transmission electron microscope equipped with an Advantage HR camera system (Advanced Microscopy Techniques) were used to capture the images. In order to stain callose at plasmodesmata connections, single cell isolates were prepared by following the protocol described in Osteryoung et al. (1998). 1-2 mm^2 leaf segments were first fixed with 1 % glutaraldehyde solution in 0.05 M sodium cacodylate buffer (pH 6.9), then the glutaraldehyde was subsequently replaced with 0.1 M Na_2EDTA (pH 9) and incubated at 55 °C water bath for two hours. Samples were then washed two times with water and then incubated with 0.01 % aniline blue fluorochrome solution (Biosupplies) in 0.01M K_3PO_4 (Zavaliev and Epel, 2015). Finally, leaf samples were placed on microscopic slides and individual cells were separated by gently applying mechanical pressure from the top of the coverslip. Bundle sheath cells were distinguished from mesophyll cells by their elongated shape.

(V) Acknowledgments

This work was funded by 'Cluster of Excellence on Plant Sciences' (CEPLAS). We thank Dr. Florian Döring for providing *A. thaliana GFP* reporter line.

(VI) References

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

Aukerman, M.J. and Sakai, H. (2003) Regulation of flowering time and floral organ identity by a MicroRNA and its *APETALA2*-like target genes. *Plant Cell*, **15**, 2730–2741.

Botha, C.E.J. (1992) Plasmodesmatal distribution, structure and frequency in relation to assimilation in C_3 and C_4 grasses in southern Africa. *Planta*, **187**, 348–358.

Bougourd, S., Marrison, J. and Haseloff, J. (2008) An aniline blue staining procedure for confocal microscopy and 3D imaging of normal and perturbed cellular phenotypes in mature *Arabidopsis* embryos. *Plant J.*, **24**, 543–550.

Bräutigam, A., Kajala, K., Wullenweber, J., et al. (2011) An mRNA blueprint for C_4 photosynthesis derived from comparative transcriptomics of closely related C_3 and C_4 species. *Plant Physiol.*, **155**, 142–156.

Bray, N., Dubchak, I. and Pachter, L. (2003) AVID: A global alignment program. *Genome Res.*, **13**, 97–102.

Brooks, A. and Farquhar, G.D. (1985) Effect of temperature on the CO_2/O_2 specificity of ribulose-1,5-bisphosphate carboxylase/oxygenase and the rate of respiration in the light.

Planta, **165**, 397–406.

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

Cegelski, L. and Schaefer, J. (2006) NMR determination of photorespiration in intact leaves using in vivo 13 CO_2 labeling. *J. Magn. Reson.*, **178**, 1–10.

Christin, P.-A., Osborne, C.P., Chatelet, D.S., et al. (2013) Anatomical enablers and the evolution of C_4 photosynthesis in grasses. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 1381 – 1386.

Covshoff, S., Majeran, W., Liu, P., Kolkman, J.M., Wijk, K.J. van and Brutnell, T.P. (2008) Deregulation of maize C_4 photosynthetic development in a mesophyll cell-defective mutant. *Plant Physiol.*, **146**, 1469–1481.

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

Curtis, M.D. and Grossniklaus, U. (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.*, **133**, 462–469.

Danila, F.R., Quick, W.P., White, R.G., Furbank, R.T. and Caemmerer, S. von (2016) The Metabolite Pathway between Bundle Sheath and Mesophyll: Quantification of plasmodesmata in leaves of C_3 and C_4 Monocots. *Plant Cell*, **28**, 1461–1471.

Danila, F.R., Quick, W.P., White, R.G., Kelly, S., Caemmerer, S. von and Furbank, R.T. (2018) Multiple mechanisms for enhanced plasmodesmata density in disparate subtypes of C_4 grasses. *J. Exp. Bot.*, **69**, 1135–1145.

Earp, D.J., Lowe, B. and Baker, B. (1990) Amplification of genomic sequences flanking transposable elements in host and heterologous plants: a tool for transposon tagging and genome characterization. *Nucleic Acids Res.*, **18**, 3271–3179.

