Transcriptomic and genetic studies of leaf development in the C_4 monocot *Sorghum bicolor* and the C_3 dicot *Arabidopsis thaliana*

Inaugural dissertation

for the attainment of the title of doctor

Faculty of Mathematics and Natural Sciences

at Heinrich Heine University Düsseldorf

Presented by

Zahida Bano

from Chitral, Pakistan

Düsseldorf, August 2022

From the institute of “Developmental and Molecular Biology of Plants”

at Heinrich Heine University Düsseldorf

Published by the permission of the

Faculty of Mathematics and Natural Sciences at

Heinrich Heine University Düsseldorf

Supervisor: **Prof. Dr. Peter Westhoff**

Co-supervisor: **Prof. Dr. Maria Von Korff Schmising**

Date of the oral examination: 28.11.2022

Declaration

Herewith, I declare that I have written this thesis by myself and independently following the principle of good scientific practice at Heinrich Heine University Düsseldorf. I have used no other sources except due references are made. I assure that the work contained in the thesis is not previously submitted to any platform for examination.

Düsseldorf, 09.08.2022 ----------

Zahida Bano

Table of Contents

Abbreviation

A	Adanine
ABA	Abscisic acid
A. thaliana	*Arabidopsis thaliana*
Ala	Alanine
AlaAT	Alanine aminotransferase
AS1	Asymmetric leaves
Asp	Aspartate
ATP	Adenosine triphosphate
bHlH	Basic helix-loop-helix
BR	Brassinosteriod
BS	Bundle sheath
C	Cytosine
CA	Carbon anhydrase
Cas9	CRISPR associated endonuclease 9
CDKs	Cyclin dependent kinases
CK	Cytokinin
CO_2	Carbon dioxide
CRISPR	Cluster regularly interspaced short palindromic repeats
CYCs	Cyclins
Cys	Cysteine
Cyt b6f	Cytochrome b6/f
CZ	Central zone
DEGs	Differentially expressed genes
DP	Dimerization protein
E2F	Elongation factor 2
EMS	Ethyl methanesulfonate
FD	Ferredoxin
FUSE	Far up stream element
FUBP	Far up stream element binding protein
G	Guanine
GA	Gibberellin/gibberelic acid

GCs	Guard cells
GDC	Glycine decarboxylase
GFP	Green fluorescent protein
GGT	Glutamate: glyoxylate aminotransferase
GIF1	GRF-interacting factor1
GL1/2/3	Glabra 1, 2 and 3
Gly	Glycine
GLYK	Glycerate kinase
GLS	Glutamate synthase
Glu	Glutamate
GMCs	Guard mother cells
GRFs	Growth regulating factors
GS	Glutamine synthase
H	Hours
HCO_3^-	Bicarbonate (Hydrogen carbonate)
JA	Jasmonic acid
Kb	Kilobase
KH	K-homology
KNOX1	Knotted like homebox1
Leu	Leucine
LUC	Luciferase
M	Mesophyll
M1, M2	1^{st}, 2^{nd} mutant plant generation
MAS	M phase-specific activator
MAPKs	Mitogen activated protein kinases
Mb	Megabase
MDH	Malate dehydrogenase
MMCs	Meristemoid mother cells
NADPH	Nicotinamide adenine dinucleotide phosphate
NADP-ME	NADP-dependent malic enzyme
NH_3	Ammonia
OAA	Oxaloacetate
PC	Plastocyanin

PCA	Principal component analysis
PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvate
PEPC	Phosphoenolpyruvate carboxylase
PEPCK	Phosphoenolpyruvate carboxykinase
PIFs	Phytochrome interacting factors
3-PGA	3-Phosphoglycerate
2-PG	2-Phosphoglycolate
PGLP	Phosphoglycolate phosphatase
PLBs	Polymers body
PPDK	Pyruvate phosphate dikinase
PQ	Plastoquinone
Pro	Proline
PSI	Photosystem I
PSII	Photosystem II
Pyr	Pyruvate
PZ	Periferal zone
ROS	Reactive oxygen species
RuBP	Ribulose-1, 5-bisphosphate
Rubisco	Ribulose-1, 5-bisphosphate carboxylase/oxygenase
SAM	Shoot apical meristem
SAUR	Small auxin up-RNA
S. bicolor	*Sorghum bicolor*
Ser	Serine
SGT	Serine: glyoxylate aminotransferase
sgRNA	Small guide RNA
SHMT	Serine hydroxymethyltransferase
SIM/SMR	Siamese/Siamese related
SLGC	Stomatal lineage ground cell
SNP	Single nucleotide polymorphism
SPCH	Speechless
SUVH1	Su(var) 3-9 homolog
tasiRNA	Trans-acting small interfering RNA

TCL1/2/3	Trichomeless 1, 2, and 3
T1, T2, T3, T4	1^{st}, 2^{nd}, 3^{rd}, 4^{th}, transgenic plant generation
T	Thymine
TF	Transcription factors
TP	Transit peptide
Tyr	Tyrosine

I. Introduction

Leaves are essential organs of vascular plants that are specialized for capturing solar energy and converting it into chemical energy and metabolic compound through a process known as photosynthesis. In another way, we can say it is the primary site of a plant that manufactures food and oxygen, which in turn nourish and sustain all land animals. Without leaves, plants would not be able to perceive environmental conditions, such as light quality and quantity. In most angiosperms, the floral organs are modified from leaves. Besides photosynthesis, leaves are also involved in plant respiration and photoperception (Tsukaya, 2013). Plant leaves possess a huge diversity in their structure, and this variation is an important factor to track the evolution of their biological shapes. Most variations are observed in the leaf margin, such as the serrated or lobed margin (Runions *et al.,* 2017). In general, leaves are divided into two categories simple leaves and compound leaves. Simple leaves are composed of a single continuous lamina, while in compound leaves, the lamina is divided into multiple leaflets, each resembling a simple leaf (Efroni *et al.,* 2010).

1. Leaf growth and morphogenesis

The development of plant leaves follows a basic mechanism that is flexible in relation to the plant species, environmental circumstances, and developmental stages (Bar and Ori, 2014). It is a dynamic process where independent regulatory pathways instruct component cells to make differentiation switches at various stages of development in order to execute developmental processes (Kalve *et al.,* 2014). In general, leaves originate from shoot apical meristem (SAM), which harbours stem cell niches at its center, from which cells continuously recruit its growing organ during post-embryonic development. SAM is composed of two different functional domains named as central zone (CZ) and peripheral zone (PZ) that deliver founder cells (Du *et al.,* 2018, Nikolov *et al.,* 2019). These founder cells differentiate into leaf primordia that grow out to a lateral organ, such as leaves, stem, and floral structure. At the beginning of leaf initiation, all cells divide, but with the passage of time, cells at the tip cease to divide and start expanding while cells at the base continue dividing (Nelisson *et al.,* 2016). In SAM, such alteration in cell division activity leads to leaf

initiation. This pattern of cell division was observed in both monocots and dicots that reflect very similar temporal regulation of leaf growth.

In eudicots, leaves growth and differentiation are divided into four stages based on the developmental processes. In the initial stages, the founder cells of the leaf primordial are assembled from the peripheral zone of SAM to the site of primordial initiation and bulge out. Second, initiation of leaf primordia from SAM is followed by distal growth and establishment of adaxial-abaxial and proximal-distal axes. The third is the development of the leaf lamina. Leaf-blade and petiole are distinguished in this stage, and the leaf lamina is further modified by a coordinated process of cell division and expansion (Du *et al.,* 2018, Nakata and Okada, 2013, Andriankaja *et al.,* 2012, Sinha, 1999, Xiong and Jiao, 2019).

1.1. Leaf development in Arabidopsis

Arabidopsis leaves are simple leaves which are connected to the stem through petioles. *Arabidopsis* leaves comprise the upper (adaxial) layer of epidermis bearing trichomes and lower (abaxial) epidermal cells bearing stomata. Between upper and lower epidermises lays photosynthetically active mesophyll (M), i.e., one or more layers of palisade mesophyll followed by multiple layers of spongy mesophyll with air space in between to facilitate gas exchange. Embedded in the M, there is a vasculature (xylem and phloem) surrounded by a layer of bundle sheath (BS) cells. The vasculature facilitates the transport of water and photosynthates from one part of a plant to the other (Figure 1, Wuyts *et al.,* 2010).

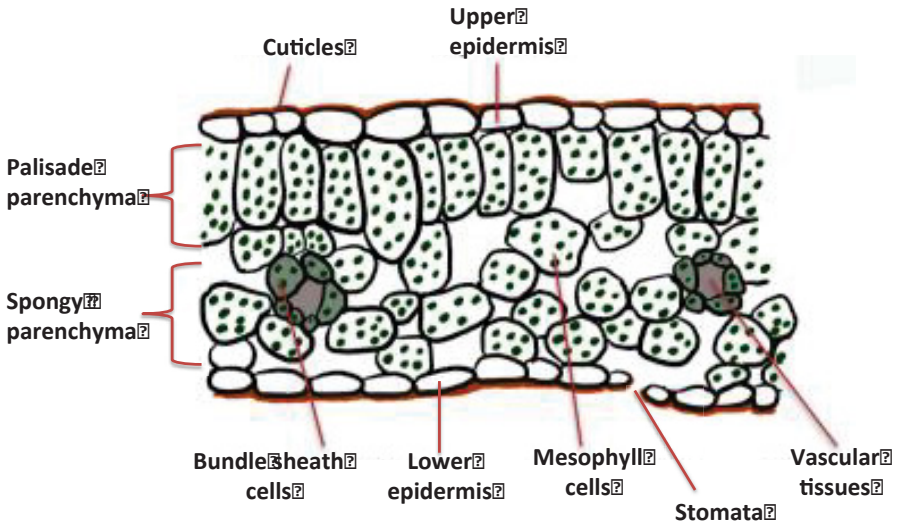


Figure 1. Schematic diagram of a cross-section of an *Arabidopsis* leaf.

Display represents upper and lower epidermis, mesophyll and bundle sheath cells, palisade and spongy parenchyma, stomata, vascular tissue, and cuticles. Image adapted from Kirschner, 2017 Ph.D. thesis.

To understand the process of leaf development in *Arabidopsis thaliana,* many mutants have been isolated, and various aspects of leaf development have been addressed so far, including leaf initiation, leaf differentiation, cell cycle regulation, trichome formation, stomata, and vascular development (Talbert *et al.,* 1995, Kinsman and Pyke, 1998, Cheng *et al.,* 2013, Eloy *et al.,* 2011, Wang *et al.,* 2018, Tsuge *et al.,* 1996, Wang *et al.,* 2007, Tsuji and Coe, 2013, Walker *et al.,* 2000, Yang and Sack, 1995). However, the regulatory network that controls growth and final organ size is still poorly understood (Andirankaja *et al.,* 2011). Leaf morphogenesis is a process regulated by many genes and pathways that lead to the generation of an organ of various shapes and sizes; therefore, differences can be seen between various species as well as within a species (Rodrigues *et al.,* 2014).

1.2 Development of leaf primordia

Leaf development is also regulated by the concentrations of hormones and their interplay. Such as, high auxin concentration plays a pivotal role in the regulation of leaf initiation. During leaf initiation, the influx carriers AUXIN RESISTANT (AUX1), LIKE AUXIN RESISTANT (LAX) and an efflux transporter PIN-FORMED1 (PIN1) establish a high concentration of auxin in the region of leaf primordial founder cells (Du *et al.,* 2018, Rodregues *et al.,* 2013, Klave *et al.,* 2014, Yang *et al.,* 2006, Geisler *et al.,* 2003, Rutschow *et al.,* 2014, Xiong and Jioa, 2019). A high level of auxin transport at the flank of the SAM inhibits cytokinin formation by repression of KNOTTED LIKE HOMEBOX1 (KNOX1) and triggers primordium development. KNOX1 positively regulates cytokinin biosynthesis and prevents early cell differentiation and gibberellin (GA) synthesis. Unlike KNOX1, which prevents cell differentiation in SAM, ASYMMETRIC LEAF1/ROUGH SHEATH2/PHANTHASTICA (ARP) expressed in the developing primordium initiates differentiation of leaf primordium (Klave *et al.,* 2014, Bryne *et al.,* 2002, Xiong and Jiao, 2019).

Besides auxin, other transcription factor networks distinguish leaf primordia from the rest of the meristem. The MYB transcription factor ASYMMETRIC LEAVES (AS1), together with AS2, is involved in the repression of the SHOOTMERISTEMLESS (STM) and other KNOX1 genes in leaf primordia (Rodrigues *et al.,* 2014, Machida *et al.,* 2015, Matsumura *et al.,* 2016, Ikezaki *et al.,* 2009, Hasson *et al.,* 2010). The adaxial-abaxial polarity of growing leaf primordium is maintained and strengthened via domain-specific expression and mutual repression of adaxial-abaxial promoting genes. Such as, adaxial cell fate is promoted by REVOLUTA (REV) and HD-ZIP III transcription factors PHAVOLUTA (PHV), PHABULOSA (PHB), AS1/2, LOB domain transcription factor (Du and Wang, 2015, Hasson *et al.,* 2010) and trans-acting small interfering RNA (tasiRNA). At the same time the abaxial domain is promoted by KANADI1 (KAN1) and YABBY (YAB) (DU *et al.,* 2018, Eshed *et al.,* 2004, McConnell *et al.,* 2001, Iwakawa *et al.,* 2002, Lin *et al.,* 2003, Ichihashi and Tsukaya, 2015, Kerstetter *et al.,* 2001).

1.3 Coordination of cell division and leaf proliferation during the growth of leaf blade

Unlike the morphology of leaf primordium, the final size and shape of leaf lamina differ enormously among species due to the various degrees of cell division and expansion. During primary morphogenesis, cell division occurs throughout the primordia and specific structures, such as trichomes, stomata, and vasculature, begin to form. In parallel, dividing cells also grow; this joint activity is known as cell proliferation. Cell division and growth are tightly regulated to maintain the average size of the proliferating cells fairly constant (Skalák *et al.,* 2019, Vercruysse *et al.,* 2020).

The process of cell division is called the cell cycle and is subdivided into four phases, G1, S, G2, and M phase. During the S phase, the nuclear DNA is duplicated; the M phase, also known as the mitotic phase, where the chromosomes are separated and distributed into daughter cells, and the G1 and G2 phase or gap phase to prepare the cells for DNA replication or mitosis respectively (Qi and Zhang, 2020, Noir *et al.,* 2015). To ensure accurate transmission of genetic information through these phases, they are controlled by different groups of core cell cycle proteins, including CYCLINs (CYCs), Cyclin-dependent kinases (CDKs), E2F/DIMERIZATION

PROTEINS (E2F/DP), SIAMESE/SIAMESE RELATED (SIM/SMR) and KIP RELATED PROTEIN/ INTERACTOR OF CDKS (KRP/ICKs.) The core machinery-controlled cell cycle is cyclin-depended kinases and cyclin complexes. *Arabidopsis* contains 12 CDKs named from A to E, together with many cyclin genes (Gutierrez, 2009, Kalve *et al.,* 2014). The involvement of all CDKs except C and E and around 32 cyclins (A-10 cyclin, B- 11 cyclin, D-10 cyclin, and H-1 cyclin) appear to have a role in cell cycle activities. However, CDKA is a key player in cell cycle progression.

The composition and activity of the CDK/CYC complex are highly phase regulated, such as CDKA and CYCD involved in the G1 progression and G1 to S transition, and CYCBs are mainly regulated in G2 to M transition. The CDKA/CYCD complex is activated by CDKF and CDKD coupling with CYCH through a phosphorylation cascade. The active CDKA/CYCD complex dissociates retinoblastoma-related protein (RBR) from E2F/DP complex because RBR regulates the activity of E2F-TF to control cell proliferation. Dissociation of RBR from the E2F/DP complex leads to the initiation of G1/S transition by activating the transcription of genes involved in cell cycle progression, transcription, chromatin dynamic, and DNA replication. Previous studies revealed that overexpression of *E2FA* and *E2FB* transcription factors lead to an enlarged phenotype in *Arabidopsis* due to enhance leaf proliferation. After the S phase, the cell enters the G2 phase to prepare for division through mitosis (Figure 2). CDKA/CDKB and CYCA/B/D are involved in this process. Like the G1 phase, the CDK/CYC complex is activated in the G2 phase, which promotes MYB repeat (MYB3R) transcription factors to bind M phase-specific activator (MAS) elements. That, in turn, activates the expression of M phase-specific genes, such as *KNOLLE*, *CYCA, CYCB, CDC20* and *NACK1* (Kalve *et al.,* 2014a, Maugarny-Cales and Laufs, 2018, Verkruesse *et al.,* 2020).

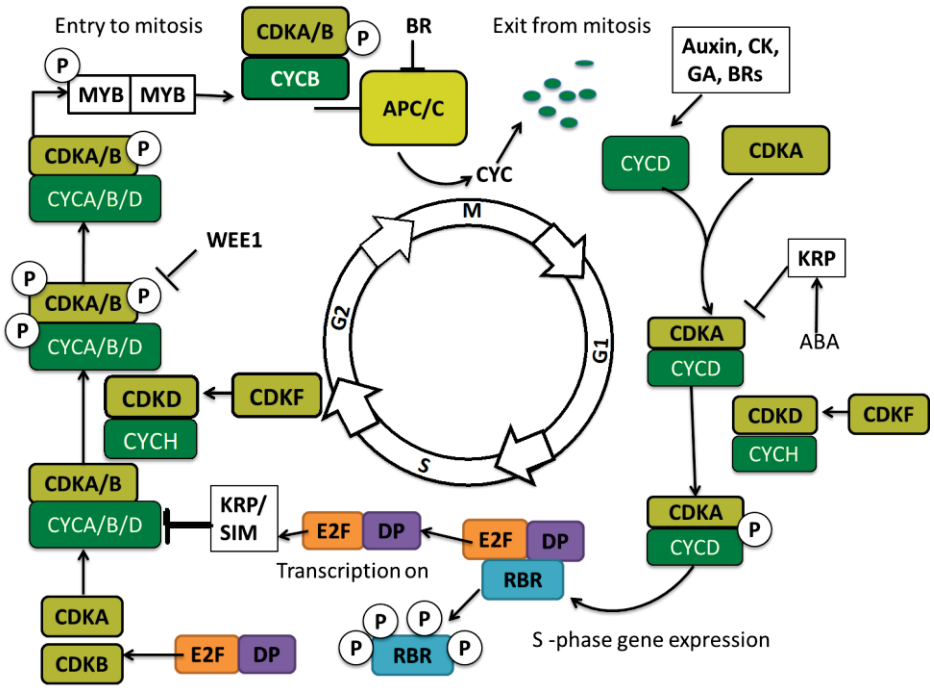


Figure 2. Generalized model indicating the molecular mechanism of cell cycle regulation. The regulation of four phases of the cell cycle is achieved via successive activation and deactivation of cyclin-dependent kinases (CDKs). During the cell cycle, the CDKs are incorporated with cyclin (CYC) and activated by CDK activators (CDKD and CDKF). In contrast, KRP act as a negative regulator of CDK/CYC complexes. CDKA/CYCD complex regulates G1 to S transition by phosphorylating RBR and releasing the E2F transcription factor. E2F actives transcription of S phase-related genes. CDKA/B and CYCA/B/D complexes regulate G2 to M transition. The CDK complex is inactivated by WEE1 through phosphorylation. The exit from the M phase occurs via the degradation of CYCs mediated by anaphase-promoting complex/cyclosome (APC/C). Various phytochromes are also involved in the regulation of the cell cycle, such as Auxin, cytokinin (CK), gibberellin (GA), and brassinosteroid (BR). Image modified from Maugarny-Cales and Laufs, 2018 and Kalve *et al.,* 2014a.

The activity of CDK/CYCs complexes is strictly regulated by multiple mechanisms, such as phosphorylation by a WEE1 protein kinase and interaction with cell cycle inhibitor proteins KRP/ICK, SIM/SMR family proteins, and APC/C (Noir *et al.,* 2015). WEE1 protein kinase inactivates CDKA/CYCD complex activity through phosphorylation. KRP inhibits CDK/CYC complex formation because down-regulation of KRP leads to increased cell proliferation and enlarged leaf area. SIM/SMR inhibits CDKA/CYCD and CDKB/CYCB complexes and promotes endoreduplication. Similarly, APC/C is involved in the degradation of mitotic

CYCB;1 through ubiquitination and regulates faster exit from the M phase (Kalve *et al.,* 2014a).

Various other genes have been reported as positive regulators of cell proliferation, such as *ANGUSTIFOLIA3 (AN3), STRUWWELPETER (SWP),* and *KLUH (KLU).* Conversely, *PEAPODS* and *BIG BROTHER* genes negatively regulate cell proliferation and organ growth (Lee *et al.,* 2009). AN3, also known as GRF-INTERACTING FACTOR1 (GIF1), interacts with GROWTH REGULATING FACTOR5 (GRF5) and promotes cell proliferation (Ichihachi and Tsukaya 2015, Horiguchi *et al.,* 2005). It binds with the chromatin binding complex SWITCH/SUCROSE NONFERMENTING (SWI/SNF) to regulate transcription during leaf development (Verkruyssen *et al.,* 2014). In addition to the role of various genes and transcription factors in cell division and proliferation, various hormones and growth factors also regulate cell cycle progression because they positively regulate the expression/ activity of CDKs. GA is involved in the repression of cell cycle inhibitor KRP2, SIAMESE, and BRs control exit from mitosis (Du *et al.,* 2018). Similarly, cytokinin is not only involved in the maintenance of SAM but also acts as a regulatory signal of leaf proliferation together with auxin. Cytokinin, together with auxin, induces the expression of CDKA that regulates the transition of the G1/S or G2/M phase of cell division (Mogarney-Cales and Laufs, 2018, Kalve *et al.,* 2014a).

1.4. Transition from cell division to expansion

During secondary morphogenesis, cell division and proliferation are replaced by cell expansion. The processes of cell proliferation and expansion are strictly and timely regulated for proper organogenesis and to acquire final leaf size. The transition of cell proliferation to expansion is a complicated process that occurs gradually, with various factors interacting to create a network of growth control at both transcription and posttranscriptional levels. Several regulators are involved during the transition of cell proliferation to expansion, such as auxin. Downstream auxin signaling gene *AUXIN RELATED GENE INVOLVED IN ORGAN SIZE (ARGOS)* enhances the expression of APETALA like the TF family gene *AINTEGUMENTA (ANT)*. ANT is involved in regulating organ size by controlling cell numbers in proliferating tissue through the activation of cell cycle driver CYCD3;1 (Kalve *et al.,* 2014a). Genetic studies have shown that separate regulatory pathways are involved in both longitudinal and lateral

leaf expansion. ROTUNDIFOLIA (ROT3 or ROT4) regulates longitudinal expansion, and ANGUSTIFOLIA (AN*)* regulates lateral expansion (Kalve *et al.,* 2014b).

Cell expansion requires the loosening of a cell wall, which is mediated by the activity of cell wall loosening enzymes, including EXPANSINs (EXPs), XYLOGLUCAN ENDOTRANSGLUCOSEYLASE/HYDROLASEs (XTHs), XYLOGLUCAN ENDOHYDROLASE (XEH), XYLOGLUCAN ENDOTRANSGLUCOSYLASE (XET), PECTIN METHYLESTERASEs (PMEs) (Majda and Robert, 2018, Kalve *et al.,* 2014a, Verkruyssen *et al.,* 2014) and reactive oxygen species (ROS) (Schmidt *et al.,* 2016). Auxin stimulates the activity of plasma membrane H^+-ATPase proton pumps through auxin-inducible SMALL AUXIN UP RNA (SAUR) protein, which pumps out proton to the wall matrix resulting in acidification of the apoplast and the loosening of the cell wall (Cosgrove 2005, Kalve *et al.,* 2014, Verkruyssen *et al.,* 2014, Majda and Robert, 2018). SAUR36 acts as a negative regulator of cell expansion, and SAUR53 positively regulates cell elongation (Kathare *et al.,* 2018). Besides SAUR family genes and EXPs, numbers of other proteins that have also been described to involve in leaf expansion are ARABIDOPSIS THALIANA HOMEOBOX 12 (ATHB12), EOD3/CYP78A6, and KUODA 1 (KUA1) (Lu *et al.,* 2014, Hur *et al.,* 2015).

1.5 Cell differentiation in Arabidopsis leaves

1.5.1. Differentiation of vascular tissue

Vascular tissues are composed of xylem and phloem that transport nutrients and signals internally and play an essential role in plant development and survival. In *Arabidopsis* leaves, the vascular system is composed of a continuous network of interconnected veins that develop a hierarchical reticulate venation pattern starting from the central primary vein, followed by the successive addition of secondary veins and, finally, higher veins (Wenzel *et al.,* 2007, Biedroń and Banasiak, 2018). Vein formation starts in the early stage of leaf initiation and proceeds simultaneously with expansion. Generally, vascular cells differentiate from procambial cells under the control of increased auxin flow. Auxin plays a vital role during vein formation that is distributed by carrier protein PIN1 (Lee *et al.,* 2014, Scarpella *et al.,* 2006, Du *et al.,* 2018). Besides PIN1, several other proteins are considered to involve in vascular

differentiation, such as ARABIDOPSIS *THALIANA* HOMEBOX8 (ATHB8), DNA BINDING WITH ONE ZINC FINGER (DOF) and MONOPTEROS (MP) (Biedroń and Banasiak, 2018). The *ATHB8*, a member of the *HD-ZIP III* gene family is induced by auxin and involved in early vascular development, such as the development of pre-procambium and procambium. Together with other transcription factors of this family, such as REV, PHB, PHV, and CORONA (CNA), ATHB8 is involved in vascular pattern formation, tissue specification, cell division, and differentiation (Biedroń and Banasiak, 2018, Kalve *et al.,* 2014a).

1.5.2. Trichome formation

During leaf development, specific epidermal cells convert into specialized cells known as leaf hairs or trichomes. Trichomes are an excellent model system for studying cell fate determination, cell polarity and cell expansion, and cell cycle regulation because the differentiation of trichomes follows the same sequence of the event as observed in other cells that form complex organs and tissues (Kalve *et al.,* 2014a, Gutierrez, 2009, Marks *et al.,* 1989, Pattanaik *et al.,* 2014). *Arabidopsis* contains large unicellular trichomes on the areal parts of the plant, such as leaves, stems, branches, and floral organs, with increased DNA content compared to the normal diploid cell (Hülskamp, 2004). Various genes are involved in trichome development to determine its morphology, position and spacing. The initiator complex is composed of the R3R2 MYB transcription factor, GLABRA1 (GL1), bHLH transcription factor GLABRA3 (GL3), and the WD40 repeat factor TRANSPARENT TESTA GLABRA (TTG1), while CAPRICE (CPC), TRIPTYCHON (TRY), ENHANCER OF TRY AND CPCs (ETC1, 2 and 3) and TRICHOMELESS (TCL1, 2 and 3) act as an inhibitor of trichome initiation. The initiator complex (GL1, GL3, and TTG1) regulates GLABRA2 (GL2) to control trichome development. Meanwhile, the inhibitor complex moves to the neighbouring cell and assembles there with TTG1 and GL3, resulting in the dissociation of GL1 and thereby leading to the inhibition of GL2 expression (Doroshkov *et al.,* 2019, Pattanaik *et al.,* 2014, Kalve *et al.,* 2014a, Gutierrez, 2009, Cox and Smith, 2019).

1.5.3. Formation of guard cells

Guard cells are specialized cells surrounding stomata and regulate gas and water

exchange by controlling the stomatal aperture. The formation of guard cells is initiated by protodermal cells that undergo a cellular transition to become meristemoid mother cells (MMCs). The MMCs undergo division asymmetrically and produce large stomatal lineage ground cells (SLGC) and small meristemoids. The SLGC gives rise to a new meristemoid by subsequent asymmetric division. The meristemoids eventually differentiate in guard mother cells (GMCs), which undergo a single symmetric division and cell fate transition to produce a pair of terminally differentiated guard cells (GCs). Various genes have been identified as a regulator of the transition of GMC to guard cells, including bHLH proteins SPEECHLESS (SPCH), MUTE, and FAMA. SPCH regulates the initial symmetric division, while MUTE transforms meristemoids to GMC, and FAMA is involved in the conversion of GMC to GCs. Two additional bHLH proteins, INDUCER OF CFB EXPRESSION1/SCREAM1 (ICE1/SCRM1) and SCREAM2 (SCRM2), also regulate GMCs to GC transition and cell proliferation. Besides that, CDKB1;1, CDKB1;2, and members of MITOGEN ACTIVATED PROTEIN KINASE (MAPK), YODA, MAPK kinase 7, and MAPK kinase 9 are also expressed in stomatal cell lineage and regulate the transition of GMCs to guard cells (Pillitter and Torii, 2012, Hachez *et al.,* 2011, Lau and Bergmann, 2012, Kalve *et al.,* 2014, Kanaoka *et al.,* 2008).

2. An overview of photosynthesis and photorespiration

In higher plants, photosynthesis occurs in two steps: the light-dependent reactions and the Calvin-Benson cycle. The light-dependent reactions occur in the thylakoid membrane of the chloroplast, which is the main site of photosynthesis, containing the reaction centers of the photosystems surrounded by light-harvesting complexes (LHC). During light reactions, water molecules split into protons, electrons, and oxygen on the luminal side. The manganese cluster present at the center of PSII extracts the electrons, which then reduces a tyrosine found at the D1 subunit of PSII. Here the electron is transferred to the core chlorophyll of PSII. Upon excitation, the linear electron transport takes place from photosystem II (PSII) to plastoquinone (PQ), cytochrome b6/f (Cyt b6f), plastocyanin (PC), and finally to photosystem I (PSI), where it is accepted by the terminal acceptor ferredoxin (FD) and used to reduce NADP+ to NADPH. The electron transfer reaction across the membrane is coupled to proton transport from the chloroplast stroma to the thylakoid lumen

generating changes in pH gradient, which drives ATP synthase for ATP production (Nelson and Yucum, 2006). The ATP and NADPH produced during the light reaction are used during CO_2 assimilation in the Calvin-Benson cycle (Foyer *et al.,* 2012, Rochaix, 2011).

The Calvin-Benson cycle starts with the carboxylation of ribulose-1, 5-bisphosphate (RuBP) by the enzyme Rubisco and produces 3-Phosphoglycerate (PGA) (Raines, 2003, Hugler and Sievert, 2011, Reumann and Weber, 2006). In the reduction phase, the resulting PGA is immediately phosphorylated by ATP, followed by reduction into glyceraldehyde 3-phosphate. Later, this triose phosphate is converted into either sucrose or starch. In the regeneration phase, triose phosphate is used to regenerate RuBP again under the consumption of ATP (Figure 3, Raines, 2003).

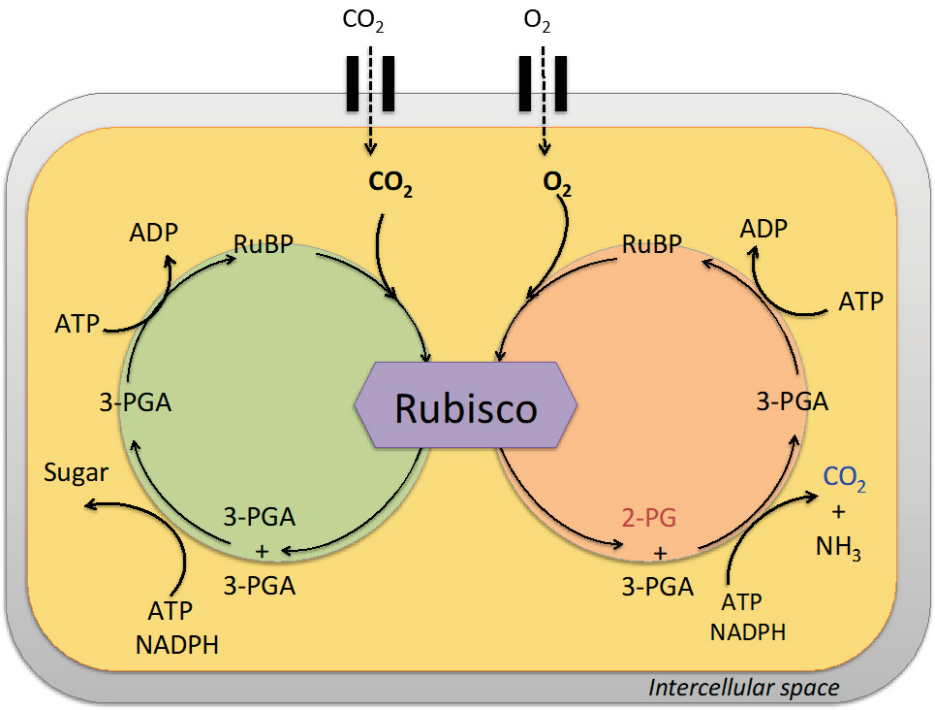


Figure 3. Schematic representation of Rubisco dual activity.

During the Calvin cycle, Rubisco reacts with CO_2 in the presence of RuBP and generates two molecules of 3-PGA, which in turn is used in sugar synthesis as well as in RuBP regeneration. While in photorespiration, it reacts with O_2 and produces one molecule of 2-PG and 3-PGA. 2-PG go through a series of energy-requiring reactions and finally releases CO_2 and NH_3.

Rubisco is considered to be the most abundant protein on earth (Bar-On *et al.,* 2019). It has a pivotal role in carbon fixation in photoautotrophic plants. Rubisco is a dual-

functional enzyme. Besides carboxylation reaction, it also reacts with oxygen, especially under lower CO_2 conditions and produces one molecule of 3-PGA and one molecule of 2-phosphoglycolate (2-PG) (Hatch 1987). This side reaction is very unfavorable for plants because 2-PG is a toxic compound (Bowes *et al.,* 1971); however, it is immediately converted into 3-PGA by a process called photorespiration, as shown in figure 3 (Fernie *et al.,* 2013, Peterhansel *et al.,* 2010, Hodges *et al.,* 2016).

2.1 Photorespiration is an energy-demanding but essential process in plants

Photorespiration is an energy-demanding process and can lead to the loss of up to 25% of assimilated CO_2, whereas the rate is increased in hot and dry conditions (Andersson, 2008). Photorespiration requires enzymatic reactions in three organelles chloroplast, peroxisome, and mitochondria (Peterhansel *et al.,* 2010). During photorespiration, 2-PG is hydrolyzed into glycolate in the chloroplast, which is then exported to the cytosol by glycolate/glyoxylate transporter and subsequently diffused into peroxisome (Pick *et al.,* 2013). In peroxisome, the glycolate is oxidized into glyoxylate by glycolate oxidase (GOX), which is transaminated to 2-glycin by glutamate: glyoxylate aminotransferase (GGT) and serine: glyoxylate aminotransferase (SGT). Glycin is transported from the peroxisome to mitochondria, where it is converted to serine by glycine decarboxylase (GDC) and serine hydroxymethyltransferase (SHMT), and the remaining nitrogen and carbon are released as ammonia and CO_2. Serine molecule is transported back to peroxisome and converted into hydroxypyruvate. Hydroxypyruvate is reduced to glycerate and subsequently phosphorylated into 3-PGA in the chloroplast that re-enters in the Calvin Benson cycle. The amine group generated in the mitochondria is reused to make new glycin from glyoxylate (Figure 4, Peterhansel and Maurino, 2011, Hu *et al.,* 2012, Reumann and Weber, 2006). Though the toxic 2-PG is successfully eliminated during photorespiration, but it costs a lot of energy because ATP and NADPH are needed to operate photorespiration. It is assumed that photorespiration cost approximately 50% extra energy for photosynthesis (Peterhansel *et al.,* 2010) and in C_3 species, this process is responsible for the loss of up to 30% of assimilated CO_2. As the rate of oxygenation of Rubisco increases, in consequence, the efficiency of photosynthesis will be decreased. However, plants developed different ways to deal with photorespiration, such as the development of C_4 photosynthesis (Gowik and

Westhoff, 2011). The C_4 pathway is a mechanism to overcome the limitation of photorespiration and suppresses oxygenation reaction by elevating CO_2 concentration around the Rubisco and acting as a biochemical CO_2 pump (Furbank and Taylor, 1995, Sage, 2004, Kajala *et al.,* 2011).

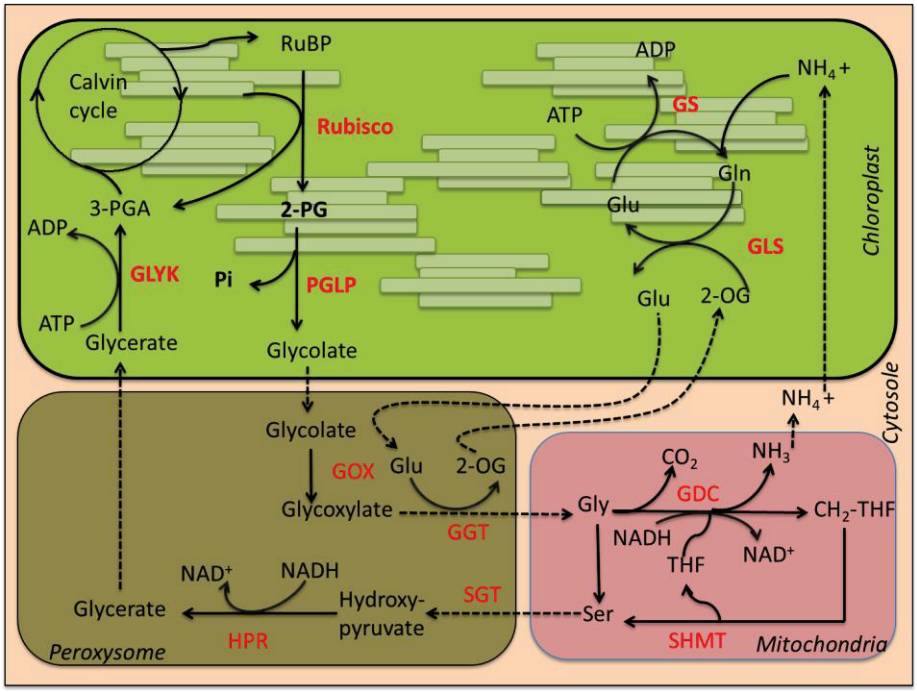


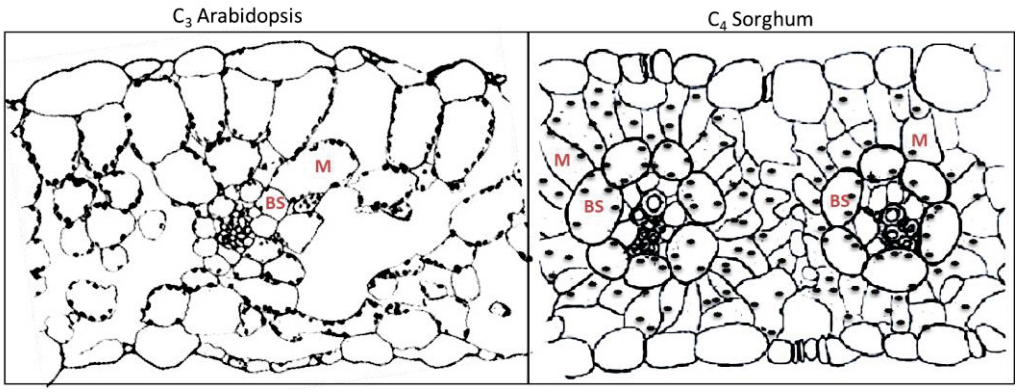
Figure 4. Photorespiratory cycle.

Distribution of photorespiratory reactions between chloroplast, mitochondria, and peroxisomes. Enzymes involved in the pathway are highlighted in red. Abbreviation of metabolites and enzymes are as follow; RuBP, Ribulose 1,5 bisphosphate; 3-PGA, 3-phosphoglycerate; 2-PG, 2-phosphoglycolate; α-KG, α-ketoglutarate; Glu, glutamate; Gly, glycine; Ser, serine; CH_2-THF, methylenetetrahydrofolate; 2-OG, 2-oxoglutarate; Gln, glutamine; Rubisco, ribulose I, 5 bisphosphates carboxylase/oxygenase; PGLP, phosphoglycolate phosphatase; GLYK, glycerate kinase; GOX, glycolate oxidase; HPR, hydroxypyruvate reductase; GGT, glutamate-glyoxylate aminotransferase; SHMT, serine hydroxymethyltransferase; GDC, glycine decarboxylase complex; GLS, glutamate synthase; GS, glutamine synthase; SGT, serine glyoxylate aminotransferase.

2.2 The C_4 cycle

2.2.1 Kranz anatomy; A CO_2 concentrating mechanism compensating photorespiration

All higher plants are classified into C_3, C_4, and Crassulacean Acid Metabolism (CAM) based on their initial carbon fixation mechanism. In C_3 species, the CO_2 is fixed in the form of a three-carbon compound by Calvin Benson cycle enzyme Rubisco in the mesophyll cells. While in C_4 plants, the atmospheric CO_2 is initially fixed into four carbons compounds in the M cell, which is transported into BS cells, where CO_2 is released by the action of the decarboxylating enzyme and is refixed by Rubisco within the Calvin cycle. C_4 plants are characterized by a high rate of photosynthesis and better adaptation to an environment of elevated temperatures, intense sunlight, and drought compared to C_3 ancestors. These characteristics are associated with the special leaf anatomy of C_4 species, where M and BS cells form successive layers around the vascular bundles. This typical leaf anatomy is known as Kranz anatomy, first described by Haberlandt (1881) (Dengler and Nelson, 1999, Fouracre *et al.,* 2014, Westhoff and Gowik, 2010).



,

Figure 5. Comparative leaf anatomy of C_3 and C_4 plants.

Unlike C_3 relatives, C_4 leaf have a higher bundle sheath/mesophyll cell ratio with more contact between both cell types. M, mesophyll cell; BS, bundle sheath cells. *Arabidopsis* leaf cross-section is taken from a self-made cross-section and *Sorghum* is adapted from El-Sharkawy, 2016.

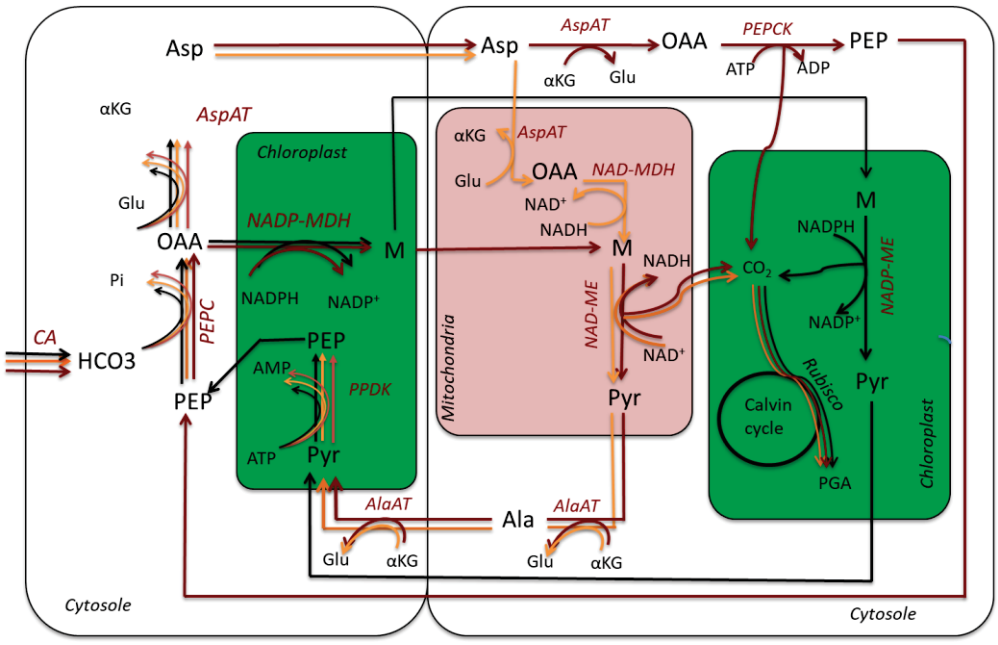
The anatomical features of C_4 plants show great alteration compared to the C_3 ancestors (Figure 5). The specialized anatomy of C_4 leaves is assumed to be developed by spatial and quantitative differentiation of the C_3 leaves (Nelson, 2010, McKown and Dengler, 2007). Such as, C_4 leaves are characterized by dense venation (McKown and Dengler, 2010), usually leading to a 1:1 ratio of M and BS cells and abundant connection between mesophyll and BS cells via plasmodesmata to exchange

metabolites (Dengler and Nelson, 1999, Nelson, 2010). In the majority of C_4 plants, the mesophyll cells are arranged in a radial direction (Ueno *et al.,* 2006) that allows each cell to be in contact with BS cells and serve to supercharge the bundle sheath cells with CO_2, which enables highly efficient and productive assimilation of CO_2 into carbohydrates by the Calvin cycle in the BS cells. The volume of intercellular space, interveinal distance, and leaf thickness are reduced compared to C_3 relatives. The BS cells of C_3 and C_4 species also differ in their size and numbers of organelles, e.g., the C_3 BS cells are smaller in size and constitute fewer chloroplasts (≃15% in *A. thaliana*) compared to C_4 BS cells, which are large in size and rich in chloroplasts and other organelles to carry out photosynthetic activities, giving C_4 leaf distinctly different appearance than a typical C_3 leaf (Leegood, 2008, Muhaidat *et al.,* 2011, Stata *et al.,* 2014).

2.2.2. The basic biochemical pathway of C_4 photosynthesis

C_4 photosynthesis is more complicated than C_3-type photosynthesis. Usually, in C_4 photosynthesis, all relevant enzymes are split between mesophyll cells and bundle sheath cells, while in C_3 leaves, the photosynthesis occurs in all photosynthetic cells (Hatch, 1987). C_4 photosynthesis is a series of biochemical and structural adjustments to concentrate CO_2 around Rubisco by the exploitation of phosphoenolpyruvate carboxylase (PEPC) and other enzymes (Sage, 2004). C_4 photosynthesis starts with the conversion of CO_2 into bicarbonate (HCO_3) by carbonic anhydrase (CA). HCO_3 is used as a substrate by the enzyme PEPC to carboxylate the 3-carbon compound phosphoenolpyruvate (PEP), resulting in the four-carbon compound oxaloacetate (OAA); hence the name C_4. OAA is immediately reduced to malate by NADP-malate dehydrogenase (NADP-MDH). Malate is then translocated from mesophyll to bundle sheath cells, where Rubisco is located and is degraded to CO_2 and pyruvate by NADP-ME (NADP-dependent malic enzyme) (Hatch *et al.,* 1975, Kanai and Edward, 1999). The enzyme Rubisco refixes the released CO_2 in the bundle sheath in the Calvin-Benson cycle. The three-carbon compound pyruvate produced because of CO_2 release in the bundle sheath, diffuses back to mesophyll cells and is regenerated into PEP by pyruvate phosphate dikinase (PPDK) to maintain the C_4 cycle as shown in figure 1 (Hatch, 1987, Sage, 2004)

However, C_4 plants also exhibit NAD-ME (NAD-malic enzyme) and PEPCK (Phosphoenolpyruvate carboxykinase) types of photosynthesis. Unlike the NADP-ME type, in NAD-ME and PEPCK subtypes, the OAA is transaminated to aspartate by the enzyme aspartate-amino transferase (AspAT) in the M cytosol and is translocated to the bundle sheath cells. However, in both NAD-ME and PEPCK types, the aspartate is converted back to OAA in the bundle sheath mitochondria and cytosol, respectively. In NAD-ME subtypes, the OAA is first reduced to malate by the enzyme NADP-ME and then de-carboxylated into pyruvate by NAD-malic enzyme (NAD-ME), releasing CO_2 into the mitochondria, which is then diffuses into chloroplasts. In the PEPCK subtypes, the OAA is de-carboxylated by PEPC, yielding PEP and CO_2 in the cytosol. The CO_2 released as a result of decarboxylation in all three types is used in the Calvin cycle, while the PEP (PEPCK types) and pyruvate (NAD-ME types) are converted into alanine and transported into mesophyll cells, as shown in the figure 6 (Hatch *et al.,* 1975, Kanai and Edward, 1999, Furbank, 2011). Various studies have reported that depending on the environmental conditions, some C_4 species with NADP-ME subtypes (maize, sugarcane, and sorghum) showed some flexibility using the three-carboxylation enzyme. Such as, maize showed a considerable amount of PEPCK activity in the bundle sheath and gene expression analysis level of PEPCK in mature sugarcane leaves.



C4 subtypes: NADP-ME, NAD-ME and PEPCK

Figure 6. Illustration of carbon fixation and concentrating mechanism in three C_4 subtypes; NADP-ME, NAD-ME, and PEPCK types.

Abbreviation of enzymes PEPC, phosphoenol-pyruvate carboxylase; NAD-MDH, NAD-malate dehydrogenase; NADP-MDH, NADP-malate dehydrogenase; NAD-ME, NAD-malic enzyme; NADP-ME, NADP malic enzyme; Rubisco, ribulose-1,5-bisphosphate carboxylase/oxygenase; AstAT, aspartate aminotransferase; AlaAT, alanine aminotransferase; PPDK, Pyruvate orthophosphate dikinase. Metabolites abbreviations: CA, carbonic anhydrase; PEP, phosphoenol-pyruvate; OAA, oxaloacetate; Asp, aspartate; Ala, alanine; Pyr, pyruvate; M, malate; PGA, 3-phosphoglyceric acid. Source image adapted from Reeves *et al.,* 2017

Unlike C_3 species, the Rubisco in C_4 has located in the BS cells and decarboxylation in this compartment leads to an elevated concentration of CO_2 around the Rubisco (Furbank, 2011). Therefore, the oxygenation reaction in C_4 has reduced drastically, leading to photorespiration repression since Rubisco's activity mainly depends on the surrounding CO_2 /O_2 ratio. Hence photorespiration in C_4 plants is not entirely abolished but occurs at a significantly low rate compared to C_3 plants.

2.2.3. Evolution of C_4 syndrome

C_4 photosynthesis has evolved at least 66 times independently from C_3 pathways over the last 35 million years (Sage *et al.,* 2011a). So far, 8145 species have been identified, which are distributed to 19 families of mono- and dicotyledonous. These species include 5044 grasses, 1322 sedges, and 1777 eudicots (Sage, 2017). These families are physiologically distinct from each other. Such diverse occurrence of C_4 plants indicates that C_4 photosynthesis evolved on several occasions independently from the C_3 ancestor during the evolution of angiosperm (Westhoff and Gowik, 2004) and the discovery of the C_3- C_4 intermediate has been hypothesized as an evolutionary intermediate to C_4 pathway (von-Caemmerer *et al.,* 2017). Though the number of species performing C_4 photosynthesis is very low, around 8000 species of more than 250,000 species worldwide, it constitutes 23% of the terrestrial biomass production on earth. C_4 evolution is assumed as a consequence of the abrupt atmospheric CO_2 decline from 1000 ppm to 390 ppm (Christin *et al.,* 2008). This dramatic reduction of

CO_2 led to the transition of C_3 to C_4 photosynthesis, first reported in a primitive C_4 grass lineage Chloridoidea, which evolved around 30 Mya (Christin *et al.,* 2008, Liu *et al.,* 2013). Besides atmospheric CO_2 concentration, high temperature, salinity, and arid condition-driven stomata closure could also promote C_4 evolution (Sage, 2004, Osborne and Sack, 2012).

The transition from C_3 to C_4 did not require drastic changes but occurred gradually with small changes in a stepwise sequence, and each step produced a new trait which was evolutionary beneficial (Sage, 2012, Willuda *et al.,* 2012). Studies on closely related species showed that during C_4 evolution, the transcription abundance of approximately 600 or even more genes was altered in C_4 compared to C_3 species (Gowik *et al.,* 2011, Külahoglu *et al.,* 2014, Bräutigam *et al.,* 2011). Sage (2004) classified the evolution of the C_4 trait into five phases, preconditioning, the evolution of proto-kranz anatomy, the establishment of C_2 photosynthesis, the establishment of the C_4 cycle in M and BS cells and optimizing the C_4 cycle. The preconditioning step starts with genome or single-gene duplications resulting in genetic redundancy. Multiple copies of a gene allow the evolutionary modification of one copy without disturbing the original function of that gene (Gowik and Westhoff, 2011), and the resulting modification leads to obtaining a new function to increase fitness toward changing environment. For example, C_3 ancestor genes CA, PEPC, PEPCK, NADP-ME, and NADP-MDH have been reported to experience duplication/neo-functionalization before they were recruited in the C_4 pathway (Liu *et al.,* 2013). Such as C_3, C_3 - C_4 intermediate, and C_4 *Flaveria* species contain two CA genes, cytosolic and chloroplastic CA, performing a distinct function. Gene duplication is followed by anatomical preconditioning, such as increasing venation in C_3 leaves and reducing the distance between M and BS cells by decreasing the interveinal space or/and enlargement of BS that allows efficient metabolites changes between two cells. Increased venation also facilitates efficient water supply, reduces evaporation, and promote adaptation toward hot and dry condition (Sage, 2004, 2014, Liu *et al.,* 2013). Increased vein density has been observed in C_3 species from the hot and semi-arid regions. It seems particular for C_3 sister clades of C_4, such as *Cleome*, *Antcharis, Euphorbia, Heliotropium, Mullogo, and Salsola* (Khoshraveh *et al.,* 2012, Marshall *et al.,* 2007, Muhaidat *et al.,* 2011, Sage *et al.,* 2011b, 2014, Voznesenskaya *el al.,* 2013,

Christin *et al.,* 2011). In *Flaveria,* enlarged BS has been observed in C_3 species, *F. cronquistii* (Sage *et al.,* 2014).

After preconditioning, the plant enters the next step of C_4 evolution, known as the establishment of Proto-kranz anatomy. In a typical C_3 ancestor, BS have few chloroplasts with little photosynthetic activity. In this step, the activation of BS cells happened, such as an increasing number of chloroplasts, mitochondria as well as enhanced Rubisco in the BS cells (Sage *et al.,* 2013, Gowik and Westhoff, 2011) first described in *Helitropium* species *H. procumbens* and *H. karwinskyi* and in *Flaveria* species, *F. pringli* and *F. rubusta* (Sage *et al.,* 2014). An increased venation raised the BS to M ratio, leading to the loss of M tissue as well as an overall photosynthetic activity since M cells show high photosynthetic activity in ancestor C_3. The evolutionary pressure to maintain photosynthetic activity would favour an increasing number of chloroplasts in BS cells. Similarly, enhanced Rubisco expression in BS cells leads to elevate photorespiration, consequently, more mitochondria and peroxisome are required to metabolize photorespiratory glycine (Liu *et al.,* 2013).

Once the proto-kranz anatomy is established, it influences the relocation of photorespiratory GDC from M to BS cells. That involved the restriction of GDC into BS tissue, achieved by molecular manipulation of the existing process in two steps. Firstly, duplication of GDC followed by sub-functionalization of two copies expressed separately in M and BS cells. Secondly, loss of expression of GDC of expression in M cells by mutation. This type of modification is observed in $C_3 - C_4$ intermediate species *Moricandia arvensis*, where the P subunit of GDC is no more functional in M cells since the enzyme is inactive without this subunit (Rylott *et al.,* 1998). Loss of GDC in M cells induced photorespiratory glycine transport to BS cells because, in M cells, the glycine is no more carboxylated. Hence, a constant transport of glycine from M to BS takes place. Decarboxylation of glycine in BS cells prevents photorespiratory CO_2 loss, thereby creating a photorespiratory CO_2 pump in the BS cells. Under the high photorespiratory condition, the glycine shuttle causes three-fold more CO_2 in the BS compared to mesophyll cells, thus enhancing photosynthetic efficiency by 30%. Elevated CO_2 may also lead to a further increase in organelle number in BS cells (Gowik and Westhoff, 2011, Sage 2004, Liu *et al.,* 2013, Bräutigam and Gowik, 2016). However, the photorespiratory pump does accumulate

not CO_2 only but also nitrogen in the BS cells, creating a nitrogen imbalance between M and BS cells. This led the plant to develop a recycling mechanism to shuttle ammonia back to the M cells. All these conditions finally led to the evolution of the C_4 cycle (Figure 7, Mallmann *et al.,* 2014).

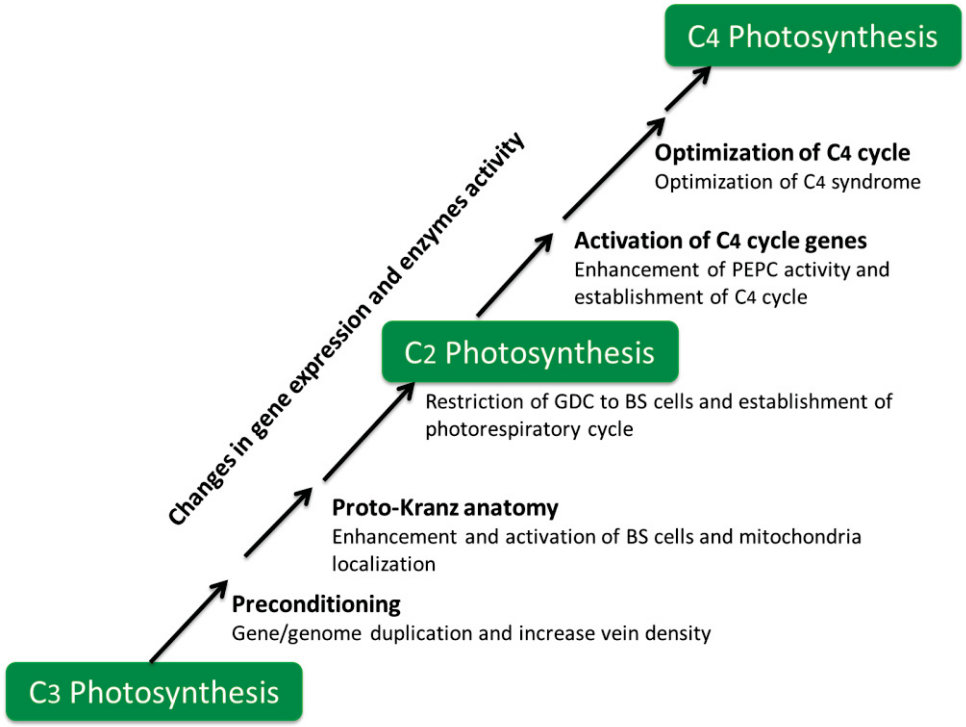


Figure 7. Model of stepwise evolution of C_4 photosynthesis.

BS, bundle sheath; GDC, glycine decarboxylase; PEPC, phosphoenolpyruvate carboxylase. Modified from Sage *et al.,* 2012 and Gowik and Westhoff, 2011.

The key step toward true C_4 photosynthesis is the enhanced expression and strict compartmentalization of key enzymes between M and BS cells, such as CA, PEPC in the M cell and Rubisco in the BS cells. Various studies have shown that the differential expression of C_4 genes is regulated at different levels, i.e., transcriptional, post-transcriptional, translational, and post-translational levels (Reeves *et al.,* 2017) and a single change in the cis/trans-regulatory element could lead to a tissue-specific gene expression. For example, in *Flaveria,* the MEM (M expression module) at the distal end *ppcA* promoter control the integrity and location of PEPC expression that went through a minimal alteration of base pair changes during C_4 evolution leading to a M-specific expression of PEPC. Similarly, the BS-specific expression of NADP-ME

and NAD-ME is also under the control of trans-acting regulatory factors (Sage *et al.,* 2012, Westhoff and Gowik, 2004). As the glycine shuttle results in elevated CO_2 in the BS cells, in consequence, the PEPC activity in the M cell could also rise to scavenge CO_2 escaping from BS cells or diffusing from intercellular space. Correspondingly, other enzymes required for the C_4 cycle would also increase to sustain PEP regeneration (Sage, 2004). This type of modification has been observed in C_3 - C_4 intermediate *Flaveria,* where PEPC activity has increased five to 20 folds compared to C_3 species (Liu *et al.,* 2013). The final step toward C_4 photosynthesis is the optimization of photosynthetic efficiency, which includes alteration in the kinetics and regulatory set point of C_4 enzymes for efficient operation in M and BS cells (Sage *et al.,* 2012).

3. Photosynthesis in *Sorghum bicolor*

Sorghum bicolor is a typical African C_4 cereal and the fourth most important cereal worldwide after rice, maize, and wheat. It is grown for fiber, food, and fodder in Africa and many other developing countries and is a source of biofuel in the US. Being a C_4 plant species, it has high photosynthetic efficiency and high water use efficiency, grown in arid, semi-arid regions and those areas which are not suitable for the growth of other cereals (Li *et al.,* 2019, Zhang *et al.,* 2011, Jagtap *et al.,* 1998, Christine *et al.,* 2005). *S. bicolor* is a C_4 plant exhibiting NADP-ME type biochemistry. Leaf development in *Sorghum* requires light contrasting other monocotyledonous species, i.e., barley, maize and oat. When *Sorghum* seedlings are grown in the dark, the etiolated leaves remain tiny and enclosed by the coleoptile. After a prolonged dark period, some leaves begin to emerge from the coleoptile but remain small and unfolded (Schrubar *et al.,* 1990). The polypeptides of water splitting apparatus, cytochrome b6f and light-harvesting complex of PSII and PSI were not detected in dark-grown seedlings, and their accumulation started only after illumination. Also, the level of plastid DNA and RNA and β-subunit of the plastid-encoded RNA polymerase increased only after illumination (Schrubar *et al.,* 1990). In addition, *S. bicolor* is characterized by a relatively small genome (~730 Mb) that is completely sequenced, consists of ten chromosomes, and shows a low frequency of gene duplications (Paterson, 2009). The gene expression patterns of M and BS cells differ drastically at a very early stage of seedlings (Wyrich *et al.,* 1998). These entire

characteristics made *S. bicolor* an important candidate to study light-induced leaf development and gene expression.

II. Scientific aims

Plants using C_4 photosynthesis have higher CO_2 and radiation use efficiency than those using the C_3 pathway. It is estimated that C_4 plants have approximately 50% higher photosynthetic efficiency than C_3 plants, but it is only used by about 3% of all vascular plant species (Edwards *et al.,* 2010). However, most crops are still using the C_3 pathway, and this process is responsible for the loss of up to 30% of assimilated CO_2. It is an ambitious goal to install the C_4 pathway into C_3 crops to boost agricultural yield. Though C_4 photosynthesis evolved from C_3 ancestors with the altered expression of thousands of genes, therefore, most enzymes required for C_4 photosynthesis are already present in C_3 species, but they acquired cell specific expression patterns. Therefore, the integration of a fully functional C_4 cycle in C_3 hence requires a prior understanding of the genetic basis, particularly the central regulators. Therefore, we pursued two different strategies.

(1) Since leaf development and chloroplast development in *Sorghum* are strictly light-dependent. We wanted to exploit this fact by recording developmental transcriptome patterns after inducing leaf development by light in *Sorghum* seedlings that had been etiolated. We aimed to identify regulators, such as transcription factors, that could explain how leaf development is regulated in C_4 species. The transcriptome data set was prepared from etiolated and illuminated *Sorghum* seedlings at six-time points starting from 1h to 24h of illumination. The transcriptome data was assessed using the pairwise comparison between etiolated and illuminated seedlings at each time point. The results of this work are presented in the Manuscript I of the thesis.

(2) In parallel, we pursued a genetic approach with the C_3 genetic model *Arabidopsis thaliana,* explicitly aiming to identify regulators of bundle-sheath development and hoping to use this knowledge for the engineering of Kranz anatomy in current C_3 crops. Our study identified genes responsible for the mutant phenotype of *ebss1* and *fbss1* using CRISPR/Cas9 gene modification. The *ebss1* and *fbss1* mutants were isolated from EMS forward genetic screening showing altered reporter gene signal in the bundle sheath cell. The results of this work are presented in the Manuscript II of the thesis.

III. Summary

Leaf morphogenesis is a crucial process for a plant to adjust the cellular arrangement of photosynthetic tissue in an optimized way to improve photosynthetic activity. During evolution, leaves of C_3 plants underwent a few anatomical and cellular adaptations to acquire C_4 anatomy. The changes include the development of Kranz anatomy, enlargement of mesophyll (M) and bundle sheath size (BS), increased number of chloroplasts in BS cells, enrichment of plasmodesmata between M and BS cell to facilitate metabolic exchange, followed by compartmentation of CO_2 assimilation and fixation by increasing PEPC activity in the M cell and Rubisco in the BS chloroplast. To identify novel regulators underpinning C_4-specific leaf development, in this study, we presented two different approaches.

In the first approach, we analyzed leaf anatomy and transcriptome dataset of etiolated and illuminated *Sorghum* seedlings using light as a trigger of leaf development. We observed that the leaf development in *Sorghum* is arrested in the dark because, during etiolation, the leaves remained pale and enclosed in the coleoptile. Leaves emergence and greening were observed only after light exposure. Transcriptome data analysis showed a strong correlation between leaf development and chloroplast biogenesis in *Sorghum*. Genes involved in chloroplast development and differentiation, assembly of photosynthetic apparatus and C_4 cycle genes showed light-induced accumulation. Moreover, genes involved in various pathways do not accumulate sequentially but rather coordinately. More interestingly, the nuclear-encoded genes expressed at early exposure, while plastid-encoded genes required a longer exposure time.

In the second approach, numerous BS mutants were produced using a forward genetic screen of EMS-induced mutant lines in *Arabidopsis thaliana* using a bundle sheath labelled with *GFP* reporter genes. The BS mutants were selected based on the altered reporter gene signal intensity and hypothesized that the altered signal intensity of the BS could be due to changes in BS numbers and/or chloroplast inside the BS cells. By pursuing a mapping-by-sequencing approach, the genomic segments containing mutated candidate genes were identified. Two EMS-induced mutant lines, *ebss1* (high reporter genes signal in the BS cell) and *fbss1* (low reporter genes signal in the BS cell), were selected for further investigation to identify the putative genes responsible for *ebss1* and *fbss1* mutant phenotype. By pursuing CRISPR/Cas9-induced genetic

modification of candidate genes, it was revealed that the mutant allele of At2g25970 was responsible for high *GFP* signals in *ebss1* BS cells and the elevated *GFP* signal intensity was due to an increased number of BS cells. Furthermore, At5g04940 (*SUVH1*) was proven to be a responsible gene for the *fbss1* mutant phenotype.

IV. Literature

Andersson, I. (2008). Catalysis and regulation in Rubisco. *Journal of Experimental Botany, 59*, 1555-1568.

Andriankaja, M., Dhondt, S., De Bodt, S., Vanhaeren, H., Coppens, F., De Milde, L., . . . Beemster, G. T. (2012). Exit from proliferation during leaf development in *Arabidopsis thaliana*: a not-so-gradual process. *Developmental Cell, 22*, 64-78.

Bar, M., and Ori, N. (2014). Leaf development and morphogenesis. *Development, 141*, 4219-4230.

Bar-On YM and Milo R. (2019). The global mass and average rate of rubisco. *Proceedings of the National Academy of Sciences*. 116, 4738-43.

Biedroń, M., and Banasiak, A. (2018). Auxin-mediated regulation of vascular patterning in *Arabidopsis thaliana* leaves. *Plant Cell Reports, 37*, 1215-1229.

Bowes, G., Ogren, W., and Hageman, R. (1971). Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochemical and biophysical research communications, 45*, 716-722.

Bräutigam, A., and Gowik, U. (2016). Photorespiration connects C3 and C4 photosynthesis. *Journal of Experimental Botany, 67*, 2953-2962.

Bräutigam, A., Kajala, K., Wullenweber, J., Sommer, M., Gagneul, D., Weber, K. L., . . . Lercher, M. J. (2011). An mRNA blueprint for C4 photosynthesis derived from comparative transcriptomics of closely related C3 and C4 species. *Plant Physiology, 155*, 142-156.

Byrne, M. E., Simorowski, J., and Martienssen, R. A. (2002). ASYMMETRIC LEAVES1 reveals knox gene redundancy in *Arabidopsis*. *Development, 129*, 1957-1965.

Cheng, X., Ruyter-Spira, C., and Bouwmeester, H. (2013). The interaction between strigolactones and other plant hormones in the regulation of plant development. *Frontiers in Plant Science, 4*, 199.

Christin, P.-A., Besnard, G., Samaritani, E., Duvall, M. R., Hodkinson, T. R., Savolainen, V., and Salamin, N. (2008). Oligocene CO2 decline promoted C4 photosynthesis in grasses. *Current biology, 18*, 37-43.

Christin, P. A., Sage, T. L., Edwards, E. J., Ogburn, R. M., Khoshravesh, R., and Sage, R. F. (2011). Complex evolutionary transitions and the significance of C3–C4 intermediate forms of photosynthesis in Molluginaceae. *Evolution: International Journal of Organic Evolution, 65*, 643-660.

Christine, K., Springob, K., Schmidt, J., Nicholson, R. L., Chu, I. K., Yip, W. K., and Lo, C. (2005). A stilbene synthase gene (SbSTS1) is involved in host and nonhost defense responses in *Sorghum*. *Plant Physiology, 138*, 393-401.

Cosgrove, D. J. (2005). Growth of the plant cell wall. *Nature reviews molecular cell biology, 6*, 850-861.

Cox, N., and Smith, L. M. (2019). A novel upstream regulator of trichome development inhibitors. In: Am Soc Plant Biol.

Dengler, N. G., and Nelson, T. (1999). Leaf structure and development in C4 plants. *C, 4*, 133-172.

Doroshkov, A. V., Konstantinov, D. K., Afonnikov, D. A., and Gunbin, K. V. (2019). The evolution of gene regulatory networks controlling *Arabidopsis thaliana* L. trichome development. *BMC Plant Biology, 19*, 71-85.

Du, F., Guan, C., and Jiao, Y. (2018). Molecular mechanisms of leaf morphogenesis. *Molecular Plant, 11*, 1117-1134.

Du, Q., and Wang, H. (2015). The role of HD-ZIP III transcription factors and miR165/166 in vascular development and secondary cell wall formation. *Plant signaling and behavior, 10*, e1078955.

Edwards, E. J., Osborne, C. P., Strömberg, C. A., Smith, S. A., and Consortium, C. G. (2010). The origins of C4 grasslands: integrating evolutionary and ecosystem science. *Science, 328*, 587-591.

Efroni, I., Eshed, Y., and Lifschitz, E. (2010). Morphogenesis of simple and compound leaves: a critical review. *The Plant Cell, 22*, 1019-1032.

El-Sharkawy, M. A. (2016). Prospects of photosynthetic research for increasing agricultural productivity, with emphasis on the tropical C 4 Amaranthus and the cassava C 3-C 4 crops. *Photosynthetica, 54*, 161-184.

Eloy, N. B., de Freitas Lima, M., Van Damme, D., Vanhaeren, H., Gonzalez, N., De Milde, L., . . . Ferreira, P. C. (2011). The APC/C subunit 10 plays an essential role in cell proliferation during leaf development. *The Plant Journal, 68*, 351-363.

Eshed, Y., Izhaki, A., Baum, S. F., Floyd, S. K., and Bowman, J. L. (2004). Asymmetric leaf development and blade expansion in *Arabidopsis* are mediated by KANADI and YABBY activities. *Development, 131*, 2997-3006.

Fernie, A. R., Bauwe, H., Eisenhut, M., Florian, A., Hanson, D. T., Hagemann, M., . . . Peterhänsel, C. (2013). Perspectives on plant photorespiratory metabolism. *Plant Biology, 15*, 748-753.

Fouracre, J. P., Ando, S., and Langdale, J. A. (2014). Cracking the Kranz enigma with systems biology. *Journal of Experimental Botany, 65*, 3327-3339.

Foyer, C. H., Neukermans, J., Queval, G., Noctor, G., and Harbinson, J. (2012). Photosynthetic control of electron transport and the regulation of gene expression. *Journal of Experimental Botany, 63*, 1637-1661.

Furbank, R. T. (2011). Evolution of the C4 photosynthetic mechanism: are there really three C4 acid decarboxylation types? *Journal of Experimental Botany, 62*, 3103-3108.

Furbank, R. T., and Taylor, W. C. (1995). Regulation of photosynthesis in C3 and C4 plants: a molecular approach. *The Plant Cell, 7*, 797.

Geisler, M., Kolukisaoglu, H. U., Bouchard, R., Billion, K., Berger, J., Saal, B., . . . Dudler, R. (2003). TWISTED DWARF1, a unique plasma membrane-anchored immunophilin-like protein, interacts with *Arabidopsis* multidrug resistance-like transporters AtPGP1 and AtPGP19. *Molecular Biology of the Cell, 14*, 4238-4249.

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

Gowik, U., and Westhoff, P. (2011). The path from C3 to C4 photosynthesis. *Plant Physiology, 155*, 56-63.

Gutierrez, C. (2009). The *Arabidopsis* cell division cycle. *The Arabidopsis Book/American Society of Plant Biologists, 7*.

Hachez, C., Ohashi-Ito, K., Dong, J., and Bergmann, D. C. (2011). Differentiation of *Arabidopsis* guard cells: analysis of the networks incorporating the basic helix-loop-helix transcription factor, FAMA. *Plant Physiology, 155*, 1458-1472.

Hasson, A., Blein, T., and Laufs, P. (2010). Leaving the meristem behind: the genetic and molecular control of leaf patterning and morphogenesis. *Comptes Rendus Biologies, 333*, 350-360.

Hatch, M., Kagawa, T., and Craig, S. (1975). Subdivision of C4-pathway species based on differing C4 acid decarboxylating systems and ultrastructural features. *Functional Plant Biology, 2*, 111-128.

Hatch, M. D. (1987). C4 photosynthesis: a unique elend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta (BBA)-Reviews on Bioenergetics, 895*, 81-106.

Hodges, M., Dellero, Y., Keech, O., Betti, M., Raghavendra, A. S., Sage, R., . . . Weber, A. P. (2016). Perspectives for a better understanding of the metabolic integration of photorespiration within a complex plant primary metabolism network. *Journal of Experimental Botany, 67*, 3015-3026.

Horiguchi, G., Kim, G. T., and Tsukaya, H. (2005). The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of *Arabidopsis thaliana*. *The Plant Journal, 43*, 68-78.

Hu, J., Baker, A., Bartel, B., Linka, N., Mullen, R. T., Reumann, S., and Zolman, B. K. (2012). Plant peroxisomes: biogenesis and function. *The Plant Cell, 24*, 2279-2303.

Hur, Y. S., Um, J. H., Kim, S., Kim, K., Park, H. J., Lim, J. S., . . . Lim, J. (2015). Arabidopsis thaliana homeobox 12 (ATHB 12), a homeodomain-leucine zipper protein, regulates leaf growth by promoting cell expansion and endoreduplication. *New Phytologist, 205*, 316-328.