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

Ehleringer, J.R. and Monson, R.K. (1993) Evolutionary and ecological aspects of photosynthetic pathway variation. *Annu. Rev. Ecol. Syst.*, **24**, 411–39.

Erdmann, V.A., Szymanski, M., Hochberg, A., Groot, N. and Barciszewski, J. (2000) Non-coding, mRNA-like RNAs database Y2K. *Nucleic Acids Res.*, **28**, 197–200.

Ernst, K. and Westhoff, P. (1997) The phospho*enol*pyruvate carboxylase (ppc) gene family of *Flaveria trinervia* (C_4) and *F. pringlei* (C_3): molecular characterization and expression analysis of the *ppcB* and *ppcC* genes. *Plant Mol. Biol.*, **34**, 427–443.

Furbank, R.T. (2011) Evolution of the C_4 photosynthetic mechanism: are there really three C_4 acid decarboxylation types? *J. Exp. Bot.*, **62**, 3103–3108.

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

Gowik, U. and Westhoff, P. (2011) The path from C_3 to C_4 photosynthesis. *Plant Physiol.*, **155**, 56–63.

Grefen, C., Donald, N., Hashimoto, K., Kudla, J., Schumacher, K. and Blatt, M.R. (2010) A ubiquitin-10 promoter-based vector set for fluorescent protein tagging facilitates temporal stability and native protein distribution in transient and stable expression studies. *Plant J.*, **64**, 355–365.

Haberlandt, G. (1904) *Physiologische Pflanzenanatomie*, Leipzig, Germany: Verlag von Wilhelm Engelmann.

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

Harlow, E. and Lane, D. (1999) Using Antibodies. A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

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

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

Hohmann, N., Wolf, E.M., Lysak, M.A. and Koch, M.A. (2015) A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell*, **27**, 2770–2784.

Jeong, D.-H., An, S., Park, S., et al. (2006) Generation of a flanking sequence-tag database for activation-tagging lines in japonica rice. *Plant J.*, **45**, 123–132.

Kawakatsu, T., Huang, S.-S.C., Jupe, F., et al. (2016) Epigenomic diversity in a global collection of *Arabidopsis thaliana* Accessions. *Cell*, **166**, 492–505.

Khoshravesh, R., Lundsgaard-Nielsen, V., Sultmanis, S. and Sage, T.L. (2017) Light microscopy, transmission electron microscopy, and immunohistochemistry protocols for studying photorespiration. In: Fernie A., Bauwe H., Weber A. (eds) Photorespiration. *Methods Mol. Biol.*, **1653**, 243-270. Humana Press, New York, NY

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

Klepikova, A. V., Kasianov, A.S., Gerasimov, E.S., Logacheva, M.D. and Penin, A.A. (2016) A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J.*, **88**, 1058–1070.

Kondou, Y., Higuchi, M. and Matsui, M. (2010) High-throughput characterization of plant gene functions by using gain-of-function technology. *Annu. Rev. Plant Biol.*, **61**, 373–393.

Li, R., Zhu, H. and Luo, Y. (2016) Understanding the functions of long non-coding RNAs through their higher-order structures. *Int. J. Mol. Sci.*, **17**, 702.

Li, W., Notani, D. and Rosenfeld, M.G. (2016) Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat. Rev. Genet.*, **17**, 207–223.

Logemann, E., Birkenbihl, R.P., Ülker, B. and Somssich, I.E. (2006) An improved method for preparing Agrobacterium cells that simplifies the *Arabidopsis* transformation protocol. *Plant Methods*, **2**, 1–5.

Lundgren, M.R., Osborne, C.P. and Christin, P.-A. (2014) Deconstructing Kranz anatomy to understand C_4 evolution. *J. Exp. Bot.*, **65**, 3357–3369.

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

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

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

Mersereau, M., Pazour, G.J. and Das, A. (1990) Efficient transformation of *Agrobacterium tumefaciens* by electroporation. *Gene*, **90**, 149–151.