Hügler, M., and Sievert, S. M. (2011). Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. *Annual Review of Marine Science, 3*, 261-289.

Hülskamp, M. (2004). Plant trichomes: a model for cell differentiation. *Nature reviews molecular cell biology, 5*, 471-480.

Ichihashi, Y., and Tsukaya, H. (2015). Behavior of leaf meristems and their modification. *Frontiers in Plant Science, 6*, 1060.

Ikezaki, M., Kojima, M., Sakakibara, H., Kojima, S., Ueno, Y., Machida, C., and Machida, Y. (2010). Genetic networks regulated by ASYMMETRIC LEAVES1 (AS1) and AS2 in leaf development in *Arabidopsis thaliana*: KNOX genes control five morphological events. *The Plant Journal, 61*, 70-82.

Iwakawa, H., Ueno, Y., Semiarti, E., Onouchi, H., Kojima, S., Tsukaya, H., . . . Machida, C. (2002). The ASYMMETRIC LEAVES2 gene of *Arabidopsis thaliana*, required for formation of a symmetric flat leaf lamina, encodes a member of a novel family of proteins characterized by cysteine repeats and a leucine zipper. *Plant and cell physiology, 43*, 467-478.

Jagtap, V., Bhargava, S., Streb, P., and Feierabend, J. (1998). Comparative effect of water, heat and light stresses on photosynthetic reactions in *Sorghum* bicolor (L.) Moench. *Journal of Experimental Botany, 49*, 1715-1721.

Kajala, K., Covshoff, S., Karki, S., Woodfield, H., Tolley, B. J., Dionora, M. J. A., . . . Hibberd, J. M. (2011). Strategies for engineering a two-celled C4 photosynthetic pathway into rice. *Journal of Experimental Botany, 62*, 3001-3010.

Kalve, S., De Vos, D., and Beemster, G. T. (2014). Leaf development: a cellular perspective. *Frontiers in Plant Science, 5*, 362.

Kalve, S., De Vos, D., and Beemster, G. T. (2014). Leaf development: a cellular perspective. *Frontiers in Plant Science, 5*, 362.

Kalve, S., De Vos, D., Nektarios, M. M., Domagalska, M. A., Stelmaszewska, J., De, L., . . . Beemster, G. T. (2014). Downregulation of KRPs promotes cell division and endoreduplication in the leaves of *Arabidopsis thaliana*. A Systems Biology Approach to Understand Leaf Growth Regulation in *Arabidopsis thaliana*, 183.

Kanai, R., and Edwards, G. E. (1999). The biochemistry of C4 photosynthesis. *C4 plant biology, 49*, 87.

Kanaoka, M. M., Pillitteri, L. J., Fujii, H., Yoshida, Y., Bogenschutz, N. L., Takabayashi, J., . . . Torii, K. U. (2008). SCREAM/ICE1 and SCREAM2 specify three cell-state transitional steps leading to *Arabidopsis* stomatal differentiation. *The Plant Cell, 20*, 1775-1785.

Kathare, P. K., Dharmasiri, S., and Dharmasiri, N. (2018). SAUR53 regulates organ elongation and apical hook development in *Arabidopsis*. *Plant signaling and behavior, 13*, e1514896.

Kerstetter, R. A., Bollman, K., Taylor, R. A., Bomblies, K., and Poethig, R. S. (2001). KANADI regulates organ polarity in *Arabidopsis*. *Nature, 411*, 706-709.

Khoshravesh, R., Akhani, H., Sage, T. L., Nordenstam, B., and Sage, R. F. (2012). Phylogeny and photosynthetic pathway distribution in Anticharis Endl.(Scrophulariaceae). *Journal of Experimental Botany, 63*, 5645-5658.

Kirschner, S. (2017). Identification of a cis-regulatory module for bundle sheath-specific gene expression and transcriptome analysis of the chilling response of the C4 grass Zea mays. *Heinrich-Heine-Universität Düsseldorf.*

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

Klein, R., and Mullet, J. (1990). Light-induced transcription of chloroplast genes. psbA transcription is differentially enhanced in illuminated barley. *Journal of Biological Chemistry, 265*, 1895-1902.

Külahoglu, C., Denton, A. K., Sommer, M., Maß, J., Schliesky, S., Wrobel, T. J., . . . Simon, R. (2014). Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae C3 and C4 plant species. *The Plant Cell, 26*, 3243-3260.

Lau, O. S., and Bergmann, D. C. (2012). Stomatal development: a plant's perspective on cell polarity, cell fate transitions and intercellular communication. *Development, 139*, 3683-3692.

Lee, B. H., Ko, J.-H., Lee, S., Lee, Y., Pak, J.-H., and Kim, J. H. (2009). The *Arabidopsis* GRF-INTERACTING FACTOR gene family performs an overlapping function in determining organ size as well as multiple developmental properties. *Plant Physiology, 151*, 655-668.

Lee, S.-W., Feugier, F. G., and Morishita, Y. (2014). Canalization-based vein formation in a growing leaf. *Journal of theoretical biology, 353*, 104-120.

Leegood, R. C. (2008). Roles of the bundle sheath cells in leaves of C3 plants. *Journal of Experimental Botany, 59*, 1663-1673.

Li, Y., Wang, W., Feng, Y., Tu, M., Wittich, P. E., Bate, N. J., and Messing, J. (2019). Transcriptome and metabolome reveal distinct carbon allocation patterns during internode sugar accumulation in different *Sorghum* genotypes. *Plant biotechnology journal, 17*, 472-487.

Liu, Z., Sun, N., Yang, S., Zhao, Y., Wang, X., Hao, X., and Qiao, Z. (2013). Evolutionary transition from C3 to C4 photosynthesis and the route to C4 rice. *Biologia, 68*, 577-586.

Lu, D., Wang, T., Persson, S., Mueller-Roeber, B., and Schippers, J. H. (2014). Transcriptional control of ROS homeostasis by KUODA1 regulates cell expansion during leaf development. *Nature communications, 5*, 1-9.

Machida, C., Nakagawa, A., Kojima, S., Takahashi, H., and Machida, Y. (2015). The complex of ASYMMETRIC LEAVES (AS) proteins plays a central role in antagonistic interactions of genes for leaf polarity specification in *Arabidopsis*. *Wiley Interdisciplinary Reviews: Developmental Biology, 4*, 655-671.

Majda, M., and Robert, S. (2018). The role of auxin in cell wall expansion. *International Journal of Molecular Sciences, 19*, 951.

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

Marshall, D. M., Muhaidat, R., Brown, N. J., Liu, Z., Stanley, S., Griffiths, H., . . . Hibberd, J. M. (2007). Cleome, a genus closely related to *Arabidopsis*, contains species spanning a developmental progression from C3 to C4 photosynthesis. *The Plant Journal, 51*, 886-896.

Marks, M. D., and Feldmann, K. A. (1989). Trichome development in *Arabidopsis thaliana*. I. T-DNA tagging of the GLABROUS1 gene. *The Plant Cell, 1*, 1043-1050.

McKown, A. D., and Dengler, N. G. (2010). Vein patterning and evolution in C4 plants. *Botany, 88*, 775-786.

Matsumura, Y., Ohbayashi, I., Takahashi, H., Kojima, S., Ishibashi, N., Keta, S., . . . Echeverria, M. (2016). A genetic link between epigenetic repressor AS1-AS2 and a putative small subunit processome in leaf polarity establishment of *Arabidopsis*. *Biology open, 5*, 942-954.

Maugarny-Calès, A., and Laufs, P. (2018). Getting leaves into shape: a molecular, cellular, environmental and evolutionary view. *Development, 145*.

McConnell, J. R., Emery, J., Eshed, Y., Bao, N., Bowman, J., and Barton, M. K. (2001). Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. *Nature, 411*, 709-713.

McKown, A. D., and Dengler, N. G. (2007). Key innovations in the evolution of Kranz anatomy and C4 vein pattern in Flaveria (Asteraceae). *American Journal of Botany, 94*, 382-399.

Muhaidat, R., Sage, T. L., Frohlich, M. W., Dengler, N. G., and Sage, R. F. (2011). Characterization of C3–C4 intermediate species in the genus Heliotropium L.(Boraginaceae): anatomy, ultrastructure and enzyme activity. *Plant, Cell and Environment, 34*, 1723-1736.

Nakata, M., and Okada, K. (2013). The leaf adaxial-abaxial boundary and lamina growth. *Plants, 2*, 174-202.

Nelissen, H., Gonzalez, N., and Inze, D. (2016). Leaf growth in dicots and monocots: so different yet so alike. *Current Opinion in Plant Biology, 33*, 72-76.

Nelson, N., and Yocum, C. F. (2006). Structure and function of photosystems I and II. *Annu. Rev. Plant Biol., 57*, 521-565.

Nelson, T. (2010). Development of leaves in C 4 plants: Anatomical features that support C 4 metabolism. In C4 photosynthesis and related CO2 concentrating mechanisms. Springer, 147-159.

Nelson, T., Harpster, M. H., Mayfield, S. P., and Taylor, W. C. (1984). Light-regulated gene expression during maize leaf development. *The Journal of Cell Biology, 98*, 558-564.

Nelson, T., and Langdale, J. A. (1989). Patterns of leaf development in C4 plants. *The Plant Cell, 1*, 3.

Nikolov, L. A., Runions, A., Gupta, M. D., and Tsiantis, M. (2019). Leaf development and evolution. *Current Topics in Developmental Biology, 131*, 109-139.

Noir, S., Marrocco, K., Masoud, K., Thomann, A., Gusti, A., Bitrian, M., . . . Genschik, P. (2015). The control of *Arabidopsis thaliana* growth by cell proliferation and endoreplication requires the F-box protein FBL17. *The Plant Cell, 27*, 1461-1476.

Oakley, J. C., Sultmanis, S., Stinson, C. R., Sage, T. L., and Sage, R. F. (2014). Comparative studies of C3 and C4 Atriplex hybrids in the genomics era: physiological assessments. *Journal of Experimental Botany, 65*, 3637-3647.

Osborne, C. P., and Sack, L. (2012). Evolution of C4 plants: a new hypothesis for an interaction of CO2 and water relations mediated by plant hydraulics.

Philosophical Transactions of the Royal Society B: Biological Sciences, 367, 583-600.

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., . . . Poliakov, A. (2009). The *Sorghum* bicolor genome and the diversification of grasses. *Nature, 457*, 551-556.

Pattanaik, S., Patra, B., Singh, S. K., and Yuan, L. (2014). An overview of the gene regulatory network controlling trichome development in the model plant, *Arabidopsis*. *Frontiers in Plant Science, 5*, 259.

Peterhansel, C., Horst, I., Niessen, M., Blume, C., Kebeish, R., Kürkcüoglu, S., and Kreuzaler, F. (2010). Photorespiration. *The Arabidopsis Book/American Society of Plant Biologists, 8*.

Pick, T. R., Bräutigam, A., Schulz, M. A., Obata, T., Fernie, A. R., and Weber, A. P. (2013). PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a unique class of metabolite transporters. *Proceedings of the National Academy of Sciences, 110*, 3185-3190.

Pillitteri, L. J., and Torii, K. U. (2012). Mechanisms of stomatal development. *Annual Review of Plant Biology, 63*, 591-614.

Qi, F., and Zhang, F. (2020). Cell cycle regulation in the plant response to stress. *Frontiers in Plant Science, 10*, 1765.

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

Reumann, S., and Weber, A. P. (2006). Plant peroxisomes respire in the light: some gaps of the photorespiratory C2 cycle have become filled—others remain. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 1763*, 1496-1510.

Reeves, G., Grangé-Guermente, M. J., and Hibberd, J. M. (2017). Regulatory gateways for cell-specific gene expression in C4 leaves with Kranz anatomy. *Journal of Experimental Botany, 68*, 107-116.

Rochaix, J.-D. (2011). Reprint of: Regulation of photosynthetic electron transport Bioenergetics.

Rodriguez, R. E., Debernardi, J. M., and Palatnik, J. F. (2014). Morphogenesis of simple leaves: regulation of leaf size and shape. *Wiley Interdisciplinary Reviews: Developmental Biology, 3*, 41-57.

Runions, A., Tsiantis, M., and Prusinkiewicz, P. (2017). A common developmental program can produce diverse leaf shapes. *New Phytologist, 216*, 401-418.

Rutschow, H. L., Baskin, T. I., and Kramer, E. M. (2014). The carrier AUXIN RESISTANT (AUX1) dominates auxin flux into *Arabidopsis* protoplasts. *New Phytologist, 204*, 536-544.

Rylott, E. L., Metzlaff, K., and Rawsthorne, S. (1998). Developmental and Environmental Effects on the Expression of the C3-C4 Intermediate Phenotype inMoricandia arvensis. *Plant Physiology, 118*, 1277-1284.

Sage, R. F. (2004). The evolution of C4 photosynthesis. *New Phytologist, 161*, 341-370.

Sage, R. F. (2017). A portrait of the C4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *Journal of Experimental Botany, 68*, e11-e28.

Sage, R. F., Christin, P.-A., and Edwards, E. J. (2011). The C4 plant lineages of planet Earth. *Journal of Experimental Botany, 62*, 3155-3169.

Sage, R. F., Khoshravesh, R., and Sage, T. L. (2014). From proto-Kranz to C4 Kranz: building the bridge to C4 photosynthesis. *Journal of Experimental Botany, 65*, 3341-3356.

Sage, R. F., Sage, T. L., and Kocacinar, F. (2012). Photorespiration and the evolution of C4 photosynthesis. *Annual Review of Plant Biology, 63*, 19-47.

Sage, R. F., and Zhu, X.-G. (2011). Exploiting the engine of C4 photosynthesis. *Journal of Experimental Botany, 62*, 2989-3000.

Sage, T. L., Busch, F. A., Johnson, D. C., Friesen, P. C., Stinson, C. R., Stata, M., . . . Sage, R. F. (2013). Initial events during the evolution of C4 photosynthesis in C3 species of Flaveria. *Plant Physiology, 163*, 1266-1276.

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 C2 photosynthesis in Euphorbia subgenus Chamaesyce (Euphorbiaceae). *Journal of Experimental Botany, 62*, 3183-3195.

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

Schmidt, R., Kunkowska, A. B., and Schippers, J. H. (2016). Role of reactive oxygen species during cell expansion in leaves. *Plant Physiology, 172*, 2098-2106.

Schrubar, H., Wanner, G., and Westhoff, P. (1991). Transcriptional control of plastid gene expression in greening *Sorghum* seedlings. *Planta, 183*, 101-111.

Shu, G., Pontieri, V., Dengler, N. G., and Mets, L. J. (1999). Light induction of cell type differentiation and cell-type-specific gene expression in cotyledons of a C4Plant, Flaveria trinervia. *Plant Physiology, 121*, 731-741.

Sinha, N. (1999). Leaf development in angiosperms. *Annual Review of Plant Biology, 50*, 419-446.

Skalák, J., Vercruyssen, L., Claeys, H., Hradilová, J., Černý, M., Novák, O., . . . Coppens, F. (2019). Multifaceted activity of cytokinin in leaf development shapes its size and structure in *Arabidopsis*. *The Plant Journal, 97*, 805-824.

Stata, M., Sage, T. L., Rennie, T. D., Khoshravesh, R., Sultmanis, S., Khaikin, Y., . . . Sage, R. F. (2014). Mesophyll cells of C4 plants have fewer chloroplasts than those of closely related C3 plants. *Plant, Cell and Environment, 37*, 2587-2600.

Talbert, P. B., Adler, H. T., Parks, D. W., and Comai, L. (1995). The REVOLUTA gene is necessary for apical meristem development and for limiting cell divisions in the leaves and stems of *Arabidopsis thaliana. Development, 121*, 2723-2735.

Tsukaya, H. (2013). Leaf development. *The Arabidopsis Book/American Society of Plant Biologists, 11*.

Tsuge, T., Tsukaya, H., and Uchimiya, H. (1996). Two independent and polarized processes of cell elongation regulate leaf blade expansion in *Arabidopsis thaliana* (L.) Heynh. *Development, 122*, 1589-1600.

Tsuji, J., and Coe, L. (2013). The glabra1 mutation affects the stomatal patterning of *Arabidopsis thaliana* rosette leaves. *Bios, 84*, 92-97.

Ueno, O., Kawano, Y., Wakayama, M., and Takeda, T. (2006). Leaf vascular systems in C3 and C4 grasses: a two-dimensional analysis. *Annals of Botany, 97*, 611-621.

Varoquaux, N., Cole, B., Gao, C., Pierroz, G., Baker, C. R., Patel, D., . . . Sievert, J. (2019). Transcriptomic analysis of field-droughted *Sorghum* from seedling to maturity reveals biotic and metabolic responses. *Proceedings of the National Academy of Sciences, 116*, 27124-27132.

Vercruyssen, L., Verkest, A., Gonzalez, N., Heyndrickx, K. S., Eeckhout, D., Han, S.-K., . . . Andriankaja, M. (2014). ANGUSTIFOLIA3 binds to SWI/SNF chromatin remodeling complexes to regulate transcription during *Arabidopsis* leaf development. *The Plant Cell, 26*, 210-229.

Vercruysse, J., Baekelandt, A., Gonzalez, N., and Inzé, D. (2020). Molecular networks regulating cell division during *Arabidopsis* leaf growth. *Journal of Experimental Botany, 71*, 2365-2378.

Von Caemmerer, S., Ghannoum, O., and Furbank, R. T. (2017). C4 photosynthesis: 50 years of discovery and innovation. *Journal of Experimental Botany, 68*, 97.

Voznesenskaya, E. V., Koteyeva, N. K., Akhani, H., Roalson, E. H., and Edwards, G. E. (2013). Structural and physiological analyses in Salsoleae (Chenopodiaceae) indicate multiple transitions among C3, intermediate, and C4 photosynthesis. *Journal of Experimental Botany, 64*, 3583-3604.

Walker, J. D., Oppenheimer, D. G., Concienne, J., and Larkin, J. C. (2000). SIAMESE, a gene controlling the endoreduplication cell cycle in *Arabidopsis thaliana* trichomes. *Development, 127*, 3931-3940.

Wang, G., Zhong, M., Shuai, B., Song, J., Zhang, J., Han, L., . . . Song, R. (2017). E+ subgroup PPR protein defective kernel 36 is required for multiple mitochondrial transcripts editing and seed development in maize and *Arabidopsis*. *New Phytologist, 214*, 1563-1578.

Wang, R., Zhao, J., Jia, M., Xu, N., Liang, S., Shao, J., . . . Yu, F. (2018). Balance between cytosolic and chloroplast translation affects leaf variegation. *Plant Physiology, 176*, 804-818.

Wang, C., Guo, L., Li, Y., and Wang, Z. (2012). *Systematic comparison of C3 and C4 plants based on metabolic network analysis.* Paper presented at the BMC systems biology.

Wenzel, C. L., Schuetz, M., Yu, Q., and Mattsson, J. (2007). Dynamics of MONOPTEROS and PIN-FORMED1 expression during leaf vein pattern formation in *Arabidopsis thaliana. The Plant Journal, 49*, 387-398.

Westhoff, P., and Gowik, U. (2004). Evolution of C4 phosphoenolpyruvate carboxylase. Genes and proteins: a case study with the genus Flaveria. *Annals of Botany, 93*, 13-23.

Westhoff, P., and Gowik, U. (2010). Evolution of C4 photosynthesis—looking for the master switch. *Plant Physiology, 154*, 598-601.

Wuyts, N., Palauqui, J.-C., Conejero, G., Verdeil, J.-L., Granier, C., and Massonnet, C. (2010). High-contrast three-dimensional imaging of the *Arabidopsis* leaf enables the analysis of cell dimensions in the epidermis and mesophyll. *Plant Methods, 6*, 1-14.

Wyrich, R., Dressen, U., Brockmann, S., Streubel, M., Chang, C., Qiang, D., . . . Westhoff, P. (1998). The molecular basis of C 4 photosynthesis in *Sorghum*: isolation, characterization and RFLP mapping of mesophyll-and bundle-sheath-specific cDNAs obtained by differential screening. *Plant Molecular Biology, 37*, 319-335.

Xiong, Y., and Jiao, Y. (2019). The diverse roles of auxin in regulating leaf development. *Plants, 8*, 243.

Yang, Y., Hammes, U. Z., Taylor, C. G., Schachtman, D. P., and Nielsen, E. (2006). High-affinity auxin transport by the AUX1 influx carrier protein. *Current biology, 16*, 1123-1127.

Yang, M., and Sack, F. D. (1995). The too many mouths and four lips mutations affect stomatal production in *Arabidopsis*. *The Plant Cell, 7*, 2227-2239.

Zhang, L., Zheng, Y., Jagadeeswaran, G., Li, Y., Gowdu, K., and Sunkar, R. (2011). Identification and temporal expression analysis of conserved and novel microRNAs in *Sorghum*. *Genomics, 98*, 460-468.

V. Manuscripts

1. Transcriptome profiling of light induced leaf development in *Sorghum bicolor* L. Tx430

2. Identifying and isolating genes affecting leaf development in *Arabidopsis thaliana*

Manuscript I

Transcriptome profiling of light induced leaf development in *Sorghum bicolor* L. Tx430

Transcriptome profiling of light-induced leaf development in *Sorghum bicolor* L. Tx430

Zahida Bano* and Peter Westhoff

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

To whom correspondence should be addressed*

Zahida Bano*
Email: Zahida.Bano@hhu.de

Peter Westhoff
Email: West@hhu.de

Abstract

Aiming to understand light regulation of C_4 leaf development, including chloroplast development in *Sorghum*, we investigated differential gene expression between seven days old, etiolated and etiolated seedlings illuminated for 1h, 2h, 4h, 6h, 8h and 24h using RNA sequencing approaches. Among 4742 differentially expressed genes between etiolated and de-etiolated seedlings at six-time points, large number of genes were associated with chloroplast development and organization, regulation and assembly of the photosynthetic machinery, and genes involved in the regulation of the C_4 cycle. The expression of genes involved in chloroplast division and differentiation increased in the early hours of illumination, i.e., *HY5* was observed at its peak after 1h of illumination. Conversely, the repressor of photomorphogenesis, *PIF3* and *PIF4,* were downregulated with increased exposure time. In addition, photosystem-related genes show high sensitivity to light. Many of them were not detected in the etiolated *Sorghum* seedlings, such as PSII and PSI subunits, light-harvesting complexes, ATP synthase, and cytochrome b/f complex but increased substantially after illumination, suggesting that the biogenesis of the photosynthetic apparatus correlates with the

transition from dark-grown to light-grown morphology. We also observed light-dependent accumulation of carbon assimilation genes, including *RBCS* and *PEPC*. However, the amount of transcript increase was not uniform. No increase in the expression of *RBCL* was observed during 24h of illumination. Furthermore, we observed two distinct regulatory phases of the establishment of photosynthesis during chloroplast development, i.e., the nuclear-encoded genes of photosynthetic apparatus expressed at early chloroplast development (2-4h of illumination). In contrast, plastid-encoded genes do not express during 24h of illumination except for two transcripts of *CP47* and number of ribosomal subunits were seen in 6h and 8h of illumination, respectively.

Keywords: *Sorghum bicolor*, differential gene expression, etiolation, de-etiolation, transcriptome

Introduction

Leaves are among the most specialized plant organs where photosynthetic light capture occurs. In addition, leaves perceive and transmit environmental signals to other plant organs. During leaf development, the mesophyll cells differentiate to become chloroplast-filled, a process involving morphogenesis and cell-specific organelle biogenesis. Understanding the mechanism of photosynthetic activity requires intimate knowledge of how cells and chloroplast within are produced and developed (Leech, 1984, Pogson and Albrecht, 2011). The establishment of functional chloroplasts is a complex process, and the molecular intricacies have not been fully elucidated. The differentiation of chloroplasts from plastid precursors to photosynthetically active chloroplasts follows a clear developmental program, where chloroplast is derived either from non-photosynthetic protoplast or via dark-grown intermediate known as etioplasts. This transformation of proplastids to mature chloroplasts requires an array of complex processes, including coordinated expression of plastid and nuclear genes, lipid synthesis, import of protein into plastids, and assembly of protein into thylakoid membrane, etc. (Pogson and Albrecht, 2011, Jarvis and Lopez-Juez, 2013, Cackett *et al.,* 2022). Chloroplast comprises approximately 3000 nuclear-encoded proteins involved in a wide range of functional processes required for chloroplast biogenesis (Richly *et al.,* 2003, Rolland *et al.,* 2012). The mechanisms of how chloroplast and nuclear gene expression are coordinated during the transition of proplastids into chloroplasts is still not very well understood.

Various experimental systems have been pursued to investigate the biogenesis of photosynthetically active chloroplasts at the biochemical and /or gene regulatory level.

The study of chloroplast development in dicot leaves grown under natural conditions has proven to be difficult because meristematic activities occur in various leaf regions, and hence the various stages of plastid development cannot be easily separated for analysis (Donnelly *et al.,* 1999). Because of its easily experimental handling, the light-induced greening of an etiolated dicot, as well as of monocot leaves, has been used extensively as a substitute for studying chloroplast differentiation (Mullet, 1988).

In barley and maize, the leaf grows and expands in the dark, and the plastids turn into fully elaborated etioplasts (Klein and Mullet, 1987, Westhoff *et al.,* 1988). Etioplasts accumulate substantial amounts of photosynthetic proteins, such as the two subunits of cytochrome b-559 (psbE/F), the luminal subunits of the oxygen-evolving complex (PsbO, PsbP and PsbQ) (Müller and Eichacker., 1999), two polypeptides of the NDH complex (NDH-A and NDH-F) (Guera *et al.,* 2000) and two subunits of ribulose - 1.5-bisphosphate carboxylase/oxygenase (Rubisco) (Nelson *et al.,* 1984). In contrast, all subunits of photosystem I, the chlorophyll-containing subunit of photosystem II, i.e., D1 and D2, CP43 and CP47, as well as the subunits of the peripheral light-harvesting complexes of the two photosystems (LCHI and LCHII) are absent from etioplast (Klein and Mullet, 1986, 1990, Bedbrook *et al.,* 1978, Shen *et al.,* 2009). Light-induced greening, therefore, does not reflect the natural trajectory of chloroplast differentiation from proplastids into chloroplasts (Leech, 1984). Conversely, the development of proplastids into chloroplasts can be easily studied in grass leaves when exploiting the developmental gradient along the leaf blade. Only the cells at the leaf base divide, and their proplastids mature into chloroplasts along the longitudinal axis of the leaf towards the tip (Pogson *et al.,* 2015, Leech, 1984).

The developmental gradient of grass leaves has been used extensively to study the gene expression programs associated with chloroplast differentiation both in C_3 and C_4 photosynthesis (Li *et al.,* 2010, Wang *et al.,* 2014, Loudya *et al.,* 2021). Only recently has the transition of proplastids into chloroplasts been investigated within the shoot apical meristem of *Arabidopsis* at cellular resolution (Charuvi *et al.,* 2012)

Schrubar *et al.,* 1990, showed that the etiolation and de-etiolation of *Sorghum*

seedlings differed from that of barley and maize. Firstly, in contrast to barley or maize, the leave of *Sorghum* did not expand in the dark and remained confined to the coleoptile. Secondly, in *Sorghum*, chloroplast biogenesis is stalled at the proplastid or at least an early etioplast stage (depending upon the variety). Only trace amounts of photosynthetic proteins, therefore, accumulated in etiolated *Sorghum* seedlings, thus mimicking the situation in proplastids. Illumination of the etiolated seedlings initiated both leaf development and the build-up of the photosynthetic apparatus as proplastids/early etioplasts differentiate into chloroplasts.

We were interested in understanding the regulation of chloroplast biogenesis and therefore used transcriptome analyses of light-induced greening of etiolated *Sorghum* seedlings to get insight into the dynamic of photosynthetic gene expression and its integration into leaf development. Since *Sorghum* is a NADP malic enzyme type C_4 species (Kanai and Edward, 1999, Sage *et al.,* 1999), we expected that the information obtained could also be useful for C_4 engineering (Von Caemmerer *et al.,* 2012).

Material and methods

Seed growth, light treatment, sample collection, and phenotyping

Seeds from *Sorghum bicolor* L. Tx430 were soaked in water for 24hr at room temperature and planted in four trays (100 seeds per tray) containing soil (Floraton 1) in a dark chamber for seven days at a constant temperature of 25°C. After seven days, two trays containing etiolated seedlings were exposed to continuous light (9 am), and the other two remained in the dark chamber. Samples were harvested after 1h (10am), 2h (11am), 4h (1pm), 6h (3pm), 8h (5pm), and 24h (9am) of illumination and directly frozen in liquid nitrogen. In parallel, etiolated seedlings in the dark were also harvested simultaneously using green light for dark control. All samples were harvested in three biological replicates. The biological replicates were prepared by pooling three seedlings each to reduce variation between replicates and increase the statistical power of analysis. The harvested samples were stored at -80°C for further use.

For phenotyping and light microscopy, seedlings grown in a similar condition were

used. For light microscopy, seedlings were fixed in 3.5% glutaraldehyde for 2h (Haswell and Meyerowitz, 2006). After 2h the treated samples were rinsed with water, and cross-sections were made by hand with a sharp blade. The cross-sections were imaged under a light microscope for chlorophyll fluorescence using ZEISS light microscope Axio Imager.M2 connected to the digital camera AxioCamMR3.

RNA isolation and quality assessment

Total RNA was isolated from the harvested samples following the manufacturer's protocol (RNeasy Mini Handbook, QIAGEN GmbH) with the RNeasy plant mini kit. The concentration and quality of RNA (A260/280 ~ 2.0) were determined with the NanoDrop Spectrophotometer (PeQLaB Biotechnology ND1000). The integrity of RNA was tested in the Agilent 2100 Bioanalyzer, and RNAs with RNA Integrity Number (RIN>8) were used for library preparation. Samples showing DNA contamination were treated with DNase and retested for RNA integrity.

Library preparation, sequencing, and data analysis

The library preparation was performed following the protocol of the Illumina TruSeq RNA Sample Prep Kit V2. The quality and quantity of the libraries were assessed with the Agilent 2100 Bioanalyzer and the NanoDrop Spectrophotometer (PeQLaB Biotechnology ND1000). Single-end reads were sequenced with the HiSEQ3000 Illumina platform of the Biologisch-Medizinisches Forschungszentrum (BMFZ) of Heinrich Heine University. An average of 38 million reads were obtained per library. The quality of raw reads was assessed using the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were aligned against the transcriptome of *Sorghum bicolor* as available in the Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html) using the Bowtie2 tool (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). The differentially expressed genes were analyzed using EdgeR (Robinson *et al.,* 2010) as implemented in RStudio (*Version 3.3.3; https://cran.r-project.org/bin/macosx/*). EdgeR generated expression value fold changes, p-values, and false discovery rate corrected p values from read counts. The false discovery rates in EdgeR were calculated using the Benjamini-Yekutieli correction (Reiner *et al.,* 2003). The quality of RNA was assessed using principal component analysis (PCA) on the MultiExperiment Viewer (MeV) platform

(https://bioinformaticshome.com/tools/rna-seq/descriptions/MeV.html) and PCAGO (Gerst and Hölzer, 2019; https://pcago.bioinf.uni-jena.de). All differentially expressed genes (DEGs) were transported into an Excel sheet and were annotated against *Arabidopsis* ids using the VLOOKUP function. A list of transcripts showing log_2fold changes >2 or log_2fold changes < -2 and false discovery rate < 0.05 were classified as DEGs, DEGs were clustered using K-means and Hierarchical clustering (HCL) as implemented in the MultiExperiment Viewer platform. The DEGs were categorized with MapMan software (Thimm *et al.,* 2004).

Results

Morphological analysis of light-induced leaf development in *Sorghum bicolor*

Sorghum bicolor L. Tx430 were grown in the dark for seven days to investigate light-induced leaf development. No leaf expansion was observed in the dark, but a long mesocotyl was formed. Seedlings were pale, and the leaf remained enclosed within the coleoptile. After 4h of illumination, the leaves emerged from the coleoptile and were fully expended after 24h of illumination (Figure 1A). In order to visualize the greening process more precisely, leaf cross-sections were prepared from various time points of illumination and examined with a fluorescence microscope. Figure 1B shows that chlorophyll accumulation was clearly detectable after 2 h of illumination.

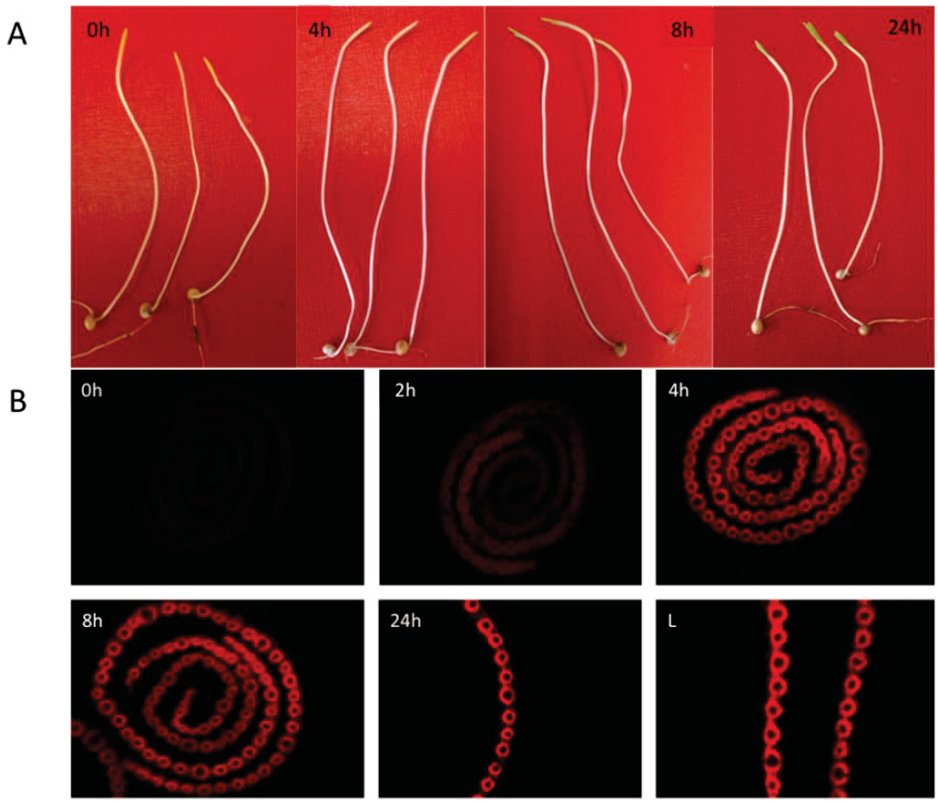


Figure 1: Leaf development and imaging of chlorophyll fluorescence in etiolated and illuminated *Sorghum* seedlings.

(**A)** Etiolated (0h) and illuminated seedling (4, 8 and 24h) of *S. bicolor L.* Tx430. (**B**) Images of chlorophyll fluorescence in etiolated (0h) and de-etiolated seedlings after 4, 6, 8, and 24 h of illumination. L represents a picture taken from a light-grown seedling without prior etiolation.

Design of transcriptome analysis of etiolated and de-etiolated *Sorghum* seedlings

Seven days old dark-grown *S. bicolor* L. Tx430 seedlings were exposed to the light, and samples were harvested after 1h, 2h, 4h, 6h, 8h, and 24h for the analysis of transcriptome profile using RNA-seq technology. In parallel, samples were also collected from seedlings that continued growing in the dark at the same time points (Figure 2). RNA-seq reads were generated from three independent biological replicates from each time point. The sequences were aligned to the *Sbicolor-v3.1* reference sequence (Phytozome V11) comprising 47,205 protein-coding transcripts, and the data were analyzed as outlined in figure 2 and described in Material and methods.

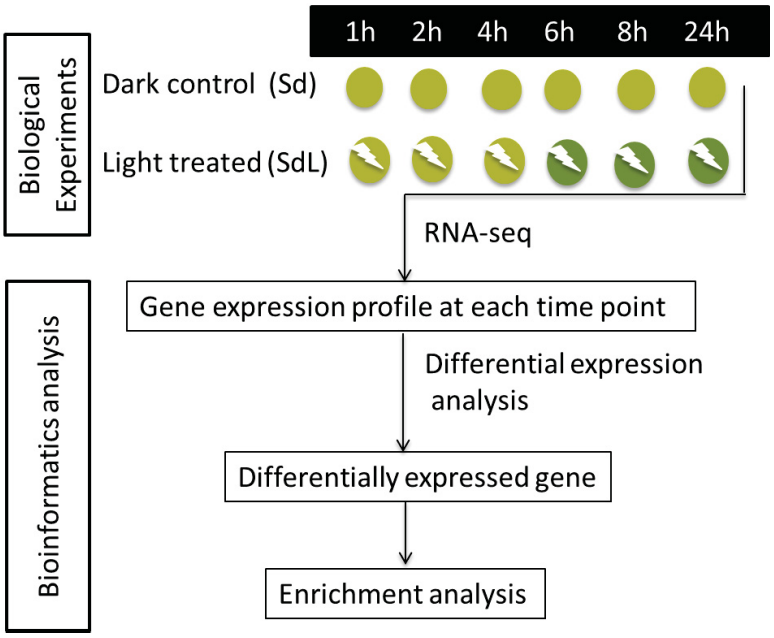


Figure 2: Schematic diagram of the experimental design of etiolated and de-etiolated seedlings at various time points were subjected to RNA sequencing. The Differentially expressed genes (DEGs) were identified by comparing illuminated samples (SdL) with dark-grown samples (Sd) at each time point, followed visualization of DEGs via heatmap in MEV.

The two-dimensional PCA plot visualizes the relevance of biological data and the similarity between biological replicates. Such as, the dark control groups appeared separately on one axis and showed a slight difference in each time point, while samples within each illuminated group of 1, 2, 6, 8, and 24 h aggregated, respectively, and separated from one another moving away from dark samples (Figure S1).

An overview of differentially expressed genes

The differential expression analysis of each group was carried out between dark control (Sd) and de-etiolated seedling (SdL) at each time point using EdgeR in Rstudio v.3.3.3. We used pairwise comparison (i.e., 1h Sd versus 1h SdL), as shown in figure 2, to identify differentially expressed genes (DEGs) at each time point. We considered differentially expressed genes based on FDR < 0.05, $Log_2FC>2$ for significantly up-regulated and Log2FC <-2 for down-regulated genes at each time point compared to the dark control. In total, 4742 DEGs were identified for at least one-time point. In detail, there were 281, 694, 1258, 1279, 1800, and 1921 DEGs for 1, 2, 4, 6, 8 and 24h, respectively. That indicates more genes are differentially expressed with the time of illumination.

Genes encoding components of the thylakoid membrane

Chloroplast development requires a coordinated expression of both nuclear and plastid genes to achieve a stoichiometric accumulation of polypeptides for photosynthetic enzymes or membrane complexes, respectively. Most of the enzymes involved in the light-independent reaction and almost half of the proteins involved in a light reaction are nuclear-encoded and require a specialized transit peptide for targeting chloroplasts (Tyagi and Gaur, 2003, Piechulla *et al.,* 1986). During etioplast to chloroplast conversion, one of the critical steps is the initiation of transcription of photosynthetic genes. We investigated the expression of genes encoding the component building photosynthetic machinery. We generated a heatmap to visualize the expression level over the period of chloroplast development. The expression profile of large numbers of nuclear-encoded chloroplast genes associated with PSII, cytochrome b6f, electron transport, PSI, and ATP synthase display strong induction after 2 to 4h of illumination when compared to the etiolated sample. Such as light-harvesting complexes (LCH1 and LCHII), *PetM* of cytochrome b6f*, PGR5L, AtpC* and *AtpD* of ATP complex and genes encoding components of oxygen-evolving complex (*PsbO, PsbP* and *PsbQ*) are upregulated after 4h of light exposure. Following the strong induction after 4h, most of the genes related to the light-harvesting complex and PSII and PSI subunits showed circadian regulation of gene expression (Figure 3B). In addition, the accumulation of PSII subunit *PsbS* and, *PetB* and *PetC* of cytochrome b6f complex has seen after 6h followed by the expression of NDH complex *(NdhM, NdhN* and *NdhO*) after 8h of light exposure. In contrast to the nuclear encoded-chloroplast genes, the accumulation of plastid-encoded genes of photosynthetic apparatus during 24 of illumination except for two subunits of PSII (*PsbB* and *PsbH*), whose expression was observed after 6h of illumination (Figure 3A and B). These results suggest that the accumulation of plastid-encoded genes requires prolonged exposure to light.

Besides nuclear genes encoding structural components of the photosystems, the nuclear genes encoding photosystem assembly factors also showed a light-induced expression profile. Notably, PSII assembly genes *HCF136* (High chlorophyll fluorescence 136) and *HCF244* were observed after 2hr of illumination. Both Hcf136 and HCF244 are involved in the assembly of precursor D1 and D2 precomplex (Plöchinger *et al.,* 2016). Similarly, the *PSB28* and *LPA3* are involved in the assembly

of CP47 and CP43, respectively (Lu Y., 2016), and PSI assembly gene *HCF101* and *ALB3* showed induction between 6-8h of illumination (Figure 3A).

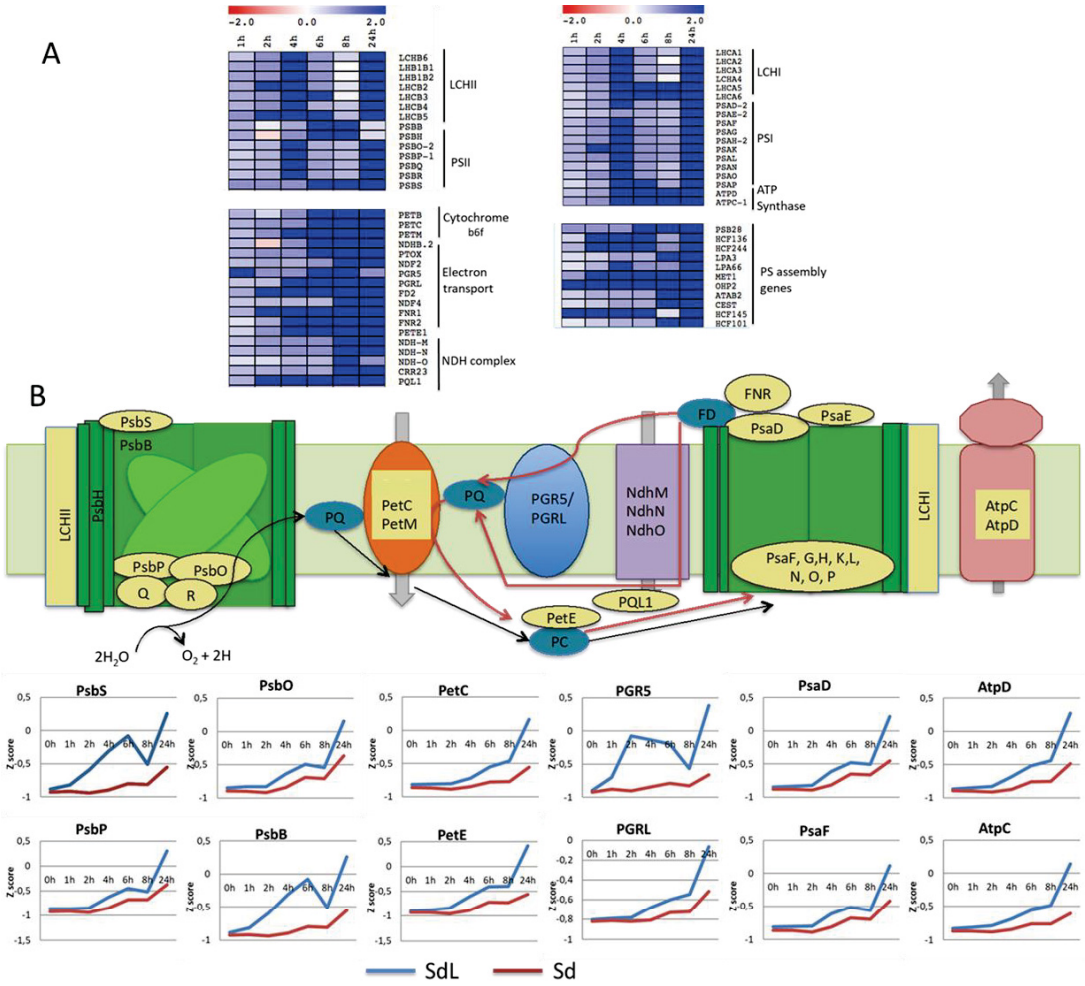


Figure 3. Detailed expression profile of genes involved in the photosynthetic machinery. (A) Heatmaps illustrating the relative expression of gene encoding light-harvesting complexes (LHCI and LHCII), PSII and PSI core proteins, Cytochrome b6f complex, cyclic electron transport chain, NDH complex, ATP synthase and photosystem assembly. Blue and red colors depict up-and downregulation relative to the dark control, respectively. The scale represents log_2fold changes. (B) Schematic representation of signature gene in photosynthetic apparatus that showed differential expression between etiolated and illuminated samples at various time points and the expression pattern of signature photosynthetic genes are shown in the graph. The graphs were generated using z-score values.

Calvin-Benson cycle and photorespiratory genes

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) is a major enzyme in the Calvin cycle that catalyzes the first step of carbon fixation. Here we investigated the

transcript abundance of a Rubisco large subunit (*RBCL*), Rubisco small subunit (*RBCS*) and Rubisco activase (*RCA*). The transcript level of these genes, particularly *RCA* and *RBCS,* increased in abundance by many folds after 2h and 4h of exposure, respectively. In contrast, the *RBCL* did not show differential expression during de-etiolation (Figure 4B). Besides that, several genes encoding key enzymes of Calvin cycle and photorespiration were observed after 4 to 8h of light exposure, including phosphoglycerate kinase (*PGK1*), phosphoribulokinase (*PRK*), glycerate kinase (*GLYK*), phosphoglycolate phosphatase (*PGLP*), hydroxypyruvate (*HPR1*) and glyceraldehyde-3-phosphate dehydrogenase a (*GAPA*). Overall, these results suggested that the C_4 photosynthetic pathway is strictly regulated by light via the modulation of the expression of these key enzymes (Figure 4B).

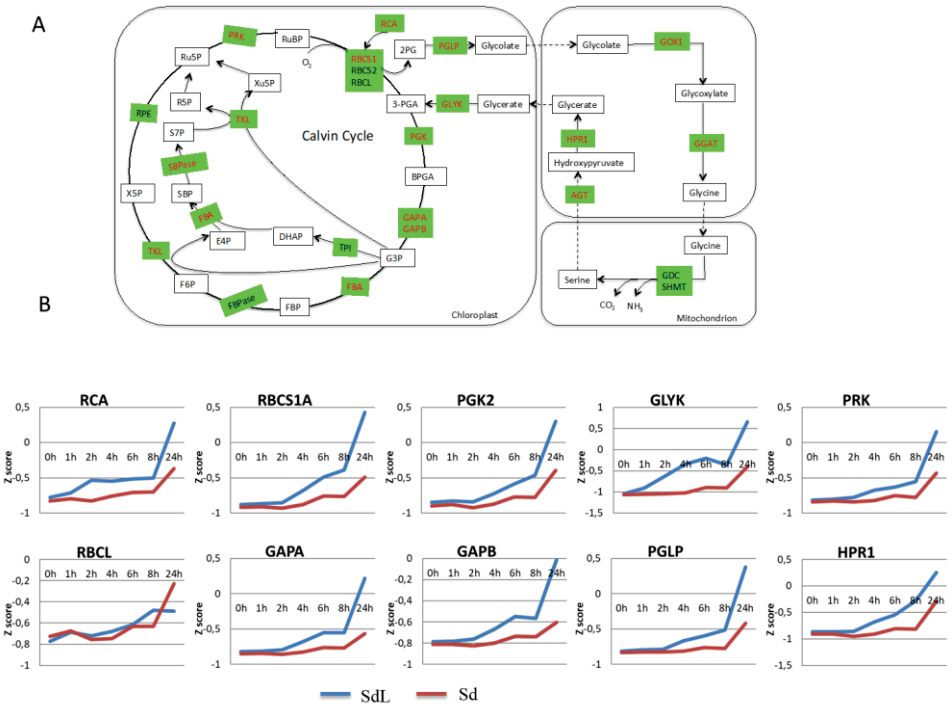


Figure 4. Detailed expression profile of genes involved in Calvin-Benson Bessham cycle and Photorespiration.

Schematic representation of Calvin cycle and photorespiratory pathway. DEGs involved in the pathways are marked in red. *RCA*- Rubisco activase; *RBCS1A* - Ribulose bisphosphate carboxylase small chain 1a; *PGK* - Phosphoglycerate kinase; *GAPA/B* - Glyceraldehyde-3-phosphate dehydrogenase a/b subunit; *FBA* - Fructose bisphosphate aldolase; *FBPASE* - Fructose-1, 6-bisphosphatase; *SBPASE* - Sedoheptulose-1, 7-bisphosphatase; *RPI* - ribose-5-p isomerase; *PRK* - Phosphoribulokinase; *PGLP* - Phosphoglycolate phosphatase; *GOX* - Glycolate oxidase; *GLYK* - Glycerate kinase; *GGAT* – Glutamate glyoxylate

aminotransferase; *HPR* - Hydroxypyruvate reductase; *AGT* – Alanine glyoxylate aminotransferase. Relative transcript levels of selected DEGs genes in Sd and SdL are visualized using a z score plot.

C_4 cycle genes

Sorghum has been recognized as NADP-ME subtype C_4 plant (Rao and Dixon, 2016). We identified the expression of key enzymes involved in the C_4 cycle. Figure 5 shows changes in the transcript abundance of genes involved in initial carbon assimilation during the greening of etiolated *Sorghum*. In the C_4 photosynthetic reaction, the CO_2 is initially fixed into HCO3 by carbonic anhydrase (CA) in the mesophyll cells. The HCO3 is subsequently fixed via phosphoenolpyruvate carboxylase (PEPC) and NADP-malic dehydrogenase (NADP-MDH) into malate that diffuses to the bundle sheath cells for decarboxylation. The NADP- dependent malic enzyme catalyzes the four-carbon acid malate conversion to pyruvate in bundle sheath cells. Pyruvate orthophosphate dikinase (PPDK) catalyzes the interconversion of pyruvate into phosphoenolpyruvate (PEP) in mesophyll chloroplast also showed light-dependent expression (Hatch and Burnell, 1990, Rao and Dixon, 2016). An increase in the transcript abundance of the majority of enzymes, including *PEPC2, MDH, ME4* and *PPDK* has been observed after 4h. Notably, *CA* showed elevated transcript abundance after 2h of illumination (Figure 5B).

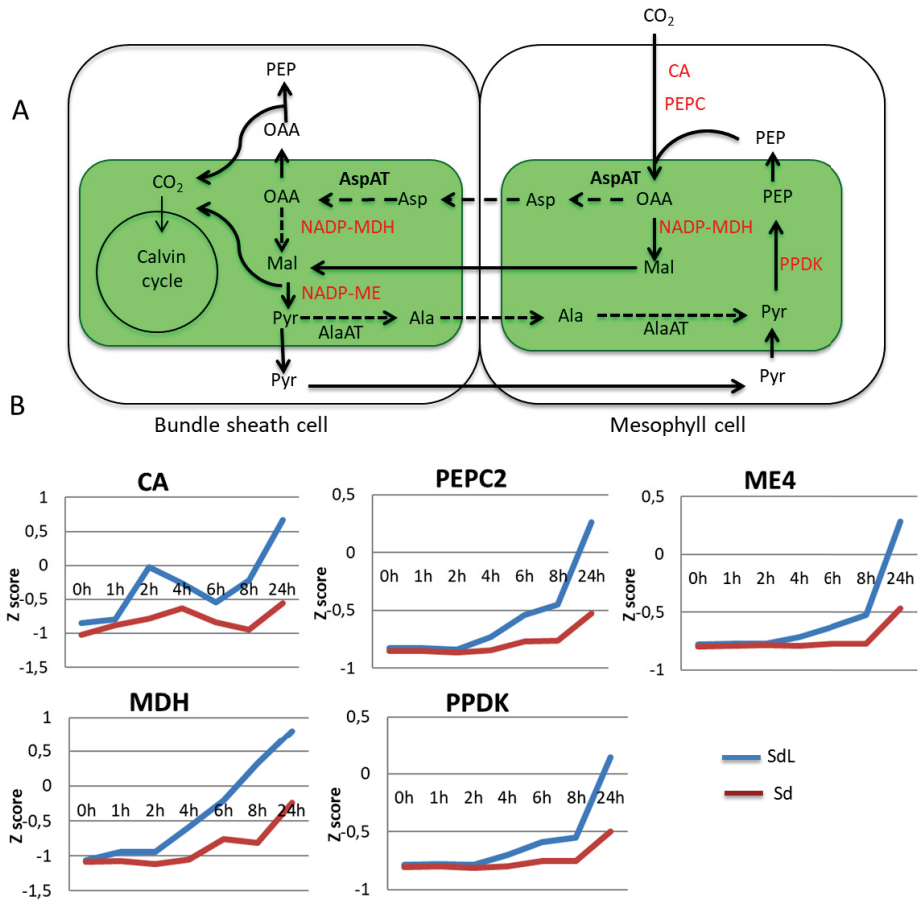


Figure 5. Detailed expression profile of genes involved in the C4 pathway

(A) Schematic view of the NADP-ME type C_4 pathway. DEGs are marked in red. *CA* - Carbonic anhydrase; *PEPC* - Phosphoenolpyruvate carboxylase; *PPDK* - Pyruvate orthophosphate dikinase; *NADP-ME* – NADP dependent malic enzyme; *NADP-MDH* - NADP dependent malate dehydrogenase; **(B)** Relative transcript levels of selected DEGs genes in Sd and SdL are visualized using a z score plot.

Chloroplast located/associated metabolic pathways

The chloroplast is the site of various metabolic reactions, including the biosynthesis of amino acid, starch, lipid and fatty acid, hormone metabolism, as well as the reduction of nitrate and sulfur also take place in the chloroplast. Light plays an essential role in the accumulation of these metabolites in plants (Chen *et al.,* 2018). Here we investigated the expression of genes encoding chloroplast-associated metabolites.

Carbohydrate metabolism

During the daytime, carbohydrate is assimilated in the chloroplast, with some stored in the form of starch or degraded at night to supply sugar for growth. Various enzymes have been involved in the synthesis and degradation of starch, such ADP-

glucose pyrophosphorylase (AGPase), starch synthase (SS), β-amylase (BAM1 and BAM3) (Skryhan *et al.,* 2018, Ball and Deschamp 2009). *BAM1* and *BAM3* showed differential expression after 1h of illumination, while APGase large subunit *APL2* and *APL1*, and APGase small subunit *APS1* (Orzechowski, 2008) upregulated after 4, 8 and 24h, respectively (Figure 6).

Amino acid and nitrogen metabolism

Chloroplast has an impact on primary metabolisms, such as syntheses of phenylamine, lysine and tryptophane (Kretschmer *et al.,* 2017). Our RNA-Seq analysis revealed substantial changes in the levels of transcripts for enzymes involved in amino acid biosynthesis during light-induced chloroplast development. These enzymes include arogenate dehydratase (ADT6), meso-diaminopimelate decarboxylase (DAPDC1 and DAPDC2), and 3-phosphoglycerate dehydrogenase (PGDH) (Figure 6). Among them, ADT6 is involved in L phenylalanine biosynthesis (Höhner *et al.,* 2018), DAPDC1 and DAPDC are involved in lysine biosynthesis (Crowther *et al.,* 2019), and 3- PGDH is involved in serine biosynthesis (Reyes-Prieto and Moustafa, 2012). Besides that, the gene involved in nitrogen metabolism, nitrate reductase 1 (*NR1*) (Olas and Wahl, 2019) expressed after 2h of illumination (Figure 6).

Hormone metabolism

Plant hormones, such as auxin, gibberellic acid (GA), ethylene and brassinolide, play an important role in light-regulated development (Chen *et al.,* 2018, Cackett *et al.,* 2022). Abscisic acid (ABA) is involved in seed development and plays an essential role in stress response. The biosynthesis of ABA and GA is closely associated with plastids (Yamburenko *et al.,* 2013, Cackett *et al.,* 2022). Gene involved in ABA biosynthesis, including nine-cis-epoxycarotenoid dioxygenase (*NCED9*), aldehyde oxidase isoenzymes (*ABA1*) and carotenoid cleavage dioxygenase 8 (*CCD8*) (Iuchi *et al.,* 2001, Sano and Marion-Poll, 2021) are upregulated after 1-2h of illumination (Figure 6).

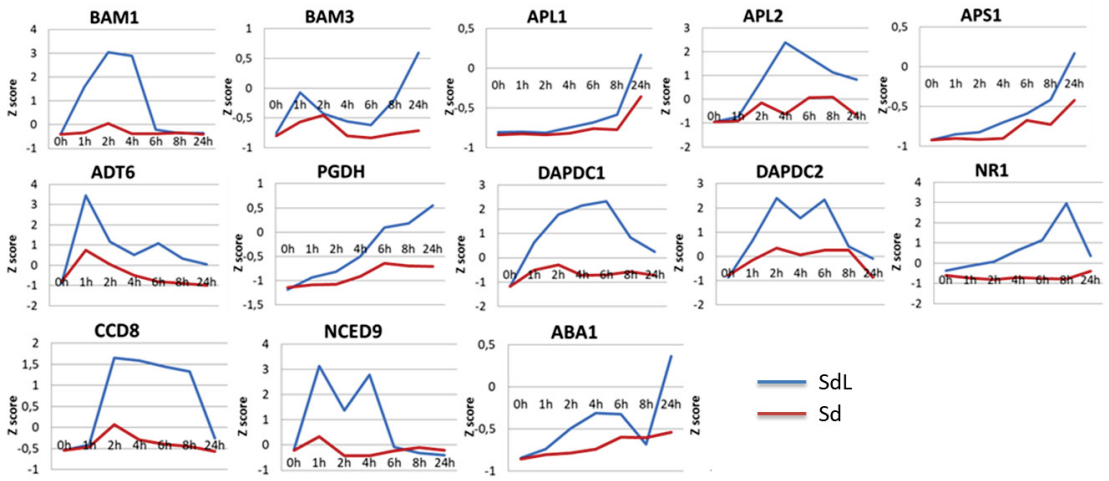


Figure 6. Expression pattern of genes encoding chloroplast located and/or associated metabolites.

APL1/2 - ADP-glucose pyrophosphorylase large subunit; *APS1*- ADP-glucose pyrophosphorylase small subunit; *BAM1*/3 - β-amylase; *ADT6* - Arogenate dehydratase; *DAPDC1/2* - Meso-diaminopimelate decarboxylase; *PGDH* – 3 -phosphoglycerate dehydrogenase; *NR1* - Nitrate reductase; *NCED9* - Nine-cis-epoxycarotenoid dioxygenase; *ABA1* - Aldehyde oxidase isoenzymes; *CCD8* - Carotenoid cleavage dioxygenase 8; *GA2* - GA requiring 2. The graphs were made from z-score values.

Genes with putative regulatory function

We have identified several regulators of chloroplast development and/or biogenesis that are differentially expressed during light-induced chloroplast development. These regulators include light signaling cascade gene ELONGATED HYPOCOTYL 5 (HY5) and PHYTOCHROME INTERACTING FACTORS (PIFs). HY5 is a class B GATA transcription factor that promotes photomorphogenesis in light (Xiao *et al.,* 2021, Waters and Langdale, 2009), while PIFs are a member of a subfamily of bHLH transcription factors that acts as constitutive repressors of photomorphogenesis in the dark (Leiver *et al.,* 2008, Dong *et al.,* 2017). *HY5* expression displayed a strong induction after 2h of light exposure and then the expression level declined. Conversely, *PIF3* and *PIF4* are downregulated after exposure, as expected (Figure 7). GOLDEN2-LIKE transcription factors GLK1 and GLK2, positively regulate chloroplast biogenesis (Hall *et al.,* 1998; Fitter *et al.,* 2002, Water *et al.,* 2009), and SIGMA FACTORS (SIGs), involved in the transcription regulation of plastid-encoded genes (Puthiyaveetil *et al.,* 2021). The expression of *GLK1* showed strong induction after 24h, and *SIG1, SIG2, SIG3, SIG5* and *SIG6* peaked in expression after

2-4h of illumination. Additionally, VARIGATED 2 (VAR2/FTSH2), which is responsible for the degradation of the D1 protein of the photosystem II (PSII) reaction center during its repair from photoinhibition (Zaltsman *et al.,* 2005) showed strong induction after 2h of illumination. In addition, genes encoding components of the chlorophyll biosynthesis pathway, in particular glutamyl-tRNA reductase (*HEMA1*), a subunit of MG-chelatase (*CHLH/GUN5*), genome uncoupled 4 *(GUN4),* and chlorophyll a oxygenase (*CAO*) were found to be upregulated after 8h (Figure 7). HEMA1 catalyzes the first step in tetrapyrrole biosynthesis, CHLH/GUN5 diverts tetrapyrroles toward chlorophyll biosynthesis, GUN4 is required for efficient Mg-chelatase activity, and CAO is involved in the import of LHCB monomers into isolated chloroplasts and catalyzes the conversion of chlorophyllide *a* to chlorophyllide *b* (Water *et al.,* 2009). *GUN4* and *GUN5* also participate in plastid-to-nucleus retrograde signaling (Liu *et al.,* 2020, Adhikari *et al.,* 2011, Mochizuki *et al.,* 2001).

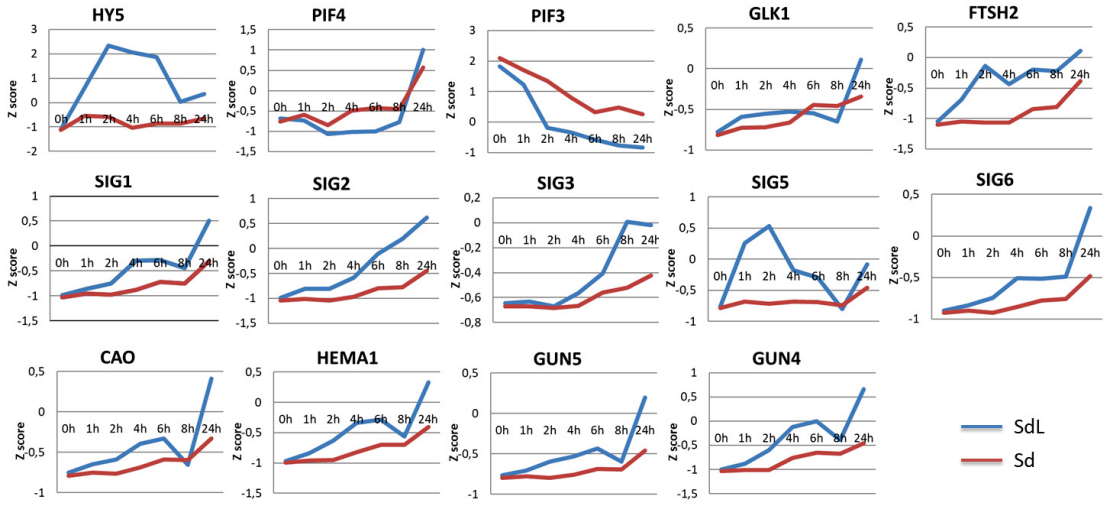


Figure 7. Expression of genes and transcription factors associated with chloroplast biogenesis. *GLK1* - Golden2-like 1*; HY5* - Elongated hypocotyl 5*; PIF3/4* - Pytochrome interacting factor genes*; SIG1/2/3/5/6* - Sigma transcription factors*; HEMA1* - Glutamyl-tRNA reductase; *GUN4/5* - Genome uncoupled*; CAO* - Childe a oxygenase. The graphs were made from z-score values.

Chloroplast translational apparatus

One of the crucial aspects of chloroplast development is the accumulation of ribosomes which is essential for protein synthesis. Plastid ribosomes constitute 70S ribosomes that comprise small (30S; 16 rRNA and 24 plastid ribosomal proteins) and

large (50S; 23S, 5S, 4.5S rRNA and 33 plastid ribosomal proteins) ribosomal subunit. All four rRNAs are plastid-encoded, while the majority of the plastid ribosomal proteins (RPs) are nuclear-encoded (Zhang *et al.,* 2016). Like chloroplast genomes, numbers of chloroplast ribosomes were present in very low amounts in the early stage of exposure but peaked in expression after 8h of illumination (Figure 8).

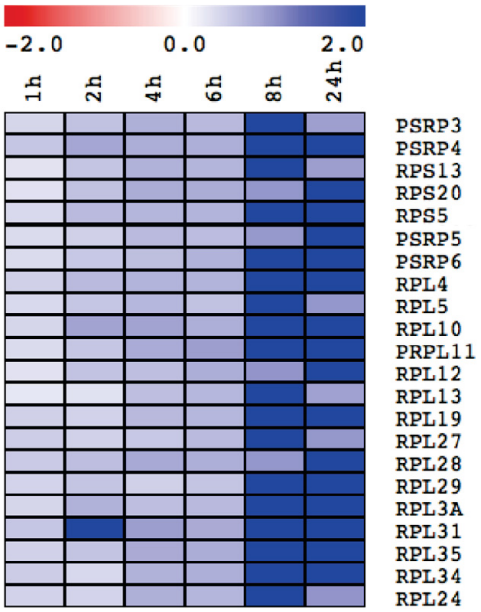


Figure 8. Heatmap shows the expression profile of genes associated with the chloroplast translational apparatus 50S and 30S ribosomal subunits. Blue and red colors depict up- and downregulation relative to the dark control, respectively. The scale represents log2fold changes.

Discussion

This study aimed to understand how and to what extent the biogenesis of the photosynthetic apparatus, and hence chloroplast biogenesis, is intertwined with leaf development. We suspected that the developmental program of the leaf essentially controls chloroplast biogenesis. The developmental gradient of grass leaves provides a good system for investigating the development of proplastids into chloroplasts. Various transcriptome studies along the leaf gradient of maize and rice have uncovered a few potential regulators of photosynthetic development (Wang *et al.,* 2014, Pick *et al.,* 2011). The standard etiolation/de-etiolation systems of dicot plants (Arabidopsis, spinach, etc.) or grasses, such as barley and maize, have been

extensively used for studying chloroplast differentiation. However, in these species, the plastids turn into fully elaborated etioplasts in the dark and, at least in barley and maize, the leaves fully expand in darkness, and photosynthetic proteins already accumulate. For instance, Rubisco and the thylakoid membrane protein complexes of ATP synthase and the cytochrome b6f entity and oxygen-evolving complex (Klein and Mullet, 1987, Müller and Eichacker, 1999, Nechushtai and Nelson, 1985, Nelson *et al.,* 1984). Therefore, the etiolation/de-etiolation systems in these species are not suitable for studying the dependencies of the differentiation programs of leaves and chloroplasts. Schrubar *et al.,* 1990 showed that the etiolation/de-etiolation in *Sorghum* seedlings are different to the above-mentioned species, i.e., leaves of *Sorghum* seedlings do not expand in the dark, and their development is drastically stalled. This means *Sorghum* leaf development needs light. Additionally, in *Sorghum*, chloroplast biogenesis is stalled at the proplastid or at least an early etioplast. Schrubar *et al.,* 1990 also showed that dark-grown *Sorghum* shows only traces of photosynthetic proteins. Their accumulation requires light, suggesting that *Sorghum* leaf development and chloroplast biogenesis is intimately intertwined and can be coordinately induced by light.

Based on these findings, we assumed that the light-induced greening of *Sorghum* seedlings might give insight into on how and which timely order the various components of the photosynthetic apparatus located in the thylakoid membrane, i.e., antenna complex, PSI, PSII, ATP synthase, the cytb6f complex and the NDH complex as well as the stromal components, i.e., Rubisco and the Calvin-Benson cycle enzymes are building up. Which regulators/transcription factors might trigger the expression of these genes? Therefore, in the present study, we performed a comparative transcriptome investigation during *Sorghum* seedling de-etiolation by using light as a trigger of leaf development and the transition of proplastids/early etioplasts into a chloroplast. By comparing gene expression between seven days old etiolated and illuminated seedlings after 1h, 2h, 4h, 6h, 8h, and 24h, we identified 4742 differentially expressed genes (DEGs) in at least one time point. Among them, a large number of genes involved in photomorphogenesis and chloroplast development were revealed. That indicates that the presence or absence of light dramatically affects chloroplast development in *Sorghum*. This finding is consistent with the fact that proteins involved in chlorophyll synthesis and photosynthesis, especially the

photosynthetic apparatus, must synthesize and accumulate rapidly as the etioplasts turn into chloroplasts.

Chloroplast biogenesis and development required coordinated activity of plastid- and nuclear-encoded genes to achieve stoichiometric accumulation of polypeptides for photosynthetic enzymes or membrane complexes. In the first phase of chloroplast development, light triggers an increased expression of the photosynthesis-associated nuclear genes (Dubreuil *et al.,* 2018; Yang *et al.,* 2019, Hernández *et al.,* 2022). This initial response is assumed to be entirely under nuclear control, but the second phase of development required to establish fully photosynthetically active chloroplasts relies upon a retrograde signal originating in the plastids (Dubreuil *et al.,* 2018). Consistent with the fact the nuclear-encoded genes of the photosystem associated with the reaction centers, antenna complex or electron transport showed their peak expression after 4h (Figure 4A). The PS assembly genes, including *HCF136*, *HCF244*, *OHP2* and *HCF145* (Plöchinger *et al.,* 2016, Li *et al.,* 2019), also upregulated soon after illumination (Figure 4). Most genes involved in the Calvin cycle, photorespiration and C_4 cycle genes showed a similar pattern of expression, consistent with shared regulation, except *RBCL* of the Calvin cycle, whose transcript abundance did not change significantly during 24h of illumination (Figure 5B and 6B). Contrary to the nuclear-encoded genes, plastid-encoded photosynthetic genes were not detectable in the dark or during 24 hours of illumination.

Two classes of transcription factors and signal transduction cascade, HY5 and PIFs, regulate chlorophyll biosynthesis in opposing manners. PIFs accumulate in dark-grown seedlings and negatively regulate chlorophyll biosynthesis and photosynthetic genes. Namely, PIF3 is involved in the regulation of HEMA, CHLH and CAO during photomorphogenesis, as *pif3* mutant line showed strong induction of these genes (Stephenson *et al.,* 2009), while HY5 promote photomorphogenic development and responsive gene expression (Liu *et al.,* 2020, Zhou *et al.,* 2011). We observed strong induction of *HY5* after 1h of illumination. Conversely, *PIF3* and *PIF4* are strongly downregulated after light exposure, indicating the onset of photomorphogenesis.

GLK transcription factor that promotes the formation of stable LHCII-PSII super complexes and upregulation of components of the chlorophyll pathway, particularly *HEMA1, CHLH*, and *CAO.* The expression of these genes increased significantly during the early hours of illumination; surprisingly, the strong induction of *GLK1* was

observed only after 24h in the light. Plastidic SIGMA FACTORS control the initiation of PEP-mediated transcription of chloroplast genes (Che *et al.,* 2015). SIG1 has been shown to bind to the *PsbA* and *RBCL* promoter in vitro (Privat *et al.,* 2003), and rice *sig1* knockout line results decrease *PsaA* transcript accumulation (Tozawa *et al.,* 2007). SIG2 and SIG6 appeared to be involved in the phytochrome-dependent coordinated control of plastid gene expression in anterograde signaling during photomorphogenesis. However, they also showed a partially redundant role in retrograde signaling to control nuclear gene expression, mimicking dual function in plastid to nucleus communication (Che *et al.,* 2015). The expression of *SIG1-SIG6* significantly upregulated after 2-4h of illumination (Figure 7).

Plastid ribosomes play an important role in plastid development and differentiation. We observed a substantial number of genes encoding ribosomal protein had been augmented after 8h, including RPS5, which is involved in the accumulation and processing of chloroplast 16S rRNA. Comparative proteome analysis showed that RPS5 also affect the accumulation of many PSI and II proteins and plastid RPs (Zhang *et al.,* 2016). Disruption of PRSP20, RPL12, and RPL13 function in rice lead to suppression of chloroplast development (Gong *et al.,* 2013, Zhao *et al.,* 2016, Song *et al.,* 2014). Mutation in RPL11 and L24 in *Arabidopsis* results decrease in the synthesis of plastid proteins and photosynthesis (Pesaresi *et al.,* 2001, Romani *et al.,* 2012). These observations suggest the two distinct regulatory phases of the establishment of photosynthesis during chloroplast development, and a major part of early chloroplast development begins after 4h of illumination in the developing leaves.

In conclusion, our RNA sequencing approach provided precise and reproducible quantification of genes during the de-etiolation of *Sorghum* seedlings. This technique lets us demonstrate light's effect on the synthesis of the chloroplast transcript during light-induced chloroplast biogenesis. We found a correlation between leaf development and chloroplast biogenesis in *Sorghum*. All photosynthetic-related genes appeared to be strictly light-dependent. Moreover, genes involved in various pathways do not accumulate sequentially but rather coordinately, i.e., light-harvesting complex or ATP synthase gene, *PEPC* and *RBCS* expressed simultaneously. In addition, nuclear-encoded genes expressed at early exposure, while plastid-encoded genes required a longer exposure time.

Acknowledgement

We thank the Genomics and Transcriptomics laboratory of the 'Biologisch-Medizinisches Forschungszentrum' (BMFZ) at the Heinrich-Heine-University of Duesseldorf, for conducting Illumina sequencing.

References

Adhikari, N. D., Froehlich J. E., Strand D. D., Buck S. M., Kramer D. M., and Larkin R. M. (2011). "GUN4-porphyrin complexes bind the ChlH/GUN5 subunit of Mg-Chelatase and promote chlorophyll biosynthesis in *Arabidopsis*." *The Plant Cell,* 23, 1449-1467.

Ball, S. G., and Deschamps, P. (2009). Starch metabolism. *In The Chlamydomonas Sourcebook* (pp. 1-40), *Elsevier*.

Bedbrook, J. R., Link G., Coen D. M., and Bogorad, L. (1978). Maize plastid gene expressed during photoregulated development. *Proceedings of the National Academy of Sciences, 75*, 3060-3064.

Cackett, L., Luginbuehl, L. H., Schreier, T. B., Lopez-Juez, E., and Hibberd, J. M. (2022). Chloroplast development in green plant tissues: the interplay between light, hormone, and transcriptional regulation. *New Phytologist,* 233, 2000-2016.