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

Nakazawa, M., Ichikawa, T., Ishikawa, A., Kobayashi, H., Tsuhara, Y., Kawashima, M., Suzuki, K., Muto, S. and Matsui, M. (20013) Activation tagging, a novel tool to dissect the functions of a gene family. *Plant J.*, **34**, 741–750.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C. and Weigel, D. (2003) Control of leaf morphogenesis by microRNAs. *Nature*, **425**, 257-263

Quinn, J.J. and Chang, H.Y. (2016) Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.*, **17**, 47–62.

Radford, J.E., Vesk, M. and Overall, R.L. (1998) Callose deposition at plasmodesmata. *Protoplasma*, **201**, 30–37.

Raines, C.A. (2011) Increasing photosynthetic carbon assimilation in C_3 plants to improve crop yield: current and future strategies. *Plant Physiol.*, **155**, 36–42.

Rawat, V., Abdelsamad, A., Pietzenuk, B., Seymour, D.K., Koenig, D., Weigel, D., Pecinka, A. and Schneeberger, K. (2015) Improving the annotation of *Arabidopsis lyrata* using RNA-Seq Data. *PLoS One*, **10**, 1–12.

Sage, R.F. (2016) A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *J. Exp. Bot.*, **67**, 4039–4056.

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

Sage, R.F., Khoshravesh, R. and Sage, T.L. (2014) From proto-Kranz to C_4 Kranz: building the bridge to C_4 photosynthesis. *J. Exp. Bot.*, **65**, 3341–3356.

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

Sage, T.L., Busch, F.A., Johnson, D.C., et al. (2013) Initial events during the evolution of C_4 photosynthesis in C_3 species of *Flaveria*. *Plant Physiol.*, **163**, 1266–76.

Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*, **9**, 671–675.

Schägger, H. and von Jagow, G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem* **166**: 368-379.

Simpson, G.G., Laurie, R.E., Dijkwel, P.P., Quesada, V., Stockwell, P.A., Dean, C. and Macknight, R.C. (2010) Noncanonical translation initiation of the *Arabidopsis* flowering time and alternative polyadenylation regulator FCA. *Plant Cell*, **22**, 3764–3777.

Slewinski, T.L., Anderson, A.A., Price, S., Withee, J.R., Gallagher, K. and Turgeon, R. (2014) *SHORT-ROOT1* plays a role in the development of vascular tissue and kranz anatomy in maize leaves. *Mol. Plant*, **7**, 1388–1392.

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

Stata, M., Sage, T.L., Hoffmann, N., Covshoff, S., Ka-Shu Wong, G. and Sage, R.F. (2016) Mesophyll chloroplast investment in C_3 , C_4 and C_2 Species of the genus *Flaveria*. *Plant Cell Physiol.*, **57**, 904–918.

Stata, M., Sage, T.L., Rennie, T.D., Khoshravesh, R., Sultmanis, S., Khaikin, Y., Ludwig, M. and Sage, R.F. (2014) Mesophyll cells of C_4 plants have fewer chloroplasts than those of closely related C_3 plants. *Plant. Cell Environ.*, **37**, 2587–2600.

Stockhaus, J. (1997) The Promoter of the gene encoding the C_4 Form of phosphoe*nol*pyruvate carboxylase directs mesophyll-specific expression in transgenic C_4 *Flaveria* spp. *Plant Cell*, **9**, 479–489.

Stone, B.A., Evans, N.A., Bonig, I. and Clarke, A.E. (1984) The application of Sirofluor, a chemically defined fluorochrome from aniline blue for the histochemical detection of callose. *Protoplasma*, **122**, 191–195.

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

Walker, B.J., Vanloocke, A., Bernacchi, C.J. and Ort, D.R. (2016) The costs of photorespiration to food production now and in the future. *Annu. Rev. Plant Biol*, **67**, 107–29.