Chi, W., He, B., Mao, J., Jiang, J., and Zhang, L. (2015). Plastid sigma factors: their individual functions and regulation in transcription. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, *1847*, 770-778.

Chen, Y., Zhou, B., Li, J., Tang, H., Tang, J., and Yang, Z. (2018). Formation and change of chloroplast-located plant metabolites in response to light conditions. *International Journal of Molecular Sciences*, 19, 654.

Charuvi D., Kiss V., Nevo R., Shimoni E., Adam Z., and Reich Z. (2012). Gain and Loss of Photosynthetic Membranes during Plastid Differentiation in the Shoot Apex of *Arabidopsis*. *Plant Cell,* 24, 1143–1157.

Crowther, J. M., Cross, P. J., Oliver, M. R., Leeman, M. M., Bartl, A. J., Weatherhead, A. W., . . . Suzuki, H. (2019). Structure–function analyses of two plant meso-diaminopimelate decarboxylase isoforms reveal that active-

site gating provides stereochemical control. *Journal of Biological Chemistry*, 294, 8505-8515.

Dalal, V., Dagan, S., Friedlander, G., Aviv, E., Bock, R., Charuvi, D., . . . Adam, Z. (2018). Transcriptome analysis highlights nuclear control of chloroplast development in the shoot apex. *Scientific Reports, 8*, 1-9.

Dong, J., Ni, W., Yu, R., Deng, X. W., Chen, H., and Wei, N. (2017). Light-dependent degradation of PIF3 by SCFEBF1/2 promotes a photomorphogenic response in *Arabidopsis*. *Current Biology, 27*, 2420-2430. e2426.

Donnelly PM, Bonetta D, Tsukaya H, Dengler RE & Dengler NG (1999). Cell cycling and cell enlargement in developing leaves of *Arabidopsis*. *Developmentl Biology,* 215, 407-419.

Dubreuil, C., Jin, X., Barajas-López, J. d. D., Hewitt, T. C., Tanz, S. K., Dobrenel, T., . . . Grönlund, A. (2018). Establishment of photosynthesis through chloroplast development is controlled by two distinct regulatory phases. *Plant Physiology, 176*, 1199-1214.

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. *The Plant Journal, 31*, 713-727.

Gerst, R., and Hölzer, M. (2019). PCAGO: An interactive tool to analyze RNA-Seq data with principal component analysis. *BioRxiv*, 433078.

Gong, X., Jiang, Q., Xu, J., Zhang, J., Teng, S., Lin, D., and Dong, Y. (2013). Disruption of the rice plastid ribosomal protein S20 leads to chloroplast developmental defects and seedling lethality. *G3: Genes, Genomes, Genetics*, *3*, 1769-1777.

Guera, A., de Nova, P. G., and Sabater, B. (2000). Identification of the Ndh (NAD (P) H-plastoquinone-oxidoreductase) complex in etioplast membranes of barley: changes during photomorphogenesis of chloroplasts. *Plant and Cell Physiology, 41*, 49-59.

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. *The Plant Cell, 10*, 925-936.

Haswell, E. S., and Meyerowitz, E. M. (2006). MscS-like proteins control plastid size and shape in *Arabidopsis thaliana*. *Current Biology, 16*, 1-11.

Hatch, M. D., and Burnell, J. N. (1990). Carbonic anhydrase activity in leaves and its role in the first step of C4 photosynthesis. *Plant Physiology, 93*, 825-828.

Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P., and Minghim, R. (2015). InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics, 16*, 1-7.

Hernández, V., Tamara, L Vuorijoki, X. Jin, A. Vergara, C. Dubreuil, A. (2022) Strand. GENOMES UNCOUPLED1 plays a key role during the de-etiolation process in *Arabidopsis*. *New Phytologist*.

Höhner, R., Marques, J. V., Ito, T., Amakura, Y., Budgeon Jr, A. D., Weitz, K., . . . Lewis, N. G. (2018). Reduced arogenate dehydratase expression: ramifications for photosynthesis and metabolism. *Plant Physiology*, 177, 115-131.

Iuchi, S., Kobayashi, M., Taji, T., Naramoto, M., Seki, M., Kato, T., . . . Shinozaki, K. (2001). Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in *Arabidopsis*. *The Plant Journal*, 27, 325-333.

Jarvis P., and Lopez-Juez E. (2013). Biogenesis and homeostasis of chloroplasts and other plastids. *Nature Reviews Molecular Cell Biolog,* 14, 787-802.

Kanai R., and Edwards G.E. (1999) The biochemistry of C4 photosynthesis. In *C4 Plant Biology* (Sage, R.F. and Monson, R.K. eds). *San Diego: Academic Press*, pp. 49-87.

Klein, R., and Mullet, J. (1987). Control of gene expression during higher plant chloroplast biogenesis. Protein synthesis and transcript levels of psbA, psaA-psaB, and rbcL in dark-grown and illuminated barley seedlings. *Journal of Biological Chemistry, 262*, 4341-4348.

Klein, R. R., and Mullet, J. E. (1986). Regulation of chloroplast-encoded chlorophyll-binding protein translation during higher plant chloroplast biogenesis. *Journal of Biological Chemistry, 261*, 11138-11145.

Kretschmer M., Croll D., and Kronstad J.W. (2017) "Chloroplast-associated metabolic functions influence the susceptibility of maize to Ustilago maydis." *Molecular Plant Pathology,* 18, 1210-1221.

Leech, R. (1984). Chloroplast development in angiosperms: current knowledge and future prospects. *Topics in Photosynthesis*.

Leivar, P., Monte, E., Oka, Y., Liu, T., Carle, C., Castillon, A., . . . Quail, P. H. (2008). Multiple phytochrome-interacting bHLH transcription factors repress premature seedling photomorphogenesis in darkness. *Current Biology, 18*, 1815-1823.

Liu, L., Lin, N., Liu, X., Yang, S., Wang, W., and Wan, X. (2020). From chloroplast biogenesis to chlorophyll accumulation: the interplay of light and hormones on gene expression in Camellia sinensis cv. Shuchazao Leaves. *Frontiers in Plant Science,* 11, 256.

Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S.L., Kebrom, T.H., Provart, N., Patel, R., Myers, C.R. and Reidel, E.J., (2010). The developmental dynamics of the maize leaf transcriptome. *Nature Genetics*, 42, 1060-1067.

Li, Y., Liu, B., Zhang, J., Kong, F., Zhang, L., Meng, H., Li, W., Rochaix, J.D., Li, D. and Peng, L. (2019). OHP1, OHP2, and HCF244 form a transient functional complex with the photosystem II reaction center. *Plant Physiology*, 179, 195-208.

Loudya, N., Mishra, P., Takahagi, K., Uehara-Yamaguchi, Y., Inoue, K., Bogre, L., . . . López-Juez, E. (2021). Cellular and transcriptomic analyses reveal two-staged chloroplast biogenesis underpinning photosynthesis build-up in the wheat leaf. *Genome Biology*, 22, 1-30

Lu Y. (2016). Identification and roles of photosystem II assembly, stability, and repair factors in *Arabidopsis*. *Frontiers in Plant Science, 7*, 168.

Majeran W., Friso, G., Ponnala, L., Connolly, B., Huang, M., Reidel, E., . . . Sun, Q. (2010). Structural and metabolic transitions of C4 leaf development and differentiation defined by microscopy and quantitative proteomics in maize. *The Plant Cell*. 22, 3509-3542.

Mochizuki, N., Brusslan, J. A., Larkin, R., Nagatani, A., and Chory, J. (2001). *Arabidopsis* genomes uncoupled 5 (GUN5) mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. *Proceedings of the National Academy of Sciences.* 98, 2053-2058.

Müller, B., and Eichacker, L. A. (1999). Assembly of the D1 precursor in monomeric photosystem II reaction center precomplexes precedes chlorophyll a–triggered

accumulation of reaction center II in barley etioplasts. *The Plant Cell, 11*, 2365-2377.

Mullet, J. E. (1988). Chloroplast development and gene expression. *Annual Review of Plant Physiology and Plant Molecular Biology*, *39*, 475-502.

Nechushtai, R., and Nelson, N. (1985). Biogenesis of photosystem I reaction center during greening of oat, bean and spinach leaves. *Plant molecular biology*, *4*, 377-384.

Nelson, T., Harpster, M. H., Mayfield, S. P., and Taylor, W. C. (1984). Light-regulated gene expression during maize leaf development. *The Journal of Cell Biology, 98*, 558-564.

Olas, J. J., and Wahl, V. (2019). Tissue-specific NIA1 and NIA2 expression in *Arabidopsis thaliana. Plant Signaling and Behavior,* 14, 1656035.

Orzechowski, S. (2008). Starch metabolism in leaves. *Acta Biochimica Polonica,* 55, 435-445.

Pesaresi, P., Varotto, C., Meurer, J., Jahns, P., Salamini, F., & Leister, D. (2001). Knock-out of the plastid ribosomal protein L11 in Arabidopsis: effects on mRNA translation and photosynthesis. *The Plant Journal*, *27*, 179-189

Pick, T. R., Bräutigam, A., Schlüter, U., Denton, A. K., Colmsee, C., Scholz, U., ... and Weber, A. P. (2011). Systems analysis of a maize leaf developmental gradient redefines the current C4 model and provides candidates for regulation. *The Plant Cell*, *23*, 4208-4220.

Piechulla, B., Pichersky, E., Cashmore, A. R., and Gruissem, W. (1986). Expression of nuclear and plastid genes for photosynthesis-specific proteins during tomato fruit development and ripening. *Plant Molecular Biology, 7*, 367-376.

Plöchinger, M., S. Schwenkert, L. Von Sydow, W. P. Schröder, J. Meurer. (2016). Functional update of the auxiliary proteins PsbW, PsbY, HCF136, PsbN, TerC and ALB3 in maintenance and assembly of PSII. *Frontiers in Plant Science, 7*.

Pogson, B. J., Ganguly, D., and Albrecht-Borth, V. (2015). Insights into chloroplast biogenesis and development. *Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1847*, 1017-1024.

Pogson, B.J. and Albrecht, V., 2011. Genetic dissection of chloroplast biogenesis and development: an overview. *Plant Physiology,* 155, 1545-1551.

Privat, I., Hakimi, M. A., Buhot, L., Favory, J. J., and Lerbs-Mache, S. (2003). Characterization of Arabidopsisplastid sigma-like transcription factors SIG1, SIG2 and SIG3. *Plant molecular biology*, *51*, 385-399.

Puthiyaveetil, S., McKenzie, S. D., Kayanja, G. E., and Ibrahim, I. M. (2021). Transcription initiation as a control point in plastid gene expression. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1864*, 194689.

Rao X., and Dixon, R. A. (2016). The differences between NAD-ME and NADP-ME subtypes of C4 photosynthesis: more than decarboxylating enzymes. *Frontiers in Plant Science, 7*.

Reyes-Prieto A., and MoustafaA. (2012). Plastid-localized amino acid biosynthetic pathways of Plantae are predominantly composed of non-cyanobacterial enzymes. *Scientific Reports,* 2, 1-12.

Reiner, A., Yekutieli, D., and Benjamini, Y. (003). Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics, 19*, 368-375.

Richly E., Dietzmann A., Biehl A., Kurth J., Laloi C., Apel K., Salamini F., and Leister D. (2003). Covariations in the nuclear chloroplast transcriptome reveal a regulatory master-switch. *EMBO Reports,* 4, 491-498.

Rolland N., Curien G., Finazzi G., Kuntz M., Marechal E., Matringe M., Ravanel S., and Seigneurin-Berny D. (2012). The biosynthetic capacities of the plastids and integration between cytoplasmic and chloroplast processes. *Annual Review Genetics,* 46, 233–264.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*, 139-140.

Romani, I., Tadini, L., Rossi, F., Masiero, S., Pribil, M., Jahns, P., ... and Pesaresi, P. (2012). Versatile roles of Arabidopsis plastid ribosomal proteins in plant growth and development. *The Plant Journal*, *72*, 922-934.

Sano, N., and Marion-Poll, A. (2021). ABA metabolism and homeostasis in seed dormancy and germination. *International Journal of Molecular Sciences*, 22: 5069.

Sage R.F., Li M., and Monson R.K. (1999). The taxonomic distribution of C4 photosynthesis. *In C4 Plant Biology (Sage, R.F. and Monson, R.K. eds). San Diego: Academic Press,* 551-584.

Song, J., Wei, X., Shao, G., Sheng, Z., Chen, D., Liu, C., ... and Hu, P. (2014). The rice nuclear gene WLP1 encoding a chloroplast ribosome L13 protein is needed for chloroplast development in rice grown under low temperature conditions. *Plant molecular biology*, *84*, 301-314.

Skryhan, K., Gurrieri, L., Sparla, F., Trost, P., and Blennow, A. (2018). Redox regulation of starch metabolism. *Frontiers in Plant Science,* 9, 1344.

Schrubar, H., Wanner, G., and Westhoff, P. (1991). Transcriptional control of plastid gene expression in greening *Sorghum* seedlings. *Planta, 183*, 101-111.

Shen, Z., Li, P., Ni, R.-J., Ritchie, M., Yang, C.-P., Liu, G.-F., . . . Li, S.-J. (2009). Label-free quantitative proteomics analysis of etiolated maize seedling leaves during greening. *Molecular and Cellular Proteomics, 8*, 2443-2460.

Stephenson, P. G., Fankhauser, C., and Terry, M. J. (2009). PIF3 is a repressor of chloroplast development. *Proceedings of the National Academy of Sciences*, *106*, 7654-7659.

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., . . . Stitt, M. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal, 37*, 914-939.

Tozawa, Y., Teraishi, M., Sasaki, T., Sonoike, K., Nishiyama, Y., Itaya, M., ... and Hirochika, H. (2007). The plastid sigma factor SIG1 maintains photosystem I activity via regulated expression of the psaA operon in rice chloroplasts. *The Plant Journal*, *52*, 124-132.

Tyagi, A. K., and Gaur, T. (2003). Light regulation of nuclear photosynthetic genes in higher plants. *Critical Reviews in Plant Sciences, 22*, 417-452.

Von Caemmerer, S., Quick, W. P., and Furbank, R. T. (2012). The development of C4 rice: current $_{\text{progress}}$ and future challenges. *Science*, 336, 1671-1672.

Wang L., Czedik-Eysenberg A., Mertz R.A., ... Brutnell T.P. (2014). Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of maize and rice. *Nature Biotechnology,* 32, 1158-1165.

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*. *The Plant Cell, 21*, 1109-1128.

Westhoff, P., Grüne, H., Schrubar, H., Oswald, A., Streubel, M., Ljungberg, U., and Herrmann, R. (1988). Mechanisms of plastid and nuclear gene expression during thylakoid membrane biogenesis in higher plants. *Photosynthetic Light-Harvesting Systems Structure and Function*, 261-276.

Xiao, Y., Chu, L., Zhang, Y., Bian, Y., Xiao, J., and Xu, D. (2021). HY5: A pivotal regulator of light-dependent development in higher plants. *Frontiers in Plant Science, 12*.

Yamburenko, M. V., Zubo, Y. O., Vanková, R., Kusnetsov, V. V., Kulaeva, O. N., and Börner, T. (2013). Abscisic acid represses the transcription of chloroplast genes. *Journal of Experimental Botany*, 64, 4491-4502.

Yang, E. J., Yoo, C. Y., Liu, J., Wang, H., Cao, J., Li, F.-W., . . . Zhou, P. (2019). NCP activates chloroplast transcription by controlling phytochrome-dependent dual nuclear and plastidial switches. *Nature Communications, 10*, 1-13.

Zaltsman, A., Ori, N., and Adam, Z. (2005). Two types of FtsH protease subunits are required for chloroplast biogenesis and photosystem II repair in *Arabidopsis*. *The Plant Cell,* 17, 2782-2790.

Zhang, J., H,Yuan, Y. Yang, T. Fish, S. M. Lyi, T. W. Thannhauser, L. Zhang, L. Li. (2016). Plastid ribosomal protein S5 is involved in photosynthesis, plant development, and cold stress tolerance in *Arabidopsis*. *Journal of Experimental Botany, 67*, 2731-2744.

Zhou, X., Fei, Z., Thannhauser, T. W., and Li, L. (2011). Transcriptome analysis of ectopic chloroplast development in green curd cauliflower (Brassica oleracea L. var. botrytis). *BMC Plant Biology, 11*, 1-12.

Zhao, D. S., Zhang, C. Q., Li, Q. F., Yang, Q. Q., Gu, M. H., and Liu, Q. Q. (2016). A residue substitution in the plastid ribosomal protein L12/AL1 produces defective plastid ribosome and causes early seedling lethality in rice. *Plant molecular biology*, *91*, 161-177

Supplementary Data

Supplementary Figures

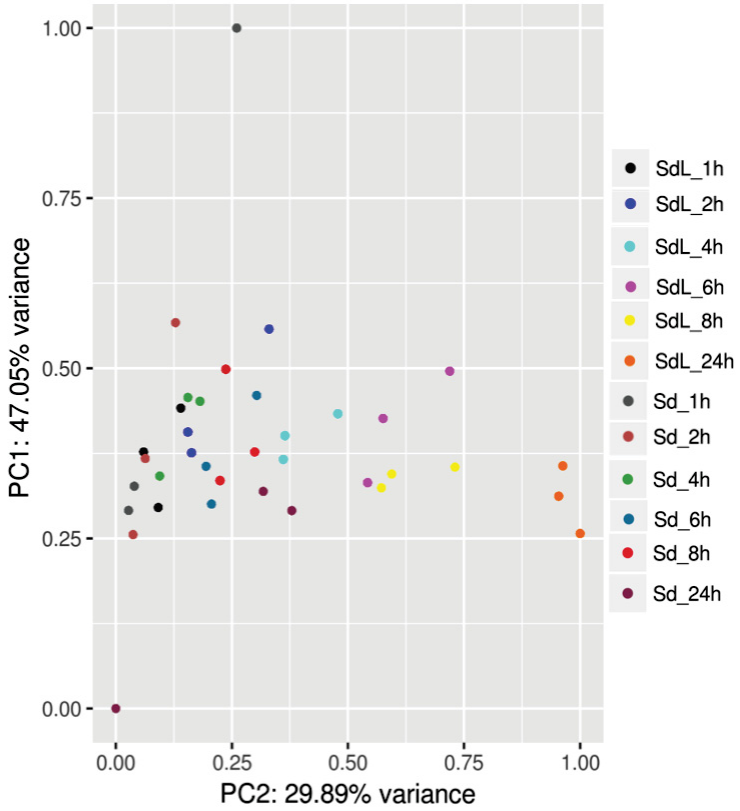


Figure S1. Principal component analysis **(**PCA) displays the variance among each biological replicate of each sample. The PCA plot was generated using the online tool PCAGO.

Supplementary Table 1

Supplementary Table 1: An overview of Log2 fold changes and p-values of differentially expressed genes involved in chloroplast biogenesis

PS light reaction														
Locus name	Arabidospis_ID	Gene symbol	Log2FC_1h	Log2FC_2h	Log2FC_4h	Log2FC_6h	Log2FC_8	Log2FC_24	1h_PValue	2h_PValue	4h_PValue	6h_Pvalue	8h_Pvalue	24h_PValue
Sobic.003G203000.1	AT1G34000.1	OHP2	1,03E+14	1,71E+13	1,73E+14	1,51E+14	1,13E+06	1,23E+06	0,000419021	1,51465E-11	2,29898E-10	2,58114E-08	0,000120005	0,000321978
Sobic.001G314200.1	AT2G46820.2	PSAP	5,01E-01	8,53E-01	1,10E+14	8,64E-01	1,13E+06	1,37E+06	0,061684981	0,001057821	2,73948E-05	0,00204567	4,16176E-05	0,0000151
Sobic.001G046900.1	AT1G03130.1	PSAD-2	5,14E-01	8,17E-01	1,34E+13	8,60E-01	7,53E-01	1,66E+06	0,129766877	0,007395935	0,000416264	0,03511086	0,059439462	0,000115106
Sobic.002G119500.3	AT2G20260.1	PSAE-2	4,39E-01	6,11E-01	9,14E-01	7,61E-01	5,97E-01	1,17E+06	0,259348864	0,076084709	0,019876186	0,07095962	0,173823849	0,021762633
Sobic.001G070500.1	AT1G31330.1	PSAF	6,59E-01	1,03E+14	1,25E+14	8,29E-01	6,82E-01	1,59E+06	0,077124109	0,003098995	0,001151219	0,03661672	0,094276586	0,001716399
Sobic.002G242000.1	AT1G55670.1	PSAG	6,73E-01	9,57E-01	1,62E+14	1,12E+14	8,43E-01	1,83E+06	0,102990658	0,007821015	3,53428E-05	0,006830012	0,049123568	0,000411953
Sobic.009G229700.1	AT1G52230.1	PSAH-2	7,05E-01	9,75E-01	1,35E+14	9,48E-01	8,54E-01	1,70E+06	0,129918686	0,009098216	0,002156108	0,0311697	0,07348428	0,00324513
Sobic.002G033900.1	AT1G30380.1	PSAK	7,95E-01	1,54E+14	1,59E+14	1,15E+14	9,18E-01	1,82E+06	0,066525935	0,000216227	0,000517074	0,008879034	0,059191437	0,002986303
Sobic.003G052500.1	AT4G12800.1	PSAL	7,38E-01	1,11E+14	1,41E+14	1,03E+14	8,97E-01	1,76E+06	0,086003792	0,001190811	0,000592956	0,008653367	0,027633059	0,000581825
Sobic.008G063500.1	AT5G64040.1	PSAN	6,42E-01	6,47E-01	1,21E+14	9,15E-01	5,91E-01	1,37E+06	0,049331032	0,011411617	4,70109E-05	0,006681928	0,066574979	0,001113113
Sobic.006G073500.1	AT1G08380.1	PSAO	5,34E-01	1,04E+13	1,47E+14	9,63E-01	6,09E-01	1,67E+06	0,194925997	0,003518512	0,000176291	0,02147501	0,203935331	0,001542895
Sobic.007G136900.1	AT3G47470.1	CAB4	6,37E-01	1,29E+14	1,37E+14	8,64E-01	1,06E-01	2,11E+06	0,088020034	0,00021919	0,00010155	0,01554955	0,902067526	0,00000389
Sobic.006G264201.1	AT1G15820.1	CP24	5,51E-01	1,20E+14	1,52E+13	9,47E-01	4,37E-01	2,28E+05	0,126750327	0,000388704	5,64702E-05	0,01284708	0,404952168	7,16E-08
Sobic.002G289100.1	AT2G34430.1	LHB1B1	8,24E-01	1,27E+14	1,53E+14	9,11E-01	-8,31E-03	2,37E+06	0,005558382	5,19194E-05	1,98358E-06	0,01085829	0,656185451	1,66E-11
Sobic.009G234600.1	AT2G34420.1	LHB1B2	6,49E-01	8,54E-01	1,17E+14	8,58E-01	2,19E-01	1,53E+06	0,112051173	0,053218395	0,005145508	0,05218604	0,817309197	0,009148965
Sobic.004G056900.1	AT3G54890.1	LHCA1	6,26E-01	1,03E+14	1,21E+14	8,09E-01	5,44E-01	1,85E+06	0,12037052	0,004713832	0,002991432	0,04782756	0,261711707	0,000253378
Sobic.002G352100.1	AT3G61470.1	LHCA2	6,77E-01	1,23E+14	1,53E+14	8,43E-01	-7,01E-02	2,19E+06	0,077906394	0,000638521	7,6385E-05	0,04248195	0,493336691	0,00001
Sobic.002G215000.2	AT1G19150.1	LHCA2*1	4,03E-01	8,27E-01	1,19E+14	1,14E+13	1,60E+06	1,59E+06	0,166291978	0,00429303	0,001134445	0,001420363	3,1065E-06	0,0000442
Sobic.010G189300.1	AT1G61520.1	LHCA3	7,20E-01	1,06E+14	1,16E+13	8,14E-01	6,78E-01	1,80E+06	0,057196427	0,001638673	0,001595376	0,02644568	0,100334096	0,000204003
Sobic.004G308700.1	AT1G45474.1	LHCA5	5,51E-01	8,42E-01	1,42E+14	1,23E+13	1,45E+06	1,78E+06	0,025793281	8,37948E-05	1,38831E-07	1,95387E-06	5,55175E-09	2,31E-09
Sobic.001G177000.1	AT3G27690.1	LHCB2	7,89E-01	1,45E+14	1,38E+14	9,63E-01	2,82E-01	1,80E+06	0,036267371	5,38807E-06	0,000139873	0,005230291	0,729055295	0,0000934
Sobic.002G339200.2	AT5G54270.1	LHCB3	8,02E-01	1,56E+14	1,64E+14	1,03E+14	9,40E-02	2,60E+04	0,06242873	0,000270352	0,000197848	0,02147773	0,814703485	0,000000361
Sobic.002G338000.1	AT3G08940.2	LHCB4.2	7,50E-01	1,06E+13	1,21E+14	7,60E-01	5,65E-01	1,84E+06	0,024762032	0,000557839	0,000393185	0,03545301	0,159302236	0,00000821
Sobic.005G087000.1	AT4G10340.1	LHCB5	9,97E-01	1,63E+14	1,74E+14	1,11E+13	6,74E-01	2,43E+06	0,003362142	4,19529E-07	6,69545E-08	0,000734304	0,052792509	1,26E-09
Sobic.003G169100.1	ATCG00680.1	PSbB, CP47	7,71E-01	2,56E-01	2,04E+14	2,28E+14	2,67E+06	1,18E+05	0,234365331	0,954997732	0,000327171	0,007900195	0,000253996	0,399896434

Locus name	Arabidospis_ID	Gene symbol	Log2FC_1h	Log2FC_2h	Log2FC_4h	Log2FC_6h	Log2FC_8	Log2FC_24	1h_PValue	2h_PValue	4h_PValue	6h_Pvalue	8h_Pvalue	24h_PValue
Sobic.003G370000.1	AT1G44575.1	NPQ4	1,61E+14	2,73E+14	3,11E+14	2,44E+12	1,61E+06	2,08E+06	0,000853468	2,23722E-14	1,9512E-13	5,54235E-10	8,86198E-05	0,0000313
Sobic.002G033800.1	AT5G02120.1	OHP	5,28E-01	1,51E+14	1,34E+13	9,13E-01	7,41E-01	1,25E+06	0,041673601	3,52896E-09	1,10303E-05	0,001518969	0,013430766	0,0000478
Sobic.004G191200.1	AT1G14150.1	PQL1	7,42E-01	1,29E+14	1,65E+14	1,49E+14	1,88E+06	1,61E+06	0,0069618	2,14323E-07	2,88642E-08	1,63154E-06	5,81162E-11	2,69E-08
Sobic.003G419900.1	AT4G28660.1	PSB28	1,13E+14	2,23E+14	1,88E+14	1,68E+14	1,25E+06	1,65E+06	0,000373949	2,05828E-15	1,16691E-08	1,36997E-08	0,000145029	0,00000373
Sobic.002G329600.1	AT4G05180.1	PSBQ	5,86E-01	7,19E-01	1,25E+14	8,81E-01	7,04E-01	1,50E+06	0,075045304	0,006693967	1,41369E-05	0,006560071	0,029472665	0,0000918
Sobic.007G071500.1	AT1G79040.1	PSBR	6,08E-01	7,73E-01	1,21E+14	8,58E-01	7,05E-01	1,17E+06	0,094815095	0,005441611	0,000112197	0,01233756	0,042684175	0,00834873
ATP Synthase														
Sobic,003G253300,1	AT2G31040,1	CGL160	6,46E-01	7,59E-01	1,14E+14	1,32E+13	1,19E+06	1,42E+06	0,000487005	2,81122E-05	9,3846E-13	3,46376E-12	6,392E-12	1,49E-16
Sobic,004G235200,1	AT4G09650,1	ATPD	5,84E-01	8,62E-01	1,34E+14	1,16E+14	1,30E+06	1,74E+05	0,045088808	0,00147174	8,71388E-06	0,000136356	3,61991E-05	0,00000115
Sobic,002G316700,1	AT4G04640,1	ATPC1	7,46E-01	1,19E+14	1,42E+14	1,20E+14	1,36E+06	2,07E+06	0,016182848	4,07431E-05	2,4235E-05	0,000206892	2,90619E-05	8,78E-08
Sobic,001G417200,1	AT4G32260,1	PDE334	7,94E-01	1,08E+13	1,37E+14	1,31E+14	1,35E+06	1,92E+06	0,018347669	0,000312558	2,61475E-05	6,31309E-05	7,84902E-05	0,000000341
Cyclic electron transcport														
Sobic,010G201266,1	ATCG01250,1	0	6,44E-01	-4,81E-01	1,39E+14	2,72E+14	1,76E+06	1,33E+06	0,210988248	0,610024011	0,009159709	0,0364459	0,00247144	0,260837629
Sobic,006G258700,1	AT4G22260,1	PTOX	7,79E-01	1,85E+14	1,90E+13	1,63E+14	1,50E+05	1,39E+06	0,006231067	3,05586E-14	1,37436E-16	2,37405E-10	3,875E-08	0,00000237
Sobic,001G498300,1	AT1G64770,1	NDF2	6,17E-01	8,31E-01	7,59E-01	9,50E-01	1,50E+06	1,71E+06	0,102556673	0,007057854	0,084383477	0,01149021	9,95954E-05	0,000109546
Sobic,007G161500,1	AT2G05620,1	PGR5	1,75E+14	3,47E+14	2,85E+14	2,11E+14	1,57E+06	2,75E+14	9,74814E-06	1,38605E-23	3,33242E-15	3,61428E-10	7,44638E-06	1,78E-09
Sobic,001G005200,2	AT4G22890,4	PGR5-LIKE A	5,34E-01	9,29E-01	1,48E+14	1,04E+13	1,08E+06	1,65E+06	0,013968162	3,05831E-07	2,90504E-12	2,02983E-06	5,37371E-09	1,35E-12
Sobic,001G179100,1	AT3G15840,1	PIFI	8,03E-01	1,64E+14	1,45E+14	1,41E+13	1,73E+06	1,37E+06	0,018104418	7,60661E-08	8,93351E-06	2,47984E-05	6,00324E-07	0,000129744
Cytochrome b6f														
Sobic,001G078000,1	AT2G26500,1	PETM	6,13E-01	1,14E+14	1,48E+14	1,30E+14	1,43E+06	2,05E+06	0,080423018	0,000939803	0,000180999	0,000596193	0,000233921	0,0000253
Sobic,003G440500,1	AT5G36120,1	atylmg3	2,73E-01	5,84E-01	8,82E-01	8,59E-01	9,58E-01	1,12E+06	0,452742449	0,004729676	0,000343702	0,001383114	0,000126301	0,0000195
Sobic,003G168700,1	ATCG00720,1	PETB	7,29E-01	4,21E-01	1,36E+14	2,45E+14	2,99E+06	1,44E+04	0,222618373	0,704457311	0,098882147	0,004713586	8,62549E-05	0,198164161
Sobic,009G143100,1	AT4G03280,1	PETC	7,30E-01	9,55E-01	1,08E+11	1,25E+14	1,52E+06	2,01E+06	0,049015712	0,001759946	0,001725684	0,000241858	2,33451E-05	0,00000468
NADH related genes														
Sobic,003G378600,1	AT5G58260,1	NDHN	6,80E-01	8,79E-01	1,05E+14	1,04E+14	1,52E+06	1,84E+06	0,052415556	0,00155502	0,002319514	0,001525719	4,17382E-06	0,0000113
Sobic,002G198600,1	AT1G70760,1	CRR23	6,41E-01	7,49E-01	9,06E-01	8,22E-01	1,44E+06	1,70E+06	0,046395254	0,006136205	0,003634854	0,01155679	4,04022E-06	0,00000589
Sobic,006G162200,1	AT4G37925,1	NDH-M	5,83E-01	9,38E-01	1,06E+13	1,16E+14	1,70E+06	1,67E+06	0,127883497	0,005427892	0,006436517	0,001547771	4,24401E-06	0,000178782
Sobic,003G433900,1	AT1G74880,1	NDH-O	2,98E-01	3,40E-01	6,62E-01	6,66E-01	1,09E+06	9,64E-01	0,409733895	0,105168785	0,038520951	0,07584521	0,000313409	0,006250582
PS electron carrier														
Sobic,007G004600,1	AT1G60950,1	ATFD2	8,13E-01	1,37E+14	1,64E+14	1,49E+14	1,28E+06	1,26E+06	0,074512046	0,000113974	0,000102752	0,000152494	0,002188818	0,010755577
Sobic,002G305700,1	AT3G16250,1	NDF4	5,49E-01	6,46E-01	6,97E-01	6,77E-01	1,56E+06	1,16E+06	0,005971764	0,001492636	0,021913273	0,01896133	2,8756E-11	0,000000277
Sobic,003G431700,1	AT5G66190,1	ATLFNR1	7,56E-01	1,03E+14	1,44E+14	1,35E+14	1,74E+06	1,90E+06	0,020332071	8,77469E-06	4,54109E-07	1,84091E-06	1,42775E-09	0,00000002
Sobic,010G005500,1	AT1G20020,3	ATLFNR2	3,03E-01	7,71E-01	1,36E+14	1,17E+14	1,09E+06	1,78E+06	0,209698751	9,95388E-07	4,7583E-11	3,99943E-08	7,09448E-09	1,21E-14
Sobic,010G000500,1	AT1G76100,1	PETE1	4,57E-01	9,77E-01	1,44E+14	1,09E+14	1,02E+06	1,87E+06	0,169118521	0,001635706	0,000104933	0,003506242	0,004793724	0,00000448

Calvin cycle														
Locus name	Arabidospis_ID	Gene symbol	Log2FC_1h	Log2FC_2h	Log2FC_4h	Log2FC_6h	Log2FC_8h	Log2FC_24h	Pvalue_1h	Pvalue_2h	Pvalue_4h	Pvalue_6h	Pvalue_8h	Pvalue_24h
Sobic,010G082800,1	AT2G47400,1	CP12	1,42E+14	2,34E+14	2,68E+14	2,59E+14	2,58E+06	2,33E+06	1,66111E-05	8,96556E-14	1,36046E-15	2,23902E-12	1,06141E-14	5,04E-11
Sobic,005G056400,1	AT4G38970,1	FBA2	3,91E-01	1,10E+14	1,47E+14	1,25E+14	1,25E+04	2,46E+14	0,08981401	2,75704E-07	3,39117E-09	2,55182E-07	8,61787E-09	4,37E-19
Sobic,009G152700,1	AT1G43670,1	FBP	5,04E-01	5,85E-01	8,66E-01	1,34E+14	1,82E+06	1,04E+06	0,069121675	0,013861025	0,00029954	9,96148E-09	6,69295E-16	0,0002988
Sobic,001G425400,1	AT3G54050,1	HCEF1	4,18E-01	7,98E-01	1,01E+14	8,85E-01	1,27E+06	2,00E+06	0,159283645	0,004294568	0,004023248	0,005246526	1,36741E-05	0,00000024
Sobic,006G105900,1	AT1G12900,1	GAPA-2	6,38E-01	1,20E+14	1,70E+13	1,46E+14	1,51E+06	2,33E+05	0,10530579	0,000261365	8,10116E-06	5,53567E-05	3,81648E-05	0,000000627
Sobic,001G519800,1	AT1G42970,1	GAPB	6,92E-01	1,19E+14	1,68E+14	1,45E+14	1,37E+06	2,19E+06	0,059105073	7,19789E-05	2,55611E-06	2,46669E-05	8,9814E-05	0,000000194
Sobic,010G221800,1	AT1G79550,1	PGK	1,16E-01	4,10E-01	2,21E-01	-4,90E-01	-1,24E+06	-1,46E+06	0,545527351	0,54291398	0,039510871	0,07847497	3,60627E-08	0,000000346
Sobic,009G183700,1	AT1G56190,1	PGK2	3,85E-01	5,09E-01	8,23E-01	8,02E-01	1,02E+06	1,44E+05	0,055575908	0,000897565	3,64271E-05	6,52653E-05	5,02076E-08	3,79E-10
Sobic,004G272100,1	AT1G32060,1	PRK	4,06E-01	9,30E-01	1,44E+14	1,06E+14	1,34E+05	1,72E+06	0,113818793	7,83319E-05	1,82364E-08	0,000127915	7,26997E-09	0,000000143
Sobic,002G199000,1	AT1G14030,1	LSMT-L	3,77E-01	5,62E-01	8,76E-01	7,97E-01	1,07E+05	8,35E-01	0,016061494	0,001596533	2,38814E-05	0,00065067	2,07472E-08	0,00000104
Sobic,005G231600,1	AT2G39730,1	RCA	7,70E-01	1,74E+14	1,32E+14	9,18E-01	1,05E+06	1,68E+05	0,088876375	1,28232E-06	0,002632745	0,01241053	0,015792431	0,00145499
Sobic,005G042000,1	AT1G67090,1	RBCS1A	5,40E-01	9,27E-01	1,38E+14	1,20E+14	1,39E+06	1,87E+06	0,0466345	0,000599698	4,01021E-06	5,31062E-05	1,2326E-06	3,53E-08
Sobic,003G359100,1	AT3G55800,1	SBPASE	6,83E-01	1,11E+14	1,58E+14	1,50E+13	1,72E+06	2,02E+06	0,045733571	6,97489E-05	1,39378E-07	1,61378E-06	3,83435E-09	8,63E-08
Sobic,006G034300,1	AT2G45290,1	TKL2	2,99E-01	1,81E+14	5,83E-01	6,67E-01	2,93E-01	2,74E+06	0,664816303	0,579541467	4,90946E-05	0,2671678	0,042084732	0,008576724
Photorespiration														
Sobic,007G213700,1	AT2G13360,2	AGT	1,35E+13	2,22E+14	2,62E+14	2,25E+14	2,00E+06	1,71E+06	4,3586E-05	1,65396E-13	1,91259E-15	4,61313E-12	2,04867E-09	0,0000136
Sobic,009G004100,1	AT1G80380,2	GLYK	6,23E-01	1,18E+14	1,64E+13	1,47E+13	1,21E+06	1,40E+06	0,004264122	8,78833E-11	7,28357E-16	1,06295E-12	8,40145E-09	1,27E-09
Sobic,001G065600,1	AT3G14420,1	GLO1	4,15E-01	7,44E-01	1,26E+14	1,23E+14	1,56E+05	1,71E+06	0,198046032	0,000659345	2,39448E-06	6,05503E-07	3,47435E-10	0,000000109
Sobic,002G036000,4	AT3G14420,1	GLO1	6,39E-01	1,37E+14	2,21E+14	1,83E+14	2,34E+06	2,40E+06	0,055399592	8,42349E-07	2,1931E-13	4,97963E-10	7,87834E-16	1,44E-11
Sobic,004G001300,1	AT1G68010,1	ATHPR1	5,17E-01	8,05E-01	1,46E+14	1,27E+13	1,74E+06	1,26E+06	0,172389835	0,000535081	1,52309E-07	3,6058E-05	4,03571E-10	0,000505216
C4 related genes														
Sobic,010G160700,1	AT2G42600,1	PEPC2	1,23E+05	1,17E+14	2,32E+14	1,92E+14	2,12E+06	2,11E+06	0,224047727	0,110971127	0,012863299	0,04686518	0,027396148	0,051422961
Sobic,007G166200,1	AT5G58330,1	NADP-MDH	5,22E-01	6,11E-01	1,26E+14	1,13E+14	1,61E+06	1,09E+06	0,04006312	0,000534347	1,90011E-08	3,42107E-06	3,37458E-13	0,000000471
Sobic,007G166300,1	AT5G58330,1	NADP-MDH	7,64E-01	1,25E+14	2,01E+14	1,77E+14	1,92E+05	1,54E+06	0,011282807	7,34929E-09	1,17556E-13	4,1389E-11	1,55361E-13	3,02E-08
Sobic,004G088800,1	AT2G02030,1	0	4,96E-01	2,08E-01	1,59E+14	6,71E-01	1,18E+06	6,47E-01	0,590324839	0,468701265	0,016212845	0,1077894	0,246938815	0,608071807
Sobic,003G292400,1	AT2G19900,1	ATNADP-ME1	-3,16E-01	1,44E+14	1,99E+14	6,12E-01	9,15E-01	3,31E+05	0,450977413	0,310061952	4,92383E-08	0,658769	0,344836788	0,008232234
Sobic,003G036200,1	AT1G79750,1	ATNADP-ME4	8,57E-01	7,60E-01	1,92E+14	2,26E+14	2,72E+06	2,32E+06	0,003416812	0,005730749	5,79724E-15	3,39106E-15	1,46693E-23	2,09E-10

Others															
Locus name	Arabidospis_ID	Gene symbol	Log2FC_1h	Log2FC_2h	Log2FC_4h	Log2FC_6h	Log2FC_8	Log2FC_24	1h_PValue	2h_PValue	4h_PValue	6h_Pvalue	8h_Pvalue	24h_PValue
Sobic,002G316900,1	AT2G05710,1	ACO3	1,45E-01	1,06E+14	1,53E+14	7,54E-01	-3,98E-01	2,32E-01	0,773632851	0,660944808	7,55792E-11	0,07733702	0,015657394	0,894449072
Sobic,010G101300,2	AT1G24180,1	IAR4	-1,57E-01	1,10E+14	1,43E+14	4,84E-02	2,47E-01	8,94E-01	0,700880092	0,801218083	7,82444E-09	0,9464297	0,27115457	0,09922212
Sobic,001G372600,1	AT5G18660,1	PCB2	3,29E-01	9,96E-01	1,15E+14	1,12E+14	1,01E+06	1,60E+06	0,197401212	0,000172038	0,000500971	0,00065293	0,002060117	0,000000519
Sobic,006G234100,1	AT1G58290,1	HEMA1	8,91E-01	1,63E+14	1,59E+13	1,17E+14	5,81E-01	1,59E+05	0,000182808	1,50274E-13	4,6144E-11	4,90701E-07	0,032352685	0,00000011
Sobic,010G184800,1	AT2G26670,1	GUN2	1,02E+14	-2,12E-01	-6,50E-01	-2,41E-01	-2,62E+14	-1,25E+06	0,002332507	0,497300469	0,845529712	0,223623	8,52878E-18	0,005519445
Sobic,006G264900,1	AT5G13630,1	CHLH, GUN5	9,21E-01	1,94E+14	1,73E+14	1,47E+14	7,54E-01	1,99E+06	0,001784682	2,33402E-13	1,25875E-09	4,4089E-07	0,017032277	8,07E-08
Sobic,010G022100,1	AT4G25080,1	CHLM	4,41E-01	8,21E-01	9,63E-01	8,51E-01	8,54E-01	1,54E+06	0,11924757	0,007771883	0,032371518	0,02514268	0,039189063	0,0000267
Sobic,008G038300,1	AT3G59400,1	GUN4	7,99E-01	1,66E+14	1,58E+14	1,22E+14	7,69E-01	1,88E+06	0,007018639	8,84561E-10	9,23732E-07	5,58467E-06	0,016913357	0,00000027
Sobic,001G243300,2	AT5G43860,1	CLH2	7,55E-01	1,64E+14	1,54E+14	1,31E+14	1,46E+05	8,81E-01	0,010507056	5,54278E-05	0,000162057	0,000171428	0,000313592	0,039371691
Sobic,007G173800,1	AT4G35250,1	HCF244	7,23E-01	1,55E+14	1,81E+14	1,67E+14	1,73E+06	1,32E+06	0,00466818	5,16293E-15	5,40408E-17	2,88175E-14	6,66492E-15	0,000000133
Sobic,001G020600,1	AT5G08720,1	HCF145	1,48E+14	1,58E+14	1,87E+14	1,55E+14	3,10E-01	1,72E+06	1,04056E-06	3,4896E-15	5,07257E-24	1,2897E-14	0,129302746	8,06E-09
Sobic,003G421400,1	AT4G31560,1	HCF153	2,72E-01	3,06E-01	7,14E-01	1,06E+14	1,14E+06	5,40E-01	0,20272291	0,17746281	0,006061134	2,60452E-05	6,38821E-06	0,03579092
Sobic,006G010900,1	AT3G24430,1	HCF101	2,19E-01	4,90E-01	7,25E-01	7,46E-01	1,03E+06	1,29E+05	0,075289975	0,004368275	0,000378459	0,001100762	1,11127E-08	5,84E-13
Sobic,010G277600,1	AT5G23120,1	HCF136	4,62E-01	1,05E+14	1,18E+13	1,13E+14	8,55E-01	1,02E+06	0,053622662	3,37582E-07	1,04559E-06	5,63428E-06	0,000899011	0,000161838
Sobic,004G085600,1	AT5G11260,1	HY5	1,69E+14	2,38E+14	3,15E+14	3,14E+14	1,79E+06	1,52E+06	3,94591E-11	6,94754E-31	3,06226E-59	1,39424E-41	3,2078E-14	1,49E-11
Sobic,006G217800,1	AT1G09530,1	PIF3	-2,24E-01	-1,83E+14	-1,85E+14	-1,99E+14	-3,71E+13	-4,65E+14	0,259062851	7,82156E-12	5,33542E-12	6,14659E-16	5,18384E-41	5,98E-49
Sobic,008G163700,3	AT2G43010,1	PIF4	-0,07559	-0,6442314	-0,9363379	-1,097E+14	-0,5027909	0,7579238	0,323766726	0,0173768	2,56371E-06	2,4995E-06	0,016554231	0,027027951
Sobic.008G060200.1	AT5G65700.1	BAM1	-1,44E-01	8,99E-02	-5,07E-01	-7,19E-01	-1,00E+06	-5,51E-01	0,871894471	0,711017184	0,101673566	0,03132036	0,00359457	0,020919887
Sobic.009G259500.1	AT4G20270.1	BAM3	1,84E-01	1,24E+13	3,96E-01	5,29E-01	1,06E-02	2,27E-01	0,494015739	0,009041514	0,3104773	0,1560516	0,981835139	0,753407443
Sobic.009G245000.2	AT1G27680.1	APL2	2,84E-01	9,04E-01	1,28E+14	1,08E+14	7,81E-01	1,13E+06	0,283337238	0,001777187	9,23025E-15	9,63216E-06	0,00465625	5,58E-08
Sobic.007G101500.1	AT5G48300.1	APS1	1,95E-01	4,06E-01	6,44E-01	4,80E-01	6,56E-01	1,06E+06	0,169191448	0,008996697	5,67349E-05	0,06860371	2,77954E-05	6,1E-09
Sobic.001G100000.1	AT5G19220.1	APL1	5,51E-01	5,34E-01	9,75E-01	8,57E-01	1,27E+06	1,50E+06	0,108647769	0,011514001	0,000506924	0,006379072	2,02193E-07	0,00000727
Sobic.006G066200.1	AT1G08250.1	ADT6	1,58E+14	6,36E-01	1,02E+14	1,53E+14	1,71E+06	2,01E+06	0,000871627	0,014018395	0,003273927	4,48238E-08	9,14644E-06	0,0000113
Sobic.002G302100.1	AT3G14390.1	DAPDC1	1,39E+14	1,65E+14	2,01E+14	2,28E+14	1,46E+06	1,41E+06	0,00091809	7,00728E-06	9,65705E-11	3,03015E-11	8,56428E-05	0,000237467
Sobic.002G284000.1	AT5G11880.1	DAPDC2	1,07E+14	1,36E+14	1,23E+14	1,32E+14	2,91E-01	1,50E+06	0,021438499	0,000511086	0,001991515	0,000316498	0,710156173	0,00000482
Sobic.010G214100.1	AT1G17745.1	PGDH	4,22E-01	6,46E-01	9,22E-01	1,05E+14	1,14E+06	1,33E+06	0,00573214	4,59536E-06	7,84429E-07	3,31411E-10	4,30036E-12	3,17E-20
Sobic.004G196101.1	AT1G77760.1	GNR1	-2,74E-01	2,24E+14	1,67E+14	1,37E+14	2,70E+14	2,26E+06	0,731137953	7,2119E-05	0,001633081	1,84141E-06	4,26599E-07	0,00000307
Sobic.007G153900.1	AT1G77760.1	GNR1	1,11E+14	2,61E+14	3,30E+14	2,46E+14	2,69E+06	2,07E+06	0,03883849	3,15189E-09	2,23151E-09	2,02001E-07	3,61224E-07	0,000370164
Sobic.004G312500.1	AT1G77760.1	GNR1	1,83E+14	1,96E+13	2,71E+14	2,79E+14	4,03E+14	1,32E+06	0,0026058	5,64153E-06	5,28488E-06	8,54294E-09	1,35029E-12	0,018721363
Sobic.007G170300.1	AT4G32810.1	ATCCD8	7,08E-01	2,19E+14	2,41E+14	2,72E+13	3,36E+14	2,48E+06	0,430594374	0,000154894	3,74271E-08	2,57131E-10	7,99942E-11	0,0000224
Sobic.001G155300.1	AT1G78390.1	ATNCED9	2,61E+13	1,69E+14	3,67E+14	4,97E-01	-5,23E-01	-3,02E-01	8,09337E-05	7,58166E-09	1,02255E-11	0,3178849	0,153736802	0,419262931
Sobic.006G097500.1	AT5G67030.1	ABA1	5,62E-01	1,52E+14	1,57E+14	1,04E+14	-1,69E-01	2,11E+06	0,014778359	4,54924E-13	4,60222E-14	1,06372E-05	0,11600089	3,71E-16

Ribosomal subunits														
Locus name	Arabidospis_ID	Gene symbol	Log2FC_1h	Log2FC_2h	Log2FC_4h	Log2FC_6h	Log2FC_8	Log2FC_24	1h_PValue	2h_PValue	4h_PValue	6h_Pvalue	8h_Pvalue	24h_PValue
Sobic.001G122400.1	AT5G14320.1	RPS13	2,84E-01	5,62E-01	7,33E-01	7,05E-01	1,05E+06	9,01E-01	0,044395092	0,000447402	0,000444142	0,000793143	1,52212E-08	0,0000003
Sobic.003G256200.1	AT3G15190.1	RPS20	2,96E-01	5,85E-01	7,70E-01	7,72E-01	9,73E-01	1,04E+06	0,050133841	0,000616416	0,000333758	0,000495721	3,21511E-06	2,51E-08
Sobic.001G191000.1	AT2G33800.1	RPS5	3,62E-01	6,48E-01	6,98E-01	7,06E-01	1,02E+06	1,28E+06	0,110750854	0,004779353	0,023754485	0,01407116	0,00033052	0,0000207
Sobic.001G017000.1	AT1G64510.1	RPS6	2,24E-01	6,75E-01	5,55E-01	7,09E-01	9,84E-01	1,28E+06	0,431369735	0,068326031	0,328475913	0,1071254	0,029078889	0,004010176
Sobic.009G123900.1	AT3G63490.1	RPL1	2,72E-01	5,30E-01	1,01E+13	7,76E-01	9,80E-01	7,70E-01	0,071023708	0,000441978	1,18646E-08	0,000239564	1,52738E-07	0,00000317
Sobic.001G411300.1	AT5G13510.1	RPL10	3,98E-01	8,55E-01	8,63E-01	7,50E-01	1,16E+06	1,27E+06	0,082846219	0,002029676	0,012871582	0,01585725	0,000161644	0,00007
Sobic.001G527100.1	AT1G32990.1	PRPL11	3,50E-01	5,39E-01	7,91E-01	8,97E-01	1,25E+06	1,08E+06	0,079406661	0,005446022	0,000458368	0,000160959	4,19215E-08	0,000000317
Sobic.003G246400.1	AT3G27850.1	RPL12-C	2,90E-01	5,78E-01	6,21E-01	7,63E-01	9,82E-01	1,10E+06	0,212495108	0,018021061	0,108137838	0,01366285	0,001622361	0,000345485
Sobic.010G233800.1	AT5G54600.1	RPL24	4,33E-01	4,27E-01	7,41E-01	6,94E-01	1,14E+06	9,95E-01	0,016779088	0,007226452	0,000629185	0,003724576	2,54973E-08	0,000000188
Sobic.003G408000.2	AT5G40950.1	RPL27	4,78E-01	4,44E-01	5,18E-01	6,45E-01	1,11E+06	9,57E-01	0,096563186	0,044076093	0,062705109	0,02132915	3,53186E-05	0,001075642
Sobic.001G240000.1	AT2G33450.1	RPL28	5,13E-01	6,26E-01	8,20E-01	7,47E-01	9,78E-01	1,16E+06	0,017479746	0,003409655	0,002074135	0,004093788	0,000214962	0,000012
Sobic.004G231800.1	AT5G65220.1	RPL29	4,24E-01	5,64E-01	4,61E-01	5,60E-01	1,15E+06	1,00E+06	0,016179797	0,00180024	0,117224391	0,02226806	1,62424E-07	0,00000304
Sobic.004G034500.1	AT2G43030.1	PRPL3	4,05E-01	7,37E-01	6,13E-01	6,74E-01	1,07E+06	1,24E+06	0,069932415	0,001786322	0,057796736	0,01641418	7,88534E-05	0,0000222
Sobic.001G427000.1	AT1G07320.1	RPL4	4,50E-01	6,53E-01	7,20E-01	7,11E-01	1,01E+05	1,08E+06	0,042746063	0,000360338	0,008398049	0,004197875	2,5531E-05	0,000032
Sobic.001G524900.1	AT4G01310.1	RPL5	3,70E-01	5,42E-01	6,83E-01	5,81E-01	1,00E+06	9,74E-01	0,144432751	0,004351113	0,018415801	0,02965632	6,34181E-05	0,000519278
Sobic.010G208000.1	AT2G24090.1	RPL35	4,43E-01	5,69E-01	8,05E-01	7,74E-01	1,21E+06	1,09E+06	0,044056642	0,002115458	0,003022038	0,002759032	6,45747E-07	0,00000645
Sobic.003G328000.1	AT1G29070.1	RPL34	4,83E-01	3,86E-01	7,36E-01	7,50E-01	1,12E+06	1,03E+06	0,013288683	0,013585539	0,002535479	0,001488997	5,24075E-07	0,00000153

Manuscript II

Identifying and isolating genes affecting leaf development in *Arabidopsis thaliana*

Manuscript II

Identifying and isolating genes affecting leaf development in *Arabidopsis thaliana*

Zahida Bano and Peter Westhoff *

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

To whom correspondence should be addressed*

Zahida Bano
Email: Zahida.Bano@hhu.de

Peter Westhoff
Email: West@hhu.de

Summary

Understanding the mechanism of leaf growth and development is an important objective in biology in order to improve crop productivity and resistance to the changing climate in the future. Leaf functions as a solar panel due to its photosynthetic capacity that provides the basis for the growth of the entire plant. Although the molecular mechanisms of leaf development have been well studied in *Arabidopsis* compared to grass species, still a lot of information is needed about the key genes that regulate cell division and differentiation, particularly differentiation of mesophyll (M) and bundle sheath (BS) cells. Genetic approaches provide powerful tools for understanding the mechanism of complex processes. We were specifically interested in understanding the differentiation of mesophyll and bundle sheath cells in *Arabidopsis thaliana,* and therefore, we established a forward genetic screen of EMS-induced mutant lines using bundle sheath-labelled reporter genes. As a result, numerous mutants were produced with an altered bundle sheath cell layer. By pursuing a mapping-by-sequencing approach, the genomic segments containing mutated candidate genes were identified (Döring *et al.,* 2019). In this study, we identified putative genes responsible for two EMS-induced mutants, *ebss1*

(*ELEVATED BUNDLE SHEATH CELLS SIGNAL 1*) and *fbss1* (*FAINT BUNDLE SHEATH CELLS SIGNAL 1*), by modification of candidate genes using the CRISPR/Cas9 tool. Screening of Cas9-induced mutant lines revealed that the mutant allele of At2g25970 was responsible for high GFP signals in *ebss1* BS cells and that of At5g04940 for the *fbss1* phenotype.

Keywords, CRISPR/Cas9, ethyl methanesulfonate, bundle sheath cells, *Arabidopsis thaliana,* green fluorescent protein, KH domain, AT5G04940

Introduction

Leaves are the organs of photosynthesis, capturing light energy and converting it into chemical energy via the photosynthetic machinery located in the chloroplast of the leave's chlorenchyma cells. The development of a leaf is a complex mechanism that relies on the precise regulation of cell division and proliferation. The vast majority of research has been done on *Arabidopsis* to identify genes that function in the control of leaf development, such as cell division, differentiation, and expansion. Numerous regulators of leaf development have been identified with various genetic approaches, but the scenario is far from complete due to highly complex growth regulatory networks (Street *et al.,* 2008, Tsukaya, 2013, Kalve *et al.,* 2014, Nikolov *et al.,* 2019). The development of a forward genetic screen allowed us to further identify additional players involved in the leaf regulatory network. Forward genetic approaches have proven to be effective tools to explore biological processes and discover their genetic basis. T-DNA insertions and Ethyl methanesulfonate (EMS) induced mutations are commonly used in forward genetic screening with the model plant *A. thaliana* (Qu and Qin, 2014). EMS-based forward genetic screening is the most robust and frequently used technique because it induces point mutations with a high frequency and is easy to handle as compared to other chemical mutagens (Espina *et al.,* 2018, Addo-Quaye *et al.,* 2017, Wilson-Sanchez *et al.,* 2019). Unlike insertional mutagenesis (T-DNA), EMS mutagenesis results in single nucleotide polymorphism (SNP) that, when occurring in the coding region of a gene, may help us to understand the protein function (Qu and Qin, 2014) by substitution of amino acids or by a generation of novel stops codon (Addo-Quaye *et al.,* 2017).

Understanding the genetic basis of leaf development and anatomy is imperative if one aims to introduce a C_4-like photosynthetic cycle into C_3 crops (Mitchell and Sheehy, 2006, Ermakowa *et al.,* 2020). The high efficiency of C_4 photosynthesis is intimately

associated with the special anatomic feature of the leaf where mesophyll cells form a successive layer around BS cells (V-BS-M-M-BS-V) known as “Kranz anatomy” (Garner *et al.,* 2001, 2008, Lundgern *et al.,* 2014, Sedelnikova *et al.,* 2018, Sage *et al.,* 2014, Gowik and Westhoff, 2011). The photosynthetic reactions are compartmentalized between the two distinct cell types (Wang *et al.,* 2017, Sage, 2004, Sedelnikova *et al.,* 2018). Anatomical studies showed that C_4 bundle sheaths are large in size and enriched with organelles, such as chloroplasts and mitochondria as compared to the C_3 BS (Sage *et al.,* 2014, Döring *et al.,* 2019, Van-Rooijen *et al.,* 2020).

Promoter reporter gene studies have shown that bundle sheath specific/preferential promoters of the genes encoding the P and the T subunits of glycine decarboxylase the Asteracean C_4 species *Flaveria trinervia,* maintain the BS expression in the Brassicacean C_3 species *A. thaliana* (Engelmann *et al.,* 2008, Wiludda *et al.,* 2012, Emmerling, 2018). The other way around, the bundle sheath specific/preferential promoter of the sulfate transporter gene (*SULTR2;2*) of *A. thaliana* was shown to retain BS specificity in the C_4 species *Flaveria bidentis* (Kirschner *et al.,* 2018). These findings demonstrated that the gene regulatory systems of the bundle sheath in these two families share a large degree of conservation, although the two families are separated about 125 million years ago. Because of its easy genetic tractability, *A. thaliana* should therefore be a straightforward system to identify genes affecting the development and function of the bundle sheath in dicots (Westhoff and Gowik, 2010, Döring *et al.,* 2019).

With the goal of finding genes that affect bundle sheath development and/or anatomy, the C_3 model species *A. thaliana* was used in this study. For easy identification of candidate bundle sheath mutants, we labeled the bundle sheath by using the bundle sheath preferential *GLDPA* promoter of *F. trinervia*, driving a reporter gene encoding a chloroplast targeted Green Fluorescent Protein (GFP; Figure 1A) (Döring *et al.,* 2019). We expected that an increase in bundle size/volume or in chloroplast numbers due to a mutation should lead to an enhanced GFP fluorescence. Seeds from homozygous lines of *A. thaliana* carrying the p*GLDPA*$_{Ft}$::TP$_{RbcS}$-*sGFP* reporter gene construct (called the reporter line) were mutagenized with EMS, and stable mutant lines of the M2 generation showing aberrant GFP expression (variation > 30 % compared to reporter line) were chosen for further analysis (Döring *et al.,* 2019). Two

of these mutants, *ebss1* and *fbss1* (Fig 1A) were selected for identifying causative candidates for the *ebss1* and *fbss1* mutant phenotypes, respectively, by SHOREmap backcross analysis (Döring *et al.,* 2019). The identified candidate genes were tested for being responsible for the mutant phenotype by using CRISPR/Cas9 genetic modification. By this approach, At5g04940 encoding a SU(var) 3-9 homolog (Li *et al.,* 2016) was identified as the causative agent of the *fbss1* mutation, while At2g25970 encoding a K homology domain (KH) protein was found to be responsible for the *ebss1* mutant phenotype.

Material and methods

Mutant screening and growth of plants

The mutagenesis of the p*GLDPA*$_{Ft}$::TP$_{RbcS}$-*sGFP* reporter line of *A. thaliana* (ecotype Columbia) with EMS, the screening and isolation of candidate mutants as well as conditions for growing *Arabidopsis* plants have been described by Döring *et al.* 2019.

Preparation of CRISPR/Cas9 constructs

For CRISPR/Cas9 constructs 20bp long single guide RNAs (sgRNAs) were designed from the gene of interest (GOI) containing PAM motif ($GN_{19}NGG$) sequences. The sgRNAs were designed in such a way that they targeted an exon of GOI and preferably no off-target sequence with maximum mismatches number 3 (http://www.rgenome.net/cas-offinder/). Primers (Suppl. Table S1. 1/2 and 3/4 for *AT5G04940*, 9/10 and 31/32 for At2g24610, 25/26 for At2g25220, and 11/12 for At2g25970) were annealed, and the annealed primers were ligated (50 ng PFH6, 1ul T_4 ligase, 2 ul T_4 buffer, 1.16 ng sgRNA, and H2O up to 20 ul) into *Bbs1*-digested plasmid PFH6 (GenBank accession number KY080689 (Hahn *et al.,* 2017). After transformation into competent cell DH5α *E. coli* cells (ThermoFisher Scientific), colonies of recombinant bacteria were selected and the correct insertion of the guide RNA was confirmed by Sanger sequencing (LGC Genomics, Berlin, Germany). sgRNA, including the entire U6-26 promoter-sgRNA cassette was amplified from PFH6 using Phusion PCR polymerase (Suppl. Table S1, 42/44 and 43/45), and after purification on agarose gels, the sgRNA was cloned into *KpnI/HindIII*-digested vector pUB-Cas9 (GenBank accession number KY080691 (Hahn *et al.,* 2017) through Gibson cloning (Gibson *et al.,* 2009). Recombinant colonies were identified by PCR, and plasmid DNA was isolated from positive colonies by using the Qiagen MiniPrep

Kit and verified by sequencing (Hahn *et al.,* 2017).

Transformation of *A. thaliana* with CRISPR/Cas9 constructs and selection of positive transformants.

The CRISPR/Cas9 constructs carrying the sgRNA of the At2g25970*,* At2g25220*,* At2g24610, and At5g04940 genes in vector pUB-Cas9 were transformed into *Agrobacterium tumefaciens* strain AGL1 by electroporation (Lazo *et al.,* 1991). Positively transformed Agrobacterium cells were verified by colony PCR followed by digestion of the extracted plasmids by suitable restriction enzyme and transformed into the reporter line via the floral dipping method (Zhang *et al.,* 2006).

Surface-sterilized seeds of the CRISPR/Cas9- mutated reporter plants were sown on Petri dishes containing half-strength Murashige and Skoog medium containing 50 µg ml^{-1} hygromycin B (H0654, Sigma Aldrich) and hygromycin resistant plants were finally transferred to soil as describe in Döring *et al.* 2019. Whole genomic DNA was isolated from five weeks old T1 plants mutated in At2g24610, At2g25220, At2g25970, or At5g04940, and the presence/absence of Cas9 was assessed with PCR amplification (Suppl. Table S1. 40/41). All T1 plants were self-pollinated and screened in the T2 and T3 generation to get homozygous mutant lines. The mutant screening was initially based on GFP signal followed by PCR amplification of the mutated gene and confirmation by Sanger sequencing.

Mutant screening in the T2 generation and quantification of GFP signal

The first leaf pairs of 17 days old T2 plants of At2g24610*,* At2g25220*,* At2g25970, and AT5G04940 were screened for aberrant GFP expression under a fluorescent binocular microscope (Axio Imager M2m Zeiss, Oberkochen, Germany). After that, leaf genomic DNA was isolated from T2 plants (ten plants for each line) and amplified with gene-specific primers (Suppl. Table S1. 13/14, 27/28, 33/34*,* 5/6) using Phusion High Fidelity DNA polymerase. PCR products were purified using the Qiagen PCR Purification Kit and then sequenced to detect Cas9-induced mutations. Selected homozygous mutant lines were screened again in T3 and T4 generations to confirm the aberrant GFP phenotype compared to the reporter line, *ebss1,* and *fbss1*. The GFP signal intensity was measured from whole leaves using Image J and normalized to the leaf area (Schneider *et al.,* 2012, Döring *et al.,* 2019).

Microscopic analysis of leaf tissue

Leaf tissue was prepared for light microscopy according to Akhani and Khoshravesh (2013). Fully expanded second leaves of 25 days old plants were used, which were grown under optimal light conditions (16 h light and 8 h dark, by 22 °C). Leaf edges and the midvein were removed, and the remaining of the leaf was cut horizontally to use the mid area (3x5 mm) of the leaf. Leaf samples were immediately fixed in the solution containing 1 % glutaraldehyde, 1 paraformaldehyde, and 0.1 M sodium cacodylate. The fixed leaf tissues were washed twice with sodium cacodylate (30 min incubation each) followed by post-fixation in 1.5 % osmium tetraoxide (OsO4) for 3 h. Leaf samples were then rinsed twice (each 30 min) with a fixative solution followed by dehydration in a gradient of ethanol starting from 10 % and ending at 2x100 % (1 h incubation in each step). Tissues were then infiltrated and embedded subsequently in Araldite resin: propylene oxide solutions (v/v 1:3, 1:1, 3:1, and 100 % Araldite) followed by final polymerization for 72 h at 60°C. The resin-embedded blocks were then cut with a microtome to obtain two micro-meter thick sections and stained with toluidine blue.

RNA quantification by quantitative real-time PCR (qPCR)

Total RNA was isolated using RNeasy Plant Mini Kit (Qiagen) from rosette leaf of 25 days old plants grown under standard growth conditions. cDNA was synthesized from 1ug RNA using the Quantitect Reverse Transcription Kit (Qiagen, Hilden, Germany), and the purity and integrity of the cDNA were verified by agarose gel electrophoresis. qPCR was performed with KAPA SYBR FAST qPCR Master Mix Universal (KAPA Biosystem, Roche Sequencing and Life Science) using the ABI7500 Fast Real-Time PCR system following the standard procedure. The gene-specific primers used in the qPCR amplified 120 to 150 bp coding region of the gene (Suppl. Table S1. 17/18, 19/20, 21/22, 23/24, and 47/48). The abundance of actin primers was used as a reference (Suppl. Table S1. 49/50).

Leaf size measurement

The second leaf pairs of 24 days old mutant lines *ebss1*, *At2g25970-1, At2g25970-2, At2g25970-3,* and reporter line were photographed and measured the leaf area with image J (Version 2.0.0-rc-69/1.52p).

Results

The EMS mutant lines *ebss1* and *fbss1* were isolated from the forward genetic screening of *A. thaliana* as explained in Döring *et al.* 2019. Mutant *ebss1* showed an elevated level of signal intensity of the *GFP* reporter gene in bundle sheath and vascular tissue, whereas the *fbss1* showed a remarkably reduced GFP signal compared to the reporter line (Figure 1A). Quantification of GFP signal intensity showed more than a 30 % increase (*ebss1*) or decrease (*fbss1*) expression of the reporter gene (Figure 1B). Applying an allele frequency > 0.9 in SHOREmapping three candidate genes each for *ebss1* (At2g24610, At2g25220, and At2g25970) and *fbss1* (AT5G04940, At5g05930, and At5g03495) were identified that are located on chromosome 2 and 5, respectively.

ELEVATED BUNDLE SHEATH CELLS SIGNAL 1 (ebss1)

Location and effects of the mutations in the candidate genes At2g24610, At2g25220, and At2g25970

All three genes identified as candidate genes responsible for the *ebss1* mutant phenotype were mapped on chromosome II and further narrowed down to between 10 Mb to 12 Mb (Figure 2). In each of the candidate genes, the EMS-induced mutations resulted in a single nucleotide polymorphism (SNP; G to A substitution) within the coding region. In At2g24610, the SNP was located in exon-3, changing the 716th codon from GGA to GAA and resulting in an amino acid substitution of Glycine to Glutamic acid. In At2g25220, the mutation was observed in exon-5, altering the 716th codon from CCT to CTT and causing an amino acid substitution of Proline by Leucine. In At2g25970, the SNP was identified in exon-6 within the codon 1456th codon, converting CAG, the codon for Glutamine into the stop codon UAG.

Analysis of *ebss1* candidate genes by CRISPR/Cas9 mutagenesis

To identify the gene responsible for the *ebss1* mutant phenotype, we generated CRISPR/Cas9-induced mutant lines for At2g24610, At2g25220, and At2g25970 as described in the Materials and methods. The CRISPR/Cas9 vector used in this study encoded sgRNAs that targeted exon-2 of At2g24610, exon-3 of the At2g25220, and exon-1 of At2g25970, respectively (Figure 3A).

As a result, various mutant alleles were detected, ranging from one or two base pair indels to a large deletion. The information regarding the various types of mutations obtained and their effects is shown in figure 3B.

Screening of CRISPR/Cas9 induced mutant alleles of all three candidate genes for *ebss1* in the T3 generation revealed that the GFP fluorescence signals of the mutant alleles At2g24610 and At2g25220 were similar to that of the reporter line, i.e., they did not show the enhanced fluorescence phenotype typical for *ebss1*. In contrast, the mutant alleles of the At2g25970 gene exhibited a high GFP signal phenotype (Figure 4A). To corroborate the finding, first leaves from 50 T3 plants homozygous for mutant alleles of *At2g25970-1*/*2*/*3* were harvested, and GFP fluorescence was quantified. Figure 4B shows that the GFP signal intensity in the CRISPR/Cas9 induced mutant lines (*At2g25970-1*, *At2g25970-2,* and *At2g25970-3*) is about 30% higher than the reporter line and like that of *ebss1*. We concluded from this finding that At2g25970 if mutated, leads to an increased GFP fluorescence and thus causes the *ebss1* phenotype.

At2g25970 encodes a K homology (KH) domain-containing protein

At2g25970 encodes an unknown protein that contains K homology (KH) domains. KH domains were first detected in the heterogeneous nuclear ribonucleoprotein K (hnRNP K) of metazoan (Siomi *et al.,* 1993) and later in other organisms; among them are plants (Lorkovic and Barta, 2002). KH domains function as nucleic acid recognition motifs and are found as single or multiple copies within a protein (Valverde *et al.,* 2008, Nicatro *et al.,* 2015). The At2g25970 protein contains two KH type 1 domains and is closely related to the At1g33680 and At4g10070 proteins of *Arabidopsis* (see Figure 5)

KH domains are considered as nucleic acid recognition motifs in proteins that perform various cellular functions. KH domains are found as single or multiple copies within a protein (Valverde *et al.,* 2008). KH domains are grouped into four classes based on the presence of specific binding motifs, such as PCB-like KH (poly C binding protein), SF1-like KH (splicing factor 1 like), KH-1 superfamily, and KH type 1 (Karlsson *et al.,* 2015, Buckner *et al.,* 2008).

Phenotypic analysis of *At2g25970-1/2/3*

At2g25970 transcript level

To determine whether mutations in At2g25970 affect transcript abundances of the gene, we performed quantitative RT-PCR analysis. As shown in figure 6B, the expression of the At2g25970 gene is highly reduced in the CRISPR/Cas9- induced

mutant lines (*At2g25970-1*, *At2g25970-2,* and *At2g25970-3*) as well as in the *ebss1* line compared to the reporter line.

Growth characteristics

Major differences in the plant growth and height were not observed in the mutant lines; however, in all mutants, the leaf area was found to be increased as compared to the reporter line (Figs. 7B/C). Though this difference in leaf area was not observed in *ebss1* at an early stage of development (Figure 7A) but became prominent in later stages. Detailed examinations revealed that the increasing leaf area was due to an increase in both the length and width of the leaf blade (Figure 7D).

The enlarged leaf area of the mutants *At2g25970-1/2/3* and *ebss1* as compared to the reporter line, raised the question of whether this increase is due to cell enlargement, cell numbers or, a combination of both. To determine the contribution of cell division and expansion to the *At2g25970-1/2/3* mutants, transverse sections from leaf blades of the *At2g25970-1*, *ebss1,* and reporter line were analyzed by light microscopy. In both types of mutants, the vascular bundle was found to be expanded, and concomitantly the number of bundle sheath cells had increased by two to three more cells (Figure 8). These findings suggest that the overall increase in the leaf area is caused by a general increase in the leaf cell numbers due to the expansion of the vascular bundle and the bundle sheath. Since the *GLDPA* promoter-driven GFP gene is active only in the bundle sheath and the vascular bundle. The finding also explains the enhanced GFP fluorescence as compared to the original reporter line.

Trichome density

To investigate whether other leaf features have been altered in the At2g25970 mutant lines, we examined the leaf surface by fluorescent microscopy. Trichome number and leaf area were measured in *At2g25970-1/2/3* and *ebss1*mutant plants as well as in the reporter line plants (Figure 9A), and trichome density was calculated. The results obtained revealed a two-fold increase in trichome density in all At2g25970 mutants, including *ebss1* as compared to the reporter line (Figure 9B).

Flowering time

Visual inspection of the growth behavior of *At2g25970-1/2/3* and *ebss1* mutant plants revealed a delay in flowering time (Figure 10). Bolting started 5-7 days earlier in the plants of the reporter line, and flowering of the mutant plants began about 7-8 days

later (Table 1). The finding suggests that defects in At2g25970 function influence extend the period of growth and thereby causing a delay in floral induction.