Wang, H.-L. V and Chekanova, J.A. (2017) Long noncoding RNAs in Plants. In: Rao M. (eds) Long Non Coding RNA Biology. *Adv. Exp. Med. Biol*. Springer, Singapore: **1008**, 133-154.

Wang, P., Khoshravesh, R., Karki, S., Furbank, R., Sage, T.L. and Langdale, J.A. (2017) Re-creation of a key step in the evolutionary switch from C_3 to C_4 leaf anatomy. *Curr. Biol.*, **27**, 3278–3287.

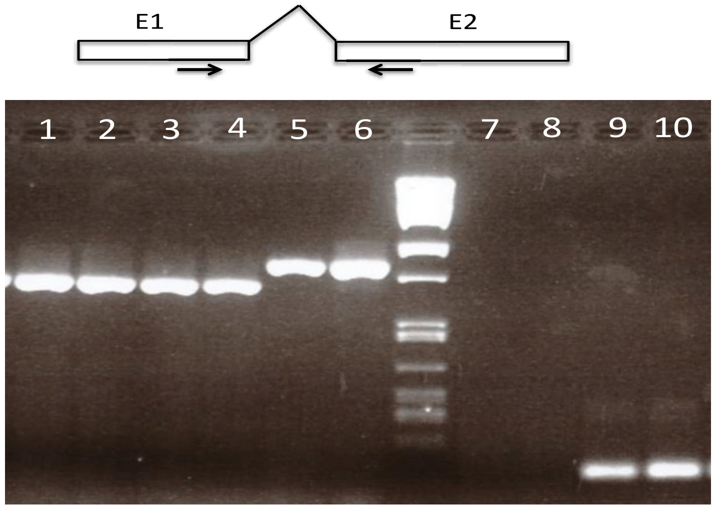
Weigel, D., Ahn, J.H., Blázquez, M.A., et al. (2000) Activation tagging in *Arabidopsis*. *Plant Physiol.*, **122**, 1003–1013.

Wen, J., Parker, B.J. and Weiller, G.F. (2007) In Silico identification and characterization of mRNA-like noncoding transcripts in *Medicago truncatula*. *In Silico Biol.*, **7**, 485–505.

Zavaliev, R. and Epel, B.L. (2015) Imaging callose at plasmodesmata using Aniline Blue: Quantitative confocal microscopy. In: Heinlein M. (eds) Plasmodesmata. *Methods in*

Mol. Biol., **1217**, 105-119. Humana Press, New York, NY.

(VII) Supplementary Data



Supplementary Figure 1: Semi-Quantitative PCR amplification of *BSOM4* gene using cDNA synthesized from total leaf RNA of both reference and *bsom4* mutant line. Actin 7 was used as endogenous control. Arrows marks indicate primer-binding sites of *BSOM4.* 1, 2 – actin 7 (reference line); 3, 4 – actin 7 (*bsom4*); 5, 6 – actin 7 amplification using genomic DNA as template. The *BSOM4* gene amplification from reference (7, 8) and mutant line (9, 10).

Supplementary Table 1: Oligonucleotides used in this study.