FAINT BUNDLE SHEATH CELLS SIGNAL 1 (fbss1)

Location and effects of EMS mutation in the candidate genes At5g05930, AT5G04940 and At5g03495

The candidate genes for *fbss1* (At5g05930, AT5G04940, and At5g03495) are located on chromosome V (Figure 11). The EMS-induced mutations resulted in SNPs (C-T substitution) in the coding region of all three candidate genes. The AT5G04940 gene is intron-less, and the SNP was detected at the 788th codon GGA to GAA, resulting in the exchange of amino acid Glycine to Glutamic acid. In At5g05930, the SNP was observed at exon 2, altering the 260th codon TGC to TAC and causing the substitution of Cysteine by Tyrosine. In At5g03495, the SNP was observed at exon 6 converting the 499th codon GGT to AGT, which leads to the substitution of Glycine to Serine.

CRISPR/Cas9 mutagenesis of one of the candidate genes AT5G04940

AT5G04940 encoded a SU (var)3-9 homologous protein named SUVH1 and was described as an anti-silencing factor because of its inhibitory activity on DNA methylation (Li *et al.,* 2016). Since the *fbss1* phenotype is due to a reduced expression of the GFP reporter gene, it was sensible to speculate that a mutation in AT5G04940 is causing the low GFP signal in the bundle sheath of *fbss1*. Therefore, we generated a CRISPR/Cas9-induced mutant line of AT5G04940 in the reporter line background as explained in Materials and methods.

The three mutant alleles recovered i.e., *At5g04940-1, At5g04940-2, and At5g04940-3*, each had inserted one additional base pair. Mutant alleles *At5g04940-1* showed a 'T' insertion, while *At5g04940-2* and *At5g04940-3* showed 'A' insertions three base pairs upstream of the PAM sequence at the 839 bp coding region (Figure 12A). The resulting one base pair insertion led to premature stop codons. All three mutant alleles showed a reduced GFP phenotype, which is typical for *fbss1*. We concluded, therefore, that the mutation in the AT5G04940 was responsible for the *fbss1* phenotype (Figure 12B) and did not pursue to further generate CRISPR/Cas9-induced mutation in At5g05930 and At5g03495.

Discussion

We aimed at identifying genes for bundle sheath anatomy by pursuing a forward genetic approach using *A. thaliana* as a genetic model system and ethyl

methanesulfonate as a mutagen. To easily recognize putative bundle sheath defective mutants, our genetic screen was based on a reporter line containing a chloroplast targeted GFP gene that was driven by a bundle sheath preferential promoter. Using deviations in GFP fluorescence as a proxy for alteration in bundle sheath size or photosynthetic activity, we were able to identify a number of stable mutants with an elevated or a faint GFP fluorescence phenotype, with many of them showing defects in bundle sheath anatomy (Döring *et al.,* 2019). Pursuing a SHOREmap approach with nine selected mutants, we could map candidate genes in single genomic regions (Döring *et al.,* 2019). For the study presented in this report, we have chosen the mutant lines *ebss1* and *fbss1* to demonstrate that the mutant candidate genes identified by SNP calling in the mapping intervals can be successfully scrutinized by using CRISPR/Cas9 technology, finally resulting in the molecular identification of the responsible genes.

Mutant *ebss1* exhibited a high GFP signal in the bundle sheath and vasculature, while *fbss1* showed a low GFP signal in the bundle sheath compared to the reporter line (Figure1A). SHOREmapping revealed three candidate genes, each for *ebss1* and *fbss1*. To identify the mutated genes responsible for the GFP phenotype in *ebss1* and *fbss1*, respectively, we analyzed three candidate genes for *ebss1* (At2g24610, At2g25220, and At2g25970) and one candidate gene for *fbss1* (At5g04940). The use of CRISPR/Cas9 technology allowed us to generate three independent mutant alleles for each candidate gene except for At2g24610, where only one mutant allele was obtained (Figs. 3B and 12A). By quantitative phenotyping of plants homozygous for the various mutant alleles, we identified At2g25970 as the causative gene for the high GFP fluorescence phenotype of *ebss1* (Figure 4A) and At5g04940 for the low GFP fluorescence phenotype of *fbss1* (Figure 12B). At2g25970 encodes a KH domain-containing protein (Lorkovic and Barta, 2002), while At5g04940, named *SUVH1*, a SET domain protein involved in the epigenetic control of gene expression by regulating DNA methylation (Li *et al.,* 2016). Transcripts from both At2g25970 and At5g04940 are found to accumulate ubiquitously in both time and space (Klepikova *et al.,* 2016, CATd.b http://urgv.evry.inra.fr/cgi-bin/projects/CATdb/catdb_projects_Ath.pl), suggesting that the two genes have a general function.

At2g25970 -a KH domain protein and its putative target

At2g25970 encodes one of the 26 KH domain-containing proteins present in the *A. thaliana* genome. The protein is 80-90% conserved in Brassicacean species, a sequence conservation to homologues from monocot species is mainly restricted to the KH-domain (Figure 13)

Five of the KH-domain proteins of *A. thaliana*, i.e., HEN4 (Cheng *et al.,* 2003), BTR1 (Fujisaki and Ishikawa, 2008), PEPPER (Ripoll *et al.,* 2006), FLK (Lim *et al.,* 2004) and RCF3 (Karlsson *et al.,* 2015), have been functionally characterized, and all turned out to be multi-functional. Each of these proteins had multiple KH-domain, which may explain their multi-functionality. RCF3 (At5g53060) for example, has two PCBP-like KH-domian, two KH-1 superfamily KH-domains, and one KH type-1 KH-domain, and is involved in jasmonate signaling (Thatcher *et al.,* 2015), miRNA biogenesis (Karlsson *et al.,* 2015), pre-mRNA splicing (Cheng *et al.* 2003), and regulation of abiotic signaling processes (Jeong *et al.,* 2013).

The presence of multiple domains for RNA binding and protein-protein interactions (Mackereth and Sattler, 2012, Gronland and Ramos, 2017, Ottoz and Berchowitz, 2020) and the versatility of these proteins in recognizing both RNA and DNA (Hudson and Ortlund, 2014, Debaize and Troadec, 2019) make KH-domain proteins ideal candidates for integrators of and in regulatory networks (Pawson and Nash, 2003, Csizmok *et al.,* 2016, Alvariz *et al.,* 2021). Similar to these five KH-domain proteins, At2g25970 turned out to be multi-functional. In At2g25970 mutants, the *GLDPA* promoter-driven GFP reporter gene is more strongly expressed, the sizes of leaves and the density of trichomes increase, and the flowering date is delayed. The rise in leaf area and concomitantly in GFP fluorescence appears to be caused by an increased cell number in the bundle sheath and the vasculature suggesting that At2g25970 is involved in the regulation of cell division and/or the cell cycle.

In humans, the KH-domain protein FUBP1 (Debaize and Troadec, 2019) has been shown to be involved in cell proliferation by activating c-Myc gene expression (Duncan *et al.,* 1994). In plants, Myc-like, i.e., basic helix-loop-helix (bHLH) proteins do not seem to play a major role in regulating cell division and /or cell cycle. In contrast, Myb proteins, their frequent interaction partners (Millard *et al.,* 2019, 2021), are involved in cell cycle regulation (Xie *et al.,* 2010, Verkruysse *et al.,* 2020). Interestingly a Myc/Myb regulatory module has been shown to be involved in bundle sheath specific gene expression in *Arabidopsis* (Dickinson *et al.,* 2020).

Co-expression analyses of transcriptome data (https://www.michalopoulos.net/act/; https://atted.jp) support a possible role of At2g25970 in cell cycle regulation. Fluctuation in the expression of At2g25970 transcript has been observed during the G1 phase of the cell cycle (Menges *et al.,* 2002), and in addition, they co-accumulate with CDKC;2 at RNA levels. CDKC;2 is a cyclin-dependent protein kinase that phosphorylates the carboxyterminal domain of large subunit of RNA polymerase II and is part of the so-called transcript elongation factor P-TEFb (Antosz *et al.,* 2017). These transcript elongation factors attached to the carboxyterminal domain and are required for transcript elongation but are also involved in co-transcriptional pre-mRNA processing (Harlen and Churchman, 2017). Zhao *et al.* (2017) found that the loss of *CDKC;2* function leads to an increase in lateral organ size due to a rise in cell division activity, indicating that *CDKC;2* acts as a negative regulator of cell division. A similar phenotype, i.e., broader leaves and an increased number of bundle sheath and vascular cells, is also true for *At2g250970* mutant lines, suggesting that the At2g25970 protein is functionally associated with *CDKC;2* activity. Along these lines, *CDKC;2* transcript abundance was found to be substantially reduced in the *At2g25970* mutants as compared to the reporter line (Fig 14B).

CDKC;2 as the target of *At2g25970* action may also explain the floral retardation observed for the mutant lines. Floral induction is a major developmental transition in plants and is regulated by various external and internal cues (Amasino, 2010). The floral repressor *FLOWERING LOCUS C* (*FLC*) plays a central role in floral induction and is the hub on which the autonomous and the vernalization pathways converge (Whittakar and Dean, 2017). The autonomous pathway is comprised of a variety of factors involved in RNA processing, among them the KH-domain protein FLK and chromatin modification that together repress *FLC* activity. A long antisense RNAs transcribed from the 3' end into the *FLC* locus is processed, with the help of RNA processing components of the autonomous pathway, to various alternatively spliced and polyadenylated *FLC* antisense RNA, which are collectively called *COOLAIR* and that are involved in repressing *FLC* transcription by chromatin silencing (Wu *et al.,* 2020).

COOLAIR transcription is promoted by CDKC;2 activity (Wang *et al.,* 2014), and the loss of *CDKC;2* function, therefore, leads to a delay in flowering (Cui *et al.,* 2007, Zhao *et al.,* 2017) (Figure 14A). Quantification of transcript amounts from both

CDKC;2 and *FLK* revealed that both transcript levels were significantly reduced in *At2g25970* mutants (Figure 14B and C). In contrast, but as to expected, *FLC* transcript abundance was higher in the mutants as compared to the reporter line (Figure 14D).

Taken together, the available evidence suggests that the mutant phenotypes are brought about via *CDKC;2,* which in turn is functionally connected to *At2g25970* activity. It is completely unclear at present how this functional connection could be achieved mechanistically. Quite recently, another KH-domain protein encoded by *FLOWERING LOCUS Y* (At1g33680) has been reported of delaying flowering when mutated (Dai *et al.,* 2020) suggesting redundancy and overlapping functions in this family of RNA-binding proteins.

SUVH1- an anti-silencing factor of the *GLDPA* promotor

SUVH1 (AT5G04940*)* is one of the ten genes in the *Arabidopsis* genome encoding homologue of *Drosophila* SU(var) 3-9 (Baumbusch *et al.,* 2001). The *Arabidopsis* gene was functionally discovered forward genetic EMS screen as a factor promoting the expression of a luciferase reporter gene driven by 35S promoter. It was found that the luciferase expression was remarkably reduced in a *suvh1* mutant line, and it was further shown that genes whose promoter regions are subject to DNA methylation were also affected. This study adds the *GLDPA* promoter of *F. trinervia* when inserted into the *Arabidopsis* genome to the list of promoters that can undergo DNA methylation. It should be noted in this context that the GLDPA promoter is not a simple promoter but consists of two tandem promoters. These tandem promoters interact with each other in a complex manner and, moreover, post-transcriptional processes via nonsense-mediated mRNA decay are intimately intertwined with the transcriptional output of this double promoter (Wiludda *et al.,* 2012). SUVH1 and three DNAJ homologs (SDJ1, SDJ2, and SDJ3) form a complex that is required for the expression of a subset of promoter-methylated genes (Zhao *et al.,* 2019).

Outlook and conclusion

This study was undertaken as part of an endeavor to identify novel regulators of bundle sheath anatomy (Döring *et al.,* 2019). The two mutant candidates selected after SHOREmapping (Schneeberger *et al.,* 2009) could be identified and verified at the gene level by systematically scrutinizing mutant gene candidates in the mapping interval. From the technical point of view our strategy was therefore successful.

However, the two genes identified as being responsible for the *ebss1* and *fbss1* mutant phenotype i.e., *At2g25970* and *At5g04940*, cannot be considered as novel regulators of bundle sheath anatomy, reinforcing our conclusions based on the phenotypic analysis of all the mutants recovered in the screen (Döring *et al.,* 2019). This raises the question of whether the reporter gene designed was specific enough for the intended goal. The *GLDPA* promoter of *F. trinervia* driving the gene encoding a chloroplast- targeted GFP is active in the bundle sheath cells but also in the vasculature, i.e., the promoter lacks the desired specificity necessary for a proxy-based screen. Consequently, the majority of the mutants identified have an increased bundle sheath volume because of an enlargement of the vasculature, including *ebss1*. The second lesson learned concerns the architecture of the promoter. As discussed above, the *GLDPA* promoter is complex in its regulatory properties, and the *fbss1* mutant and the underlying gene may be indicative of this fact. It is therefore imperative to use the most simple promoter with respect to its regulatory wiring that is available. To conclude, reporter gene-based forward genetic screens using chemical mutagenesis with *Arabidopsis thaliana* may be powerful tools (Page and Grossniklaus, 2002), they are technically relatively easy, and the mutated genes can be identified and verified via CRISPR/Cas9 technology. However, the basis of such a genetic screen, the specificity of the reporter gene in relation to the biological process to be studied, is of paramount importance and has to be reflected on with the greatest care possible.

Acknowledgement

We thank Dr. Florian Döring for advice and providing EMS mutant lines. We are indebted to Dr. Udo Gowik who unfortunately passed away ultimately for help in the analysis of the genotyping by sequencing data. Work reported in this study was supported by the European Union through the 3to4 project - Converting C3 to C4 Photosynthesis for Sustainable Agriculture, the Deutsche Forschungsgemeinschaft through the Cluster of Excellence on Plant Sciences ‘From Complex Traits towards Synthetic Modules’ and by institutional funds of Heinrich Heine University to PW. Z.B thanks the German Academic Exchange Service (DAAD) and the Schlumberger Foundation for stipends.

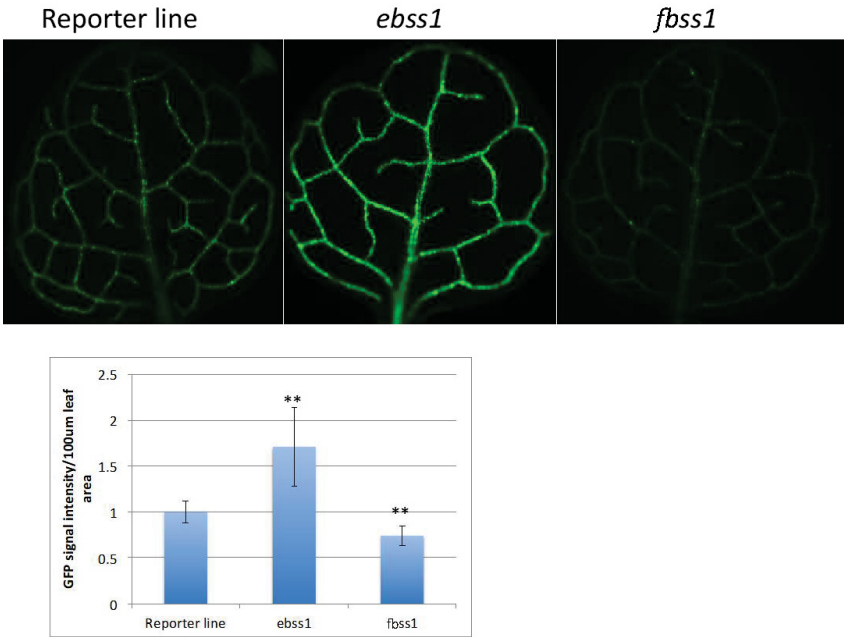


Figure 1. Fluorescent characteristics of *ebss1* and *fbss1*

(A) GFP fluorescence images of first leaf of 3 weeks old EMS lines *ebss1*, *fbss1,* and reporter line. (B) Relative changes of the GFP signal intensity in EMS lines *ebss1* and *fbss1* compared to the reporter line. The relative signal intensity was measured from the whole leaves of 15 plants and normalized with the leaf area. The data shows mean ±SD from 15 individuals, * *indicates t-test P< .001

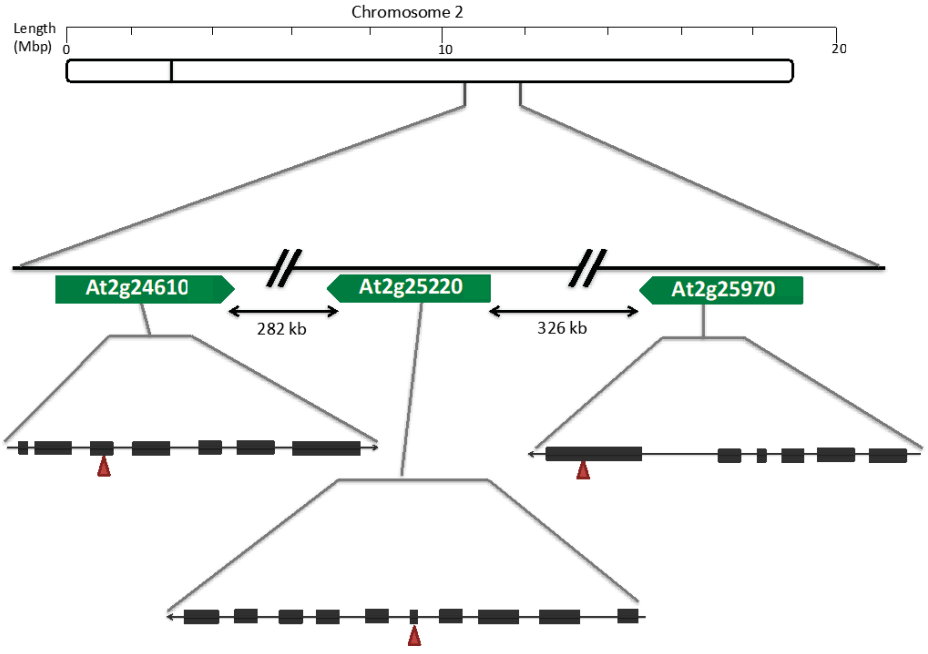


Figure 2. Mapping locus of candidate genes for EMS mutant line *ebss1* on chromosome II.

Location of *ebss1* candidate genes on the chromosome. Red triangle denoted SNP region. Allelic frequencies (AF) for all SNP were isolated from whole genomic sequencing of the reporter line and backcrossed F2 mutants. Gene showing AF > 0.9 were selected as candidate genes.

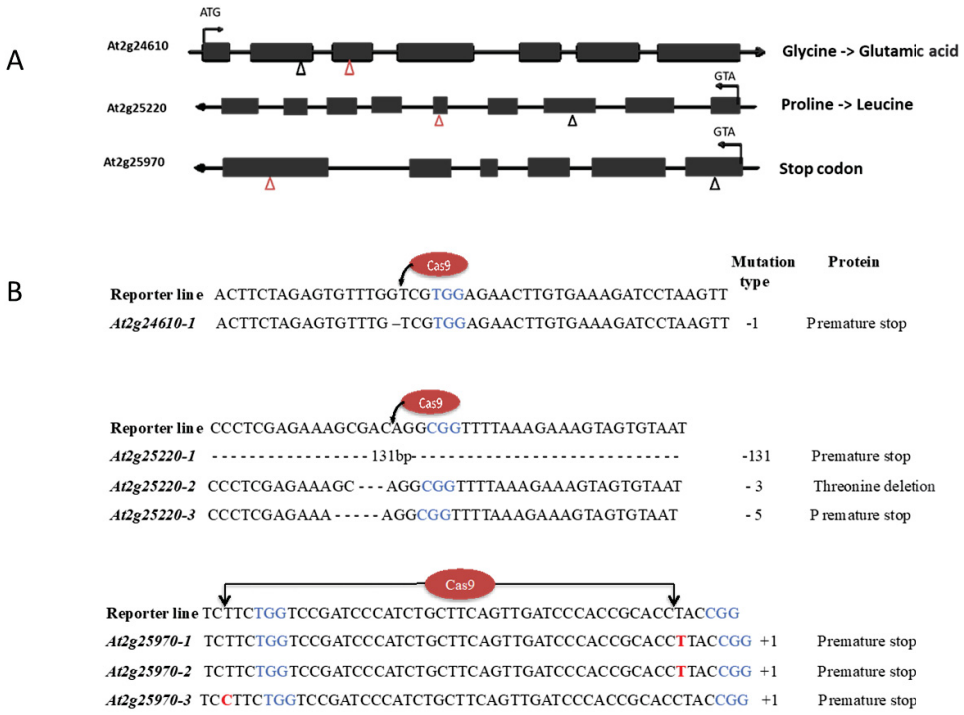


Figure 3. EMS and CRISPR/Cas9 induced mutation in the *ebss1* candidate genes

(A) Location of the EMS and CRISPR/Cas9-induced mutation in At2g24610, At2g25220, and At2g25970. The red triangle indicates the EMS target region, and the black triangle indicates CRISPR/Cas9 target region. (B) CRISPR/Cas9-induced sequence changes and their effects on protein level in *ebss1* candidate genes. Cas9 target region is marked with the arrow and PAM motif in light blue; deletions are indicated by dashes and insertions by red letters

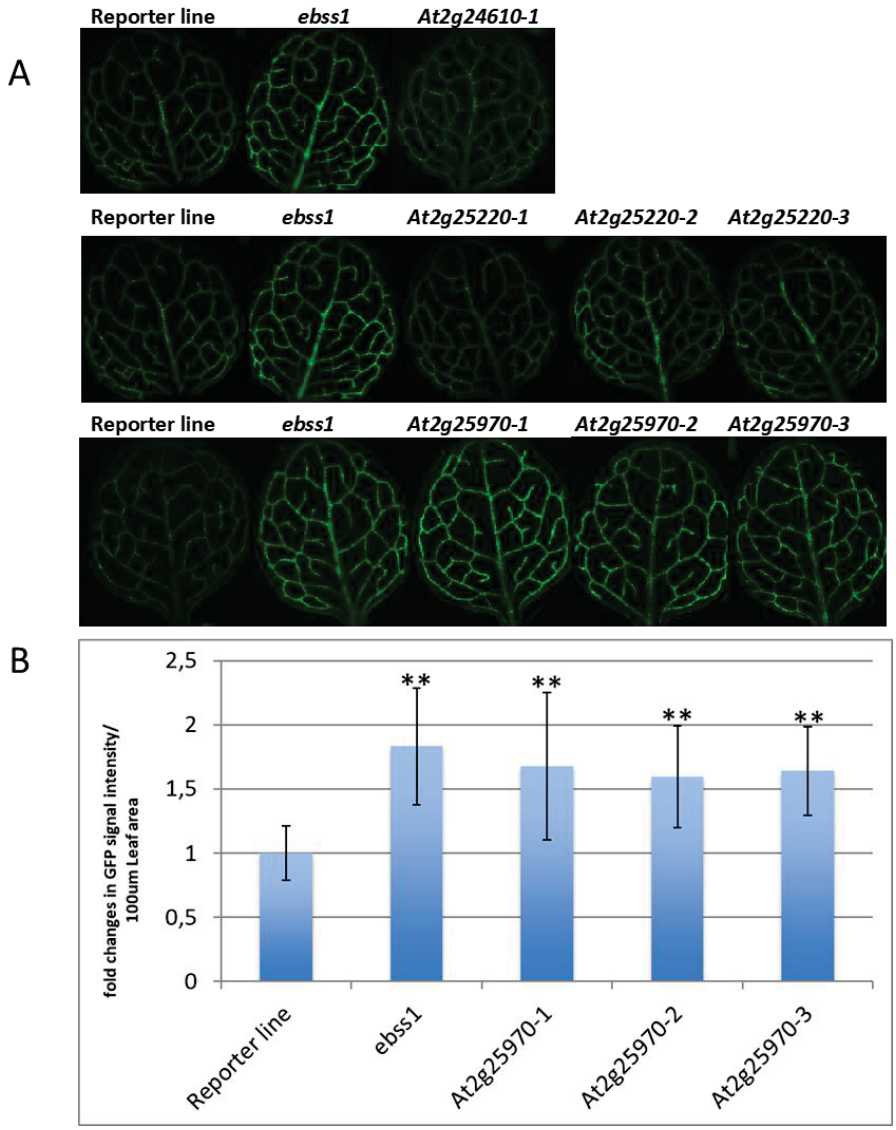


Figure 4. GFP fluorescent phenotype of CRISPR/Cas9-induced mutation in At2g24610, At2g25220 and At2g25970

(A) Leaf images of GFP fluorescence of reporter line, *ebss1* and Cas9-induced mutant alleles of At2g24610, At2g25220, and At2g25970. (B) Quantification of GFP fluorescence of reporter line, *ebss1* and various mutant alleles of AT2g25970. The relative signal intensity was measured from the first leaves of 17 days old plants, normalized with the leaf area, and compared with the reporter line. The data are shown as mean± standard deviation from 50 plants with standard t-test ** indicate P< .001.

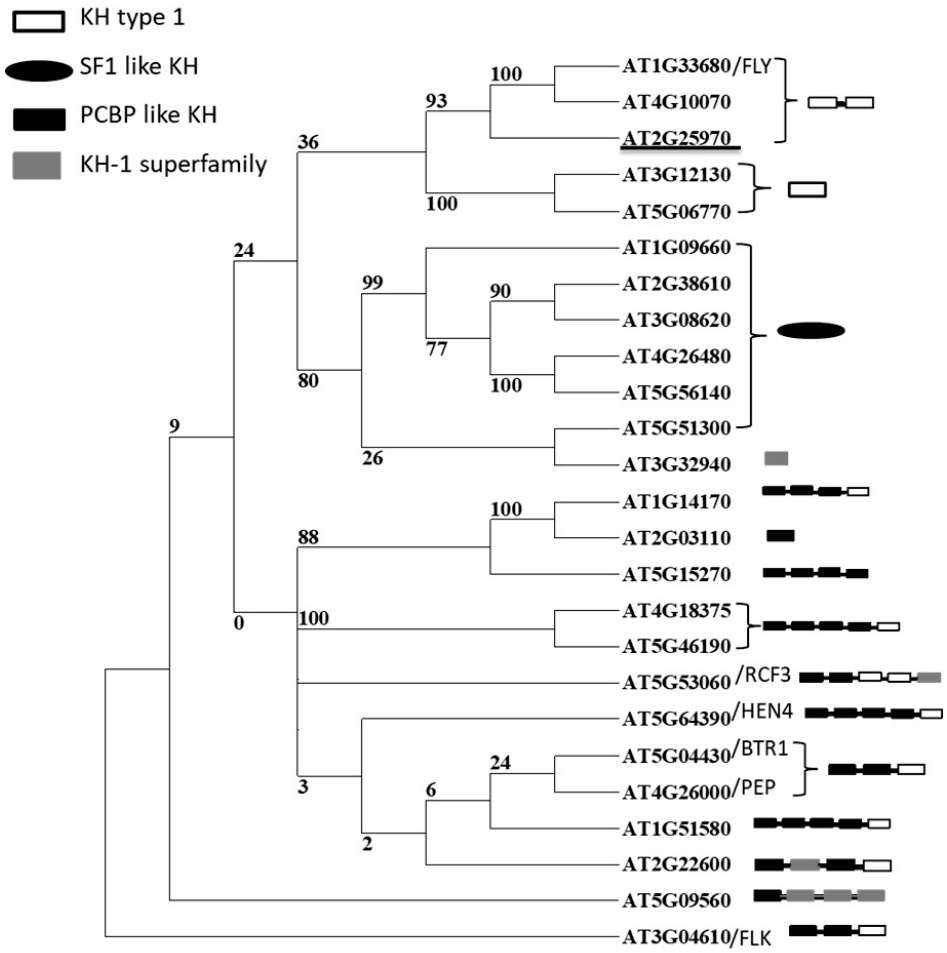


Figure 5. Cladogram of 26 KH domain-containing proteins in *A. thaliana*

Gene IDs of KH domain-containing proteins were taken from Lorkovic and Barta, 2002 and used these IDs to get amino acid sequences from the tair website. Full-length amino acid sequences were aligned using an online tool for multiple sequence alignment, MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/). The ClustalW formatted alignments were loaded into Seasaw version 5 to generate a phylogenetic tree using Parsimony and bootstrapping (100X). The functional domains were analyzed using National Center for Biotechnology Information (NCBI) conserved domain search engine (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

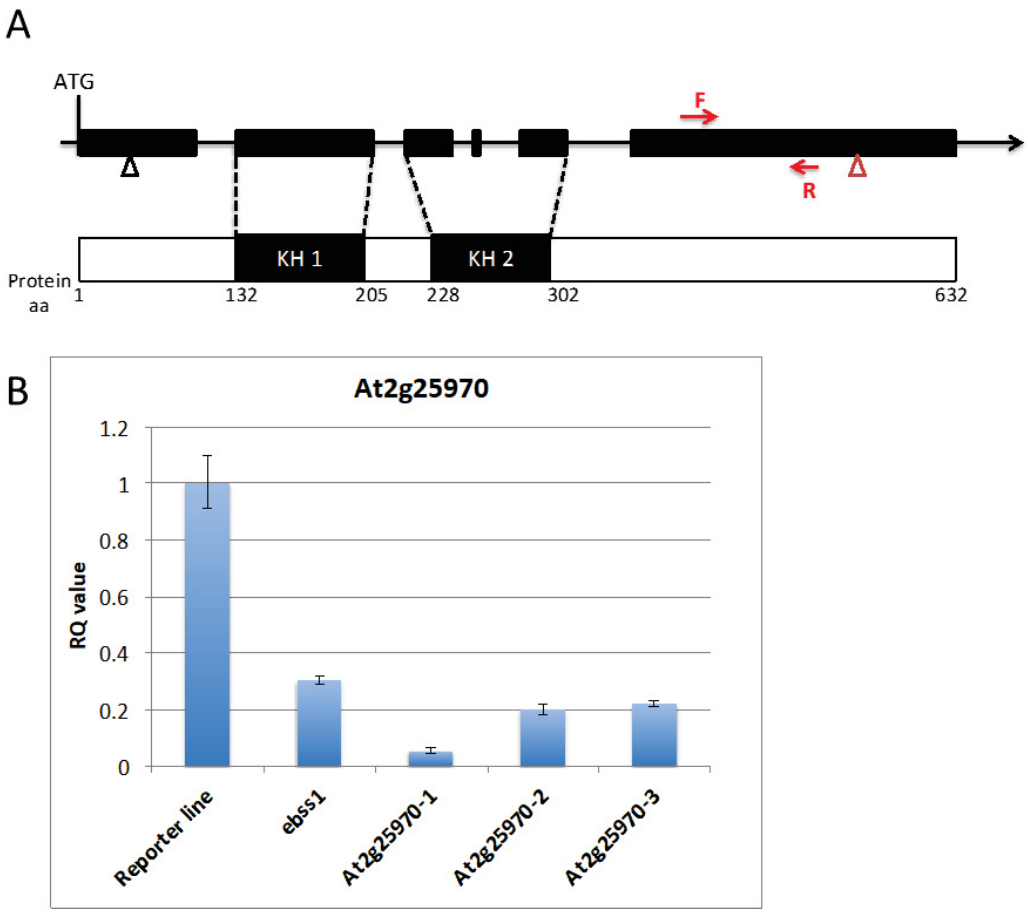


Figure 6. Characterization of At2g25970 gene

(A) Schematic diagram of At2g25970 gene with EMS (red triangle) and Cas9 target sites (black triangle). Filled boxes represent exons. The red arrows on exon-6 note the forward (F) and reverse (R) primers used to detect At2g25970 transcript level in reporter and mutant lines. The KH1 and KH2 in the diagram depicting K-homology domain region in the At2g25970 encoded protein (B) Transcript abundance of At2g25970 in the reporter and mutant lines from 25 days old seedlings.

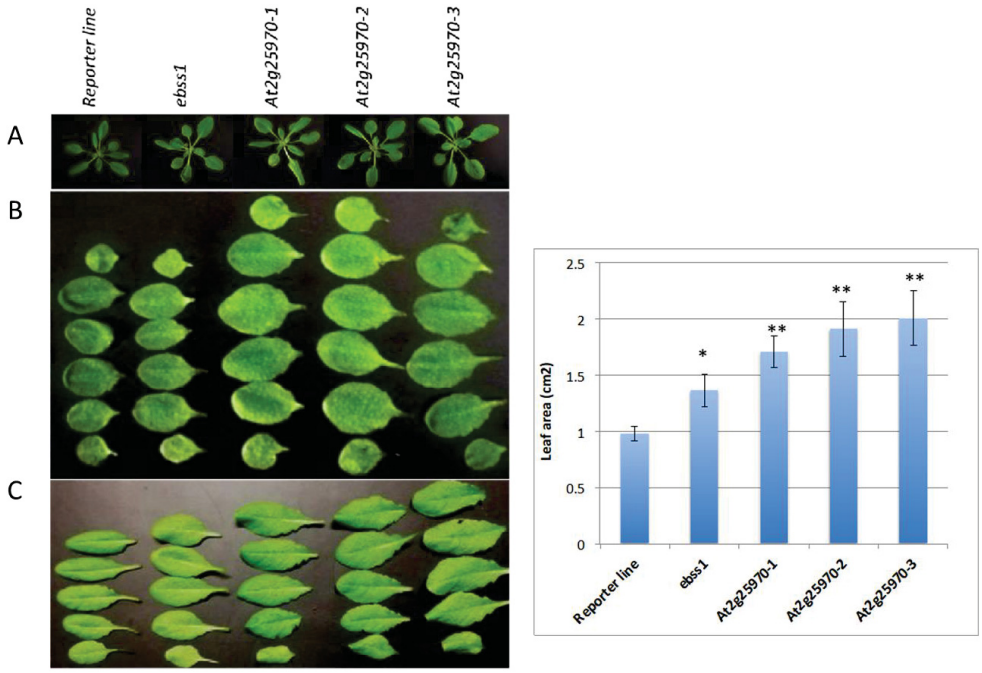


Figure 7. Growth phenotype of reporter and mutant lines (*ebss1, At2g25970-1/2/3*) at various ages

(A) Morphology of 28 days old *Arabidopsis* rosette (B) Morphology of the first, second and third leaf pairs of the reference line, *ebss1* and *At2g25970-1/2/3.* (C) Morphology of second to the sixth leaf of 36 days old plants. (D) The differences between leaf area of 24 days old reference and mutant lines. The leaf area was measured from 10 individuals each. Data are shown as a mean± standard deviation (SD) for each genotype, t-test: *P < .05 ** P< .001

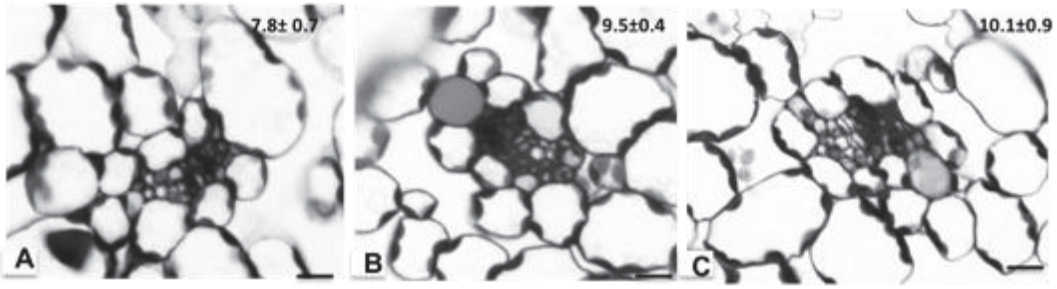


Figure 8. Anatomical analysis of bundle sheath cells in the third vein order of 25 days old plant

Microscopic images of (A) reference line (B) EMS line (*ebss1*) and (C) CRISPR/Cas9 induced mutant lines (*At2g25970-1*). The data showed a mean value of five individuals each. The mean value ± SD of bundle sheath numbers is depicted in the upper right corner of the respective line. Scale bar 10um.

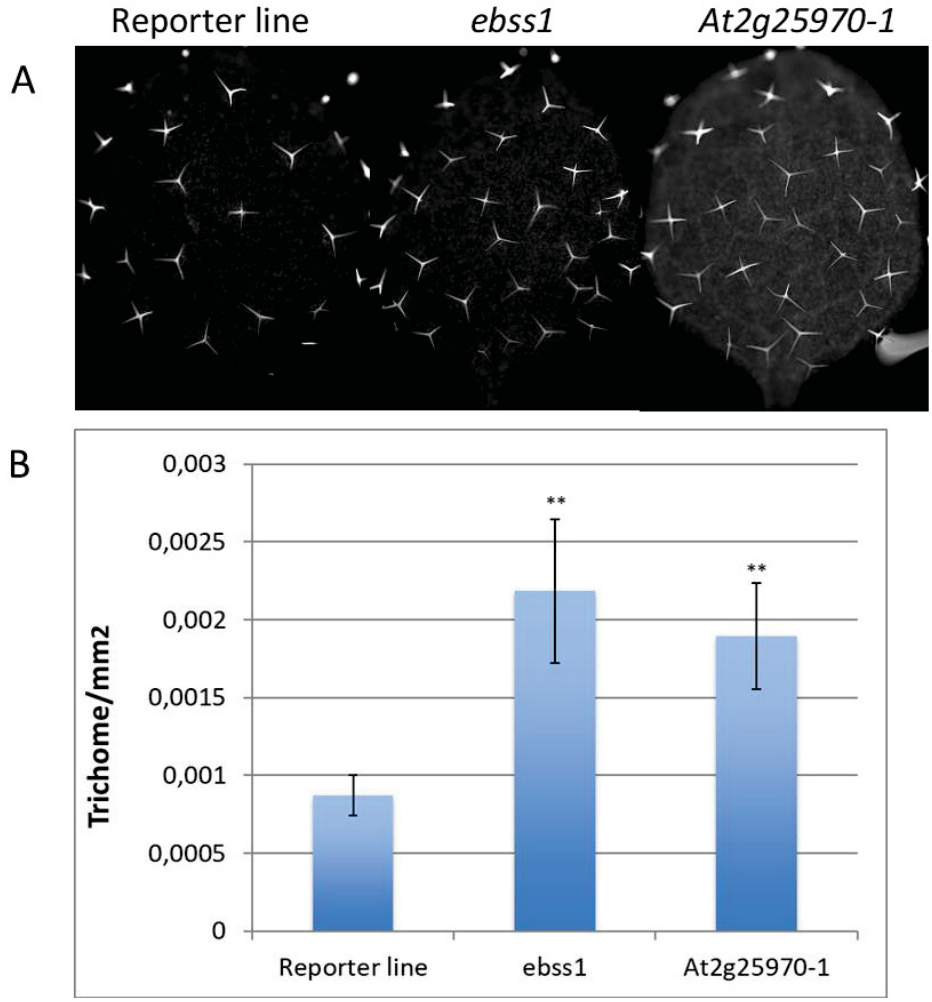


Figure 9. Analysis of trichomes density in *ebss1 and At2g25970-1* compared to the reporter line

(A) Trichome numbers in the first leaf of 17 days old reference and mutant lines, *ebss1* and *At2g25970-1*. (B) Quantification of trichome numbers from the first leaf pairs of reference and mutant lines. Trichomes were counted from 20 individuals. ** Indicate t-test P < .001

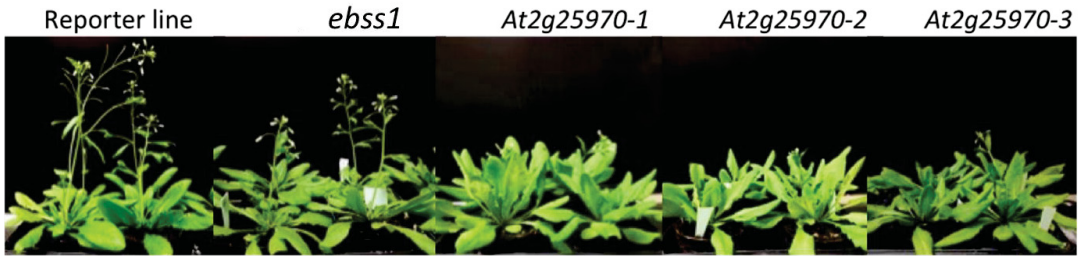


Figure 10. Analysis of flowering time control of EMS line (*ebss1*) and Cas9 induced mutant lines (*At2g25970-1/2/3*) in comparison with the reporter line.

The flowering phenotype of 36 days old mutant lines (*ebss1, At2g25970-1/2/3)* as compared to the reporter line

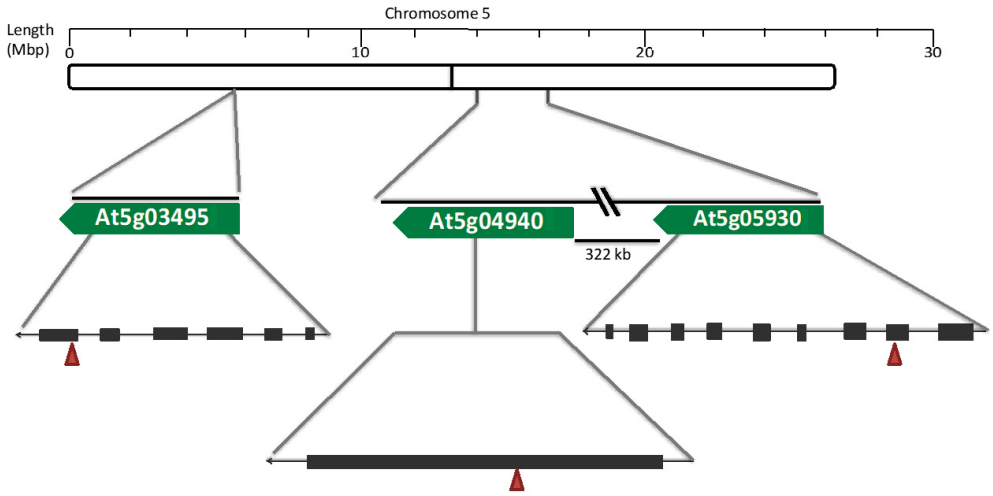


Figure 11. Mapping locus of candidate genes for *fbss1* mutant lines on chromosome V.

Location of candidate genes for *fbss1* on chromosome V with SNP region (red triangle). Allelic frequencies (AF) for all SNP were isolated from whole genomic sequencing of reference line and backcrossed F2 mutants. Gene showing AF >.9 were selected as candidate genes.

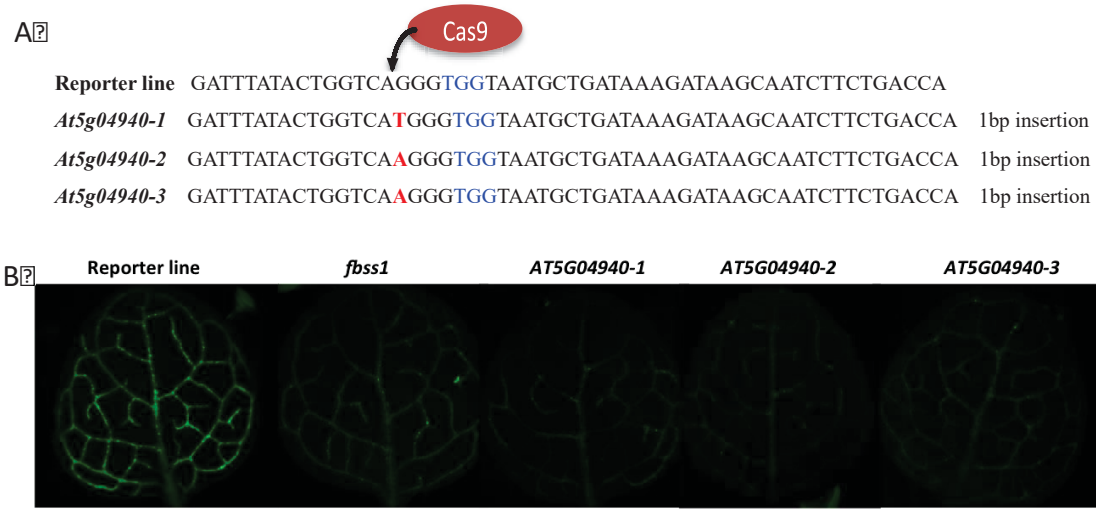


Figure 12. The target site of Cas9 in *fbss1* candidate gene At5g04940 and the GFP fluorescent phenotype of the resulting mutant line

(A) CRISPR/Cas9-induced sequence changes in *fbss1* candidate genes. Cas9 target region is marked with the arrow and PAM motif in light blue, and red letters indicates

nucleotide insertions. (B) Leaf images of GFP fluorescence of reporter line, *fbss1* and Cas9-induced mutant alleles of *At5g04940-1, At5g04940-2,* and *At5g04940-3.*

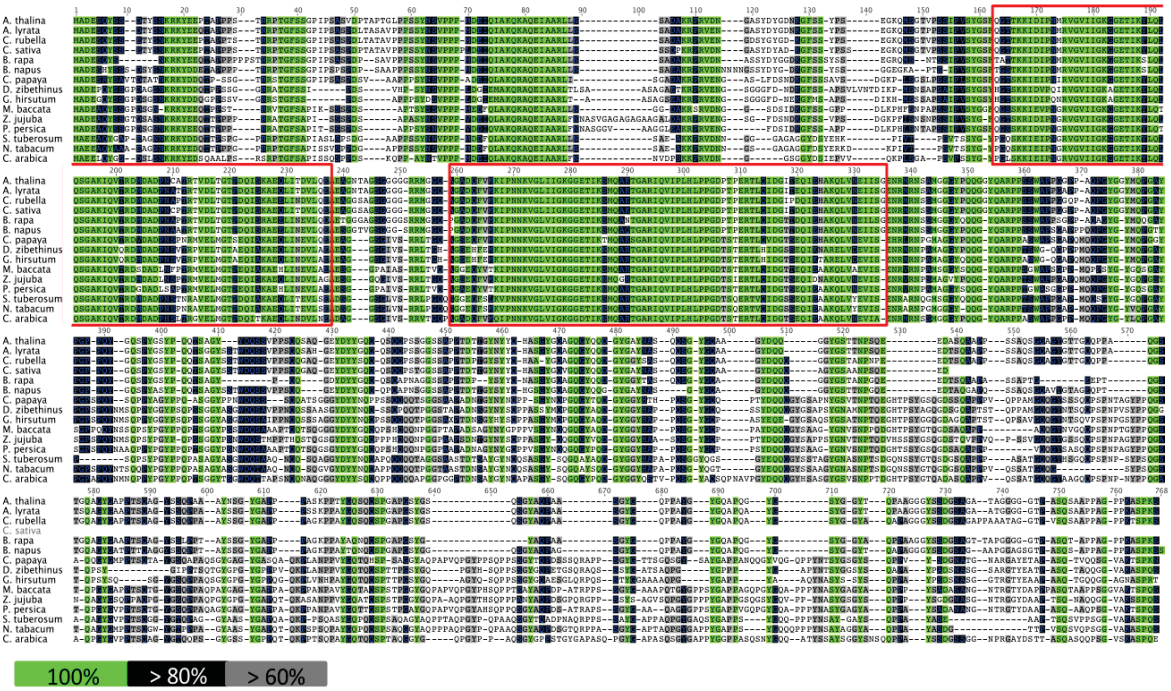


Figure 13. Sequence conservation of KH domain-containing protein (At2g25970) and its orthologs in selected plant species

Sequence alignment illustrating the evolutionary conservation of amino acid sequences of KH domain-containing protein of *Arabidopsis* and its orthologous in other plant species. The sequence conservation was mainly observed in the KH domain region (red box). Amino acid residue in green colors showed 100% sequence similarity, while in black and gray color showed greater than 80% and 60%, respectively. We got orthologs sequence from protein blast in NCBI and confirm the orthologs with reciprocal blasts.

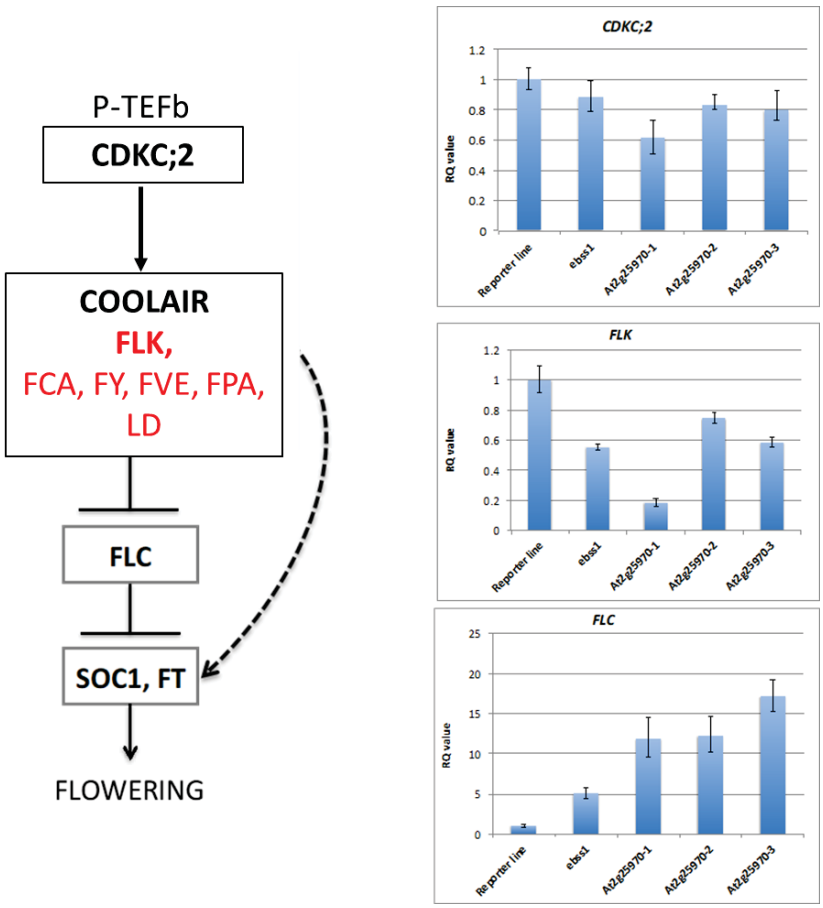


Figure 14. Impact of At2g25970 mutation on CDKC;2 transcript level and CDKC;2 regulated flowering genes

(A) Schematic pathway of genes controlling flowering time in *A. thaliana*. This pathway is modified from Wang *et al.* (2014) and Henderson and Dean (2004). (B) Relative quantification of transcript abundance of *CDKC;2* from 25 days old rosette leaf of *At2g25970* mutant alleles. (C) Transcript abundance of flowering regulator gene *FLK* and (D) floral repressor gene *FLC* in mutant lines *ebss1 and At22g25970-1/2/3* compared to the reporter line.

Table 1. Flowering time of EMS and CRISPR/Cas9 induced mutant lines compared to the reference

Genotype	Rosette leaves at bolting (n)	Time to bolting (d)	Time to Flowering (d)
Reporter line	13.5 ± 1.2	27.9 ± 1.2	30.9 ± 1.8
ebss1	13.6 ± 1.1	28.9 ± 1.5	32.5 ± 2.0*
At2g25970-1	16.4 ± 1.1	35.3 ± 2.7	38.2 ± 2.1**
At2g25970-2	15.0 ± 2.4	32.2 ± 3.3	35.8 ± 3.2**
At2g25970-3	15.4 ± 1.2	34.0 ± 2.6	37.9 ± 2.7**
n= 20	ttest * P< .05	**P< .001	

Rosette leaves were counted at the time of bolting. Bolting time was counted with the appearance of 1 cm long inflorescence stalk and flowering time was scored as the day when first flower opened. Flowering time were observed in 20 individual values are expressed in mean± SD

Supplementary data

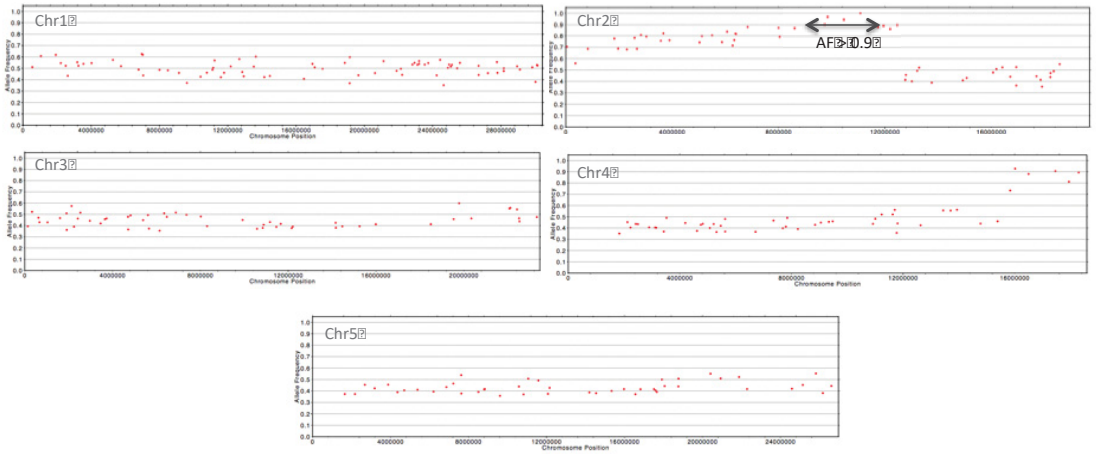


Figure S1. Allelic frequency of mutant line *ebss1*

Allelic frequencies (AF) for all SNP were isolated from whole genomic sequencing of reference line and backcrossed F2 mutants. Gene showing AF >.9 were selected as candidate genes and genomic region with AF>0.9 are highlighted with arrow in the diagram

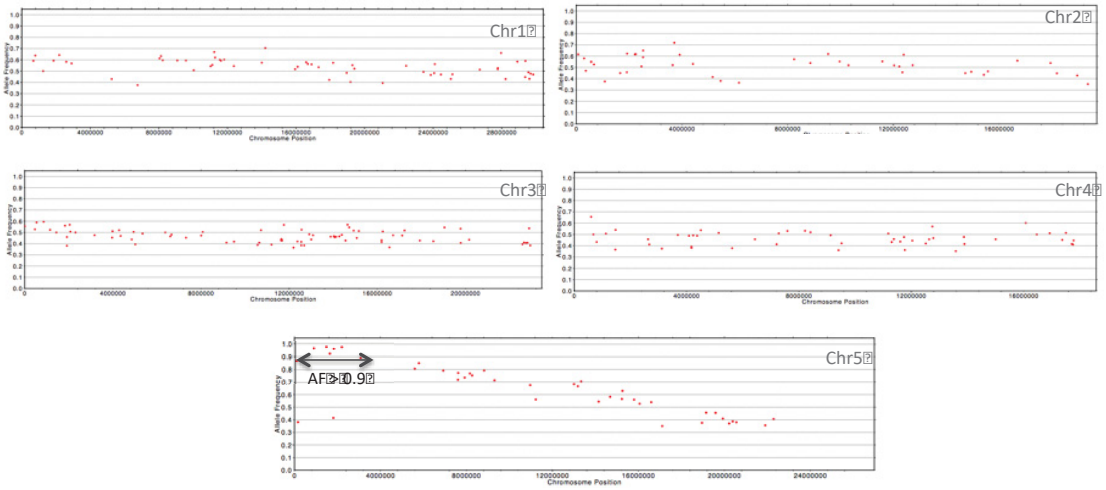


Figure S2. Allelic frequency of mutant line *fbss1*

Allelic frequencies (AF) for all SNP were isolated from whole genomic sequencing of reference line and backcrossed F2 mutants. Gene showing AF >.9 were selected as candidate genes and genomic region with AF>0.9 are highlighted with arrow in the diagram.

Table S1. List of oligonucleotide used in this study

Sr.	Primer name	Sequence 5´→3´	Orientation
	ZB1→ At5g04940		
	ZB2→ At2g25970		
	ZB3→ At2g25220		
	ZB4→ At2g24610		
1	ZB1-sgRNA1-F	TTCGTTGATTTATACTGGTCAGGG	F
2	ZB1-sgRNA-1-R	AAACCCCTGACCAGTATAAATCAA	R
3	ZB1-sgRNA-2-F	TTCGTGTGCCAACTTATGCAAACC	F
4	ZB1-sgRNA-2-R	AAACGGTTTGCATAAGTTGGCACA	R
5	ZB1-sgRNA-amp-F	TCCTCCTGGGTTCTCATCGT	F
6	ZB1-sgRNA-amp-R	ATAGCATCCCATGACCGCAA	R
7	ZB1-sgRNA-1-seq	TGGTGTTGAGTGTTCTTATG	
8	ZB1-sgRNA-1-seq	AATAAGGGGCTTGAAAGAGGCT	
9	ZB2-sgRNA-1-F	TTCGAGACCTACCGGCTTCTCTTC	F
10	ZB2-sgRNA-1-R	AAACGAAGAGAAGCCGGTAGGTCT	R
11	ZB2-sgRNA-2-F	TTCGGTTGATCCCACCGCACCTAC	F
12	ZB2-sgRNA-2-R	AAACGTAGGTGCGGTGGGATCAAC	R
13	ZB2-sgRNA-amp-F	CTTGATGGAGAAGATCAAGAGAAT	F
14	ZB2-sgRNA-amp-R	CTGAATCTTAGCTCCAGACTGA	R
15	ZB2-EMS-F	CGCGCAAGGTGAGTATGATTAT	F
16	ZB2-EMS-R	CAGCACTCTGTGAAGCTGGT	R
17	ZB2-qPCR-F	CAAGCAGTGGTGGTAGCTCA	F
18	ZB2-qPCR-R	CCATATCCCGATTGCTGCGA	R
19	CDKC;2-qPCR-F	TCCAGGTCGAGATAGGGATGA	F
20	CDKC;2-qPCR-R	CTCAGTCCAGGACGATCAGC	R
21	SF1-qPCR-F	GAACGGTGTTAGCAAGACGC	F
22	SF1-qPCR-R	CCTTCGGCGCTAGGGTTATT	R
23	FLK-qPCR-F	CAGTTGCAGGACCAGGCTAA	F
24	FLK-qPCR-R	TTGTACATCACCTTGCGCCT	R
25	ZB3-sgRNA-F	TTCGACCCTCGAGAAAGCGACAGG	F

Sr.	Primer name	Sequence 5′→3′	Orientation
26	ZB3-sgRNA-R	AAACCCTGTCGCTTTCTCGAGGGT	R
27	ZB3-sgRNA-amp-F	TATCGCAAGAACCAATCTCCAA	F
28	ZB3-sgRNA-amp-R	CGACAATGCTCATGGAGATACTC	R
29	ZB3-EMS-F	TGTGTTGATCTTGGTGCAATGT	F
30	ZB3-EMS-R	CGACAATGCTCATGGAGATACTC	R
31	ZB4-sgRNA-F	TTCGACTTCTAGAGTGTTTGGTCG	F
32	ZB4-sgRNA-R	AAACCGACCAAACACTCTAGAAGT	R
33	ZB4-sgRNA-amp-F	CAAATCAACAACGGCTCCATTTATG	F
34	ZB4-sgRNA-amp-R	ATACCTTAGATTCTGCAGTCCAA	R
35	ZB4-sgRNA-seq	TCTGGTGTCTCGTTGCACTT	
36	ZB4-EMS-F	TCTGGTGTCTCGTTGCACTT	F
37	ZB4-EMS-R	ATACCTTAGATTCTGCAGTCCAA	R
38	M13-F	TGTAAAACGACGGCCAGT	F
39	M13-R	CAGGAAACAGCTATGACCATG	R
40	Cas9-F	CAGCTGGTGCAGACCTACAA	F
41	Cas9-R	CGTAGTAGGGGATGCGGAAC	R
42	FH41	AAACGACGGCCAGTGCCAGAATTGGGC CCGACGTCG	
43	FH42	TACTGACTCGTCGGGTACCAAGCTATGC ATCCAACGCG	
44	FH254	GCCCAATTCCAAGCTATGCATCCAACGC G	
45	FH255	CATAGCTTGGAATTGGGCCCGACGTCG	
46	FH179	TATTACTGACTCGTCGGGTA	
47	FLC-qPCR-F	CGGCGATAACCTGGTCAAGA	F
48	FLC-qPCR-R	TCCCACAAGCTTGCTATCCA	R
49	Actin-qPCR-F	TCAGATGCCCAGAAGTCTTGTTCC	F
50	Actin-qPCR-R	CCGTAAGATCCTTCCTGATATCC	R

References

Addo-Quaye, C., Buescher, E., Best, N., Chaikam, V., Baxter, I., and Dilkes, B. P. (2017). Forward genetics by sequencing EMS variation-induced inbred lines. *G3: Genes, Genomes, Genetics, 7,* 413-425.

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

Alvarez, J. M., Brooks, M. D., Swift, J., and Coruzzi, G. M. (2021). Time-based systems biology approaches to capture and model dynamic gene regulatory networks. *Annual Review of Plant Biology, 72*.

Amasino, R. (2010). Seasonal and developmental timing of flowering. *The Plant Journal, 61*, 1001-1013.

Antosz, W., Pfab, A., Ehrnsberger, H. F., Holzinger, P., Köllen, K., Mortensen, S. A., . . . Griesenbeck, J. (2017). The composition of the *Arabidopsis* RNA polymerase II transcript elongation complex reveals the interplay between elongation and mRNA processing factors. *The Plant Cell, 29*, 854-870.

Baumbusch, L. O., Thorstensen, T., Krauss, V., Fischer, A., Naumann, K., Assalkhou, R., . . . Aalen, R. B. (2001). The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Research, 29*, 4319-4333.

Buckner, B., Swaggart, K. A., Wong, C. C., Smith, H. A., Aurand, K. M., Scanlon, M. J., . . . Janick-Buckner, D. (2008). Expression and nucleotide diversity of the maize RIK gene. *Journal of Heredity, 99*, 407-416.

Cheng, Y., Kato, N., Wang, W., Li, J., and Chen, X. (2003). Two RNA binding proteins, HEN4 and HUA1, act in the processing of AGAMOUS pre-mRNA in *Arabidopsis thaliana*. *Developmental Cell, 4*, 53-66.

Csizmok, V., Follis, A. V., Kriwacki, R. W., and Forman-Kay, J. D. (2016). Dynamic protein interaction networks and new structural paradigms in signaling. *Chemical Reviews, 116*, 6424-6462.

Cui, X., Fan, B., Scholz, J., and Chen, Z. (2007). Roles of *Arabidopsis* cyclin-dependent kinase C complexes in cauliflower mosaic virus infection, plant growth, and development. *The Plant Cell, 19*, 1388-1402.

Dai, G.-Y., Chen, D.-K., Sun, Y.-P., Liang, W.-Y., Liu, Y., Huang, L.-Q., . . . Yao, N. (2020). The *Arabidopsis* KH-domain protein FLOWERING LOCUS Y delays flowering by upregulating FLOWERING LOCUS C family members. *Plant Cell Reports, 39*, 1705-1717.

Debaize, L., and Troadec, M.-B. (2019). The master regulator FUBP1: its emerging role in normal cell function and malignant development. *Cellular and Molecular Life Sciences, 76*, 259-281.

Dickinson, P. J., Kneřová, J., Szecowka, M., Stevenson, S. R., Burgess, S. J., Mulvey, H., . . . Hibberd, J. M. (2020). A bipartite transcription factor module controlling expression in the bundle sheath of *Arabidopsis thaliana*. *Nature Plants, 6*, 1468-1479.

Döring, F., Billakurthi, K., Gowik, U., Sultmanis, S., Khoshravesh, R., Das Gupta, S., . . . Westhoff, P. (2019). Reporter-based forward genetic screen to identify bundle sheath anatomy mutants in A. thaliana. *The Plant Journal, 97*, 984-995.

Duncan, R., Bazar, L., Michelotti, G., Tomonaga, T., Krutzsch, H., Avigan, M., and Levens, D. (1994). A sequence-specific, single-strand binding protein activates the far upstream element of c-myc and defines a new DNA-binding motif. *Genes and Development, 8*, 465-480.

Emmerling, J. (2019). Studies Into the Regulation of C_4 Photosynthesis-Towards Factors Controlling Bundle Sheath Expression and Kranz Anatomy Development. *Universitäts-und Landesbibliothek der Heinrich-Heine-Universität Düsseldorf*.

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

Ermakova, M., Danila, F. R., Furbank, R. T., and von Caemmerer, S. (2020). On the road to C_4 rice: advances and perspectives. *The Plant Journal, 101*, 940-950.

Espina, M. J., Ahmed, C., Bernardini, A., Adeleke, E., Yadegari, Z., Arelli, P., . . . Taheri, A. (2018). Development and phenotypic screening of an ethyl methane sulfonate mutant population in soybean. *Frontiers in Plant Science, 9*, 394.

Garner, D. M., Mure, C. M., Yerramsetty, P., and Berry, J. O. (2001). Kranz anatomy and the C_4 pathway. *eLS*, 1-10.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods, 6*, 343-345.

Gowik, U., and Westhoff, P. (2011). The path from C3 to C4 photosynthesis. *Plant Physiology, 155*, 56-63.

Gronland, G. R., and Ramos, A. (2017). The devil is in the domain: understanding protein recognition of multiple RNA targets. *Biochemical Society Transactions, 45*, 1305-1311.

Hahn, F., Eisenhut, M., Mantegazza, O., and Weber, A. P. (2017). Generation of targeted knockout mutants in *Arabidopsis thaliana* using CRISPR/Cas9. *Bio-Protocol, 7*, e2384-e2384.

Harlen, K. M., and Churchman, L. S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nature Reviews Molecular Cell Biology, 18*, 263-273.

Henderson, I. R., and Dean, C. (2004). Control of *Arabidopsis* flowering: the chill before the bloom. *Development, 131*, 3829-3838.

Hudson, W. H., and Ortlund, E. A. (2014). The structure, function and evolution of proteins that bind DNA and RNA. *Nature Reviews Molecular Cell Biology, 15*, 749-760.

Kalve, S., De Vos, D., and Beemster, G. T. (2014). Leaf development: a cellular perspective. *Frontiers in Plant Science, 5*, 362.

Karlsson, P., Christie, M. D., Seymour, D. K., Wang, H., Wang, X., Hagmann, J., . . . Manavella, P. A. (2015). KH domain protein RCF3 is a tissue-biased regulator of the plant miRNA biogenesis cofactor HYL1. *Proceedings of the National Academy of Sciences, 112*, 14096-14101.

Kirschner, S., Woodfield, H., Prusko, K., Koczor, M., Gowik, U., Hibberd, J. M., and Westhoff, P. (2018). Expression of SULTR2; 2, encoding a low-affinity sulphur transporter, in the *Arabidopsis* bundle sheath and vein cells is mediated by a positive regulator. *Journal of Experimental Botany, 69*, 4897-4906.

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. *The Plant Journal, 88*, 1058-1070.

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

Li, S., Liu, L., Li, S., Gao, L., Zhao, Y., Kim, Y. J., and Chen, X. (2016). SUVH1, a Su (var) 3–9 family member, promotes the expression of genes targeted by DNA methylation. *Nucleic Acids Research, 44*, 608-620.

Lorković, Z. J., and Barta, A. (2002). Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant *Arabidopsis thaliana*. *Nucleic Acids Research, 30*, 623-635.

Lundgren, M. R., Osborne, C. P., and Christin, P. A. (2014). Deconstructing Kranz anatomy to understand C4 evolution. *Journal of Experimental Botany, 65*, 3357-3369.

Mackereth, C. D., and Sattler, M. (2012). Dynamics in multi-domain protein recognition of RNA. *Current Opinion in Structural Biology, 22*, 287-296.

Menges, M., Hennig, L., Gruissem, W., and Murray, J. A. (2002). Cell cycle-regulated gene expression in*Arabidopsis*. *Journal of Biological Chemistry, 277*, 41987-42002.

Millard, P. S., Kragelund, B. B., and Burow, M. (2019). R2R3 MYB transcription factors-functions outside the DNA-binding domain. *Trends in Plant Science, 24*, 934-946.

Mitchell, P., and Sheehy, J. E. (2006). Supercharging rice photosynthesis to increase yield. *New Phytologist, 171*, 688-693.

Nicastro, G., Taylor, I. A., and Ramos, A. (2015). KH–RNA interactions: back in the groove. *Current Opinion in Structural Biology, 30*, 63-70.

Ottoz, D. S., and Berchowitz, L. E. (2020). The role of disorder in RNA binding affinity and specificity. *Open Biology, 10*, 200328.

Page, D. R., and Grossniklaus, U. (2002). The art and design of genetic screens: *Arabidopsis thaliana. Nature Reviews Genetics, 3*, 124-136.

Pawson, T., and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. *Science, 300*, 445-452.

Qu, L.-J., and Qin, G. (2014). Generation and identification of *Arabidopsis* EMS mutants. *Methods in Molecular Biology, 1062,* 225-239

Sage, R. F. (2004). The evolution of C4 photosynthesis. *New Phytologist, 161*, 341-370.

Sage, R. F., Khoshravesh, R., and Sage, T. L. (2014). From proto-Kranz to C4 Kranz: building the bridge to C4 photosynthesis. *Journal of Experimental Botany, 65*, 3341-3356.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A. H., Nielsen, K. L., . . . Andersen, S. U. (2009). SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nature Methods, 6*, 550-551.

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

Sedelnikova, O. V., Hughes, T. E., and Langdale, J. A. (2018). Understanding the genetic basis of C4 Kranz anatomy with a view to engineering C3 crops. *Annual Review of Genetics, 52*, 249-270.

Siomi, H., Matunis, M. J., Michael, W. M., and Dreyfuss, G. (1993). The pre-mRNA binding K protein contains a novel evolutionary conserved motif. *Nucleic Acids Research, 21*, 1193-1198.

Street, N. R., Sjödin, A., Bylesjö, M., Gustafsson, P., Trygg, J., and Jansson, S. (2008). A cross-species transcriptomics approach to identify genes involved in leaf development. *BMC Genomics, 9*, 1-18.

Tsukaya, H. (2013). Leaf development. *The Arabidopsis Book/American Society of Plant Biologists, 11*.

Valverde, R., Edwards, L., and Regan, L. (2008). Structure and function of KH domains. *The FEBS Journal, 275*, 2712-2726.

Van Rooijen, R., Schulze, S., Petzsch, P., and Westhoff, P. (2020). Targeted misexpression of NAC052, acting in H3K4 demethylation, alters leaf morphological and anatomical traits in *Arabidopsis thaliana*. *Journal of Experimental Botany, 71*, 1434-1448.

Vercruysse, J., Baekelandt, A., Gonzalez, N., and Inzé, D. (2020). Molecular networks regulating cell division during *Arabidopsis* leaf growth. *Journal of experimental botany, 71*(8), 2365-2378.

Wang, P., Khoshravesh, R., Karki, S., Tapia, R., Balahadia, C. P., Bandyopadhyay, A., . . . Langdale, J. A. (2017). Re-creation of a key step in the evolutionary switch from C3 to C4 leaf anatomy. *Current Biology, 27*, 3278-3287.

Wang, Z.-W., Wu, Z., Raitskin, O., Sun, Q., and Dean, C. (2014). Antisense-mediated FLC transcriptional repression requires the P-TEFb transcription elongation factor. *Proceedings of the National Academy of Sciences, 111*, 7468-7473.

Westhoff, P., and Gowik, U. (2010). Evolution of C4 photosynthesis—looking for the master switch. *Plant Physiology, 154*, 598-601.

Whittaker, C., and Dean, C. (2017). The FLC locus: a platform for discoveries in epigenetics and adaptation. *Annual Review of Cell and Developmental Biology, 33*, 555-575.

Wilson-Sánchez, D., Lup, S. D., Sarmiento-Mañús, R., Ponce, M. R., and Micol, J. L. (2019). Next-generation forward genetic screens: using simulated data to improve the design of mapping-by-sequencing experiments in *Arabidopsis*. *Nucleic Acids Research, 47*, e140-e140.

Wiludda, C., Schulze, S., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., . . . Westhoff, P. (2012). Regulation of the photorespiratory GLDPA gene in C4 Flaveria: an intricate interplay of transcriptional and posttranscriptional processes. *The Plant Cell, 24*, 137-151.

Wu, Z., Fang, X., Zhu, D., and Dean, C. (2020). Autonomous pathway: FLOWERING LOCUS C repression through an antisense-mediated chromatin-silencing mechanism. *Plant Physiology, 182*, 27-37.

Xie, Z., Lee, E., Lucas, J. R., Morohashi, K., Li, D., Murray, J. A., . . . Grotewold, E. (2010). Regulation of cell proliferation in the stomatal lineage by the *Arabidopsis* MYB FOUR LIPS via direct targeting of core cell cycle genes. *The Plant Cell, 22*, 2306-2321.

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

Zhao, L., Li, Y., Xie, Q., and Wu, Y. (2017). Loss of CDKC; 2 increases both cell division and drought tolerance in *Arabidopsis thaliana*. *The Plant Journal, 91*, 816-828.

Zhao, Q. Q., Lin, R. N., Li, L., Chen, S., and He, X. J. (2019). A methylated-DNA-binding complex required for plant development mediates transcriptional activation of promoter methylated genes. *Journal of Integrative Plant Biology, 61*, 120-139.

Acknowledgement

First, I would like to thank **Prof**. **Dr. Peter Westhoff** for providing me an opportunity to conduct PhD research in his group and for his supervision and critical review of my thesis. Most importantly, for his encouragement and constant support in every single step during this journey.

I want to thank **Prof. Dr. Maria von Korff Schmising** for her invaluable time in reviewing this thesis.

My profound gratitude goes to **Dr. Udo Gowik (late)** for helping me in preparing and analysing the transcriptome data.

Many thanks to **Dr. Florian Döring** for providing EMS mutant lines and **Dr. Shipon Das Gupta** for carefully reading the manuscript-II and for the valuable comment and suggestions.

Thanks to the **German Academic Exchange Service (DAAD)** and **Schlumberger Foundation,** this work would not be complete without their financial support.

Special thanks to **Maria Koczor, Susanne Paradies, Dr**. **Elena Pestova**, **Dr. Christian Wiludda**, **Dr. Christian Wever** and **Jasmine Mensing** for sharing a memorable moment in the office.

I want to acknowledge all members of Botanik-IV**,** including **Dr. Kumari Billakurthi, Dr. Karin Ernst, Monika Straubel, Dr. Stefanie Schulze**, for a friendly environment and appropriate working atmosphere.

I want to express my gratitude to my **parents** and **siblings** for their unconditional love and invaluable support in achieving my goal and to my **in-laws** for their constant support and belief.

Finally, huge thanks to my beloved husband **Dr. Ramzan Ali,** for his encouragement, believing in me, and continuous support. Love to my daughters **Mahin Ali** and **Maya Ali** for giving me a tough time during experimental work and thesis writeup, respectively.