F – forward primer; R- reverse primer

Bold letters – restriction or gateway (*attB*) sites

Primer	Oligonucleotide sequence (5'→3')	Orientation
P1	GCTGAAATGGGTTGTTTTTG	F
P2	TCAAACCACAATCCGTTTAAG	R
P3	AGGGTTGGAGGGGAATTAAG	F
P4	CCCAATAGATACTGTAAACCCAACA	R
P5	**AAGCTT**ATGTTTGTTGGTAGTTTTTC	F
P6	**GGCGCGCC**TACTCACACCCTTGCTTAATACTT	R
P7	**GGGGACAAGTTTGTACAAAAAAGCAGGCTTA**ATGAAGATTGCTACGACATCAAGTGCTTTCGATTTG	F
P8	**GGGGACCACTTTGTACAAGAAAGCTGGGTA**ATGAAGATTGCTACGACATCAAGTGCTTTCGAT	R
P9	**GGGGACAAGTTTGTACAAAAAAGCAGGCTTA**ACATCAAGTGCTTTCGATTTGTT	F
P10	**GGGGACAAGTTTGTACAAAAAAGCAGGCTTA**GACCTCTCCTTCTTACCCGT	F
P11	**GGGGACCACTTTGTACAAGAAAGCTGGGTA**TCAATGTTCTTTGACATCTGTAGG	R
P12	**GGGGACAAGTTTGTACAAAAAAGCAGGCTTA**CTGGAAGAAGAAGTGGACTC	F
P13	**GGGGACAAGTTTGTACAAAAAAGCAGGCTTA**CTTCAAGAAACAAAGACGAATA	F
P14	**GGGGACAAGTTTGTACAAAAAAGCAGGCTTA**GCGAAATCTAGCAGGACGAGT	F
P15	**GGGGACAAGTTTGTACAAAAAAGCAGGCTTA**ATGGTAGCAACCGTTAGCTTTGA	F
P16	ACCTCTCCTTCTTACCCGTAGT	F
P17	GGTAAGACTCGTCCTGCTAGA	R
P18	GTAGCCACGCTGTTCCATTT	R
ACTIN 7	TTCAATGTCCCTGCCATGTA	F
ACTIN 7	TGAACAATCGATGGACCTGA	R

(VIII) Author contributions

KB wrote the manuscript and performed all experiments except

TLS helped with transmission electron microscopy analysis and provided images of aniline blue stained single cell isolates.

V. Acknowledgments

Many thanks to…

Prof. Dr. Peter Westhoff for giving the opportunity to pursue my Ph.D. study with him and for his continuous support, motivation, and encouragement. His guidance not only helped me in improving my scientific skills but also improved self-confidence.

Prof. Dr. Maria von Korff Schmising for kindly accepting to review my thesis.

My thesis committee advisors **Prof. Dr. Rüdiger Simon** and **Prof. Dr. Ute Höcker** for their time, useful comments and suggestions.

Dr. Udo Gowik for helping me in analyzing the transcriptome data.

Prof. Tammy Sage for allowing visiting her lab and helping me with the microscope work.

Thomas J. Wrobel and **Dr. Florian Döring** for their collaboration and help.

Jan Emmerling and **Dr. Stefanie Schulze** for good scientific conversations, valuable suggestions and help.

I am especially thankful to **Dr. Stefanie Schulze** and **Daniela Lichtblau** for their kind help when I arrived in Germany and would never forget that.

Parag Bhople and **Dr. Jathish Ponnu** for carefully reading my thesis and making helpful comments and suggestions.

My present and former **Botanik IV** colleagues for their help in many aspects and friendly conversations. Special thanks goes to **Dr. Karin Ernst** and **Dr. Sandra Kirschner** for helping me in many situations.

Entire **CEPLAS** team members for such a wonderful environment and organizing many events to place plant science research in a better position. I am personally thankful to **Brigitte**

Haumann, **Dr. Petra Fackendahl** and **Dr. Esther Jawurek** for helping me with all the formal documentation work during and after my arrival in Germany and making my life easy. **Susan Jefferies** for such a wonderful accommodation during my visit to **Prof. Tammy Sage** lab.

My teacher **Appala Naidu Kimidi** for his continuous encouragement, motivation, and for wise suggestions, right from my intermediate education to till this moment. There is no doubt without you I would not have reached this position.

Dr. Naveen Kumar Singh for encouraging me to explore the research outside of India and definitely it was a good decision.

Satish Kumar Eeda for his unconditional care, help and support since last six years.

Udhaya Ponraj and **Angelo Romeo Agossou** for their wonderful company in these five years.

Asu, **Githa**, **Appu** and **Srinu** for their love and care. I am lucky to have you as my best friends and sure that no one else can replace your position.

My parents **Janardhana Rao** and **Viyaya Lakshmi** for their unconditional love, support and for giving me the freedom to achieve my goals. From my deep heart, I feel lucky being your daughter.

Last but not least to all my family members for their love, care, and continuous support.